The Infinity Project

A 2009–2011 Research Programme

Edited by

Sy-David Friedman

Martin Koerwien

Moritz Müller

Kurt Gödel Research Center for Mathematical Logic

Universität Wien

CENTRE DE RECERCA MATEMÀTICA

Centre de Recerca Matemàtica
Campus de Bellaterra, Edifici C
08193 Bellaterra (Barcelona)

First edition: December 2012

ISSN: 2014-2323 (printed edition)
ISSN: 2014-2331 (electronic edition)

ISBN: 978-84-616-3307-4

Legal deposit:

Foreword

The *Infinity Project* was hosted at the Centre de Recerca Matemàtica (CRM) during 2009–2011 with a specific focus on interdisciplinarity within the field of mathematical logic. Support for the project was provided by John Templeton Foundation Collaborative Research Project #13152, *The Myriad Aspects of Infinity*, and by the CRM. At the July 2011 *Infinity Conference*, the project participants and other logicians surveyed the achievements of the project and pointed the way toward future interdisciplinary work. Support for the Infinity Conference was provided by the CRM, the Spanish government (through its *Ingenio Mathematica* programme), the European Science Foundation (through its Research Networking Programme *New Frontiers of Infinity*) and the Association for Symbolic Logic.

The research of the project was carried out in the form of one or two week collaborations among the 21 project visitors (coming from the fields of computation theory, model theory, proof theory, philosophy and history of set theory), the Barcelona logicians, the two project postdocs Martin Koerwien (model theory) and Moritz Müller (computational complexity theory), and the project leader Sy-David Friedman of the Kurt Gödel Research Center, Vienna. In addition, 17 lectures were presented at the *Infinity Project Seminar* at the time of these collaborations to elaborate on the work of the project.

Twenty years ago, the principal subfields of mathematical logic (computation theory, model theory, proof theory and set theory) were closely related, sharing many ideas and results. Then, due to rapid developments of these subfields, connections between them were lost. As more recent work in mathematical logic suggested that a concerted effort to bring interdisciplinarity back into the field could be of great value, the *Infinity Project* was launched with this specific goal in mind.

The project was a big success. We organized our work around six interdisciplinary themes:

CM	Computations and Models
CP	Computations and Proofs
CS	Computations and Sets
MS	Models and Sets
PS	Proofs and Sets
HPS	History and Philosophy of Set Theory

In each of these areas we found exciting new results, contained in the resulting 37 articles included in this volume, and initiated fruitful collaborations between researchers, some of whom never worked together before. Here are some of the highlights of this work:

1. *Set-Theoretic Proof Theory.* We formulated proof-theoretic results concerning provably recursive functions and formal consistency in analogy to results in set theory. Specifically, in PS1 and CP11 we establish results about provably recursive functions in arithmetic in analogy to the construction of models of set theory using forcing, and in PS2, PS4 and CP9 we study consistency in proof theory by means of an analogue of the set-theoretic concept of jump operator.

2. *Computational Complexity in Set Theory.* We generalised computational complexity theory from the finite to the infinite. Specifically, in CS2 we define functions on sets which are computable in polynomial time.

3. *Higher Descriptive Model Theory.* We developed a generalised descriptive set theory and related it to the model-theoretic classification of theories. Specifically, in MS3, MS5 and MS9 we study Borel reducibility for equivalence relations on the generalised Baire space and show that, for a first-order theory, the complexity of isomorphism for its uncountable models correlates well with its model-theoretic complexity.

4. *Philosophical Implications of Absoluteness.* We explored the significance of absoluteness principles for the foundations of set theory from a philosophical perspective. Specifically, in HPS1 we argue that the Inner Model Hypothesis, a particular absoluteness principle, undermines the claim that Gödel maximality entails the existence of large cardinals.

5. *Descriptive Set Theory and Computational Complexity.* We adapted concepts from the descriptive set theory of equivalence relations to the finite, with implications for computational complexity theory. Specifically, in CS3 we develop finitary analogues of Borel reducibility and relate them to known computational complexity classes.

6. *Descriptive Set Theory and Computable Model Theory.* We developed analogues of the descriptive set theory of equivalence relations for both countable and uncountable computable models. Specifically, in CM2, CM3, CM4 and CS4 we study isomorphism relations on classes of computable models defined by computable infinitary sentences and show that, unlike in the set-theoretic context, there is no nontrivial bound on their complexity.

7. *The Set-Theoretical Foundation for Model Theory.* We explored the effect of different axioms of set theory on the fundamental concepts of model theory. Specifically, in MS1, MS2, MS4, MS6, MS7 and MS8 we study the effects of forcing axioms and the GCH on model existence and uniqueness for infinitary model theory and the model theory of abstract elementary classes.

8. *Forcing and Computational Complexity.* We developed a version of forcing for bounded arithmetic which can be used to attack questions in computational complexity. Specifically, CS1 presents a wide spectrum of such forcing notions, with numerous applications.

9. *Computations and Proofs.* We related the study of algorithms to properties of proof systems. Specifically, CP1 studies the extraction of algorithms from proofs, and CP4–CP8 are concerned with optimal algorithms and their relationship to optimal proof systems.

The unusually lively Infinity Conference, which served as the culmination to the project, was comprised of 29 widely-accessible lectures which left the participants with an appreciation of the potential of interdisciplinary work in mathematical logic, as well as a strong sense of enthusiasm for a continuation of the Infinity Project. Plans for such a continuation are currently in the making.

We wish to offer our deep thanks to Hyung Choi of the Templeton Foundation for his enouragement and support, and most of all to CRM Director Joaquim Bruna together with the CRM staff (especially Neus Portet and Consol Roca) for their gracious hospitality and constant willingness to help with the many needs of this project.

Sy-David Friedman
Martin Koerwien
Moritz Müller
March 2012

Participants

Invited Research Visitors

Tatiana Arrigoni	Fondazione Bruno Kessler, Trento
John T. Baldwin	University of Illinois at Chicago
Arnold Beckmann	Swansea University
Samuel R. Buss	University of California at San Diego
Yijia Chen	Shanghai Jiao Tong University
Fred Drueck	University of Illinois at Chicago
Jörg Flum	Universität Freiburg
Ekaterina Fokina	Kurt Gödel Research Center, Vienna
Loren Graham	Harvard University
Rami Grossberg	Carnegie Mellon University
Tapani Hyttinen	University of Helsinki
Jean-Michel Kantor	Université de Paris VII
Julia F. Knight	University of Notre Dame
Lars Kristiansen	University of Oslo
Vadim Kulikov	University of Helsinki
Russell Miller	Queens College, New York
Antonio Montalbán	University of Chicago
Michael Rathjen	University of Leeds
Andrés Villaveces	Universidad Nacional de Colombia
Albert Visser	Universiteit Utrecht
Andreas Weiermann	Universiteit Gent

Infinity Project Postdocs

Martin Koerwien	Centre de Recerca Matemàtica, Barcelona
Moritz Müller	Centre de Recerca Matemàtica, Barcelona

Barcelona Logicians

Albert Atserias	Universitat Politècnica de Catalunya
Joan Bagaria	ICREA – Universitat de Barcelona
María Luisa Bonet	Universitat Politècnica de Catalunya
Enrique Casanovas	Universitat de Barcelona
Rafel Farré	Universitat Politècnica de Catalunya
Ignasi Jané	Universitat de Barcelona
Juan Carlos Martínez	Universitat de Barcelona

Project Leader

Sy-David Friedman	Kurt Gödel Research Center, Vienna

Infinity Project Seminar

Centre de Recerca Matemàtica
Room: C1/028
Campus de Bellaterra, Edifici C
08193 Bellaterra (Barcelona)

(1) September 23, 2009

Jean-Michel Kantor, Institut de Mathématiques de Jussieu, Paris

A survey of Naming in mathematics and philosophy: the case of the birth of descriptive set theory, later developments and current perspectives

(2) September 30, 2009

Loren Graham, Harvard University

The power of names

(3) December 3, 2009

Jörg Flum, Universität Freiburg

Optimal proof systems and finite model theory

(4) December 10, 2009

Sam Buss, University of California at San Diego

Experiments with a SAT solver

(5) January 13, 2010

Michael Rathjen, University of Leeds

Relativised ordinal analysis

(6) January 20, 2010

Andreas Weiermann, Universiteit Gent

Phase transitions in proof theory

(7) June 1, 2010

Tapani Hyttinen, University of Helsinki

Borel sets on uncountable cardinals and classification theory

(8) June 9, 2010 (Double session)

John Baldwin, University of Illinois at Chicago

Exploring Cantor's paradise: model theory and set theory

Tatiana Arrigoni, Fondazione Bruno Kessler, Trento

On the epistemological status of (infinite) sets

(9) September 23, 2010

John Baldwin, University of Illinois at Chicago
Martin Koerwien, Centre de Recerca Matemàtica, Barcelona

The Infinity Project theme "Sets and Models": a progress report

(10) February 15, 2011

Arnold Beckmann, Swansea University

Proof notations and definable search problems

(11) February 23, 2011

Sam Buss, University of California at San Diego

Time-space tradeoffs and lower bounds for satisfiability

(12) May 12, 2011

Albert Visser, Universiteit Utrecht

Consistency statements and the Wilkie hierarchy

(13) June 1, 2011

Julia Knight, University of Notre Dame

Computable structure theory in the setting of ω_1

(14) June 7, 2011

Russell Miller, City University of New York

Fields and computable categoricity

(15) June 9, 2011

Antonio Montalbán, University of Chicago

A fixed point for the jump operator on structures

Infinity Conference

Dates

July 18–22, 2011

Scientific Committee

John T. Baldwin	University of Illinois at Chicago
María Luisa Bonet	Universitat Politècnica de Catalunya
Sy-David Friedman	Kurt Gödel Research Center, Vienna
Juan Carlos Martínez	Universitat de Barcelona
Michael Rathjen	University of Leeds

Program

CM	Computations and Models
CP	Computations and Proofs
CS	Computations and Sets
MS	Models and Sets
PS	Proofs and Sets
HPS	History and Philosophy of Set Theory

Monday, July 18

09:30–09:50	Sy-David Friedman Universität Wien	Opening remarks	
09:50–10:40	Andreas Weiermann Universiteit Gent	Degree theory for provably recursive functions and unprovability phase transitions	PS
11:00–11:50	Andrés Villaveces Universidad Nacional de Colombia	Categoricity and amalgamation at low cardinalities: weak diamonds versus forcing	MS
12:00–12:50	Antonio Montalbán University of Chicago	The boundary of determinacy within second order arithmetic	CS
15:00–15:50	Arnold Beckmann Swansea University	Safe recursive set functions	CS
16:00–16:50	Colin McLarty Case Western Reserve University	Grothendieck's reflection principle: number theory with a set that the operations of set theory do not go beyond	HPS
17:10–18:00	Tapani Hyttinen Helsinki University	Constructing groups and fields from a geometry	MS
18:10–18:40	Vadim Kulikov Helsinki University	Borel equivalence relations on 2^κ, $\kappa > \omega$	

Tuesday, July 19

09:30–10:20	Jörg Flum Universität Freiburg	The myriad applications of a halting problem	CS
10:40–11:30	John Baldwin University of Illinois at Chicago	Calculating Hanf numbers	MS
11:50–12:40	Julia Knight University of Notre Dame	Comparing classes of countable structures	CM
15:00–15:50	Lars Kristiansen University of Oslo	Subrecursive degrees of honest functions and provably recursive functions	CP
16:00–16:30	Alexander Gavryushkin Irkutsk State University	Finite structures, Fraïssé limits, Ehrenfeucht theories. Computability aspects	CM
16:40–17:10	Bing Kai Lin Shanghai Jiao Tong University	The parametrised complexity of k-edge induced subgraphs	
17:30–18:00	Denis Saveliev Moscow State University	Ultrafilters without choice	
18:10–18:40	Gunnar Wilken Okinawa University	Infinitary concepts and Gödel's T	

Wednesday, July 20

09:30–10:20	Martin Koerwien CRM	Absoluteness considerations in $L_{\omega_1\omega}$	MS
10:40–11:30	Michael Rathjen University of Leeds	Ordinal analysis for powerset and the existence property	PS
11:50–12:40	Moritz Müller CRM	Partially definable forcing and bounded arithmetic	CS
Guided Barcelona visit followed by conference dinner			

Thursday, July 21

09:30–10:20	Philip Welch Bristol University	Transfinite machines, analysis and determinacy	CS
10:40–11:30	Ignasi Jané Universitat de Barcelona	On Cantor's account of the distinction between sets and inconsistent multiplicities	HPS
11:50–12:40	Russell Miller City University of New York	Local computability and uncountable structures	CM
15:00–15:50	Sam Buss University of California at San Diego	Towards NP-P via proof complexity and search	CP
16:00–16:30	Jesse Johnson University of Notre Dame	Computable categoricity for uncountable structures	
16:40–17:10	Michael Lieberman University of Pennsylvania	Category-theoretic foundations of abstract model theory	
17:30–18:00	Robert Lubarsky Florida Atlantic University	Weak weak Koenig's lemma does not imply decidable fan	
18:10–18:40	Stefan Vatev Sofia University	Conservative extensions of abstract structures	

Friday, July 22

09:30–10:20	Joan Bagaria ICREA – Universitat de Barcelona	Structural complexity, reflection, and topologies on ordinals	HPS
10:40–11:30	Tatiana Arrigoni Fondazione Bruno Kessler	Sy Friedman's inner model hypothesis. Philosophical and foundational reflections	HPS
11:50–12:40	Yijia Chen Shanghai Jiao Tong University	Consistency, incompleteness, and optimality	CP

Conference Participants

Adler, Hans	Kurt Gödel Research Center, Vienna
Arrigoni, Tatiana	Fondazione Bruno Kessler, Trento
Bagaria, Joan	ICREA – Universitat de Barcelona
Baldwin, John T.	University of Illinois at Chicago
Beckmann, Arnold	Swansea University
Blasco, José María	Universitat de Barcelona
Bonet, María Luisa	Universitat Politècnica de Catalunya
Boney, Will	Carnegie Mellon University
Buss, Samuel R.	University of California at San Diego
Castells, Neus	Universitat de Barcelona
Chen, Yijia	Shanghai Jiao Tong University
Flum, Jörg	Universität Freiburg
Friedman, Sy-David	Kurt Gödel Research Center, Vienna
García Avila, Luz María	Universitat de Barcelona
Gavryushkin, Alexander	Irkutsk State University
Gavryushkina, Alexandra	Irkutsk State University
Harizanov, Valentina	George Washington University
Hyttinen, Tapani	University of Helsinki
Jané, Ignasi	Universitat de Barcelona
Johnson, Jesse	University of Notre Dame
Knight, Julia F.	University of Notre Dame
Koerwien, Martin	CRM, Barcelona
Kristiansen, Lars	University of Oslo
Kulikov, Vadim	University of Helsinki
Lieberman, Michael	University of Pennsylvania
Lin, Bing Kai	Shanghai Jiao Tong University
Lubarsky, Robert	Florida Atlantic University
Martínez, Juan Carlos	Universitat de Barcelona
McLarty, Colin S.	Case Western Reserve University
Miller, Russell Geddes	City University of New York
Montalbán, Antonio	University of Chicago
Mota, Miguel Ángel	Kurt Gödel Research Center, Vienna
Müller, Moritz	CRM, Barcelona
Müller, Sebastian	Charles University, Prague
Nabutovsky, Alexander	University of Toronto
Nemoto, Takako	Universität Bern
Phillips, Youyu	Keystone College
Rathjen, Michael	University of Leeds
Rittberg, Colin Jakob	Università degli studi di Palermo
Rivello, Edoardo	Università degli studi di Torino
Sato, Kentaro	Universität Bern
Saveliev, Denis	Moscow State University
Schupp, Paul	University of Illinois at Urbana
Tonti, Fabio	Universität Wien

Vatev, Stefan	Sofia University
Villaveces, Andrés	Universidad Nacional de Colombia
Weiermann, Andreas	Universiteit Gent
Welch, Philip D.	Bristol University
Wilken, Gunnar	Okinawa Institute of Science and Technology

Table of Contents

Preface

Foreword . . . i

Participants . . . iv

Infinity Project Seminar . . . v

Infinity Conference . . . vii

Table of Contents . . . xiii

Computations and Models

CM1. Hubie Chen and Moritz Müller . . . 3
An algebraic preservation theorem for $\aleph_0$-categorical quantified constraint satisfaction

CM2. Ekaterina Fokina, Sy-David Friedman, Valentina Harizanov, Julia F. Knight, Charles McCoy and Antonio Montalbán . . . 27
Isomorphism relations on computable structures

CM3. Ekaterina Fokina, Sy-David Friedman, Julia F. Knight, Russell Miller and Antonio Montalbán . . . 39
Classes of structures with universe a subset of ω_1

CM4. Sy-David Friedman . . . 51
Equivalence relations in set theory, computation theory, model theory and complexity theory

CM5. Alexander Gavryushkin . . . 67
Computable models of Ehrenfeucht theories

Computations and Proofs

CP1. Arnold Beckman and Samuel R. Buss . . . 81
Improved witnessing and local improvement principles for second-order bounded arithmetic

CP2. María Luisa Bonet and Samuel R. Buss . . . 115
An improved separation of regular resolution from pool resolution and clause learning

CP3. Samuel R. Buss . . . 133
Sharpened lower bounds for cut elimination

CP4. Yijia Chen and Jörg Flum . . . 145
From almost optimal algorithms to logics for complexity classes via listings and a halting problem

CP5. Yijia Chen and Jörg Flum . . . 179
On p-optimal proof systems and logics for PTIME

CP6. Yijia Chen and Jörg Flum 191
On slicewise monotone parametrized problems and optimal proof systems for TAUT

CP7. Yijia Chen, Jörg Flum and Moritz Müller 205
Hard instances of algorithms and proof systems

CP8. Yijia Chen, Jörg Flum and Moritz Müller 225
On optimal probabilistic algorithms for SAT

CP9. Yijia Chen, Jörg Flum and Moritz Müller 231
Consistency, optimality, and incompleteness

CP10. Jörg Flum and Moritz Müller 243
Some definitorial suggestions for parameterized proof complexity

CP11. Lars Kristiansen, Robert S. Lubarsky, Jan-Christoph Schlage-Puchta and Andreas Weiermann 255
On the structure of honest elementary degrees

Computations and Sets

CS1. Albert Atserias and Moritz Müller 283
Partially definable forcing and bounded arithmetic

CS2. Arnold Beckmann, Samuel R. Buss and Sy-David Friedman 313
Safe recursive set functions

CS3. Samuel R. Buss, Yijia Chen, Jörg Flum, Sy-David Friedman and Moritz Müller 337
Strong isomorphism reductions in complexity theory

CS4. Ekaterina Fokina and Sy-David Friedman 359
On Σ^1_1 equivalence relations over the natural numbers

CS5. Philip D. Welch 375
Transfinite machines, analysis and determinacy

History and Philosophy of Set Theory

HPS1. Tatiana Arrigoni and Sy-David Friedman 383
Foundational implications of the inner model hypothesis

Models and Sets

MS1. John T. Baldwin 395
Amalgamation, absoluteness and categoricity

MS2. John T. Baldwin, Tapani Hyttinen and Meeri Kesälä 417
Beyond first order logic: from number of structures to structure of numbers, parts I and II

MS3. Sy-David Friedman and Tapani Hyttinen 451
On Borel equivalence relations in generalized Baire space

MS4. Sy-David Friedman, Tapani Hyttinen and Martin Koerwien 457
Non-absoluteness of model existence in uncountable cardinals for $L_{\omega_1,\omega}$

MS5. Sy-David Friedman, Tapani Hyttinen and Vadim Kulikov 471
Generalized descriptive set theory and classification theory

MS6. Sy-David Friedman, Tapani Hyttinen and Agatha C. Walczak-Typke 547
Potential isomorphism of elementary substructures of a strictly stable homogeneous model

MS7. Sy-David Friedman and Martin Koerwien 565
On absoluteness of categoricity in AEC's

MS8. Martin Koerwien and Stevo Todorcevic 573
Two examples concerning $\aleph_1$-categoricity in abstract elementary classes

MS9. Vadim Kulikov 579
Borel reductions on the generalized Cantor space

MS10. Denis Saveliev 599
On ultrafilter extensions of models

Proofs and Sets

PS1. Sy-David Friedman, Michael Rathjen and Andreas Weiermann 619
Some results on PA-provably recursive functions

PS2. Sy-David Friedman, Michael Rathjen and Andreas Weiermann 623
Slow consistency

PS3. Michael Rathjen 637
Relativized ordinal analysis: The case of Power Kripke–Platek set theory

PS4. Vladimir Yu. Shavrukov and Albert Visser 659
Uniform density in Lindenbaum algebras

PS5. Gunnar Wilken and Andreas Weiermann 671
Derivation lengths classification of Gödel's T extending Howard's assignment

Part I

Computations and Models

An algebraic preservation theorem for $\aleph_0$-categorical quantified constraint satisfaction

Hubie Chen*, **Moritz Müller**†

* Universitat Pompeu Fabra, Barcelona, Spain
hubie.chen@upf.edu

† Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
moritz.mueller@univie.ac.at

Abstract. We prove a preservation theorem for positive Horn definability in $\aleph_0$-categorical structures. In particular, we define and study a construction which we call the *periodic power* of a structure, and define a *periomorphism* of a structure to be a homomorphism from the periodic power of the structure to the structure itself. Our preservation theorem states that, over an $\aleph_0$-categorical structure, a relation is positive Horn definable if and only if it is preserved by all periomorphisms of the structure. We give applications of this theorem, including a new proof of the known complexity classification of quantified constraint satisfaction on equality templates.

Introduction

Model checking —deciding if a logical sentence holds on a structure— is a basic computational problem which is in general intractable; for example, model checking first-order sentences on finite structures is well-known to be PSPACE-complete. In the context of model checking, fragments of first-order logic based on restricting the connectives $\{\wedge, \vee, \neg\}$ and quantifiers $\{\exists, \forall\}$ have been considered in a variety of settings. For instance, the problem of model checking *primitive positive* sentences, sentences formed using $\{\wedge, \exists\}$, is an NP-complete problem that is a formulation of the *constraint satisfaction problem* (CSP), and admits a number of other natural characterizations, as shown in the classical work of Chandra and Merlin [**18**]. The problem of model checking *positive Horn* sentences, sentences formed using $\{\wedge, \exists, \forall\}$, is known as the *quantified constraint satisfaction problem* (QCSP), and is PSPACE-complete; indeed, certain cases of this problem are canonical complete problems for PSPACE [**42**, Chapter 19]. Another natural fragment consists of the *existential positive* sentences, which are formed from $\{\wedge, \vee, \exists\}$.

Such syntactically restricted fragments of first-order logic can be naturally parameterized by the structure [**41**]. As examples, consider the following problems for a structure $\mathfrak{A}$:

- $\mathsf{CSP}(\mathfrak{A})$: decide the primitive positive theory of $\mathfrak{A}$.
- $\mathsf{QCSP}(\mathfrak{A})$: decide the positive Horn theory of $\mathfrak{A}$.
- $\mathsf{EXPOS}(\mathfrak{A})$: decide the existential positive theory of $\mathfrak{A}$.
- $\mathsf{EFPOS}(\mathfrak{A})$: decide the equality-free positive theory of $\mathfrak{A}$.

Via this parameterization, one obtains four *families* of problems, and is prompted with classification programs: for each of the families, classify the problems therein according to their computational complexity. On finite structures, comprehensive classifications are known for the families $\mathsf{EXPOS}(\mathfrak{A})$ and $\mathsf{EFPOS}(\mathfrak{A})$. Each problem $\mathsf{EXPOS}(\mathfrak{A})$ is either in L or NP-complete [**9**], and each problem $\mathsf{EFPOS}(\mathfrak{A})$ is either in L, NP-complete, coNP-complete, or PSPACE-complete [**40**]. Moreover, each of these two classifications is effective in that for each there exists an algorithm that, given a finite structure, tells what the complexity of the corresponding problem is. For the family of problems $\mathsf{CSP}(\mathfrak{A})$, Feder and Vardi [**26**] famously conjectured that there is a dichotomy in the finite: for each finite structure $\mathfrak{A}$, the problem $\mathsf{CSP}(\mathfrak{A})$ is either polynomial-time tractable or NP-complete. Investigation of the complexity-theoretic properties of the problem families $\mathsf{CSP}(\mathfrak{A})$ and $\mathsf{QCSP}(\mathfrak{A})$ on finite structures is a research theme of active interest [**1, 2, 15, 20, 22, 31, 36**].

At the heart of the work on these classification programs are *algebraic preservation theorems* which state that, relative to a finite structure, the relations definable in a given fragment are precisely those preserved by a suitable set of operations. As an example, one such theorem states that a relation is primitive positive definable on a finite structure $\mathfrak{A}$ if and only if all polymorphisms of $\mathfrak{A}$ are polymorphisms of the relation [**14, 29**]. (A polymorphism of a structure $\mathfrak{A}$ is a homomorphism from a finite power $\mathfrak{A}^k$ to $\mathfrak{A}$ itself.) On finite structures there are analogous preservation theorems connecting positive Horn definability to surjective polymorphisms [**15**], existential positive definability to endomorphisms [**35**], and equality-free positive definability to so-called surjective hyper-endomorphisms [**39**]. For the purposes of complexity classification, these preservation theorems are relevant in that they allow one to pass from the study of structures to the study of algebraic objects. For instance, it follows from the preservation theorem for primitive positive definability that two finite structures $\mathfrak{A}, \mathfrak{B}$ having the same polymorphisms are primitive positively interdefinable, from which it readily follows that the problems $\mathsf{CSP}(\mathfrak{A})$ and $\mathsf{CSP}(\mathfrak{B})$ are interreducible and share the same complexity (under many-one logspace reduction); thus, insofar as one is interested in CSP complexity, one can focus on investigating the polymorphisms of structures.

Given the import and reach of these algebraic preservation theorems for finite structures, a natural consideration is to generalize them to infinite structures. Although it is known that these preservation theorems do not hold on *all* infinite structures (see the discussion in [**10**] as well as [**4**, Theorem 4.7]), Bodirsky and Nešetřil [**13**, Theorem 5.1] established that the preservation theorem characterizing primitive positive definability via polymorphisms does hold on $\aleph_0$-categorical structures, which have countably infinite universes. An $\aleph_0$-categorical structure is "finite-like", in that, for each fixed arity, there are a finite number of first-order definable relations; indeed, this is one of the characterizations of $\aleph_0$-categoricity given by the classical theorem of Ryll-Nardzewski. The class of $\aleph_0$-categorical structures includes many structures of computational interest, including those whose relations are first-order definable over one of the following structures: equality on a countable universe, the ordered rationals $(\mathbb{Q}; <)$, and the countable random graph; see [**5**] for a survey.

In this paper, we present an algebraic preservation theorem for positive Horn definability on $\aleph_0$-categorical structures. This theorem characterizes positive Horn definability by making use of a construction which we call the *periodic power*. In particular, we define a *periomorphism* of a structure $\mathfrak{A}$ as a homomorphism from the periodic power of $\mathfrak{A}$ to $\mathfrak{A}$ itself, and show that a relation is positive Horn definable over an $\aleph_0$-categorical structure $\mathfrak{A}$ if and only if all surjective periomorphisms of $\mathfrak{A}$ are periomorphisms of the relation.

The periodic power of a structure $\mathfrak{A}$ is the substructure of $\mathfrak{A}^{\mathbb{N}}$ whose universe is the set of all periodic tuples in $\mathfrak{A}^{\mathbb{N}}$. A tuple $(a_0, a_1, \ldots)$ is *periodic* if there exists an integer $k \geq 1$ such that the tuple *repeats mod k*, by which is meant $a_n = a_{n \bmod k}$ for all $n \in \mathbb{N}$. As we discuss in the paper, the periodic power arises as the direct limit of an appropriately defined system of embeddings. Despite the extremely natural character of the periodic power, we are not aware of previous work where this construction has been explicitly considered. We believe it possible that the periodic power may find applications in other areas of mathematics. Indeed, one basic fact that we demonstrate is that the positive Horn theory of a structure holds in the structure's periodic power; this readily implies that the class of groups is closed under the taking of periodic powers, and likewise one has closure under periodic powers for other classes of classical algebraic structures such as rings, lattices, and Boolean algebras. Our introduction and study of the periodic power also forms a contribution of this paper.

A direct corollary of our preservation theorem is that for two $\aleph_0$-categorical structures $\mathfrak{A}, \mathfrak{B}$ with the same universe $A = B$, if $\mathfrak{A}$ and $\mathfrak{B}$ have the same surjective periomorphisms, then the structures $\mathfrak{A}$ and $\mathfrak{B}$ are positive Horn interdefinable, and the computational problems $\mathsf{QCSP}(\mathfrak{A})$ and $\mathsf{QCSP}(\mathfrak{B})$ are interreducible (under many-one logspace reduction). This permits the use of surjective periomorphisms in the study of the complexity of the QCSP on $\aleph_0$-categorical structures. As an application of our preservation theorem and the associated theory that we develop, we give a new proof of the known complexity classification of *equality templates*, which are structures whose relations are first-order definable over the equality relation on a countable set.

Related work

An algebraic preservation theorem for positive Horn definability via surjective polymorphisms was shown for the special case of equality templates [**6**]. The presented proof crucially depends on results on the clones of equality templates given there and in [**8**].

In model theory, there are *classical preservation theorems* that show that a sentence is equivalent to one in a given fragment if and only if its model class satisfies some suitable closure properties. Such theorems have been shown for positive Horn logic. A well-known instance is Birkhoff's HSP theorem characterizing universally quantified equations. And in 1955, Bing [**3**] showed that a positive sentence is preserved by direct products if and only if it is equivalent to a positive Horn sentence. Later, assuming the continuum hypothesis (CH), Keisler[1] proved that a sentence is equivalent to a positive Horn sentence if and only if it is preserved (in the parlance of [**28, 45**]) by the following binary relation: relate $\mathfrak{A}$ to $\mathfrak{B}$ when $\mathfrak{B}$ is a homomorphic image of $\mathfrak{A}^{\mathbb{N}}$ [**33**, Corollary 3.8] (see also [**19**, Section 6.2]). Absoluteness considerations can be used to eliminate the assumption of CH when one has ZFC provability of the stated closure property. More recently, Madelaine and Martin [**38**, Theorem 1] showed, without relying on CH, that Keisler's result holds when one considers preservation under the relation defined as above, but where $\mathfrak{B}$ is required to be finite.

In some cases, an algebraic preservation theorem can be derived from a corresponding classical preservation theorem. Such a derivation has been given for Bodirsky and Nešetřil's theorem in [**5**], and Bodirsky and Junker [**11**] derived algebraic preservation theorems for existential positive definability and positive definability in $\aleph_0$-categorical structures from well-known classical preservation theorems of Lyndon. Roughly speaking, these methods need the preservation relation to be PC_Δ (cf. [**28**]) and thus cannot be applied to Keisler's classical preservation theorem mentioned above. To the best of our knowledge, prior to this work no algebraic preservation theorem for positive Horn formulas on $\aleph_0$-categorical structures has been known (neither in the presence nor absence of CH).

[1] In fact, Keisler could do assuming only the existence of some cardinal $\kappa \geq \aleph_0$ such that $2^\kappa = \kappa^+$.

1 Basics from model theory

1.1 First-order logic

Throughout the paper, L will denote a countable first-order language. If not explicitly stated otherwise, by a structure (formula) we always mean an L-structure (first-order L-formula). Throughout, we use the letters $\mathfrak{A}$, $\mathfrak{B}$, etc. to denote structures and φ, ψ, χ, etc. to denote formulas. For a structure $\mathfrak{A}$ and a (finite) tuple $\overline{a}$ from A, by $(\mathfrak{A}, \overline{a})$ we denote, as usual, the expansion of $\mathfrak{A}$ interpreting new constants by the components of $\overline{a}$. We do not distinguish between constants outside L and variables. For a formula $\varphi = \varphi(\overline{x})$ and a structure $\mathfrak{A}$, writing $(\mathfrak{A}, \overline{a}) \models \varphi(\overline{x})$ or $\mathfrak{A} \models \varphi(\overline{a})$ (with $\overline{x}$ clear from context) means that $\mathfrak{A}$ satisfies $\varphi(\overline{x})$ under the assignment $\overline{a}$ to $\overline{x}$. By $\varphi(\mathfrak{A})$ we denote the relation $\{\overline{a} \mid \mathfrak{A} \models \varphi(\overline{a})\}$ on A; this relation is said to be *defined by* φ *in* $\mathfrak{A}$. A relation is *first-order (positive Horn, primitive positively) definable in* $\mathfrak{A}$ if it is defined by some first-order (positive Horn, primitive positive) formula φ in $\mathfrak{A}$ (see Section 2 for definitions of positive Horn and primitive positive).

Let L' be another first-order language, $\mathfrak{B}$ an L'-structure and $\mathfrak{A}$ an L-structure such that $A = B$. Then $\mathfrak{B}$ is *first-order (positive Horn, primitive positively) definable in* $\mathfrak{A}$ if for every atomic L'-formula the relation $\varphi(\mathfrak{B})$ is (positive Horn, primitive positively) definable in $\mathfrak{A}$.

1.2 Direct products

For a family of (L-)structures $(\mathfrak{A}_i)_{i\in I}$, we denote its direct product by $\prod_{i\in I} \mathfrak{A}_i$. Recall that this structure

- has universe $\prod_{i\in I} A_i$, which is the set of functions mapping each $i \in I$ into the universe A_i of $\mathfrak{A}_i$;
- interprets a k-ary relation symbol $R \in L$ by those k-tuples $(\vec{a}_0, \ldots, \vec{a}_{k-1})$ from $\prod_{i\in I} A_i$ such that $\mathfrak{A}_i \models R\vec{a}_0(i) \cdots \vec{a}_{k-1}(i)$ for all $i \in I$; and
- interprets a k-ary function symbol $f \in L$ by the function mapping a k-tuple $(\vec{a}_0, \ldots, \vec{a}_{k-1})$ from $\prod_{i\in I} A_i$ to the element $\vec{a} \in \prod_{i\in I} A_i$ such that, for all $i \in I$, $\mathfrak{A}_i \models f(\vec{a}_0(i), \ldots, \vec{a}_{k-1}(i)) = \vec{a}(i)$.

We write $\mathfrak{A}^I$ for $\prod_{i\in I} \mathfrak{A}_i$ with $\mathfrak{A}_i = \mathfrak{A}$ for all i, and we write $\mathfrak{A}^k$ to denote $\mathfrak{A}^I$ when $I = \{0, \ldots, k-1\}$ for $k \in \mathbb{N}$, $k > 0$. We consider $\mathfrak{A}^k$ to have universe A^k, the set of k-tuples over A. We do not distinguish between 1-tuples and elements, so $\mathfrak{A}^1 = \mathfrak{A}$. The direct product of two structures $\mathfrak{A}$ and $\mathfrak{B}$ is denoted by $\mathfrak{A} \times \mathfrak{B}$ and considered to have universe $A \times B$.

1.3 Direct limits

We recall the definitions associated with direct limits. Let $(I, \prec)$ be a directed strict partial order (i.e., every two elements in I have a common upper bound). An $(I, \prec)$-*system of embeddings* is a family of embeddings $e_{(i,j)} \colon \mathfrak{A}_i \to \mathfrak{A}_j$ for $i \prec j$ such that

$$e_{(i,k)} = e_{(j,k)} \circ e_{(i,j)}$$

for all $i \prec j \prec k$. A *cone* of the system is a family of *limit embeddings* $e_i^* \colon \mathfrak{A}_i \to \mathfrak{A}^*$ such that $e_j^* \circ e_{(i,j)} = e_i^*$. It is known that, for a system, there exists a cone satisfying the following universal property: for every other cone, say given by $\tilde{\mathfrak{A}}$ and $(\tilde{e}_i)_{i\in I}$, there exists a unique embedding $e\colon \mathfrak{A}^* \to \tilde{\mathfrak{A}}$ such that $e \circ e_i^* = \tilde{e}_i$. A structure $\mathfrak{A}^*$ with these properties is unique up to isomorphism and called the *direct limit* of the system; if $(I, \prec)$ and the $e_{(i,j)}$s are clear from context, it is denoted $\lim_i \mathfrak{A}_i$.

1.4 $\aleph_0$-categoricity

A structure $\mathfrak{A}$ is $\aleph_0$-*categorical* if it is countable and every countable structure $\mathfrak{B}$ that satisfies the same first-order sentences as $\mathfrak{A}$ is isomorphic to $\mathfrak{A}$. We assume basic familiarity with $\aleph_0$-categoricity as covered by any standard course in model theory (e.g. [**19**]). Here, we briefly recall some facts that we are going to use.

The *Ryll-Nardzewski theorem* states that a countable structure $\mathfrak{A}$ is $\aleph_0$-categorical if and only if for every $k \in \mathbb{N}$ there are at most finitely many k-ary relations that are first-order definable in $\mathfrak{A}$. It is straightforward to verify that this implies that for an $\aleph_0$-categorical structure $\mathfrak{A}$, when $\overline{a}$ is an arbitrary finite-length tuple from A, the structure $(\mathfrak{A}, \overline{a})$ is also $\aleph_0$-categorical. Further, it implies that for an $\aleph_0$-categorical structure $\mathfrak{A}$, the structure $\mathfrak{A}^k$ is $\aleph_0$-categorical for any $k \in \mathbb{N}$; in fact, every structure that is first-order interpretable in an $\aleph_0$-categorical structure is also $\aleph_0$-categorical.

Another easy consequence of this theorem is that $\aleph_0$-categorical structures are $\aleph_0$-*saturated*, by which is meant that for every finite tuple $\overline{a}$ from A and every set of formulas $\Phi = \Phi(x)$ in the language of $(\mathfrak{A}, \overline{a})$ (that is, having constants for $\overline{a}$) one has: if $\Phi(x)$ is finitely satisfiable in $(\mathfrak{A}, \overline{a})$, then it is satisfiable in in $(\mathfrak{A}, \overline{a})$. Here, a set of formulas $\Phi = \Phi(x)$ with one free variable x is *satisfiable* in $\mathfrak{A}$ if there is $a \in A$ such that $(\mathfrak{A}, a) \models \Phi(x)$ (that is, a satisfies every $\varphi(x) \in \Phi(x)$ in $\mathfrak{A}$); Φ is *finitely satisfiable in* $\mathfrak{A}$ if every finite subset of Φ is satisfiable in $\mathfrak{A}$.

Finally, we mention the fact that for an $\aleph_0$-categorical structure $\mathfrak{A}$, a relation over A is first-order definable if and only if it is preserved by all automorphisms of $\mathfrak{A}$ (see Section 2.3 for the definition of preservation).

2 Basics from constraint complexity

2.1 Positive Horn formulas

As noted in the introduction, a *positive Horn* formula is a first-order formula built from atoms, conjunction, and the two quantifiers. Existential such formulas are *primitive positive*. For simplicity, we assume that first-order logic contains a propositional constant $\bot$ for falsehood; formally, $\bot$ is a 0-ary relation symbol always interpreted by $\emptyset$. Note that $\bot$ is a positive atomic sentence. If any positive Horn sentence true in $\mathfrak{A}$ is also true in $\mathfrak{B}$, we write $\mathfrak{A} \Rrightarrow_{\mathrm{pH}} \mathfrak{B}$.

A formula $\varphi(\overline{x})$ is *preserved by direct products* if it holds in $(\mathfrak{A}, \overline{a}) \times (\mathfrak{B}, \overline{b})$ whenever it holds in both $(\mathfrak{A}, \overline{a})$ and $(\mathfrak{B}, \overline{b})$. Positive Horn formulas are preserved by direct products; in fact, the following is straightforward to verify.

Lemma 2.1 *Let $(\mathfrak{A}_i)_{i \in I}$ be a family of structures. A positive Horn sentence holds in $\prod_{i \in I} \mathfrak{A}_i$ if and only if it holds in every $\mathfrak{A}_i$, $i \in I$.*

2.2 Quantified constraints

The quantified constraint satisfaction problem (QCSP) on a structure $\mathfrak{A}$, denoted by $\mathsf{QCSP}(\mathfrak{A})$, is the problem of deciding the positive Horn theory of $\mathfrak{A}$. The following proposition relates positive Horn definability to the complexity of the QCSP.

Proposition 2.2 *Let $\mathfrak{A}$ be an L-structure and $\mathfrak{B}$ be an L_0-structure for some finite first-order language L_0. If $\mathfrak{B}$ is positive Horn definable in $\mathfrak{A}$, then the problem* $\mathsf{QCSP}(\mathfrak{B})$ *many-one logspace reduces to* $\mathsf{QCSP}(\mathfrak{A})$.

Proof. For every function symbol $f \in L_0$, each constant $c \in L_0$ and each relation symbol $R \in L_0$, choose some fixed positive Horn L-formulas $\psi_f(\overline{x}, y), \psi_c(x), \psi_R(\overline{x})$ that respectively define, in $\mathfrak{A}$, the relations given by the formulas $f(\overline{x}) = y$, $x = c$, $R\overline{x}$ interpreted over $\mathfrak{B}$. Let φ be an instance of $\mathsf{QCSP}(\mathfrak{B})$, that is, a positive Horn sentence in the language L_0. In a first step, compute in logspace an equivalent sentence φ^* in which every atomic subformula contains at most one symbol from L_0, that is, has the form $x = y$, $f(\overline{x}) = y$, $x = c$ or $R\overline{x}$. This can be done by successively replacing atomic subformulas of φ; for example, replacing $Rxcf(f(x))$ by

$$\exists y_0 y_1 y_2 (Rxy_0y_2 \wedge y_0 = c \wedge f(y_1) = y_2 \wedge f(x) = y_1).$$

In a second step, replace in φ^* every atomic subformula that mentions $s \in L_0$ by the formula ψ_s (with the right choice of variables). This can also be done in logspace: note that we may hardwire the finite list of the formulas ψ_s in the code of the algorithm. Finally, recall that the composition of two logspace algorithms can be implemented in logspace. □

Remark 2.3 In the literature, the CSP and QCSP are typically defined in *relational* first-order logic. We take a more general stance and allow the language to contain function symbols if not explicitly stated otherwise. In particular, our preservation theorem (Theorem 5.1) holds true in the presence of function symbols.

2.3 Preservation

Let $\mathfrak{A}$ be a structure, I a nonempty set and h a partial function from A^I to A. Then h is said to *preserve* an r-ary relation $R \subseteq A^r$ if it is a partial homomorphism from $(A, R)^I$ to (A, R). This means the following: whenever $\vec{a}_0, \ldots, \vec{a}_{r-1}$ are in the domain of h and $(\vec{a}_0(i), \ldots, \vec{a}_{r-1}(i)) \in R$ for all $i \in I$, then $(h(\vec{a}_0), \ldots, h(\vec{a}_{r-1})) \in R$. Further, we say that h preserves a formula φ if it preserves the relation $\varphi(\mathfrak{A})$. If h is defined on all of A^I (and I is finite), it is called a *(finitary) operation* on A.

2.4 Clones and polymorphisms

A *clone on A* is a set of finitary operations on A that is closed under composition and contains all projections. A set F of operations on A *interpolates* an operation g on A if for all finite sets $B \subseteq A$ there exists an operation $f \in F$ such that $f \upharpoonright B = g \upharpoonright B$. A set of operations is *locally closed* if it contains every operation that it interpolates.

A *polymorphism of* $\mathfrak{A}$ is a homomorphism from $\mathfrak{A}^k$ to $\mathfrak{A}$ for some $k \in \mathbb{N}$, $k > 0$; k is the *arity* of the polymorphism. Equivalently, a polymorphism of $\mathfrak{A}$ is a finitary operation on A that preserves each $\mathfrak{A}$-relation, $\mathfrak{A}$-constant, and graph of an $\mathfrak{A}$-function; or, a polymorphism of $\mathfrak{A}$ is a finitary operation on A that preserves all atomic formulas. It is straightforward to verify that the set of polymorphisms of $\mathfrak{A}$ forms a locally closed clone on A.

An operation $h\colon A^k \to A$ is a polymorphism of a relation $R \subseteq A^\ell$ if h is a polymorphism of the structure (A, R). In a picture, this means the following. If every column of

$$\begin{matrix} a_0^0 & a_1^0 & \cdots & a_{k-1}^0 \\ a_0^1 & a_1^1 & \cdots & a_{k-1}^1 \\ \vdots & \vdots & \ddots & \vdots \\ a_0^{\ell-1} & a_1^{\ell-1} & \cdots & a_{k-1}^{\ell-1} \end{matrix}$$

is a tuple contained in R, then so is the ℓ-tuple obtained by applying h to each row.

We have the following polymorphism-based characterization of primitive positive definability.

Theorem 2.4 ([**13**]) *Let $\mathfrak{A}$ be $\aleph_0$-categorical. A relation R over A is primitive positively definable in $\mathfrak{A}$ if and only if it is preserved by all polymorphisms of $\mathfrak{A}$.*

3 Periodic powers

In this section, we present the notion of *periodic power* of a structure, and identify some basic properties thereof. We also discuss how the periodic power arises as the direct limit of a system of embeddings. Throughout this section, we use $\mathfrak{A}, \mathfrak{B}$ to denote structures.

Definition 3.1 A function $\vec{a}\colon \mathbb{N} \to A$ is *periodic* if there exists $k \in \mathbb{N}$, $k > 0$ such that for all $i \in \mathbb{N}$, it holds that $\vec{a}(i) = \vec{a}(i \bmod k)$; in this case the function $\vec{a}$ is said to be *k-periodic*, and we write $\langle \vec{a}(0) \cdots \vec{a}(k-1)\rangle$ to denote $\vec{a}$. The set of periodic functions A^{per} carries a substructure in $\mathfrak{A}^{\mathbb{N}}$: the set A^{per} is closed under all $\mathfrak{A}^{\mathbb{N}}$-interpretations of function symbols. We define the *periodic power* of $\mathfrak{A}$, denoted $\mathfrak{A}^{\mathrm{per}}$, to be the substructure of $\mathfrak{A}^{\mathbb{N}}$ induced on A^{per}.

Lemma 3.2 *Assume that $\varphi(\overline{x})$ is a positive Horn formula. Then $(\mathfrak{A}^{\mathrm{per}}, \overline{\vec{a}}) \models \varphi(\overline{x})$ if and only if $(\mathfrak{A}, \overline{\vec{a}}(i)) \models \varphi(\overline{x})$ for all $i \in \mathbb{N}$.*

Here, $\overline{\vec{a}}(i)$ denotes the tuple obtained by evaluating the functions in $\overline{\vec{a}}$ at i; more precisely, if $\overline{\vec{a}} = \vec{a}_0 \cdots \vec{a}_{\ell-1}$ and $i \in \mathbb{N}$, then $\overline{\vec{a}}(i) = \vec{a}_0(i) \cdots \vec{a}_{\ell-1}(i) \in A^\ell$.

Proof of Lemma 3.2. Call a formula φ *good* if it satisfies the claimed equivalence. Clearly, conjunctions of atoms are good. Assume $\varphi(\overline{x}, y)$ is good. It is easy to see that also $\forall y \varphi(\overline{x}, y)$ is good. We show that $\exists y \varphi(\overline{x}, y)$ is good, via the following equivalences:

$$\begin{aligned} & (\mathfrak{A}^{\mathrm{per}}, \overline{\vec{a}}) \models \exists y \varphi(\overline{x}, y) \\ & \iff \exists \vec{b} \in A^{\mathrm{per}} : (\mathfrak{A}^{\mathrm{per}}, \overline{\vec{a}}, \vec{b}) \models \varphi(\overline{x}, y) \\ & \iff \exists \vec{b} \in A^{\mathrm{per}}\ \forall i \in \mathbb{N} : (\mathfrak{A}, \overline{\vec{a}}(i), \vec{b}(i)) \models \varphi(\overline{x}, y) \qquad (3.1) \\ & \iff \forall i \in \mathbb{N}\ \exists b \in A : (\mathfrak{A}, \overline{\vec{a}}(i), b) \models \varphi(\overline{x}, y) \qquad (3.2) \\ & \iff \forall i \in \mathbb{N} : (\mathfrak{A}, \overline{\vec{a}}(i)) \models \exists y \varphi(\overline{x}, y). \end{aligned}$$

The second equivalence follows from $\varphi(\overline{x}, y)$ being good. The rest being trivial, we show that (3.2) implies (3.1). By (3.2) there is a function $\vec{b}\colon \mathbb{N} \to A$ such that $(\mathfrak{A}, \overline{\vec{a}}(i), \vec{b}(i)) \models \varphi(\overline{x}, y)$ for all $i \in \mathbb{N}$. For any component $\vec{a}$ of $\overline{\vec{a}}$ choose $n_{\vec{a}} \in \mathbb{N}$ such that $\vec{a}$ is $n_{\vec{a}}$-periodic. Let $n \in \mathbb{N}$ be a common multiple of the $n_{\vec{a}}$s. Then any component of $\overline{\vec{a}}$ is n-periodic and, in particular,

$$\overline{\vec{a}}(i) = \overline{\vec{a}}(i \bmod n)$$

for all $i \in \mathbb{N}$. Define $\vec{b}^*\colon \mathbb{N} \to A$ by

$$\vec{b}^*(i) := \vec{b}(i \bmod n).$$

Then $\vec{b}^* \in A^{\mathrm{per}}$ and $(\mathfrak{A}, \overline{\vec{a}}(i), \vec{b}^*(i)) \models \varphi(\overline{x}, y)$ for all $i \in \mathbb{N}$; this is (3.1). □

Consider the following embeddings:

- The function $e_1\colon \mathfrak{A} \to \mathfrak{A}^{\mathrm{per}}$ defined by $e_1(a) := \langle a \rangle$, that is, the function mapping each $a \in A$ to the constant sequence $(a)_{i \in \mathbb{N}}$, is a canonical embedding of $\mathfrak{A}$ into $\mathfrak{A}^{\mathrm{per}}$.
- More generally, for each $k \in \mathbb{N}$, the function $e_k\colon \mathfrak{A}^k \to \mathfrak{A}^{\mathrm{per}}$ defined by $e_k((a_0, \ldots, a_{k-1})) := \langle a_0 \cdots a_{k-1} \rangle$ is a canonical embedding from $\mathfrak{A}^k$ into $\mathfrak{A}^{\mathrm{per}}$.

In the following proposition we identify $a \in A$ with $e_1(a) \in A^{\mathrm{per}}$ for notational simplicity. We use $\mathfrak{A} \preceq_{\mathrm{pH}} \mathfrak{B}$ to indicate that $\mathfrak{A} \subseteq \mathfrak{B}$ (i.e., $\mathfrak{A}$ is a substructure of $\mathfrak{B}$) and that for every positive Horn formula $\varphi(\overline{x})$ and all tuples $\overline{a}$ from A, it holds that

$$(\mathfrak{A}, \overline{a}) \models \varphi(\overline{x}) \iff (\mathfrak{B}, \overline{a}) \models \varphi(\overline{x}).$$

Lemmas 2.1 and 3.2 imply:

Proposition 3.3 $\mathfrak{A} \preceq_{\mathrm{pH}} \mathfrak{A}^{\mathrm{per}} \preceq_{\mathrm{pH}} \mathfrak{A}^{\mathbb{N}}$.

The next two propositions explain how the periodic power relates to finite powers.

Proposition 3.4 *Let $k \in \mathbb{N}$, $k > 0$. Then $\mathfrak{A}^{\mathrm{per}} \cong (\mathfrak{A}^k)^{\mathrm{per}}$ via an isomorphism that maps $\langle a_0 \cdots a_{k-1} \rangle$ to $\langle (a_0, \ldots, a_{k-1}) \rangle$ for all $a_0, \ldots, a_{k-1} \in A$.*

To make clear the notation used in the statement of this proposition, let us look at an example: the notation $\langle ab \rangle$ denotes the 2-periodic sequence $ababab \cdots \in A^{\mathrm{per}}$, whereas the notation $\langle (a,b) \rangle$ denotes the constant, 1-periodic sequence $(a,b)\ (a,b)\ (a,b) \cdots \in (A^2)^{\mathrm{per}}$.

Proof of Proposition 3.4. Choose for any $\vec{a} \in \mathfrak{A}^{\mathrm{per}}$ some $n_{\vec{a}} \in \mathbb{N}$ such that $\vec{a}$ is $n_{\vec{a}}$-periodic. Define the map $f\colon \mathfrak{A}^{\mathrm{per}} \to (\mathfrak{A}^k)^{\mathrm{per}}$ to map $\vec{a} \in \mathfrak{A}^{\mathrm{per}}$ to

$$i \mapsto (\vec{a}(ik), \ldots, \vec{a}((i+1)k - 1)).$$

The map f is clearly injective. For $j < k$, let π^k_j denote the projection of k-tuples to their $(j+1)$th component. An element $\vec{b} \in (\mathfrak{A}^k)^{\mathrm{per}}$ has

$$i \mapsto \pi^k_{i \bmod k}(\vec{b}(\lfloor i/k \rfloor))$$

as preimage under f, so f is surjective. It is straightforward to verify that f is an isomorphism. □

Proposition 3.5 *Let $k \in \mathbb{N}$, $k > 1$. Then $\mathfrak{A}^{\mathrm{per}} \cong (\mathfrak{A}^{\mathrm{per}})^k$.*

The proof relies on the following observation.

Lemma 3.6 $\mathfrak{A}^{\mathrm{per}} \times \mathfrak{B}^{\mathrm{per}} \cong (\mathfrak{A} \times \mathfrak{B})^{\mathrm{per}}$.

Proof. Map a pair of functions $(\vec{a}, \vec{b}) \in A^{\mathrm{per}} \times B^{\mathrm{per}}$ to $((\vec{a}(i), \vec{b}(i)))_{i \in \mathbb{N}}$; note this function is nm-periodic whenever $\vec{a}$ and $\vec{b}$ are n- and m-periodic respectively. The map is clearly injective. It is surjective as $((a_i, b_i))_{i \in \mathbb{N}} \in (A \times B)^{\mathrm{per}}$ has preimage $((a_i)_{i \in \mathbb{N}}, (b_i)_{i \in \mathbb{N}}) \in$

$A^{\text{per}} \times B^{\text{per}}$. To see that it is an isomorphism, let α be an atom. For simplicity assume $\alpha = \alpha(x, y)$, and let $(\vec{a}, \vec{b}), (\vec{a}', \vec{b}') \in A^{\text{per}} \times B^{\text{per}}$. Then

$$\begin{aligned}
&(\mathfrak{A}^{\text{per}} \times \mathfrak{B}^{\text{per}}, (\vec{a}, \vec{b}), (\vec{a}', \vec{b}')) \models \alpha(x, y)\\
&\quad\iff (\mathfrak{A}^{\text{per}}, \vec{a}, \vec{a}') \models \alpha(x, y) \text{ and } (\mathfrak{B}^{\text{per}}, \vec{b}, \vec{b}') \models \alpha(x, y)\\
&\quad\iff \forall i \in \mathbb{N}: \ (\mathfrak{A}, \vec{a}(i), \vec{a}'(i)) \models \alpha(x, y) \text{ and } (\mathfrak{B}, \vec{b}(i), \vec{b}'(i)) \models \alpha(x, y)\\
&\quad\iff \forall i \in \mathbb{N}: \ (\mathfrak{A} \times \mathfrak{B}, (\vec{a}(i), \vec{b}(i)), (\vec{a}'(i), \vec{b}'(i))) \models \alpha(x, y)\\
&\quad\iff ((\mathfrak{A} \times \mathfrak{B})^{\text{per}}, (\vec{a}(i), \vec{b}(i))_{i\in\mathbb{N}}, ((\vec{a}'(i), \vec{b}'(i)))_{i\in\mathbb{N}}) \models \alpha(x, y),
\end{aligned}$$

where the first and third equivalences hold by definition of direct products, and the second and fourth equivalences hold by Lemma 3.2. □

Proof of Proposition 3.5. By induction on k: we have the isomorphisms

$$(\mathfrak{A}^{\text{per}})^{k+1} = (\mathfrak{A}^{\text{per}})^k \times \mathfrak{A}^{\text{per}} \cong \mathfrak{A}^{\text{per}} \times \mathfrak{A}^{\text{per}} \cong (\mathfrak{A}^2)^{\text{per}} \cong \mathfrak{A}^{\text{per}}$$

by induction, the previous lemma and Proposition 3.4. □

Observe that for $n, m > 0$ there is a natural embedding $e_{(n,m)}\colon \mathfrak{A}^n \to \mathfrak{A}^m$ whenever $n < m$ and n divides m, namely the embedding that maps the n-tuple $\overline{a} \in A^n$ to the m-tuple

$$e_{(n,m)}(\overline{a}) = \underbrace{\overline{a}\overline{a}\cdots\overline{a}}_{m/n \text{ times}} \in A^m.$$

Clearly, these embeddings are compatible in the sense that $e_{(\ell,m)} \circ e_{(n,\ell)} = e_{(n,m)}$ whenever $n < \ell < m$, n divides ℓ and ℓ divides m. In other words, the $e_{(n,m)}$s determine an $(I, \prec)$-system of embeddings where $I = \mathbb{N}_{>0}$ and $\mathfrak{A}_n := \mathfrak{A}^n$ and $\prec$ denotes divisibility.

Proposition 3.7 $\mathfrak{A}^{\text{per}} \cong \lim_n \mathfrak{A}^n$.

Proof. Let $(e_n^*)_{n>0}$ denote the limit homomorphisms into the direct limit $\lim_n \mathfrak{A}^n$ of the directed system of embeddings given by the $e_{(n,m)}$s (for $n < m$ and n divides m). Observe that the embeddings e_n from $\mathfrak{A}^n$ into $\mathfrak{A}^{\text{per}}$ satisify the requirement for limit embeddings, so the e_ns are also a cone of the directed system. By the universal property of $\lim_n \mathfrak{A}^n$ there is an embedding $e\colon \lim_n \mathfrak{A}^n \to \mathfrak{A}^{\text{per}}$ such that $e \circ e_n^* = e_n$ for all $n > 0$. But every element of A^{per} is in the image of some e_n, so e has to be surjective and thus is an isomorphism. □

Remark 3.8 If $J \subseteq I$ is dense in $(I, \prec)$ (i.e., every $i \in I$ has an upper bound in J), then the direct limit of an I-system of embeddings is isomorphic to the direct limit of the subsystem restricted to J. For every $k > 0$ the set of multiples of k is dense in $\mathbb{N}_{>0}$ with respect to divisibility. Thus

$$\mathfrak{A}^{\text{per}} \cong \lim_n \mathfrak{A}^n \cong \lim_\ell \mathfrak{A}^{\ell\cdot k} \cong \lim_\ell (\mathfrak{A}^k)^\ell \cong (\mathfrak{A}^k)^{\text{per}}.$$

Here, the third occurrence of lim is to be understood with respect to the embeddings imported from the system of the $\mathfrak{A}^{\ell\cdot k}$s via the natural isomorphisms $\mathfrak{A}^{\ell\cdot k} \cong (\mathfrak{A}^k)^\ell$.

Propositions 3.3 and 3.7 imply:

Corollary 3.9 *Every positive Horn sentence true in* $\mathfrak{A}$ *and every* $\forall\exists$*-sentence true in all finite powers of* $\mathfrak{A}$*, is true in* $\mathfrak{A}^{\text{per}}$.

Recall that a $\forall\exists$-sentence is a sentence of the form $\forall\overline{x}\,\exists\overline{y}\,\psi$ with ψ quantifier free.

4 Periomorphisms

In this section, we introduce and study the notion of *periomorphism.* Throughout this section, let $\mathfrak{A}$ be a structure.

Definition 4.1 A *periomorphism of* $\mathfrak{A}$ is a homomorphism from $\mathfrak{A}^{\text{per}}$ to $\mathfrak{A}$.

In other words, a periomorphism of $\mathfrak{A}$ is a partial function from $A^{\mathbb{N}}$ to A with domain A^{per} that preserves all atomic formulas. The following lemma follows straightforwardly from the definitions.

Lemma 4.2 *A periomorphism h of $\mathfrak{A}$ preserves a relation $R \subseteq A^\ell$ if and only if for any choice of finitely many tuples $\overline{a}_0 = (a_0^0, \ldots, a_0^{\ell-1}), \ldots, \overline{a}_{k-1} = (a_{k-1}^0, \ldots, a_{k-1}^{\ell-1})$ from R, we have*

$$\big(h(\langle a_0^0 a_1^0 \cdots a_{k-1}^0\rangle), \ldots, h(\langle a_0^{\ell-1} a_1^{\ell-1} \cdots a_{k-1}^{\ell-1}\rangle)\big) \in R.$$

Proof. The forward direction is trivial. Conversely, assume the right hand side of the claimed equivalence and let $\vec{a}_0, \ldots, \vec{a}_{\ell-1} \in \mathfrak{A}^{\text{per}}$ be such that $(\vec{a}_0(i), \ldots \vec{a}_{\ell-1}(i)) \in R$ for all $i \in \mathbb{N}$. We claim $h(\vec{a}_0) \cdots h(\vec{a}_{\ell-1}) \in R$. Choose a sufficiently large $k \in \mathbb{N}$ such that all $\vec{a}_j$ are k-periodic, that is, $\vec{a}_j = \langle \vec{a}_j(0) \cdots \vec{a}_j(k-1)\rangle$ for all $j < \ell$. Applying the assumption yields the claim. □

To see the lemma's statement with a picture, let h be a periomorphism of $\mathfrak{A}$, and consider the following:

$$\begin{matrix} \langle a_0^0 & a_1^0 & \cdots & a_{k-1}^0\rangle \\ \langle a_0^1 & a_1^1 & \cdots & a_{k-1}^1\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle a_0^{\ell-1} & a_1^{\ell-1} & \cdots & a_{k-1}^{\ell-1}\rangle. \end{matrix}$$

The right hand side of the lemma states that if the column ℓ-tuples $\overline{a}_i = (a_i^0, \ldots, a_i^{\ell-1})$ are contained in R for all $i < k$, then so is the ℓ-tuple $\overline{b}$ obtained by applying h to each row.

For later use we introduce the following mode of speech.

Definition 4.3 In the situation above, if h is a *surjective* periomorphism of the structure under study, then we call $\overline{b}$ *a surjective periomorphic image of the tuples $\overline{a}_i$, $i < k$.*

Proposition 4.4 *With respect to a structure $\mathfrak{A}$, every positive Horn formula is preserved by all surjective periomorphisms of $\mathfrak{A}$.*

Proof. Let $\varphi(\overline{x})$ be a positive Horn formula and h be a surjective periomorphism of $\mathfrak{A}$. For notational simplicity assume $\overline{x} = xx'$ and let $a_0a_0', \ldots, a_{k-1}a_{k-1}'$ be any finitely many pairs in $\varphi(\mathfrak{A})$. We have to show that $\varphi(xx')$ is true in $(\mathfrak{A}, h(\langle a_0 \cdots a_{k-1}\rangle), h(\langle a_0' \cdots a_{k-1}'\rangle))$; see the previous lemma. But $\varphi(xx')$ is true in $(\mathfrak{A}^{\text{per}}, \langle a_0 \cdots a_{k-1}\rangle, \langle a_0' \cdots a_{k-1}'\rangle)$ by Lemma 3.2, and, being positive, is preserved by surjective homomorphisms. □

The periomorphisms and the polymorphisms of a structure contain the same information. If one knows the periomorphisms of a structure, then one also knows its polymorphisms, and vice-versa. Why is this? For $k \in \mathbb{N}$, $k > 0$, define

$$\pi_{<k} : A^{\text{per}} \to A^k : \pi_{<k}(\vec{a}) := (\vec{a}(0), \ldots, \vec{a}(k-1)).$$

This operation is clearly a homomorphism from $\mathfrak{A}^{\mathrm{per}}$ to $\mathfrak{A}^k$. Now, if someone hands us an operation $h\colon A^k \to A$, we can decide if it is a polymorphism of $\mathfrak{A}$ by checking if

$$h^{\mathrm{per}} := h \circ \pi_{<k}$$

is a periomorphism of $\mathfrak{A}$. For, if h is a polymorphism of $\mathfrak{A}$, then by composing homomorphisms we have that h^{per} is a periomorphism of $\mathfrak{A}$; and, if h^{per} is a periomorphism of $\mathfrak{A}$, by composing homomorphisms, we have that $h^{\mathrm{per}} \circ e_k$, which is equal to h, is a homomorphism from $\mathfrak{A}^k$ to $\mathfrak{A}$.

Going the other way, suppose that someone places in our hands an operation $h\colon A^{\mathrm{per}} \to A$. It can be seen from Lemma 4.2 that h is a periomorphism of $\mathfrak{A}$ if and only if each of the operations

$$h_{<k} := h \circ e_k \tag{4.1}$$

is a polymorphism of $\mathfrak{A}$.

It is thus no surprise that preservation by periomorphisms coincides with preservation by polymorphisms. Preservation by *surjective* periomorphisms, however, is an a priori stronger property than preservation by surjective polymorphisms.

Proposition 4.5 *Let φ be a formula. Then*

(1) *φ is preserved by all periomorphisms of $\mathfrak{A}$ if and only if φ is preserved by all polymorphisms of $\mathfrak{A}$;*

(2) *if φ is preserved by all surjective periomorphisms of $\mathfrak{A}$, then φ is preserved by all surjective polymorphisms of $\mathfrak{A}$.*

Proof. To see the forward directions, observe that if h is a (surjective) polymorphism of $\mathfrak{A}$ that does not preserve φ, then h^{per} is a (surjective) periomorphism of $\mathfrak{A}$ that does not preserve φ. For the converse direction in (1) assume h is a periomorphism that does not preserve $\varphi = \varphi(x_0, \dots, x_{\ell-1})$. Then by Lemma 4.2 there are $k \in \mathbb{N}$ and $(a_0^0, \dots, a_0^{\ell-1}), \dots, (a_{k-1}^0, \dots, a_{k-1}^{\ell-1}) \in \varphi(\mathfrak{A})$ such that

$$(h(\langle a_0^0 a_1^0 \cdots a_{k-1}^0 \rangle), \dots, h(\langle a_0^{\ell-1} a_1^{\ell-1} \cdots a_{k-1}^{\ell-1} \rangle)) \notin \varphi(\mathfrak{A}),$$

that is,

$$\big(h(e_k(a_0^0, a_1^0, \dots a_{k-1}^0)), \dots, h(e_k(a_0^{\ell-1}, a_1^{\ell-1}, \dots a_{k-1}^{\ell-1}))\big) \notin \varphi(\mathfrak{A}).$$

Hence, $h_{<k}$ is a k-ary polymorphism of $\mathfrak{A}$ that does not preserve φ. □

Remark 4.6 Inspection of the above proof shows that the converse of (2) is true in case $\mathfrak{A}$ satisfies the following condition: for every surjective periomorphism h of $\mathfrak{A}$ there exists $k \in \mathbb{N}$ such that $h_{<k}$ is surjective. For example, finite structures satisfy this condition.

We saw that a periomorphism h gives rise to a sequence of polymorphisms $(h_{<k})_{k>0}$. In fact, this gives a one-to-one correspondence with those polymorphism sequences that satisfy the following property.

Definition 4.7 A sequence $(g_k)_{k>0}$ is a *cone of polymorphisms of* $\mathfrak{A}$ if every g_k is a k-ary polymorphism of $\mathfrak{A}$ and $g_\ell = g_k \circ e_{(\ell,k)}$ whenever $\ell < k$ and ℓ divides k.

Proposition 4.8 *A sequence $(g_k)_{k>0}$ is a cone of polymorphisms of $\mathfrak{A}$ if and only if there is a periomorphism h of $\mathfrak{A}$ such that $h_{<k} = g_k$ for all $k > 0$.*

Proof. For the backward direction, let h be a periomorphism of $\mathfrak{A}$. Clearly, $(h_{<k})_{k>0}$ is a sequence of polymorphisms of $\mathfrak{A}$ —and it is a cone:

$$h_{<\ell} = h \circ e_\ell = h \circ (e_k \circ e_{(\ell,k)}) = h_{<k} \circ e_{(\ell,k)}.$$

Here, the second equality follows from the e_ℓs being limit embeddings (see the previous section).

Conversely, assume that $(g_k)_{k>0}$ is a cone of polymorphisms of $\mathfrak{A}$. Assume that $(n_{\vec{a}})_{\vec{a}} = (n_{\vec{a}})_{\vec{a}\in A^{\mathrm{per}}}$ is a *nice family*, that is, every $\vec{a} \in A^{\mathrm{per}}$ is $n_{\vec{a}}$-periodic. With respect to this family, define

$$h(\vec{a}) := g_{n_{\vec{a}}} \circ \pi_{<n_{\vec{a}}}(\vec{a}).$$

We claim that h is independent from the choice of the nice family. More precisely, let $(n'_{\vec{a}})_{\vec{a}}$ be another nice family and define h' with respect to this family as h is defined with respect to $(n_{\vec{a}})_{\vec{a}}$. We claim that $h = h'$.

By symmetry, it suffices to show $h = h''$ where h'' is analogously defined with respect to $(n_{\vec{a}} \cdot n'_{\vec{a}})_{\vec{a}}$ (observe that with $(n_{\vec{a}})_{\vec{a}}$ and $(n'_{\vec{a}})_{\vec{a}}$ also $(n_{\vec{a}} \cdot n'_{\vec{a}})_{\vec{a}}$ is a nice family). But this is true:

$$\begin{aligned} h(\vec{a}) &= g_{n_{\vec{a}}} \circ \pi_{<n_{\vec{a}}}(\vec{a}) \\ &= (g_{n_{\vec{a}}\cdot n'_{\vec{a}}} \circ e_{(n_{\vec{a}}, n_{\vec{a}}\cdot n'_{\vec{a}})}) \circ \pi_{<n_{\vec{a}}}(\vec{a}) \\ &= g_{n_{\vec{a}}\cdot n'_{\vec{a}}} \circ \pi_{<n_{\vec{a}}\cdot n'_{\vec{a}}}(\vec{a}) \\ &= h''(\vec{a}). \end{aligned}$$

Here, the first and the last equality follow from the definition of h and h'' respectively, the second follows from the cone property, and the third follows from

$$e_{(k,\ell)} \circ \pi_{<k} = \pi_{<\ell}$$

whenever $k < \ell$ and k divides ℓ.

It is routine to check that h is a periomorphism of $\mathfrak{A}$. Hence, we are left to verify that $h_{<k} = g_k$ for every $k > 0$. So, given $k > 0$ and $\overline{a} \in A^k$, we have to show $h_{<k}(\overline{a}) = g_k(\overline{a})$. By definition $h_{<k}(\overline{a}) = g_{n_{\vec{b}}} \circ \pi_{<n_{\vec{b}}}(\vec{b})$ for $\vec{b} := e_k(\overline{a}) \in A^{\mathrm{per}}$.

Now $\vec{b}$ is k-periodic, so there exists a nice family $(\tilde{n}_{\vec{a}})_{\vec{a}}$ with $\tilde{n}_{\vec{b}} = k$. For the corresponding $\tilde{h}$ we know $h = \tilde{h}$. Hence

$$h_{<k}(\overline{a}) = \tilde{h}_{<k}(\overline{a}) = g_k \circ \pi_{<k} \circ e_k(\overline{a}) = g_k(\overline{a}),$$

as claimed. □

Intuitively speaking, just as the periodic power is a cone of finite powers, any periomorphism “is” a cone of (finitary) polymorphisms. Note that the last proposition just details a special case of “lifting cones of homomorphisms to direct limits”.

5 Preservation theorem

Theorem 5.1 (Main) *Let $\mathfrak{A}$ be an $\aleph_0$-categorical structure. A relation R over A is positive Horn definable in $\mathfrak{A}$ if and only if it is preserved by all surjective periomorphisms of $\mathfrak{A}$.*

The following is a straightforward generalization of Proposition 4.4.

Proposition 5.2 *If $\mathfrak{A}$ and $\mathfrak{B}$ are structures such that there is a surjective homomorphism from $\mathfrak{A}^{\mathrm{per}}$ onto $\mathfrak{B}$, then $\mathfrak{A} \Rrightarrow_{\mathrm{pH}} \mathfrak{B}$.*

The main lemma in the proof of Theorem 5.1 states that a converse of this proposition holds true in the $\aleph_0$-categorical case:

Lemma 5.3 *If $\mathfrak{A}$ and $\mathfrak{B}$ are $\aleph_0$-categorical structures such that $\mathfrak{A} \Rrightarrow_{\mathrm{pH}} \mathfrak{B}$, then there is a surjective homomorphism from $\mathfrak{A}^{\mathrm{per}}$ onto $\mathfrak{B}$.*

Proof. Let I be the set of finite partial functions f from $\mathfrak{A}^{\mathrm{per}}$ to $\mathfrak{B}$ such that

$$(\mathfrak{A}^{\mathrm{per}}, \overline{\vec{a}}) \Rrightarrow_{\mathrm{pH}} (\mathfrak{B}, \overline{b}), \tag{5.1}$$

where $\overline{\vec{a}}$ is a (finite) tuple from $\mathfrak{A}^{\mathrm{per}}$ listing all elements of the domain of f and $\overline{b}$ is a tuple from $\mathfrak{B}$ such f maps $\overline{\vec{a}}$ to $\overline{b}$.

Observe that $\mathfrak{A}^{\mathrm{per}}$ is countable. Hence, by a standard back and forth argument, it suffices to verify the following two claims.

Claim 1 For all $f \in I$ and $\vec{a} \in A^{\mathrm{per}}$ there is $b \in B$ such that $f \cup \{(\vec{a}, b)\} \in I$.

Claim 2 For all $f \in I$ and $b \in B$ there is $\vec{a} \in A^{\mathrm{per}}$ such that $f \cup \{(\vec{a}, b)\} \in I$.

Proof of Claim 1. Given $f \in I$, choose tuples $\overline{\vec{a}}$ and $\overline{b}$ as above. Let $\vec{a} \in A^{\mathrm{per}}$ be arbitrary. It sufficies to find $b \in B$ such that

$$(A^{\mathrm{per}}, \overline{\vec{a}}, \vec{a}) \Rrightarrow_{\mathrm{pH}} (\mathfrak{B}, \overline{b}, b). \tag{5.2}$$

Note in particular that $x = y$ is positive Horn, so (5.2) implies that $f \cup \{(\vec{a}, b)\}$ is a function. To find such b, consider the set $\Delta(x)$ of all positive Horn formulas $\psi(x)$ (in the language of $(\mathfrak{A}^{\mathrm{per}}, \overline{\vec{a}})$) satisfied by $\vec{a}$ in $(\mathfrak{A}^{\mathrm{per}}, \overline{\vec{a}})$. It suffices to show that this set is satisfiable in $(\mathfrak{B}, \overline{b})$. Since $\mathfrak{B}$ is $\aleph_0$-categorical, it is $\aleph_0$-saturated (recall Section 1.4), and hence it suffices to show that $\Delta(x)$ is finitely satisfable in $(\mathfrak{B}, \overline{b})$. But for a finite $\Delta_0(x) \subseteq \Delta(x)$ the positive Horn sentence $\exists x \bigwedge \Delta_0(x)$ is true in $(\mathfrak{A}^{\mathrm{per}}, \overline{\vec{a}})$, so it is also true in $(\mathfrak{B}, \overline{b})$ by (5.1). Hence $(\mathfrak{B}, \overline{b})$ contains some b satisfying $\Delta_0(x)$. ⊣

Proof of Claim 2. Let $f \in I$ and again choose $\overline{\vec{a}}$ and $\overline{b}$ as above; say, these tuples have length k. Again it suffices, given any $b \in B$, to find some $\vec{a} \in A^{\mathrm{per}}$ such that (5.2) holds. As $\mathfrak{A}$ is $\aleph_0$-categorical by Ryll-Nardzewski, there are up to equivalence in $\mathfrak{A}$ only finitely many formulas in the variables $\overline{y}x$ where $\overline{y}$ is a tuple of k variables. Let

$$\psi_0(\overline{y}, x), \ldots, \psi_{m-1}(\overline{y}, x)$$

list all positive Horn formulas that are in $\mathfrak{A}$ equivalent to some positive Horn formula $\psi(\overline{y}, x)$ such that

$$\mathfrak{B} \not\models \psi(\overline{b}, b). \tag{5.3}$$

Subclaim For every $j < m$ we have $(\mathfrak{A}^{\mathrm{per}}, \overline{\vec{a}}) \not\models \forall x \psi_j(\overline{y}, x)$.

Proof of the subclaim. Otherwise there is $j < m$ such that for all $i \in \mathbb{N}$ we have $(\mathfrak{A}, \overline{\vec{a}}(i)) \models \forall x \psi_j(\overline{y}, x)$ (by Lemma 3.2). Choose a positive Horn formula $\psi(\overline{y}, x)$ that is equivalent to $\psi_j(\overline{y}, x)$ in $\mathfrak{A}$ and such that (5.3) holds. Then $(\mathfrak{A}, \overline{\vec{a}}(i)) \models \forall x \psi(\overline{y}, x)$ holds for all $i \in \mathbb{N}$ and hence $(\mathfrak{A}^{\mathrm{per}}, \overline{\vec{a}}) \models \forall x \psi(\overline{y}, x)$ (by Lemma 3.2). As $f \in I$, $\mathfrak{B} \models \forall x \psi(\overline{b}, x)$ follows and this contradicts (5.3). ⊣

By the subclaim and Lemma 3.2 there are $i_0 \in \mathbb{N}$ and $a_0 \in A$ such that

$$(\mathfrak{A}, \overline{\vec{a}}(i_0)) \not\models \psi_0(\overline{y}, a_0).$$

Similarly, there are $i_1 \in \mathbb{N}$ and $a_1 \in A$ such that

$$(5.4) \qquad (\mathfrak{A}, \overline{\vec{a}}(i_1)) \not\models \psi_1(a_1).$$

Moreover, we can choose i_1 such that $i_1 > i_0$ by periodicity: if $i_1 \leq i_0$ replace it by $i_1 + i_0 \cdot n$ where $n \in \mathbb{N}$ is large enough such that all components of $\overline{\vec{a}}$ are n-periodic; then $\overline{\vec{a}}(i_1) = \overline{\vec{a}}(i_1 + i_0 \cdot n)$ and (5.4) remains true.

Continuing in this manner we get sequences $i_0 < i_1 < \cdots < i_{m-1}$ and $a_0, a_1, \ldots, a_{m-1}$ such that for all $j < m$

$$(5.5) \qquad (\mathfrak{A}, \overline{\vec{a}}(i_j)) \not\models \psi_j(\overline{y}, a_j).$$

Choose a periodic $\vec{a} \colon \mathbb{N} \to A$ such that, for all $j < m$,

$$(5.6) \qquad \vec{a}(i_j) = a_j.$$

We verify (5.2) for this $\vec{a}$: let $\psi(\overline{y}, x)$ be a positive Horn formula such that $\mathfrak{B} \not\models \psi(\overline{b}, b)$. Then there exists $j < m$ such that $\psi(\overline{y}, x)$ is in $\mathfrak{A}$ equivalent to $\psi_j(\overline{y}, x)$. By (5.5) and (5.6) we get $(\mathfrak{A}, \overline{\vec{a}}(i_j)) \not\models \psi_j(\overline{y}, \vec{a}(i_j))$ and hence $(\mathfrak{A}, \overline{\vec{a}}(i_j)) \not\models \psi(\overline{y}, \vec{a}(i_j))$. By Lemma 3.2 we conclude $(\mathfrak{A}^{\mathrm{per}}, \overline{\vec{a}}) \not\models \psi(\overline{y}, \vec{a})$. □

Proof of Theorem 5.1. The forward direction follows from Proposition 4.4 (the $\aleph_0$-categoricity of $\mathfrak{A}$ is not needed).

Conversely, assume that a relation $R \subseteq A^\ell$ is preserved by all surjective periomorphisms of $\mathfrak{A}$. By Proposition 4.5(2) it is preserved by all surjective polymorphisms, and in particular by all automorphisms of $\mathfrak{A}$. Since $\mathfrak{A}$ is $\aleph_0$-categorical, R is first-order definable in $\mathfrak{A}$ (recall Section 1.4). Let $\varphi_R(\overline{x}) = \varphi_R(x_0, \ldots, x_{\ell-1})$ be a formula such that $R = \varphi_R(\mathfrak{A})$.

By Ryll-Nardzewski there is a finite list of positive Horn formulas

$$\psi_0(\overline{x}), \ldots, \psi_{m-1}(\overline{x})$$

in the free variables $\overline{x} = x_0 \cdots x_{\ell-1}$ such that every such formula is in $\mathfrak{A}$ equivalent to one from the list. Some of these formulas are implied by $\varphi_R(\overline{x})$ (in $\mathfrak{A}$) and others not, and we may suppose that precisely the first k are not:

$$(5.7) \qquad \begin{aligned} &\forall i < k \; \exists \overline{a}_i \in A^\ell \; : \; \overline{a}_i \in \varphi_R(\mathfrak{A}) \setminus \psi_i(\mathfrak{A}); \\ &\forall k \leq j < m \; : \; \varphi_R(\mathfrak{A}) \subseteq \psi_j(\mathfrak{A}). \end{aligned}$$

We can assume that $k \neq 0$ as otherwise $(\varphi_R \leftrightarrow \bot)$ holds in $\mathfrak{A}$ and then we are done. We claim that the positive Horn formula $\bigwedge_{k \leq j < m} \psi_j(\overline{x})$ is equivalent to $\varphi_R(\overline{x})$ in $\mathfrak{A}$. Therefore, it suffices to show

$$\mathfrak{A} \models \forall \overline{x} \big(\textstyle\bigwedge_{k \leq j < m} \psi_j(\overline{x}) \to \varphi_R(\overline{x}) \big).$$

So we assume that $\overline{b}$ satisfies $\bigwedge_{k \leq j < m} \psi_j(\overline{x})$ in $\mathfrak{A}$ and have to show that $\overline{b} \in \varphi_R(\mathfrak{A})$.

Choose for $i < k$ a tuple $\overline{a}_i \in A^\ell$ according to (5.7).

Claim $\prod_{i<k} (\mathfrak{A}, \overline{a}_i) \Rrightarrow_{\mathrm{pH}} (\mathfrak{A}, \overline{b})$.

Proof of the claim. Let $\psi(\overline{x})$ be a positive Horn formula that is not satisfied by $\overline{b}$ in $\mathfrak{A}$. Choose $i < m$ such that $\psi_i(\overline{x})$ is equivalent to $\psi(\overline{x})$ in $\mathfrak{A}$. Then $\overline{b}$ does not satisfy $\psi_i(\overline{x})$ in $\mathfrak{A}$, so $i < k$. But then $(\mathfrak{A}, \overline{a}_i) \not\models \psi_i(\overline{x})$ by (5.7) and thus $(\mathfrak{A}, \overline{a}_i) \not\models \psi(\overline{x})$. As $\psi(\overline{x})$ is positive Horn, $\prod_{i<k} (\mathfrak{A}, \overline{a}_i) \not\models \psi(\overline{x})$ by Lemma 2.1. ⊣

Write $\overline{a}_i = a_i^0 \cdots a_i^{\ell-1}$ for $i < k$. Then $\prod_{i<k}(\mathfrak{A}, \overline{a}_i)$ equals

$$\bigl(\mathfrak{A}^k, (a_0^0, \ldots, a_{k-1}^0)(a_0^1, \ldots, a_{k-1}^1) \cdots (a_0^{\ell-1}, \ldots, a_{k-1}^{\ell-1})\bigr).$$

With $\mathfrak{A}$ also $(\mathfrak{A}, \overline{b})$ is $\aleph_0$-categorical, and the structure $\bigl(\mathfrak{A}^k, (a_0^0, \ldots, a_{k-1}^0) \cdots\bigr)$ is $\aleph_0$-categorical, because $\mathfrak{A}^k$ is (see Section 1.4). By the Claim we can thus apply Lemma 5.3 and conclude that there is a surjective homomorphism

$$h : \bigl(\mathfrak{A}^k, (a_0^0, \ldots, a_{k-1}^0) \cdots (a_0^{\ell-1}, \ldots, a_{k-1}^{\ell-1})\bigr)^{\mathrm{per}} \twoheadrightarrow (\mathfrak{A}, \overline{b}).$$

By Proposition 3.4 there is an isomorphism g from the left hand side structure onto

$$\bigl(\mathfrak{A}^{\mathrm{per}}, \langle a_0^0 \cdots a_{k-1}^0\rangle \cdots \langle a_0^{\ell-1} \cdots a_{k-1}^{\ell-1}\rangle\bigr).$$

Then $h \circ g^{-1}$ is a surjective homomorphism from $\mathfrak{A}^{\mathrm{per}}$ onto $\mathfrak{A}$, i.e., a surjective periomorphism of $\mathfrak{A}$, such that

$$h \circ g^{-1}(\langle a_0^0 \cdots a_{k-1}^0\rangle) \cdots h \circ g^{-1}(\langle a_0^{\ell-1} \cdots a_{k-1}^{\ell-1}\rangle) = \overline{b}.$$

By (5.7) we have $\overline{a}_i \in \varphi_R(\mathfrak{A})$ for all $i < k$. By Lemma 4.2 and the assumption that R and hence $\varphi_R(\overline{x})$ is preserved by surjective periomorphisms of $\mathfrak{A}$, we conclude that $\overline{b} \in \varphi(\mathfrak{A})$, as was to be shown. □

Theorem 5.4 *For a finite language L_0, let $\mathfrak{B}$ be an L_0-structure and $\mathfrak{A}$ an L-structure on the same universe. If every surjective periomorphism of $\mathfrak{A}$ is a periomorphism of $\mathfrak{B}$, then the problem* $\mathsf{QCSP}(\mathfrak{B})$ *many-one logspace reduces to* $\mathsf{QCSP}(\mathfrak{A})$.

Proof. If $\varphi(\overline{x})$ is an atomic L_0-formula, then $\varphi(\mathfrak{B})$ is preserved by all polymorphisms of $\mathfrak{B}$, hence also by all periomorphisms of $\mathfrak{B}$ (by Proposition 4.5(1)), and hence by all surjective periomorphisms of $\mathfrak{A}$ (by assumption). By the Main Theorem 5.1 the relation $\varphi(\mathfrak{B})$ is positive Horn definable in $\mathfrak{A}$. Hence $\mathfrak{B}$ is positive Horn definable in $\mathfrak{A}$. Now apply Proposition 2.2. □

6 Characterization of the pH-hull

A central tool in constraint complexity is the description of the smallest primitive positive definable relation containing a given relation R as the smallest relation that contains all polymorphic images of R; this description follows readily from Theorem 2.4. Here we provide a similar tool for quantified constraint complexity. The proof of this uses most of the results we established so far.

Recall Definition 4.3.

Theorem 6.1 *Let $\mathfrak{A}$ be $\aleph_0$-categorical and let R be a relation over A. Then*

$$\{\overline{a} \mid \exists k \in \mathbb{N}\ \exists \overline{a}_0, \ldots, \overline{a}_{k-1} \in R : \\ \overline{a} \textit{ is a surjective periomorphic image of } \overline{a}_i,\ i < k\}$$

is the smallest positive Horn definable relation containing R.

Proof. For notational simplicity, we assume that R is binary. It is easy to see that the described relation $\tilde{R}$ contains R. We have to show

(i) $\tilde{R} \subseteq \psi(\mathfrak{A})$ for any positive Horn formula ψ such that $R \subseteq \psi(\mathfrak{A})$;

(ii) $\tilde{R}$ is positive Horn definable in $\mathfrak{A}$.

To show (i), let $aa' \in \tilde{R}$. Choose $a_i a'_i$, $i < k$ in R such that some surjective periomorphism of $\mathfrak{A}$ maps $\langle a_0 \cdots a_{k-1}\rangle\langle a'_0 \cdots a'_{k-1}\rangle$ to aa'. Then $a_i a'_i \in \psi(\mathfrak{A})$ as $R \subseteq \psi(\mathfrak{A})$, so $aa' \in \psi(\mathfrak{A})$ by Proposition 4.4 as ψ is positive Horn.

We now prove (ii). By Theorem 5.1 it suffices to show that $\tilde{R}$ is preserved by all surjective periomorphisms of $\mathfrak{A}$. We use Lemma 4.2, so let $a_i a'_i$, $i < k$ be k tuples in $\tilde{R}$ and h be a surjective periomorphism that maps $\langle a_0 \cdots a_{k-1}\rangle\langle a'_0 \cdots a'_{k-1}\rangle$ to aa'. We have to show that $aa' \in \tilde{R}$.

For $i < k$, choose ℓ_i pairs $b_{ij} b'_{ij}$, $j < \ell_i$ in R such that there is a surjective periomorphism h_i that maps $\langle b_{i0} \cdots b_{i(\ell_i-1)}\rangle\langle b'_{i0} \cdots b'_{i(\ell_i-1)}\rangle$ to $a_i a'_i$. Letting the h_js act componentwise we get a surjective homomorphism

$$h' : \textstyle\prod_{i<k}(\mathfrak{A}^{\mathrm{per}}, \langle b_{i0} \cdots b_{i(\ell_i-1)}\rangle\langle b'_{i0} \cdots b'_{i(\ell_i-1)}\rangle) \twoheadrightarrow \prod_{i<k}(\mathfrak{A}, a_i a'_i). \tag{6.1}$$

By Proposition 3.4, the left-hand side structure is isomorphic to

$$\textstyle\prod_{i<k}(\mathfrak{A}^{\ell_i}, (b_{i0} \cdots b_{i(\ell_i-1)})(b'_{i0} \cdots b'_{i(\ell_i-1)}))^{\mathrm{per}}$$

and thus, by Lemma 3.6, to the periodic power of

$$\left(\mathfrak{A}^{\sum_{i<k}\ell_i}, (b_{00} \cdots b_{(k-1)(\ell_{k-1}-1)}), (b'_{00} \cdots b'_{(k-1)(\ell_{k-1}-1)})\right).$$

Denote this structure by $\mathfrak{B}$. By (6.1) and Proposition 5.2 we get

$$\mathfrak{B} \Rrightarrow_{\mathrm{pH}} \textstyle\prod_{i<k}(\mathfrak{A}, a_i a'_i). \tag{6.2}$$

By Proposition 3.4, $(\prod_{i<k}(\mathfrak{A}, a_i a'_i))^{\mathrm{per}}$ is isomorphic to $(\mathfrak{A}^{\mathrm{per}}, \langle a_0 \cdots a_{k-1}\rangle, \langle a'_0 \cdots a'_{k-1}\rangle)$ which maps surjectively onto $(\mathfrak{A}, aa')$ by h. Hence, by Proposition 5.2 again,

$$\textstyle\prod_{i<k}(\mathfrak{A}, a_i a'_i) \Rrightarrow_{\mathrm{pH}} (\mathfrak{A}, aa'). \tag{6.3}$$

By (6.2) and (6.3) we conclude $\mathfrak{B} \Rrightarrow_{\mathrm{pH}} (\mathfrak{A}, aa')$. But these two structures are $\aleph_0$-categorical (by Ryll-Nardzewski), so Lemma 5.3 applies and there is a surjective homomorphism

$$h'' : \mathfrak{B}^{\mathrm{per}} \twoheadrightarrow (\mathfrak{A}, aa').$$

By Proposition 3.4, $\mathfrak{B}^{\mathrm{per}}$ is isomorphic to

$$\left(\mathfrak{A}^{\mathrm{per}}, \langle b_{00} \cdots b_{(k-1)(\ell_{k-1}-1)}\rangle\langle b'_{00} \cdots b'_{(k-1)(\ell_{k-1}-1)}\rangle\right),$$

so aa' is a surjective periomorphic image of the $\sum_{i<k}\ell_i$ many pairs

$$b_{00}b'_{00}, \ldots, b_{(k-1)(\ell_{k-1}-1)}b'_{(k-1)(\ell_{k-1}-1)} \in R.$$

Thus $aa' \in \tilde{R}$, as was to be shown. □

7 Equality templates

Fix a countably infinite set A and define an *equality template* to be a relational structure $\mathfrak{A}$ that is first-order definable in (A), the structure interpreting the empty language; that is, every relation of $\mathfrak{A}$ is definable by a pure equality formula. A complexity classification of the QCSPs of equality templates was given in previous work **[6]** (see Theorem 7.9 below): it was shown that each such QCSP is either in L, NP-complete or coNP-hard. In this section, we re-examine this classification theorem. Based on our Main Theorem 5.1 we give a new proof of this classification which is, in our view, more modular, conceptually cleaner, and shorter than the original proof.

7.1 Clone analysis

Our proof follows the algebraic approach to constraint complexity and thereby relies on an analysis of the polymorphism clones of equality templates. Such clones are locally closed and contain all permutations (note that every permutation of A is an automorphism of $\mathfrak{A}$). Bodirsky et al. [**8**], building on the work of Bodirsky and Kara [**12**], performed a study of these clones. Here we state only what we shall need from their analysis.

We define an operation to be *elementary* if it is contained in the smallest locally closed clone containing all permutations; a set of operations is *elementary* if each of its operations is elementary. Let us say that an operation f *generates* another operation g if g is contained in the smallest locally closed clone that contains f and all permutations of A. As an example, an operation is elementary if and only if it is generated by the identity on A. Finally, recall that an *essentially unary* operation is one that can be written as the composition of a unary operation and a projection; and an *essential* operation is one that is not essentially unary.

Lemma 7.1 (Clone analysis)

(1) *A non-elementary operation generates either a binary injective operation or a unary constant operation.*

(2) *An operation with infinite image that does not preserve $\neq$ generates all unary operations.*

(3) *Let $k \geq 3$. An essential operation with image size k generates all operations with image size at most k.*

Proof. The lemma can be derived from results in [**8, 12**] as follows. To prove (1), let f be a non-elementary operation. If f is essentially unary, then f generates a unary non-elementary operation h. The operation h is not injective, since all unary injective operations can be interpolated by permutations. By the proof of [**12**, Lemma 10], h generates a unary constant operation.

Now suppose that f is essential. By [**12**, Lemma 12], f generates an essential binary operation. By [**12**, Theorem 13], f generates either a unary constant operation or a binary injective operation.

Statement (2) follows from [**8**, Lemma 38] and statement (3) is [**8**, Lemma 36]. □

7.2 Classification

We now start the proof of the classification theorem for equality templates.

Theorem 7.2 *Let $\mathfrak{A}$ be an equality template such that $\neq$ is not positive Horn definable in $\mathfrak{A}$. Then every unary operation on A is a polymorphism of $\mathfrak{A}$.*

Proof. If $\neq$ is not positive Horn definable in $\mathfrak{A}$, then, by our Main Theorem 5.1, the relation $\neq$ is not preserved by some surjective periomorphism h of $\mathfrak{A}$. Recall that according to (4.1) with h there is a naturally associated sequence of polymorphisms $(h_{<k})_{k\geq 1}$. Because h does not preserve $\neq$, there exists k_0 such that $h_{<k_0}$ does not either. Suppose there exists some k_1 such that $h_{<k_1}$ has infinite image. Then $h_{<k_0 \cdot k_1}$ does not preserve $\neq$ and has infinite image. Then our claim follows from Lemma 7.1(2). We thus assume that all $h_{<k}$ have finite image. By local closure it suffices to show:

Claim For every $k \in \mathbb{N}$, every partial unary operation $g\colon A \to A$ that is defined on k points can be extended to a (unary) polymorphism of $\mathfrak{A}$.

We prove the claim by induction on k. For $k = 0$ there is nothing to show. Suppose that the claim is true for k and let g be a unary operation defined on $k+1$ points. If g has image size $k + 1$, then there exists a permutation g' extending g, and the claim follows; recall that all permutations are automorphisms of $\mathfrak{A}$. So suppose that g has image of size at most k.

It suffices to show that the polymorphism clone of $\mathfrak{A}$ contains a unary operation that has finite image of size $\geq k$, for this implies that the clone contains a unary operation that maps $k + 1$ points to k points; by composing this unary operation with itself and suitable permutations, one obtains the claim.

Since h has infinite image, there exists $\ell > 0$ such that $h_{<\ell}$ has image size $\geq k$. Let $\overline{a}_0, \ldots, \overline{a}_{k-1} \in A^\ell$ be k many ℓ-tuples on which $h_{<\ell}$ is injective. Assume for the sake of notation that $0, \ldots, k-1 \in A$. Consider the maps $u_0, \ldots, u_{\ell-1}$ defined on $\{0, \ldots, k-1\}$ such that u_j maps each $i < k$ to the jth component of $\overline{a}_i$. Note that $u_0(i) \cdots u_{\ell-1}(i) = \overline{a}_i$. By induction every u_j can be extended to a polymorphism u'_j of $\mathfrak{A}$. Define $u \colon A \to A$ to map $a \in A$ to $h_{<\ell}(u'_0(a), \ldots, u'_{\ell-1}(a))$. Then $u(i) = h_{<\ell}(\overline{a}_i)$ for every $i < k$, so u is injective on the set $\{0, \ldots, k-1\}$. Thus the image of u has size $\geq k$ and is finite because it is contained in the image of $h_{<\ell}$. □

The following simple lemma will be useful. It appears as Lemma 11 in [**12**]; we supply a proof for self-containment.

Lemma 7.3 *Let $\mathfrak{A}$ be an equality template. Either $\mathfrak{A}$ has a constant polymorphism, or the relation $\neq$ is primitive positively definable in $\mathfrak{A}$.*

Proof. Suppose that $\mathfrak{A}$ does not have a constant polymorphism. Then there is a relation $R^{\mathfrak{A}}$ that is non-empty and does not contain the constant tuple. Let k be the arity of $R^{\mathfrak{A}}$. Let us say that an equivalence relation σ on $\{0, \ldots, k-1\}$ is *realized* if there exists a tuple $(a_0 \ldots, a_{k-1}) \in R^{\mathfrak{A}}$ such that $a_i = a_j$ if and only if $(i, j) \in \sigma$. (Note that if there exists one tuple in $R^{\mathfrak{A}}$ satisfying the given condition, then all tuples satisfying the given condition are in $R^{\mathfrak{A}}$.) Let τ be a coarsest realized equivalence relation. Consider the relation defined in $\mathfrak{A}$ by the primitive positive formula

$$\varphi(x_0, \ldots, x_{k-1}) := Rx_0 \cdots x_{k-1} \wedge \textstyle\bigwedge_{(i,j)\in\tau} x_i = x_j;$$

in this relation, τ is realized, and it is the only equivalence relation that is realized. Since $R^{\mathfrak{A}}$ does not contain the constant tuple, τ contains more than one equivalence class. Fix $i, j \in \{0, \ldots, k-1\}$ to be values such that $(i, j) \notin \tau$. The formula $\psi(x_i, x_j)$ derived from φ by existentially quantifying all variables other than x_i and x_j defines the relation $\neq$. □

Let us say that a relation over A is *negative* if it is definable as the conjunction of (i) equalities and (ii) disjunctions of disequalities; by a disequality, we mean a formula of the form $\neg x = y$. Let us say that a relation is *positive* if it is definable using equalities and the binary connectives $\{\wedge, \vee\}$. We call an equality template *negative* or *positive* if each of its relations is negative or positive respectively.

Example 7.4 The ternary relation $P \subseteq A^3$ defined by the formula

$$\varphi_P(x, y, z) := (x = y \vee y = z)$$

in (A) is positive; it can be verified from the definition that it is not negative.

Example 7.5 The ternary relation $I \subseteq A^3$ defined by the formula

$$\varphi_I(x, y, z) := (x = y \to y = z)$$

in (A) is neither positive not negative; this can be verified from the definitions.

Positivity can be characterized algebraically as follows. This has been shown in [**6**, Proposition 7.3].

Proposition 7.6 *Let $\mathfrak{A}$ be an equality template, and fix f to be any non-injective surjective unary operation on A. The following are equivalent:*

- *$\mathfrak{A}$ is positive.*
- *Every unary operation is a polymorphism of $\mathfrak{A}$.*
- *The operation f is a polymorphism of $\mathfrak{A}$.*

We have the following fact.

Corollary 7.7

(1) *If $\mathfrak{A}$ is a positive equality template, then every positive Horn definable relation in $\mathfrak{A}$ is positive.*
(2) *If $\mathfrak{A}$ is a negative equality template, then every positive Horn definable relation in $\mathfrak{A}$ is negative.*

Proof. From Proposition 7.6 it follows that, for any fixed non-injective surjective unary operation f, a relation is positive if and only if it is preserved by f; this characterization of positivity implies (1).

Likewise, (2) follows from the fact that negativity can be characterized by preservation by a surjective operation (see [**8**, Proposition 68]). □

The following is known ([**6**, Lemma 8.8]):

Lemma 7.8 *If R is a relation over A that is not negative and is preserved by a binary injective operation, then I is primitive positively definable in $(A, R, \neq)$.*

We are ready to state and prove the classification.

Theorem 7.9 ([**6**]) *Let $\mathfrak{A}$ be an equality template.*

(1) *If $\mathfrak{A}$ is negative, then $\mathsf{QCSP}(\mathfrak{A})$ is in* L.
(2) *If $\mathfrak{A}$ is not negative but positive, then the relation P is positive Horn definable in $\mathfrak{A}$ and $\mathsf{QCSP}(\mathfrak{A})$ is* NP*-complete.*
(3) *If $\mathfrak{A}$ is neither negative nor positive, then the relation I is positive Horn definable in $\mathfrak{A}$ and $\mathsf{QCSP}(\mathfrak{A})$ is* coNP*-hard.*

Proof. We take as given the complexity results: it is shown in [**6**] that a negative template $\mathfrak{A}$ has $\mathsf{QCSP}(\mathfrak{A})$ in L, that $\mathsf{QCSP}((A, P))$ is NP-hard, and $\mathsf{QCSP}((A, I))$ is coNP-hard; and it follows from [**34**] that a positive template $\mathfrak{A}$ has $\mathsf{QCSP}(\mathfrak{A})$ in NP. By Proposition 2.2 and Corollary 7.7, it thus suffices to show that for an equality template $\mathfrak{A}$ one of the following three conditions holds:

(i) $\mathfrak{A}$ is negative.
(ii) $\mathfrak{A}$ is positive and P is positive Horn definable in $\mathfrak{A}$.
(iii) I is positive Horn definable in $\mathfrak{A}$.

Let $\mathfrak{A}$ be an equality template and let $[\mathfrak{A}]_{\mathrm{pH}}$ denote its expansion by all relations that are positive Horn definable in $\mathfrak{A}$. Further, let C denote the clone of polymorphisms of $[\mathfrak{A}]_{\mathrm{pH}}$. By Lemma 7.1(1), the following three cases are exhaustive.

Case 1: C is elementary. Then C preserves I, so this relation is primitive positively definable in $[\mathfrak{A}]_{\mathrm{pH}}$ by Theorem 2.4 and hence positive Horn definable in $\mathfrak{A}$.

Case 2: C contains a constant operation. Then $\neq$ is not contained in $[\mathfrak{A}]_{\mathrm{pH}}$, since $\neq$ is not preserved by a constant operation. Applying Theorem 7.2 to $[\mathfrak{A}]_{\mathrm{pH}}$, we obtain that C contains all unary operations. Proposition 7.6 implies that $[\mathfrak{A}]_{\mathrm{pH}}$ (and hence $\mathfrak{A}$) is positive. We claim that either $[\mathfrak{A}]_{\mathrm{pH}}$ (and hence $\mathfrak{A}$) is negative or P is positive Horn definable in $\mathfrak{A}$.

Case 2.1: Suppose that there exists a surjective periomorphism h of $\mathfrak{A}$ and a $k > 0$ such that the polymorphism $h_{<k}$ is essential. We claim that in this case C contains all operations. It is known (and easy to verify) that each relation preserved by this clone can be defined by a conjunction of equalities, so then $[\mathfrak{A}]_{\mathrm{pH}}$ will be negative. By local closure, it suffices to show that C contains all finite image operations. Hence, by Lemma 7.1(3), it also suffices to show that C contains a sequence of polymorphisms that is *desirable* in the sense that each polymorphism is essential and has finite image, and that the sequence has unbounded image size. Now, $(h_{<\ell\cdot k})_{\ell>0}$ is such a desirable sequence in case each $h_{<\ell\cdot k}$ has finite image. And otherwise there is $\ell_0 > 0$ such that $h_{<\ell_0\cdot k}$ has infinite image, and then one obtains a desirable sequence $(u_i \circ h_{<\ell_0\cdot k})_{i>0}$ for suitable unary operations u_i (recall that all unary operations are in C).

Case 2.2: Suppose otherwise that for every surjective periomorphism h and all $k > 0$ the polymorphism $h_{<k}$ is essentially unary. We claim that then the relation P is positive Horn definable in $\mathfrak{A}$. By our Main Theorem 5.1 it suffices to show that P is preserved by all surjective periomorphisms of $\mathfrak{A}$. But if a surjective periomorphism h of $\mathfrak{A}$ does not preserve P, then there exists $k > 0$ such that $h_{<k}$ does not preserve P. Since $h_{<k}$ is essentially unary, this is impossible.

Case 3: C contains a binary injective operation and does not contain a constant operation. In this case, $[\mathfrak{A}]_{\mathrm{pH}}$ contains $\neq$ by Lemma 7.3. It follows immediately from Lemma 7.8 that either $[\mathfrak{A}]_{\mathrm{pH}}$ (and hence $\mathfrak{A}$) is negative or I is primitive positively definable in $[\mathfrak{A}]_{\mathrm{pH}}$ and hence positive Horn definable in $\mathfrak{A}$. □

8 Discussion

Bing's theorem **[3]** involves a clever, technical argument that allows us to strengthen our main preservation theorem for structures that are isomorphic to their finite powers. Such structures have gained some attention in constraint complexity **[7, 10]**. We have the following theorem.

Theorem 8.1 *Let $\mathfrak{A}$ be a countable $\aleph_0$-categorical structure such that $\mathfrak{A} \cong \mathfrak{A}^2$. Then a formula $\varphi(\overline{x})$ is equivalent to a positive Horn formula in $\mathfrak{A}$ if and only if it is preserved by all surjective polymorphisms of $\mathfrak{A}$.*

Proof. Let $\mathfrak{A}$ accord the assumption of the theorem. We only prove the backward direction. Assume $\varphi(\overline{x})$ is preserved by all surjective polymorphisms of $\mathfrak{A}$. In particular, $\varphi(\overline{x})$

is preserved by all surjective homorphisms from $\mathfrak{A}$ to $\mathfrak{A}$. It is not hard to see that Lyndon's Theorem implies that there exists a positive formula $\varphi^+(\overline{x})$ such that $\varphi(\mathfrak{A}) = \varphi^+(\mathfrak{A})$ (see [**11**, Proposition 2(c)] for details). We can assume that φ^+ has the form of some quantifier prefix followed by a quantifier free formula

$$\psi = \bigwedge_{i\in I} \bigvee_{j\in J} \alpha_{ij},$$

where the α_{ij}s are atoms. For each $f \in J^I$ write

$$\psi_f := \bigwedge_{i\in I} \alpha_{if(i)}.$$

Bing's argument. Let $\overline{Q}\overline{y}$ be an arbitrary quantifier prefix. Assume for every $f \in J^I$ the tuple $\overline{a}_f$ in $\mathfrak{A}$ is an assignment to the free variables in $\overline{Q}\overline{y}\psi$ such that $\prod_{f\in J^I}(\mathfrak{A}, \overline{a}_f) \models \overline{Q}\overline{y}\psi$. Then there exists $f \in J^I$ such that $(\mathfrak{A}, \overline{a}_f) \models \overline{Q}\overline{y}\psi_f$.

Proof of Bing's argument. This can be proved by a straightforward induction on the length of $\overline{Q}\overline{y}$. See [**3**, Lemma 3] for details. ⊣

Write $\varphi^+(\overline{x}) = \overline{Q}\overline{y}\psi(\overline{y}, \overline{x})$.

Claim There exists $f \in J^I$ such that $\mathfrak{A} \models \forall\overline{x}(\varphi^+(\overline{x}) \to \overline{Q}\overline{y}\psi_f(\overline{y}, \overline{x}))$.

Proof of the claim. Otherwise we find for every $f \in J^I$ an $\overline{a}_f \in \varphi^+(\mathfrak{A})$ such that

$$(\mathfrak{A}, \overline{a}_f) \not\models \overline{Q}\overline{y}\psi_f(\overline{y}, \overline{x}).$$

Then $\prod_{f\in J^I}(\mathfrak{A}, \overline{a}_f) \not\models \varphi^+(\overline{x})$ by Bing's argument. As $\mathfrak{A} \cong \mathfrak{A}^2$, there is an isomorphism

$$h : \mathfrak{A}^{J^I} \cong \mathfrak{A}.$$

Write $\overline{x} = x_0 \cdots x_{\ell-1}$ and $\overline{a}_f = a_f^0 \cdots a_f^{\ell-1}$. Then

$$\begin{aligned} h : \textstyle\prod_{f\in J^I}(\mathfrak{A}, \overline{a}_f) &= \big(\mathfrak{A}^{J^I}, (a_f^0)_{f\in J^I}, \ldots, (a_f^{\ell-1})_{f\in J^I}\big) \\ &\cong (\mathfrak{A}, h((a_f^0)_{f\in J^I}), \ldots, h((a_f^{\ell-1})_{f\in J^I})). \end{aligned}$$

Since h is an isomorphism, $\varphi^+(\overline{x})$ is false in the right hand side structure. Hence h is (up to a renaming of indices) a surjective polymorphism of $\mathfrak{A}$ that does not preserve $\varphi(\overline{x})$, a contradiction. ⊣

Since $(\overline{Q}\overline{y}\psi_f \to \varphi^+)$ is logically valid, the Claim implies that φ^+ is equivalent in $\mathfrak{A}$ to the positive Horn formula $\overline{Q}\overline{y}\psi_f$. □

Examples 8.2 An example of a structure satisfying the assumption of the theorem is the countable atomless Boolean algebra (cf. [**5**, Section 5.2]). This template is of central importance for spatial reasoning in artificial intelligence. Another example is an infinite dimensional vector space over some finite field (cf. [**5**, Section 5.3], [**15**, Example 2.10]). More generally, it is easy to see that every countable $\aleph_0$-categorical structure $\mathfrak{A}$ whose theory is Horn axiomatizable satisfies $\mathfrak{A} \cong \mathfrak{A}^2$.

We conclude with some remarks and questions.

Very recently, Bodirsky, Hils and Martin [**10**] explored the possibilities to extend the algebraic machinery for constraint satisfaction to structures that are not necessarily $\aleph_0$-categorical; they established a variant of the preservation theorem for primitive positive definability via ω-polymorphisms for structures that are in a certain sense sufficiently saturated. (An ω-polymorphism of a structure $\mathfrak{A}$ is a homomorphism from $\mathfrak{A}^{\mathbb{N}}$ to $\mathfrak{A}$.)

The first author showed [**21**, Lemma 7.5] that, in finite structures, positive Horn definability coincides with Π_2 positive Horn definability (see [**23, 38**] for a related result). Using the method of the proof, one can infer that Boolean QCSPs with quantifier alternation rank restricted to some even $t \geq 2$ are either Π_t^{P}-complete or in P (cf. [**21**, Theorem 7.2]). An open issue is to study $\aleph_0$-categorical QCSPs with bounded alternation rank.

One can ask the following concrete question. Let $\mathfrak{A}$ be an $\aleph_0$-categorical structure and φ a Π_t formula that is preserved by the surjective periomorphisms of $\mathfrak{A}$. Is φ equivalent to a positive Horn formula that is also Π_t?

A related question is posed by Y. Chen and Flum in [**24**]. They ask for an alternation rank preserving version of Lyndon's preservation theorem: is any Π_t sentence that is preserved by surjective homomorphisms equivalent to a positive Π_t sentence? This is known to be true for $t \leq 2$ [**43**]. By a well-known trick of Lyndon [**37**] (see also Feferman's survey [**27**]) a positive answer would follow from a proof of the following: any implication between Π_t formulas has a Π_t Lyndon-interpolant. The usual argument constructs an interpolant by recursion on a cut-free proof of the given implication. But again for $t > 2$ there seems to be no control on the alternation rank of an interpolant constructed in this way.

Acknowledgements

Manuel Bodirsky and Barnaby Martin made valuable comments on an early version of this paper. The authors also thank Manuel for useful literature pointers. The first author is supported by the Spanish program "Ramón y Cajal" and MICINN grant TIN2010-20967-C04-02. The second author thanks the John Templeton Foundation for its support through Grant #13152, *The Myriad Aspects of Infinity*, and the FWF (Austrian Research Fund) for its support through Grant P 23989 - N13.

References

[1] E. Allender, M. Bauland, N. Immerman, H. Schnoor and H. Vollmer. The complexity of satisfiability problems: refining Schaefer's Theorem. *Journal of Computer and System Sciences* 75(4):245–254, 2009.

[2] L. Barto and M. Kozik. Constraint sastisfaction problems of bounded width. *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science*, pp. 595–603, 2009.

[3] K. Bing. On arithmetical classes not closed under direct union. *Proceedings of the American Mathematical Society* 6:836–846, 1955.

[4] *Constraint Satisfaction with Infinite Domains.* PhD Thesis, Humboldt Universität Berlin, 2004.

[5] M. Bodirsky. Constraint satisfaction problems with infinite templates. In N. Creignou et al. (eds.), *Complexity of Constraints – An Overview of Current Research Themes*, LNCS 5250, pp. 196–228, 2008.

[6] M. Bodirsky and H. Chen. Quantified equality constraints. *SIAM Journal on Computing* 39(8):3682–3699, 2010.

[7] M. Bodirsky, H. Chen, J. Kara and T. von Oertzen. Maximal infinite-valued constraint languages. *Theoretical Computer Science* 410: 1684–1693, 2009.

[8] M. Bodirsky, H. Chen and M. Pinsker. The reducts of equality up to primitive positive interdefinability. *The Journal of Symbolic Logic* 75(4):1249–1292, 2010.

[9] M. Bodirsky, M. Hermann and F. Richoux. Complexity of existential positive first-order logic. *Proceedings of Computability in Europe*, pp. 31–36, 2009.

[10] M. Bodirsky, M. Hils and B. Martin. On the scope of the universal-algebraic approach to constraint satisfaction. *Proceedings of the 25th IEEE Symposium on Logic in Computer Science*, 2010.

[11] M. Bodirsky and M. Junker. Aleph0-categorical structures: interpretations and endomorphisms. *Algebra Universalis* 64(3-4):403–417, 2010.

[12] M. Bodirsky and J. Kára. The complexity of equality constraint languages. *Theory of Computing Systems* 3(2):136–158, 2008.

[13] M. Bodirsky and J. Nešetřil. Constraint satisfaction with sountable homogeneous templates. *Journal of Logic and Computation* 16(3):359–373, 2006.

[14] V. G. Bodnarčuk, L. A. Kalužnin, V. N. Kotov and B. A. Romov. Galois theory for post algebras, part I and part II. *Cybernetics* 5:243–252, 531–539, 1969.

[15] F. Börner, A. Bulatov, H. Chen, P. Jeavons and A. Krokhin. The complexity of constraint satisfaction games and QCSP. *Information and Computation* 207: 923–944, 2009.

[16] A. Bulatov. A dichotomy theorem for constraints on a three-element set. *Proceedings of the 43rd IEEE Symposium on Foundations of Computer Science*, pp. 649–658, 2002.

[17] A. Bulatov, P. Jeavons and A. Krokhin. Classifying the complexity of constraints using finite algebras. *SIAM Journal on Computing* 34(3):720–742, 2005.

[18] A. K. Chandra and P. M. Merlin. Optimal implementation of conjunctive queries in relational databases. *Proceedings of the 9th Annual ACM Symposium on Theory of Computing*, pp. 77–90, 1977.

[19] C. C. Chang and H. J. Keisler. *Model Theory.* Studies in Logic and the Foundations of Mathematics 73. North-Holland Publishing Co., Amsterdam, third edition, 1990.

[20] H. Chen. The complexity of quantified constraint satisfaction: collapsibility, sink algebras, and the three-element case. *SIAM Jornal on Computing* 37(5):1674–1701, 2008.

[21] H. Chen. A rendez-vous of logic, complexity and algebra. *ACM Computing Surveys* 42(1), 2009.

[22] H. Chen. Quantified constraint satisfaction and the polynomially generated powers property. *Algebra Universalis* 65:213–241, 2011.

[23] H. Chen, F. Madelaine and B. Martin. Quantified constraints and containment problems. *Proceedings of the 23rd IEEE Symposium on Logic in Computer Science*, pp. 317–328, 2008.

[24] Y. Chen and J. Flum. The parameterized complexity of maximality and minimality problems. *Proceedings of the 2nd International Workshop on Parameterized and Exact Computation*, pp. 25–37, 2006.

[25] N. Creignou, S. Khanna and M. Sudan. Complexity Classification of Boolean Constraint Satisfaction Problems. *SIAM Monographs on Discrete Mathematics and Applications. Society for Industrial and Applied Mathematics*, 2001.

[26] T. Feder and M. Vardi. The computational structure of monotone monadic SNP and constraint satisfaction: A study through datalog and group theory. *SIAM Journal on Computing* 28:57–104, 1999.

[27] S. Feferman. Harmonious logic: Craig's interpolation theorem and its descendants. *Synthese* 164:341–357, 2008.

[28] J. Flum. First order logic and its extensions. In G. H. Müller et al. (eds.), Logic Conference Kiel 1974, *Lecture Notes in Mathematics* 499, 1975.

[29] D. Geiger. Closed systems of functions and predicates. *Pacific Journal of Mathematics* 27:95–100, 1968.

[30] P. Hell and J. Nešetřil. On the complexity of H-colouring. *Journal of Combinatorial Theory Series B* 48:92–110, 1990.

[31] P. Idziak, P. Markovic, R. McKenzie, M. Valeriote and R. Willard. Tractability and learnability arising from algebras with few subpowers. *SIAM Journal on Computing* 39(7):3023–3037, 2010.

[32] P. Jeavons. On the algebraic structure of combinatorial problems. *Theoretical Computer Science* 200:185–204, 1998.

[33] H. J. Keisler. Reduced products and Horn classes. *Transactions of the American Mathematical Society* 117:307–328, 1965.

[34] D. Kozen. Positive first-order logic is NP-complete. *IBM Journal of Research and Development* 25(4):327–332, 1981.

[35] M. Krasner. Endothéorie de Galois abstraite. *Séminaire P. Dubreil (Algèbre et Théorie des Nombres)* 1(6), 1968.

[36] B. Larose and P. Tesson. Universal algebra and hardness results for constraint satisfaction problems. *Theoretical Computer Science* 410(18):1629–1647, 2009.

[37] R. C. Lyndon. Properties preserved under homomorphism. *Pacific Journal of Mathematics* 9(1):143–154, 1959.

[38] F. Madelaine and B. Martin. The preservation properties of positive Horn logic. Manuscript, available at `www.dur.ac.uk/barnaby.martin/publications.html`, 2009.

[39] F. Madelaine and B. Martin. The complexity of positive first-order logic without equality. *Proceedings of the 24th IEEE Symposium on Logic in Computer Science*, pp. 429–438, 2009.

[40] F. Madelaine and B. Martin. A tetrachotomy for positive first-order logic without equality. *Proceedings of the 26th IEEE Symposium on Logic in Computer Science*, pp. 311–320, 2011.

[41] B. Martin. First-order model checking problems parameterized by the model. *Proceedings of Computability in Europe 2008: Logic and Theory of Algorithms*, pp. 417–427, 2008.

[42] C. H. Papadimitriou. *Computational Complexity.* Addison-Wesley, 1995.

[43] C. Ritter. *Fagin-Definierbarkeit.* Diplomarbeit, Universität Freiburg, 2005.

[44] T. J. Schaefer. The complexity of satisfiability problems. *Proceedings of the ACM Symposium on Theory of Computing*, pp. 216–226, 1978.

[45] H. Vogler. A unifying approach to theorems on preservation and interpolation for binary relations between structures. *Archive of Mathematical Logic* 21(1): 101–112, 1981.

Isomorphism relations on computable structures

Ekaterina Fokina[†], Sy-David Friedman[‡], Valentina Harizanov[§], Julia F. Knight[¶], Charles McCoy[¶], Antonio Montalbán[||]

[†] Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
efokina@logic.univie.ac.at

[‡] Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
sdf@logic.univie.ac.at

[§] Department of Mathematics, George Washington University, USA
harizanv@gwu.edu

[¶] Department of Mathematics, University of Notre Dame, USA
knight.1@nd.edu, Charles.F.McCoy.9@nd.edu

[||] Department of Mathematics, University of Chicago, USA
antonio@math.uchicago.edu

Abstract. We study the complexity of the isomorphism relation on classes of computable structures. We use the notion of FF-reducibility introduced in [**9**] to show completeness of the isomorphism relation on many familiar classes in the context of all Σ^1_1 equivalence relations on hyperarithmetical subsets of ω.

Introduction

We develop the theory for computable structures analogous to the theory of isomorphism relations introduced by H. Friedman and Stanley in [**13**]. Our languages are computable, and our structures have universes contained in ω. In measuring complexity, we identify structures with their atomic diagrams. In particular, a structure is *computable* if its atomic diagram is computable.

In descriptive set theory, the study of Borel equivalence relations under Borel reducibility has developed into a rich area. The notion of Borel reducibility allows one to compare the complexity of equivalence relations on Polish spaces; for details see, for example, [**15, 19, 21**]. In particular, natural equivalence relations such as isomorphism and bi-embeddability on classes of countable structures have been widely studied, e.g., [**13, 14, 18, 25**]. An effective version of this study was introduced in [**4**] and [**24**]. The complexity of the isomorphism relation on various classes of countable structures was measured using the idea of effective transformations. In the recent work [**11**] the general theory of effectively Borel (i.e., Δ^1_1) equivalence relations on effectively presented Polish

Appeared in the Journal of Symbolic Logic vol. 77(1) (2012), pp. 122–132.

[†,‡,§,¶,||] We express our gratitude to the John Templeton Foundation for its support through Project #13152, Myriad Aspects of Infinity.

[†,‡] Supported by the FWF through projects M 1188 - N13 and P 19898 - N18.

[‡,§,¶] Partially supported by the NSF binational grant DMS-1101123.

[§] Partially supported by the NSF grant DMS-0904101.

[||] Partially supported by the NSF grant DMS-0901169.

spaces was developed via the notion of effective Borel reducibility. The resulting structure turned out to be much more complex than in the classical case.

In computable model theory, equivalence relations have also been a subject of study, e.g., [**3, 7, 23**], etc. In these papers, equivalence relations of rather low complexity were studied (computable, in the Ershov hierarchy, Σ^0_1, Π^0_1). In [**9**] Σ^1_1 equivalence relations on computable structures were investigated. The notion of hyperarithmetical and computable reducibility of Σ^1_1 equivalence relations on ω was used to estimate the complexity of natural equivalence relations on hyperarithmetical classes of computable structures within the class of Σ^1_1 equivalence relations on hyperarithmetical subsets of ω as a whole.

In this paper we continue the study of the theory of Σ^1_1 equivalence relations on computable structures. Our work here shows that this theory behaves very differently than the theory initiated in Friedman–Stanley [**13**] for isomorphism relations and further developed for arbitrary Borel equivalence relations on Polish spaces [**15, 19, 21**]. In particular we show that isomorphism of computable graphs is complete with respect to the chosen effective reducibility in the context of *all* Σ^1_1 equivalence relations on ω. This is false in the context of countable structures and Borel reducibility [**22**]: there are examples of Borel equivalence relations that are not Borel-reducible to isomorphism of graphs. We also show that the isomorphism relation on computable torsion abelian groups is complete among Σ^1_1 equivalence relations on ω, while in the classical case it is known to be incomplete among isomorphism relations on classes of countable structures [**13**]. The same holds for isomorphism of computable torsion-free abelian groups, which in the case of countable structures is not known to be complete for isomorphism relations.

1 Background

1.1 Trees

Here we give some definitions useful for describing computable trees. Our trees are isomorphic to subtrees of $\omega^{<\omega}$. For the language, we take a single unary function symbol, interpreted as the predecessor function. We write $\emptyset$ for the top node (our trees grow down), and we think of $\emptyset$ as its own predecessor. Thus, our trees are defined on ω with their structure given by the predecessor function, but we often consider them as subtrees of $\omega^{<\omega}$ and treat their elements as finite sequences.

Definition 1.1 Let $S, T \subseteq \omega^{<\omega}$ be trees. Define the tree $S * T$ in the following way. We think of the elements of $S * T$ as ordered pairs (σ, τ), where $\sigma \in S$, $\tau \in T$. At level 0 of $S * T$, we have $(\emptyset, \emptyset)$. For an element (σ, τ) at level k of $S * T$, σ and τ are at level k of S and T, respectively. The successors of (σ, τ) are the pairs (σ', τ'), where σ' is a successor of σ in S and τ' is a successor of τ in T.

Definition 1.2 Let T be a subtree of $\omega^{<\omega}$. We define the *tree rank* of $x \in T$, denoted by $\mathrm{tr}(x)$, by induction:

(1) $\mathrm{tr}(x) = 0$ if x has no successor;
(2) For $\alpha > 0$, $\mathrm{tr}(x) = \alpha$ if α is the least ordinal greater than $\mathrm{tr}(y)$ for all successors y of x;
(3) $\mathrm{tr}(x) = \infty$ if x does not have ordinal tree rank.

The *tree rank* of the tree T is defined to be the rank of the top node $\emptyset$.

Note that all computable trees have rank ∞ or rank some computable ordinal. Moreover, for any node $x \in T$, $\mathrm{tr}(x) = \infty$ iff x extends to an infinite path through T [**27**].

Remark The tree rank of the tree $S * T$ is the minimum of the tree ranks of S and T. In particular, $S * T$ has an infinite path iff both S and T have infinite paths. More generally, for $\sigma \in S$ and $\tau \in T$, where σ and τ lie at the same level in their respective trees, $\mathrm{tr}((\sigma, \tau)) = \min(\mathrm{tr}(\sigma), \mathrm{tr}(\tau))$.

Definition 1.3 (Rank-saturated tree) A computable subtree T of $\omega^{<\omega}$ is *rank-saturated* provided that, for all x in T,

(1) if $\mathrm{tr}(x)$ is an ordinal α, then for all $\beta < \alpha$, x has infinitely many successors z such that $\mathrm{tr}(z) = \beta$;

(2) if $\mathrm{tr}(x) = \infty$, then for all computable β, x has infinitely many successors z such that $\mathrm{tr}(z) = \beta$ and x has infinitely many successors z with $\mathrm{tr}(z) = \infty$.

Lemma 1.4 *There is a computable rank-saturated tree T^∞ such that $rk(T^\infty) = \infty$.*

Proof. In [**17**] Harrison proved the existence of a computable linear ordering $\mathcal{H}$ of type $\omega_1^{\mathrm{CK}}(1 + \eta)$. We let T^∞ be the set of finite sequences $((a_0, k_0), \ldots, (a_n, k_n))$, where $a_0 > \cdots > a_n$ in $\mathcal{H}$ and $k_0, \ldots, k_n \in \omega$. It is easy to see that if a_i corresponds to an ordinal α in $\mathcal{H}$, then $\mathrm{tr}((a_0, k_0), \ldots, (a_i, k_i)) = \alpha$, and if a_i lies in the non-well-ordered part of $\mathcal{H}$, then $\mathrm{tr}((a_0, k_0), \ldots, (a_i, k_i)) = \infty$. □

Proposition 1.5 *If T is a computable tree, then $T * T^\infty$ is a computable rank-saturated tree of the same tree rank as T.*

Proof. The top node in $T*T^\infty$ clearly has the proper rank, by Remark 1.1. For $x \in T*T^\infty$ of rank α and $\beta < \alpha$, we show that x has infinitely many successors of rank β. Say $x = (\sigma, \tau)$; by Remark 1.1, $\mathrm{tr}(\tau) \geq \alpha$ and because T^∞ is rank-saturated, τ has infinitely many successors τ' of rank β. Also, $\mathrm{tr}(\sigma) \geq \alpha$, so σ has a successor σ' of rank at least β. Then for all such pairs (σ', τ'), $\mathrm{tr}(\sigma', \tau') = \beta$. □

Remark Computable rank-saturated trees are a special case of computable *rank-homogeneous* trees, defined in [**5**].

Proposition 1.6

(1) *For every computable α, if T^α and T_1^α are computable rank-saturated trees of tree rank α, then $T^\alpha \cong T_1^\alpha$.*

(2) *If T_1^∞ is a computable rank-saturated tree of tree rank ∞, then $T^\infty \cong T_1^\infty$.*

Proof. By induction on α. □

We will fix the notation T^α for the computable rank-saturated tree of rank α, and we recall that T^∞ is a computable rank-saturated tree with infinite paths.

1.2 Σ_1^1 sets and relations

We assume that the reader is familiar with basic concepts of recursion theory. However, here we list some definitions and facts that will be useful for the future proofs. Detailed information can be found, for example, in [**1, 27**].

Definition 1.7

(1) A relation $S(\overline{x})$ is Σ_1^1 if there is an arithmetical relation $R(\overline{x}, u)$, on tuples of numbers, such that $\overline{x} \in S$ iff $(\exists f \in \omega^\omega)\,(\forall s)\, R(\overline{x}, f \upharpoonright s)$ —we identify $f \upharpoonright s$ with its code.

(2) A relation $S(\overline{x})$ is Π^1_1 if there is an arithmetical relation $R(\overline{x}, u)$, on tuples of numbers, such that $\overline{x} \in S$ iff $(\forall f \in \omega^\omega)\,(\exists s)\,R(\overline{x}, f \upharpoonright s)$.

(3) A relation $S(\overline{x})$ is Δ^1_1 if it is both Σ^1_1 and Π^1_1.

By the Kleene–Suslin Theorem, a relation is Δ^1_1 iff it is hyperarithmetical.

If $S(\overline{x})$ is a k-place relation, we may consider the set S' of codes for k-tuples belonging to S. It is clear that S is Σ^1_1 iff S' is Σ^1_1. The next result gives familiar conditions equivalent to being Σ^1_1 [**1, 27**]. We identify finite sequences with their codes.

Proposition 1.8 (Kleene) *The following are equivalent:*

(1) *S is Σ^1_1.*

(2) *There is a computable relation $R(n, u)$, on pairs of numbers, such that $n \in S$ iff $(\exists f)\,(\forall s)\,R(n, f \upharpoonright s)$.*

(3) *There is a computable sequence of computable trees $(T_n)_{n\in\omega}$ such that $n \in\ S$ iff T_n has an infinite path.*

Theorem 1.9 (Bounding) *Let CWF denote the set of codes for computable well-founded trees on ω and, for each computable ordinal α, let CWF_α denote the set of codes for computable trees of tree rank less than α. Then if F is a hyperarithmetical function from a hyperarithmetical subset of ω into CWF, there exists a computable α such that the range of F is contained in CWF_α.*

We now give a notion of effective reducibility of Σ^1_1 equivalence relations on hyperarithmetical subsets of ω. The idea is the following. A relation E is effectively reducible to a relation E' if there is an effective procedure which allows us to answer any question about E-equivalence using information about E'-equivalence. We want to use partial computable functions as witnesses for reducibilities.

Definition 1.10 Let E, E' be Σ^1_1 equivalence relations on hyperarithmetical subsets $X, Y \subseteq \omega$, respectively. The relation E is *FF-reducible* to E' iff there exists a partial computable function f with $X \subseteq \mathrm{dom}(f)$, $Y \subseteq f(X)$ such that, for all $x, y \in X$,

$$xEy \iff f(x)E'f(y).$$

We denote this fact by $E \leq_{FF} E'$.

The notion of FF-reducibility was first used in [**9**] where it was called "tc-reducibility". In the next section we will explain the relationship between FF-reducibility and the notion of tc-reducibility introduced in [**4**] to compare the classes of countable structures.

1.3 Computable characterization and classification

Here we review two equivalent approaches, from [**16**], to the problems of computable characterization and classification. The goal is to be able to measure the complexity of a set of computable structures or an equivalence relation on a set of computable structures.

The first approach is based on the notion of computable infinitary formulas. Roughly speaking, computable infinitary formulas are $L_{\omega_1\omega}$ formulas in which the infinite disjunctions and conjunctions are over c.e. sets. For a formal definition see [**1**]. Computable infinitary formulas form a hierarchy: a *computable* Σ_0 or Π_0 formula is a finitary quantifier-free formula. For $\alpha > 0$, a *computable* Σ_α formula is a c.e. disjunction of formulas of the form $\exists\overline{u}\,\psi$, where ψ is computable Π_β for some $\beta < \alpha$, and a *computable* Π_α formula is a c.e. conjunction of formulas of the form $\forall\overline{u}\,\psi$, where ψ is computable Σ_β for some $\beta < \alpha$.

Following [**16**], we say that a class K of structures closed under isomorphism *has a computable characterization* if the set K^c of its computable members consists exactly of all computable models of a computable infinitary sentence. This definition expresses the idea that the set of all computable members of K can be nicely defined among all other structures for the same language.

The second approach uses the notion of an index set. For a computable structure $\mathcal{M}$, an *index* is a number a such that $\varphi_a = \chi_{D(\mathcal{M})}$, where $(\varphi_a)_{a\in\omega}$ is a computable enumeration of all unary partial computable functions. The *index set* for $\mathcal{M}$ is the set $I(\mathcal{M})$ of all indices for computable (isomorphic) copies of $\mathcal{M}$. For a class K of structures, closed under isomorphism, the *index set* is the set $I(K)$ of all indices for computable members of K. As in [**16**], we say that a class K *has a computable characterization* if its index set is hyperarithmetical.

Proposition 1.11 (Goncharov–Knight [**16**]) *Let K be a class of countable structures closed under isomorphism, and let K^c be the set of computable members of K. Then the following are equivalent:*

(1) *The index set $I(K)$ of K is hyperarithmetical.*

(2) *There is a computable infinitary sentence ψ such that $K^c = \mathrm{Mod}^c_\psi$, where Mod^c_ψ is the set of all computable models of ψ.*

For a relation E on a class K of structures, denote by $I(E, K)$ the set of pairs of indices

$$\{(m,n) \mid m,n \in I(K) \text{ and } \mathcal{M}_m E \mathcal{M}_n\}.$$

We measure the complexity of various relations on computable structures via the complexity of the corresponding sets of pairs of indices. In what follows we will often identify E with $I(E, K)$ considered as a relation on indices. Thus, it will make sense to compare relations on classes of computable structures with relations on subsets of ω. The most studied cases are that of isomorphism and bi-embeddability relations, e.g., [**2, 6, 9, 16**].

We are interested in studying the relations on classes that are nicely defined. For this reason we will require the index set of each class K to be hyperarithmetical. Equivalently, $K^c = \mathrm{Mod}^c_\psi$ for some computable infinitary ψ. Let K and K' be two classes of *countable* structures, such that $K = \mathrm{Mod}_\psi$ and $K' = \mathrm{Mod}_{\psi'}$ for some computable infinitary ψ, ψ'. Suppose the isomorphism relation on K is tc-reducible to the isomorphism relation on K' in the sense of [**4**]. Then $I(\cong, K) \leq_{FF} I(\cong, K')$ and the reduction is exactly the restriction to computable structures of the reduction of K to K'.

2 Isomorphism is complete among Σ^1_1 equivalence relations

If $I(K)$ is hyperarithmetical and E is the isomorphism or bi-embeddability relation, then the corresponding equivalence relation $I(E, K)$ on indices is a Σ^1_1 set. In this section we prove completeness of the isomorphism relation on various familiar classes of structures in the context of all Σ^1_1 equivalence relations on hyperarithmetical subsets of ω under FF-reducibility. These results show the difference of our theory from the classical theory of Borel equivalence relations since, by [**22**], some Borel equivalence relations cannot be reduced to isomorphism relations.

Definition 2.1 A relation E on a hyperarithmetical subset of ω is an *FF-complete Σ^1_1 equivalence relation* if E is Σ^1_1 and every Σ^1_1 equivalence relation E' on a hyperarithmetical subset of ω is FF-reducible to E.

Note that an equivalence relation E on a hyperarithmetical class K^c of computable structures is complete if and only if for every Σ^1_1 relation E' there exists a computable sequence of computable structures $(\mathcal{M}_n)_{n\in\omega}$ from K^c such that, for all $m, n \in \omega$,

$$mE'n \iff \mathcal{M}_m E \mathcal{M}_n.$$

2.1 Trees and graphs

Theorem 2.2 *The isomorphism relation on computable trees is an FF-complete Σ^1_1 equivalence relation.*

Proof. Let E be a Σ^1_1 equivalence relation on ω. To prove that E is FF-reducible to the isomorphism relation on computable trees, we will build a computable sequence of computable trees $(T_n)_{n\in\omega}$ such that, for every $m, n \in \omega$,

$$mEn \iff T_m \cong T_n.$$

By Proposition 1.8, since E is Σ^1_1, there exists a uniformly computable sequence of trees $(T_{m,n})_{m,n\in\omega}$ such that $\neg mEn$ if and only if $T_{m,n}$ is well founded. Then we say that *$\neg mEn$ is witnessed by stage α* if and only if $T_{m,n}$ has tree-rank less than α.

The strategy to build $(T_n)_{n\in\omega}$ is the following. First, uniformly in m, n, we will build a computable tree $T^*_{m,n}$ with the following properties:

(1) $T^*_{m,n} \cong T^*_{n,m}$;

(2) $mEn \Rightarrow T^*_{m,n} \cong T^\infty$, where T^∞ is the rank-saturated tree with an infinite path;

(3) $\neg mEn \Rightarrow T^*_{m,n} \cong T^\alpha$, where T^α is the rank-saturated tree of tree rank α, for a computable ordinal α such that for all $m' \in [m]_E$ and $n' \in [n]_E$ the relation $\neg m'En'$ is witnessed by stage α. This α will be the least ordinal such that for all $m' \in [m]_E$, $n' \in [n]_E$ and all finite sequences $a_0 = m', a_1, \ldots, a_s = n'$, $\alpha \geq \min\{\mathrm{tr}(T_{a_i, a_{i+1}}) : i \leq s-1\} + 1$, where $(T_{m,n})_{m,n\in\omega}$ is the sequence fixed for E in the previous paragraph.

We start from the computable sequence of computable trees $(T_{m,n})_{m,n\in\omega}$ mentioned above: $T_{m,n}$ is well founded if and only if $\neg mEn$. For every $m, n \in \omega$, we construct (effectively and uniformly) a new tree $T'_{m,n}$ in the following way. Let $\sigma_0, \sigma_1, \ldots$ be an enumeration of all finite sequences of natural numbers. Suppose $\sigma_s = (a_0, \ldots, a_{l_s})$. Then under the s-th node on level 1 (i.e., under the element of the form (s), $s \in \omega$) of $T'_{m,n}$ we put the tree $P_s = T_{m,a_0} * T_{a_0,a_1} * \cdots * T_{a_{l_s},n}$, identifying the top node of P_s with s. Then

$$\mathrm{tr}(T'_{m,n}) = \sup\{\mathrm{tr}(P_s) + 1 \mid s \in \omega\}.$$

If mEn, then $T_{m,n}$ has an infinite path, i.e., $\mathrm{tr}(T_{m,n}) = \infty$. Thus, $\mathrm{tr}(T'_{m,n}) = \infty$. If $\neg mEn$, then for every $\sigma = (a_0, \ldots, a_l)$, $\mathrm{tr}(T_{m,a_0} * T_{a_0,a_1} * \cdots * T_{a_l,n})$ is a computable ordinal. Indeed, fix $m, n \in \omega$ such that $\neg mEn$. For every finite sequence σ_s consider the corresponding tree $P_s = T_{m,a_0} * T_{a_0,a_1} * \cdots * T_{a_{l_s},n}$. Consider the function F from the set of finite sequences into CWF such that $F(s)$ is the code of P_s. The function F is hyperarithmetical, its domain is computable. By Bounding, there is a computable bound on the range of F. Therefore, $T'_{m,n}$ has rank α for some computable α. Note that for all $m' \in [m]_E$ and $n' \in [n]_E$, we get the same bound α. Indeed, let $m'Em, n'En$ and let β be the computable bound on the ranks of trees constructed using finite sequences starting with m' and ending with n'. Let $P_s = T_{m,a_0} * T_{a_0,a_1} * \cdots * T_{a_{l_s},n}$ be as above. Then $\mathrm{tr}(T_{m',m} * P_s * T_{n,n'}) = \mathrm{tr}(P_s)$, thus $\alpha \leq \beta$. Similarly, one can show that $\beta \leq \alpha$.

Let $T^*_{m,n} = T'_{m,n} * T^\infty$. As shown in Proposition 1.5, the tree $T^*_{m,n}$ is a computable rank-saturated tree, $\mathrm{tr}(T^*_{m,n}) = \mathrm{tr}(T'_{m,n})$, and the construction is uniform.

Now we build the desired sequence $(T_n)_{n\in\omega}$. Take the tree T consisting exactly of the sequences $(m, m, \dots, m)$ of length $i \le m$, for $m \in \omega$. Now fix n and, for every m, attach $T^*_{m,n}$ to the m-th leaf of T. The resulting tree is T_n. The sequence $(T_n)_{n\in\omega}$ witnesses the reducibility: mEn iff $T_m \cong T_n$. Indeed, suppose mEn. Then

(1) for every $k \in [m]_E = [n]_E$, $\mathrm{tr}(T'_{k,m}) = \mathrm{tr}(T'_{k,n}) = \infty$, thus $T^*_{k,m} \cong T^*_{k,n} \cong T^\infty$;

(2) for every $k \notin [m]_E$, $\mathrm{tr}(T'_{k,m}) = \mathrm{tr}(T'_{k,n}) = \alpha$, thus $T^*_{k,m} \cong T^*_{k,n} \cong T^\alpha$.

Therefore, $T_m \cong T_n$.

Suppose now that $\neg mEn$. Then $T^*_{m,m} \cong T^\infty$, while $T^*_{m,n} \cong T^\alpha$ for some computable α. Thus $T_m \not\cong T_n$. □

Corollary 2.3 *The isomorphism relation on computable graphs is an FF-complete Σ^1_1 equivalence relation.*

2.2 Torsion-free abelian groups

Torsion-free abelian groups are subgroups of $\mathbb{Q}$-vector spaces. Hjorth [**18**] gave a transformation from trees to torsion-free abelian groups, which enabled him to show that the isomorphism relation on these groups is not Borel. Downey and Montalbán [**8**] built on Hjorth's ideas to show that the isomorphism problem on these groups is complete among Σ^1_1 *sets*. In this paper we use the transformation from [**18**] and [**8**] to show that the isomorphism relation on computable torsion-free abelian groups is, in fact, complete as a Σ^1_1 equivalence relation. First we describe the transformation.

We consider the elements of $\omega^{<\omega}$ as a basis for a $\mathbb{Q}$-vector space V^*. Let T be a subtree of $\omega^{<\omega}$, and let V be the subspace of V^* with basis T. Let T_n be the set of elements at level n of T. If u is at level $n > 0$, let u^- be the predecessor of u. Let $(p_n)_{n\in\omega}$ be a computable list of distinct primes. We let $G(T)$ be the subgroup of V generated by the vector space elements of the following forms:

(1) $v/(p_{2n})^k$, where $v \in T_n$, and $k \in \omega$;

(2) $(v + v')/(p_{2n+1})^k$, where $v \in T_n$, v' is a successor of v, and $k \in \omega$.

Theorem 2.4 *The isomorphism relation on computable torsion-free abelian groups is FF-complete among Σ^1_1 equivalence relations.*

Proof. It follows from [**12**] that if we restrict the class of trees to only rank-saturated trees, then the transformation from the class of trees into torsion-free abelian groups described above is 1-1 on isomorphism types. Thus, given a Σ^1_1 equivalence relation E for every $n \in \omega$, we first construct the sequence of rank-saturated trees $(T^*_{m,n})_{m\in\omega}$ as in Theorem 2.2. We want to pass effectively from the sequence to a group G_n such that $G_n \cong G_{n'}$ iff for all m, $T^*_{m,n} \cong T^*_{m,n'}$.

For $m \in \omega$, let $(p_{m,k})_{k\in\omega}$ be uniformly computable lists of primes such that, for distinct m, the lists are disjoint. For each m, we apply the transformation described above, taking $T_{m,n}$ to a torsion-free abelian group $G_{m,n}$, using the list of primes $(p_{m,k})_{k\in\omega}$. The resulting sequence $(G_{m,n})_{n\in\omega}$ will satisfy the property

$$T^*_{m,n} \cong T^*_{m',n'} \iff G_{m,n} \cong G_{m',n'}.$$

Let $G_n = \oplus_m G_{m,n}$.

Using the fact that the sequences of primes are disjoint, we can see that $G_n \cong G_{n'}$ iff for all m, $G_{m,n} \cong G_{m,n'}$. The reason is that $G_{m,n}$ is the subgroup of G_n generated by the set of elements divisible by all the powers of some prime in the list $(p_{m,k})_{k\in\omega}$ (for more details, see [**8**] or [**12**]). □

2.3 Abelian *p*-groups

Let p be a prime number. A *p-group* is a group such that each element has some power of p for its order. Countable abelian p-groups are classified up to isomorphism in terms of Ulm invariants (see [**20**] for details).

In this section we use the transformation from trees into abelian p-groups to get completeness of the isomorphism relation for this class. Note that in the classical theory of Borel equivalence relations the analogous result is false (see [**13**] and a proof for Turing computable embeddings in [**12**]).

Theorem 2.5 *The isomorphism relation on abelian p-groups is an FF-complete Σ^1_1 equivalence relation.*

Proof. By Theorem 2.2, for any Σ^1_1 equivalence relation E on ω, we have a uniformly computable sequence of trees $(T_n)_{n\in\omega}$ such that mEn iff $T_m \cong T_n$. Each tree T_n is the result of combining a family of trees $T^*_{m,n}$. Each $T^*_{m,n}$ is rank-saturated, so it is really determined by its tree rank. We may modify our trees, if necessary, so that the tree rank, if it exists, is a limit ordinal.

Let $\mathcal{T} = (T^m)_{m\in\omega}$ be a sequence of rank-saturated trees. We need a transformation taking such sequences $\mathcal{T}$ to abelian p-groups $G(\mathcal{T})$, such that $G(\mathcal{T}) \cong G(\mathcal{T}')$ iff the sequences of ranks for the trees in $\mathcal{T}$ and $\mathcal{T}'$ match. We replace T^m by a tree T^m_* such that each single successor in T^m becomes a chain of p_m successors in T^m_*. Then $\mathrm{tr}(T^m_*) = p_m \mathrm{tr}(T^m)$. We form a single tree with infinitely many nodes at level 1, with a copy of T^0_* below the first, a copy of T^1_* below the second, etc. Denote the resulting tree by T. Let G be the abelian p-group generated by the elements of T in a standard way [**20**]: the top node is the identity, and if x' is a successor of x, then $px' = x$.

Rogers [**28**] described how to calculate (non-effectively, of course) the Ulm sequence for G from the tree ranks of elements in the corresponding tree T. We describe her scheme briefly. For each node of successor rank, apart from the top node, we choose a successor witnessing the rank. Now, for each α, $u_G(\alpha)$ is the number of nodes of rank α that are not chosen as witnesses. In computing $u_G(\alpha)$, we count all x at level 1 such that $\mathrm{tr}(x) = \alpha$. Suppose x is an element at level $n > 1$, where $\mathrm{tr}(x) = \alpha$. Let y be the predecessor of x. If $\mathrm{tr}(y) > \alpha + 1$, then x cannot witness the rank of y, so we count x. If $\mathrm{tr}(y) = \alpha + 1$, then x may be the chosen successor of y witnessing the rank. We count x just in case it is not chosen.

Using Rogers' scheme, we can see that our group G has the following features. For all computable α, the Ulm invariant $u_\alpha(G)$ is either ∞ or 0. For limit α, $u_\alpha(G) = 0$. If $\alpha = \omega\beta + p_m$, then $u_\alpha(G) = \infty$ iff $\mathrm{tr}(T^m) \geq \omega\beta$. □

Corollary 2.6 *The isomorphism relation on torsion abelian groups is an FF-complete Σ^1_1 equivalence relation.*

Suppose K and K' are classes of countable structures, with universe a subset of ω, closed under isomorphism. We write $K \leq_{tc} K'$ if there is a Turing computable operator $\Phi = \varphi_e$ taking the atomic diagram of each $\mathcal{A} \in K$ to the atomic diagram of some $\mathcal{B} \in K'$, such that Φ is 1-1 on isomorphism types. This notion was introduced in [**4**]. If $I(K)$ and $I(K')$ are hyperarithmetical, and $K \leq_{tc} K'$, then $I(\cong, K) \leq_{FF} I(\cong, K')$. If Φ is the computable operator reducing the isomorphism relation on structures in K to that on structures in K', then for computable $\mathcal{A} \in K$ we can effectively compute an index for $\Phi(\mathcal{A})$ from an index for $\mathcal{A}$.

H. Friedman and Stanley [**13**] introduced the study of Borel reductions $\leq_B$ of isomorphism relations on classes of structures with universe ω. They showed that the class of undirected graphs, the class of fields of any fixed characteristic, the class of 2-step nilpotent groups, and the class of linear orderings all lie "on top" in this setting. In [**4**], it was observed that the Borel transformations are all effective. Moreover, the transformations work perfectly well for structures with universe an arbitrary subset of ω. Therefore, these classes are also "on top" under the relation $\leq_{tc}$ in [**4**]. We have shown that, for the class K of trees, the relation $I(E, K)$ (the set of pairs of indices for computable members of K that are isomorphic) lies "on top" under the relation $\leq_{FF}$ on Σ^1_1 equivalence relations on ω. From this, we immediately get the following.

Theorem 2.7 *For each of the following classes K, $I(E, K)$ is an FF-complete Σ^1_1 equivalence relation:*

- *undirected graphs;*
- *fields of characteristic* 0 *or p;*
- 2*-step nilpotent groups;*
- *linear orderings.*

3 Open problems

In [**9**] equivalence relations were compared not only via FF-reducibility but also via hyperarithmetical reducibility (h-reducibility):

Definition 3.1 Let E, E' be Σ^1_1 equivalence relations on hyperarithmetical subsets $X, Y \subseteq \omega$, respectively. The relation E is *h-reducible* to E' iff there exists a hyperarithmetical function f such that, for all $x, y \in X$,

$$xEy \iff f(x)E'f(y).$$

By [**14**] the following theorem is true for the bi-embeddability relation on computable structures. Here we mean the standard model-theoretic notion of embeddings on structures.

Theorem 3.2 *For every Σ^1_1 equivalence relation E on ω there exists a hyperarithmetical class K of structures which is closed under isomorphism and such that E is h-equivalent to the bi-embeddability relation on computable structures from K.*

Remark 3.4 of [**14**] provides the result for Σ^1_1 preorders on the reals, but the result for preorders on ω follows almost immediately.

In [**10**] it was proved that the general structure of Σ^1_1 equivalence relations on hyperarithmetical subsets of ω (under FF- or h-reducibility) is rich. The above theorem states that the structure of bi-embeddability relations on hyperarithmetical classes of computable structures is as complex as the whole structure of Σ^1_1 equivalence relations under h-reducibility. It would be interesting to get the following refinement of Theorem 3.2:

Question 3.3 If E is a Σ^1_1 equivalence relation on ω, does there exist a hyperarithmetical class K of structures, closed under isomorphism and such that E is FF-equivalent to the bi-embeddability relation on computable structures from K?

Let K be a class of structures closed under isomorphism such that the index set $I(K)$ is hyperarithmetical. Consider the following statements:

(1) $I(\cong, K)$ is properly Σ^1_1;
(2) $I(\cong, K)$ is m-complete Σ^1_1;
(3) $I(\cong, K)$ is Σ^1_1 complete under FF-reducibility;
(4) $I(\cong, K \restriction \text{highSR})$ is not hyperarithmetical within $K \restriction \text{highSR}$, where highSR is the class of structures of high (i.e., noncomputable) Scott rank;
(5) K has infinitely many non-isomorphic computable structures of high Scott rank.

The following implications are true: (1) $\Leftarrow$ (2) $\Leftarrow$ (3) $\Rightarrow$ (4) $\Rightarrow$ (5).

Question 3.4 Which of these arrows are reversible?

One of the approaches to give a negative answer to the question "(1) $\Rightarrow$ (3)?" would be to positively answer the following:

Question 3.5 Is there a hyperarithmetical class of structures with a unique (up to isomorphism) computable structure of high Scott rank?

If the answer to this question is positive, we see immediately that (1) does not imply (5). Since (3) implies (5), we also conclude that (1) does not imply (3).

Remark It is known that up to bi-embeddability this is true in the following sense. In the class of computable linear orderings, the equivalence class of linear orderings bi-embeddable with the rationals is Σ^1_1-complete, but every computable scattered linear ordering (i.e., not bi-embeddable with the rationals) has a hyperarithmetical equivalence class. For more information on the bi-embeddability relation in the class of countable linear orderings see [**26**].

This question may be also considered as a weaker version of the question from [**16**] where the authors asked about the existence of a computable structure with high Scott rank and a hyperarithmetical index set.

Question 3.6 Are there isomorphism relations on hyperarithmetical classes of computable structures which are not hyperarithmetical and not FF-complete?

References

[1] C. J. Ash, J. F. Knight, *Computable Structures and the Hyperarithmetical Hierarchy*, Elsevier, 2000.

[2] W. Calvert, *Algebraic structure and computable structure*, PhD Dissertation, University of Notre Dame, 2005.

[3] W. Calvert, D. Cenzer, V. Harizanov, A. Morozov, *Effective categoricity of equivalence structures*, Ann. Pure Appl. Logic 141 (2006), 61–78.

[4] W. Calvert, D. Cummins, J. F. Knight, S. Miller, *Comparing classes of finite structures*, Algebra and Logic 43 (2004), 374–392.

[5] W. Calvert, J. Knight, J. Millar, *Computable trees of Scott rank ω_1^{CK} and computable approximations*, J. Symb. Logic 71 (2006), 283–298.

[6] J. Carson, E. Fokina, V. S. Harizanov, J. F. Knight, S. Quinn, C. Safranski, J. Wallbaum, *Computable embedding problem*, submitted.

[7] D. Cenzer, V. Harizanov, J. Remmel, *Σ^0_1 and Π^0_1 equivalence structures*, Proceedings of "Computability in Europe 2009", Heidelberg, Germany, Lecture Notes in Computer Science 5635, 99–108, 2009.

[8] R. Downey, A. Montalbán, *The isomorphism problem for torsion-free Abelian groups is analytic complete*, J. Algebra 320 (2008), 2291–2300.

[9] E. Fokina, S. Friedman, *Equivalence relations on classes of computable structures*, Proceedings of "Computability in Europe 2009", Heidelberg, Germany, Lecture Notes in Computer Science 5635, 198–207, 2009.

[10] E. Fokina, S. Friedman, Σ^1_1 *equivalence relations on* ω, submitted.

[11] E. Fokina, S. Friedman, A. Törnquist, *The effective theory of Borel equivalence relations*, Ann. Pure Appl. Logic 161 (2010), 837–850.

[12] E. Fokina, J. Knight, A. Melnikov, S. Quinn, C. Safranski, *Ulm type, and coding rank-homogeneous trees in other structures*, to appear in J. Symb. Logic.

[13] H. Friedman, L. Stanley, *A Borel reducibility theory for classes of countable structures*, J. Symb. Logic 54 (1989), 894–914.

[14] S. D. Friedman, L. Motto Ros, *Analytic equivalence relations and bi-embeddability*, to appear in J. Symb. Logic.

[15] S. Gao, *Invariant Descriptive Set Theory*, Pure and Applied Mathematics, CRC Press/Chapman & Hall, 2009.

[16] S. S. Goncharov, J. F. Knight, *Computable structure and non-structure theorems*, Algebra and Logic 41 (2002), 351–373 (English translation).

[17] J. Harrison, *Recursive pseudo well-orderings*, Trans. Amer. Math. Soc. 131 (1968), 526–543.

[18] G. Hjorth, *The isomorphism relation on countable torsion-free Abelian groups*, Fund. Math. 175 (2002), 241–257.

[19] V. Kanovei, *Borel Equivalence Relations. Structure and Classification*, University Lecture Series 44, American Mathematical Society, 2008.

[20] I. Kaplansky, *Infinite Abelian Groups*, University of Michigan Press, Ann Arbor, 1954.

[21] A. Kechris, *New directions in descriptive set theory*, Bull. Symb. Logic 5 (1999), 2, 161–174.

[22] A. Kechris, A. Louveau, *The classification of hypersmooth Borel equivalence relations*, J. Amer. Math. Soc. 10 (1997), 1, 215–242.

[23] B. Khoussainov, F. Stephan, Y. Yang, *Computable categoricity and the Ershov hierarchy*, Ann. Pure Appl. Logic 156 (2008), 86–95.

[24] J. F. Knight, S. Miller (Quinn), M. Vanden Boom, *Turing computable embeddings*, J. Symb. Logic 73 (2007), 901–918.

[25] A. Louveau, C. Rosendal, *Complete analytic equivalence relations*, Trans. Amer. Math. Soc. 357 (2005), 12, 4839–4866.

[26] A. Montalbán, *On the equimorphism types of linear orderings*, Bull. Symb. Logic 13 (2007), 71–99.

[27] H. Rogers, *Theory of Recursive Functions and Effective Computability*, McGraw-Hill, 1967.

[28] L. Rogers, *Ulm's theorem for partially ordered structures related to simply presented Abelian p-groups*, Trans. Amer. Math. Soc. 227 (1977), 333–343.

Classes of structures with universe a subset of ω_1

Ekaterina Fokina[†], **Sy-David Friedman**[†], **Julia F. Knight**[‡],
Russell Miller[§], **Antonio Montalbán**[¶]

[†] Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
`efokina@logic.univie.ac.at, sdf@logic.univie.ac.at`

[‡] Department of Mathematics, University of Notre Dame, USA
`Julia.F.Knight.1@nd.edu`

[§] Department of Mathematics, Queens College, City University of New York, USA
`Russell.Miller@qc.cuny.edu`

[¶] Department of Mathematics, University of Chicago, USA
`antonio@math.uchicago.edu`

Abstract. In this paper, we add to the collection of recent results on computable structure theory in the setting of ω_1. In the standard setting, there are results comparing the classification problems for different classes of countable structures [**10**]. In [**8**], it is shown that Σ^1_1 equivalence relations on ω are reducible under an effective relation $\leq_{FF}$, to isomorphism of certain classes of computable structures, where the structures are identified with their indices. In the present paper, we lift this result to ω_1. In the standard setting, there are many results on computable categoricity and relative computable categoricity. In particular, there are results saying which structures of various familiar kinds are relatively computably categorical and which are not even computably categorical. In the setting of ω_1, the real number field and the complex number field are both relatively computably categorical. The present paper gives some results on categoricity for some further uncountable fields.

1 Introduction

We begin by summarizing some definitions and results from [**3, 12**] on computability in ω_1. We assume that all subsets of ω are constructible. In certain places, we assume that all subsets of ω_1 are constructible. The basic definitions come from *α-recursion theory*, where $\alpha = \omega_1$.

Definition 1.1

- A set or relation on ω_1 is *computably enumerable*, or *c.e.*, if it is defined in $(L_{\omega_1}, \in)$ by a Σ_1 formula $\varphi(\bar{c}, x)$ with finitely many parameters —a Σ_1 formula is finitary, with only existential and bounded quantifiers, occurring only positively.
- A set or relation is *computable* if it and its complement are both computably enumerable.
- A (partial) function is *computable* if its graph is c.e.

Results of Gödel provide a 1-1 function g from ω_1 onto L_{ω_1} such that the relation $g(\alpha) \in g(\beta)$ is computable. There is also a computable function ℓ taking α to the code for L_α. The function g gives ordinal codes for sets, so that computing on ω_1 is really the

The authors wish to thank the John Templeton Foundation for its generous support of this research through the CRM Infinity Project (Grant ID 13152).

same as computing on L_{ω_1}. We may allow relations and functions of arity α, where α is any countable ordinal.

As in the standard setting, we have indices for c.e. sets. There is a c.e. set C of codes for pairs $(\varphi, \overline{c})$, representing Σ_1 definitions —$\varphi(\overline{u}, x)$ is a Σ_1 formula and $\overline{c}$ is a tuple of parameters appropriate for $\overline{u}$. We have a computable function h mapping ω_1 onto C. Then the ordinal α is a *c.e. index* for the set X if $h(\alpha)$ is the code for a pair $(\varphi, \overline{c})$, where $\varphi(\overline{c}, x)$ is a Σ_1 definition of X in $(L_{\omega_1}, \in)$. We write W_α for the c.e. set with index α.

Suppose W_α is determined by the pair $(\varphi, \overline{c})$; i.e., $\varphi(\overline{c}, x)$ is a Σ_1 definition. We say that *x is in W_α at stage β*, and we write $x \in W_{\alpha,\beta}$, if L_β contains x, the parameters $\overline{c}$, and witnesses making the formula $\varphi(\overline{c}, x)$ true. The relation $x \in W_{\alpha,\beta}$ is computable. Let $U \subseteq (\omega_1)^2$ consist of the pairs (α, β) such that $\beta \in W_\alpha$. Then U is m-complete c.e. It is not computable, since the "halting set" $K = \{\alpha : \alpha \in W_\alpha\}$ is c.e. and not computable.

We define relative computability as follows.

Definition 1.2

- A relation is *c.e. relative to X* if it is Σ_1-definable in $(L_{\omega_1}, \in, X)$.
- A relation is *computable relative to X* if it and its complement are both c.e. relative to X.
- A (partial or total) function is *computable relative to X* if the graph is c.e. relative to X.

A *c.e. index* for R relative to X is an ordinal α such that $g(h(\alpha)) = (\varphi, \overline{c})$, where φ is a Σ_1 formula (in the language with $\in$ and a predicate symbol for X), and $\varphi(\overline{c}, x)$ defines R in $(L_{\omega_1}, \in, X)$. We write W_α^X for the c.e. set with index α relative to X. As in the standard setting, we have a universal c.e. set of partial computations using oracle information. Let U consist of the codes for triples (σ, α, β) such that $\sigma \in 2^\rho$ (for some countable ordinal ρ), and for X with characteristic function extending σ, $\beta \in W_\alpha^X$. Then U is c.e.

Definition 1.3 The *jump* of X is $X' = \{\alpha : \alpha \in W_\alpha^X\}$.

We can iterate the jump function through countable levels. For this, we let $X^{(0)} = X$, $X^{(\alpha+1)} = (X^{(\alpha)})'$, and for limit α, $X^{(\alpha)}$ is the set of codes for pairs (β, x) such that $\beta < \alpha$ and $x \in X^{(\beta)}$.

1.1 Computable structures

We consider structures with universe a subset of ω_1. As in the standard setting, we identify a structure with its atomic diagram. The ordered field of reals has a computable copy with universe ω_1. In fact, if we think of the reals as a subset of L_{ω_1}, where each real is identified with a rational cut, then the field of real numbers itself is a computable structure. The field of complex numbers is also computable. We may add to $\mathbb{R}$ a total analytic function such as exp, or sin, and the resulting expansion is still computable. To see this, we note that the analytic function is determined by the countable sequence of coefficients of a power series.

In the standard setting, Metakides and Nerode [**13**] showed that there is a computable infinite-dimensional $\mathbb{Q}$-vector space with no infinite c.e. linearly independent set, while any computable $\mathbb{Z}_p$-vector space has a computable basis. In the setting of ω_1, there is a computable $\mathbb{R}$-vector space with no uncountable c.e. independent set. By contrast, any $\mathbb{Q}$-vector space has a computable basis.

One of the earliest results in computable structure theory says that there is a computable field F of characteristic 0 with no "splitting algorithm"; that is, the set of irreducible polynomials over F is not computable. The idea, due to Van der Waerden [**17**], and made more precise by Fröhlich and Shepherdson [**11**], is to put into F a primitive (p_n)-th root of unity iff $n \in K$. Thus, for any $F' \cong F$, K is computable relative to the set of polynomials that are irreducible over F'. Note that the set K is coded in the isomorphism type of F. Hirschfeldt gave an analogous result in the setting of ω_1, showing that there is a computable field F of characteristic 0 with a uniform procedure which, for any $G \cong F$, computes K from G and the set of polynomials that are irreducible over G. Hirschfeldt's construction uses ideas of H. Friedman and Stanley [**10**].

In the standard setting, Morley [**16**] and Millar [**14**] showed that for any countable complete decidable elementary first order theory T, there is a decidable saturated model iff there is a computable enumeration of the complete types consistent with T. In the setting of ω_1, we have the following.

Proposition 1.4 *For any countable complete elementary first order theory T (with infinite models), T has a decidable saturated model with universe ω_1.*

In the standard setting, Nurtazin showed that there is a "computable numbering" of the computable linear orderings; i.e., there is a uniformly computable sequence of linear orderings representing all computable order types at least once. The same is true in the setting of ω_1. In the standard setting, the first non-computable ordinal, ω_1^{CK}, is the next admissible ordinal after ω. In the setting of ω_1, the first non-computable ordinal comes much before the next admissible after ω_1, as the set of codes for computable wellorderings is definable over L_{ω_1}. In the standard setting, the *Harrison ordering* is a computable ordering of type $\omega_1^{CK}(1+\eta)$. This ordering has initial segments isomorphic to all computable well orderings. In the setting of ω_1, we have the following.

Theorem 1.5 (Greenberg–Knight–Shore) *There is a computable ordering $\mathcal{H}$ with initial segments isomorphic to all computable ordinals.*

We take a uniformly computable list of linear orderings, representing all computable isomorphism types, and carry out a finite-injury priority construction to produce $\mathcal{H}$ with an initial segment that is a sum of intervals representing the well ordered $\mathcal{A}_\alpha$, in order, followed by various other intervals that are not well ordered.

Definition 1.6 Let $\mathcal{A}$ be a computable structure, and let R be a relation on $\mathcal{A}$.

- R is *relatively intrinsically c.e. on $\mathcal{A}$* if for all $\mathcal{B} \cong \mathcal{A}$ the image of R is c.e. relative to $\mathcal{B}$.
- R is *intrinsically c.e. on $\mathcal{A}$* if for all computable $\mathcal{B} \cong \mathcal{A}$ the image of R is c.e.

In [**1**], [**4**], it is shown in the standard setting that for a relation R on a computable structure $\mathcal{A}$, for any computable $\alpha \geq 1$, R is relatively intrinsically Σ^0_α iff it is definable in $\mathcal{A}$ by a "computable Σ_α" formula with a finite tuple of parameters. In particular, R is relatively intrinsically c.e. if it is defined by a c.e. disjunction of finitary existential formulas, with a finite tuple of parameters. In [**12**] the result is lifted to the setting of ω_1. Here the computable Σ_1 formula has a countable tuple of parameters, and each disjunct has a countable block of existential variables followed by a quantifier-free formula of $L_{\omega_1\omega}$. In [**3**], there is a definition of the arithmetical hierarchy for the setting of ω_1, extending through the countable ordinals (coinciding with the levels of Σ_n definability over the $L_{\omega_1+\alpha}$, α countable), and there is a corresponding definition of "computable Σ_α

formula". It is shown that a relation on a computable structure is relatively intrinsically Σ_α iff it has a computable Σ_α definition, with a countable tuple of parameters.

Definition 1.7 Let $\mathcal{A}$ be a computable structure.

- $\mathcal{A}$ is *relatively computably categorical* if, for all copies $\mathcal{B}$, there is an isomorphism from $\mathcal{A}$ onto $\mathcal{B}$ which is computable relative to $\mathcal{B}$.
- $\mathcal{A}$ is *computably categorical* if, for all computable copies $\mathcal{B}$, there is a computable isomorphism from $\mathcal{A}$ onto $\mathcal{B}$.

In the standard setting, there is a syntactical characterization of relative computable categoricity [**1, 4**]. A structure is relatively computably categorical iff there is a formally c.e. Scott family, where this is a c.e. set Φ of existential formulas, with a fixed finite tuple of parameters $\bar{c}$, such that each tuple in $\mathcal{A}$ satisfies some formula in Φ, and any two tuples satisfying the same formula in Φ are automorphic. In the setting of ω_1, the field of real numbers and the field of complex numbers are both relatively computably categorical. Jesse Johnson, in work for his Ph.D. thesis, has some results on computable categoricity. Two simple examples that are not computably categorical are the ordering of type $\eta \cdot \omega_1$ and the equivalence structure with $\aleph_1$ classes, all of size $\aleph_0$. Johnson has shown that the "Zilber field" of size $\aleph_1$ is not computably categorical. Associated with each Zilber field is a "cover", and Johnson has shown that the cover associated with the Zilber field is computably categorical. In [**12**], there is a characterization of the structures that are relatively computably categorical, involving a "continuous", formally c.e. Scott family.

In the standard setting, there is work comparing classification problems for various classes of countable structures. Friedman and Stanley [**10**] considered structures with universe ω, and Borel classes, closed under isomorphism. If K, K' are two such classes, then $K \leq_B K'$ if there is a Borel function $\Phi\colon K \to K'$ such that, for $\mathcal{A}, \mathcal{B} \in K$, $\mathcal{A} \cong \mathcal{B}$ iff $\Phi(\mathcal{A}) \cong \Phi(\mathcal{B})$. Among the classes that lie on top under $\leq_B$ are linear orderings and fields (of any desired characteristic). In [**2**], we consider effective variants of the notions from [**10**]. In [**7**] there are some results comparing Σ^1_1 equivalence relations on ω under a relation $\leq_{FF}$. Andrea Sorbi pointed out that the relation $\leq_{FF}$ has been denoted simply by $\leq$ by people working on the theory of "numberings". The notion goes back to early work of Ershov [**6**] in this setting.

Definition 1.8 For Σ^1_1 equivalence relations E, E' on ω, $E \leq_{FF} E'$ if there is a computable function f such that aEb iff $f(a)E'f(b)$.

If we identify computable members of a class K with their indices, and the index set $I(K)$ is hyperarithmetical, then the isomorphism relation becomes a Σ^1_1 equivalence relation on numbers. We add a class for the numbers that are not indices for computable elements of K. The following result, for the standard setting, is proved in [**8**].

Theorem 1.9 (Fokina–Friedman–Harizanov–Knight–McCoy–Montalbán) *For every Σ^1_1 equivalence relation E on ω, there exists a uniformly computable sequence of structures $(\mathcal{A}_n)_{n\in\omega}$ such that mEn iff $\mathcal{A}_m \cong \mathcal{A}_n$.*

In Section 2, we lift this result to the setting of ω_1. In Section 3, we give some results on computable categoricity of fields in the setting of ω_1.

2 Equivalence relations

We can show, in the setting of ω_1, that all Σ^1_1 sets $S \subseteq \omega_1$ are m-reducible to the isomorphism relation on computable subtrees of $\omega_1^{<\omega_1}$. In fact, there is a particular tree T such that for any Σ^1_1 set S there is a uniformly computable sequence of trees $(T_\alpha)_{\alpha<\omega_1}$ such that $\alpha \in S$ iff $T_\alpha \cong T$. We begin with the following analogue of Kleene normal form.

Lemma 2.1 *For any Σ^1_1 set $S \subseteq \omega_1$, there is a uniformly computable sequence $(T_\alpha)_{\alpha<\omega_1}$ of subtrees of $\omega_1^{<\omega_1}$ such that $\alpha \in S$ iff T_α has an ω_1-branch.*

Next, we define the special tree T. There is just one node $\emptyset$ at level 0. This node has $\aleph_1$ successors. For each node above level 0, there are $\aleph_1$ copies. Half of the copies are terminal, while the other half have $\aleph_1$ successors. We think of T as a set of functions σ from countable ordinals to $\omega_1 \times \{0,1,2\}$ such that if σ has last term $(\beta,0)$, then σ is terminal, and if σ has limit length α, with terms $(\beta,1)$ for arbitrarily large $\beta < \alpha$, then σ is also terminal. The elements of T are the sequences σ mapping countable ordinals α to $\omega_1 \times \{0,1,2\}$ such that if there is a term $(\beta,0)$, then σ has length $\beta+1$, and if there are infinitely many terms $(\beta_i,1)$ and $\beta = \sup\{\beta_i\}$, then σ has length β.

Definition 2.2 Let T_1, T_2 be subtrees of $\omega_1^{<\omega_1}$. Then $T_1^* T_2$ is the subtree of $(\omega_1 \times \omega_1)^{<\omega_1}$ consisting of the functions τ such that for some $\sigma_1 \in T_1$ and $\sigma_2 \in T_2$, both of length α, τ has length α and for all $\beta < \alpha$, $\tau(\beta) = (\sigma_1(\beta), \sigma_2(\beta))$.

It is easy to see that $T_1^* T_2$ has an ω_1-branch iff T_1 and T_2 each have an ω_1-branch.

Lemma 2.3 *For any tree $R \subseteq \omega_1^{<\omega_1}$, if R has an ω_1-branch, then $R^*T \cong T$, and if R has no ω_1-branch, then R^*T also has no ω_1-branch.*

Combining the two lemmas, we get the following.

Theorem 2.4 *For any Σ^1_1 set $S \subseteq \omega_1$, there is a uniformly computable sequence of trees $(T_\alpha)_{\alpha<\omega_1}$ such that $\alpha \in S$ iff $T_\alpha \cong T$.*

2.1 Passing from graphs to fields, linear orderings

Here we consider arbitrary structures with universe a subset of ω_1, not just computable structures. We write $K \leq_{tc} K'$ if there is a computable operator Φ taking structures in K to structures in K' such that, for $\mathcal{A}, \mathcal{B} \in K$, $\mathcal{A} \cong \mathcal{B}$ iff $\Phi(\mathcal{A}) \cong \Phi(\mathcal{B})$.

Proposition 2.5 *If K is the class of undirected graphs, and K' is the class of fields of characteristic* 0 *(or any other desired characteristic), then $K \leq_{tc} K'$.*

Proof. We use the Friedman–Stanley embedding of undirected graphs in fields of the fixed characteristic (see [**10**]). □

Proposition 2.6 *If K is the class of undirected graphs and K' is the class of linear orderings, then $K \leq_{tc} K'$.*

Idea. We use the analogue of the Friedman–Stanley embedding [**10**], replacing $\mathbb{Q}$ by the saturated model of the theory of dense linear orderings without endpoints of cardinality $\aleph_1$, and replacing the finite sequences by sequences of arbitrary countable ordinal length. Instead of finite discrete sets to code atomic types, we use sets having the order type of countable ordinals. □

We get a linear ordering L (on ω_1) that resembles the Harrison ordering in the following way.

Proposition 2.7 *There is a linear ordering L such that, for any Σ^1_1 set S, there is a uniformly computable sequence $(L_\alpha)_{\alpha<\omega_1}$ such that $\alpha \in S$ iff $L_\alpha \cong L$.*

2.2 Main result

Theorem 2.8 *Assume $V = L$. For any Σ^1_1 equivalence relation E on ω_1, there is a uniformly computable sequence of structures $M^*(\alpha)_{\alpha<\omega_1}$ (with universe ω_1) such that $\alpha E \beta$ iff $M^*(\alpha) \cong M^*(\beta)$.*

The structures $M^*(\alpha)$ will not be members of any familiar class. Each structure will code a sequence of sets $(X_\beta)_{\beta<\omega_1}$, up to an equivalence relation $\sim$ defined as follows.

Definition 2.9 For $X, Y \subseteq \omega_1$, $X \sim Y$ iff $X \Delta Y$ is not stationary.

Lemma 2.10 *For any Σ^1_1 set $X \subseteq \omega_1$, there is a uniformly computable sequence $(S_\alpha)_{\alpha<\omega_1}$ of subsets of ω_1 such that $\alpha \in X$ iff S_α contains a club.*

Proof. Let $(T_\alpha)_{\alpha<\omega_1}$ be a uniformly computable sequence of trees resulting from applying Theorem 2.4 to X; thus $\alpha \in X$ iff T_α has an ω_1 branch. Let S_α be the set of countable ordinals β such that, for some countable $\gamma > \beta$,

(1) $L_\gamma \models ZF^-$;

(2) $\omega_1^{L_\gamma} = \beta$;

(3) $T_\alpha^{L_\gamma}$ has a branch of length β in L_γ, where $T_\alpha^{L_\gamma}$ is the tree that in L_γ satisfies the definition of T_α. (Note that this tree is independent of the choice of γ satisfying 1 and 2.)

First, suppose that T_α has an ω_1-branch b. We must show that S_α contains a club. Choose $\gamma > \omega_1$ such that $b \in L_\gamma$ and $L_\gamma \models ZF^-$. We form a continuous elementary chain of models $M_0 \prec M_1 \prec \ldots \prec L_\gamma$ with $\alpha, b \in M_0$. Let $c = \{\beta_i : \beta_i = \omega_1^{M_i}\}$. We take the transitive collapse $\pi_i(M_i) = \overline{M}_i$. Then $\overline{M}_i = L_{\gamma_i}$ for some γ_i, and $\beta_i = \omega_1^{L_{\gamma_i}}$. Then $\pi_i(b)$ is a β_i-branch through $T_\alpha^{L_{\gamma_i}}$. We may suppose that the β_i are strictly increasing, $\beta_i \cap \omega_1 \subseteq M_{i+1}$, and for limit i, β_i is the sup of the β_j for $j < i$. Then c is the required club.

We must show that if T_α has no ω_1 branch, then S_α does not contain a club. Let c be a club and choose a limit ordinal $\gamma > \omega_2$ such that $c \in L_\gamma$. In L_γ, T_α (or $T_\alpha^{L_\gamma}$) has no ω_1-branch. Let M be the Skolem hull of c, α, ω_1 in L_γ. Again, we take the transitive collapse $\pi(M) = \overline{M} = L_{\overline{\gamma}}$. We have $\beta = M \cap \omega_1 = \omega_1^{\overline{M}} \in c$. We can see that $\beta \notin S_\alpha$, since in $L_{\overline{\gamma}+\omega}$ we have that β is countable. And in $L_{\overline{\gamma}}$, the tree $T_\alpha^{L_{\overline{\gamma}}}$ has no β branch, using the isomorphism π. It follows that no γ can witness that β belongs to S_α. □

Let E be a Σ^1_1 equivalence relation on ω_1. We identify pairs of ordinals with single ordinals and let S be as above, so that $\alpha E \beta$ iff $S_{\alpha,\beta}$ contains a club. For any $X \subseteq \omega_1$, let $L(X)$ be the $\aleph_1$-like linear order formed by stacking ω_1 many copies of the rational order and at limit stage α putting in a supremum iff $\alpha \in X$.

Lemma 2.11 *For $X, Y \subseteq \omega_1$, $L(X) \cong L(Y)$ iff $X \sim Y$.*

Now, we use the trick from [**8**]. For any finite chain $c = (\alpha, \gamma_1, \gamma_2, \ldots, \gamma_n, \beta)$, let $S^*(\alpha, \beta)$ consist of the sets of the form

$$S(c) = S_{\alpha,\gamma_1} \cap S_{\gamma_1,\gamma_2} \cap \ldots \cap S_{\gamma_n,\beta}.$$

If $\alpha' E \alpha$, then $S_{\alpha',\alpha}$ contains a club. Therefore, for each finite chain c from α to β, $S_{\alpha',\alpha} \cap S(c) \sim S(c)$. It follows that if we define $S^*(\alpha, \beta)$ to be the set of the $S(c)$ where c is a chain starting with α and ending with β, and $\alpha E \alpha'$, then $S^*(\alpha, \beta)$ agrees with $S^*(\alpha', \beta)$, in the sense that they have the same elements modulo the ideal of nonstationary sets.

Let $M(\alpha, \beta)$ be the structure that is the "free union" of ω_1 copies of the linear orders $L(X)$ for $X \in S^*(\alpha, \beta)$. One way to make this precise is to let $M(\alpha, \beta)$ consist of two disjoint sets A, B of size ω_1, with a relation $R(a, b_0, b_1)$ for a in A and b_0, b_1 in B so that, for each fixed a, $R(a, -, -)$ defines a linear order of B isomorphic to one of the $L(X)$, for $X \in S^*(\alpha, \beta)$, and each such order occurs for exactly ω_1-many such a in A. Alternatively, we may let $M(\alpha, \beta)$ have equivalence relation with an ordering on each equivalence class, so that, for each set $X \in S^*(\alpha, \beta)$, the ordering $L(X)$ is copied in uncountably many equivalence classes, and for each equivalence class, the ordering on the equivalence class is isomorphic to $L(X)$ for some $X \in S^*(\alpha, \beta)$.

Lemma 2.12

(1) *If $\alpha E \alpha'$, then $M(\alpha, \beta) \cong M(\alpha', \beta)$ for all β.*
(2) *If NOT $\alpha E \alpha'$, then $M(\alpha, \alpha) \ncong M(\alpha', \alpha)$.*

Proof. For (2), we note that if NOT $\alpha E \alpha'$, then there is no set $X \in S^*(\alpha, \alpha')$ that contains a club, but there is such a set in $S^*(\alpha, \alpha)$. From this, it follows that $M(\alpha, \alpha)$ is not isomorphic to $M(\alpha, \alpha')$. □

Finally, let $M^*(\alpha)$ be the *sequence* (not the free union) of the structures $M(\alpha, \beta)$, for $\beta < \omega_1$.

Lemma 2.13 *For all α, α', $\alpha E \alpha'$ iff $M^*(\alpha) \cong M^*(\alpha')$.*

Proof. If $\alpha E \alpha'$, then $M(\alpha, \beta) \cong M(\alpha', \beta)$ for all β. Therefore, $M^*(\alpha) \cong M^*(\alpha')$. If NOT $\alpha E \alpha'$, then $M(\alpha, \alpha) \ncong M(\alpha', \alpha)$. Therefore, $M^*(\alpha) \ncong M^*(\alpha')$. □

3 Results on fields

Here we consider arbitrary ω_1-computable fields of characteristic 0. The domain of the field is either ω_1 or possibly just ω, and the field operations are all ω_1-computable. We believe that our results carry over equally well to fields of positive characteristic.

Lemma 3.1 *Every ω_1-computable field has a computable transcendence basis over its prime subfield $\mathbb{Q}$. ($\mathbb{Q}$ itself, being countable, is also ω_1-computable.)*

Proof. For each $\alpha \in F$ we define $\alpha \in B$ iff

$$\begin{gathered}(\forall \langle \beta_1, \ldots, \beta_n \rangle \in \alpha^{<\omega})(\forall p \in \mathbb{Q}[X_1, \ldots, X_n, Y]) \\ [p(\beta_1, \ldots, \beta_n, \alpha) = 0 \rightarrow p(\beta_1, \ldots, \beta_n, Y) = 0].\end{gathered}$$

This statement quantifies only over countable sets which we can enumerate uniformly and know when we have finished enumerating each one. It says that α lies in B iff α satisfies no nonzero polynomial over the subfield $\mathbb{Q}(\beta : \beta < \alpha)$ generated by all elements $< \alpha$. Clearly this B is a transcendence basis for F. □

Using a finite-time algorithm given by Kronecker, we will quickly infer the ω_1-computable categoricity of the field of complex numbers. First, here is the lemma.

Lemma 3.2 *In the context of finite-time computation, let F be a computable field which is a purely transcendental extension of $\mathbb{Q}$ of infinite transcendence degree. If F has a computable transcendence basis B, then F has a splitting algorithm. (That is, reducibility of polynomials in $F[X]$ is decidable.)*

Proof. The work of Kronecker showed that $\mathbb{Q}$ itself has a splitting algorithm, and that, whenever a computable field E has a splitting algorithm and x is transcendental over E within some larger computable field, then the subfield $E(x)$ also has a splitting algorithm, uniformly in the splitting algorithm for E. (For a modern explanation of the details, see [**5**].) Write the computable basis $B = \{b_0 < b_1 < \cdots\}$. Now, given any polynomial $p(X) \in F[X]$, search for a finite set $B_0 = \{b_{\alpha_i} : i < n\}$ such that all coefficients of p are algebraic over $\mathbb{Q}(B_0)$. By Kronecker, $\mathbb{Q}(B_0)$ has a splitting algorithm (uniformly in B_0), which we can use to determine whether p factors over $\mathbb{Q}(B_0)$. However, since every monic factor of p is of the form $(X - r_1)(X - r_2) \cdots (X - r_n)$ for some roots $r_1, \ldots, r_n$ of p and all such roots are algebraic over $\mathbb{Q}(B_0)$, we see that every monic factor of p has all coefficients algebraic over $\mathbb{Q}(B_0)$ as well. Therefore, p factors over F iff it factors over $\mathbb{Q}(B_0)$. □

Corollary 3.3 *The field $\mathbb{C}$ is relatively ω_1-computably categorical.*

Proof. Given any two ω_1-computable fields $E \cong F \cong \mathbb{C}$, fix the computable transcendence bases B for E and C for F described in Lemma 3.1. Let f be the unique bijection from B onto C preserving order (viewing the field elements as ordinals in ω_1). This f is ω_1-computable and extends effectively to an isomorphism from E onto F: go through each element $x \in E$ in their order as ordinals, determine whether $x \in B$ (in which case $f(x)$ is already defined), and if not, find the minimal polynomial $p(X)$ of x over the subfield E_x generated by all elements $< x$. (Using Lemma 3.2, we can find the minimal polynomial of x over $\mathbb{Q}(B \cap E_x)$, and then we simply adjoin each $y < x$ to this subfield, one at a time, and check at each step whether the former minimal polynomial of x factors over the new subfield. Thus, after countably many steps, we have the desired $p(X)$.) Then map x to the least root in F of the image of $p(X)$ in $F[X]$ under the map f on the coefficients of $p(X)$. By normality of F over $\mathbb{Q}(C)$, at every step this map still extends to an isomorphism from E into F, so we always find such a root in F. Moreover, since f maps B onto the transcendence basis C for F, f must map E onto all of F: every $y \in F$ has a minimal polynomial $p(X) \in \mathbb{Q}(C)[X]$ of some degree d, and the roots $x_1, \ldots, x_d$ of its preimage in $E[X]$ must map one-to-one to the d-many roots of $p(X)$ in F, forcing $y \in \mathrm{rg}(f)$.

The foregoing proof relativizes to the degree of any field $E \cong \mathbb{C}$, yielding relative ω_1-computable categoricity. □

Theorem 3.4 *Let F be any ω_1-computable field with a subfield K isomorphic to $\mathbb{C}$, and assume that F is countably generated over K. Then F is relatively ω_1-computably categorical.*

Proof. Say $F = K(C)$, where C is countable and $C \cap K = \emptyset$. In general K will not be computable. Notice that then each $c \in F - K$ must be transcendental over K, and so F cannot contain the algebraic closure of $K(c)$, because this field is not countably generated over K. So we may fix a countable set $S \subseteq K$ with the property that, for every $c \in C$, S contains some tuple $x_0, \ldots, x_n$ such that F does not contain the algebraic closure of the set $\{x_0, \ldots, x_n, c\}$. Therefore, an arbitrary element $x \in F$ lies in K iff K contains the algebraic closure of $S \cup \{x\}$. Since $S \cup \{x\}$ is countable, we will recognize at some countable stage that F contains this algebraic closure (if indeed $x \in F$). Therefore, K is computably enumerable within F.

We next define a subfield F_0 of K as follows. Write $C = \{c_1, c_2, \ldots\}$. For each $i > 0$, if c_i is transcendental over the field $K(c_1, \ldots, c_{i-1})$, then add nothing to F_0; otherwise, add to F_0 a finite set of elements $y_1, \ldots, y_n$ from K such that the minimal polynomial of c_i over $K(c_1, \ldots, c_{i-1})$ has coefficients in $\mathbb{Q}(y_1, \ldots, y_n, c_1, \ldots, c_{i-1})$. Since C is countable, this only adds countably many elements in all, and we let $F_0 \subseteq K$ be the algebraic closure of the subfield of K generated by these elements along with the elements of S. Thus F_0 is also countable.

We claim that every automorphism of K which fixes F_0 pointwise extends to an automorphism of F which is the identity on C. To see this, let h_0 be such an automorphism of K. Define h to extend h_0 by setting $h(c_i) = c_i$ for all i. We claim that this h extends to an automorphism of all of F. For each $s > 0$, if c_s is transcendental over $K(c_1, \ldots, c_{s-1})$, then it is clear that setting $h_s(c_s) = c_s$ extends to an automorphism h_s of $K(c_1, \ldots, c_s)$. If c_s is algebraic over $K(c_1, \ldots, c_{s-1})$, then by our choice of F_0, the minimal polynomial $p(X)$ of c_s over all of $K(c_1, \ldots, c_{s-1})$ lies in $F_0(c_1, \ldots, c_{s-1})[X]$. So when we apply h_{s-1} to the coefficients of p, we just get p itself, and therefore, in defining $h_s(c_s) = c_s$, we are mapping c_s to a root of the image (under h_{s-1}) of its own minimal polynomial, and so h_s is again seen to be an automorphism. Thus, the union h of all these h_s is an automorphism of $K(C)$, which is to say, of F.

Now let E be any field isomorphic to F, with domain ω_1, and suppose ρ is a noncomputable isomorphism from F onto E. We give the details for the case where E is computable; they relativize directly to an arbitrary E. Let E_0 be the countable image $\rho(F_0)$, and let $T = \rho(S)$. We start be defining $f_0 = \rho \upharpoonright (F_0(C))$, which is computable because F_0 and C are countable. Next, we enumerate K as defined above, and similarly enumerate its image $\rho(K)$ within E, using the set T. At stage $\sigma + 1$, we wait for a new element x to appear in K (using our enumeration) on which f_σ is not defined. When this happens, we find the first element y to appear in our enumeration of $\rho(K)$ which is not already in $\mathrm{rg}(f_\sigma)$, and define $f_{\sigma+1}(x) = y$. At this stage we also find all elements of F which are algebraic over the portion of K which has appeared so far (including x) but not in the domain of f_σ, and define $f_{\sigma+1}$ of each of these to be a root of the corresponding polynomial in E. (We can do this effectively, simply enumerating F until all of the countably many polynomials over this portion of K have their full complement of roots.) Thus we extend the domain of $f_{\sigma+1}$ to include a larger algebraically closed subfield $K_{\sigma+1}$ of K than previously. In addition, we extend $f_{\sigma+1}$ to have the appropriate values on all elements generated by C over $K_{\sigma+1}$; again, it is not difficult to find all of these elements in countably many steps. This completes stage $\sigma + 1$.

It is clear that this defines a map $f = \cup f_\sigma$ on all of F, whose restriction to K is an isomorphism from K onto $\rho(K)$ (because we always chose the next new element of $\rho(K)$ in our enumeration to be the image of the next new element of K). The map f is also defined and equal to ρ on C (as well as on F_0), and is defined on all elements generated by C over K as well. That is, f is defined on all of F. Since ρ and f are equal on F_0, our argument above shows that the automorphism $(\rho^{-1} \circ f) \upharpoonright K$ of K extends to an automorphism τ of all of F, which is the identity on C. But then $f = \rho \circ \tau$, so f is an isomorphism from F onto E, as desired. □

At the other extreme from algebraically closed fields, namely fields purely transcendental over $\mathbb{Q}$, the opposite result holds.

Proposition 3.5 *The purely transcendental field extension $F = \mathbb{Q}(X_\alpha : \alpha \in \omega_1)$ is not ω_1-computably categorical.*

Proof. We may assume that F is a presentation with transcendence basis $\{X_\alpha : \alpha < \omega_1\}$ computable. (Lemma 3.1 only guarantees the existence of some computable transcendence basis, not necessarily of one generating the entire field.) We build a computable field $E \cong F$ with no computable isomorphism from E onto F. Then X_α will be our witness that the computable function φ_α is not such an isomorphism.

At the start, we build E_0 to be F itself, although we only use half the elements of ω_1 to do so. (Let E_0 be the isomorphic image of F under the map $\lambda+n \mapsto \lambda+2n$ for all limit ordinals λ.) We write $y_\alpha \in E_0$ for the image of x_α under this map. Then, for each α, we wait for $\varphi_\alpha(y_\alpha)$ to converge, say to some $z_\alpha \in F$. When this happens, we find $\beta_1, \ldots, \beta_n$ such that $z_\alpha \in \mathbb{Q}(x_{\beta_1}, \ldots, x_{\beta_n})$, and ask whether the polynomial $p(X) = X^2 - z_\alpha$ factors over the subfield $\mathbb{Q}(x_{\beta_1}, \ldots, x_{\beta_n})$. (Kronecker gives a splitting algorithm for this field, since we know x_{β_i} to be algebraically independent over $\mathbb{Q}$.) If so, then z_α has a square root in F, and so we do not change anything in E, but define $y'_\alpha = y_\alpha$. If not, then we adjoin to E a new element y'_α whose square in E is y_α, and use half of the currently unused elements to close E under the field operations. This completes the construction.

Now $E = \mathbb{Q}(y'_\alpha : \alpha < \omega_1)$ is isomorphic to F via the map $y'_\alpha \mapsto x_\alpha$. However, if $\varphi_\alpha(y_\alpha)\downarrow$, then y_α has a square root in E iff $\varphi_\alpha(y_\alpha)$ has no square root in F. Thus no φ_α can be an isomorphism from E onto F. □

Finally, we believe that we can exploit the fact that $2^\omega \geq \omega_1$ to produce an ω_1-computably categorical field of transcendence degree ω_1 which is very far from algebraically closed. The proof, which still needs to be checked, mixes the technique of the Friedman–Stanley embedding from [**10**] with the Miller–Schoutens idea of "tagging" transcendentals, described in [**15**].

Conjecture 3.6 There exists an ω_1-computable, relatively ω_1-computably categorical field of uncountable transcendence degree which does not even contain a copy of the algebraic closure $\overline{\mathbb{Q}}$, let alone the closure of any of its transcendentals.

Proof. The idea of the proof is as follows. We start with the field generated over $\mathbb{Q}$ by an uncountable transcendence basis $B = \{x_n : n \in \omega\} \cup \{x_\alpha : \alpha < \omega_1\}$. The countably many elements z_n in this basis are used to identify the other elements x_α, using an ω_1-computable bijective function $h : \omega_1 \to 2^\omega$. (This function does not actually need to be onto the power set 2^ω; an injective computable function would suffice. So we do not require **CH** here.)

To complete the construction of F, we adjoin certain elements algebraic over the basis B. First, for every $\alpha < \omega_1$, we adjoin an element y_α satisfying $x_\alpha^5 + y_\alpha^5 = 1$, the Fermat equation of degree 5. It must be checked (by commutative algebraists) that this does not cause any solutions of $X^5 + Y^5 = 1$ to appear in F except the six solutions generated by each pair (x_α, y_α), and the two pre-existing solutions $(0, 1)$ and $(1, 0)$. (Details appear in [**15**].) Assuming this is so, we define t_α, for each α, to be the sum

$$t_\alpha = x_\alpha + y_\alpha + \frac{1}{y_\alpha} - \frac{x_\alpha}{y_\alpha} + \frac{1}{x_\alpha} - \frac{y_\alpha}{x_\alpha}$$

of the six elements used for these six pairs. (Since the Fermat polynomial is symmetric, the solutions actually consist of three distinct pairs, involving these six elements, along with the three pairs symmetric to these.) Thus the set $\{t_\alpha : \alpha < \omega_1\}$ is intrinsically computably enumerable, and since it forms a transcendence basis for F, it is in fact intrinsically computable. (Every computably enumerable transcendence basis for any computable field is computable.) Now we adjoin the following set of elements to F, similar to the process invented by Friedman and Stanley:

$$\{\sqrt{z_{2n} + t_\alpha} : n \in h(\alpha)\} \cup \{\sqrt{z_{2n+1} + t_\alpha} : n \notin h(\alpha)\}.$$

As shown in [**10**], this adjoinment does not cause any other square roots of this form to appear in F. We must check that it also does not cause any more solutions of the Fermat polynomial to appear there (which seems reasonable, since all these extensions have degree 2). Assuming these facts, the set T of elements t_α remains intrinsically computable in this field F.

Therefore, in any ω_1-computable field E isomorphic to F via some (noncomputable) function ρ, we may compute the image $\rho(T)$ of the set of all elements t_α. We may also take the countable function $n \mapsto \rho(z_n)$ as given. Then, for each $\alpha < \omega_1$, we then search for an element $t \in \rho(T)$ such that, for every $n \in h(\alpha)$, F contains a square root of $t + \rho(z_{2n})$, and for every $n \notin h(\alpha)$, F contains a square root of $t + \rho(z_{2n+1})$. Within countably many steps, we find such a $t \in \rho(T)$, and we define it to be $f(t_\alpha)$. We also choose $f(x_\alpha)$ and $f(y_\alpha)$ in E to be any of the six solutions to the Fermat equation whose sum is $f(t_\alpha)$; [**15**] shows that the six choices are equivalent. Finally, we define $f(z_n) = \rho(z_n)$ for each n, using countably much information. We have thus defined f on a generating set for F, and it is clear that this map extends to an isomorphism from F onto E. Thus F is ω_1-computably categorical, and the relativization goes through in exactly the same manner.

We again remind the reader that the commutative algebra in this argument must be checked before we can claim to have proven Conjecture 3.6. Even if it fails for the Fermat polynomial of degree 5, however, we do believe that there exist polynomials which would allow the proof to go through as described here. □

References

[1] C. J. Ash, J. F. Knight, M. Mannasse, T. Slaman, "Generic copies of countable structures", *Ann. Pure Appl. Logic*, vol. 42 (1989), pp. 195–205.

[2] W. Calvert, D. Cummins, J. F. Knight, S. Miller, "Comparing classes of finite structures", *Algebra and Logic*, vol. 43 (2004), pp. 374–392.

[3] J. Carson, J. Johnson, J. F. Knight, K. Lange, C. McCoy CSC, J. Wallbaum, "The arithmetical hierarchy in the setting of ω_1", preprint.

[4] J. Chisholm, "Effective model theory versus recursive model theory", *J. Symb. Logic*, vol. 55 (1990), pp. 1168–1191.

[5] H. M. Edwards, *Galois Theory*, New York: Springer-Verlag, 1984.

[6] Yu. L. Ershov, *Theory of Numberings*, Moscow: Nauka, 1977 (in Russian).

[7] E. Fokina, S. Friedman, "On Σ^1_1 equivalence relations over the natural numbers", *Math. Log. Quart.*, vol. 58 (2012), pp. 113–124.

[8] E. Fokina, S. Friedman, V. Harizanov, J. F. Knight, C. McCoy CSC, A. Montalbán, "Isomorphism relations on computable structures", *J. Symb. Logic*, vol. 77 (2012), pp. 122–132.

[9] E. Fokina, S. Friedman, A. Törnquist, "The effective theory of Borel equivalence relations", *Ann. Pure Appl. Logic*, vol. 161 (2010), pp. 837–850.

[10] H. Friedman, L. Stanley, "A Borel reducibility theory for classes of countable structures", *J. Symb. Logic*, vol. 54 (1989), pp. 894–914.

[11] A. Frölich, J. C. Shepherdson, “Effective procedures in field theory”, *Phil. Trans. Royal Soc. London*, Series A 248 (1956) 950, pp. 407–432.

[12] N. Greenberg, J. F. Knight, “Computable structure theory in the setting of ω_1”, paper for Proceedings of first EMU workshop.

[13] G. Metakides, A. Nerode, “Effective content of field theory”, *Ann. Math. Logic*, vol. 17 (1979), pp. 289–320.

[14] T. Millar, “Foundations of recursive model theory”, *Ann. Math. Logic*, vol. 13 (1978), pp. 45–72.

[15] R. Miller, H. Schoutens, “Computably categorical fields via Fermat’s Last Theorem”, submitted for publication.

[16] M. Morley, “Decidable models”, *Israel J. Math.*, vol. 25 (1976), pp. 233–240.

[17] B. L. van der Waerden, “Eine Bemerkung über die Unzerlegbarkeit von Polynomen”, *Math. Ann.*, vol. 102 (1930), pp. 738–739.

Equivalence relations in set theory, computation theory, model theory and complexity theory

Sy-David Friedman[†]

[†] Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
sdf@logic.univie.ac.at

Introduction

One of Harvey's most influential articles is his joint work with Lee Stanley [**8**] in which he introduces a notion of *Borel reducibility* between isomorphism relations on the countable models of a theory in infinitary logic. Through the work of many researchers, this theory later blossomed into a rich field devoted to the more general study of Borel reducibility between Borel and analytic equivalence relations (and quasi-orders). For a look at some of this work see [**11, 12, 17, 19, 23, 26, 27, 30**].

The aim of the present article is to illustrate how a similar idea has recently been used to good effect in four new contexts: *effective* descriptive set theory, computation theory, model theory and complexity theory. This work has deepened research in these fields, produced a number of unexpected results and raised a host of interesting new open problems.

1 Effective descriptive set theory

We begin with a brief description of the classical, non-effective setting, before turning to the more recent work [**6**] in the effective context. The principal objects of study in the classical theory are analytic (Σ^1_1 with parameters) equivalence relations on Polish spaces (think of the reals). Such equivalence relations are compared using *Borel reducibility* in the following way: E_0 *is Borel reducible to* E_1 iff there is a Borel function $f\colon X_0 \to X_1$ such that xE_0y iff $f(x)E_1f(y)$.

E_0 and E_1 are *Borel bireducible* if each Borel reduces to the other. Then $\mathcal{B}$ denotes the resulting set of degrees, ordered under Borel reducibility. When discussing Borel reducibility we sometimes identify an equivalence relation with its degree. Work of Silver [**37**] and of Harrington–Kechris–Louveau [**16**] identifies an interesting initial segment of $\mathcal{B}$:

Theorem 1.1 *$\mathcal{B}$ has the initial segment*

$$1 < 2 < \cdots < \omega < \mathrm{id} < E_0,$$

where n denotes Borel equivalence relations with exactly n classes; ω denotes Borel equivalence relations with exactly $\aleph_0$ classes; id *is* $({}^{\omega}\omega, =)$ *(equality on reals); and E_0 is*

This work has been published: Equivalence relations in set theory, computation theory, model theory and complexity theory, in *Foundational Adventures: Essays in Honor of Harvey M. Friedman*, College Publications, London, 2011.

[†]The author wishes to thank the John Templeton Foundation for its support through the project *Myriad Aspects of Infinity*, Project ID# 13152, and the Austrian Science Fund (FWF) for its support through research project P 20835-N13.

the equivalence relation xE_0y iff $x(n) = y(n)$ for all but finitely many n. In fact, any Borel equivalence relation is Borel equivalent to one of the above or lies strictly above E_0 under Borel reducibility.

The question for the effective theory is: What happens if we replace "Borel" by "effectively Borel"? In what follows we simply write "Hyp" for "effectively Borel" (that is, lightface Δ^1_1). We define:

Definition 1.2 If E and F are Hyp equivalence relations on the reals, then E is *Hyp reducible to* F, written $E \leq_H F$, iff for some Hyp function f, xEy iff $f(x)Ff(y)$.

The relation $\leq_H$ is reflexive and transitive. We write $E \equiv_H F$ for $E \leq_H F$ and $F \leq_H E$.

So the new object of study is $\mathcal{H}$, the degrees of Hyp equivalence relations on the reals under Hyp reducibility.

There are some surprises! Again we have degrees

$$1 < 2 < \cdots < \omega < \mathrm{id} < E_0,$$

defined as follows: n is represented by xE^ny iff $x(0) = y(0) < n-1$ or $x(0), y(0) \geq n-1$; ω is represented by $xE^\omega y$ iff $x(0) = y(0)$; id, E_0 are as before: $x\mathrm{id}y$ iff $x = y$, xE_0y iff $x(n) = y(n)$ for all but finitely many n.

Proposition 1.3 *There are Hyp equivalence relations strictly between* 1 *and* 2*.*

Here is why: Let E be a Hyp equivalence relation. Recall that the $\mathcal{H}$-degree n is represented by the equivalence relation E^n where:

$$xE^ny \text{ iff } x(0) = y(0) < n-1 \text{ or } x(0), y(0) \geq n-1.$$

Fact 1. E^n is Hyp reducible to E iff at least n distinct E-equivalence classes contain Hyp reals.

Proof. Suppose that E^n Hyp reduces to E via the Hyp function f. Each of the n equivalence classes of E^n contains a Hyp real; let $x_0, \ldots, x_{n-1}$ be Hyp, pairwise E^n-inequivalent reals. Then the reals $f(x_i)$, $i < n$ are Hyp, pairwise E-inequivalent reals. Conversely, if $y_0, \ldots, y_{n-1}$ are Hyp, pairwise E-inequivalent reals, then send the E^n-equivalence class of x_i to the real y_i; this is a Hyp reduction of E^n to E. □

Fact 2. E is Hyp reducible to E^2 iff E has at most 2 equivalence classes.

Proof. If E is Hyp reducible to E^2, then E has at most 2 equivalence classes because E^2 has only 2 equivalence classes. Conversely, suppose that the equivalence classes of E are A_0 and A_1. We may assume that A_0 has a Hyp element x. Then A_0 is Hyp as it consists of those reals E-equivalent to x and A_1 is Hyp as it consists of those reals not E-equivalent to x. Now we can reduce E to E^2 by choosing E^2-inequivalent Hyp reals y_0, y_1 and sending the elements of A_0 to y_0 and the elements of A_1 to y_1. □

So to get a Hyp equivalence relation between 1 and 2 we need only find one with two equivalence classes but with all Hyp reals in just one class. The existence of such an equivalence relation follows from a classical fact from Hyp theory (see [**35**, page 52, Theorem 1.1]):

Fact 3. There are nonempty Hyp sets of reals which contain no Hyp element.

Proof. Let A be the set of non-Hyp reals. Then A is Σ_1^1 and therefore the projection of a Π_1^0 subset P of Reals × Reals. P is nonempty. A Hyp real $h = (h_0, h_1)$ in P would give a Hyp real h_0 in A, contradiction. □

Now we ask a harder question: Are there incomparable degrees between 1 and 2? To answer this we prove:

Theorem 1.4 ([**6**]) *There exist Hyp sets of reals A, B such that for* no *Hyp function F do we have $F[A] \subseteq B$ or $F[B] \subseteq A$.*

Given this theorem, define E_A to be the equivalence relation with equivalence classes A and $\sim A$ (the complement of A); define E_B similarly. Note that the sets A, B contain no Hyp reals, else there would be a constant Hyp function F mapping one of them into the other. So a Hyp reduction of E_A to E_B would have to send the elements of $\sim A$ (which contains Hyp reals) to elements of $\sim B$, and therefore the elements of A to elements of B, contradicting the Theorem. Similarly there is no Hyp reduction of E_B to E_A.

Proof sketch of Theorem 1.4. First we quote a result of Harrington [**15**] (see also [**33**, Theorem XIII.3.5]). For reals a, b and a recursive ordinal α, we say that a is *α-below* b iff a is recursive in the α-jump of b.

Fact. For any recursive ordinal α there are Π_1^0 singletons a, b such that a is not α-below b and b is not α-below a.

Using Barwise Compactness, find a nonstandard ω-model M of ZF^- with standard ordinal ω_1^{CK} in which there are Π_1^0 singletons a, b such that, for all recursive α, a is not α-below b and b is not α-below a (i.e., a and b are Hyp incomparable). Let a, b be the unique solutions in M to the Π_1^0 formulas φ_0, φ_1, respectively. The desired sets A, B are $\{x \mid \varphi_0(x)\}$ and $\{x \mid \varphi_1(x)\}$. If F were a Hyp function mapping A into B, then it would send the element a of A to an element $F(a)$ of $B \cap M$; but then $F(a)$ must equal b and therefore b is Hyp in a, contradicting the choice of a, b. □

Now fix A, B as in the Theorem. Using them we can get incomparable Hyp equivalence relations between n and $n+1$ for any finite n, by considering E_A, E_B where the equivalence classes of E_A are A together with a split of $\sim A$ (the complement of A) into n classes, each of which contains a Hyp real (similarly for E_B).

We now consider Hyp equivalence relations with infinitely many equivalence classes. Recall the Silver and Harrington–Kechris–Louveau dichotomies:

Theorem 1.5

(a) (Silver) *A Borel equivalence relation is either Borel reducible to ω or Borel reduces* id.

(b) (Harrington–Kechris–Louveau) *A Borel equivalence relation is either Borel reducible to* id *or Borel reduces E_0.*

How effective are these results? Harrington's proof of (a) and the original proof of (b) show the following:

Theorem 1.6

(a) *A Hyp equivalence relation is either Hyp reducible to ω or Borel reduces* id.

(b) *A Hyp equivalence relation is either Hyp reducible to* id *or Borel reduces E_0.*

The sets A, B of Theorem 1.4 can be used to show that the Silver and Harrington–Kechris–Louveau dichotomies are *not* fully effective:

Theorem 1.7 ([**6**])

(a) *There are incomparable Hyp equivalence relations between* ω *and* id.

(b) *There are incomparable Hyp equivalence relations between* id *and* E_0.

Proof sketch. For (a), consider the relations

$$E_A(x,y) \text{ iff } (x \in A \text{ and } x = y) \text{ or } (x, y \notin A \text{ and } x(0) = y(0));$$

$$E_B\text{: The same, with } A \text{ replaced by } B.$$

Now E^ω Hyp reduces to E_A by $n \mapsto (n, 0, 0, ...)$. Also E_A Hyp reduces to id via the map $G(x) = x$ for $x \in A$, $G(x) = (x(0), 0, 0, ...)$ for $x \notin A$ (same for B).

There is no Hyp reduction of E_A to E_B: If F were such a reduction, then let C be $F^{-1}[\sim B]$. As $\sim B$ is Hyp, C is also Hyp and therefore $A \cap C$ is also Hyp. But $A \cap C$ must be countable as F is a reduction. So if $A \cap C$ were nonempty it would have a Hyp element, contradicting the fact that A has no Hyp element. Therefore F maps A into B, which is impossible by the choice of A, B. By symmetry, there is no Hyp reduction of E_B to E_A.

For (b), we define E_A on $\mathbb{R} \times \mathbb{R}$ by: $(x, y)E_A(x', y')$ iff $x = x'$ and either $x \notin A$ or $(x \in A$ and $yE_0y')$. E_B is the same, with A replaced by B.

We need two facts (see [**18**, Lemma 2.49] and [**24**, Theorem 2.2.5(a)], respectively):

1. If $h\colon \mathbb{R} \to \mathbb{R}$ is Baire measurable and constant on E_0 classes, then h is constant on a comeager set.
2. If $B \subseteq \mathbb{R}^2$ is Hyp, then so is $\{x \mid \{y \mid (x, y) \in B\}$ is comeager$\}$.

Now suppose that F were a Hyp reduction of E_A to E_B. Let $\pi(x, y) = x$ for all x and define $h\colon \mathbb{R} \to \mathbb{R}$ by $h(x) = z$ iff $\{y \mid \pi(F(x, y)) = z\}$ is comeager.

Using 1 and 2, h is a total Hyp function. We claim that $h[A] \subseteq B$, contradicting the choice of A, B: Assume $x \in A$. Then for comeager-many y, $\pi(F(x, y)) = h(x)$. So if $h(x) \notin B$ then F maps more than one E_A class into a single E_B class, contradiction. By symmetry there is no Hyp reduction of E_B to E_A. □

The overall picture of the degrees of Hyp sets of reals under Hyp reducibility is the following: Call a degree *canonical* if it is one of $1 < 2 < \cdots < \omega < \text{id} < E_0$. For any two canonical degrees $a < b$ there is a rich collection of degrees which are above a, below b and incomparable with all canonical degrees in between.

However at least one nice thing happens: If a degree is above n for each finite n, then it is also above ω.

Because this field is so new (like the others introduced in this paper), there remain many open questions. Here are several:

1. If a Hyp equivalence relation is Borel reducible to E_0, then must it also be Hyp reducible to E_0? (This is true for finite n, ω, id.)
2. Are there any nodes other than 1? That is, is there a Hyp equivalence relation with more than one equivalence class which is comparable with all Hyp equivalence relations under Hyp reducibility?
3. Is there a minimal degree? Are there incomparables above each degree?

There is also a jump operation, which is in need of further study.

2 Computation theory

We now turn to equivalence relations not on the reals but on the natural numbers, where computation theory plays a central role. As seen in the last section, Hyp-reducibility for Hyp equivalence relations on the real numbers has a rich structure; however the analogous theory in the context of the natural numbers is trivial:

Proposition 2.1 ([**4**, Section 2.2, Fact 2.10.2]) *Any Hyp equivalence relation on the natural numbers is Hyp reducible to the equality relation on ω.*

Therefore the central objects of interest in our study of equivalence relations on the natural numbers are not the Hyp equivalence relations but instead the Σ^1_1 equivalence relations. Indeed, in the classical theory of Borel reducibility one considers not only the Borel equivalence relations but more generally analytic (Σ^1_1 with parameters) equivalence relations which are not Borel; indeed these appeared already in [**8**].

Let T be any theory in first-order logic (or any sentence of the infinitary logic $\mathcal{L}_{\omega_1\omega}$). Then the isomorphism relation on the countable models of T is an analytic equivalence relation which need not be Borel.

There are analytic equivalence relations which are not Borel reducible to such an isomorphism relation; an example is E_1, the equivalence relation on $\mathbb{R}^\omega$ defined by:

$$\vec{x}E_1\vec{y} \text{ iff } \vec{x}(n) = \vec{y}(n) \text{ for almost all } n.$$

Note that E_1 is even Hyp. A motivating question for our study is the following:

Question. Is every Σ^1_1 equivalence relation on the natural numbers reducible to isomorphism on a Hyp class of *computable* structures?

Of course we can identify a computable structure with a natural number which serves as an index for it. The reducibility we use is: $E_0 \leq_H E_1$ iff there is a Hyp function $f\colon \mathcal{N} \to \mathcal{N}$ such that mE_0n iff $f(m)E_1f(n)$. (We say that E_0 is *Hyp reducible* to E_1.)

Theorem 2.2 ([**5**]) *Every Σ^1_1 equivalence relation on $\mathcal{N}$ is Hyp reducible to isomorphism on computable trees.*

This answers the above question positively.

Proof sketch: Let E be a Σ^1_1 equivalence relation on $\mathcal{N}$ and choose a computable

$$f\colon \mathcal{N}^2 \longrightarrow \{\text{computable trees}\}$$

such that $\sim mEn$ iff $f(m,n)$ is well-founded.

Now associate to pairs m,n computable trees $T(m,n)$ so that:

- $T(m,n)$ is isomorphic to $T(n,m)$;
- mEn implies that $T(m,n)$ is isomorphic to the "canonical" non-well-founded computable tree;
- $\sim mEn$ implies that $T(m,n)$ is isomorphic to the "canonical" computable tree of rank α, where α is least so that $f(m',n')$ has rank at most α for all $m' \in [m]_E$, $n' \in [n]_E$.

Now to each n associate the tree T_n gotten by gluing together the $T(n,i)$, $i \in \omega$. If mEn, then T_m is isomorphic to T_n as they are obtained by gluing together isomorphic trees. And if $\sim mEn$ then T_m, T_n are not isomorphic as they are obtained by gluing together trees which on some component are non-isomorphic. □

It can be shown that the isomorphism relation on computable trees (and therefore any Σ^1_1 equivalence relation on $\mathcal{N}$) Hyp-reduces to the isomorphism relation on each of the following Hyp classes:

1. Computable graphs.
2. Computable torsion-free Abelian groups.
3. Computable Abelian p-groups for a fixed prime p.
4. Computable Boolean algebras.
5. Computable linear orders.
6. Computable fields.

These results came as a surprise, because, in the classical setting, the analogue of 2 is an open problem and the analogue of 3 is false!

Fokina and I show in [**4**] that the global structure of Σ^1_1 equivalence relations on $\mathcal{N}$ under Hyp reducibility is very rich: it embeds the partial order of Σ^1_1 sets under Hyp many-one reducibility. But it is not known if there is a single isomorphism relation on computable structures which is neither Hyp nor complete under Hyp-reducibility! However we do have:

Theorem 2.3 ([**4**]) *Every Σ^1_1 equivalence relation is Hyp bireducible to a bi-embeddability relation on computable structures.*

The proof is based on the analagous result in the non-effective setting:

Theorem 2.4 ([**11**]) *Every analytic equivalence relation on the reals is Borel bireducible to a bi-embeddability relation on countable structures.*

I should also mention that there has been considerable prior work on *computably enumerable* equivalence relations, of which provable equivalence is a natural example. For those interesting results we refer to [**13**] and the references therein.

3 Model theory

It is natural to expect that insights into the model-theoretic properties of a first-order theory could be derived from the descriptive set-theoretic behaviour of the isomorphism relation on its countable models under Borel reducibility. This idea was pursued by Laskowski [**29**], Marker [**31**] and in depth by Koerwien [**28**]. But the conclusion was rather negative: theories can be complicated model-theoretically and simple descriptive set-theoretically (an example is dense linear orderings), or vice-versa (an example is described in [**28**]).

A solution to this difficulty emerged through the study of isomorphism on a theory's *uncountable* models. The work of [**10**] (see Chapter V, Theorem 64) shows, for example, that a theory is classifiable and shallow in Shelah's model-theoretic sense exactly if the isomorphism relation on its models of size κ (for an appropriate choice of regular uncountable cardinal κ) is "Borel" in a generalised sense.

Naturally, a prerequisite for this study is the development of a suitable descriptive set theory of the uncountable, which has turned out to be a fascinating area of independent interest. Armed with such a theory, it becomes possible to bring in the methods of model-theoretic stability theory to uncover deep connections between the model theory and descriptive set theory of first-order theories.

I begin with the uncountable descriptive set theory. It is favourable to choose κ to be uncountable and such that $\kappa^{<\kappa} = \kappa$. The *Generalised Baire Space* κ^κ is the space of all functions $f\colon \kappa \to \kappa$ topologised with basic open sets of the form $N_s = \{f \mid s \subseteq f\}$, s an element of $\kappa^{<\kappa}$. In this context the *Borel* sets are obtained by closing the open sets under the operations of complementation and unions of size at most κ. The Σ^1_1 sets are the projections of Borel sets, the Π^1_1 sets are the complements of the Σ^1_1 sets and the Δ^1_1 sets are those which are both Σ^1_1 and Π^1_1. Borel sets are Δ^1_1 but the converse is false. As usual, a set is *nowhere dense* if its closure contains no nonempty open set; a set is *meager* if it is the union of κ-many nowhere dense sets. The Baire Category Theorem holds in the sense that the intersection of κ-many open dense sets is dense. A set has the *Baire Property (BP)* if its symmetric difference with some open set is meager. Borel sets have the BP. A *perfect set* is the range of a continuous injection from 2^κ (the Generalised Cantor Space) into κ^κ. A set has the *Perfect Set Property (PSP)* iff it either has size at most κ or contains a perfect subset.

Theorem 3.1 ([**10**])

(a) *It is consistent that all Δ^1_1 sets have the BP.*

(b) *For any stationary subset S of κ, the filter* CUB(S)*, the closed unbounded filter restricted to S, is a Σ^1_1 set without the BP.*

(c) *In L,* CUB(S) *for stationary S is not Δ^1_1, but there are nevertheless Δ^1_1 sets without the BP and without the PSP.*

(d) *It is consistent relative to an inaccessible cardinal that all Σ^1_1 sets have the PSP (and the use of an inaccessible is necessary).*

Remark. Part (a) was proved independently by Lücke–Schlicht; in the case $S = \kappa$, part (b) is due to Halko–Shelah and part (d) was proved independently by Schlicht.

I turn now to Borel reducibility. Suppose that X_0, X_1 are Borel subsets of κ^κ. Then $f\colon X_0 \to X_1$ is a *Borel function* iff $f^{-1}[Y]$ is Borel whenever Y is Borel. This implies that the graph of f is Borel, as (x, y) belongs to the graph of f iff for all $s \in \kappa^{<\kappa}$, either y does not belong to N_s or x belongs to $f^{-1}[N_s]$.

If E_0, E_1 are equivalence relations on Borel sets X_0, X_1 respectively, then we say that E_0 is *Borel reducible to* E_1, written $E_0 \leq_B E_1$, iff for some Borel $f\colon X_0 \to X_1$,

$$x_0 E_0 y_0 \text{ iff } f(x_0) E_1 f(x_1).$$

Now recall the following picture from the classical case:

$$1 <_B 2 <_B \cdots <_B \omega <_B \mathrm{id} <_B E_0$$

forms an initial segment of the Borel equivalence relations under $\leq_B$ where n denotes an equivalence relation with n classes for $n \leq \omega$, id denotes equality on ω^ω and E_0 denotes equality modulo finite on ω^ω.

At κ we easily get the initial segment

$$1 <_B 2 <_B \cdots <_B \omega <_B \omega_1 <_B \cdots <_B \kappa$$

where for each nonzero cardinal $\lambda \leq \kappa$ we identify λ with the $\equiv_B$ class of Borel equivalence relations with exactly λ-many classes. What happens above these equivalence relations? We might hope for:

Silver Dichotomy. The equivalence relation id (equality on κ^κ) is the strong successor of κ under $\leq_B$, i.e., if a Borel equivalence relation E has more than κ classes then id is Borel reducible to E.

Theorem 3.2

(a) *The Silver Dichotomy implies the PSP for Borel sets. Therefore it fails in L and its consistency requires at least an inaccessible cardinal.*

(b) *The Silver Dichotomy is false with Borel replaced by* Δ^1_1.

Is the Silver Dichotomy consistent? This question remains open.

We can also consider what happens above id. In the case $\kappa = \omega$ we have:

Classical Glimm–Effros Dichotomy. $E_0 =$ (equality mod finite) is the strong successor of id, i.e., if a Borel equivalence relation E is not Borel reducible to id (i.e., E is not *smooth*) then E_0 Borel-reduces to E.

At κ, what shall we take E_0 to be? For infinite regular $\lambda \leq \kappa$, define $E_0^{<\lambda} =$ equality for subsets of κ modulo sets of size $< \lambda$.

Proposition 3.3 *For* $\lambda < \kappa$, $E_0^{<\lambda}$ *is Borel bireducible with* id.

So we can forget about $E_0^{<\lambda}$ for $\lambda < \kappa$ and set $E_0 = E_0^{<\kappa}$, equality modulo bounded sets.

As in the classical case, we have:

Proposition 3.4 $E_0 = E_0^{<\kappa}$ *is* not *Borel reducible to* id.

There are other versions of E_0: For regular $\lambda < \kappa$ define $E^\kappa_\lambda =$ equality modulo the ideal of λ-nonstationary sets. These equivalence relations are key for connecting model-theoretic stability with uncountable descriptive set theory.

How do the relations E^κ_λ compare to each other under Borel reducibility for different λ? For simplicity, consider the special case $\kappa = \omega_2$.

Theorem 3.5 ([**10**])

(a) *It is consistent that* $E^{\omega_2}_{\omega}$ *and* $E^{\omega_2}_{\omega_1}$ *are incomparable under Borel reducibility.*

(b) *Relative to a weak compact it is consistent that* $E^{\omega_2}_{\omega}$ *is Borel reducible to* $E^{\omega_2}_{\omega_1}$.

It is not known if it is consistent for $E^{\omega_2}_{\omega_1}$ to be Borel reducible to $E^{\omega_2}_{\omega}$.

What is the relationship between E_0 and E^κ_λ?

Theorem 3.6

(a) *The relations* E^κ_λ *do not Borel reduce to* E_0*, as* E_0 *is Borel and the* E^κ_λ *are not.*

(b) *If* $\kappa = \mu^+$ *for some cardinal* μ*, then* E_0 *reduces to* E^κ_λ*, unless* λ *is the cofinality of* μ.

(c) *In L, the condition in* (b) *that* λ *not be the cofinality of* μ *can be dropped.*

The structure of the Δ^1_1 equivalence relations under Borel reducibility is (consistently) very rich:

Theorem 3.7 *Consistently, there is an injective, order-preserving embedding from* $(\mathcal{P}(\kappa), \subseteq)$ *into the partial order of* Δ^1_1 *equivalence relations under Borel reducibility.*

The above summarises the current state of knowledge regarding uncountable descriptive set theory. As has been mentioned, there remain many open questions, some of which we list at the end of this section.

Now we return to the connection between uncountable descriptive set theory and model theory. Let T be a countable, complete and first-order theory. Then T is *classifiable* iff there is a "structure theory" for its models; example: algebraically closed fields (transcendence degree). T is *unclassifiable* otherwise; example: dense linear orderings.

Shelah's Characterisation (Main Gap). T is classifiable iff T is superstable without the OTOP and without the DOP.

A classifiable T is *deep* iff it has the maximum number of models in all uncountable powers; example: acyclic undirected graphs —every node has infinitely many neighbours. T is *shallow* otherwise. (Remark: Actually, Shelah defined "deep" differently, in terms of rank. The fact that his definition is equivalent to the previous is one of the most profound results of his classification theory.)

Now for simplicity assume $\kappa = \lambda^+$ where λ is uncountable and regular and the GCH holds at λ. Isom^κ_T is the isomorphism relation on the models of T of size κ.

Theorem 3.8 ([**10**])

(a) *T is classifiable and shallow iff* Isom^κ_T *is Borel.*

(b) *T is classifiable iff for all regular $\mu < \kappa$, $E_{S^\kappa_\mu}$ is not Borel reducible to* Isom^κ_T.

(c) *In L, T is classifiable iff* Isom^κ_T *is* Δ^1_1.

The proof uses Ehrenfeucht–Fraïssé games. The Game $\mathrm{EF}^\kappa_t(\mathcal{A}, \mathcal{B})$ is defined as follows, where $\mathcal{A}$, $\mathcal{B}$ are structures of size κ and t is a tree. Player I chooses size $< \kappa$ subsets of $A \cup B$ and nodes along an initial segment of a branch through t; player II builds a partial isomorphism between $\mathcal{A}$ and $\mathcal{B}$ which includes the sets that player I has chosen. Player II wins iff he survives until a cofinal branch is reached.

The tree t *captures* Isom^κ_T iff for all size κ models $\mathcal{A}$, $\mathcal{B}$ of T, $\mathcal{A} \simeq \mathcal{B}$ iff player II has a winning strategy in $\mathrm{EF}^\kappa_t(\mathcal{A}, \mathcal{B})$.

Now there are four cases:

Case 1: *T is classifiable and shallow.*

Then Shelah's work [**36**] shows that some well-founded tree captures Isom^κ_T. We use this to show that Isom^κ_T is Borel.

Case 2: *T is classifiable and deep.*

Then Shelah's work shows that no fixed well-founded tree captures Isom^κ_T. We use this to show that Isom^κ_T is not Borel.

Shelah's work also shows that $L_{\infty\kappa}$ equivalent models of T of size κ are isomorphic. This means that the tree $t = \omega$ (with a single infinite branch) captures Isom^κ_T. As the games $\mathrm{EF}^\kappa_\omega(\mathcal{A}, \mathcal{B})$ are determined, this shows that Isom^κ_T is Δ^1_1.

We must also show: $E_{S^\kappa_\mu}$ (equality modulo the μ-nonstationary ideal) is not Borel reducible to Isom^κ_T for any regular $\mu < \kappa$. This is because (in this case) Isom^κ_T is absolutely Δ^1_1, whereas μ-stationarity is not.

Now we look at the unclassifiable cases. Recall that classifiable means superstable without DOP and without OTOP.

Case 3: *T is unstable, superstable with DOP or superstable with OTOP.*

Work of Hyttinen–Shelah [**20**] and Hyttinen–Tuuri [**21**] shows that in this case no tree of size κ without branches of length κ captures Isom^κ_T. This can be used to show Isom^κ_T is not Δ^1_1.

But $E_{S^\kappa_\lambda} \leq_B \mathrm{Isom}^\kappa_T$ is harder. Following Shelah, there is a Borel map $S \mapsto \mathcal{A}(S)$ from subsets of κ to Ehrenfeucht–Mostowski models of T built on linear orders so that $\mathcal{A}(S_0) \simeq \mathcal{A}(S_1)$ iff $S_0 = S_1$ modulo the λ-nonstationary ideal.

Case 4: *T is stable but not superstable.*

This is the hardest case and requires some new model theory. In our joint paper [**10**], Hyttinen replaces Ehrenfeucht–Mostowski models built on linear orders with primary models built on trees of height $\omega + 1$ to show $E_{S^\kappa_\omega} \leq_B \mathrm{Isom}^\kappa_T$. (We do not know if $E_{S^\kappa_\lambda} \leq_B \mathrm{Isom}^\kappa_T$ or if Isom^κ_T could be Δ^1_1 in this case.)

Now we have all we need to prove the theorem mentioned earlier:

(a) T is classifiable and shallow iff Isom^κ_T is Borel.

We mentioned that if T is classifiable and shallow then Isom^κ_T is Borel and if it is classifiable and deep it is not. If T is not classifiable, then some $E_{S^\kappa_\mu}$ Borel reduces to Isom^κ_T, so the latter cannot be Borel.

(b) T is classifiable iff for all regular $\mu < \kappa$, $E_{S^\kappa_\mu}$ is not Borel reducible to Isom^κ_T.

We mentioned that if T is not classifiable then $E_{S^\kappa_\mu}$ is Borel reducible to Isom^κ_T where μ is either λ or ω. We also mentioned that if T is classifiable and deep then no $E_{S^\kappa_\mu}$ is Borel reducible to Isom^κ_T, by an absoluteness argument. When T is classifiable and shallow there is no such reduction as Isom^κ_T is Borel.

(c) In L, T is classifiable iff Isom^κ_T is Δ^1_1.

We mentioned that if T is classifiable then Isom^κ_T is Δ^1_1 in ZFC. If T is not classifiable, then $E_{S^\kappa_\mu}$ Borel reduces to Isom^κ_T for some μ, and in L, $E_{S^\kappa_\mu}$ is not Δ^1_1.

This summarises the work in [**10**]. Some surprisingly basic and very interesting open questions remain in this new area. Below are some of them. Assume $\kappa^{<\kappa} = \kappa$, as before.

1. Under what conditions on an uncountable κ does Vaught's Conjecture hold in the following form: If an isomorphism relation on the models of size κ has more than κ classes, then id is Borel reducible to it?

2. Is the Silver Dichotomy for uncountable κ consistent?

3. Is it consistent for there to be Borel equivalence relations which are incomparable under Borel reducibility for an uncountable κ?

4. Is it consistent that $S^{\omega_2}_{\omega_1}$ Borel reduces to $S^{\omega_2}_{\omega}$?

5. We proved that the isomorphism relation of a theory T is Borel if and only if T is classifiable and shallow. Is there a connection between the depth of a shallow theory and the Borel degree of its isomorphism relation? Is one monotone in the other?

6. Can it be proved in ZFC that if T is stable unsuperstable then isomorphism for the size κ models of T (κ uncountable) is not Δ^1_1?

7. If $\kappa = \lambda^+$ with λ regular and uncountable, then does equality modulo the λ-nonstationary ideal Borel reduce to isomorphism for the size κ models of T for all stable unsuperstable T?

8. Let DLO be the theory of dense linear orderings without end points and RG the theory of random graphs. Does the isomorphism relation of RG Borel reduce to that of DLO for an uncountable κ?

4 Complexity theory

We consider NP equivalence relations on finite strings. One motivation for this topic is the following: Borel reducibility allows us to compare isomorphism relations on Borel classes of countable structures. Is there an analogous reducibility for "nice" classes of *finite* structures?

The resulting theory of "strong isomorphism reductions" is introduced in [**9**] and studied systematically in [**2**]. We consider polynomial-time definable classes C of structures for a finite vocabulary τ, where the structures in C have universe $\{1, \dots, n\}$ for some finite $n > 0$ and where C is *invariant*, i.e., closed under isomorphism. To avoid trivialities we also assume that C contains arbitrarily large structures. Some examples of such classes are:

1. The classes SET, BOOLE, FIELD, GROUP, ABELIAN and CYCLIC of sets (structures of empty vocabulary), Boolean algebras, fields, groups, abelian groups, and cyclic groups, respectively.
2. The class GRAPH of (undirected and simple) graphs.
3. The class ORD of linear orderings.
4. The classes LOP of linear orderings with a distinguished point and LOU of linear orderings with a unary relation.

Let C and D be classes. We say that C is *strongly isomorphism reducible to* D and write $C \leq_{\text{iso}} D$ if there is a function $f\colon C \to D$ computable in polynomial time such that, for all $\mathcal{A}, \mathcal{B} \in C$, $\mathcal{A} \simeq \mathcal{B}$ iff $f(\mathcal{A}) \simeq f(\mathcal{B})$. We then say that f is a *strong isomorphism reduction* from C to D and write $f\colon C \leq_{\text{iso}} D$. If $C \leq_{\text{iso}} D$ and $D \leq_{\text{iso}} C$, denoted by $C \equiv_{\text{iso}} D$, then C and D *have the same strong isomorphism degree.*

Examples

(a) The map sending a field to its multiplicative group shows that FIELD $\leq_{\text{iso}}$ CYCLIC.

(b) CYCLIC $\leq_{\text{iso}}$ ABELIAN $\leq_{\text{iso}}$ GROUP; more generally, if $C \subseteq D$, then $C \leq_{\text{iso}} D$ via the identity.

(c) SET $\equiv_{\text{iso}}$ FIELD $\equiv_{\text{iso}}$ ABELIAN $\equiv_{\text{iso}}$ CYCLIC $\equiv_{\text{iso}}$ ORD $\equiv_{\text{iso}}$ LOP. (For the proof, see [**2**].)

Proposition 4.1 $C \leq_{\text{iso}}$ GRAPH *for all classes* C.

The structure of $\leq_{\text{iso}}$ between LOU and GRAPH is linked with central open problems of descriptive complexity. Before turning to that, I will first consider the structure below LOU. That structure, even below LOP, is quite rich.

Theorem 4.2 *The partial ordering of the countable atomless Boolean algebra is embeddable into the partial ordering induced by* $\leq_{\text{iso}}$ *on the degrees of strong isomorphism reducibility below* LOP. *More precisely, let* $\mathcal{B}$ *be the countable atomless Boolean algebra. Then there is a one-to-one function* $b \mapsto C_b$ *defined on* B *such that, for all* $b, b' \in B$,

(i) C_b *is a subclass of* LOP*;*

(ii) $b \leq b'$ *iff* $C_b \leq_{\text{iso}} C_{b'}$.

This result is obtained by comparing the number of isomorphism types of structures with universe of bounded cardinality in different classes. For a class C, we let $C(n)$ be the subclass consisting of all structures in C with universe of cardinality $\leq n$ and we let $\#C(n)$ be the number of isomorphism types of structures in $C(n)$.

Examples

(a) $\#\text{BOOLE}(n) = [\log n]$, $\#\text{CYCLIC}(n) = n$, $\#\text{SET}(n) = \#\text{ORD}(n) = n+1$.

(b) $\#\text{LOP}(n) = \sum_{i=1}^{n} i = (n+1)\cdot n/2$ and $\#\text{LOU}(n) = \sum_{i=0}^{n} 2^i = 2^{n+1} - 1$.

(c) $\#\text{GROUP}(n)$ is superpolynomial but subexponential (more precisely, it is bounded by $n^{O(\log^2 n)}$); see [**1**].

A class C is *potentially reducible* to a class D, written $C \leq_{\text{pot}} D$, iff there is some polynomial p such that $\#C(n) \leq \#D(p(n))$ for all $n \in \mathbb{N}$. Of course, by $C \equiv_{\text{pot}} D$ we mean $C \leq_{\text{pot}} D$ and $D \leq_{\text{pot}} C$.

Lemma 4.3 *If $C \leq_{\text{iso}} D$, then $C \leq_{\text{pot}} D$.*

Proof. Let $f\colon C \leq_{\text{iso}} D$. As f is computable in polynomial time, there is a polynomial p such that for all $\mathcal{A} \in C$ we have $|f(A)| \leq p(|A|)$, where $f(A)$ denotes the universe of $f(\mathcal{A})$. As f strongly preserves isomorphisms, it therefore induces a one-to-one map from $\{\mathcal{A} \in C : |A| \leq n\}/_{\simeq}$ to $\{\mathcal{B} \in D : |B| \leq p(n)\}/_{\simeq}$. □

We state some consequences of this simple observation:

Proposition 4.4

1. CYCLIC $\not\leq_{\text{iso}}$ BOOLE *and* LOU $\not\leq_{\text{iso}}$ LOP.
2. $C \leq_{\text{pot}}$ LOU *for all classes C and* LOU $\equiv_{\text{pot}}$ GRAPH.
3. *The strong isomorphism degree of* GROUP *is strictly between that of* LOP *and* GRAPH.
4. *The potential reducibility degree of* GROUP *is strictly between that of* LOP *and* LOU.

The following concepts are used in the proof of Theorem 4.2. We call a function $f\colon \mathbb{N} \to \mathbb{N}$ *value-polynomial* iff it is increasing and $f(n)$ can be computed in time $f(n)^{O(1)}$. Let VP be the class of all value-polynomial functions. For $f \in \text{VP}$, the set

$$C_f = \{\mathcal{A} \in \text{LOP} : |A| \in \text{im}(f)\}$$

is in polynomial time and is closed under isomorphism. As there are exactly $f(k)$ pairwise non-isomorphic structures of cardinality $f(k)$ in LOP, we get

$$\#\mathcal{C}_f(n) = \sum_{k\in\mathbb{N} \text{ with } f(k)\leq n} f(k).$$

The following proposition contains the essential idea underlying the proof of Theorem 4.2. Loosely speaking, it says that if the gaps between consecutive values of $f \in \text{VP}$ "kill" every polynomial, then there are classes C and D with $C \not\leq_{\text{pot}} D$.

Proposition 4.5 *Let $f \in VP$ and assume that for every polynomial $p \in \mathbb{N}[X]$ there is an $n \in \mathbb{N}$ such that*

$$\sum_{k\in\mathbb{N} \text{ with } f(2k)\leq n} f(2k) > \sum_{k\in\mathbb{N} \text{ with } f(2k+1)\leq p(n)} f(2k+1).$$

Then $\mathcal{C}_{g_0}$ is not potentially reducible to $\mathcal{C}_{g_1}$, where $g_0, g_1 \colon \mathbb{N} \to \mathbb{N}$ are defined by $g_0(n) := f(2n)$ and $g_1(n) := f(2n+1)$.

Proof. For contradiction assume that there is some polynomial p such that $\#\mathcal{C}_{g_0}(n) \le \#\mathcal{C}_{g_1}(p(n))$ for all $n \in \mathbb{N}$. Choose n to satisfy the hypothesis. Then

$$\#\mathcal{C}_{g_0}(n) = \sum_{f(2k)\le n} f(2k) > \sum_{f(2k+1)\le p(n)} f(2k+1) = \#\mathcal{C}_{g_1}(p(n)),$$

a contradiction. □

The other needed ingredient for the proof of Theorem 4.2 is:

Lemma 4.6 *The images of the functions in VP together with the finite subsets of $\mathbb{N}$ are the elements of a countable Boolean algebra $\mathcal{V}$ (under the usual set-theoretic operations). The factor algebra $\mathcal{V}/_{\equiv_{\mathrm{pot}}}$, where, for $b, b' \in V$,*

$$b \equiv b' \iff (b \setminus b') \cup (b' \setminus b) \text{ is finite},$$

is a countable atomless Boolean algebra.

This lemma shows that the set of images of functions in VP has a rich structure. To complete the proof of Theorem 4.2, the functions in VP are composed with a "stretching" function h, which guarantees that the gaps between consecutive values "kill" every polynomial. Then we can apply the idea of the proof of Proposition 4.5 to show that the set of the $\le_{\mathrm{pot}}$-degrees has a rich structure too. For the details, see [**2**].

So far, in all concrete examples of classes C and D for which we know the status of $C \le_{\mathrm{iso}} D$ and of $C \le_{\mathrm{pot}} D$, we have $C \le_{\mathrm{iso}} D$ iff $C \le_{\mathrm{pot}} D$. So the question arises whether the relations of strong isomorphism reducibility and potential reducibility coincide. We believe that they are distinct but have only the following partial result:

Theorem 4.7 *If* UEEXP$\cap$coUEEXP $\neq$ EEXP, *then the relations of strong isomorphism reducibility and that of potential reducibility are distinct.*

Recall that EEXP = DTIME$\Big(2^{2^{n^{O(1)}}}\Big)$ and NEEXP := NTIME$\Big(2^{2^{n^{O(1)}}}\Big)$.

The complexity class UEEXP consists of those $Q \in$ NEEXP for which there is a non-deterministic Turing machine of type NEEXP that for every $x \in Q$ has exactly one accepting run. Finally, coUEEXP := $\{\sim Q \mid Q \in \text{UEEXP}\}$.

Here is the idea of the proof: Assume $Q \in$ UEEXP$\cap$coUEEXP. We construct classes C and D which contain structures in the same cardinalities and which contain exactly two non-isomorphic structures in these cardinalities. Therefore they are potentially reducible to each other. While it is trivial to exhibit two non-isomorphic structures in C of the same cardinality, from any two non-isomorphic structures in D we obtain information on membership in Q for all strings of a certain length. If $C \le_{\mathrm{iso}} D$ held, then we would get non-isomorphic structures in D (in time allowed by EEXP) by applying the strong isomorphism reduction to two non-isomorphic structures in C and therefore obtain $Q \in$ EEXP.

In the other direction we have:

Theorem 4.8 *If strong isomorphism reducibility and potential reducibility are distinct, then $P \neq \#P$.*

Recall that $P = \#P$ means that for every polynomial time non-deterministic Turing machine $\mathbb{M}$ the function $f_{\mathbb{M}}$ such that $f_{\mathbb{M}}(x)$ is the number of accepting runs of $\mathbb{M}$ on $x \in \Sigma^*$ is computable in polynomial time. The class $\#P$ consists of all the functions $f_{\mathbb{M}}$.

Until now we have focused exclusively on isomorphism relations on invariant polynomial time classes of finite structures. But this theory can be put into the broader context of NP equivalence relations in general. If E and E' are NP equivalence relations, then we say that E is *strongly equivalence reducible to* E', and write $E \leq_{\text{eq}} E'$, iff there is a function f computable in polynomial time such that for all strings x, y: xEy iff $f(x)E'f(y)$. We then say that f is a *strong equivalence reduction* from E to E' and write $f\colon E \leq_{\text{eq}} E'$. The following natural question then arises: Is there a *maximal* NP equivalence relation under the reducibility $\leq_{\text{eq}}$? The final section of [**2**] relates this question to enumerations of clocked Turing machines, to p-optimal proof systems as well as to other central questions in complexity theory.

Another natural question is whether, in analogy to the computability theory context, every NP equivalence relation is reducible to an isomorphism relation on a polynomial time invariant class of finite structures, or equivalenty, whether graph isomorphism is $\leq_{\text{eq}}$ complete among NP equivalence relations. For this we have the following partial result:

Proposition 4.9 ([**2**]) *Assume that the polynomial time hierarchy does not collapse. Then* not *every* NP *equivalence relation reduces to graph isomorphism.*

Indeed there are many worthy open questions in this area waiting to be explored.

In conclusion, after decades of work focusing on the "unary" case, definability theory has been dramatically deepened by the study of binary relations, most importantly equivalence relations. An important step in this process was taken in Harvey's fundamental paper with Lee Stanley [**8**]. The extent to which the different areas of logic have been enriched through the study of analogues of Harvey's idea is only now being understood, and I look forward to seeing much exciting work in this direction during the coming years.

References

[1] H. U. Besche, B. Eick and E. A. O'Brien, The groups of order at most 2000, Electronic Research Announcements of the American Mathematical Society, 7 (2001), 1–4.

[2] S. Buss, Y. Chen, J. Flum, S. Friedman and M. Müller, Strong isomorphism reductions in complexity theory, Journal of Symbolic Logic, December 2011.

[3] E. Fokina and S. Friedman, Equivalence relations on classes of computable structures, Proceedings of Computability in Europe 2009, Heidelberg, Germany, Lecture Notes in Computer Science 5635, 198–207, 2009.

[4] E. Fokina and S. Friedman, On Σ^1_1 equivalence relations over the natural numbers, to appear in Mathematical Logic Quarterly.

[5] E. Fokina, S. Friedman, V. Harizanov, J. Knight, C. McCoy and A. Montalbán, Isomorphism relations on computable structures, Journal of Symbolic Logic, March 2012.

[6] E. Fokina, S. Friedman and A. Törnquist, The effective theory of Borel equivalence relations, Annals of Pure and Applied Logic, 161 (2010), 837–850.

[7] E. Fokina, J. Knight, C. Maher, A. Melnikov and S. Quinn, Classes of Ulm type, and relations between the class of rank-homogeneous trees and other classes, submitted.

[8] H. Friedman and L. Stanley, A Borel reducibility theory for classes of countable structures, Journal of Symbolic Logic, 54 (1989), 894–914.

[9] S. Friedman, Descriptive set theory for finite structures, Lecture at the Kurt Gödel Research Center, 2009. Available at http://www.logic.univie.ac.at/ sdf/papers/wien-spb.pdf.

[10] S. Friedman, T. Hyttinen and V. Kulikov, Generalized descriptive set theory and classification theory, submitted, see http://www.logic.univie.ac.at/ sdf/papers/joint.tapani.vadim.pdf.

[11] S. D. Friedman and L. Motto Ros, Analytic equivalence relations and bi-embeddability, Journal of Symbolic Logic, 76 (2011), no. 1, 1581–1587.

[12] S. Gao, Invariant Descriptive Set Theory, Pure and Applied mathematics, CRC Press/Chapman & Hall, 2009.

[13] S. Gao and P. M. Gerdes, Computably enumerable equivalence relations, Studia Logica, 67 (2001), 27–59.

[14] L. Harrington, McLaughlin's Conjecture, Handwritten notes, 1976.

[15] L. Harrington, Arithmetically Incomparable Arithmetical Singletons, Handwritten notes, 1975.

[16] L. Harrington, A. Kechris and A. Louveau, Glimm–Efros dichotomy for Borel equivalence relations, Journal of the American Mathematical Society, 3 (1990), no. 4, 903–928.

[17] G. Hjorth, The isomorphism relation on countable torsion-free Abelian groups, Fundamenta Mathematicae, 175 (2002), 241–257.

[18] G. Hjorth, Classification and orbit equivalence relations, Mathematical Surveys and Monographs 75, American Mathematical Society, 2000.

[19] G. Hjorth and A. Kechris, Recent developments in the theory of Borel reducibility, Fundamenta Mathematicae, 170 (2001), no. 1–2, 21–52.

[20] T. Hyttinen and S. Shelah, Constructing strongly equivalent nonisomorphic models for unsuperstable theories, Part C, Journal of Symbolic Logic, 64 (1999), no. 2, 634–642.

[21] T. Hyttinen and H. Tuuri, Constructing strongly equivalent nonisomorphic models, Annals of Pure and Applied Logic, 52 (1991), no. 3, 203–248.

[22] S. Jackson, A. Kechris and A. Louveau, Countable Borel equivalence relations, Journal of Mathematical Logic, 2 (2002), no. 1, 1–80.

[23] V. Kanovei, Borel Equivalence Relations. Structure and Classification, University Lecture Series 44, American Mathematical Society, 2008.

[24] A. Kechris, Measure and category in effective descriptive set theory, Annals of Pure and Applied Logic, 5 (1973), 337–384.

[25] A. Kechris, Classical Descriptive Set Theory, Graduate Texts in Mathematics, Springer-Verlag, 1995.

[26] A. Kechris, New directions in descriptive set theory, Bulletin of Symbolic Logic, 5 (1999), no. 2, 161–174.

[27] A. Kechris and A. Louveau, The classification of hypersmooth Borel equivalence relations, Journal of the American Mathematical Society, 10 (1997), no. 1, 215–242.

[28] M. Koerwien, A complicated ω-stable depth 2 theory, to appear in Journal of Symbolic Logic.

[29] C. Laskowski, An old friend revisited: Countable models of omega-stable theories, Proceedings of the Vaught's Conjecture Conference, Notre Dame Journal of Formal Logic, 48 (2007) 133–141.

[30] A. Louveau and C. Rosendal, Complete analytic equivalence relations, Transactions of the American Mathematical Society, 357 (2005), no. 12, 4839–4866.

[31] D. Marker, The Borel complexity of isomorphism for theories with many types, preprint.

[32] A. Montalbán, On the equimorphism types of linear orderings, Bulletin of Symbolic Logic, 13 (2007), 71–99.

[33] P. G. Odifreddi, Classical Recursion Theory, vol. II, North-Holland, 1999.

[34] H. Rogers, Theory of Recursive Functions and Effective Computability, McGraw-Hill, 1967.

[35] G. Sacks, Higher Recursion Theory, Springer-Verlag, 1989.

[36] S. Shelah, Classification Theory, revised edition, North Holland, 1990.

[37] J. H. Silver, Counting the number of equivalence classes of Borel and coanalytic equivalence relations, Annals of Mathematical Logic, 18 (1980), 1–18.

Computable models of Ehrenfeucht theories

Alexander Gavryushkin[*]

[*] Irkutsk State University, Russia
gavryushkin@gmail.com

Abstract. In 1976, M. Morley posed the following question: If T is a hereditarily decidable theory with only finitely many countable models, are all countable models of T necessarily decidable? Since then, there were many attempts to solve the problem but the success is yet to come. We present several new results on computable (and decidable) models of Ehrenfeucht theories together with a historical survey. We prove that there are many examples of Ehrenfeucht theories having arbitrarily many homogenous models. This enables us to solve the Morley Problem positively for a large subclass of the class of hereditarily decidable Ehrenfeucht theories. Also, we study an analogue of the Problem for computable models and present some results on the computable complexity of Ehrenfeucht theories. We use a classification of countable models involving almost prime (over a type) models and limit (over a type) models.

Introduction

This paper contains a survey of the current state of research concerning computable models of Ehrenfeucht theories together with several new results. We start with definitions, preliminaries, and history of the subject.

We use standard notions of computable model theory. We use canonical notation from [**1, 2, 17**].

All signatures in the paper are computable and structures are countable. We shall use letters like $\mathfrak{A}$, $\mathfrak{B}$ (sometimes with indices —$\mathfrak{B}_m$) for structures, and respective letters like A, B (sometimes with indices —B_m) for their domains.

When we talk about computability properties of some set of formulas, we identify a formula with its Gödel number. Due to historical reasons, we call computable theories *decidable*. We say that a structure $\mathfrak{A}$ is *computable* if its domain is a subset of ω (the set of natural numbers), and its atomic diagram, denoted by $\mathcal{D}(\mathfrak{A})$, is computable. It is equivalent to say that the domain of $\mathfrak{A}$ is computable and the relations and operations are uniformly computable. We say that a structure $\mathfrak{A}$ (whose domain is a subset of ω) is *decidable* if its complete diagram, denoted by $\mathcal{D}_c(\mathfrak{A})$, is computable. We say that a structure $\mathfrak{B}$ is a *presentation* of the structure $\mathfrak{A}$ if $\mathfrak{B}$ is isomorphic to $\mathfrak{A}$, written $\mathfrak{B} \cong \mathfrak{A}$. We shall usually use “is computable” instead of “has a computable presentation” and “is decidable” instead of “has a decidable presentation”, but it will be clear from the context what exactly we mean.

Denote by $S(T)$ the set of all types (over $\varnothing$) consistent with a theory T. A complete theory T in a countable language is *small* if the set $S(T)$ is countable. Denote by $\omega(T)$ the number of countable models up to isomorphism of a theory T. A complete theory T is an *Ehrenfeucht theory* if $1 < \omega(T) < \omega$. The class of Ehrenfeucht theories is one of the most mysterious subclasses of the class of small theories. There are many longstanding

[*]The author was partially supported by the Federal Target Program “Research and Training Specialists in Innovative Russia, 2009–2013”, Contract no. П1227 and 16.740.11.0567. He expresses his gratitude to the John Templeton Foundation for its support through Project #13152, Myriad Aspects of Infinity.

open questions about it in both model theory and in computability theory. We are going to mention the most important of them in the current paper.

The story began in Ithaca, New York. A. Nerode, inspired by a result of L. Harrington [**9**] and N. Khissamiev [**10**], who proved that if a decidable theory is $\aleph_1$-categorical then all of its countable models are decidable, asked the following question. If T is a decidable Ehrenfeucht theory, are all the countable models of T decidable? M. Morley and A. Lachlan [**15**] answered this question by giving an example of a theory with six countable models of which only the prime one was decidable. Later, M. Peretyat'kin [**16**] gave for all $n \geqslant 3$ an example of a theory with exactly n models of which only the prime one was decidable. By a theorem of R. Vaught (see [**2**]) no complete theory has exactly two countable models. To achieve these results, the authors produced decidable Ehrenfeucht theories with a non-computable non-principal type and then used an effective version of the type omitting theorem. That was the reason for Morley [**15**] to modify Nerode's question in the following way.

Problem 1 (M. Morley, 1976) Suppose that a decidable theory T has exactly $n < \omega$ countable models and every type consistent with T is computable.

(1) Is every countable model of T decidable?
(2) If not, what is the least n giving a counterexample?

Morley himself solved the problem positively for $n = 3$. For $n \geqslant 4$ the problem is still unsolved. Investigations of decidable models of Ehrenfeucht theories are closely related to the study of decidability of prime and saturated models, in particular, and homogeneous models in general.

Harrington [**9**] found a criterion for a prime model to be decidable. Suppose T has a prime model. A necessary and sufficient condition that it be decidable is that there be a computable list of the principal types. One corollary of the computable version of the omitting types theorem is the following. If a decidable theory does not have a decidable prime model then it has an infinite number of nonisomorphic decidable models. (Indeed, in this case every decidable model of the theory realizes a non-principal type, which can be omitted in some other decidable model.) So, prime models of decidable Ehrenfeucht theories are decidable. Morley proved a criterion for a saturated model to be decidable. A necessary and sufficient condition is the existence of a computable list of all the finite types. From this, he noticed that the answer to the second question of his Problem 1 is not three. Another question Morley posed is whether the results for prime and saturated models can be extended to homogeneous models. S. Goncharov [**7**] and T. Millar [**14**] independently found counterexamples. Also, they found necessary and sufficient conditions for a homogeneous model to be decidable. These conditions require additional properties on computability of extensions of types. But it follows from Vaught's theorem that every Ehrenfeucht theory has a model that is not homogeneous. However, all known models of Ehrenfeucht theories are homogeneous over tuples of its elements, we call such models *almost homogeneous*. More precisely, a model $\mathfrak{A}$ is *almost homogeneous* if there is a tuple of elements $\overline{a}$ in $\mathfrak{A}$ such that the expansion of $\mathfrak{A}$ by these elements $\langle \mathfrak{A}, \overline{c} \rangle$ is a homogeneous model (here we interpret constants $\overline{c}$ as elements $\overline{a}$). This led Goncharov and Millar to pose the following question.

Problem 2 Suppose that an almost homogeneous model $\mathfrak{A}$ has an Ehrenfeucht theory and realizes only computable types. Is $\mathfrak{A}$ decidable?

This question is still open. It might be equivalent to Problem 1. "Might be" because of the following model-theoretic question, which was also asked by Goncharov and Millar.

Problem 3 Suppose that a model $\mathfrak{A}$ has Ehrenfeucht theory. Is $\mathfrak{A}$ almost homogeneous?

We are going to look at these problems from a modern point of view.

1 Model-theoretic preliminaries

In this section, we describe results about isomorphism types of countable models of Ehrenfeucht theories.

A type p of a theory T is said to be *powerful* (in the theory T) if every model $\mathfrak{A}$ of T realizing p also realizes every type from $S(T)$. If a complete theory does not have a powerful type, then it has infinitely many models. Indeed, take a type p_0; since it is not powerful, there exist a type p_1 and a model $\mathfrak{A}_0$ that realizes p_0 and omits p_1; since p_0, p_1 are not powerful, again there exist a type p_2 and a model $\mathfrak{A}_1$ that realizes p_0, p_1 and omits p_2; since p_0, p_1, p_2 are not powerful... Continuing as such would produce infinitely many non-isomorphic countable models. Thus, every Ehrenfeucht theory has a powerful type.

A model $\mathfrak{A}$ is said to be *prime over a type* p if there is a tuple of elements $\overline{a}$ in $\mathfrak{A}$ such that $\overline{a}$ is a realization of p in $\mathfrak{A}$, i.e., $\mathfrak{A} \models p(\overline{a})$, and the model $\langle \mathfrak{A}, \overline{a} \rangle$ is prime. A model $\mathfrak{M}$ is *almost prime* if it is prime over a realization of some type. As we shall see later, almost prime models form a sort of basis in the class of *Ehrenfeucht models* (that is, the class of models whose theories are Ehrenfeucht).

If p is a type and two models $\mathfrak{A}$ and $\mathfrak{B}$ are prime over p, then $\mathfrak{A} \cong \mathfrak{B}$. So we denote by $\mathfrak{A}_p$ some model which is prime over p. Denote[1] by $RK(T)$ the set of all pairwise nonisomorphic models $\mathfrak{A}_p$ over all $p \in S(T)$. This set is preordered by the relation $\hookrightarrow$ defined as follows: $\mathfrak{A}_p \hookrightarrow \mathfrak{A}_q$ if and only if $\mathfrak{A}_q \models p$. Equivalently, there exists an elementary submodel $\mathfrak{B}$ of the model $\mathfrak{A}_q$, written $\mathfrak{B} \preceq \mathfrak{A}_q$, such that $\mathfrak{A}_p \cong \mathfrak{B}$. Also, we say that a type p *is dominated by* a type q if $\mathfrak{A}_p \hookrightarrow \mathfrak{A}_q$, written $p \hookrightarrow q$.

Let T be an Ehrenfeucht theory. Note that $RK(T)$ has a least element. Indeed, if p is a principal type then for all $q \in S(T)$ we have $\mathfrak{A}_q \models p$. Of course, in this case $\mathfrak{A}_p$ is a prime model of the theory T, and hence the least element of $RK(T)$ is unique. Also, $RK(T)$ has a greatest element. To see that, take a powerful type p. Then $\mathfrak{A}_p \models q$ for all $q \in S(T)$. As we shall see later, a greatest element is not necessary unique. So we have a finite set of $\hookrightarrow$-equivalent greatest elements.

Of course, not every model of an Ehrenfeucht theory is almost prime. For example, if $\mathfrak{A}$ is saturated then it is not almost prime.

Lemma 1.1 (Sudoplatov [**18**]) *Suppose a theory T is Ehrenfeucht and $\mathfrak{A}$ is a model of T. Then there exist a type $p \in S(T)$ and an elementary chain of isomorphic models $\mathfrak{A}_0 \preceq \ldots \preceq \mathfrak{A}_n \preceq \ldots$, each of which is prime over p, such that $\mathfrak{A} = \bigcup\limits_n \mathfrak{A}_n$.*

Proof. Since every model of T is presentable as a union of an elementary chain of almost prime models, and there are only finitely many models in $RK(T)$, there is an infinite sub-chain of models isomorphic to some $\mathfrak{A}_p$. □

[1] RK stands for the Rudin–Keisler order. As we shall see, it is the preorder on types given by domination.

Due to this lemma, if a model is as in the last sentence of Lemma 1.1 and is not almost prime, call it a *limit model over the type* p. If a model is limit over some type p, call it a *limit model.* As we shall see later, there can be nonisomorphic models of an Ehrenfeucht theory that are limit over the same type.

Lemma 1.2 (Sudoplatov [**18**]) *Suppose* $\mathfrak{A}_p$ *and* $\mathfrak{A}_q$ *are nonisomorphic* $\hookrightarrow$*-equivalent (that is,* $\mathfrak{A}_p \hookrightarrow \mathfrak{A}_q \hookrightarrow \mathfrak{A}_p$*) almost prime models. Then there exists a model which is limit over* p *and limit over* q*.*

Proof. Form the chain $\mathfrak{A}_0 \preceq \mathfrak{A}_1 \preceq \ldots \preceq \mathfrak{A}_i \preceq \ldots$ where $\mathfrak{A}_i \cong \mathfrak{A}_p$ if i is even and $\mathfrak{A}_i \cong \mathfrak{A}_q$ if i is odd.

Consider the model $\mathfrak{A} = \bigcup\limits_n \mathfrak{A}_n$. This model is limit. Indeed, if it were almost prime, then it would be isomorphic to both $\mathfrak{A}_p$ and $\mathfrak{A}_q$, which is impossible. □

Define a function[2] $ln\colon RK(T) \to \omega$ as follows: $ln(\mathfrak{A}_p)$ is the number of pairwise nonisomorphic models $\mathfrak{M}_1, \ldots, \mathfrak{M}_t$ which are limit over types $q_1, \ldots, q_k$ such that each q_i is $\hookrightarrow$-equivalent to p. Because a union of an elementary chain of prime models of a fixed theory is also a prime model of this theory, $ln(\mathfrak{A}_p) = 0$ when p is principal.

Thus, if we have a finite preordered set $\langle X, \leqslant \rangle$ and a function $f\colon X \to \omega$ and we want to construct an Ehrenfeucht theory T such that there is an isomorphism

$$\varphi\colon \langle X, \leqslant \rangle \longrightarrow \langle RK(T), \hookrightarrow \rangle$$

that preserves f, i.e., $f(x) = ln(\varphi(x))$, then X and f must possess the following properties:

(1) There exists a unique least element a_0.
(2) There exists a greatest element $z_0 \neq a_0$. If z_1 and z_2 are both greatest, then $z_1 \leqslant z_2 \leqslant z_1$.
(3) $f(a_0) = 0$.
(4) $f(z_0) > 0$.
(5) If $x \leqslant y \leqslant x$ and $x \neq y$, then $f(x) > 0$.

If T is an Ehrenfeucht theory, then $RK(T)$ is the *E-order of the theory* T, $ln(T)$ is the *E-function of the theory* T, and the pair $(RK(T), ln)$ is the *E-parameters of the theory* T. We showed that E-parameters of every Ehrenfeucht theory possess the properties above.

Now we are going to describe one easy construction that is very useful for producing Ehrenfeucht theories with complicated E-parameters.

Let $\mathfrak{A} = \langle A; \Sigma_{\mathfrak{A}} \rangle$ and $\mathfrak{B} = \langle B; \Sigma_{\mathfrak{B}} \rangle$ be countable models of signatures $\Sigma_{\mathfrak{A}}$ and $\Sigma_{\mathfrak{B}}$ respectively such that $\Sigma_{\mathfrak{A}} \cap \Sigma_{\mathfrak{B}} = \varnothing$, both signatures have no function symbols, and $A \cap B = \varnothing$. Put $\mathfrak{M} = \langle A \cup B, \Sigma_{\mathfrak{A}}, \Sigma_{\mathfrak{A}}, P, Q \rangle$, where $P(x) \leftrightarrow x \in A$, $Q(x) \leftrightarrow x \in B$, the values for the constants in $\Sigma_{\mathfrak{A}}$ and in $\Sigma_{\mathfrak{B}}$ remain as before, and the predicates in $\Sigma_{\mathfrak{A}}$ ($\Sigma_{\mathfrak{B}}$) are thought of as false at any tuples not in A (B). The model $\mathfrak{M}$ is called a *direct sum* of the models $\mathfrak{A}$ and $\mathfrak{B}$ and is denoted by $\mathfrak{A} \oplus \mathfrak{B}$. The definition is naturally extended to the case of arbitrary structures; we need only replace function symbols by their graphs and add appropriate predicates. If T_1 and T_2 are theories of respective signatures Σ_1 and Σ_2, and $\mathfrak{N}_i \models T_i$, $i = 1, 2$, then the theory $\mathrm{Th}(\mathfrak{N}_1 \oplus \mathfrak{N}_2)$ is called *a direct sum* of the

[2] ln stands for the number of limit models.

theories T_1 and T_2 and is denoted by $T_1 \oplus T_2$. It is not hard to see that the definition of the direct sum of theories is sound (see, for example, [**3**][3]).

To understand the usefulness of the construction for producing new Ehrenfeucht theories, one can take the classic example of Ehrenfeucht theory T having three models and find E-parameters of the theory $T \oplus T \oplus T$ that indeed has 27 models. And we use direct sums in Section 3 to produce a non-arithmetical Ehrenfeucht theory having a computable model.

We conclude this section with an easy but useful result saying that usually an Ehrenfeucht theory has many homogeneous models.

Theorem 1.3 *Suppose T is an Ehrenfeucht theory and p is a type consistent with T. If $\mathfrak{A}_{p_1}, \ldots, \mathfrak{A}_{p_k}$ are all the elements of $RK(T)$ that are $\hookrightarrow$-equivalent to $\mathfrak{A}_p$, and $\mathfrak{N}_1, \ldots, \mathfrak{N}_m$ are all the models, each of which is limit over some of the types $p_1, \ldots, p_k$, then there exists a unique homogeneous model among the models $\mathfrak{A}_{p_1}, \ldots, \mathfrak{A}_{p_k}, \mathfrak{N}_1, \ldots, \mathfrak{N}_m$.*

Proof. Consider the set of types $S = \{q \mid \mathfrak{A}_p \models q\}$. It is not hard to check that the set S possesses the following properties:

(1) S is closed under rearrangements of variables in types;
(2) S is closed under the taking of a subtype;
(3) Any two types $p_1, p_2 \in S$ are subtypes of some type $q \in S$;
(4) For any type $p(\overline{x}) \in S$ and any formula $\varphi(\overline{x}, y)$, if $\exists y \varphi(\overline{x}, y) \in p$ then there exists a type $q(\overline{x}, y) \in S$ such that $p \cup \{\varphi\} \subseteq q$;
(5) For any two types $p_1(x_1, \ldots, x_k, y)$, $p_2(x_1, \ldots, x_k, z) \in S$, if

$$p_1 \upharpoonright \{x_1, \ldots, x_k\} = p_2 \upharpoonright \{x_1, \ldots, x_k\}$$

then there exists a $(k+2)$-type $q \in S$ such that $p_1 \subseteq q$, $p_2 = q \upharpoonright \{x_1, \ldots, x_k, z\}$.

Therefore, there exists a countable homogeneous model of T realizing precisely the types in S (see, for example, [**8**]). Because $\mathfrak{A}_{p_1}, \ldots, \mathfrak{A}_{p_k}, \mathfrak{N}_1, \ldots, \mathfrak{N}_m$ are all the models of T that realize precisely the set S, there must be a homogeneous model among them. It is unique because no two of the models are isomorphic. □

Note that if the models $\mathfrak{A}_{p_1}, \ldots, \mathfrak{A}_{p_k}, \mathfrak{N}_1, \ldots, \mathfrak{N}_m$ are as in the Theorem and $m \geqslant 1$ then the homogeneous model is among $\mathfrak{N}_1, \ldots, \mathfrak{N}_m$, that is, it is a limit model. Indeed, if we have an elementary chain of isomorphic homogeneous models, then its union is also homogeneous and is in fact isomorphic to the models of the chain. But, as we know, if there are no limit models over some type p, then there must be only one (up to isomorphism) almost prime model in the equivalence class of almost prime models containing $\mathfrak{A}_p$. Thus, we have a corollary:

Corollary 1.4 *(An almost prime model $\mathfrak{A}_p$ is homogeneous) $\Longleftrightarrow$ (there are no limit models over p) $\Longrightarrow$ (the element of $RK(T)$ containing $\mathfrak{A}_p$ contains nothing else).*

As we shall see, Theorem 1.3 has several useful corollaries.

[3] Please note that there is a translation mistake on the first page of this paper —it should be “not countably” instead of “uncountably”.

2 Decidable models

This section addresses the following general problem: Describe decidable models of Ehrenfeucht theories. More precisely, suppose we are given an Ehrenfeucht theory T; which models of T have decidable presentations?

We note first of all that decidable models are downward closed in $RK(T)$.

Proposition 2.1 *If* $\mathfrak{A}_p \hookrightarrow \mathfrak{A}_q$ *and* $\mathfrak{A}_q$ *is decidable then* $\mathfrak{A}_p$ *is decidable. Particularly, if* $\mathfrak{A}_p \hookrightarrow \mathfrak{A}_q \hookrightarrow \mathfrak{A}_p$*, and* $\mathfrak{A}_q$ *is decidable then* $\mathfrak{A}_p$ *is decidable.*

Proposition 2.2 *If* $\mathfrak{A}$ *is a decidable model that is limit over p then* $\mathfrak{A}_p$ *is decidable.*

Proof. Take $\overline{a}$ such that $\langle \mathfrak{M}_p, \overline{a}\rangle$ is a prime model. The set of types realized in $\langle \mathfrak{M}_p, \overline{a}\rangle$ is decidable. Hence, $\mathfrak{M}_p$ is decidable. □

Morley and Lachlan [**15**] were the first to show that decidable models are not upward closed in $RK(T)$. The following theorem, which is proved in [**5**], shows, in particular, that any finite linear ordering can be realized as $RK(T)$ for some *hereditarily decidable* (that is, all consistent types are computable) Ehrenfeucht theory T.

Theorem 2.3 *Let* $n \geqslant 1$ *and* $0 \leqslant k \leqslant n$ *be natural numbers and* $L_n = \{x_0 \leqslant \ldots \leqslant x_n\}$ *be a linear ordering having* $n+1$ *elements. Then there exists an Ehrenfeucht theory* T_{nk} *such that*

(1) $RK(T_{nk}) = \{\mathfrak{A}_0 \hookrightarrow \ldots \hookrightarrow \mathfrak{A}_n\} \cong L_n$*;*

(2) *The models* $\mathfrak{A}_0, \ldots, \mathfrak{A}_k$ *are decidable, and the models* $\mathfrak{A}_{k+1}, \ldots, \mathfrak{A}_n$ *do not even have computable presentations.*

We believe that the following statement, saying that the downward closedness in $RK(T)$ along with closedness under taking least upper bounds are the only restrictions for decidable almost prime models of a theory T, is true. Denote by $RK_d(T)$ the suborder of $RK(T)$ composed of decidable almost prime models of the theory T. It is not hard to see that if some models $\mathfrak{A}_p$ and $\mathfrak{A}_q$ are decidable and a model $\mathfrak{A}_r$ is the least upper bound of $\mathfrak{A}_p$ and $\mathfrak{A}_q$ in $RK(T)$, then $\mathfrak{A}_r$ is decidable.

Conjecture 2.4 If X is an E-order of some Ehrenfeucht theory and Y is a downward-closed sub-order of X which is closed under taking least upper bounds in Y, then there exists an Ehrenfeucht theory T such that

(1) $RK(T) \cong X$;

(2) $RK_d(T) \cong Y$.

And what about limit models? Limit models are a big problem. Almost every statement about decidability of limit models is either trivial or equivalent to the Morley Problem. Indeed, if we have a decidable saturated model (which is limit indeed) then all the types are decidable, then all the almost prime models are decidable, then anything you ask about decidability of the rest of the models depends on the Morley Problem. Thus, one can easily produce a number of open questions. We mention only two of them.

Question 2.5 Suppose that p is a decidable type of an Ehrenfeucht theory T and there is a model $\mathfrak{A}$ that is limit over p. Is there a decidable model which is limit over p?

Question 2.6 Suppose that p is a decidable type of an Ehrenfeucht theory T and there is a decidable model $\mathfrak{A}$ that is limit over p. Are all limit over p models decidable?

Question 2.5 was asked by Goncharov after the author's talk at the Logic Colloquium 2010. We answer this question positively. The answer follows from Theorem 1.3.

Answer. By Goncharov's theorem [**7**], every homogeneous model of an Ehrenfeucht theory that realizes only computable types is decidable. By Theorem 1.3, there is a homogeneous model which is limit over p. By Proposition 2.1, all the types this model realizes are decidable. □

Thus, Question 2.6 is actually equivalent to the Morley Problem.

Another corollary of Theorem 1.3 solves the Morley Problem positively for a large class of Ehrenfeucht theories. More precisely,

Corollary 2.7 *Suppose that T is an Ehrenfeucht theory and every type consistent with T is computable. Also, suppose that if p is a type consistent with T, then there is at most one limit model over p. Then every countable model of T is decidable.*

Proof. By Theorem 1.3, every countable model of T is almost prime or homogeneous. □

3 Computable models

In this section we shall consider computable presentations instead of decidable.

We start from the following question. If a model $\mathfrak{A}$ has a computable presentation, what is the complexity of its theory $\mathrm{Th}(\mathfrak{A})$? The upper bound is provided by the standard model of true arithmetic $\langle\omega; \leqslant, +, \times, s, 0\rangle$. But what about Ehrenfeucht theories? How complex could an Ehrenfeucht theory with a computable model be? How complex could an Ehrenfeucht theory be if all of its models are computable? As it is proved in [**3**], it can be of arbitrary arithmetical complexity (that is, Turing equivalent to $\mathbf{0}^{(n)}$). Recently B. Khoussainov and A. Montalbán [**12**] constructed an ω-categorical theory that has a computable model and is 1-equivalent to $\mathrm{Th}(\omega; \leqslant, +, \times, s, 0)$. This example can be easily reconstructed into an Ehrenfeucht theory. The only tool we need is direct sums. It is not hard to see that $\omega(T_1 \oplus T_2) = \omega(T_1) \times \omega(T_2)$ and that the Turing degree of $T_1 \oplus T_2$ equals the join of Turing degrees of T_1 and T_2 (see [**3**] for details). Thus we have

Theorem 3.1 *For each $n \geqslant 3$ there exists a complete theory T having exactly n countable models such that T has a computable model and is Turing equivalent to true first order arithmetic.*

The rest of the section is focused on the question of which models of an Ehrenfeucht theory can be presented computably.

Khoussainov, Nies, and Shore [**13**] were the first to show that computability of a limit over a type p model does not imply computability of the model which is prime over p. More precisely, they constructed an Ehrenfeucht theory having three countable models, the only computable model of which is the saturated. Gavryushkin [**4, 6**] shows that computable models are not downward closed under $\hookrightarrow$ in $RK(T)$ and are not necessarily intervals in $RK(T)$.

Moreover, Gavryushkin [**6**] constructs an example of an Ehrenfeucht theory T having six countable models such that

(1) $RK(T) = \{\mathfrak{A}_0 \hookrightarrow \mathfrak{A}_1 \hookrightarrow \mathfrak{A}_2\}$;

(2) $ln(\mathfrak{A}_0) = ln(\mathfrak{A}_1) = 0$, $ln(\mathfrak{A}_2) = 3$, that is, there are three nonisomorphic limit models over the powerful type;

(3) The models of T that have computable presentations are $\mathfrak{A}_0$ and two of the limit models.

Thus, we have two limit models that are limit over the same type, one of which is computable and the other is not. Note that these models are limit over the powerful type. And as we mentioned before, such a result for decidable models would imply a negative solution to the Morley Problem.

Decidability of one of $\hookrightarrow$-equivalent almost prime models implies decidability of them all. We finish the paper with a result saying that for computable models this is not true. More precisely, we shall show that for all $n \geqslant 3$ there is an Ehrenfeucht theory having exactly n countable models but only one of them is limit. (Note that *every* Ehrenfeucht theory has a limit model —the saturated one.) Moreover, such a theory can have an arbitrary number $\leqslant n$ of computable models.

Theorem 3.2 *Let $n \geqslant 1$ and m be natural numbers such that $1 \leqslant m \leqslant n+1$. There exists an Ehrenfeucht theory T such that*

(1) *T has exactly $n+2$ models;*
(2) *T has n nonisomorphic $\hookrightarrow$-equivalent almost prime models $\mathfrak{A}_1, \dots, \mathfrak{A}_n$;*
(3) *The models $\mathfrak{A}_m, \dots, \mathfrak{A}_n$ have computable presentations, yet $\mathfrak{A}_1, \dots, \mathfrak{A}_{m-1}$ do not.*

Proof. The argument is based on two ideas. The first one is existence of a Δ^0_2-set which is not the range of a limitwise monotonic function. Sets of this kind are independently constructed by Khissamiev [**11**] and Khoussainov–Nies–Shore [**13**]. And the second one is coding such sets into Ehrenfeucht theories.

Let n and m be natural numbers as in the theorem. Consider first the case $1 \leqslant m \leqslant n$.

For each tuple of cardinals $k_1, \dots, k_n \in (\omega+1)$, define a structure $Q(k_1, \dots, k_n)$ as follows.

Let $\mathbb{Q}$ be the set of rationals, and let $\mathbb{Q}_i = \{q \in \mathbb{Q} \mid i \leqslant q < i+1\}$ for $i \in \{1, \dots, n\}$. The domain of the structure $Q(k_1, \dots, k_n)$ is $\mathbb{Q}_1 \cup \ldots \cup \mathbb{Q}_n \cup C$, where

$$C = \bigcup_{i=1}^{n} \{c_{q,1}, \dots, c_{q,k_i} \mid q \in \mathbb{Q}_i\}$$

is a set of new elements. The signature of the structure is

$$\langle \leqslant; f_1, \dots, f_n, g_1, \dots, g_{n-1} \rangle,$$

where $\leqslant$ is a binary relation and $f_1, \dots, f_n, g_1, \dots, g_{n-1}$ are unary function symbols. The relation $\leqslant$ and the functions $f_1, \dots, f_n$ are defined as follows. For all x and y we have $x \leqslant y$ if and only if there is an $i \in \{1, \dots, n\}$ such that $x, y \in \mathbb{Q}_i$ and x is less than or equal to y as rational numbers. Let $i \in \{1, \dots, n\}$ and $x \in \mathbb{Q}_i \cup C$, define $f_i(x)$ in the following way.[4] If $x \in \mathbb{Q}_i$, then $f_i(x) = x$. If $x = c_{q,t}$ for some $q \in \mathbb{Q}_i$ and some t, then $f_i(x) = q$.

Let $i \in \{1, \dots, n-1\}$. Define $g_i \colon \mathbb{Q}_i \to \mathbb{Q}_{i+1}$ in the following manner:

(1) g_i is an order-preserving injection, that is, $(\forall x < y \in \mathbb{Q}_i) f_i(x) < f_i(y)$;
(2) $g_i(i) = i+1$;
(3) Both $\mathrm{Range}(g_i)$ and $\overline{\mathrm{Range}(g_i)}$ are dense in $\mathbb{Q}_{i+1}$, that is, $(\forall x < y \in \mathbb{Q}_{i+1})((\exists u, z)\, x < g_i(u) < y \;\&\; x < z < y \;\&\; (\forall v) g_i(v) \neq z)$;

[4] Here and further partial functions are allowed. If you prefer, think of functions as predicates —their graphs.

For $i \in \{0, \ldots, n\}$, denote by $Q_i(\omega)$ the structure obtained from $Q(\omega, \ldots, \omega)$ by removing the elements $1, c_{1,1}, \ldots, c_{1,\omega}; 2, c_{2,1}, \ldots, c_{2,\omega}; \ldots; i, c_{i,1}, \ldots, c_{i,\omega}$ from the domain of $Q(\omega, \ldots, \omega)$. (If $i = 0$, we do nothing and get $Q_0(\omega) = Q(\omega, \ldots, \omega)$.)

If $\mathfrak{A}$ and $\mathfrak{B}$ are isomorphic copies of the structures $Q(k_1, \ldots, k_n)$ and $Q(s_1, \ldots, s_n)$, respectively, and $A \cap B = \varnothing$, then one can naturally define the isomorphism type of the structure $Q(k_1, \ldots, k_n) + Q(s_1, \ldots, s_n)$ as follows. The domain of the new structure is the set $A \cup B$. The relation $\leqslant$ in the new structure is the least partial ordering which contains the partial ordering of $\mathfrak{A}$, the partial ordering of $\mathfrak{B}$, and the relation

$$\bigcup_{i=1}^{n} \{(x, y) \mid x \in A \;\&\; f_i^{\mathfrak{A}}(x) = x \;\&\; y \in B \;\&\; f_i^{\mathfrak{B}}(y) = y\}.$$

For $1 \leqslant i \leqslant n$ the unary function f_i in the new model is the union of the unary functions $f_i^{\mathfrak{A}}$ and $f_i^{\mathfrak{B}}$. For $1 \leqslant j \leqslant n-1$ the unary function g_j in the new model is the union of the unary functions $g_j^{\mathfrak{A}}$ and $g_j^{\mathfrak{B}}$.

If $k_1^j, k_2^j, \ldots, k_n^j$, $j < \omega$ are tuples of natural numbers then, as above, we can define the structure

$$Q(k_1^0, \ldots, k_n^0) + Q(k_1^1, \ldots, k_n^1) + \cdots + Q(k_1^j, \ldots, k_n^j) + \cdots.$$

Let S be a Δ_2^0-set which is not the range of a limitwise monotonic function **[11, 13]**. There exists a computable function β such that, for all x, the function $\alpha(x) = \lim\limits_{y \to \infty} \beta(x, y)$ is defined and $\mathrm{Range}(\alpha) = S$. For $i \in \{1, \ldots, n\} \setminus \{m\}$ put $(\forall x \in \omega)\gamma_i(x) = x$. Also, put $\gamma_m(x) = \alpha(x)$ for all x. Consider the model

$$\mathfrak{A}_0 = Q(\gamma_1(0), \gamma_2(0), \ldots, \gamma_n(0)) + Q(\gamma_1(1), \gamma_2(1), \ldots, \gamma_n(1)) + \cdots$$

and consider the theory T of the structure $\mathfrak{A}_0$.

Prove that T *has exactly* $n+2$ *models.*

$\mathfrak{A}_0$ is a prime model of T. The second model of T is the saturated model $\mathfrak{A}_{n+1} = \mathfrak{A}_0 + Q_n(\omega)$. And we have n $\hookrightarrow$-equivalent almost prime models $\mathfrak{A}_{i+1} = \mathfrak{A}_0 + Q_i(\omega)$, $0 \leqslant i \leqslant n-1$.

To see that these structures are indeed models of the theory T, prove that $\mathfrak{A}_0$ is an elementary submodel of each of these models. Clearly it is submodel. To see that it is an elementary submodel, pick some tuple of elements $\overline{a}$ from $\mathfrak{A}_0$, some formula of the form $\exists x \varphi(\overline{a}, x)$, and some model $\mathfrak{B} \in \{\mathfrak{A}_1, \ldots, \mathfrak{A}_{n+1}\}$. It can be checked that, if the formula $\exists x \varphi(\overline{a}, x)$ is true in $\mathfrak{B}$, then there is an element b from $\mathfrak{A}_0$ such that $\varphi(\overline{a}, b)$ is true in $\mathfrak{A}_0$, so the submodel is elementary.

We have to prove that any countable model of T is isomorphic to one of the $n+2$ models described above. Let $\mathfrak{A}$ be a model of T. Define by induction n sequences of elements $a_0^i, a_1^i, \ldots$, $1 \leqslant i \leqslant n$. Fix $i \in \{1, \ldots, n\}$. Let a_0^i be the minimal element with respect to the partial ordering in $\mathfrak{A}$ such that $f_i(a_0^i) = a_0^i$. Note that the set $\{b \mid b \neq a_0^i \;\&\; f_i(b) = a_0^i\}$ has exactly $\gamma_i(0)$ elements. Put $k_0^i = 0$. Suppose that the elements $a_0^i, \ldots, a_{t-1}^i$ and the numbers $k_0^i, \ldots, k_{t-1}^i$ have been defined. Let k_t^i be the least number such that $\gamma_i(k_t^i) \neq \gamma_i(k_j^i)$ for $j = 0, \ldots, t-1$. The element a_t^i is the one such that the following properties hold:

(1) The set $\{b \mid b \neq a_t^i \;\&\; f_i(b) = a_t^i\}$ has exactly $\gamma_i(k_t^i)$ elements;

(2) For each $x < a_t^i$ the cardinality of the set $\{b \mid b \neq x \;\&\; f_i(b) = x\}$ is in $\{\gamma_i(k_0^i), \ldots, \gamma_i(k_{t-1}^i)\}$.

Consider the sequences $a_0^i, a_1^i, \dots$, for $i \in \{1, \dots, n\}$. Clearly $a_0^i < a_1^i < \dots$. So we have $n+2$ cases:

Case 0. $\lim\limits_{t\to\infty} a_t^1$ does not exist and for any $x \in A$ such that $f_1(x) = x$ there exists a t such that $a_t^1 \geqslant x$.

Case 1. $\lim\limits_{t\to\infty} a_t^1$ exists.

Case i $(1 \leqslant i \leqslant n-1)$. $\lim\limits_{t\to\infty} a_t^i$ does not exist, $\lim\limits_{t\to\infty} a_t^{i+1}$ does exist.

Case $n+1$. $\lim\limits_{t\to\infty} a_t^i$ does not exist for all i, and there is an $x \in A$ such that $f_1(x) = x$ and $x \geqslant a_t^1$ for all t.

Note that there are no other options. Indeed,

Lemma 3.3

(1) $\exists i \in \{1, \dots, n\}$ *(there is an* $x \in A$ *such that* $f_i(x) = x$ *and* $x \geqslant a_t^i$ *for all* t*)* $\iff$ $\forall i \in \{1, \dots, n\}$ *(there is an* $x \in A$ *such that* $f_i(x) = x$ *and* $x \geqslant a_t^i$ *for all* t*).*

(2) *If* $\lim\limits_{t\to\infty} a_t^i$ *exists, then* $\lim\limits_{t\to\infty} a_t^{i+1}$ *exists as well.*

(3) *If* $\lim\limits_{t\to\infty} a_t^i$ *does not exist, then* $\lim\limits_{t\to\infty} a_t^{i-1}$ *does not exist either.*

Proof. For (1), if there is an $x \in A$ such that $f_i(x) = x$ and $x \geqslant a_t^i$ for all t, then $f_{i+1}(g_i(x)) = g_i(x)$ and $g_i(x) \geqslant a_t^{i+1}$ for all t. Furthermore, if $i > 1$ then there is a $y \in A$ such that $f_i(y) = y$, $y \geqslant a_t^i$ for all t, and $y = g_{i-1}(z)$ for some z. This z satisfies the required conditions.

For (2), $\lim\limits_{t\to\infty} a_t^{i+1} = g_i\Big(\lim\limits_{t\to\infty} a_t^i\Big)$. □

If Case k is realized, then $\mathfrak{A} \cong \mathfrak{A}_k$.

Now we can see directly from definitions that the model $\mathfrak{A}_0$ is prime, the model $\mathfrak{A}_{n+1}$ is limit over powerful non-principal types p_i saying that "there is an $x \in A$ such that $f_i(x) = x$ and $x \geqslant a_t^i$ for all t", $i \in \{1, \dots, n\}$, and the models $\mathfrak{A}_1, \dots, \mathfrak{A}_n$ are almost prime over these types respectively.

Let $\mathfrak{B}$ be a model of T. For $1 \leqslant i \leqslant n$, consider the restriction $\mathfrak{B}_i$ of the structure $\mathfrak{B}$ to the set $\{x \mid (\exists y) f_i(x) = y\}$. The signature of the structure $\mathfrak{B}_i$ is $\langle \leqslant, f_i \rangle$. It is not hard to see that the model $\mathfrak{B}$ has a computable presentation if and only if for every i the model $\mathfrak{B}_i$ has one. Also, it is easy that if $i \neq m$ then $\mathfrak{B}_i$ has a computable presentation; it is actually decidable. Thus, the model $\mathfrak{B}$ has a computable presentation if and only if the model $\mathfrak{B}_m$ does.

For the model $\mathfrak{B}_m$, we have three cases:

(1) $\lim\limits_{t\to\infty} a_t^m$ does not exist and for any $x \in B_m$ such that $f_m(x) = x$ there exists a t such that $a_t^m \geqslant x$;

(2) $\lim\limits_{t\to\infty} a_t^m$ exists;

(3) $\lim\limits_{t\to\infty} a_t^m$ does not exist, and there is an $x \in B_m$ such that $f_m(x) = x$ and $x \geqslant a_t^m$ for all t.

Note that these three cases are equivalent to the following three cases respectively (we recall that the Range(α) is a Δ_2^0-set that is not the range of a limitwise monotonic function):

(1) $\mathfrak{B}_m \cong Q(\alpha(0)) + Q(\alpha(1)) + \cdots$;

(2) $\mathfrak{B}_m \cong Q(\alpha(0)) + Q(\alpha(1)) + \cdots + Q(\omega)$;

(3) $\mathfrak{B}_m \cong Q(\alpha(0)) + Q(\alpha(1)) + \cdots + Q_1(\omega)$.

It can be checked (see, for example [**6, 13**]) that the model $\mathfrak{B}_m$ has a computable presentation if and only if Case 3 is realized.

Thus, an arbitrary model of the theory T has a computable presentation if and only if $\lim\limits_{t\to\infty} a_t^m$ does not exist, and there is an x such that $f_m(x) = x$ and $x \geqslant a_t^m$ for all t. But only the models $\mathfrak{A}_m, \mathfrak{A}_{m+1}, \ldots, \mathfrak{A}_{n+1}$ satisfy this property.

To finish the proof, we have to consider the case $m = n + 1$, that is, the case when all the models have no computable presentations. This case is easy —we can take the function γ_1 such that Range(γ_1) is a Δ_3^0-complete set, for instance. □

References

[1] C. Ash, J. Knight, *Computable Structures and the Hyperarithmetical Hierarchy,* Elsevier, 2000.

[2] C. C. Chang, H. J. Keisler, *Model Theory,* 3rd edition, North-Holland, 1990.

[3] A. Gavryushkin, Complexity of Ehrenfeucht models, *Algebra and Logic,* **45** (2006), 5, 289–295.

[4] A. Gavryushkin, Spectra of computable models for Ehrenfeucht theories, *Algebra and Logic,* **46** (2007), 3, 149–157.

[5] A. Gavryushkin, On constructive models of theories with linear Rudin–Keisler ordering, *Journal of Logic and Computation,* doi: 10.1093/logcom/exq043 (2010).

[6] A. Gavryushkin, Computable limit models, *Programs, Proofs, Processes —CiE* (2010), 188–193.

[7] S. S. Goncharov, Strong constructivizability of homogeneous models, *Algebra and Logic,* **17** (1973), 4, 247–263.

[8] S. S. Goncharov, A totally transcendental decidable theory without constructivizable homogeneous models, *Algebra and Logic,* **19** (1980), 2, 85–93.

[9] L. Harrington, Recursively presented prime models, *Journal of Symbolic Logic,* **39** (1974), 2, 305–309.

[10] N. G. Khissamiev, On strongly constructive models of decidable theories, *Izvestiya Akademii Nauk Kazakhskoi SSR. Seriya Fiziko-Matematicheskaya,* **1** (1974), 83–94.

[11] N. G. Khissamiev, Criterion for constructivizability of a direct sum of cyclic p-groups, *Izvestiya Akademii Nauk Kazakhskoi SSR. Seriya Fiziko-Matematicheskaya,* **86** (1981), 1, 51–55.

[12] B. Khoussainov, A. Montalbán, A computable $\aleph_1$-categorical structure whose theory computes true arithmetic, *Journal of Symbolic Logic,* **75** (2010), 2, 728–740.

[13] B. Khoussainov, A. Nies, R. Shore, Computable models of theories with few models, *Notre Dame Journal of Formal Logic,* **38** (1997), 2, 165–178.

[14] T. Millar, Homogeneous models and decidability, *Pacific Journal of Mathematics,* **91** (1980), 2, 407–418.

[15] M. Morley, Decidable models, *Israel Journal of Mathematics,* **25** (1976), 233–240.

[16] M. G. Peretyat'kin, On complete theories with a finite number of denumerable models, *Algebra and Logic,* **12** (1973), 5, 310–326.

[17] R. I. Soare, *Recursively Enumerable Sets and Degrees. A Study of Computable Functions and Computably Generated Sets,* Springer Verlag, Berlin, New York, 1987.

[18] S. V. Sudoplatov, Complete theories with finitely many countable models, *Algebra and Logic,* **43** (2004), 1, 62–69.

Part II

Computations and Proofs

Improved witnessing and local improvement principles for second-order bounded arithmetic

Arnold Beckmann[*], **Samuel R. Buss**[†]

[*] Department of Computer Science, Swansea University, UK
a.beckmann@swansea.ac.uk

[†] Department of Mathematics, University of California, San Diego, USA
sbuss@math.ucsd.edu

Abstract. This paper concerns the second order systems U^1_2 and V^1_2 of bounded arithmetic. We formulate improved witnessing theorems for these two theories by using S^1_2 as a base theory for proving the correctness of the polynomial space or exponential time witnessing functions. We develop the theory of nondeterministic polynomial space computation in U^1_2. Kołodziejczyk, Nguyen, and Thapen have introduced local improvement properties to characterize the provably total NP functions of these second order theories. We show that the strengths of their local improvement principles over U^1_2 and V^1_2 depend primarily on the topology of the underlying graph, not on the number of rounds in the local improvement games. The theory U^1_2 proves the local improvement principle for linear graphs even without restricting to logarithmically many rounds. The local improvement principle for grid graphs with only logarithmically rounds is complete for the provably total NP search problems of V^1_2. Related results are obtained for local improvement principles with one improvement round, and for local improvement over rectangular grids.

Introduction

A "multifunction" is a function which can have multiple values, namely a total relation. NP search problems are multifunctions f which have polynomial growth rate and whose graph is polynomial-time recognizable. The provably total NP search problems of a theory T of bounded arithmetic are the multifunctions which have polynomial time graph $G_f(x,y)$ such that T proves $(\forall x)(\exists y)G_f(x,y)$. If G_f is instead a Σ^b_i-formula, then f is a Σ^b_i-definable multifunction of T. The provably total NP search problems of T and the Σ^b_1-definable multifunctions of T are essentially the same, as the latter can be defined as projections of the former.

There have been a series of recent results giving new characterizations of the provably total NP search problems for theories of bounded arithmetic, and more generally the Σ^b_i-definable multifunctions of these theories. The most recent work in this direction includes [**1, 2, 8, 11, 13**]. The first four of these papers give characterizations of the Σ^b_i-definable functions of T^k_2 for all $0 \leq i \leq k$. Skelley and Thapen [**13**] introduce k-round game principles, GI_k, which characterize the provably total NP search problems of T^k_2. Beckmann and Buss [**1, 2**] used an extension of polynomial local search (PLS)

[*],[†]Both authors thank the John Templeton Foundation for supporting their participation in the CRM Infinity Project at the Centre de Recerca Matemàtica, Barcelona, Catalonia, Spain, during which this project was instigated. This research was partially done while the first author was a visiting fellow at the Isaac Newton Institute for the Mathematical Sciences in the programme "Semantics & Syntax".

[†]Supported in part by NSF grants DMS-0700533 and DMS-1101228, and by a grant from the Simons Foundation (#208717 to Sam Buss).

along with Skolemization techniques to characterize the Σ_i^b-definable multifunctions of T_2^k for $1 \leq i \leq k$. Pudlák and Thapen [**11**] gave another quite different characterization of the Σ_i^b-definable multifunctions of T_2^k based on alternating min-max principles. The fifth paper, Kołodziejczyk, Nguyen, and Thapen [**8**], extended the idea of the game principles to a "local improvement" principle and applied this to characterize the Σ_1^b-definable multifunctions of the second order theories U_2^1 and V_2^1. As we explain below in more detail, the present paper extends the results of [**8**] in several ways. The first part of the paper describes U_2^1 and V_2^1 and extends the bootstrapping of U_2^1 to show that U_2^1 can define nondeterministic polynomial space (NPSPACE) computations and can prove Savitch's theorem about the equivalence of deterministic and nondeterministic polynomial space. We then present improved witnessing theorems for U_2^1 and V_2^1. The final part of the paper improves the results of [**8**] that characterize the Σ_1^b-definable multifunctions in terms of the local improvement principles. Our two new results for local improvement principles of [**8**] are that U_2^1 can prove the principle LLI, and that the $\mathrm{LI}_{\log}$ principle is (provably) many-one complete for the total NP search problems of V_2^1. These improve results from [**8**], who had proved weaker versions of these results with $\mathrm{LLI}_{\log}$ and LI in place of LLI and $\mathrm{LI}_{\log}$, respectively.

The original witnessing theorems [**4**] for bounded arithmetic followed the following general template. These witnessing theorems were formulated to apply to a theory T, a formula class Φ, and a complexity class $\mathcal{C}$. In most cases, the complexity class $\mathcal{C}$ has complete problems, and the functions in the complexity class $\mathcal{C}$ can be enumerated by specifying an algorithm for the function that uses specified computational resources. A function that is specified in such a way is said to be "explicitly $\mathcal{C}$". The witnessing theorem then states that if $\phi \in \Phi$ and $T \vdash (\forall \vec{x})(\exists y)\phi(\vec{x}, y)$, then there is an explicitly-$\mathcal{C}$ function f such that (a) T proves the totality of f and (b) T proves $(\forall \vec{x})\phi(\vec{x}, f(\vec{x}))$. For this, T does not need to have a function symbol for f, rather there is a formula G_f defining the graph of f, and condition (a) actually means that T proves $(\forall \vec{x})(\exists y)G_f(\vec{x}, y)$. Likewise, condition (b) means that T proves $(\forall \vec{x})(\forall y)[G_f(\vec{x}, y) \supset \phi(\vec{x}, y)]$. Buss [**4, 5**] established these kinds of results for the theories S_2^k, T_2^k, U_2^1, and V_2^1, and for function classes such as polynomial time, levels of the polynomial hierarchy, polynomial space, and exponential time. Buss and Krajíček [**6**] proved a witnessing theorem for T_2^1 and PLS. And various authors have established a wide range of additional witnessing theorems; many of these are reported in a modern form in Cook–Nguyen [**7**].

In many cases, the witnessing theorem also includes a "uniqueness condition" that f is a function rather than a multifunction; namely, that T proves $(\forall \vec{x})(\exists! y)G_f(\vec{x}, y)$. However, there are some notable exceptions, namely those related to witnessing with PLS and game principles: these include (among others) [**1, 2, 6, 8**]. However, in these cases, the explicitly-$\mathcal{C}$ functions are conjectured to be inherently multifunctions rather than functions, so the (conjectured!) failure of the uniqueness condition is unavoidable.

In nearly every case, the witnessing theorem is accompanied with a converse result stating that every explicitly-$\mathcal{C}$ function is provably definable in T with its graph G_f a formula from Φ.

Some recent witnessing theorems have followed an improved paradigm, which provides an extension of the template described above. These "new-style" witnessing theorems were used implicitly in [**13**] and more explicitly in [**1, 2, 8**]. For the improved paradigm, the condition (b) of a witnessing theorem is replaced with

(b'): S_2^1 proves $(\forall \vec{x})(\forall y)[G_f(\vec{x}, y) \supset \phi(\vec{x}, y)]$.

That is, the correctness of the witnessing function f is now proved in the (weaker) theory S_2^1 rather than in T.[1] Of course, in these situations, it is generally conjectured that S_2^1 does not necessarily prove the totality of f; thus (b') includes the existence of $y = f(\vec{x})$ as a hypothesis. We shall prove two such new-style witnessing theorems for U_2^1 and V_2^1 in Section 3.

Section 1 reviews quickly the definitions of the theories U_2^1 and V_2^1. We presume, however, that the reader has basic familiarity with the bounded arithmetic theories S_2^1 and T_2^i and the syntactic classes Σ_i^b and Π_i^b. This section also introduces an alternate sequent calculus formulation of U_2^1 that will be useful for establishing normal forms for free-cut free proofs in U_2^1.

Section 2 shows that U_2^1 can formalize nondeterministic polynomial space computations. This is based on a formalization of Savitch's theorem that NSPACE(n) is contained in SPACE(n^2) and that hence PSPACE equals NPSPACE. The formalization of Savitch's theorem in U_2^1 is completely straightforward, but some care must be taken to show that it is possible for U_2^1 to pick out a particular nondeterministic computation path, including, for instance, the lexicographically first one. This construction is used in a crucial way for the proof of Theorem 4.8.

Section 3 establishes the two new-style witnessing theorems of U_2^1 and V_2^1. Of course, the two theories already have witnessing theorems linking them to polynomial space and exponential time computation, respectively. The new witnessing theorems use S_2^1 as a base theory as in (b') above, or more precisely, a conservative extension of S_2^1 to include second order variables. To formulate the witnessing theorem, we define a notion of what it means for a second order object (or, "predicate") to "canonically verify" the truth of a bounded ($\Sigma_0^{1,b}$) formula. We then prove two witnessing lemmas, over the base theory S_2^1, about the witnessing of sequents of $\Sigma_1^{1,b}$ formulas that are provable in U_2^1 or V_2^1, using polynomial space or exponential time (respectively) computable predicates.

Kołodziejczyk, Nguyen, and Thapen [**8**] already proved new-style witnessing theorems for U_2^1 and V_2^1 using closure under certain types of iteration. The results of Section 3 use a more straightforward definition for polynomial space and exponential time computation, along with the notion of canonical verification. In addition, Theorems 3.7 and 3.8 for V_2^1 use S_2^1 as a base theory, rather than the ostensibly stronger theory T_2^1 which was used by [**8**]. This improvement of using S_2^1 as the base theory will be crucial later for the proof of Theorem 4.9.

Section 4 discusses the local improvement principles of [**8**]. Loosely speaking, a local improvement principle uses a directed acyclic graph G: the vertices in the graph G are assigned labels with scores. Initially all labels have score value equal to zero, but a mechanism is provided to make local updates to labels that increments scores by one. This local update proceeds by sweeping across the graph, and is well-defined since the graph is acyclic. In essence, the local improvement principle states that the scores can be incremented for a certain number, c, of rounds. (The actual formulation of the local improvement principles will be as a set of contradictory assertions, which yields an NP search problem.)

There are two kinds of local improvement principles: the principle LI has underlying graph G on N vertices with constantly bounded in- and out-degrees, and LLI uses a

[1] So far, new-style witnessing theorems have been proved only for theories T that contain S_2^1. It should be straightforward to extend these results to use even weaker theories than S_2^1. No new-style witnessing theorems have been proved yet for theories $T \subseteq S_2^1$.

linearly ordered set of N points as its underlying graph. (The value N will be first order, but not sharply bounded.) The principles LI and LLI both use $c = N^{O(1)}$ many rounds of score increases. Limiting the number of rounds to instead be $c = O(\log N)$ gives the $\mathrm{LI}_{\log}$ and $\mathrm{LLI}_{\log}$ principles. When using exactly $c = 2$ rounds, the principles are called LI_2 and LLI_2.

Prior work [**8**] proved, for T the theory U^1_2 (respectively, V^1_2), that the $\mathrm{LLI}_{\log}$ principle (respectively, the LI principle) is provable in T, and is many one complete for the provably total NP search problems of T, provably in S^1_2. Section 4 concludes with new improved results; namely, that U^1_2 proves the LLI principle, and that the $\mathrm{LI}_{\log}$ principle is many-one complete for the provably total functions of V^1_2, provably in S^1_2. In fact, it follows that LLI and $\mathrm{LLI}_{\log}$ are equivalent over S^1_2, and that LI and $\mathrm{LI}_{\log}$ are equivalent over S^1_2. In particular, the strength of these local improvement principles depends on the underlying topology of the directed graph G, not on whether the number c of rounds is logarithmic or polynomial.

The *rectangular* local improvement principles, RLI, are the versions of LI where the graph G is a grid graph. We prove that the RLI and $\mathrm{RLI}_{\log}$ principles are equivalent to each other and to LI and $\mathrm{LI}_{\log}$, over S^1_2. For local improvement principles with two rounds, we prove that, over S^1_2, the LI_2 principle is equivalent to the last four mentioned principles, and that RLI_2 is equivalent to LLI and $\mathrm{LLI}_{\log}$. However, the strength of RLI_k for constant $k \geq 3$ remains an open question.

Sections 4.2 through 4.4 present the proofs of our results on the local improvement properties.

We thank Neil Thapen for useful discussions on the topics of this paper.

1 Preliminaries for U^1_2 and V^1_2

We assume the reader is familiar with the essentials of bounded arithmetic, for which see [**4, 9**]; however, we give a quick review to establish notation. Most of the paper is concerned with second order theories in the form defined in Chapter 9 of [**4**]. Since these second order theories are less well-known, we describe them below in a bit more detail. Our theories all use the non-logical language $0, S, +, \cdot, |\cdot|, \#, \leq$. Quantifiers of the form $(\exists x \leq t)$ and $(\forall x \leq t)$ are called *bounded quantifiers.* If the term t is of the form $|s|$, the quantifier is *sharply bounded.* The classes Σ^b_i and Π^b_i are defined by counting alternations of bounded quantifiers, ignoring sharply bounded quantifiers. The theories S^i_2 are axiomatized with a set, BASIC, of open axioms defining the non-logical symbols plus the Σ^b_i-PIND induction, namely polynomial induction, or equivalently, length induction. The theories T^i_2 are axiomatized with the axioms of BASIC plus Σ^b_i-IND, namely the usual induction axioms. Restricting to the case of $i = 1$, the main witnessing theorems for S^1_2 and T^1_2 state that S^1_2 can Σ^b_1-define precisely the polynomial time functions [**4**], and that T^1_2 can Σ^b_1-define precisely the PLS (polynomial local search) multifunctions [**6**].

Second order theories of bounded arithmetic extend the first order theories by adding second order variables, $X, Y, Z, \ldots$, intended to range over sets, also called "predicates". The membership $\in$ symbol is added to the language as well; the formula $t \in X$ denotes that t is in X. We often write $X(t)$ instead of $t \in X$. It is convenient to now let the classes Σ^b_i and Π^b_i involve free second order variables (but no quantified second order variables). Thus, a *bounded quantifier* is a bounded, first order quantifier; a *bounded formula* is a formula with no unbounded first order quantifiers and no second order quantifiers. We also let S^i_2 and T^i_2 now be defined with second order variables allowed to appear in formulas, including as free variables in induction axioms. But again, second order quantifiers are not allowed in induction formulas for S^i_2 and T^i_2. (Sometimes these extensions of S^i_2 and T^i_2 to second order logic are denoted S^{i+}_2 and T^{i+}_2, but since there is no chance of confusion, we prefer to omit the superscript "+". Likewise, we eschew the notations Σ^{b+}_1 and Π^{b+}_1.)

We reserve lower-case letters $a, b, c, \ldots$ and $z, y, x, \ldots$ for first order variables, and upper-case letters $A, B, C, \ldots$ and $Z, Y, X, \ldots$ for second order variables. Occasionally, we use Greek letters α, β, γ for second order variables as well. We use ϕ, ψ, and χ for formulas.

Second order bounded formulas are classified with the classes $\Sigma^{1,b}_i$ and $\Pi^{1,b}_i$ by counting the alternations of second order quantifiers, ignoring any first order quantifiers. The class $\Sigma^{1,b}_0$ is the set of bounded formulas, namely the set of formulas with no second order quantifiers but with arbitrary (first order) bounded quantifiers. The class $\Sigma^{1,b}_1$ is the set of formulas with all second order quantifiers essentially existential (that is, existential after negations are pushed inward), and arbitrary bounded first order quantifiers.

The theories U^1_2 and V^1_2 both contain all of T_2, plus the $\Sigma^{1,b}_0$-comprehension axioms, namely

$$(\forall\vec{x})(\forall\vec{X})(\exists Z)(\forall y \le t)[y \in Z \leftrightarrow \phi(y, \vec{x}, \vec{X})] \tag{1.1}$$

for every bounded formula ϕ and term t. This axiom states that any set (on a bounded domain) defined by a bounded formula ϕ with parameters is coded by some second order object Z.[2] The theory U^1_2 has in addition the $\Sigma^{1,b}_1$-PIND axioms. The theory V^1_2 has instead the $\Sigma^{1,b}_1$-IND axioms. It is known that $V^1_2 \vdash U^1_2$.

Note that the $\Sigma^{1,b}_0$-comprehension axiom above is a $\Pi^{1,b}_2$-sentence; or, stripping off leading universal quantifiers, it is a $\Sigma^{1,b}_1$-formula, in fact a strict $\Sigma^{1.b}_1$-formula, as will be defined momentarily.

As a side remark, we note that the second order systems can be conservatively extended to include second order function variables which range over functions with a specified polynomial growth rate. Then, $\Sigma^{1,b}_0$-comprehension implies the following $\Sigma^{1,b}_0$ function comprehension axiom for a function symbol δ with growth rate bounded by the term s:

$$(\forall\vec{x})(\forall\vec{X})(\exists\delta)(\forall y \le t)[(\exists z \le s)\phi(y, z, \vec{x}, \vec{X}) \supset \delta(y) {\le} s \wedge \phi(y, \delta(y), \vec{x}, \vec{X})]$$

where ϕ is a $\Sigma^{1,b}_0$-formula [**4**, p. 164]. The $\Sigma^{1,b}_0$ function comprehension axiom is effectively subsumed by $\Sigma^{1,b}_0$-comprehension, since $\Sigma^{1,b}_0$-comprehension can define the bit graph of δ so that $\delta(y)$ equals the least $z \le s$ satisfying $\phi(y, z, \vec{x}, \vec{X})$, if any such z exists.

However, for simplicity and without loss of generality, we formulate U^1_2 and V^1_2 with only second order predicate symbols and without second order function symbols.[3]

[2] The original definition [**4**] of U^1_2 used an unbounded version of the comprehension axiom; namely, the bounded quantifier $(\forall y \le t)$ was replaced with the unbounded quantifier $(\forall y)$. In the present paper, we are interested in only $\Sigma^{1,b}_i$-consequences of U^1_2, and, by Parikh's theorem, the unbounded version of comprehension gives no additional $\Sigma^{1,b}_i$-consequences. See alternately the discussion of the theories U^1_2(BD) and U^1_2(BD) in [**4**]. At any rate, subsequent authors have preferred the bounded versions of comprehension (e.g., [**7, 9, 10**]), perhaps because it is better behaved model-theoretically.

[3] Theorem 5 of Chapter 9 of [**4**] proves the conservativity between the theories with and without function symbols.

A function $f(\vec{x})$ is said to be $\Sigma^{1,b}_1$-*defined* by a theory T provided that T proves $(\forall\vec{x})(\exists y)\phi(\vec{x}, y)$ where $\phi(\vec{x}, y)$ defines the graph of f and $\phi \in \Sigma^{1,b}_1$. The original witnessing theorems of [**4**, Ch. 10] for U^1_2 and V^1_2 characterize their $\Sigma^{1,b}_1$-defined functions in terms of computational complexity. Namely, U^1_2 can $\Sigma^{1,b}_1$-define precisely the functions which are computable by polynomial space Turing machines, and V^1_2 can $\Sigma^{1,b}_1$-define precisely the functions which are computable in exponential time (that is, time $2^{n^{O(1)}}$). A formula $\psi(\vec{x}, \vec{X})$ is said to be $\Delta^{1,b}_1$-definable by T provided that T proves ψ is equivalent to both a $\Sigma^{1,b}_1$-formula and a $\Pi^{1,b}_1$-formula. A corollary to the witnessing theorems for U^1_2 and V^1_2 states that the $\Delta^{1,b}_1$-predicates of U^1_2 (respectively, V^1_2) are precisely the polynomial space predicates (respectively, the exponential time predicates). In addition, U^1_2 can prove the $\Delta^{1,b}_1$-IND and $\Delta^{1,b}_1$-MIN principles (see Theorem 16 of Chapter 9 of [**4**]). This means that U^1_2 can use polynomial space predicates and functions freely for induction and minimization. In short, U^1_2 can carry out a range of arguments about polynomial space predicates and functions.

We now define the notion of "strict" $\Sigma^{1,b}_1$-formula in analogy with similar notions for bounded formulas. A $\Sigma^{1,b}_1$-formula is *strict* provided that it contains at most one second order existential quantifier, and this quantifier is the outermost connective. That is, a formula is strict $\Sigma^{1,b}_1$ provided either it is either a bounded formula, or it has the form $(\exists X)\phi$ where ϕ is bounded. By comparison, a non-strict $\Sigma^{1,b}_1$-formula may have connectives and bounded quantifiers in front of the second order quantifiers. We shall sometimes use the notation $\mathrm{s}\Sigma^{1,b}_1$ to denote the class of strict $\Sigma^{1,b}_1$-formulas.

It is useful to restrict proofs to contain only *strict* $\Sigma^{1,b}_1$-formulas as this will considerably simplify the proofs of the new-style witnessing theorems of Section 3. Both U^1_2 and V^1_2 can prove that any $\Sigma^{1,b}_1$-formula is equivalent to a strict $\Sigma^{1,b}_1$-formula by using $\Sigma^{1,b}_1$-replacement principles, which are theorems of both U^1_2 and V^1_2 (see Theorem 16 of Chapter 9 of [**4**]). Thus it is reasonable to assume that any free-cut free U^1_2- or V^1_2-proof of a strict $\Sigma^{1,b}_1$ formula could be restricted to contain only strict $\Sigma^{1,b}_1$-formulas.

For V^1_2 this works readily. It is easy to check that V^1_2 can prove the $\Sigma^{1,b}_1$-replacement principles, and more generally prove the equivalence of any given $\Sigma^{1,b}_1$-formula to a strict $\Sigma^{1,b}_1$ formula, while using induction only on strict $\Sigma^{1,b}_1$ formulas. Thus, the usual free-cut elimination theorem (cf. [**3**]) gives the following.

Theorem 1.1 *Suppose V^1_2 proves a sequent $\Gamma \longrightarrow \Delta$ of strict $\Sigma^{1,b}_1$-formulas. Then there is a V^1_2-proof of $\Gamma \longrightarrow \Delta$ in which every formula is strict $\Sigma^{1,b}_1$.*

For U^1_2, the situation is less simple. The known proof in U^1_2 that every $\Sigma^{1,b}_1$-formula is equivalent to a strict $\Sigma^{1,b}_1$-formula uses induction on non-strict $\Sigma^{1,b}_1$-formulas.[4] Thus, free-cut elimination seemingly cannot be used with the usual formulation of U^1_2 to obtain proofs containing only strict $\Sigma^{1,b}_1$-formulas as cut formulas.

We can instead use a trick, and work with a slightly reformulated version of the theory U^1_2 called U^{1*}_2.

[4] The proof of $\Sigma^{1,b}_1$-replacement in U^1_2 given for Theorem 16 of Chapter 9 of [**4**] uses a doubling trick that seems to depend essentially on the use of non-strict $\Sigma^{1,b}_1$-formulas.

Definition 1.2 An $s\Sigma_1^{1,b}$-repl-$\forall$ inference is an inference of the form

$$\frac{a \le t, \Gamma {\longrightarrow} \Delta, (\exists X)\phi(X,a)}{\Gamma {\longrightarrow} \Delta, (\exists Y)(\forall x \le t)\phi(\{z\}Y(\langle x,z\rangle),x)}$$

where ϕ is a $\Sigma_0^{1,b}$-formula, a is an eigenvariable, and $\langle x,z\rangle$ is the usual pairing function used for bounded arithmetic. The notation $\{z\}Y(\langle x,z\rangle)$ denotes an *abstract* in the sense of Takeuti [**14**]. An abstract is akin to a lambda term but is not a syntactic part of the language; instead it is removed by the process of substitution. Namely, $\phi(\{z\}Y(\langle x,z\rangle),x)$ is the formula obtained from $\phi(X,a)$ by replacing every occurrence of the variable a with x, and every occurrence of any subformula $X(s)$ with $Y(\langle x,s\rangle)$.

Definition 1.3 The theory U_2^{1*} is defined to be U_2^1, but with $s\Sigma_1^{1,b}$-PIND instead of $\Sigma_1^{1,b}$-PIND, and with $s\Sigma_1^{1,b}$-repl-$\forall$ as an additional rule of inference.

It is clear that the $s\Sigma_1^{1,b}$-repl-$\forall$ inference is a derived rule of inference for U_2^1, although proving this in U_2^1 involves a cut on a non-strict $\Sigma_1^{1,b}$-formula. Therefore, U_2^1 proves all theorems of U_2^{1*}.

Theorem 1.4 *In U_2^{1*}, every $\Sigma_1^{1,b}$-formula can be proved equivalent to a strict $\Sigma_1^{1,b}$-formula.*

The theorem is straightforward to prove. The proof is by induction on the complexity of formulas and uses the $s\Sigma_1^{1,b}$-repl-$\forall$ rule to handle the hard case of moving a bounded quantifier past a second order quantifier.

As a corollary, U_2^{1*} admits PIND induction on all $\Sigma_1^{1,b}$-formulas. This immediately implies the equivalence of U_2^1 and U_2^{1*}.

Corollary 1.5 *U_2^1 and U_2^{1*} have the same consequences.*

The difference between U_2^1 and U_2^{1*} is only that they have different formalizations for sequent calculus proofs. The sequent calculus is formalized in a standard way. It uses conventional rules for weak inferences, for cut, and for first order connectives. In place of induction axioms, it uses induction rules with side formulas. The theory U_2^{1*} admits the $s\Sigma_1^{1,b}$-repl-$\forall$ rule of inference. The comprehension axiom (1.1) becomes the initial sequents

$$\longrightarrow (\exists Z)(\forall y \le t)[y \in Z \leftrightarrow \phi(y,\vec{x},\vec{X})].$$

This allows the second order $\exists$:right and $\forall$:right axioms to be formulated with only second order variables (instead of the more general substitution of abstracts). Namely, the two rules for second order existential quantifiers are:

$$\exists\text{:right}\ \frac{\Gamma {\longrightarrow} \Delta, \phi(A)}{\Gamma {\longrightarrow} \Delta, (\exists X)\phi(X)} \qquad \text{and} \qquad \exists\text{:left}\ \frac{\phi(A), \Gamma {\longrightarrow} \Delta}{(\exists X)\phi(X), \Gamma {\longrightarrow} \Delta}$$

where, for the $\exists$:left rule, the second order variable A is a eigenvariable and does not appear in the lower sequent of the inference. Dual rules are used for second order universal quantifiers. (However, second order universal quantifiers are never needed in our free-cut free proofs.)

Eliminating free-cuts from U_2^{1*}-proofs gives the following theorem.

Theorem 1.6 *Suppose U_2^1 proves a sequent $\Gamma \longrightarrow \Delta$ of strict $\Sigma_1^{1,b}$-formulas. Then there is a U_2^{1*}-proof of $\Gamma \longrightarrow \Delta$ in which every formula is strict $\Sigma_1^{1,b}$.*

It will be convenient to work with U_2^{1*} instead of U_2^1 for our witnessing constructions in Section 3. The downside of having s$\Sigma_1^{1,b}$-repl-$\forall$ as an additional inference is more than offset by the convenience of working with only strict $\Sigma_1^{1,b}$-formulas in the proof of the witnessing lemmas.

2 Nondeterministic polynomial space in U_2^1

We next formalize, in U_2^1, Savitch's theorem **[12]** that nondeterministic polynomial space is equal to polynomial space. An important consequence for us is that this means that U_2^1 can use induction (IND) and minimization (MIN) on NPSPACE predicates. It turns out that Savitch's argument can be carried out inside U_2^1 without complications; nonetheless, it is useful to check the details of exactly how it is formalized.

Figure 1 shows the usual algorithm behind Savitch's theorem. We assume that M is a nondeterministic Turing machine, running on input w, with explicit polynomial space bound $p(n)$ where $n = |w|$. For convenience, we use the convention that the input w is written on a read only input tape. A *configuration* of $M(w)$ is a complete description of M's tape contents, head positions, and current state at a given instant of time.

Let C_{init} be the initial configuration of $M(w)$. We may assume without loss of generality that if there is an accepting computation for $M(w)$, then it ends with a known configuration C_{end} after a known number of steps t_{end}. Then, to determine if $M(w)$ has an accepting computation, one merely invokes

$$\textsc{Reachable}(\ w,\ C_{\text{init}},\ 0,\ C_{\text{end}},\ t_{\text{end}}\ ). \tag{2.1}$$

It is well-known that Savitch's algorithm uses only polynomial space. In fact, it is straightforward to formalize Savitch's algorithm in U_2^1 as an explicitly polynomial space bounded computation.

With efficient coding, a configuration of $M(w)$ can be written out with $d\cdot p(n)$ many bits; thus a configuration can be coded by a number $C < 2^{d\cdot p(n)}$. For convenience, we shall use $\text{Bd}_M(n)$ to denote the term $d \cdot p(n)$ bounding the lengths of codes of configurations of M. It is not particularly important how configurations C are coded, but it is important that it be done in a straightforward matter so that information about the tape contents, tape head positions, current state, etc., can be extracted by polynomial functions of C, and so that our base theory S_2^1 can prove elementary properties about configurations, including whether one configuration succeeds another, or what the possible next moves are from a given configuration.

Note that the algorithm in Figure 1 has a line for marking a configuration C as being "identified" as the time t configuration. It can certainly happen that more than one configuration C is identified for a particular time t. Indeed, suppose a recursive call $\textsc{Reachable}(w, C_1, t_1, C, t)$ returns true. Then certainly some configuration is identified for each time $t' \in (t_1, t)$. If, however, the next call $\textsc{Reachable}(w, C, t, C_2, t_2)$ returns false, then the Savitch algorithm proceeds to the next value of C, and retries the calls with the new value for C. This of course, can cause new configurations to be identified for the times $t' \in (t_1, t)$, etc.

Accordingly, when a particular call (2.1) to $\textsc{Reachable}$ returns `TRUE`, we are interested in the *last* configuration that is identified as the time t configuration. Let $C[t]$ denote this last such configuration. We claim that the sequence of configurations $C[0]$, $C[1]$, $C[2], \ldots$, $C[t_{\text{end}}]$ is in fact an accepting computation for the Turing machine M

```
Reachable( $w$, $C_1$, $t_1$, $C_2$, $t_2$ )

// $C_1$ and $C_2$ are configurations, and $t_1 < t_2$.
if $t_2 = t_1 + 1$ then
   if ($C_2$ follows from $C_1$ by one step of $M$) then
      return TRUE
   else
      return FALSE
   end if
else
   set $t$ := $\lfloor (t_1 + t_2)/2 \rfloor$
   set $C$ := $0$
   loop while $C < 2^{\mathrm{Bd}_M(|w|)}$
      if ( $C$ codes a valid configuration of $M(w)$
               and Reachable($w$, $C_1$, $t_1$, $C$, $t$)
               and Reachable($w$, $C$, $t$, $C_2$, $t_2$) )
         Mark $C$ as the identified configuration for time $t$.
         return TRUE
      end if
      set $C$ := $C + 1$
   end loop
   return FALSE
end if
```

Figure 1. Savitch's algorithm is a recursively invoked procedure that does a depth first, divide-and-conquer, search for an accepting computation. It determines whether, starting in configuration C_1 at time t_1, the Turing machine M with input string w can reach configuration C_2 at time $t_2 > t_1$ by some nondeterministic computation.

on input w, where $C[0]$ and $C[t_{\mathrm{end}}]$ are C_{init} and C_{end}. We shall call this sequence of configurations the "Savitch computation" of $M(w)$.

A computation of $M(w)$ consists of $t_{\mathrm{end}} + 1$ many configurations, each coded by a string of $d{\cdot}p(n)$ bits. Accordingly, the entire computation can be coded by $(t_{\mathrm{end}}{+}1){\cdot}d{\cdot}p(n)$ many bits, where $n = |w|$. Since t_{end} is exponentially bounded in n, an entire computation of $M(w)$ can be coded, in U^1_2, by a second order object X. Namely, by letting $X(i)$ have truth value equal to the i-th bit of the computation, for $i < (t_{\mathrm{end}} + 1) \cdot d \cdot p(n)$.

The claim is that U^1_2 can prove that if Reachable(w, C_1, t_1, C_2, t_2) returns TRUE, then there is an X coding the entire Savitch computation of $M(w)$. A sketch of the proof is as follows. First note that, for fixed inputs w, C_1, t_1, C_2 and t_2, there must be some second order object Z coding the entire computation of the call to Reachable. Consequently, $C[t]$ is computable in polynomial space (from w and t), namely by examining Z. (In fact, without loss of generality, $C[t]$ is computable in polynomial time from Z.) The execution of Reachable as coded by Z contains many invocations of Reachable(w, C_1, t_1, C_2, t_2). Using either IND on the depth of the recursive calls, or PIND on the values $t_2 - t_1$, it can be proved that for any such invocation Reachable(w, C_1, t_1, C_2, t_2) which returns TRUE, the sequence C_1, $C[t_1 + 1], \ldots,$

$C[t_2 - 1]$, C_2 identified during the invocation is a valid computation for $M(w)$ starting in configuration C_1 and ending at C_2. The base case of the induction argument is trivial, and the induction step is immediate.

In addition, we have the following theorem.

Theorem 2.1 *Let M be an explicitly polynomial space nondeterministic Turing machine. Then U^1_2 proves the following statement: "If there is a Y coding an accepting computation of $M(w)$, then* REACHABLE$(w, C_{\mathrm{init}}, 0, C_{\mathrm{end}}, t_{\mathrm{end}})$ *returns* `TRUE`. *Conversely, if* REACHABLE$(w, C_{\mathrm{init}}, 0, C_{\mathrm{end}}, t_{\mathrm{end}})$ *returns* `TRUE`*, then there exists an X coding the entire Savitch computation, and this is an accepting computation of $M(w)$".*

The first part of the theorem is proved by noting that the REACHABLE algorithm cannot fail to accept when it reaches the computation coded by Y.

Theorem 2.1 implies further that U^1_2 can prove natural properties about the existence of nondeterministic polynomial space computations. An example of this is that U^1_2 can prove that it is possible to concatenate two partial computations. To formalize this, we can extend the notion of a Savitch computation to talk about the Savitch computation that starts at configuration C_1 at time t_1 and ends at configuration C_2 at time t_2. Then, we claim that U^1_2 can prove that if there are Savitch computations X and Y, one from C_1 at time t_1 to C_2 at time t_2 and the other from C_2 at time t_2 to C_3 at time t_3, then there is a Savitch computation from C_1 at time t_1 to C_3 at time t_3. Of course, the two computations X and Y cannot be merely concatenated to give a Savitch computation, since they may have different lengths, so their divide-and-conquer splitting points do not line up. Instead, however, their concatenation does give a (non-Savitch) computation, and then Theorem 2.1 implies the existence of the desired Savitch computation from C_1 to C_3.

Savitch computations provide a kind of canonical accepting computation; that is, if there is some accepting computation, then the Savitch computation exists and is unique. However, Savitch computations are a bit unnatural since they depend on the divide-and-conquer algorithm. An arguably more natural notion of canonical computation is a "lex-first" computation, which is defined as follows. We assume that each configuration has exactly two possible successor computations that can be reached in a single step. These two successors can be called the 0-successor and the 1-successor, say according to the order they appear in the transition relation table. In other words, we think of a nondeterministic algorithm of choosing exactly one random bit in each step, and moving according to that bit. A string Z of t_{end} many bits then fully specifies a computation. A *lex-first accepting computation* is defined to be the computation that arises from the lexicographically first Z that gives an accepting computation. Note that the string Z is exponentially long, and thus is represented in U^1_2 by the values of a second order object.

Of course the property that Z gives rise to a lex-first computation can be expressed as a $\Pi^{1,b}_1$-property since it states that there does not exist a Z' lexicographically preceding Z which specifies an accepting computation. However, U^1_2 can also express this as a $\Delta^{1,b}_1$-property. To see this, let $C_Z[i]$ be the configuration reached after making i steps according to Z, and let $C'_Z[i]$ be the computation reached after making $i-1$ steps according to Z but making the i-step with the choice opposite to Z. Then Z gives rise to a lex-first computation if and only if, for each value i such that $Z(i) = 1$, there is no computation from $C'_Z[i]$ to the accepting configuration. The last condition is an coNPSPACE property, hence PSPACE; so the entire condition is $\Delta^{1,b}_1$.

Theorem 2.2 *Let M be an explicitly polynomial space nondeterministic Turing machine. Then U^1_2 proves: "If there is a Y coding an accepting computation of $M(w)$, then there exists a lex-first accepting computation of $M(w)$".*

The idea of the proof of Theorem 2.2 is the following: The Turing machine M, including its nondeterministic choices, is simulated step-by-step by a deterministic PSPACE algorithm M'. At each step, M' invokes a PSPACE algorithm to check whether there exists an accepting computation starting from the 0-successor of the current configuration. If so, M' selects the 0-successor as the next configuration of M. Otherwise, the 1-successor is selected. It is obvious that M' selects the lex-first accepting computation of M if one exists. It is furthermore straightforward to show U^1_2 proves this.

3 Improved witnessing theorems for U^1_2 and V^1_2

This section states and proves the improved, new-style witnessing theorems for U^1_2 and V^1_2. First, we need to define what it means for a polynomial space or exponential time computation to output either a first order or second order object. Second, in Section 3.1, we define what it means for a (polynomial space) computation to "canonically evaluate" the truth of a $\Sigma^{1,b}_0$- or $\Sigma^{1,b}_1$-formula. The intuition behind this is simple: in order to canonically verify the truth of such a formula, the PSPACE algorithm does a brute force evaluation by considering all possible values for the first order quantified variables. However, the unexpected aspect is that all this must be formalizable in the weak base theory S^1_2, since the new-style witnessing theorems use S^1_2 as the base theory.

Sections 3.2 and 3.3 then state and prove the two new witnessing theorems and their associated witnessing lemmas.

We first establish some further conventions on how Turing machine computations are coded by second order objects and how they produce outputs. The previous section already discussed how configurations and complete computations are coded for polynomial space computations. This notion needs to be extended to handle exponential time computations. Suppose that M is a Turing machine, with input w of length n, and that M is either explicitly polynomial space or explicitly exponential time. The running time of M is bounded by a term t_{end} with value $t_{\text{end}} < 2^{q(n)}$ for some polynomial q. Configurations of $M(w)$ are to be coded in some straightforward way by a string of length $\leq \text{Bd}_M(w)$. For M in PSPACE, $\text{Bd}_M(w)$ equals $p'(n)$ for some polynomial p'. For M exponential time, $\text{Bd}_M(w)$ equals $2^{p'(n)}$, again for p' a polynomial. For a polynomial space computation, a configuration of $M(w)$ could be coded by a first order object $C < 2^{p'(x)}$. For exponential time machines however, a configuration is too large and must be coded as a second order object C, where $C(i)$ gives the i-th bit of the configuration. In either case, an entire computation of M can be coded by a string of $\text{Bd}_M(w) \cdot (t_{\text{end}} + 1)$ bits using a second order object X. The object X can code the computation by merely concatenating the codes $C[0], \ldots, C[t_{\text{end}}]$. As before, the exact details of the encoding are not important, however, S^1_2 must be able to define polynomial time functions that extract information about the states, tape head positions, and tape contents at any given time. Furthermore S^1_2 must be able to express other combinatorial properties about the computation; in particular, the condition that X codes a correct computation must be expressible by a Π^b_1-predicate in S^1_2.

If X codes a complete computation, $out(X)$ denotes the first order object output by the computation (if any). By encoding the computation of M in X appropriately, we can ensure that $\mathit{out}(X)$ is computable in polynomial time relative to X. We sometimes allow a polynomial space or exponential time Turing machine M to also output a second order object, and use $\mathit{Out}(X)$ to denote the second order object output (if any). The encoding X must allow the second order object $\mathit{Out}(X)$ to be polynomial time computable, in that the value of $\mathit{Out}(X)(i)$ is computable in polynomial time relative to X. For an exponential time machine, which has exponentially large configurations, this can be done by using a separate output tape for the second order output. For a polynomial space machine, this can be done by requiring M to write each value $\mathit{Out}(X)(i)$ at a special tape location at a prespecified time that is easily computed from i. This permits configurations of M to be coded by first order objects in spite of the fact that M outputs an exponentially large second order object. It also permits $\mathit{Out}(X)(i)$ to be computed in polynomial time relative to X.

3.1 Canonical evaluation and canonical verification

We now define the notion of how a second order object α can "canonically evaluate" or "canonically verify" a bounded formula. It is important for later developments that these notions make sense over the base theory S^1_2.

Let $\phi(\vec{x}, \vec{X})$ be a first order bounded formula with all free variables indicated. Without loss of generality, the formula ϕ is in prenex form and, for notational convenience, we also assume that the quantifiers are alternating existential and universal, and that they all use the same bounding term $t(\vec{x})$. (These assumptions can be made without loss of generality in any event since we are only concerned that formulas are bounded, but not concerned about what Σ^b_i or Π^b_i class they are in.) Thus, we can assume ϕ has the form

$$(\exists y_1 \le t)(\forall y_2 \le t)(\exists y_3 \le t) \cdots (Q_k y_k \le t)\psi(\vec{y}, \vec{x}, \vec{X}),$$

with ψ quantifier-free. Here we are temporarily adopting the notation that, for $i \le k$, Q_i is "$\exists$" if k is odd, and "$\forall$" otherwise. We next define what it means for α to *canonically evaluate* $\phi(\vec{x}, \vec{X})$. An input to α will be interpreted as a tuple of the form $\langle a_1, a_2, \ldots, a_\ell\rangle$ where $0 < \ell \le k+1$ and where $0 \le a_i \le t$ for each i. Any standard sequence encoding may be used for coding tuples.

The intuition is that if $\ell \le k$ then $\alpha(\langle a_1, \ldots, a_\ell, t\rangle)$ is true if and only if

$$(Q_{\ell+1} y_{\ell+1} \le t) \cdots (Q_k y_k \le t)\psi(a_1, \ldots, a_\ell, y_{\ell+1}, \ldots, y_k, \vec{x}, \vec{X})$$

is true. Note the final value is t in the tuple. However, more generally, the intuition is that if $\ell < k$, then $\alpha(\langle a_1, \ldots, a_\ell, a_{\ell+1}\rangle)$ is true if and only if

$$(Q_{\ell+1} y_{\ell+1} \le a_{\ell+1}) \cdots (Q_k y_k \le t)\psi(a_1, \ldots, a_\ell, y_{\ell+1}, \ldots, y_k, \vec{x}, \vec{X})$$

is true. We make these intuitions formal by setting the following conditions on α.

(a): For all $a_1, \ldots, a_k, b \le t$, we have

$$\alpha(\langle \vec{a}, b\rangle) \ \leftrightarrow\ \psi(\vec{a}, \vec{x}, \vec{X}).$$

Note that the value b is just a placeholder and is not actually used.

(b): For all odd $\ell \le k$ and all $a_1, \ldots, a_\ell \le t$,

$$\alpha(\langle \vec{a}\rangle) \ \leftrightarrow\ [(a_\ell > 0 \wedge \alpha(\langle a_1, \ldots, a_{\ell-1}, a_\ell - 1\rangle)) \vee \alpha(\langle \vec{a}, t\rangle)].$$

(c): For all even $\ell \le k$ and all $a_1, \ldots, a_\ell \le t$,

$$\alpha(\langle \vec{a}\rangle) \ \leftrightarrow\ [(a_\ell > 0 \supset \alpha(\langle a_1, \ldots, a_{\ell-1}, a_\ell - 1\rangle)) \wedge \alpha(\langle \vec{a}, t\rangle)].$$

Definition 3.1 The second order object α *canonically evaluates* $\phi(\vec{x},\vec{X})$ provided that all the conditions (a)–(c) above hold. And, α *canonically verifies* $\phi(\vec{x},\vec{X})$ provided that α canonically evaluates $\phi(\vec{x},\vec{X})$ and $\alpha(\langle t\rangle)$ is true.

Note that "α canonically evaluates $\phi(\vec{x},\vec{X})$" and "α canonically verifies $\phi(\vec{x},\vec{X})$" are expressible as Π^b_1 formulas.

We extend the definitions of canonical evaluation and verification to $\Sigma^{1,b}_1$-formulas as follows.

Definition 3.2 Let ϕ be a strict $\Sigma^{1,b}_1$-formula of the form $(\exists Y)C(\vec{x},\vec{X},Y)$, and let β be a second order object. Then α *canonically verifies that β witnesses ϕ* if and only α canonically verifies $C(\vec{x},\vec{X},\beta)$.

Theorem 3.3 *For $\phi(\vec{x},\vec{X})$ a $\Sigma^{1,b}_0$-formula, S^1_2 proves*

"If α canonically verifies $\phi(\vec{x},\vec{X})$, then $\phi(\vec{x},\vec{X})$ is true".

For $\phi(\vec{x},\vec{X})$ a strict $\Sigma^{1,b}_1$-formula $(\exists Z)\psi(\vec{x},\vec{X},Z)$, S^1_2 proves

"If α canonically verifies that β witnesses $\phi(\vec{x},\vec{X})$, then $\psi(\vec{x},\vec{X},\beta)$ is true".

Proof. (Sketch) This is proved using induction (outside S^1_2) on the number of quantifiers k. For $k=0$, it is immediate by condition (a). For $k>0$, suppose $\alpha(\langle t\rangle)$ holds. Arguing in S^1_2, use binary search or Δ^b_1-minimization to find the least $a_1 \le t$ such that $\alpha(\langle a_1\rangle)$ holds. By (b), this implies $\alpha(\langle a_1,t\rangle)$ holds. By the (dual of the) induction hypothesis, applied to the negations of α and the negation of $(\forall y_2 \le t)\cdots(Q_k y_k \le t)\psi(a_1,\vec{y},\vec{x},\vec{X})$, we have $\phi(\vec{x},\vec{X})$ is true with y_1 set equal to a_1. □

We shall also need second order objects to canonically evaluate or canonically verify formulas ϕ that are not in prenex form. (In particular, such formulas seem to be unavoidable in the comprehension axioms.) For this, suppose ϕ is a non-prenex formula; for the next theorem, we use ϕ^* to denote any prenex form of ϕ; that is, ϕ^* is obtained from ϕ by pulling out quantifiers using prenex operations. We claim that S^1_2 is able to prove that the canonical verifications give the same results no matter what prenex form is used. The following theorem partially formalizes this claim.

Theorem 3.4 *Let ϕ and ψ be in prenex form with canonical evaluations given by α and β. Suppose γ is a canonical evaluation of $(\phi\wedge\psi)^*$, or of $(\phi\vee\psi)^*$, or of $(\neg\phi)^*$. Then γ canonically verifies the truth of the formula if and only α and β canonically verify ϕ and ψ, or one of α or β canonically verifies ϕ or ψ, or α does not canonically verify ϕ (respectively).*

Furthermore, for any fixed choice of formulas, this statement is provable in S^1_2.

The proof of the theorem is straightforward, as the canonical verification γ is expressible in terms of α and β in a very explicit way, based on the order in which prenex operations were applied. We omit the details.

3.2 The new-style witnessing theorems

Theorem 3.5 (Witnessing Theorem for U^1_2)

(a) *Suppose U^1_2 proves $(\exists y)\phi(y,\vec{a},\vec{A})$ for ϕ a $\Sigma^{1,b}_0$-formula. Then there is a* PSPACE *oracle Turing machine M such that S^1_2 proves "If Y encodes a complete computation of $M^{\vec{A}}(\vec{a})$, then $\phi(out(Y),\vec{a},\vec{A})$ is true".*

(b) *Suppose U^1_2 proves $(\exists Z)\phi(Z,\vec{a},\vec{A})$ for ϕ a $\Sigma^{1,b}_0$-formula. Then there is a* PSPACE *oracle Turing machine M such that S^1_2 proves "If W encodes a complete computation of $M^{\vec{A}}(\vec{a})$, then Out$(W) = \langle Y, Y'\rangle$ where Y canonically verifies that Y' witnesses $\exists Z\phi(x,Z,X)$ is true".*

The notation $M^{\vec{A}}(\vec{a})$ denotes that the machine M has as inputs the first order objects $\vec{a}$, and has oracle access to the second order objects $\vec{A}$. The notation $W = \langle Y, Y'\rangle$ is the ordinary pairing on second order objects, namely it means that $W(i)$ is true precisely for those i's of the form $\langle 0, y\rangle$ with $y \in Y$ or of the form $\langle 1, y'\rangle$ such that $y' \in Y'$.

The proof of Theorem 3.5 is based on the following witnessing lemma.

Theorem 3.6 (Witnessing Lemma for U^1_2) *Suppose U^{1*}_2 proves a sequent $\Gamma \longrightarrow \Delta$ of strict $\Sigma^{1,b}_1$-formulas with free variables $\vec{a},\vec{A}$. Let Γ be $\phi_1,\ldots,\phi_k$ and Δ be $\psi_1,\ldots,\psi_\ell$ with each ϕ_i equal to $(\exists Y_i)\phi'_i(\vec{a},\vec{A},Y_i)$ and each ψ_i equal to $(\exists Z_i)\psi'_i(\vec{a},\vec{A},Z_i)$. (Some of the quantifiers may be omitted.) Then there is a* PSPACE *oracle machine M such that S^1_2 proves:*

> *"If U_i canonically verifies that Y_i is a witness for ϕ_i for $i = 1,\ldots,k$, and if W encodes a complete computation of $M^{\vec{U},\vec{Y},\vec{A}}(\vec{a})$, then this computation of M outputs a first order $j =$ out$(W) \in \{1,\ldots,\ell\}$ and encodes a second order output Out$(W) = \langle V, Z_j\rangle$ such that V canonically verifies that Z_j is a witness for Ψ_j".*

Theorem 3.5 is an immediate consequence of Theorems 1.6, 3.3, and 3.6. The proof of Theorem 3.6, given in Section 3.3 below, uses induction on the number of lines in a U^{1*}_2 sequent calculus proof which contains only strict $\Sigma^{1,b}_1$-formulas.

The witnessing theorem and lemma for V^1_2 are completely analogous to those for U^1_2.

Theorem 3.7 (Witnessing Theorem for V^1_2)

(a) *Suppose V^1_2 proves $(\exists y)\phi(y,\vec{a},\vec{A})$ for ϕ a $\Sigma^{1,b}_0$-formula. Then there is an exponential time oracle Turing machine M such that S^1_2 proves "If Y encodes a complete computation of $M^{\vec{A}}(\vec{a})$, then $\phi($out$(Y),\vec{a},\vec{A})$ is true".*

(b) *Suppose V^1_2 proves $(\exists Z)\phi(Z,\vec{a},\vec{A})$ for ϕ a $\Sigma^{1,b}_0$-formula. Then there is an exponential time oracle Turing machine M such that S^1_2 proves "If W encodes a complete computation of $M^{\vec{A}}(\vec{a})$, then Out$(W) = \langle Y, Y'\rangle$ where Y canonically verifies that Y' witnesses $\exists Z\phi(x,Z,X)$ is true".*

Theorem 3.8 (Witnessing Lemma for V^1_2) *Suppose V^1_2 proves a sequent $\Gamma \longrightarrow \Delta$ of strict $\Sigma^{1,b}_1$-formulas with free variables $\vec{a},\vec{A}$. Let Γ be $\phi_1,\ldots,\phi_k$ and Δ be $\psi_1,\ldots,\psi_\ell$ with each ϕ_i equal to $(\exists Y_i)\phi'_i(\vec{a},\vec{A},Y_i)$ and each ψ_i equal to $(\exists Z_i)\psi'_i(\vec{a},\vec{A},Z_i)$. (Some of the quantifiers may be omitted.) Then there is an exponential time oracle machine M such that S^1_2 proves:*

> *"If U_i canonically verifies that Y_i is a witness for ϕ_i for $i = 1,\ldots,k$, and if W encodes a complete computation of $M^{\vec{U},\vec{Y},\vec{A}}(\vec{a})$, then this computation of M outputs a first order $j =$ out$(W) \in \{1,\ldots,\ell\}$ and encodes a second order output Out$(W) = \langle V, Z_j\rangle$ such that V canonically verifies that Z_j is a witness for Ψ_j".*

As before, Theorem 3.7 follows from Theorems 1.6, 3.3, and 3.8. Theorem 3.8 is also proved in Section 3.3 below.

3.3 Proofs of the witnessing lemmas for U_2^1 and V_2^1

We now prove Theorem 3.6. Assume P is a U_2^{1*}-proof containing only strict $\Sigma_1^{1,b}$-formulas. The proof of Theorem 3.6 uses induction on the number of steps in the proof P, and splits into cases depending on the final inference of P. There are two base cases where P consists a single sequent, with no inferences. The first is where P is a single initial sequent of the form $A \longrightarrow A$ where, without loss of generality, A is atomic. This case is completely trivial of course. The second base case is when P consists of a single $\Sigma_0^{1,b}$-comprehension axiom of the form

$$\longrightarrow (\exists Z)(\forall y \le t(\vec{x}))[y \in Z \leftrightarrow \phi(y, \vec{x}, \vec{X})] \tag{3.1}$$

for ϕ bounded. We must describe a polynomial space Turing machine M that computes Z and a second order object V that canonically verifies that Z witnesses the truth of the comprehension axiom. We shall use informal arguments to describe M, but it will be clear that S_2^1 can formalize them in the sense that S_2^1 can prove that if a second order object encoding a complete computation of M is given, then the outputs $out(M)$ and $Out(M)$ correctly provide a canonical verification of the sequent (3.1). There is only a single formula, so $\ell = 1$ and of course $out(M) = 1$. The deterministic polynomial space algorithm for M is straightforward: for each value $y < t(\vec{x})$, $M^{\vec{X}}(\vec{x})$ computes the predicate V_y such that V_y canonically evaluates the truth of $\phi(y, \vec{x}, \vec{X})$. If V_y indicates $\phi(y, \vec{x}, \vec{X})$ is true, then $Z(y)$ is determined to be true; otherwise, $Z(y)$ is determined to be false. For each fixed value y, the V_y can be straightforwardly converted into a canonical verification (of a prenex form) of $y \in Z \leftrightarrow \phi(y, \vec{x}, \vec{X})$. Combining all these gives a canonical verification of (3.1).

The cases where the final inference of P is a weakening inference or an exchange inference are trivial. The cases where the final inference is a propositional inference are also rather trivial, but we do the case of $\wedge$:right to illustrate this. Suppose the final inference of P is

$$\frac{\phi_1, \ldots, \phi_k \!\rightarrow\! \psi_1, \ldots, \psi_{\ell-1}, \psi'_\ell \qquad \phi_1, \ldots, \phi_k \!\rightarrow\! \psi_1, \ldots, \psi_{\ell-1}, \psi''_\ell}{\phi_1, \ldots, \phi_k \!\rightarrow\! \psi_1, \ldots, \psi_{\ell-1}, \psi'_\ell \wedge \psi''_\ell}$$

Note there are no second order quantifiers in ψ'_ℓ or ψ''_ℓ since P is free-cut free and thus all formulas in the proof are strict $\Sigma_1^{1,b}$. The induction hypothesis gives two Turing machines M' and M'' which satisfy Theorem 3.6 for the two upper sequents. We describe how to form a Turing machine M that fulfills the same condition for the lower sequent. The machine M has first order inputs $\vec{a}$ and uses oracles $\vec{U}, \vec{Y}, \vec{A}$. The machine M starts by forming a canonical evaluation of a prenex form of $\psi'_\ell \wedge \psi''_\ell$: this uses only inputs $\vec{y}$ and $\vec{A}$, and involves looping through all possible values for the bounded quantifiers in this formula, and uses polynomial space. If the canonical evaluation shows that $\psi'_\ell \wedge \psi''_\ell$ is true, M halts outputting the first order value ℓ indicating the ℓth formula of the antecedent is true, and also outputting a second order object Z that canonically verifies $\psi'_\ell \wedge \psi''_\ell$. Otherwise, M canonically evaluates both ψ'_ℓ and ψ''_ℓ. By Theorem 3.4, at least one of these two formulas will be found to be false. Suppose, without loss of generality, that ψ'_ℓ is false. In this case, M simulates M' and outputs whatever it outputs. Note that M' cannot report that ψ'_ℓ is true, so it must instead output some $j < \ell$, some Z_j, and some V which canonically verifies that Z_j is a witness for ψ_j. It is clear that S_2^1 can simulate this argument sufficiently well so as to prove that, if a complete computation of M is given as a second order object W, then it gives a canonical verification either for $\psi'_\ell \wedge \psi''_\ell$ or for some ψ_j with $j < \ell$.

Now suppose the final inference of P is a bounded first order $\exists$:right inference:

$$\frac{\Gamma \longrightarrow \Delta, \psi(s)}{s \le t, \Gamma \longrightarrow \Delta, (\exists x \le t)\psi(x)}$$

Note that ψ again has no second order quantifiers. The proof idea is somewhat similar to the case of $\wedge$:right just done. The induction hypothesis gives a Turing machine M' satisfying the witnessing conditions for the upper sequent. The desired Turing machine M acts as follows. It first builds a canonical evaluation for $(\exists x \le t)\psi(x)$. If this finds the formula to be true, it outputs this fact along with the canonical verification. (As an alternate construction, it would also be enough to do this only if $\psi(s)$ is true. It must be the case that $s \le t$ since the input to M includes a canonical evaluation of this atomic formula.) Otherwise, M continues to simulate M'. The output of M' must produce an index j for a formula in Δ along with a Z_j and a V which together witness and canonically verify the truth of the jth formula of Δ. Again, S^1_2 can prove that a complete computation by M produces the desired output.

Suppose the final inference of P is a bounded first order $\exists$:left inference:

$$\frac{a_0 \le t, \phi_0(a_0), \Gamma \longrightarrow \Delta}{(\exists x \le t)\phi_0(x), \Gamma \longrightarrow \Delta}$$

Here a_0 is an eigenvariable and does not occur in the lower sequent. Of course, ϕ_0 does not have any second order quantifiers. Let M' be given by the induction hypothesis. The desired machine M has among its inputs a canonical verification U_0 of the formula $(\exists x \le t)\phi_0(x)$. M starts by extracting the least value for x for which U_0 has found that $\phi(x)$ is true, and sets a_0 equal to this value. (M can readily find a_0 either a polynomial space linear search through all values of x, or by a polynomial time binary search as in the proof of Theorem 3.3.) Once a value for a_0 is determined, M continues by simulating M' and using its outputs.

The case where the final inference of P is a bounded first order $\forall$:right inference

$$\frac{a_0 \le t, \Gamma \longrightarrow \Delta, \psi(a_0)}{\Gamma \longrightarrow \Delta, (\forall x \le t)\psi(x)}$$

is similar to the previous two cases. Namely, the Turing machine M for the lower sequent starts by forming a canonical evaluation of $(\forall x \le t)\psi(x)$. If this is found to be true, this is output by M. Otherwise, M finds a value for $a_0 \le t$ that makes $\psi(a_0)$ false, and M continues by simulating the machine M' for the upper sequent with this value for a_0.

Suppose the final inference of P is a bounded first order $\forall$:left inference

$$\frac{\phi_0(s), \Gamma \longrightarrow \Delta}{s \le t, (\forall x \le t)\phi_0(x), \Gamma \longrightarrow \Delta}$$

The Turing machine M for the lower sequent is given among its inputs a second order object U_0 (an oracle) that canonically evaluates $(\forall x \le t)\phi_0(x)$. It is easy to extract from U_0 another second order object U'_0 that canonically evaluates $\phi_0(s)$. This is because $s \le t$ must be true, and since we can define $U'_0(\langle a_1, \ldots, a_j\rangle)$ to equal $U_0(\langle s, a_1, \ldots, a_j\rangle)$. Let M' be the polynomial space Turing machine given by the induction hypothesis. The machine M acts by simulating M' using U'_0 as the canonical verification for $\phi_0(s)$.

Suppose the final inference of P is a second order $\exists$:right inference

$$\frac{\Gamma\longrightarrow\Delta, \psi(A)}{\Gamma\longrightarrow\Delta, (\exists Z)\psi(Z)}$$

The second order variable A is not an eigenvariable, and so, without loss of generality, appears in the lower sequent. Thus the desired machine M' for the lower sequent takes the same inputs as the polynomial space machine M given by the induction hypothesis for the upper sequent. The machine M will output a canonical verification either of a formula in Δ or of $\psi(A)$. In the former case, M' gives the same output as M. In the latter case, M' sets Z equal to A and outputs the canonical verification of $\psi(Z)$.

Suppose the final inference of P is a second order $\exists$:left

$$\frac{\phi(A), \Gamma\longrightarrow\Delta}{(\exists Y)\psi(Y), \Gamma\longrightarrow\Delta}$$

where now A is an eigenvariable and does not appear in the lower sequent. Let M be the polynomial space machine given by the induction hypothesis; we must define the machine M' for the lower sequent. One of the inputs to M' is a second order Y along with a canonical verification U of $\phi(Y)$. The machine M' runs by letting this Y be the value of the input A to M, using U as the canonical verification of $\phi(A)$, and then just running M.

Now suppose the final inference of P is an s$\Sigma_1^{1,b}$-repl-$\forall$ inference,

$$\frac{a \le t, \Gamma\longrightarrow\Delta, (\exists X)\psi(X, a)}{\Gamma\longrightarrow\Delta, (\exists Z)(\forall x \le t)\psi(\{z\}Z(\langle x, z\rangle), x)}$$

where a is an eigenvariable and may not occur in the lower sequent. The induction hypothesis gives a polynomial space Turing machine M' for the upper sequent. We form a new machine M which has the same inputs as M' except that a is not an input to M. The machine M runs as follows: it loops through all values of $a \le t$, and simulates M' with each of these values for a. If, for any value a, M' indicates that a formula ψ_j in Δ is true and gives a witness Z_j and a canonical verification V for ψ_j, then M halts and outputs the same values j, Z_j and V. Otherwise, for each value of a, M' produces a second order X_a and a canonical verification V_a showing that X_a witnesses $(\exists X)\psi(X, a)$. When this happens for all values of a, the second order X_a's can be combined into a single second order Z defined so that $Z(a, z)$ holds iff $X_a(z)$ holds; furthermore, the canonical verifications V_a can be straightforwardly combined to give a canonical verification that Z is a witness for $(\exists Z)(\forall x \le t)\psi(\{z\}Z(\langle x, z\rangle), x)$. It is clear that M is polynomial space bounded, since M' is.

Suppose the final inference of P is a cut inference,

$$\frac{\Gamma\longrightarrow\Delta, \chi \qquad \chi, \Gamma\longrightarrow\Delta}{\Gamma\longrightarrow\Delta}$$

Let M_1 and M_2 be the Turing machines given by the induction hypothesis for the left and right upper sequents, respectively. The machine M for the lower sequent is constructed as follows. It begins by running machine M_1, which takes the identical inputs as M. If M_1 finishes with a witness for one of the formulas in Δ, then M halts producing the same first and second order outputs as M_1. Otherwise, M_1 outputs a pair of second order objects V and $Z_{\ell+1}$ such that V canonically verifies that $Z_{\ell+1}$ is a witness for χ. In this case, M then invokes M_2 with the intent of using V and $Z_{\ell+1}$ as inputs to M_2 that provide a witness and a canonical verification for the occurrence of χ in the antecedent of the upper right sequent. The only catch is that M is allowed to use only polynomial space, and this is not sufficient space for M to save the exponentially long values of V and $Z_{\ell+1}$. Instead, as M simulates M_2, it recomputes the values of V and $Z_{\ell+1}$ as needed by running machine M_1 again. Since M_1 is deterministic, this always yields consistent values for V and $Z_{\ell+1}$. This allows M to use only polynomial space as, at any given point in time, M needs to remember only one configuration of M_1 and one configuration of M_2.

Finally, suppose the last inference of P is an $s\Sigma_1^{1,b}$-LIND induction,

$$\frac{\chi(a_0), \Gamma \longrightarrow \Delta, \chi(a_0+1)}{\chi(0), \Gamma \longrightarrow \Delta, \chi(|t|)}$$

(It is slightly more convenient to use LIND instead of PIND, but the argument is essentially the same either way.) Let M' be the Turing machine given by the induction hypothesis for the upper sequent. The intuition is that we handle the induction hypothesis by treating it as $|t|-1$ many cuts, on the formulas $\chi(1), \chi(2), \ldots, \chi(|t|-1)$. This means that M is iterating computations of M'; however, the iterations are nested only to a depth $|t|$, so M needs to remember at most $|t|$ many configurations of M' at any given point in time. Since $|t|$ is polynomially bounded in terms of the lengths of the first order free variables, this means M uses only polynomial space.

For a bit more detail, let Δ have $\ell-1$ formulas. M starts by computing M' with a_0 set equal to 0, which we denote $M'[a_0 := 0]$, and potentially continues for $a_0 = 1, 2, \ldots, |t|-1$. If $M'[a_0 := i]$ yields a first order output $< \ell$, a witness of a formula in Δ has been obtained, and M can output this. Otherwise, $M'[a_0 := i]$ outputs first order output ℓ along with a witness for $\chi(|t|)$. If this happens with $i = |t|$, then the desired output has been obtained. For $i < |t|-1$, M must instead invoke $M'[a_0 := i+1]$ using the output of $M'[a_0 := i]$ as the second order witness and canonical verification for $\chi(i)$. As in the case of cut, the output of $M'[a_0 := i]$ is exponentially large, and cannot be written out in polynomial space. Instead, whenever, $M'[a_0 := i+1]$ queries its second order inputs for $\chi(i)$, M interrupts the computation of $M'[a_0 := i+1]$ and re-simulates the entire computation of $M'[a_0 := i]$. These recomputations must be carried out recursively, but only to a depth of $|t|$. At any given point in time, M needs to remember at most configurations for one invocation of each of $M'[a_0 := i]$, for $i = 0, \ldots, |t|$.

The above completes the proof of Theorem 3.6. The proof of Theorem 3.8 is mostly identical. The various cases, based on the final inference of V_2^1-proof P, are essentially identical to the cases described above for Theorem 3.6. The cases of cut and induction merit more discussion however. In the setting of V_2^1, the exponential time machine M is allowed to use exponential space and this allows a simplification to be made in the construction of M. For the case where the final inference of P is cut, the output of the machine M_1 can be written down completely in M's memory as this requires 'only' exponential time and space. It is thus unnecessary to redo the computation of M_1 every time M_2 needs a value of M_1's second order output. Similar considerations apply to the case where the final inference of the V_2^1 is an IND induction inference. M now needs to do an exponentially long iteration; however, instead of recomputing values, M can just store them all in memory.

4 Local improvement principles

4.1 Definitions and theorems

The local improvement principles were defined by Kołodziejczyk, Nguyen, and Thapen [**8**] as an extension of the game principles of Skelley and Thapen [**13**]. The local improvement principle is specified by a set of contradictory conditions, so the local improvement principle states that it is always possible to find a counterexample to one of the conditions. Our definition of the local improvement principles below includes a minor, inessential change to the definition of [**8**] so as to make the score values a function of a single label instead of a function of the labels in a neighborhood.

Definition 4.1 An instance of the local improvement principle consists of a specification of a directed acyclic graph G with domain $[a] := \{0, 1, 2, \ldots, a-1\}$ and polynomial time computable edges, an upper bound $b > 0$ on labels, an upper bound $c > 0$ on scores, an initial labeling function E, a wellformedness predicate *wf*, and a local improvement function I. These satisfy the following conditions.

(a) The directed graph G is consistent with the usual $<$-ordering of its domain $[a]$, and has in- and out-degrees bounded by a fixed constant. The edges of G are specified by a polynomial time neighborhood function f. For each vertex $x \in [a]$, $f(x)$ outputs a set of vertices $y \in [a]$: the vertices $y < x$ (respectively, $y > x$) are the predecessors (respectively, the successors) of the vertex x. The *neighborhood* of x is the set containing x together with its successors and predecessors. The *extended neighborhood* of x is the union of the neighborhoods of the neighbors of x.

(b) Vertices in G will be assigned a series of labels. A label is in the range $[0, b)$ and includes a *score* value s in the range $[0, c)$. The score value associated with the label on vertex x is polynomial time computable as a function of the label on x.[5] The polynomial time predicate *wf* determines whether a labeling of a neighborhood of x is *wellformed*. The inputs to the predicate *wf* are the vertices in the neighborhood and their labels. A labeling of vertices is *extended-wellformed* around x if it is wellformed on the neighborhood of every vertex y in the neighborhood of x.

(c) The two functions E and I provide methods of assigning labels to vertices. To initialize the labels, the polynomial time function $E(x)$ assigns labels to vertices x with score 0 so that all neighborhoods have wellformed labelings. The improvement function I provides a method to replace a label with a label with a higher score value: I takes as input a vertex x and a wellformed labeling of the neighborhood of x, and provides a new label for x. Specifically, suppose s is even and that every predecessor of x has a label with score $s+1$ and that x and every successor of x has a label with score s; then I provides a new label for x with score $s+1$. Dually, suppose s is odd and that every successor of x has a label with score $s+1$ and that x and every predecessor of x has a label with score s; then I provides a new label for x with score $s+1$. In other cases, the

[5] This is slightly different from the convention of [**8**] which makes the score value a function of the labels in the neighborhood of x. They let the score value equal "$*$" if the labels do not constitute a wellformed local labeling. The difference in how scores are defined makes no difference to the complexity of the local improvement principle.

I function is undefined. Furthermore, whenever I is defined and the labeling is extended-wellformed around x, then the labeling obtained by replacing the label on x with the the new label given by I is still extended-wellformed around x.

The intuition behind the local improvement function is that it provides labels with higher score values. Initially, all labels have score 0, but then sweeping forward through G allows scores to increase from even to odd values, and sweeping backwards allows scores to increase from odd to even values. The preservation of the extended-wellformed properties implies that scores can increase without bound. This, however, contradicts the property that score values are $< c$. Thus, the local improvement conditions listed above are contradictory.

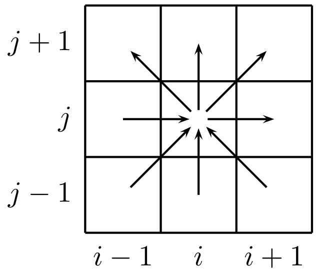
Definition 4.2 A *solution* to an instance of the local improvement property consists of either: (a) An extended-wellformed labeling of a vertex x and its extended neighborhood where the local improvement function is defined but fails to provide a new label for x with the correct score value that preserves the extended-wellformed property, or (b) a neighborhood of a vertex x where the initialization function E fails to provide an extended-wellformed labeling with scores all equal to zero.

Note that any solution to the local improvement property is polynomial time checkable.

Definition 4.3 An instance of the local improvement principle is given by a G specified with a polynomial time domain and a polynomial time neighborhood function f, first order values b and c, and polynomial time functions s, E, I and *wf*; and consists of the Σ^b_1 formula (with free variables a, b and c) that asserts that a solution exists. The notation LI denotes the set of Σ^b_1-formulas obtained from all instances of the local improvement principle. We use $\mathrm{LI}_{\log}$ to denote instances LI where c is a length, that is where $c = |c'|$ for some term c'.

The *linear local improvement* principles LLI and $\mathrm{LLI}_{\log}$ are defined in the same way, but with G restricted to be a linear graph. That is, G has vertices $[a]$, and the edges of G are the directed edges $(i-1, i)$, for $0 < i < a$.

It is also useful to define "rectangular" local improvement principles. These are instances of LI or $\mathrm{LI}_{\log}$ where the underlying graph G has domain $[a] \times [a]$, each vertex (i, j) has up to four incoming edges, namely from the vertices $(i-1, j)$, $(i-1, j-1)$, $(i, j-1)$, and $(i+1, j-1)$. Thus, the edges involving (i, j) are as pictured:



except that any edges that would involve vertices outside the domain of G are omitted. We shall call instances of LI and $\mathrm{LI}_{\log}$ based on these rectangular graphs RLI and $\mathrm{RLI}_{\log}$. (These rectangular graphs were used by **[8]**, although they did not use this terminology.)

Definition 4.4 An NP search problem Q is specified by a first order sentence

$$(\forall x)(\exists y \leq t)\phi(y,x)$$

with ϕ a Δ_1^b-formula with respect to S_2^1. A *solution* to $Q(x)$ is a value $y \leq t$ such that $\phi(y,x)$ holds. We denote this condition by $y = Q(x)$; note there may be multiple solutions y for a single input x.

The NP search problem Q is *total* provided that every x has at least one solution. It is *provably total* in a theory T provided $T \vdash (\forall x)(\exists y \leq t)\phi(y,x)$.

Any instance of the local improvement principle has a solution. This fact can be expressed as a $\forall\Sigma_1^b$-formula, and any solution can be verified in polynomial time. Thus the local improvement principles are total NP search problems.

Definition 4.5 Suppose that $(\forall x)(\exists y \leq t)\phi(y,x)$ and $(\forall x)(\exists y \leq s)\psi(y,x)$ specify NP search problems, denoted Q_ϕ and Q_ψ. A *many-one reduction* from Q_ϕ to Q_ψ consists of a pair of polynomial time functions g and h such that whenever $y = Q_\psi(g(x))$, we have $h(y,x) = Q_\phi(x)$. We write $Q_\phi \leq_m Q_\psi$ to denote that there is a many-one reduction from Q_ϕ to Q_ψ.

A theory proves that $Q_\phi \leq_m Q_\psi$ provided that it proves

$$(\forall x)(\forall y)[y = Q_\psi(g(x)) \supset h(y,x) = Q_\phi(x)].$$

We can now state the results of **[8]** about the local improvement principles and the provably total NP search problems of U_2^1 and V_2^1.

Theorem 4.6 ([**8**]) *U_2^1 proves the linear, logarithmic local improvement principle* $\mathrm{LLI}_{\log}$. *Furthermore,* $\mathrm{LLI}_{\log}$ *is many-one complete, provably in S_2^1, for the provably total* NP *search problems of U_2^1; namely, if Q is a provably total* NP *search problem of U_2^1, then S_2^1 can prove that Q is many-one reducible to an* NP *search problem in* $\mathrm{LLI}_{\log}$.

Theorem 4.7 ([**8**]) *V_2^1 proves the local improvement principle* LI. *Furthermore,* LI *is many-one complete, provably in S_2^1, for the provably total* NP *search problems of V_2^1; namely, if Q is a provably total* NP *search problem of V_2^1, then S_2^1 can prove that Q is many-one reducible to an* NP *search problem in* LI.

The same results hold for RLI *in place of* LI.

We shall improve these results below by proving the following two theorems. The first theorem states that U_2^1 can also prove the LLI formulas. This is a somewhat surprising and unexpected result, since the straightforward algorithmic way to prove the local improvement principle LLI would be to iteratively define labels with increasing score values by sweeping back and forth across the linear graph G. If this is done deterministically, this could simulate c steps of a Turing machine computation, that is to say, it could simulate exponential time algorithms. This is (conjecturally) beyond the power of U_2^1 which can only define polynomial space predicates. However, as we shall see in Section 4.2, the LLI principle can instead be proved using only (nondeterministic) polynomial space computations.

Theorem 4.8 *U_2^1 proves the linear local improvement principle* LLI. *Furthermore,* LLI *is many-one complete, provably in S_2^1, for the provably total* NP *search problems of U_2^1; namely, if Q is a provably total* NP *search problem of U_2^1, then S_2^1 can prove that Q is many-one reducible to an* NP *search problem in* LLI.

The second part of Theorem 4.8 follows already from Theorem 4.6 since LLI contains $\mathrm{LLI}_{\log}$ as a special case. The proof of first part of Theorem 4.8 is given in Section 4.2 below.

Our new result for V^1_2 states that $\mathrm{LI}_{\log}$ is already strong enough to be many-one complete for set of provably total NP search problems of V^1_2, and that the many-one completeness is provable over the base theory S^1_2.

Theorem 4.9 *V^1_2 proves the local improvement principle* $\mathrm{LI}_{\log}$*. Furthermore,* $\mathrm{LI}_{\log}$ *is many-one complete, provably in S^1_2, for the provably total* NP *search problems of V^1_2; namely, if Q is a provably total* NP *search problem of V^1_2, then S^1_2 can prove that Q is many-one reducible to an* NP *search problem in* $\mathrm{LI}_{\log}$*.*

The same results hold for $\mathrm{RLI}_{\log}$ *in place of* $\mathrm{LI}_{\log}$*.*

Theorem 4.9 will be proved in Section 4.4, using the rectangular local improvement principle $\mathrm{RLI}_{\log}$ for the many-one completeness. Of course, the first part of Theorem 4.9 follows already from Theorem 4.7.

It is interesting to observe that scores can be restricted further to a constant, for the price that the underlying graph structure will be more general than the linear structure in case of U^1_2, or than the rectangular structure in case of V^1_2. The best bound which we can obtain on scores is 2: score "0" for initialization, and score "1" for one round of improvement.

Theorem 4.10

(a) LI_2 *is many-one complete, provably in S^1_2, for the provable total* NP *search problems of V^1_2; in particular, if Q is a provably total* NP *search problem of V^1_2, then S^1_2 can prove that Q is many-one reducible to an* NP *search problem in* LI_2*.*

(b) RLI_2 *is many-one complete, provably in S^1_2, for the provable total* NP *search problems of U^1_2; namely, U^1_2 proves* RLI_2*, and if Q is a provably total* NP *search problem of U^1_2, then S^1_2 can prove that Q is many-one reducible to an* NP *search problem in* RLI_2*.*

Proof. For part (a), using Theorem 4.7, it suffices to describe how to turn an RLI problem into an equivalent LI_2 problem. Let an RLI problem be given by $a, b, c, s(\cdot), wf, E(\cdot), I(\cdot)$, that is, the underlying graph G has domain $[a] \times [a]$, and each vertex (i,j) has up to four incoming edges, namely from the vertices $(i-1,j)$, $(i-1,j-1)$, $(i,j-1)$, and $(i+1,j-1)$. We think of G aligned in a way that the origin $(0,0)$ is at the lower left corner. A simulating LI_2-problem can be constructed as follows: Let G^{-1} be G rotated by 180 degrees, so that the lower left corner $(0,0)$ of G becomes the upper right corner of G^{-1}, and that edges are pointing in the opposite direction. We create $c+1$ many copies $G_0, G_1, \ldots, G_c$, alternating between G and G^{-1}, starting with G, and place them in ascending order on the diagonal of a $[a\cdot(c+1)] \times [a\cdot(c+1)]$ grid as shown in Figure 2.

The idea of the simulation is that instead of computing initial well-founded labels of score 0, and then sweeping back and forth to compute new well-founded labels of higher scores using I, we will just sweep once over the grid and produce a well-founded labeling which at G_k scores k: For G_0 we use the initial labeling given by E. When filling in G_{k+1}, we use the already computed labels at G_k to compute a labeling, using I and additional edges between G_k and G_{k+1}. The structure of the additional edges is given in Figures 3 and 4.

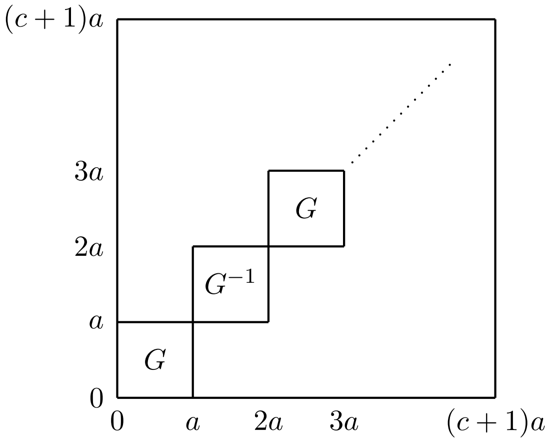


FIGURE 2. Structure of LI_2 game simulating an RLI game.

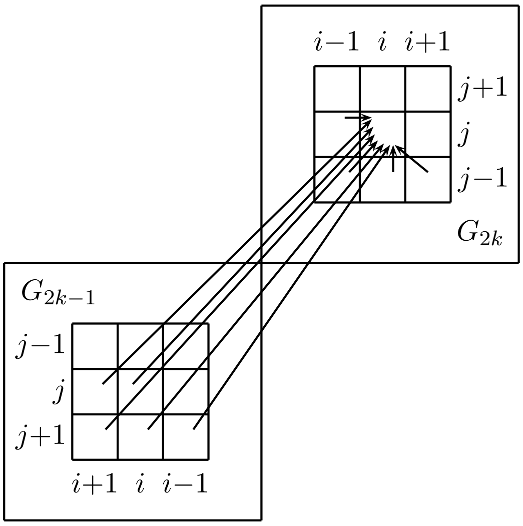


FIGURE 3. Structure of additional edges between G_{2k-1} and G_{2k} in LI_2 game which simulates an RLI game.

We call the additional edges between G_k and G_{k+1} *new* edges, and predecessors and successors based on them *new predecessors*, resp. *new successors.* Notions based on existing edges are dubbed *old*.

Specifically, in Figure 3, suppose that every old predecessor of (i, j) in G_{2k}, that is $(i-1, j-1)$, $(i, j-1)$, $(i+1, j-1)$, and $(i-1, j)$ in G_{2k}, has a label with score $2k$, and that every new predecessor of (i, j) in G_{2k}, that is (i, j), $(i+1, j)$, $(i-1, j+1)$, $(i, j+1)$, and $(i+1, j+1)$ in G_{2k-1}, has a label with score $2k-1$. Then I provides a new label for (i, j) in G_{2k} with score $2k$. That is, the neighborhood on which I bases its computation, is formed from $(i-1, j-1)$, $(i, j-1)$, $(i+1, j-1)$, $(i-1, j)$ in G_{2k}, and (i, j), $(i+1, j)$, $(i-1, j+1)$, $(i, j+1)$, $(i+1, j+1)$ in G_{2k-1}.

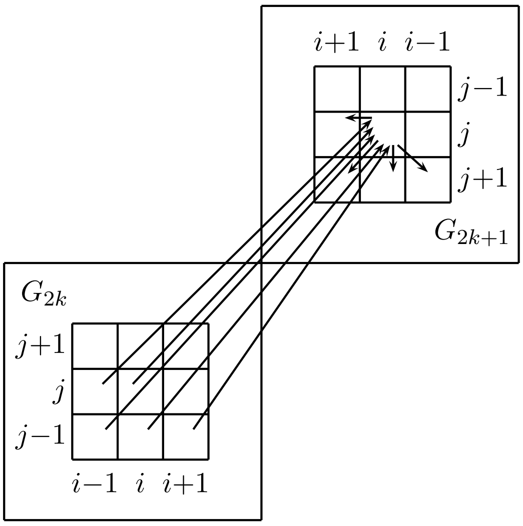


FIGURE 4. Structure of additional edges between G_{2k} and G_{2k+1} in LI_2 game which simulates an RLI game.

Dually, in Figure 4, suppose that every old successor of (i, j) in G_{2k+1}, that is $(i+1, j+1)$, $(i, j+1)$, $(i-1, j+1)$, and $(i+1, j)$ in G_{2k+1}, has a label with score $2k+1$, and that every new predecessor of (i, j) in G_{2k}, that is (i, j), $(i-1, j)$, $(i+1, j-1)$, $(i, j-1)$, and $(i-1, j-1)$ in G_{2k}, has a label with score $2k$. Then I provides a new label for (i, j) in G_{2k+1} with score $2k+1$. Here, the neighborhood on which I bases its computation, consists of $(i+1, j+1)$, $(i, j+1)$, $(i-1, j+1)$, $(i+1, j)$ in G_{2k+1}, and (i, j), $(i-1, j)$, $(i+1, j-1)$, $(i, j-1)$, $(i-1, j-1)$ in G_{2k}.

The initial labeling with score 0 will be given by $E(.)$ on G_0, and arbitrarily anywhere else, e.g. by choosing the labels to be 0. Without loss of generality, we can assume that the label 0 is used only for the initial label values.

The wellformedness predicate *wf* for the new LI_2 problem is defined with the aid of the predicate *wf* for the instance of RLI. Consider a vertex (i, j) in G_{2k} where $k > 0$; we call this vertex x. The vertex x has the nine incoming edges as shown in Figure 3, namely, one from (i, j) in G_{2k-1} plus eight additional edges. In addition, there are the corresponding nine outgoing edges. If any of the predecessors of x have label 0, then the neighborhood of x is defined to be wellformed provided that x, and all of its successors also have label 0. Otherwise, the predecessors of x all have labels different from 0. Then, if x itself has label 0, then the neighborhood of x is wellformed provided that the labels on the vertex (i, j) in G_{2k-1} and the eight other predecessors of x are wellformed according to the criteria of the RLI instance. On the other hand, if x does not have label 0, then the neighborhood of x is wellformed provided that the labels on x and its eight predecessors are wellformed according to the criteria of the RLI instance. The wellformedness predicate is defined similarly for vertices in G_{2k-1}.

It is not hard to verify that the above gives a faithful translation from the RLI problem to an LI_2 problem.

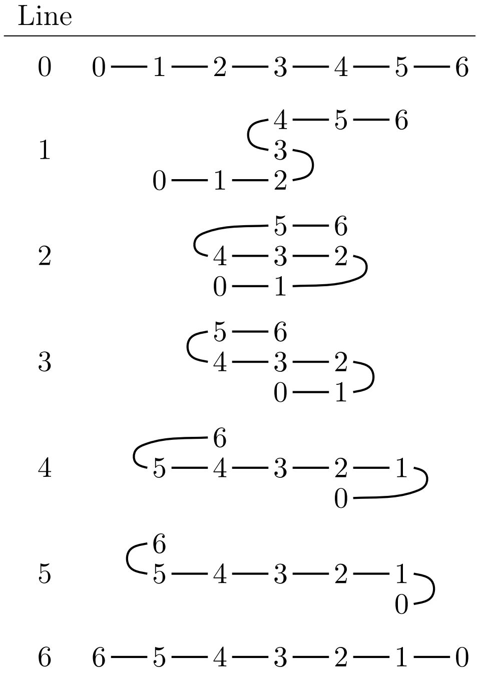


Figure 5. Inverting a line with 7 label positions using 6 additional lines.

We now turn to part (b). By Theorem 4.13 proved below, RLI_2 is provable in U_2^1. For the completeness of RLI_2 under many-one reductions, in light of Theorem 4.6, it suffices to describe how to turn an LLI problem into an equivalent RLI_2 problem. Let an LLI problem be given by $a, b, c, s(\cdot), wf, E(\cdot), I(\cdot)$, that is, the underlying graph G has vertices $[a]$, and the edges of G are the directed edges $(i-1, i)$ for $0 < i < a$. We think of G as a horizontal line with 0 to the left, and edges pointing to the right. A simulating RLI_2 problem can be constructed as follows: Let G^{-1} be the inverse of G, that is 0 is now to the right and edges are pointing to the left.

The main idea is to again create $c+1$ many copies alternating between G and G^{-1}, and stack them vertically to form a $[a] \times [c+1]$ grid. Instead of computing initial well-founded labels of score 0, and then sweeping back and forth on G to compute new well-founded labels of higher scores using I, we want to just sweep once over the grid and produce a well-founded labeling which at the k-th copy of G scores k. However, as we alternate between G and G^{-1} and the underlying graph structure shall be the rectangular one of RLI, we need, between any two alternations, $a+1$ many additional lines which allow us to invert the positions of previously computed labels. Thus, the resulting graph has dimension $[a] \times [(a+1)(c+1)]$.

The idea for inverting label positions using additional lines is as follows —see Figure 5 for an example. We can think of the original line of label positions as a rope stretching out horizontally in the plane. We transform the rope to first form a little “s” in the middle. Then, keeping the position of the middle point of the rope fixed, we stretch the two curves of the “s” horizontally to the sides, until eventually the rope is stretched out again, this time in the opposite direction.

During this process, a vertical line at an arbitrary horizontal position will intersect the rope at most three times. Thus, the labels at additional lines in our game graph, when imitating the just described transformation, will have to store at most 3 pieces of label information: one for those labels who have found their new position; one for those who are moving left to right, and one for those who are moving right to left. In addition, it is convenient to store a score value which measures how far the inversion has progressed.

It is obvious that a rectangular structure on the new $[a] \times [(a+1)(c+1)]$ grid is sufficient to imitate the above described process. It is then straightforward to define *wf*, $E(\cdot)$ and $I(\cdot)$ which exactly describe this process —details are left to the reader. □

Corollary 4.11 *Over the base theory* S^1_2*,*

(a) *the principles* LLI*,* $\mathrm{LLI}_{\log}$ *and* RLI_2 *are equivalent;*

(b) *the principles* LI*,* $\mathrm{LI}_{\log}$*,* LI_2*,* RLI*, and* $\mathrm{RLI}_{\log}$ *are equivalent.*

Proof. Part (a) is an immediate consequence of Theorems 4.6, 4.8, and 4.13, due to the fact that the LLI conditions are NP search problems. Part (b) is likewise an immediate consequence of Theorems 4.7, 4.9, and 4.10 and the fact that only $\mathrm{RLI}_{\log}$ will be used for the proof of Theorem 4.9. □

4.2 Proof of Theorem 4.8

The intuition behind the proof of LLI is based on the following exponential time algorithm: First set all vertices $x \in [a]$ in the linear directed graph to have the initial labels with score zero given by $E(x)$. Then, sweep back-and-forth through the vertices in linear order, alternating scans in left-to-right order (from 0 to $a-1$) with scans in right-to-left order (from $a-1$ to 0). Each time a vertex x is processed, its prior label, with score s, is replaced by a new label, with score $s+1$. For even values of s, this occurs while sweeping left-to-right and for odd values of s, it occurs while sweeping from right-to-left. Up to c scans are performed, by which time a contradiction to the local improvement conditions must have been found; namely, either by reaching a point where the improvement function I fails to produce an appropriate value or by obtaining a score value c.

This algorithm calculates a values of $E(x)$ and invokes the improvement function $a \cdot (c-1)$ times. Since a and c are arbitrary first order objects (not lengths), this takes exponential time. Worse, the algorithm stores the current label values for all $x \in [a]$ and this requires exponential space. Thus, the algorithm is in exponential time but not in polynomial space, and the theory U^1_2 cannot formalize the algorithm directly, unless PSPACE equals exponential time. To circumvent this barrier, we will use a non-deterministic polynomial space algorithm instead. The idea behind the NPSPACE algorithm is simple: rather than storing the labels on all vertices $x \in [a]$, it merely nondeterministically guesses them as needed. This of course does not give the "correct" labels; nonetheless, it will be sufficient to prove the theorem.

We start by describing the NPSPACE algorithm M. The algorithm M sweeps alternately from left-to-right and right-to-left setting labels on vertices x. When M is about to process the vertex numbered x, during a left-to-right sweep, it knows labels for the vertices $x-2$, $x-1$, x, $x+1$, and $x+2$ with score values $s+1$, $s+1$, s, s, and s respectively. Since it is a left-to-right sweep, the value s is even. In addition, the five known labels are wellformed around the three vertices $x-1$, x, and $x+1$; namely, according to the predicate *wf*, the labels are wellformed in the neighborhood of $x-1$, in the neighborhood of x, and in the neighborhood of $x+1$. Since the graph is linear, each neighborhood contains three vertices; for example, the neighborhood of $x-1$ contains the vertices $x-2$, $x-1$, x. M uses the local improvement function I to obtain a new label for vertex x with score value $s+1$. If I produces a label with score value unequal to $s+1$ or if the new label for x causes any of the three vertices $x-1$, x, and $x+1$ to no longer have neighborhoods with wellformed labels, then M halts in a rejecting state. Otherwise, M needs to step one vertex rightward, and for this M discards (forgets) the label for $x-2$ and needs to set a label value for $x+3$. If $s=0$, the label for $x+3$ is set to equal $E(x+3)$. For $s>0$, M merely non-deterministically guesses a label for $x+3$ with score value s. If this label for $x+3$ does not have score value s, or it makes the labels of the vertices $x+1$, $x+2$, $x+3$ in the neighborhood of $x+2$ not be wellformed, then M halts in a rejecting state.[6] Otherwise, M has finished processing vertex x and it proceeds to $x+1$, now with labels for $x-1$, x, $x+1$, $x+2$, and $x+3$.

The algorithm for sweeping right-to-left is entirely dual. In this case, s is odd. When updating the label for vertex x, M knows labels for the vertices $x-2$, $x-1$, x, $x+1$, and $x+2$ with score values s, s, s, $s+1$, and $s+1$. In the next step, to update the label for vertex $x-1$, M forgets the label for $x+2$ and has nondeterministically chosen a label for $x-3$.

At the ends of the linear order, the obvious modifications are made. If $x=a-1$ is the rightmost vertex, then there is no vertex $x+1$ or $x+2$. Or if $x=a-2$, there is no vertex $x+2$. Likewise at $x=0$, there is no vertex $x-2$ or $x-1$, and at $x=1$, no vertex $x-2$. These missing vertices cause no problem: there are fewer neighborhoods in which labels must be wellformed, and their labels are not needed by the improvement function. When reaching $x=a-2$ in a left-to-right scan, M acts purely deterministically as there is no new vertex $x+3$ which needs a label. When reaching $x=a-1$, M initially knows labels for $a-3$, $a-2$, and $a-1$ with score values $s+1$, $s+1$, and s. It updates the label on vertex $a-1$ to have score $s+1$ (unless it rejects), and switches the scan order to right-to-left while staying at the same vertex $x=a-1$. In the next step, as the first step in the right-to-left scan, it invokes I to update the label of $a-1$ to have score value $s+2$, or rejects if I fails to provide such a label. M then rejects if $s+2\geq c$. Otherwise it nondeterministically chooses a label for $a-4$: if this has the wrong score or fails the wellformedness property, M rejects; otherwise, it proceeds one vertex leftward to update the label on vertex $x=a-2$.

The vertices $x=0$ and $x=1$ at the end of right-to-left scan are handled dually.

As defined, any execution of M leads to rejection. There are three possible reasons for rejection: (a) The local improvement function I or the initialization function E may give a label with an incorrect score value or which violates the wellformedness property. (b) The nondeterministic guess of the next vertex's label (on vertex $x+3$ or $x-3$ for rightward or leftward scans, resp.) may give an incorrect score or violate the wellformedness property. (c) A score value may increase to $\geq c$. In either case (a) or (c) occurs, then M has found a point where the LLI conditions are falsified; that is, it has found a solution to the LLI problem. In case (b), no such solution is found. Our goal, thus, is to prove (arguing in U^1_2) that M has some computation that fails for reason (a) or (c).

The steps of a (nondeterministic) computation of M can be indexed with pairs $\langle x,s\rangle$, where x is the vertex number and s the score value. The evenness/oddness of s determines if the sweep is currently left-to-right or right-to-left. A pair $\langle x,s\rangle$ is *M-reachable* if there

[6] As we shall see, this is the "bad" case that we are trying to avoid. It would not happen if M remembered labels from the previous scan instead of just guessing them.

is some computation of M that reaches the point where it is considering $\langle x, s\rangle$ and trying to find a new label for x with score $s+1$. The property of $\langle x, s\rangle$ being reachable is a NPSPACE, and thus a PSPACE property. By induction (IND) on the PSPACE property of reachability, there must be some maximum value s_0 such that some $\langle x, s_0\rangle$ is reachable. If $s_0 = 0$ or $s_0 = c-1$, then M rejects at this step for one of the reasons (a) or (c). So we may assume $0 < s_0 < c-1$. Without loss of generality, s_0 is odd, so M is currently scanning right-to-left. Again using IND induction, there must be some minimum x_0 such that $\langle x_0, s_0\rangle$ is reachable.

Any computation of M that reaches $\langle x_0, s_0\rangle$ ends up with labels for x_0-2, x_0-1, x_0, x_0+1, x_0+2 with scores s_0, s_0, s_0, s_0+1, s_0+1. By choice of $\langle x_0, s_0\rangle$, M rejects while executing this step. If this happens because the improvement function fails to produce a new label for x_0 with score s_0+1 which satisfies the wellformedness properties, then it is the desired failure of type (a). Otherwise, M successfully finds a new label for x with score $s+1$ and with the necessary wellformedness properties, but there is no possible value for a label on $x-3$ with score s such that the labeling around $x-2$ is wellformed. We need to prove that this latter case, (b), can be avoided.

Definition 4.12 We continue to assume s_0 is odd. A (non-deterministic) computation of M is *s_0-consistent at x* provided that during the computation of M, there are label values u and v for vertices $x-2$ and $x-1$ such that u and v both have score s_0, and such that the vertices $x-2$ and $x-1$ have the score s_0 labels u and v at both step $\langle x, s_0-1\rangle$ and step $\langle x, s_0\rangle$.

In other words, the same labels u and v are used for $x-2$ and $x-1$ in the right-to-left scan that raises score values from s_0 to s_0+1 as in the previous left-to-right scan that raised scores from s_0-1 to s_0.

Clearly, any computation of M that reaches the s_0 scan is s_0-consistent at $x = a-1$, since the labels on $a-2$ and $a-3$ do not change when switching over from a left-to-right scan to a right-to-left scan.

For a given vertex x, the question of whether there exists a computation of M which is s_0-consistent at x can be answered by an NPSPACE, hence a PSPACE, algorithm. Thus, by induction (IND), there is a minimum value x_1 such that there is a computation U of M which is s_0-consistent at x_1. If the computation U rejects because of reason (a) at step $\langle x_1, s_0\rangle$, then we are done. Otherwise, we claim that M can continue with the computation U for an additional step so as to be s_0-consistent at x_1-1. Namely, after obtaining an appropriate new label for x_1 with score $s+1$, M existentially chooses the label on x_1-3 to be exactly the same label as in the previous scan. By choice of x_1, the vertices x_1-2 and x_1-1 already have the same label as in the previous scan, thus the labels in the neighborhood of x_1-2 are again well formed since they were well formed in the previous scan. This contradicts the choice of x_1, and completes the proof of Theorem 4.8.

4.3 RLI$_2$ is provable in U_2^1

A similar argument to the one given above shows that RLI$_2$ is provable in U_2^1.

Theorem 4.13 $U_2^1 \vdash \mathrm{RLI}_2$.

Proof. (Sketch) The idea for the NPSPACE algorithm solving RLI$_2$ is to do a similar thing as for LLI, where row numbers in RLI$_2$ now play the role of scores in LLI. The algorithm M sweeps always left-to-right: it starts setting labels in the first row from left to right, then in the second row from left to right, etc. It always guesses all necessary "previously computed" labels from the previous rows. Thus, when M is about to process vertex $\langle x, y\rangle$, it knows labels for vertices $\langle x-2, y\rangle$, $\langle x-1, y\rangle$, $\langle x-2, y-1\rangle$, $\langle x-1, y-1\rangle$, $\langle x, y-1\rangle$, $\langle x+1, y-1\rangle$, $\langle x+2, y-1\rangle$, $\langle x-2, y-2\rangle$, $\langle x-1, y-2\rangle$, $\langle x, y-2\rangle$, $\langle x+1, y-2\rangle$, and $\langle x+2, y-2\rangle$, all with score 1. These labels together with the labels of score 0 given by E for the other vertices in the extended neighborhood of $\langle x, y\rangle$ are extended wellformed. M uses the local improvement function I to obtain a new label for vertex $\langle x, y\rangle$ with score 1. If I produces a label with score value unequal to 1 or if the new label for x causes the labels of the extended neighborhood of $\langle x, y\rangle$ to not be extended wellformed, then M halts in a rejecting state. Otherwise, M needs to step to the next vertex, either one vertex to the right, or to the leftmost vertex of the next row. In the former case, i.e. when $x+1 < a$, M discards (forgets) labels for $\langle x-2, y\rangle$, $\langle x-2, y-1\rangle$ and $\langle x-2, y-2\rangle$, and guesses labels for $\langle x+3, y-1\rangle$ and $\langle x+3, y-2\rangle$. In the latter case, M discards (forgets) all labels and guesses labels for $\langle 0, y\rangle$, $\langle 1, y\rangle$, $\langle 2, y\rangle$, $\langle 0, y-1\rangle$, $\langle 1, y-1\rangle$, and $\langle 2, y-1\rangle$. If any of these labels do not have score 1, or they make the labels in the neighborhood of $\langle x+1, y\rangle$ in the former case, resp. the neighborhoods of $\langle 0, y+1\rangle$ in the latter case, not be extended wellformed, M halts in a rejecting state.

The notion of $\langle x, y\rangle$ being *M-reachable* is defined as before. A computation being *consistent* at $\langle x, y\rangle$ is also defined similarly; namely, the labels that are used when setting the label on vertex $\langle x, y\rangle$ must coincide with the labels that were guessed or computed when setting labels in row $y-1$. Then a similar argument as before shows that this computation leads to a rejection according to (a) or (c), provably in U^1_2. □

We have not been able to characterize the strength of RLI$_3$ or, more generally, the strength of RLI$_k$ for constant $k \geq 3$. In particular, we do not know if they provable in U^1_2, nor if they are many-one complete for the provably total NP search problems of V^1_2. It is also possible they are intermediate in strength.

4.4 Proof of Theorem 4.9

Our proof of Theorem 4.9 is based on the constructions of [**8**], but avoids using T^1_2 as a base theory, and correspondingly avoids a detour through polynomial local search (PLS) problems. We use instead the Witnessing Theorem 3.7 which has S^1_2 as a base theory and thus use a polynomial time computation in place of a PLS computation.

We need to prove that RLI$_{\log}$ is many-one complete for the provably total NP search problems of V^1_2. We will prove this in the following strong form, that applies to "type-2" NP search problems that have a second order input X in addition to a first order input x.

Theorem 4.14 *Suppose ϕ is Δ^b_0 and $(\forall x)(\forall X)(\exists y)\phi(y, x, X)$ is provable in V^1_2. Then, there is a many-one reduction from the* NP *search problem defined by $(\exists y)\phi(y, x, X)$ to an instance of* RLI$_{\log}$. *Furthermore, the many-one reduction is provably correct in S^1_2.*

Proof. Theorem 3.7 implies that there is an exponential time oracle Turing machine M such that S^1_2 proves:

(A): If Y encodes a complete computation of $M^X(x)$, then $\phi(out(Y), x, X)$ is true.

We make some simplifying assumptions about how $M^X(x)$ runs; namely, we assume that M uses a single tape, and that this tape contains three "tracks": the first track is read-only and holds the input x padded with blanks, the second track is also read-only and holds the input X, and the third track is read-write, initially blank. (Equivalently, M has three tapes, but the three tape heads move in lockstep.) Also without loss of generality, Y encodes the computation in some simple, direct fashion; namely, Y can be taken to be the bit-graph of the function H that maps a pair $\langle p, t\rangle$ to the tape contents at position p at time t, the head position at time t, and the state of the machine at time t. With these conventions, writing (A) out in more detail gives that S^1_2 proves

(B): $\phi(out(Y), x, X) \vee$ ("$\exists$ a place in Y where Y fails to satisfy the local conditions of being a correct computation of $M^X(x)$").

Or even more explicitly, S^1_2 proves

(B'): $(\exists y)(y = out(Y) \wedge \phi(y, x, X))$, or
$(\exists p)(\exists t)$[the values given by Y for $H(p, t+1)$, $H(p-1, t)$, $H(p, t)$, and $H(p+1, t)$ do not code consistent information for the computation], or
$(\exists p)[H(p, 0)$ does not equal valid initial tape contents and state for position p at time 0].

Therefore, by the relativized witnessing theorem for S^1_2, there is a polynomial time function f, which takes x as input and uses X and Y as oracles, and produces values for y, p and t satisfying one of the disjuncts of (B'). Without loss of generality, the function f is computed by a clocked Turing machine, so S^1_2 proves its runtime is polynomially bounded. Because of the assumption that Y codes the bit-graph of H, we can view f as asking queries to the function H. That is, rather than querying truth values of $Y(i)$, f makes queries $q = \langle p, t\rangle$ to H and receives for an answer the value $r = H(p, t)$ giving the tape contents at position p at time t, the state at time t, and the tape head position at time t. Without loss of generality, S^1_2 proves that if f outputs values p, t satisfying (B'), then f has actually queried the four values $H(p, t+1)$, $H(p-1, t)$, $H(p, t)$, and $H(p+1, t)$, and that if it outputs a value p, then it has queried $H(p, 0)$.

We will use the computation of $f^{X,Y}(x)$ to set up an instance of $\mathrm{RLI}_{\log}$. We are particularly interested in tracking the queries that f makes to H. For fixed x, X, let $q_i = \langle p_i, t_i\rangle$ be the i-th query made during the computation of $f^{X,Y}(x)$ and $r_i = H(p_i, t_i)$ be the answer received. Since f is polynomial time, without loss of generality, $i = 1, 2, \ldots, p(|x|)$ for some polynomial p. Note that $p(|x|)$ counts only queries to Y, and we do not count the queries made to X.

We next define the instance of $\mathrm{RLI}_{\log}$. The intent is that, provably in S^1_2, any solution to the $\mathrm{RLI}_{\log}$ problem will give a computation of f satisfying (B'). S^1_2 will be able to prove that there are no witnesses p, t or p satisfying the second or third disjunct of (B'); hence, the only possibility for a solution is a value y such that $\phi(y, x, X)$ holds. This will suffice to prove Theorem 4.9.

Our proof uses an amalgamation of techniques from [**8**]. Set $c := 2p(|x|) + 1$ equal to one more than twice the number of queries f makes to H. Let P be the space used by $M^X(x)$, and T the time. The directed graph G will be the rectangular graph $[P] \times [T]$. The edges of G are as described earlier, namely, the up to four edges incoming to (p, t) come from the vertices $(p-1, t)$, $(p-1, t-1)$, $(p, t-1)$, and $(p+1, t-1)$.

The vertices of G will be labeled with sequences. The initialization function E labels with each vertex (p, t) with the empty sequence $\langle\ \rangle$. This has score value 0. The empty sequence is the only label with score 0. The valid labels for a vertex (p, t) with odd score

$2s+1$ are sequences of the form

$$\langle \beta_{p,t}, q_1, r_1, q_2, r_2, \ldots, q_s, ?\rangle$$

and

$$\langle \beta_{p,t}, q_1, r_1, q_2, r_2, \ldots, q_s, r_s\rangle.$$

Here the first entry $\beta_{p,t}$ is intended to equal the value of $H(p,t)$. The q_i values are intended to be queries to H, so $q_i = \langle p_i, t_i\rangle$; the r_i values are intended to be answers to the queries, except the special symbol ? indicates that r_i is not known yet. Note that in general p_i and t_i are unequal to p and t.

The valid labels for (p,t) with even score s have the form

$$\langle \beta_{p,t}, q_1, r_1, q_2, r_2, \ldots, q_s\rangle.$$

Note that the score associated with a label is always the length of the sequence coded by the label. The score values are $\leq 2p(|x|)+1$ and the sizes of the intended entries in the sequences are polynomially bounded, hence there is a term $b = b(x)$ bounding the values of labels.

The wellformedness property, *wf*, applies to labels in the neighborhood N of a vertex (p_0, t_0). In order to be wellformed, the following four sets of conditions must be satisfied.

First the labels in the neighborhood N must agree on q_j and r_j values. Namely, if one of the labels has a value for q_j (the $2j$-th entry of the sequent), then there cannot be a different value for q_j in any other label in N. Similarly, if a non-? value for r_j appears in a label, then there cannot be a different non-? value for r_j in any label in N. Another way to state this is that wellformedness requires that, for $\ell > 1$, if two labels both have an ℓ-th entry, then their ℓ-th entries must be equal, or one of them must equal ?.

Second, the q_j, r_j values must be consistent with the $\beta_{p,t}$ values in the following way: if a vertex in the neighborhood N has coordinates $\langle p', t'\rangle$ and thus its label has first entry $\beta_{p',t'}$, and if some (possibly different) vertex $x = \langle p'', t''\rangle$ in N has an entry $q_j = \langle p', t'\rangle$ then, if r_j is present in x's label, it must satisfy:

- If r_j is the last entry of x's label, and if there is no path in the directed graph G from (p', t') to (p'', t''), then r_j equals ?.
- Otherwise r_j must equal $\beta_{p',t'}$.

Third, the $\beta_{p,t}$ values should, at least locally, behave like valid values for $H(p,t)$; that is to say, the $\beta_{p,t}$ values must be consistent with some potential computation of $M^X(x)$. Except for labels with score zero, all nine vertices in the neighborhood have $\beta_{p,t}$ values, and these values must represent some locally consistent computation of M: in particular, they should contain the correct values in the x- and X-tracks of the tape, and the values for the tape head position, the current state, and the contents of the third tape track must be consistent with the transition relation of the Turing machine M.

Fourth, the wellformedness property *wf* requires a somewhat subtle restriction on the values r_i that prevents them from recording a solution to the second or third disjunct of (B'). Namely, there must not be any $q_s = \langle p, 0\rangle$ where r_i does not equal the correct value $H(p,0)$ describing the initial tape configuration at position p at time 0. In addition, there must not be four query values $q_{i_1} = \langle p-1, t-1\rangle$, $q_{i_2} = \langle p, t-1\rangle$, $q_{i_3} = \langle p+1, t-1\rangle$, and $q_{i_4} = \langle p, t\rangle$ such that r_{i_1}, r_{i_2}, r_{i_3}, and r_{i_4}, when interpreted as values for $H(p-1, t-1)$, $H(p, t-1)$, $H(p+1, t-1)$, and $H(p,t)$, give values that would witness the second conjunct of (B').

This completes the definition of the wellformedness condition *wf*. Note that *wf* is a polynomial time computable function of the (up to) nine labels in a neighborhood, the values of p_0 and t_0 and the input x, using X as an oracle.

We next define the local improvement function I. There are four cases to consider. First, when increasing scores from 0 to 1, the function I must compute the value $\beta_{p,t}$. For t greater than zero, this is determined from $\beta_{p-1,t-1}$, $\beta_{p,t-1}$, $\beta_{p+1,t-1}$ by M's transition relation. For t equal to zero, $\beta_{p,t}$ is just the initial configuration of M for tape cell p; the appropriate values for the read-only tracks are computed by using the input x or the oracle X. Second, consider the case where scores are raised from an even value $2s > 0$ to an odd value $2s + 1$. This represents the case where we are trying to load the answer r_s to the query $q_s = \langle p_s, t_s \rangle$. If (p, t) is not reachable from (p_s, t_s) in the directed graph G, then I just sets r_s equal to ?, leaving the rest of the entries in the label unchanged. If $(p, t) = (p_s, t_s)$, I sets r_s to equal $\beta_{p,t}$. Otherwise, r_s is merely copied from a non-? r_s value of $(p - 1, t)$ or $(p + 1, t - 1)$; by wellformedness, these two values will agree if they are both present and not equal to ?. Third, when increasing a score value from an odd value $2s - 1$ to an even value with $p < P$ or $t < T$, the function I is merely propagating a query value q_s backwards through G: the value q_s for the label on (p, t) is just copied from the q_s value of either $(p+1, t)$ or $(p-1, t+1)$. Fourth, and finally, we define the local improvement function for updating the upper right vertex $(P - 1, T - 1)$ from an odd value $2s - 1$ to $2s$. For this vertex, the local improvement function I simulates $f^{X,Y}(x)$: Whenever f makes a query to $H(p, t)$, the query value $\langle p, t \rangle$ is compared to the values q_i for $i = 1, \ldots, s - 1$. If $\langle p, t \rangle$ equals some q_i in the label on $(P - 1, T - 1)$, then the corresponding r_i value is used as the query answer. Otherwise I sets the new q_s value equal to $\langle p, t \rangle$ leaving the rest of the label entries for $(P - 1, T - 1)$ unchanged. This q_s value will be propagated back and forth across G in the next scans in order to find the answer to the query. On the other hand, if $f^{X,Y}(x)$ halts without making any new query to H, then the local improvement function gives the invalid answer b: this constitutes a solution to the instance of $\mathrm{RLI}_{\log}$.

It should be clear that the local improvement function I is polynomial time in all cases. It uses the oracle X when increasing scores from 0 to 1, and also when increasing the score for $(P - 1, T - 1)$ to $2s$ while simulating the function $f^{X,Y}(x)$ querying X.

This completes the definition of the $\mathrm{RLI}_{\log}$ instance. Suppose (still arguing in S^1_2) that we have a solution to this instance. There are three possible ways to have a solution: (1) The initialization function could produce a value giving a non-wellformed neighborhood; (2) the local improvement function could produce a value giving a non-wellformed neighborhood; or (3) a score value could exceed $c = 2p(|x|) + 1$. Option (3) is impossible since this can happen only at the vertex $(P - 1, T - 1)$, and the function f is constrained to ask fewer than $p(|x|)$ queries to H. Option (1) is likewise impossible, just from the definition of the function E. Likewise, from the definition of the function I, the only way option (2) can occur is at vertex $(P - 1, T - 1)$ for the function $f^{X,Y}(x)$ to successfully halt.

This means that the only possible answer is a place where $f^{X,Y}(x)$ halted successfully while the improvement function I was attempting to update the label on vertex $(P - 1, T - 1)$. Because of the fourth wellformedness condition, this can happen only if $f^{X,Y}(x)$ outputs a value y which satisfies $\phi(y, x, X)$.

Q.E.D. Theorems 4.14 and 4.9. □

References

[1] Arnold Beckmann and Samuel R. Buss. Polynomial local search in the polynomial hierarchy and witnessing in fragments of bounded arithmetic. *Journal of Mathematical Logic*, 9(1):103–138, 2009.

[2] Arnold Beckmann and Samuel R. Buss. *Characterization of Definable Search Problems in Bounded Arithmetic via Proof Notations*, pages 65–134. Ontos Verlag, 2010.

[3] Arnold Beckmann and Samuel R. Buss. Corrected upper bounds for free-cut elimination. *Theoretical Computer Science*, 412(39):5433–5445, 2011.

[4] Samuel R. Buss. *Bounded Arithmetic.* Bibliopolis, 1986. Revision of 1985 Princeton University Ph.D. thesis.

[5] Samuel R. Buss. Axiomatizations and conservation results for fragments of bounded arithmetic. In *Logic and Computation, proceedings of a Workshop held Carnegie-Mellon University, 1987*, vol. 106 of *Contemporary Mathematics*, pages 57–84. American Mathematical Society, 1990.

[6] Samuel R. Buss and Jan Krajíček. An application of Boolean complexity to separation problems in bounded arithmetic. *Proceedings of the London Mathematical Society*, 69:1–21, 1994.

[7] Stephen A. Cook and P. Phuong Nguyen. *Foundations of Proof Complexity: Bounded Arithmetic and Propositional Translations.* ASL and Cambridge University Press, 2010. 496 pages.

[8] Leszek Aleksander Kołodziejczyk, Phuong Nguyen, and Neil Thapen. The provably total NP search problems of weak second-order bounded arithmetic. *Annals of Pure and Applied Logic*, 162(2), 2011.

[9] Jan Krajíček. *Bounded Arithmetic, Propositional Calculus and Complexity Theory.* Cambridge University Press, Heidelberg, 1995.

[10] Jan Krajíček. *Forcing with Random Variables and Proof Complexity*, vol. 232 of *London Mathematical Society Lecture Note Series*, Cambridge University Press, Cambridge, 2011.

[11] Pavel Pudlák and Neil Thapen. Alternating minima and maxima, Nash equilibria and bounded arithmetic. Typeset manuscript, November 2009.

[12] Walter J. Savitch. Relationships between nondeterministic and deterministic tape complexities. *Journal of Computer and System Sciences*, 4(2):177–192, 1970.

[13] Alan Skelley and Neil Thapen. The provably total search problems of bounded arithmetic. *Proceedings of the London Mathematical Society*, 103(1):106–138, 2011.

[14] Gaisi Takeuti. *Proof Theory.* North-Holland, Amsterdam, 2nd edition, 1987.

An improved separation of regular resolution from pool resolution and clause learning

María Luisa Bonet*, Samuel R. Buss†

* Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya, Barcelona, Spain
bonet@lsi.upc.edu

† Department of Mathematics, University of California, San Diego, USA
sbuss@math.ucsd.edu

Abstract. We prove that the graph tautology principles of Alekhnovich, Johannsen, Pitassi and Urquhart have polynomial size pool resolution refutations that use only input lemmas as learned clauses and without degenerate resolution inferences. We also prove that these graph tautology principles can be refuted by polynomial size DPLL proofs with clause learning, even when restricted to greedy DPLL search.

Introduction

The problem SAT of deciding the satisfiability of propositional CNF formulas is of great theoretical and practical interest. Even though it is NP-complete, industrial instances with hundreds of thousands variables are routinely solved by state of the art SAT solvers. Most of these solvers are based on the DPLL procedure extended with clause learning, restarts, variable selection heuristics, and other techniques.

The basic DPLL procedure without clause learning is equivalent to tree-like resolution. The addition of clause learning makes DPLL considerably stronger. In fact, clause learning together with unlimited restarts is capable of simulating general resolution proofs [**12**]. However, the exact power of DPLL with clause learning but without restarts is unknown. This question is interesting not only for theoretical reasons, but also because of the potential for better understanding the practical performance of various refinements of DPLL with clause learning.

Beame, Kautz, and Sabharwal [**4**] gave the first theoretical analysis of DPLL with clause learning. Among other things, they noted that clause learning with restarts simulates general resolution. Their construction required the DPLL algorithm to ignore some contradictions, but this situation was rectified by Pipatsrisawat and Darwiche [**12**] who showed that SAT solvers which do not ignore contradictions can also simulate resolution. These techniques were also applied to learning bounded width clauses by [**2**].

Beame et al. [**4**] also studied DPLL clause learning without restarts. Using a method of "proof trace extensions", they were able to show that DPLL with clause learning and no restarts is strictly stronger than any "natural" proof system strictly weaker than resolution. Here, a *natural* proof system is one in which proofs do not increase in length

[*] Supported in part by grant TIN2010-20967-C04-02.

[†] Supported in part by NSF grants DMS-0700533 and DMS-1101228, and by a grant from the Simons Foundation (#208717 to Sam Buss). The second author thanks the John Templeton Foundation for supporting his participation in the CRM Infinity Project at the Centre de Recerca Matemàtica, Barcelona, Catalonia, Spain, during which these results were obtained.

when variables are restricted to constants. The class of natural proof systems is known to include common proof systems such as tree-like or regular proofs. The proof trace method involves introducing extraneous variables and clauses, which have the effect of giving the clause learning DPLL algorithm more freedom in choosing decision variables for branching.

There have been two approaches to formalizing DPLL with clause learning as a static proof system rather than as a proof search algorithm. The first is pool resolution with a degenerate resolution inference, due originally to Van Gelder [**16**] and studied further by Bacchus et al. [**3**]. Pool resolution requires proofs to have depth-first regular traversal similarly to the search space of a DPLL algorithm. Degenerate resolution allows resolution inferences in which one or both of the hypotheses may be lacking occurrences of the resolution literal. Van Gelder argued that pool resolution with degenerate resolution inferences simulates a wide range of DPLL algorithms with clause learning. He also gave a proof, based on [**1**], that pool resolution with degenerate inferences is stronger than regular resolution, using extraneous variables similar to proof trace extensions.

The second approach is due to Buss–Hoffmann–Johannsen [**8**] who introduced a "partially degenerate" resolution rule called w-resolution, and a proof system regWRTI based on w-resolution and clause learning of "input lemmas". They proved that regWRTI exactly captures non-greedy DPLL with clause learning. By "non-greedy" is meant that contradictions may need to be ignored by the DPLL search.

Both [**3**] and [**8**] gave improved versions of the proof trace extension method so that the extraneous variables need depend only on the set of clauses being refuted and not on resolution refutation of the clauses. The drawback remains, however, that the proof trace extension method gives contrived sets of clauses and contrived resolution refutations.

It remains open whether any of DPLL with clause learning, pool resolution (with or without degenerate inferences), or the regWRTI proof system can polynomially simulate general resolution. One approach to answering these questions is to try to separate pool resolution (say) from general resolution. So far, however, separation results are known only for the weaker system of regular resolution, based on work of Alekhnovitch et al. [**1**], who gave an exponential separation between regular resolution and general resolution. Alekhnovitch et al. [**1**] proved this separation for two families of tautologies, variants of the graph tautologies GT' and the *Stone* pebbling tautologies. Urquhart [**15**] subsequently gave a related separation.[1] In the present paper, we call the tautologies GT' the *guarded* graph tautologies, and henceforth denote them GGT instead of GT'; their definition is given in Section 1.

Thus, an obvious question is whether pool resolution (say) has polynomial size proofs of the GGT tautologies or the *Stone* tautologies. The main result of the present paper resolves the first question by showing that pool resolution does indeed have polynomial size proofs of the graph tautologies GGT. Our proofs apply to the original GGT principles, without the use of extraneous variables in the style of proof trace extensions; our refutations use only the traditional resolution rule and do not require degenerate resolution inferences or w-resolution inferences. In addition, we use only learning of input clauses; thus, our refutations are also regWRTI proofs (and in fact regRTI proofs) in the terminology of [**8**]. As a corollary of the characterization of regWRTI by [**8**], the GGT

[1] Huang and Yu [**10**] also gave a separation of regular resolution and general resolution, but only for a single set of clauses. Goerdt [**9**] gave a quasipolynomial separation of regular resolution and general resolution.

principles have polynomial size refutations that can be found by the DPLL algorithm with clause learning and without restarts (under the appropriate variable selection order).

It is still open if there are polynomial size pool resolution refutations for the *Stone* principles. However, it is plausible that our methods could extend to give such refutations. It seems more likely that our proof methods could extend to the pebbling tautologies used by [**15**], as the hardness of those tautologies was due to the addition of randomly chosen "guard" literals, similarly to the GGT tautologies.[2] A much more ambitious project would be to show that pool resolution or regWRTI can simulate general resolution, or that DPLL with clause learning and without restarts can simulate general resolution. It is far from clear that this is true, but, if so, our methods below may represent a first step in that direction.

The outline of the paper is as follows. Section 1 begins with the definitions of resolution, degenerate resolution, and w-resolution, and then regular, tree, and pool resolution. After that, we define the graph tautologies GT_n and the guarded versions GGT_n, and state the main theorems about proofs of the GGT_n principles. Section 2 gives the proof of the theorems about pool resolution and regRTI proofs. Several ingredients are needed for the proof. The first idea is to try to follow the regular refutations of the graph tautology clauses GT_n as given by Stålmarck [**14**] and Bonet and Galesi [**6**]: however, these refutations cannot be used directly since the transitivity clauses of GT_n are "guarded" in the GGT_n clauses and this yields refutations which violate the regularity/pool property. So, the second idea is that the proof search process branches as needed to learn transitivity clauses. This generates additional clauses that must be proved: to handle these, we develop a notion of "partial bipartite order" and show that the refutations of [**6, 14**] can still be used in the presence of a bipartite partial order. The tricky part is to be sure that exactly the right set of clauses is derived by each subproof. Some straightforward bookkeeping shows that the resulting proof is polynomial size.

Section 3 discusses how to modify the refutations constructed for Section 2 so that they are "greedy". A proof is called greedy provided that, during the proof search process, if it is ever possible to give a simple (i.e., input) refutation of the current clause, then that refutation is used immediately. The greedy condition corresponds well to actual implemented DPLL proof search algorithms, since they backtrack whenever a contradiction can be found by unit propagation.

We are grateful to J. Hoffmann and J. Johannsen for a correction to an earlier version of the proof of Theorem 1.8. We also thank A. Beckmann and T. Pitassi for encouragement and useful comments.

1 Preliminaries and main results

Propositional formulas are defined over a set of variables and the connectives $\wedge$, $\vee$ and $\neg$. We use the notation $\overline{x}$ to express the negation $\neg x$ of x. A *literal* is either a variable x or a negated variable $\overline{x}$. A *clause* C is a set of literals, interpreted as the disjunction of its members. The empty clause, $\Box$, has truth value *False*. We shall only use formulas in *conjunctive normal form*, CNF; namely, a formula will be a set (conjunction) of clauses. We often use disjunction ($\vee$) and union ($\cup$) interchangeably.

[2] Subsequent to the circulation of a preliminary version of the present paper, Buss and Johanssen [in preparation] have succeeded giving polynomial size regRTI proofs of the pebbling tautologies of [**15**].

Definition 1.1 The various forms of resolution take two clauses A and B called the *premises*, and a literal x called the *resolution variable*, and produces a new clause C, called the *resolvent.*

$$\frac{\mathrm{A} \qquad \mathrm{B}}{\mathrm{C}}$$

In all cases below, it is required that $\overline{x} \notin A$ and $x \notin B$. The different forms of resolution are:

Resolution rule: The hypotheses have the forms $A := A' \vee x$ and $B := B' \vee \overline{x}$. The resolvent C is $A' \vee B'$.

Degenerate resolution rule: [**3, 16**] If $x \in A$ and $\overline{x} \in B$, we apply the resolution rule to obtain C. If A contains x, and B does not contain $\overline{x}$, then the resolvent C is B. If A does not contain x, and B contains $\overline{x}$, then the resolvent C is A. If neither A nor B contains the literal x or $\overline{x}$, then C is the lesser of A or B according to some tiebreaking ordering of clauses.

w-resolution rule: [**8**] From A and B as above, we infer $C := (A \setminus \{x\}) \vee (B \setminus \{\overline{x}\})$. It is not required that $x \in A$ or $\overline{x} \in B$.

Definition 1.2 A *resolution derivation*, or *proof*, of a clause C from a CNF formula F is a sequence of clauses $C_1, \ldots, C_s$ such that $C = C_s$ and such that each clause from the sequence is either a clause from F or is the resolvent of two previous clauses. If the derived clause, C_s, is the empty clause, this is called a *resolution refutation* of F. The more general systems of degenerate and w-resolution refutations are defined similarly.

We can represent a derivation as a directed acyclic graph (d.a.g.) on the vertices $C_1, \ldots, C_s$, where each clause from F has out-degree 0, and all the other vertices from $C_1, \ldots, C_s$ have edges pointing to the two clauses from which they were derived. The empty clause has in-degree 0. We use the terms "proof" and "derivation" interchangeably.

Resolution is sound and complete in the refutational sense: a CNF formula F has a refutation if and only if F is unsatisfiable, that is, if and only if $\neg F$ is a tautology. Furthermore, if there is a derivation of a clause C from F, then C is a consequence of F; that is, for every truth assignment σ, if σ satisfies F then it satisfies C. Conversely, if C is a consequence of F then there is a derivation of some $C' \subseteq C$ from F.

A resolution refutation is *regular* provided that, along any path in the directed acyclic graph, each variable is resolved at most once. A resolution derivation of a clause C is *regular* provided that, in addition, no variable appearing in C is used as a resolution variable in the derivation. A refutation is *tree-like* if the underlying graph is a tree; that is, each occurrence of a clause occurring in the refutation is used at most once as a premise of an inference.

We next define pool resolution, using the conventions of [**8**] who called this "tree-like regular resolution with lemmas". The idea is that clauses obtained previously in the proof, can be used freely later on. These clauses act as learned lemmas. To be able to talk about clauses previously obtained, we need to define an ordering of clauses.

Definition 1.3 Given a tree T, the *postorder* ordering $<_T$ of the nodes is defined as follows: if u is a node of T, v is a node in the subtree rooted at the left child of u, and w is a node in the subtree rooted at the right child of u, then $v <_T w <_T u$.

Definition 1.4 A *pool resolution* proof from a set of initial clauses F is a resolution proof tree that fulfills the following conditions: (a) each leaf is labeled with either a clause of F

or a clause that appears earlier in the tree in the $<_T$ ordering; (b) each internal node is labeled with a clause and a variable, and the clause is obtained by resolution from the clauses labeling the node's children, by resolving on the given variable; (c) the proof tree is regular; (d) the roof is labeled with the conclusion clause. If the labeling of the root is the empty clause $\Box$, the pool resolution proof is a pool refutation.

The notions of *degenerate pool resolution* proof and *pool w-resolution* proof are defined similarly, but allowing degenerate resolution or w-resolution inferences, respectively. Note that the two papers [**3, 16**] defined pool resolution to be the degenerate pool resolution system, so our notion of pool resolution is more restrictive than theirs. (Our definition equivalent to the one in [**7**], however.)

Next we define various graph tautologies, sometimes also called "ordering principles". They will all use a size parameter $n > 1$, and variables $x_{i,j}$ with $i, j \in [n]$ and $i \neq j$, where $[n] = \{0, 1, 2, \ldots, n{-}1\}$. A variable $x_{i,j}$ will intuitively represent the condition that $i \prec j$ with $\prec$ intended to be a total, linear order. We will thus always adopt the simplifying convention that $x_{i,j}$ and $\overline{x}_{j,i}$ are the identical literal. This identification makes no essential difference to the complexity of proofs of the tautologies, but it reduces the number of literals and clauses, and simplifies the definitions.

The following principle is based on the tautologies defined by Krishnamurthy [**11**]. These tautologies, or similar ones, have also been studied by [**1, 5, 6, 13, 14**].

Definition 1.5 Let $n > 1$. Then GT_n is the following set of clauses involving the variables $x_{i,j}$, for $i, j \in [n]$ with $i \neq j$.

($\alpha_\emptyset$) The clauses $\bigvee_{j \neq i} x_{j,i}$, for each value $i < n$.

($\gamma_\emptyset$) The *transitivity clauses* $T_{i,j,k} := \overline{x}_{i,j} \vee \overline{x}_{j,k} \vee \overline{x}_{k,i}$ for all distinct i, j, k in $[n]$.

Note that the clauses $T_{i,j,k}$, $T_{j,k,i}$ and $T_{k,i,j}$ are identical. For this reason Van Gelder [**16**] uses the name "no triangles" (NT) for a similar principle.

The next definition is from [**1**], who used the notation GT'_n. They used particular functions r and s for their lower bound proof, but since our upper bound proof does not depend on the details of r and s we leave them unspecified. We require that $r(i,j,k) \neq s(i,j,k)$ and that the set $\{r(i,j,k), s(i,j,k)\} \not\subset \{i,j,k\}$.

Definition 1.6 Let $n \geq 1$, and let $r(i,j,k)$ and $s(i,j,k)$ be functions mapping $[n]^3 \to [n]$ as above. The *guarded graph tautology* GGT_n consists of the following clauses:

($\alpha_\emptyset$) The clauses $\bigvee_{j \neq i} x_{j,i}$, for each value $i < n$.

($\gamma'_\emptyset$) The *guarded* transitivity clauses $T_{i,j,k} \vee x_{r,s}$ and $T_{i,j,k} \vee \overline{x}_{r,s}$, for all distinct i, j, k in $[n]$, where $r = r(i,j,k)$ and $s = s(i,j,k)$.

Our main result is:

Theorem 1.7 *The guarded graph tautology principles* GGT_n *have polynomial size pool resolution refutations.*

The proof of Theorem 1.7 will construct pool refutations in the form of regular tree-like refutations with lemmas. A key part of this is learning transitive closure clauses that are derived using resolution on the guarded transitivity clauses of GGT_n. A slightly modified construction, that uses a result from [**8**], gives instead tree-like regular resolution refutations with *input* lemmas. This will establish the following:

Theorem 1.8 *The guarded graph tautology principles* GGT_n *have polynomial size, tree-like regular resolution refutations with input lemmas.*

A consequence of Theorem 1.8 is that the GGT_n clauses can be shown unsatisfiable by non-greedy polynomial size DPLL searches using clause learning. This follows via Theorem 5.6 of [**8**], since the refutations of GGT_n are regRTI, and hence regWRTI, proofs in the sense of [**8**].

However, as discussed in Section 3, we can improve the constructions of Theorems 1.7 and 1.8 to show that the GGT_n principles can be refuted also by *greedy* polynomial size DPLL searches with clause learning.

2 Proof of main theorem

The following theorem is an important ingredient of our upper bound proof.

Theorem 2.1 (Stålmarck [**14**]; Bonet–Galesi [**6**]) *The sets* GT_n *have regular resolution refutations* P_n *of polynomial size* $O(n^3)$.

We do not include a direct proof of Theorem 2.1 here, which can be found in [**6**] or [**14**]. The present paper uses the proofs P_n as a "black box"; the only property needed is that the P_n's are regular and polynomial size. Lemma 2.7 below is a direct generalization to Theorem 2.1; in fact, when specialized to the case of $\pi = \emptyset$, it is identical to Theorem 2.1.

The refutations P_n can be modified to give refutations of GGT_n by first deriving each transitive clause $T_{i,j,k}$ from the two guarded transitivity clauses of $(\gamma'_\emptyset)$. This however destroys the regularity property, and in fact no polynomial size regular refutations exist for GGT_n [**1**].

As usual, a *partial order* on $[n]$ is an antisymmetric, transitive relation binary relation on $[n]$. We shall be mostly interested in "partial specifications" of partial orders: partial specifications are not required to be transitive.

Definition 2.2 A *partial specification*, τ, of a partial order is a set of ordered pairs $\tau \subseteq [n] \times [n]$ which are consistent with some (partial) order. The minimal partial order containing τ is the transitive closure of τ. We write $i \prec_\tau j$ to denote $\langle i, j\rangle \in \tau$, and write $i \prec_\tau^* j$ to denote that $\langle i, j\rangle$ is in the transitive closure of τ.

The τ-*minimal* elements are the i's such that $j \prec_\tau i$ does not hold for any j.

We will be primarily interested in particular kinds of partial orders, called "bipartite" partial orders, that can be associated with partial orders. A bipartite partial order is a partial order that does not have any chain of inequalities $x \prec y \prec z$.

Definition 2.3 A *bipartite partial order* is a binary relation π on $[n]$ such that the domain and range of π do not intersect. The set of π-minimal elements is denoted M_π.

Figure 1 shows an example. The bipartiteness of π arises from the fact that M_π and $[n] \setminus M_\pi$ partition $[n]$ into two sets. Note that if $i \prec_\pi j$, then $i \in M_\pi$ and $j \notin M_\pi$. In addition, M_π contains the isolated points of π.

Definition 2.4 Let τ be a specification of a partial order. The bipartite partial order π that is *associated with* τ is defined by letting $i \prec_\pi j$ hold for precisely those i and j such that i is τ-minimal and $i \prec_\tau^* j$.

It is easy to check that the π associated with τ is in fact a bipartite partial order. The intuition is that π retains only the information about whether $i \prec_\tau^* j$ for minimal elements i, and forgets the ordering that τ imposes on non-minimal elements. Figure 1 shows an example of how to obtain a bipartite partial order from a partial specification.

We define the graph tautology $\mathrm{GT}_{\pi,n}$ relative to π as follows.

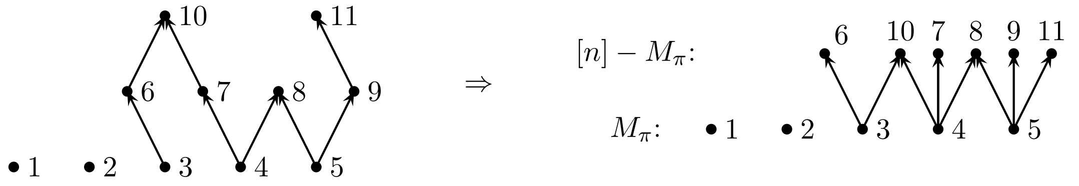


FIGURE 1. Example of a partial specification of a partial order (left) and the associated bipartite partial order (right).

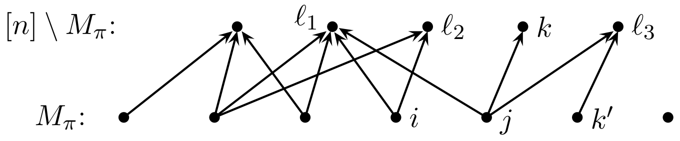


FIGURE 2. A bipartite partial order π is pictured, with the ordered pairs of π shown as directed edges. (For instance, $j \prec_\pi k$ holds.) The set M_π is the set of minimal vertices. The nodes i, j, k shown are an example of nodes used for a transitivity axiom $\overline{x}_{i,j} \vee \overline{x}_{j,k} \vee \overline{x}_{k,i}$ of type (γ). The nodes i, j, k' are an example of the nodes for a transitivity axiom of type (β).

Definition 2.5 Let π be a bipartite partial order on $[n]$. Then $\mathrm{GT}_{\pi,n}$ is the set of clauses containing:

(α) The clauses $\bigvee_{j \neq i} x_{j,i}$, for each value $i \in M_\pi$.

(β) The transitivity clauses $T_{i,j,k} := \overline{x}_{i,j} \vee \overline{x}_{j,k} \vee \overline{x}_{k,i}$ for all distinct i, j, k in M_π. (Vertices i, j, k' in Figure 2 show an example.)

(γ) The transitivity clauses $T_{i,j,k}$ for all distinct i, j, k such that $i, j \in M_\pi$ and $i \not\prec_\pi k$ and $j \prec_\pi k$. (As shown in Figure 2.)

The set $\mathrm{GT}_{\pi,n}$ is satisfiable if π is nonempty. As an example, there is the assignment that sets $x_{j,i}$ true for some fixed $j \notin M_\pi$ and every $i \in M_\pi$, and sets all other variables false. However, if π is applied as a restriction, then $\mathrm{GT}_{\pi,n}$ becomes unsatisfiable. That is to say, there is no assignment which satisfies $\mathrm{GT}_{\pi,n}$ and is consistent with π. This fact is proved by the regular derivation P_π described in the next lemma.

Definition 2.6 For π a bipartite partial order, the clause $(\bigvee\overline{\pi})$ is defined by

$$\left(\bigvee\overline{\pi}\right) := \{\overline{x}_{i,j} : i \prec_\pi j\}.$$

Lemma 2.7 *Let π be a bipartite partial order on $[n]$. Then there is a regular derivation P_π of $(\bigvee\overline{\pi})$ from the set $\mathrm{GT}_{\pi,n}$.*

The only variables resolved on in P_π are the following: the variables $x_{i,j}$ such that $i, j \in M_\pi$, and the variables $x_{i,k}$ such that $k \notin M_\pi$, $i \in M_\pi$, and $i \not\prec_\pi k$.

Note that if π is empty, $M_\pi = [n]$ and there are no clauses of type (γ). In this case, $\mathrm{GT}_{\pi,n}$ is identical to GT_n, and P_π is the same as the refutation of GT_n of Theorem 2.1.

Proof. By renumbering the vertices, we can assume without any loss of generality that $M_\pi = \{0, \ldots, m{-}1\}$. For each $k \geq m$, there is at least one value of j such that $j \prec_\pi k$: let J_k be an arbitrary such value j. Note that $J_k < m$.

Fix $i \in M_\pi$; that is, $i < m$. Recall that the clause of type (α) in $\mathrm{GT}_{\pi,n}$ for i is $\bigvee_{j \neq i} x_{j,i}$. We resolve this clause successively, for each $k \geq m$ such that $i \not\prec_\pi k$, against the clauses $T_{i,J_k,k}$ of type (γ)

$$\overline{x}_{i,J_k} \vee \overline{x}_{J_k,k} \vee \overline{x}_{k,i}$$

using resolution variables $x_{k,i}$. (Note that $J_k \neq i$ since $i \not\prec_\pi k$.) This yields a clause $T'_{i,m}$:

$$\bigvee_{\substack{k \geq m \\ i \not\prec_\pi k}} \overline{x}_{i,J_k} \;\vee\; \bigvee_{\substack{k \geq m \\ i \not\prec_\pi k}} \overline{x}_{J_k,k} \;\vee\; \bigvee_{\substack{k \geq m \\ i \prec_\pi k}} x_{k,i} \;\vee\; \bigvee_{\substack{k < m \\ k \neq i}} x_{k,i}.$$

The first two disjuncts shown above for $T'_{i,m}$ come from the side literals of the clauses $T_{i,J_k,k}$; the last two disjuncts come from the literals in $\bigvee_{j \neq i} x_{j,i}$ which were not resolved on. Since a literal $\overline{x}_{i,J_k}$ is the same literal as $x_{J_k,i}$ and since $J_k < m$, the literals in the first disjunct are also contained in the fourth disjunct. Thus, eliminating duplicate literals, $T'_{i,m}$ is equal to the clause

$$\bigvee_{\substack{k \geq m \\ i \not\prec_\pi k}} \overline{x}_{J_k,k} \;\vee\; \bigvee_{\substack{k \geq m \\ i \prec_\pi k}} x_{k,i} \;\vee\; \bigvee_{\substack{k < m \\ k \neq i}} x_{k,i}.$$

Repeating this process, we obtain derivations of the clauses $T'_{i,m}$ for all $i < m$. The final disjuncts of these clauses, $\bigvee_{i \neq k < m} x_{k,i}$, are the same as the ($\alpha_\emptyset$) clauses in GT_m. Thus, the clauses $T'_{i,m}$ give all ($\alpha_\emptyset$) clauses of GT_m, but with literals $\overline{x}_{J_k,k}$ and $x_{k,i}$ added in as side literals. Moreover, the clauses of type (β) in $\mathrm{GT}_{\pi,n}$ are exactly the transitivity clauses of GT_m. All these clauses can be combined exactly as in the refutation of GT_m described in Theorem 2.1, but carrying along extra side literals $\overline{x}_{J_k,k}$ and $x_{k,i}$, or equivalently carrying along literals $\overline{x}_{J_k,k}$ for $J_k \prec_\pi k$, and $\overline{x}_{i,k}$ for $i \prec_\pi k$. Since the refutation of GT_m uses all of its transitivity clauses and since each $\overline{x}_{J_k,k}$ literal is also one of the $\overline{x}_{i,k}$'s, this yields a resolution derivation P_π of the clause

$$\{\overline{x}_{i,k} : i \prec_\pi k\}.$$

This is the clause ($\bigvee \overline{\pi}$) as desired.

Finally, we observe that P_π is regular. To show this, note that the first parts of P_π deriving the clauses $T'_{i,m}$ are regular by construction, and they use resolution only on variables $x_{k,i}$ with $k \geq m$, $i < m$, and $i \not\prec_\pi k$. The remaining part of P_π is also regular by Theorem 2.1, and uses resolution only on variables $x_{i,j}$ with $i, j \leq m$. □

Proof of Theorem 1.7. We will show how to construct a series of "LR partial refutations", denoted R_0, R_1, $R_2, \ldots$; this process eventually terminates with a pool resolution of GGT_n. The terminology "LR partial" indicates that the refutation is being constructed in left-to-right order, with the left part of the refutation properly formed, but with many of the remaining leaves being labeled with bipartite partial orders instead of with valid learned clauses or initial clauses from GT_n. We first describe the construction of the pool refutation, and leave the size analysis to the end.

An LR partial refutation R is a tree with nodes labeled with clauses that form a correct pool resolution proof, except possibly at the leaves (its initial clauses). Furthermore, it must satisfy the following conditions.

(a) R is a tree. The root is labeled with the empty clause. Each non-leaf node in R has a left child and right child; the clause labeling the node is derived by resolution from the clauses on its two children.

(b) For each clause C occurring in R, the set of ordered pairs $\tau(C)$ is defined as

$$\tau(C) := \{\langle i,j\rangle : \overline{x}_{i,j} \text{ is introduced by resolution on the branch from the root node to } C\}.$$

In many cases, $\tau(C)$ will be a partial specification of a partial order, but this is not always true. For instance, if C is a transitivity axiom, $\tau(C)$ has a 3-cycle and is not consistent as a specification of a partial order.

Note that it follows from condition (a) that $C \subseteq \tau(C)$ since literals in C must be eliminated by resolution inferences somewhere along the path from C to the root of R.

(c) Leaves of R are flagged as “finished” or “unfinished”.

(d) Each finished leaf L is labeled with either a clause from GGT_n or a clause that occurs to the left of L in the postorder traversal of R.

(e) For an unfinished leaf labeled with clause C, the set $\tau(C)$ is a partial specification of a partial order. Furthermore, letting π be the bipartite partial order associated with $\tau(C)$, the clause C is equal to $(\bigvee\overline{\pi})$.

Property (e) is particularly crucial. As shown below, each unfinished leaf, labeled with a clause $C = (\bigvee\overline{\pi})$, will be replaced by a derivation S. The derivation S often will be based on P_π, and thus might be expected to end with exactly the clause C, but some of the resolution inferences needed for P_π might be disallowed by the pool property. This can mean that S will instead be a derivation of a clause C' such that $C \subseteq C' \subseteq \tau(C)$. The fact that $C' \subseteq \tau(C)$ is certainly required, see the comment at the end of condition (b) above. The fact that $C' \supseteq C$ will mean that enough literals are present for the derivation to use only (non-degenerate) resolution inferences —by virtue of the fact that our constructions will pick C so that it contains the literals that must be present for use as resolution literals. The extra literals in $C' \setminus C$ will be handled by propagating them down the proof to where they are resolved on.

The construction begins by letting R_0 be the “empty” refutation, containing just the empty clause. Of course, this clause is an unfinished leaf, and $\tau(\emptyset) = \emptyset$. Thus R_0 is a valid LR partial refutation.

For the induction step, R_i has been constructed already. Let C be the leftmost unfinished clause in R_i. Then R_{i+1} will be formed formed by replacing C by a refutation S of some clause C' such that $C \subseteq C' \subseteq \tau(C)$. Replacing C with C' can introduce extra literals: Since these literals are all in $\tau(C)$, they can be handled by propagating them down the refutation from C, adding each such literal ℓ to every clause below C until reaching a clause where ℓ already appears. (There will be a clause below C which contains ℓ, since $\ell \in \tau(C)$ and is resolved on below C.)

We need to describe the (LR partial) refutation S that will replace the clause C in R_{i+1}. Let π be the bipartite partial order associated with $\tau(C)$, and consider the derivation P_π from Lemma 2.7. Since C is $(\bigvee\overline{\pi})$ by condition (e), the final line of P_π is the clause C. The intuition is that we would like to let S be P_π. The first difficulty with this is that P_π is dag-like, and the LR-refutation is intended to be tree-like, This difficulty, however, can be circumvented by just expanding P_π, which is regular, into a tree-like regular derivation with lemmas by the simple expedient of using an arbitrary depth-first traversal of P_π. The second, and more serious, difficulty is that P_π is a derivation from GT_n, not GGT_n. Namely, the derivation P_π uses the transitivity clauses of GT_n instead of the guarded transitivity clauses of GGT_n. The transitivity clauses

$T_{i,j,k} := \overline{x}_{i,j} \vee \overline{x}_{j,k} \vee \overline{x}_{k,i}$ in P_π are handled one at a time as described below. We will use four separate constructions: in the first case, no change to P_π is required; the second and third cases require a small change; and in the fourth case, the subproof P_π is abandoned in favor of "learning" the transitivity clause.

Before doing the four constructions, it is worth noting that Lemma 2.7 implies that no literal in $\tau(C)$ is used as a resolution literal in P_π. To prove this, suppose $x_{i,j}$ is a resolution variable in P_π. Then, from Lemma 2.7 we have that at least one of i and j is π-minimal and that $i \not\prec_\pi j$ and $j \not\prec_\pi i$. Thus $i \not\prec_{\tau(C)} j$ and $j \not\prec_{\tau(C)} i$, so $\tau(C)$ contains neither $x_{i,j}$ nor $\overline{x}_{i,j}$.

(*i*) If an initial transivitivity clause of P_π already appears earlier in R_i (that is, to the left of C), then it is already *learned*, and can be used freely in P_π.

In the remaining cases (*ii*)–(*iv*), the transitivity clause $T_{i,j,k}$ is not yet learned. Let the guard variable for $T_{i,j,k}$ be $x_{r,s}$, so $r = r(i,j,k)$ and $s = s(i,j,k)$.

(*ii*) Suppose case (*i*) does not apply and that the guard variable $x_{r,s}$ or its negation $\overline{x}_{r,s}$ is a member of $\tau(C)$. The guard variable thus is used as a resolution variable somewhere along the branch from the root to clause C. Then, as just argued above, Lemma 2.7 implies that $x_{r,s}$ is not resolved on in P_π. Therefore, we can add the literal $x_{r,s}$ or $\overline{x}_{r,s}$ (respectively) to the clause $T_{i,j,k}$ and to every clause on any path below $T_{i,j,k}$ until reaching a clause that already contains that literal. This replaces $T_{i,j,k}$ with one of the initial clauses $T_{i,j,k} \vee x_{r,s}$ or $T_{i,j,k} \vee \overline{x}_{r,s}$ of GGT_n. By construction, it preserves the validity of the resolution inferences of R_i as well as the regularity property. Note this adds the literal $x_{r,s}$ or $\overline{x}_{r,s}$ to the final clause C' of the modified P_π. This maintains the property that $C \subseteq C' \subseteq \tau(C)$.

(*iii*) Suppose case (*i*) does not apply and that $x_{r,s}$ is not used as a resolution variable anywhere below $T_{i,j,k}$ in P_π and is not a member of $\tau(C)$. In this case, P_π is modified so as to derive the clause $T_{i,j,k}$ from the two GGT_n clauses $T_{i,j,k} \vee x_{r,s}$ and $T_{i,j,k} \vee \overline{x}_{r,s}$ by resolving on $x_{r,s}$. This maintains the regularity of the derivation. It also means that henceforth $T_{i,j,k}$ will be learned.

If all of the transitivity clauses in P_π can be handled by cases (*i*)–(*iii*), then we use P_π to define R_{i+1}. Namely, let P'_π be the derivation P_π as modified by the applications of cases (*ii*) and (*iii*). The derivation P'_π is regular and dag-like, so we can recast it as a tree-like derivation S with lemmas, by using an arbitrary depth-first traversal of P'_π. The size of S is linear in the size of P_π, since only input lemmas need to be repeated. The final line of S is the clause C', namely C plus the literals introduced by case (*ii*). The derivation R_{i+1} is formed from R_i by replacing the clause C with the derivation S of C', and then propagating each new literal $x \in C' \setminus C$ down towards the root of R_i, adding x to each clause below S until reaching a clause that already contains x. The derivation S contains no unfinished leaf, so R_{i+1} contains one fewer unfinished leaves than R_i.

On the other hand, if even one transitivity axiom $T_{i,j,k}$ in P_π is not covered by the above three cases, then case (*iv*) must be used instead. This introduces a completely different construction to form S:

(*iv*) Let $T_{i,j,k}$ be any transitivity axiom in P_π not covered by cases (*i*)–(*iii*). In this case, the guard variable $x_{r,s}$ is used as a resolution variable in P_π somewhere below $T_{i,j,k}$; in general, this means we cannot use resolution on $x_{r,s}$ to derive $T_{i,j,k}$ while maintaining the desired pool property. Hence, P_π is no longer used, and we instead will form S with a short left-branching path that "learns" $T_{i,j,k}$. This will generate two or three new unfinished leaf nodes. Since unfinished leaf nodes in a LR partial derivation must be labeled with clauses from bipartite partial orders, it is also necessary to attach short derivations to these unfinished leaf nodes to make the unfinished leaf clauses of S correspond correctly to bipartite partial orders. These unfinished leaf nodes are then kept in R_{i+1} to be handled at later stages.

There are separate constructions depending on whether $T_{i,j,k}$ is a clause of type (β) or (γ); details are given below.

First suppose $T_{i,j,k}$ is of type (γ). (Refer to Figure 2.) Let $x_{r,s}$ be the guard variable for the transitivity axiom $T_{i,j,k}$. The derivation S will have the form

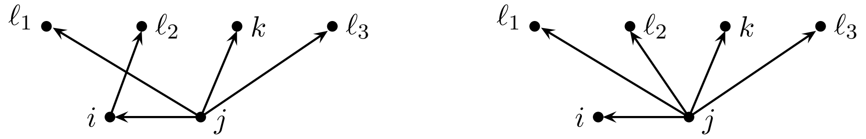
$$\dfrac{\dfrac{\dfrac{\overline{x}_{i,j}, \overline{x}_{j,k}, \overline{x}_{k,i}, x_{r,s} \qquad \overline{x}_{i,j}, \overline{x}_{j,k}, \overline{x}_{k,i}, \overline{x}_{r,s}}{\overline{x}_{i,j}, \overline{x}_{j,k}, \overline{x}_{k,i}} \qquad \begin{matrix} S_1 \cdot\cdot\vdots\cdot\cdot \\ \overline{x}_{i,j}, \overline{x}_{i,k}, \overline{\pi}_{-[jk;jR(i)]} \end{matrix}}{\overline{x}_{i,j}, \overline{x}_{j,k}, \overline{\pi}_{-[jk;jR(i)]}} \qquad \begin{matrix} S_2 \cdot\cdot\vdots\cdot\cdot \\ \overline{x}_{j,i}, \overline{x}_{j,k}, \overline{\pi}_{-[jk;iR(j)]} \end{matrix}}{\overline{x}_{j,k}, \overline{\pi}_{-[jk]}}$$

Here we are using commas instead of disjunctions to denote clauses. The notation $\overline{\pi}_{-[jk]}$ denotes the disjunction of the negations of the literals in π omitting the literal $\overline{x}_{j,k}$. We write "$iR(j)$" to indicate literals $x_{i,\ell}$ such that $j \prec_\pi \ell$. (The "$R(j)$" means "range of j".) Thus $\overline{\pi}_{-[jk;iR(j)]}$ denotes the clause containing the negations of the literals in π, omitting $\overline{x}_{j,k}$ and any literals $\overline{x}_{i,\ell}$ such that $j \prec_\pi \ell$. The clause $\overline{\pi}_{-[jk;jR(i)]}$ is defined similarly, and the notation extends to more complicated situations in the obvious way.

The upper leftmost inference of S is a resolution inference on the variable $x_{r,s}$. Since $T_{i,j,k}$ is not covered by either case (*i*) or (*ii*), the variable $x_{r,s}$ does not appear in or below clause C in R_i. Thus, this use of $x_{r,s}$ as a resolution variable does not violate regularity. Furthermore, since $T_{i,j,k}$ is of type (γ), we have $i \not\prec_{\tau(C)} j$, $j \not\prec_{\tau(C)} i$, $i \not\prec_{\tau(C)} k$, and $k \not\prec_{\tau(C)} i$. Thus the literals $x_{i,j}$ and $x_{i,k}$ do not appear in or below C, so they also can be resolved on without violating regularity.

Let C_1 and C_2 be the final clauses of S_1 and S_2, and let C_1^- be the clause below C_1 and above C. The set of literals $\tau(C_2)$ is obtained by adding $\overline{x}_{j,i}$ to $\tau(C)$, and similarly $\tau(C_1^-)$ is $\tau(C)$ plus $\overline{x}_{i,j}$. Since $T_{i,j,k}$ is type (γ), we have $i, j \in M_\pi$. Therefore, since $\tau(C)$ is a partial specification of a partial order, $\tau(C_2)$ and $\tau(C_1^-)$ are also both partial specifications of partial orders. Let π_2 and π_1 be the bipartite orders associated with these two partial specifications (respectively). We will form the subproof S_1 so that it contains the clause $(\bigvee \overline{\pi}_1)$ as its only unfinished clause. This will require adding inferences in S_1 which add and remove the appropriate literals. The first step of this type already occurs in going up from C_1^- to C_1 since this has removed $\overline{x}_{j,k}$ and added $\overline{x}_{i,k}$, reflecting the fact that j is not π_1-minimal and thus $x_{i,k} \in \pi_1$ but $x_{j,k} \notin \pi_1$. Similarly, we will form S_2 so that its only unfinished clause is $(\bigvee \overline{\pi}_2)$.

We first describe the subproof S_2 of S. The situation is pictured in Figure 3, which shows an extract from Figure 2: the edges shown in part (a) of the figure correspond to the literals present in the final line C_2 of S_2. In particular, recall that the literals $\overline{x}_{i,\ell}$ such that $j \prec_\pi \ell$ are omitted from the last line of S_2. (Correspondingly, the edge from i to ℓ_1 is omitted from Figure 3.) The last line C_2 of S_2 does not correspond to a bipartite partial order because it does not partition $[n]$ into minimal and non-minimal elements;



(a) $\overline{x}_{j,k}, \overline{x}_{i,\ell_2}, \overline{x}_{j,i}, \overline{\pi}^*$ (b) $\overline{x}_{j,k}, \overline{x}_{i,\ell_2}, \overline{x}_{j,i}, \overline{\pi}^*$

FIGURE 3. The partial orders for the fragment of S_2 shown in (2.1).

thus, the last line of S_2 does not qualify to be an unfinished node of R_{i+1}. (An example of this in Figure 3(a) is that $j \prec_{\tau(C_2)} i \prec_{\tau(C_2)} \ell_2$, corresponding to $\overline{x}_{j,i}$ and $\overline{x}_{i,\ell_2}$ being in the last line of S_2.) The bipartite partial order π_2 associated with $\tau(C_2)$ is equal to the bipartite partial order that agrees with π except that each $i \prec_\pi \ell$ condition is replaced with the condition $j \prec_{\pi_2} \ell$. (This is represented in Figure 3(b) by the fact that the edge from i to ℓ_2 has been replaced by the edge from j to ℓ_2. Note that the vertex i is no longer a minimal element of π_2; that is, $i \notin M_{\pi_2}$.) We wish to form S_2 to be a (regular) derivation of the clause $\overline{x}_{j,i}, \overline{\pi}_{-[jk;iR(j)]}$ from the clause $(\bigvee \overline{\pi}_2)$.

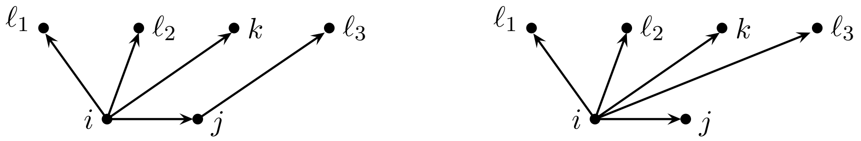
The subproof of S_2 for replacing $\overline{x}_{i,\ell_2}$ in $\overline{\pi}$ with $\overline{x}_{j,\ell_2}$ in $\overline{\pi}_2$ is as follows, letting $\overline{\pi}^*$ be $\overline{\pi}_{-[jk;iR(j);i\ell_2]}$.

$$\dfrac{\overset{S_2' \;\vdots}{\overline{x}_{j,i}, \overline{x}_{i,\ell_2}, \overline{x}_{\ell_2,j}} \qquad \overset{\vdots\; \text{rest of } S_2}{\overline{x}_{j,k}, \overline{x}_{j,\ell_2}, \overline{x}_{j,i}, \overline{\pi}^*}}{\overline{x}_{j,k}, \overline{x}_{i,\ell_2}, \overline{x}_{j,i}, \overline{\pi}^*} \tag{2.1}$$

The part labeled "rest of S_2" will handle similarly the other literals ℓ such that $i \prec_\pi \ell$ and $j \not\prec_\pi \ell$. The final line of S_2' is the transitivity axiom T_{j,i,ℓ_2}. This is a GT_n axiom, not a GGT_n axiom; however, it can be handled by the methods of cases (*i*)–(*iii*). Namely, if T_{j,i,ℓ_2} has already been learned by appearing somewhere to the left in R_i, then S_2' is just this single clause. Otherwise, let the guard variable for T_{j,i,ℓ_2} be $x_{r',s'}$. If $x_{r',s'}$ is used as a resolution variable below T_{j,i,ℓ_2}, then replace T_{j,i,ℓ_2} with $T_{j,i,\ell_2} \vee x_{r',s'}$ or $T_{j,i\ell_2} \vee \overline{x}_{r',s'}$, and propagate the $x_{r',s'}$ or $\overline{x}_{r',s'}$ to clauses down the branch leading to T_{j,i,ℓ_2} until reaching a clause that already contains that literal. Finally, if $x_{r',s'}$ has not been used as a resolution variable in R_i below C, then let S_2' consist of a resolution inference deriving (and learning) T_{j,i,ℓ_2} from the clauses $T_{j,i,\ell_2}, x_{r',s'}$ and $T_{j,i,\ell_2}, \overline{x}_{r',s'}$.

To complete the construction of S_2, the inference (2.1) is repeated for each value of ℓ such that $i \prec_\pi \ell$ and $j \not\prec_\pi \ell$. The result is that S_2 has one unfinished leaf clause, and it is labelled with the clause $(\bigvee \overline{\pi}_2)$.

We next describe the subproof S_1 of S. The situation is shown in Figure 4. As in the formation of S_2, the final clause C_1 in S_1 may need to be modified in order to correspond to the bipartite partial order π_1 which is associated with $\tau(C_1)$. First, note that the literal $\overline{x}_{j,k}$ is already replaced by $\overline{x}_{i,k}$ in the final clause of S_1. The other change that is needed is that, for every ℓ such that $j \prec_\pi \ell$ and $i \not\prec_\pi \ell$, we must make sure that π_1 is defined so that $j \not\prec_{\pi_1} \ell$ and $i \prec_{\pi_1} \ell$. Vertex ℓ_3 in Figure 4 is an example of a such a value ℓ. The ordering in the final clause of S_1 is shown in part (a), and the desired ordered pairs of π_1 are shown in part (b). Note that j is no longer a minimal element in π_1.



(a) $\overline{x}_{i,k},\overline{x}_{j,\ell_3},\overline{x}_{i,j},\overline{\pi}^*$ (b) $\overline{x}_{i,k},\overline{x}_{i,\ell_3},\overline{x}_{i,j},\overline{\pi}^*$

FIGURE 4. The partial orders for the fragment of S_1 shown in (2.2).

The replacement of $\overline{x}_{j,\ell_3}$ with $\overline{x}_{i,\ell_3}$ is effected by the following inference, letting $\overline{\pi}^*$ now be $\overline{\pi}_{-[jk;jR(i);j\ell_3]}$.

$$\frac{\overset{S_1'\ \vdots}{\overline{x}_{i,j},\overline{x}_{j,\ell_3},\overline{x}_{\ell_3,i}} \qquad \overset{\vdots\ \text{rest of } S_1}{\overline{x}_{i,k},\overline{x}_{i,\ell_3},\overline{x}_{i,j},\overline{\pi}^*}}{\overline{x}_{i,k},\overline{x}_{j,\ell_3},\overline{x}_{i,j},\overline{\pi}^*} \tag{2.2}$$

The "rest of S_1" will handle similarly the other literals ℓ such that $j \prec_\pi \ell$ and $i \not\prec_\pi \ell$. Note that the final clause of S_1' is the transitivity axiom T_{i,j,ℓ_3}. The subproof S_1' is formed in exactly the same way that S_2' was formed above. Namely, depending on the status of the guard variable $x_{r',s'}$ for T_{i,j,ℓ_3}, one of the following is done: (*i*) the clause T_{i,j,ℓ_3} is already learned and can be used as is, or (*ii*) one of $x_{r',s'}$ or $\overline{x}_{r',s'}$ is added to the clause and propagated down the proof, or (*iii*) the clause T_{i,j,ℓ_3} is inferred using resolution on $x_{r',s'}$ and becomes learned.

To complete the construction of S_1, the inference (2.2) is repeated for each value of ℓ such that $j \prec_\pi \ell$ and $i \not\prec_\pi \ell$. The result is that S_1 has one unfinished leaf clause, and it corresponds to the bipartite partial order π_1.

That completes the construction of the subproof S for the subcase of (*iv*) where $T_{i,j,k}$ is of type (γ). Now suppose $T_{i,j,k}$ is of type (β). (For instance, the values i, j, k' of Figure 2.) In this case the derivation S will have the form

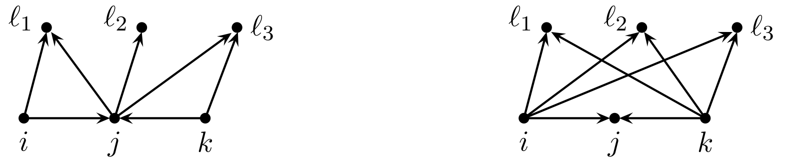
$$\frac{\dfrac{\dfrac{\dfrac{T_{i,j,k},x_{r,s} \qquad T_{i,j,k},\overline{x}_{r,s}}{T_{i,j,k}} \qquad \overset{S_3\ \vdots}{\overline{x}_{i,j},\overline{x}_{i,k},\overline{\pi}_{-[jR(i),kR(i\cup j)]}}}{\overline{x}_{i,j},\overline{x}_{j,k},\overline{\pi}_{-[jR(i),kR(i\cup j)]}} \qquad \overset{S_4\ \vdots}{\overline{x}_{i,j},\overline{x}_{k,j},\overline{\pi}_{-[jR(i\cap k)]}}}{\overline{x}_{i,j},\overline{\pi}_{-[jR(i\cap k)]}} \qquad \overset{S_5\ \vdots}{\overline{x}_{j,i},\overline{\pi}_{-[iR(j)]}}}{\overline{\pi}}$$

where $x_{r,s}$ is the guard variable for $T_{i,j,k}$. We write $[\overline{\pi}_{-[jR(i\cap k)]}]$ to mean the negations of literals in π omitting any literal $\overline{x}_{j,\ell}$ such that both $i \prec_\pi \ell$ and $k \prec_\pi \ell$. Similarly, $\overline{\pi}_{-[jR(i),kR(i\cup j)]}$ indicates the negations of literals in π, omitting the literals $\overline{x}_{j,\ell}$ such that $i \prec_\pi \ell$ and the literals $\overline{x}_{k,\ell}$ such that either $i \prec_\pi \ell$ or $j \prec_\pi \ell$.

Note that the resolution on $x_{r,s}$ used to derive $T_{i,j,k}$ does not violate regularity, since otherwise $T_{i,j,k}$ would have been covered by case (*ii*). Likewise, the resolutions on $x_{i,j}$ and $x_{j,k}$ do not violate regularity since $T_{i,j,k}$ is of type (β).

The subproof S_5 is formed exactly like the subproof S_2 above, with the exception that now the literal $\overline{x}_{j,k}$ is not present. Thus we omit the description of S_5.

We next describe the construction of the subproof S_4. Let C_4 be the final clause of S_4; it is easy to check that $\tau(C_4)$ is a partial specification of a partial order. As before, we must derive C_4 from the clause $(\bigvee\overline{\pi}_4)$ where π_4 is the bipartite partial order associated with the partial order $\tau(C_4)$. A typical situation is shown in Figure 5. As pictured there, it is necessary to add to $\overline{\pi}_4$ the literals $\overline{x}_{i,\ell}$ such that $j \prec_\pi \ell$ and $i \not\prec_\pi \ell$, while removing



(a) $\overline{x}_{i,j}, \overline{x}_{k,j}, \overline{x}_{j,\ell_2}, \overline{\pi}^*$ (b) $\overline{x}_{i,j}, \overline{x}_{i,\ell_2}, \overline{x}_{k,j}, \overline{x}_{k,\ell_2}, \overline{\pi}^*$

FIGURE 5. The partial orders as changed by S_4.

$\overline{x}_{j,\ell}$; examples of this are ℓ equal to ℓ_2 and ℓ_3 in Figure 5. At the same time, we must add the literals $\overline{x}_{k,\ell}$ such that $j \prec_\pi \ell$ and $k \not\prec_\pi \ell$, while removing $\overline{x}_{j,\ell}$; examples of this are ℓ equal to ℓ_1 and ℓ_2 in the same figure.

For a vertex ℓ_3 such that $j \prec_\pi \ell_3$ and $k \prec_\pi \ell_3$ but $i \not\prec_\pi \ell_3$, this is done similarly to the inferences (2.1) and (2.2) but without the side literal $\overline{x}_{j,k}$:

$$\frac{\overset{S_4' \;\vdots}{\overline{x}_{i,j}, \overline{x}_{j,\ell_3}, \overline{x}_{\ell_3,i}} \qquad \overset{\vdots\ \text{rest of } S_4}{\overline{x}_{i,\ell_3}, \overline{x}_{k,j}, \overline{x}_{i,j}, \overline{\pi}^*}}{\overline{x}_{j,\ell_3}, \overline{x}_{k,j}, \overline{x}_{i,j}, \overline{\pi}^*} \tag{2.3}$$

Here $\overline{\pi}^*$ is $\overline{\pi}_{-[jR(i\cap k);j\ell_3]}$. The transitivity axiom T_{i,j,ℓ_3} shown as the last line of S_4' is handled exactly as before. This construction is repeated for all such ℓ_3's.

The vertices ℓ_1 such that $j \prec_\pi \ell_1$ and $i \prec_\pi \ell_1$ but $k \not\prec_\pi \ell_1$ are handled in exactly the same way. (The side literals of π^* change each time to reflect the literals that have already been replaced.)

Finally, consider a vertex ℓ_2 such that $i \not\prec_\pi \ell_2$ and $j \prec_\pi \ell_2$ and $k \not\prec_\pi \ell_2$. This is handled by the derivation

$$\frac{\overset{S_4'' \;\vdots}{\overline{x}_{i,j}, \overline{x}_{j,\ell_2}, \overline{x}_{\ell_2,i}} \qquad \dfrac{\overset{S_4''' \;\vdots}{\overline{x}_{k,j}, \overline{x}_{j,\ell_2}, \overline{x}_{\ell_2,k}} \qquad \overset{\vdots\ \text{rest of } S_4}{\overline{x}_{i,j}, \overline{x}_{i,\ell_2}, \overline{x}_{k,j}, \overline{x}_{k,\ell_2}, \overline{\pi}^*}}{\overline{x}_{i,j}, \overline{x}_{i,\ell_2}, \overline{x}_{k,j}, \overline{x}_{j,\ell_2}, \overline{\pi}^*}}{\overline{x}_{i,j}, \overline{x}_{k,j}, \overline{x}_{j,\ell_2}, \overline{\pi}^*}$$

As before, the set π^* of side literals is changed to reflect the literals that have already been added and removed as S_4 is being created. The subproofs S_4'' and S_4''' of the transitivity axioms T_{i,j,ℓ_2} and T_{k,j,ℓ_2} are handled exactly as before, depending on the status of their guard variables.

Finally, we describe how to form the subproof S_3. For this, we must form the bipartite partial order π_3 which associated with the partial order $\tau(C_3)$, where C_3 is the final clause of S_3. To obtain $\overline{\pi}_3$, we need to add the literals $\overline{x}_{i,\ell}$ such that $i \not\prec_\pi \ell$ and such that either $j \prec_\pi \ell$ or $k \prec_\pi \ell$, while removing any literals $\overline{x}_{j,\ell}$ and $\overline{x}_{k,\ell}$. This is done by exactly the same construction used above in (2.3). The literals in $\overline{\pi}_{-[jR(i);kR(i\cup j)]}$ are exactly the literals needed to carry this out. The construction is quite similar to the above constructions, and we omit any further description.

That completes the description of how to construct the LR partial refutations R_i. The process stops once some R_i has no unfinished clauses. We claim that the process stops after polynomially many stages.

To prove this, recall that R_{i+1} is formed by handling the leftmost unfinished clause using one of cases (i)–(iv). In the first three cases, the unfinished clause is replaced by a derivation based on P_π for some bipartite order π. Since P_π has size $O(n^3)$, this means that the number of clauses in R_{i+1} is at most the number of clauses in R_i plus $O(n^3)$. Also, by construction, R_{i+1} has one fewer unfinished clauses than R_i. In case (*iv*) however, R_{i+1} is formed by adding up to $O(n)$ many clauses to R_i plus adding either two or three new unfinished leaf clauses. In addition, case (*iv*) always causes at least one transitivity axiom $T_{i,j,k}$ to be learned. Therefore, case (*iv*) can occur at most $2\binom{n}{3} = O(n^3)$ times. Consequently at most $3 \cdot 2\binom{n}{3} = O(n^3)$ many unfinished clauses are added throughout the entire process. It follows that the process stops with R_i having no unfinished clauses for some $i \leq 6\binom{n}{3} = O(n^3)$. Therefore there is a pool refutation of GGT_n with $O(n^6)$ lines.

By inspection, each clause in the refutation contains $O(n^2)$ literals. This is because the largest clauses are those corresponding to (small modifications of) bipartite partial orders, and because bipartite partial orders can contain at most $O(n^2)$ many ordered pairs. Furthermore, the refutations P_n for the graph tautology GT_n contain only clauses of size $O(n^2)$.

Q.E.D. Theorem 1.7. □

Theorem 1.8 is proved with nearly the same construction. In fact, the only change needed for the proof is the construction of S from P'_π. Recall that in the proof of Theorem 1.7, the pool derivation S was formed by using an arbitrary depth-first traversal of P. This is not sufficient for Theorem 1.8, since now the derivation S must use only input lemmas. Instead, we use Theorem 3.3 of [**8**], which states that an arbitrary (regular) dag-like resolution derivation can be transformed into a (regular) tree-like derivation with input lemmas. Forming S in this way from P'_π suffices for the proof of Theorem 1.8: the lemmas of S are either transitive closure axioms derived earlier in R_i or are derived by an input subproofs earlier in the post-ordering of S. Since the transitive closure axioms that appeared earlier in R_i were derived by resolving two GGT_n axioms, the lemmas used in S are all input lemmas.

The transformation of Theorem 3.3 of [**8**] may multiply the size of the derivation by the depth of the original derivation. Since it is possible to form the proofs $P\pi$ with depth $O(n)$, the overall size of the pool resolution refutations with input lemmas is $O(n^7)$. This completes the proof of Theorem 1.8.

3 Greedy DPLL with clause learning

We now discuss how the constructions of the refutations for Theorem 1.8 can be modified so as to ensure that the refutations are greedy.

Definition 3.1 Let R be a tree-like regular resolution refutation with input lemmas. For C a clause in R, let $\tau(C)$ be defined as in Section 2. And, let $\Gamma(C)$ be the set of clauses of Γ plus every clause $D <_R C$ in R that has been derived by an input subproof and is available as a learned clause to aid in the derivation of C.

The refutation R is *greedy* provided that, for each clause C of R, if there is an input derivation from $\Gamma(C)$ of some clause $C' \subseteq \tau(C)$ then C is derived in R by an input derivation.

Note that the condition there is an input derivation of some $C' \subseteq \tau(C)$ is equivalent to the condition that if all literals of $\tau(C)$ are set false then unit propagation yields a contradiction.

The definition of greedy is actually a bit more restrictive than necessary, since DPLL algorithms may actually learn multiple clauses at once, and this can mean that C is not derived from a single input but rather from a combination of several input proofs as described in the proof of Theorem 5.1 in [**8**].

Theorem 3.2 *The guarded graph tautology principles* GGT_n *have greedy, polynomial size, tree-like, regular resolution refutations with input lemmas.*

Proof. We indicate how to modify the proof of Theorem 1.8. We again build LR partial refutations R_0, R_1, $R_2, \ldots$ satisfying the same properties a.-e. as before. When forming R_{i+1} from R_i, let C be the unfinished leaf clause in R_i that is to be given an (LR-partial) derivation in R_{i+1}. Let π be as before, and let P_π again be the dag-like regular derivation of C.

We now give a construction that, when successful, incorporates all of transformations (*i*)–(*iv*) and also incorporates the construction of Theorem 3.3 of [**8**]. We unwind the proof P_π into a tree-like regular refutation P^*_π that is possibly exponentially big. We traverse P^*_π in a depth-first, left-to-right order (this is the preorder of P^*_π). Each time we reach a clause C' in P^*_π, we do one of the following modifications to P^*_π.

(*i'*) Suppose that some $C'' \subseteq \tau(C')$ can be derived by an input refutation from $\Gamma(C')$. Fix any such $C'' \subseteq \tau(C')$, and replace the derivation in P^*_π of the clause C' with an input derivation of C'' from $\Gamma(C)$. Any extra literals in $C'' \setminus C'$ are in $\tau(C)$ and are propagated down until reaching a clause where they already appear.

(*ii'*) If case (*i'*) does not apply, and C' is not a leaf node, then P^*_π is unchanged at this point and the depth-first traversal proceeds to the next clause.

(*iii'*) Otherwise, C' is an initial clause of the form $T_{i,j,k}$ and the guard variable $x_{r,s}$ for the clauses $T_{i,j,k} \lor x_{r,s}$ and $T_{i,j,k} \lor x_{r,s}$ is resolved on in P^*_π below C'. In this case, the modification of P^*_π is said to have failed and the traversal process stops.

If case (*iii'*) never occurs, then the modifications of P^*_π eventually terminate. As in the proof of Theorem 3.3 of [**8**], the modified version of P^*_π has polynomial size. Indeed, any clause C' in P^*_π will occur at most d_C times in the modified version of P^*_π where $d_{C'}$ is the depth of the derivation of C' in the original P_π. This is because, C' will have been learned by an input derivation once it has appeared d_C times in the modified derivation P^*_π, as can be proved by induction on d_C.

The situation where case (*iii'*) does occur is handled using the techniques of case (*iv*) of Theorem 1.8. For this case, instead of using the derivation P^*_π of C, we will use one of the derivations S shown on pages 125 and 127.

The second version of S, as shown on page 127, is used in case $T_{i,j,k}$ is type (β). Consider using this derivation fragment S in place of P^*_π to derive C. Let C'' be the clause below $T_{i,j,k}$ in P^*_π. Let D be the clause $\overline{x}_{i,j}, \overline{x}_{j,k}, \overline{\pi}_{-[jR(i),\, kR(i \cup j)]}$ shown in the derivation S on page 127. The set $\tau(C'')$ must contain two of the three literals in $T_{i,j,k}$, and without loss of generality they are the literals $\overline{x}_{i,j}$ and $\overline{x}_{j,k}$. Therefore, since there is no input derivation of any $C''' \subseteq C''$ from $\Gamma(C'')$, there is also no input derivation of any $D' \subseteq D$ from $\Gamma(D)$. Thus, we can attach the resolution fragment S, as shown on page 127, as the subderivation of C, and the resulting derivation is greedy. This leaves three new unfinished clauses, in the subderivations S_3, S_4, and S_5, to be handled at later stages. The clause $T_{i,j,k}$ is now derived by an input proof, and becomes available for later use as an initial clause.

The first version of S, on page 125, is to be used in case $T_{i,j,k}$ is type (γ) in P^*_π. Let D be the clause $\overline{x}_{i,j}, \overline{x}_{j,k}, \overline{\pi}_{-[jk;jR(i)]}$ in S. If there is no input derivation of $D' \subseteq D$ from $\Gamma(D)$, then the construction from the previous paragraph for case (β) works again, and $T_{i,j,k}$ is derived with a input derivation as before, with two new clauses in S_1 and S_2 left as unfinished clauses to be handled at later stages. On the other hand, suppose there is an input derivation of $D' \subseteq D$ from $\Gamma(D)$. This prevents the proof from reaching, and learning, the clause $T_{i,j,k}$, as the traversal of S stops at D. However, this also prunes the subproof S_1, so this generates only one new clause C_2, namely in S_2, to be handled at a later stage. In addition, the bipartite partial order associated with C_2 has one fewer minimal elements (since i is no longer minimal). Therefore, this case can occur at most n times in a row, and cannot cause a superpolynomial increase in proof size.

This completes the proof of Theorem 3.2. □

The construction for the proof of Theorem 3.2 requires only that the clauses $T_{i,j,k}$ are learned whenever possible, and does not depend on whether any other clauses are learned. This means that the following algorithm for DPLL search with clause learning will always succeed in finding a refutation of the GGT_n clauses: At each point, there is a partial assignment τ. The search algorithm must do one of the following:

(1) If unit propagation yields a contradiction, then learn a clause $T_{i,j,k}$ if possible, and backtrack.

(2) Otherwise, if there are any literals in the transitive closure of the bipartite partial order associated with τ which are not assigned a value, branch on one of these literals to set its value. (One of the assignments, true or false, yields an immediate conflict, and may allow learning a clause $T_{i,j,k}$.)

(3) Otherwise, if the regular proof P^*_π can be traversed to give a refutation from the learned clauses, then do this traversal, eventually backtracking from the assignment τ.

(4) Otherwise, find a $T_{i,j,k}$ that is blocking P^*_π from being traversed, and branch on its variables in the order given in the last proof. This either learns $T_{i,j,k}$, or replaces τ with a truth assignment whose associated bipartite partial order has one fewer minimal elements.

References

[1] Michael Alekhnovich, Jan Johannsen, Toniann Pitassi, and Alasdair Urquhart. An exponential separation between regular and general resolution. *Theory of Computation*, 3(4):81–102, 2007.

[2] Albert Atserias, Johannes Klaus Fichte, and Marc Thurley. Clause-learning algorithms with many restarts and and bounded-width resolution. *Journal of Artificial Intelligence Research*, 40:353–373, 2011.

[3] Fahim Bacchus, Philipp Hertel, Toniann Pitassi, and Allen Van Gelder. Clause learning can effectively p-simulate general propositional resolution. In *Proc. 23rd AAAI Conf. on Artificial Intelligence (AAAI 2008)*, pages 283–290. AAAI Press, 2008.

[4] Paul Beame, Henry A. Kautz, and Ashish Sabharwal. Towards understanding and harnessing the potential of clause learning. *J. Artificial Intelligence Research*, 22:319–351, 2004.

[5] Arnold Beckmann and Samuel R. Buss. Separation results for the size of constant-depth propositional proofs. *Annals of Pure and Applied Logic*, 136:30–55, 2005.

[6] María Luisa Bonet and Nicola Galesi. A study of proof search algorithms for resolution and polynomial calculus. In *40th Annual IEEE Symp. on Foundations of Computer Science*, pages 422–431. IEEE Computer Society, 1999.

[7] Samuel R. Buss. Pool resolution is NP-hard to recognise. *Archive for Mathematical Logic*, 48(8):793–798, 2009.

[8] Samuel R. Buss, Jan Hoffmann, and Jan Johannsen. Resolution trees with lemmas: Resolution refinements that characterize DLL-algorithms with clause learning. *Logical Methods of Computer Science*, 4, 4:13(4:13):1–18, 2008.

[9] Andreas Goerdt. Regular resolution versus unrestricted resolution. *SIAM Journal on Computing*, 22(4):661–683, 1993.

[10] Wenqi Huang and Xiangdong Yu. A DNF without regular shortest consensus path. *SIAM Journal on Computing*, 16(5):836–840, 1987.

[11] Balakrishnan Krishnamurthy. Short proofs for tricky formulas. *Acta Informatica*, 22(3):253–275, 1985.

[12] Knot Pipatsrisawat and Adnan Darwiche. On the power of clause-learning satisfiability solvers. *Journal of Automated Reasoning*, 44(3):277–301, 2010.

[13] Nathan Segerlind, Samuel R. Buss, and Russell Impagliazzo. A switching lemma for small restrictions and lower bounds for k-DNF resolution. *SIAM Journal on Computing*, 33(5):1171–1200, 2004.

[14] Gunnar Stålmarck. Short resolution proofs for a sequence of tricky formulas. *Acta Informatica*, 33(3):277–280, 1996.

[15] Aladair Urquhart. A near-optimal separation of regular and general resolution. *SIAM Journal on Computing*, 40(1):107–121, 2011.

[16] Allen Van Gelder. Pool resolution and its relation to regular resolution and DPLL with clause learning. In *Logic for Programming, Artificial Intelligence, and Reasoning (LPAR)*, Lecture Notes in Computer Science Intelligence 3835, pages 580–594. Springer-Verlag, 2005.

Sharpened lower bounds for cut elimination

Samuel R. Buss[*]

[*] Department of Mathematics, University of California, San Diego, USA
sbuss@math.ucsd.edu

Abstract. We present sharpened lower bounds on the size of cut free proofs for first-order logic. Prior lower bounds for eliminating cuts from a proof established superexponential lower bounds as a stack of exponentials, with the height of the stack proportional to the maximum depth d of the formulas in the original proof. Our results remove the constant of proportionality, giving an exponential stack of height equal to $d - O(1)$. The proof method is based on more efficiently expressing the Gentzen–Solovay cut formulas as low depth formulas.

Introduction

The Gentzen cut elimination procedure is a cornerstone of mathematical logic, and is one of the primary tools for establishing the consistency of proof systems, for extracting the constructive content of proofs, and for classifying the strengths of formal systems in terms of their consistency strengths or their computational complexity. It is well-known that cut free proofs may need to be superexponentially larger than proofs that contain cut, as shown originally by Statman [**21, 22**] and Orevkov [**15**]. The present paper sharpens these lower bounds to (almost) match the known upper bounds.

All proofs considered in this paper will be Gentzen-style sequent calculus (LK) proofs in first-order logic. The *depth* of a formula is defined to be the height of a formula when viewed as a tree. The *depth* of a proof is the maximum depth of a cut formula in the proof. The applications in the present paper will be for proofs that have low depth endsequents, and for these proofs, the depth will equal the maximum depth of any formula in the proof. As defined below, the *height* of a proof is the maximum number of non-weak inferences along any branch in the proof.

The base two superexponential function is defined by $2^n_0 = n$ and $2^n_{k+1} = 2^{2^n_k}$. The best known upper bounds on the size of proofs generated by cut elimination state that if a proof P has depth d, then P can be transformed into a cut free proof with size $2^{h(P)}_{d+1}$, where $h(P)$ is the height of P; for this see Orevkov [**16, 17**], Zhang [**25, 26**], Buss [**6**], and the textbook by Troelstra and Schwichtenberg [**23**]. Beckmann–Buss [**4**] give a slightly more general result that applies in the presence of non-logical axioms. Other authors have derived similar, but not quite as sharp upper bounds, including [**5, 13**]. Baaz and Leitsch [**2, 3**] have shown that better upper lower bounds hold in some special cases.

The known *lower* bounds for the size of cut free proofs are also superexponential. The sharpest lower bounds for the Gentzen sequent calculus state that there is a fixed constant ϵ, $0 \leq \epsilon < 1$, and proofs P of arbitrarily large depth d, such that any cut free proof Q with the same endsequent of P has size greater than $2^{h(P)}_{\epsilon d}$. The first such

[*]Supported in part by NSF grants DMS-0700533 and DMS-1101228, and by a grant from the Simons Foundation (#208717). The author thanks the John Templeton Foundation for supporting his participation in the CRM Infinity Project at the Centre de Recerca Matemàtica, Barcelona, Catalonia, Spain, where the main part of this paper was written.

result was proved by Orevkov [**15**], who established this with $\epsilon \approx \frac{1}{4}$, in predicate logic without function symbols. Gerhardy [**11**] obtained $\epsilon \approx \frac{1}{2}$ for first-order logic with function symbols.

The main result of this paper is to improve the lower bound on the size of cut free proofs to obtain $\epsilon \approx 1$. More precisely, we replace the bound $2_{\epsilon d}^{h(P)}$ with the bound 2_{d-c}^{0}, for $c \in \mathbb{N}$ a small constant. This is nearly optimal, as $h(P) = O(d)$.

Our new lower bound also corrects an error in the literature [**27**], which claimed to have established an upper bound of $2_{d/2}^{h(P)}$ on the size of cut free proofs.

Our lower bound can be compared to bounds obtained originally by Zhang [**25, 26**] and refined by Gerhardy [**11, 12**]. They prove that if n is an upper bound on the number of alternations of groups of $\forall$ and $\wedge$ connectives and groups of $\exists$ and $\vee$ connectives in cut formulas, then the size of a cut free proof can be bounded essentially by $2_{n+2}^{h(P)}$. (This is a somewhat simplified and weakened restatement of Zhang's and Gerhardy's upper bounds). In addition, Buss [**6**] shows upper bounds of the form $2_{n+O(1)}^{h(P)}$, where n is the number of alternation of quantifiers in cut formulas, now allowing arbitrary occurrences of intervening propositional connectives.

Our lower bound, like the earlier lower bounds of Statman, Orevkov, Gerhardy, and others, is based on proving that an inductive predicate I contains a large number 2_n^0. Loosely speaking, it is shown that there are short proofs of $I(2_n^0)$, but that any cut free proof of this requires superexponential size. These short proofs are based on defining inductive initial segments (which are sometimes called "inductive cuts", confusingly, since they have nothing to do with cut inferences). The method of defining inductive initial segments goes back essentially to Gentzen [**9**] who used it for proving transfinite induction. It became well-known from Solovay [**20**], who introduced it for use in bounded arithmetic. A number of other authors have also used this technique or similar ones, independently rediscovering it on at least two occasions. These include Statman [**21, 22**], Yessenin-Volpin [**24**], Nelson [**14**], Paris–Dimitracopoulos [**18**], Pudlák [**19**], Baaz–Leitsch [**1**], and Gerhardy [**11**].

Orevkov's lower bound [**15**] constructs short proofs of $I(2_n^0)$, with cuts, using intermediate formulas that have depth $d = O(n)$. Our principal innovation is to improve the depth of these formulas to $n + O(1)$. Section 1 establishes notation by proving a form of Statman's and Orevkov's lower bounds, but with $\epsilon \approx \frac{1}{2}$, over a first-order language with function symbols. This construction is taken almost directly from [**11, 19**]. In Section 2, we improve this to obtain our new lower bound $\epsilon \approx 1$. Section 3 outlines how to prove the same results for first-order logic without function symbols, also with $\epsilon \approx 1$.

1 Preliminaries

We begin with a short review of our formal systems, however the reader is presumed to have basic familiarity with the sequent calculus and cut elimination, as well as at least some familiarity with bounded arithmetic systems such as PV or $I\Delta_0 + \exp$. We work with a sequent calculus for classical logic over the connectives $\forall$, $\exists$, $\wedge$, $\vee$, $\supset$, and $\neg$. The only logical initial sequents are $A \longrightarrow A$, for A an atomic formula. The rules of inference are as shown below.

$$\textit{Exchange:left}\,\frac{\Gamma, A, B, \Lambda \longrightarrow \Delta}{\Gamma, B, A, \Lambda \longrightarrow \Delta} \qquad \textit{Exchange:right}\,\frac{\Gamma \longrightarrow \Delta, A, B, \Lambda}{\Gamma \longrightarrow \Delta, B, A, \Lambda}$$

$$\textit{Contraction:left}\,\frac{A, A, \Gamma \longrightarrow \Delta}{A, \Gamma \longrightarrow \Delta} \qquad \textit{Contraction:right}\,\frac{\Gamma \longrightarrow \Delta, A, A}{\Gamma \longrightarrow \Delta, A}$$

$$\textit{Weakening:left}\,\frac{\Gamma \longrightarrow \Delta}{A, \Gamma \longrightarrow \Delta} \qquad \textit{Weakening:right}\,\frac{\Gamma \longrightarrow \Delta}{\Gamma \longrightarrow \Delta, A}$$

$$\neg\textit{:left}\,\frac{\Gamma \longrightarrow \Delta, A}{\neg A, \Gamma \longrightarrow \Delta} \qquad \neg\textit{:right}\,\frac{A, \Gamma \longrightarrow \Delta}{\Gamma \longrightarrow \Delta, \neg A}$$

$$\wedge\textit{:left}\,\frac{A, B, \Gamma \longrightarrow \Delta}{A \wedge B, \Gamma \longrightarrow \Delta} \qquad \wedge\textit{:right}\,\frac{\Gamma \longrightarrow \Delta, A \qquad \Gamma \longrightarrow \Delta, B}{\Gamma \longrightarrow \Delta, A \wedge B}$$

$$\vee\textit{:left}\,\frac{A, \Gamma \longrightarrow \Delta \qquad B, \Gamma \longrightarrow \Delta}{A \vee B, \Gamma \longrightarrow \Delta} \qquad \vee\textit{:right}\,\frac{\Gamma \longrightarrow \Delta, A, B}{\Gamma \longrightarrow \Delta, A \vee B}$$

$$\supset\textit{:left}\,\frac{\Gamma \longrightarrow \Delta, A \qquad B, \Gamma \longrightarrow \Delta}{A \supset B, \Gamma \longrightarrow \Delta} \qquad \supset\textit{:right}\,\frac{A, \Gamma \longrightarrow \Delta, B}{\Gamma \longrightarrow \Delta, A \supset B}$$

$$\forall\textit{:left}\,\frac{A(t), \Gamma \longrightarrow \Delta}{(\forall x)A(x), \Gamma \longrightarrow \Delta} \qquad \forall\textit{:right}\,\frac{\Gamma \longrightarrow \Delta, A(b)}{\Gamma \longrightarrow \Delta, (\forall x)A(x)}$$

$$\exists\textit{:left}\,\frac{A(b), \Gamma \longrightarrow \Delta}{(\exists x)A(x), \Gamma \longrightarrow \Delta} \qquad \exists\textit{:right}\,\frac{\Gamma \longrightarrow \Delta, A(t)}{\Gamma \longrightarrow \Delta, (\exists x)A(x)}$$

$$\text{Cut}\,\frac{\Gamma \longrightarrow \Delta, A \qquad A, \Gamma \longrightarrow \Delta}{\Gamma \longrightarrow \Delta}$$

The $\forall$:right and $\exists$:left inferences must satisfy the usual eigenvariable condition that b does not appear in the lower sequent.

The first six inferences are called *weak inferences*: these are needed since we treat cedents as sequences of formulas, rather than as sets or multisets of formulas. However, the *size* $|P|$ of a proof is defined to be equal to the number of non-weak inferences. The *height* of P is denoted $h(P)$ and is the maximum number of non-weak inferences along any branch in the proof.

Definition 1.1 The *depth* of a formula A is defined as follows:

(a) If A is atomic, then $\mathrm{depth}(A) = 0$.
(b) If A is $\neg B$, $(\exists x)B$, or $(\forall x)B$, then $A = 1 + \mathrm{depth}(B)$.
(c) If A is $B \circ C$ for $\circ$ one of $\vee$, $\wedge$ or $\supset$, then

$$\mathrm{depth}(A) = 1 + \max\{\mathrm{depth}(B),\ \mathrm{depth}(C)\}.$$

The depth of a cut inference is the depth of its cut formula. The depth of a proof P is the maximum depth of cuts appearing in P.

We use a special notation for an "extended" superexponential function. Let $\vec{u}$ be a finite sequence $\vec{u} = \langle u_1, \ldots, u_k\rangle$, with $k \geq 1$. The value $2_{\vec{u}}$ is defined inductively. For $\vec{u} = \langle u_1\rangle$, a sequence of length one, $2_{\langle u_1\rangle} = u_1$. And, for $\vec{u} = \langle u_1, \ldots, u_k\rangle$, $2_{\vec{u}} = u_1 + 2^{2_{\langle u_2, \ldots, u_k\rangle}}$. For instance,

$$2_{\langle a,b,c,d\rangle} \ = \ a + 2^{b+2^{c+2^d}}.$$

We now review the prior superexponential lower bound for cut elimination, based on Pudlák's exposition [**19**], but with the better lower bound of $\epsilon \approx \frac{1}{2}$ as obtained by

Gerhardy [**11**]. We let T be a finitely axiomatized theory of bounded arithmetic which contains a finite fragment of Cook's theory PV plus the exponential function 2^i and the superexponential functions 2^x_i and $2_{\langle\vec{u}\rangle}$. The language of T contains function symbols for sufficiently many polynomial time computable functions to formalize the needed arguments described below: this includes sequence coding, and proving simple properties about the needed polynomial time computable functions and about the exponential and superexponential functions. The theory T is axiomatized by a finite set of purely universal formulas.

T contains an additional, uninterpreted, unary predicate symbol $I(x)$, with the two axioms $I(0)$ and $(\forall x)(I(x) \supset I(Sx))$. The predicate I is not permitted in induction axioms. The predicate $I(x)$ intuitively means that induction works up to x, or that x can be reached from zero by repeatedly adding 1. Define the formula $\psi_0(x)$ to be $I(x)$, and for $i \geq 0$, define $\psi_{i+1}(x)$ to be the formula

$$(\forall y)(\psi_i(y) \supset \psi_i(y + 2^x)).$$

There are then simple proofs of

$$\psi_i(0) \text{ and } \forall x(\psi_i(x) \supset \psi_i(Sx)). \tag{1.1}$$

These are proved for successive values of i using simple properties of zero and successor; namely, as we show below, the formulas (1.1) for $i = k + 1$ are proved from those for $i = k$. In addition, as we detail below, it is easy to prove that $\psi_{i+1}(x) \supset \psi_i(2^x)$.

Let Γ be the set of universal formulas that axiomatize T, including the two axioms for the predicate $I(x)$, and the equality axioms for the relation and functions symbols of T. As we describe below, the sequents $\psi_{i+1}(x) \longrightarrow \psi_i(2^x)$ can be proved with a proof of height $O(i)$ which contain cuts only on atomic formulas and on substitution instances of subformulas of ψ_i. Likewise, the sequent $\longrightarrow \psi_i(0)$ is proved with proofs with height $O(i)$ and with the same cut complexity. Combining these sequents with cuts, we get a proof P_ℓ of $\Gamma \longrightarrow I(2^0_\ell)$ which has height $O(\ell)$ and in which all cut formulas either are atomic or are substitution instances of subformulas of $\psi_\ell(x)$.

Let Q_ℓ be a proof with the same conclusion $\Gamma \longrightarrow I(2^0_\ell)$ as P_ℓ in which all cuts are on quantifier-free formulas. We claim that the size of Q_ℓ is $\geq 2^0_\ell$. To prove this, we modify Q_ℓ in the following fashion. Find each $\forall$:left inference in Q_ℓ, and omit this inference and instead let the auxiliary formula of the inference remain in the antecedent of that sequent and in all sequents below that sequent, down to the endsequent. For this, contractions on (formerly universal) formulas are omitted. The result is a proof Q^*_ℓ of a sequent $\Gamma^* \longrightarrow I(2^0_\ell)$ in which every formula in Γ^* is a quantifier-free substitution instance of an axiom of T. Without loss of generality, Γ^* does not contain any variables, since any variables that are present may be replaced everywhere with the constant 0. Note that the number of formulas in Γ^* is less than or equal to the number of $\forall$:right inferences in Q_i plus the number of quantifier-free axioms in the (finite) set Γ. In particular, the number of substitution instances of $I(x) \supset I(S(x))$ in Γ^* is less than the size of Q_ℓ.

Each such substitution instance of $I(x) \supset I(S(x))$ is a formula of the form $I(s) \supset I(S(s))$, for s a closed term. Let $n_0 \in \mathbb{N}$ be the least integer so that no s has value equal to n_0. Of course n_0 must be less than the size of Q_ℓ. On the other hand, we claim that $n_0 \geq 2^0_\ell$. Otherwise, we could falsify the sequent $\Gamma^* \longrightarrow I(2^0_\ell)$ in the standard model of the integers by letting $I(n)$ hold for exactly the values $n \leq n_0$. It follows that the size of Q_ℓ is greater than or equal to 2^0_ℓ.

This is enough to establish the superexponential lower bound on cut free proofs. However, it is worth examining in more detail how the proof P_ℓ can be formed. First, P_ℓ derives the sequents

$$\Gamma \longrightarrow \psi_i(0) \tag{1.2}$$

and

$$\Gamma, \psi_i(a) \longrightarrow \psi_i(S(a)) \tag{1.3}$$

for $0 \le i \le \ell$, where a is a free variable. For $i = 0$, these are simple to prove without cuts. For the induction step, P_ℓ derives (1.2) with $i = k+1$ from the three sequents

(i) $\Gamma, \psi_k(a) \longrightarrow \psi_k(S(a))$,
(ii) $\Gamma \longrightarrow S(a) = a + 2^0$,
(iii) $\Gamma, S(a) = a + 2^0, \psi_k(S(a)) \longrightarrow \psi_k(a + 2^0)$,

using cuts on the formulas $S(a) = a + 2^0$ and $\psi_k(S(a))$ followed by an $\supset$:right and a $\forall$:right. The sequent (i) is (1.3) with $i = k$. Sequent (ii) is provable by a fixed size cut free proof. And, since Γ includes equality axioms, (iii) has a cut free proof of height $O(k)$. (This last fact is readily proved by induction on the depth of ψ_k from the fact that ψ_k has depth $O(k)$.)

As the second part of the induction step, P_ℓ derives (1.3) for $i = k+1$ from the sequents

(i) $\psi_{k+1}(a), \psi_k(b) \longrightarrow \psi_k(b + 2^a)$,
(ii) $\psi_{k+1}(a), \psi_k(b + 2^a) \longrightarrow \psi_k((b + 2^a) + 2^a)$,
(iii) $\Gamma \longrightarrow (b + 2^a) + 2^a = b + 2^{S(a)}$,
(iv) $\Gamma, (b + 2^a) + 2^a = b + 2^{S(a)}, \psi_k((b + 2^a) + 2^a) \longrightarrow \psi_k(b + 2^{S(a)})$,

using cuts on the atomic formula $(b + 2^a) + 2^a = b + 2^{S(a)}$ and the formulas $\psi_k(b + 2^a)$ and $\psi_k((b + 2^a) + 2^a)$, followed by an $\supset$:right and a $\forall$:right. Note that (i) and (ii) are readily provable by fixed proof schemes without any cuts.

After proving all the instances of (1.2) and (1.3), P_ℓ derives the sequents

$$\Gamma, \psi_{k+1}(a) \longrightarrow \psi_k(2^a) \tag{1.4}$$

for $0 \le k < \ell$. This sequent is proved from the sequents

(i) $\psi_{k+1}(a), \psi_k(0) \longrightarrow \psi_k(0 + 2^a)$,
(ii) $\Gamma \longrightarrow \psi_k(0)$,
(iii) $\Gamma \longrightarrow 0 + 2^a = 2^a$,
(iv) $\Gamma, 0 + 2^a = 2^a, \psi_k(0 + 2^a) \longrightarrow \psi_k(2^a)$,

using cuts on the formulas $0 + 2^a = 2^a$, $\psi_k(0)$, and $\psi_k(0 + 2^a)$. Note that (i) is provable by a small proof with no cuts, and that (ii) is the same as (1.2).

Finally, P_ℓ derives $\Gamma \longrightarrow \psi_0(2^0_\ell)$ from the sequent (1.2) with $i = \ell$, the sequents (1.4) for $0 \le i < \ell$, and the sequents

(i) $\Gamma \longrightarrow 2^{2^0_i} = 2^0_{i+1}$,
(ii) $\Gamma, 2^{2^0_i} = 2^0_{i+1}, \psi_{\ell-i-1}(2^{2^0_i}) \longrightarrow \psi_{\ell-i-1}(2^0_{i+1})$,

using cuts on the indicated formulas.

By inspection, the height of P_ℓ is $O(\ell)$. Its depth is 2ℓ, since $\psi_\ell(0)$ is the cut formula of maximum depth. We have thus reproved, taking $d = 2\ell$, the prior results for lower bounds on cut-elimination that were described in the introduction:

Theorem 1.2 *There are proofs P_ℓ of sequents $\mathfrak{S}_\ell$ of depth d and height $O(d)$ such that any cut free proof of $\mathfrak{S}_\ell$ requires size $2^0_{(1/2)d}$. The formulas in $\mathfrak{S}_\ell$ are purely universal and have depth $O(1)$.*

The proof P_ℓ constructed above has exponential size because the formulas ψ_i have exponential size, $O(2^i)$. These formulas could be replaced by polynomial size formulas, as is done by Pudlák [**19**] using constructions from Ferrante–Rackoff [**8**]. They could even be made linear size using the refinements to [**8**] by Buss–Johnson [**7**]. With these modifications, P_ℓ would be polynomial size; its depth would become larger than 2ℓ, although it would still be $O(\ell)$.

2 Improved lower bounds for cut-elimination

We now improve Theorem 1.2 to establish the $\epsilon \approx 1$ version of the lower bounds on the size of cut free proofs. The idea is to modify the formulas ψ_i used in P_ℓ so that they have depth $i + O(1)$ instead of depth $2i$. For this we shall prove there are formulas φ_i (equivalent to ψ_i) such that $\varphi_i(x)$ has depth $i + O(1)$, and $\varphi_0(x)$ is $I(x)$, and the formulas

$$\varphi_{i+1}(x) \leftrightarrow (\forall y)(\varphi_i(y) \supset \varphi_i(y + 2^x)) \tag{2.1}$$

have proofs of height $O(i)$ and depth $i + O(1)$. The proof P_ℓ can then be carried out using the φ_i's in place of the ψ_i's, and this will give the desired lower bound on cut elimination.

Although the details will be a bit complicated, the intuition behind the construction of the φ_i's is simple. The formula $\psi_i(w)$, although exponential size, has prenex form that is a Π_i-formula after like quantifiers are collapsed. Thus, $\psi_i(w)$ can be equivalently expressed as a formula $\varphi_i(w)$ of the form

$$(\forall y_0)(\exists y_1) \cdots (Q y_{i-1}) R(\langle y_0, \ldots, y_{i-1}\rangle, w), \tag{2.2}$$

where R is a superexponential-time computable relation. We will not be able to add R as a predicate symbol to T as this seems to be precluded by the fact that the predicate symbol I cannot be used in induction axioms. Instead, we will introduce a finite set of new predicate and function symbols to the theory T, which will enable T to define R as a constant depth formula. After doing this, the principal task is to prove that the formulas (1.1) with ψ_i replaced with φ_i have T-proofs of depth $i + O(1)$.

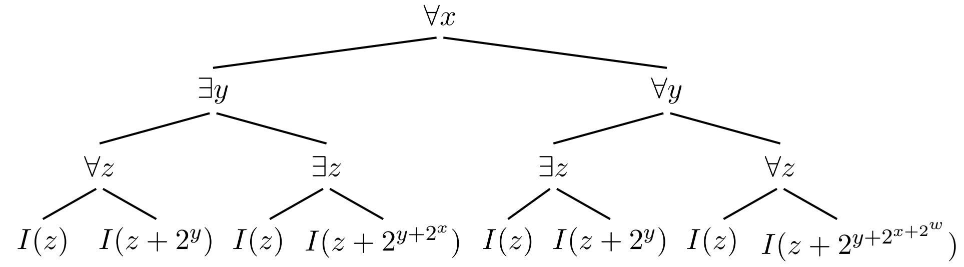
We begin by describing how to express the condition R. Recall that $\psi_0(z)$ is $I(z)$, and that $\psi_1(y)$ is $\forall z(I(z) \supset I(z + 2^y))$. Expanding further gives that $\psi_2(x)$ is

$$\forall y(\,\forall z(I(z) \supset I(z + 2^y)) \;\supset\; \forall z(I(z) \supset I(z + 2^{y+2^x}))\,),$$

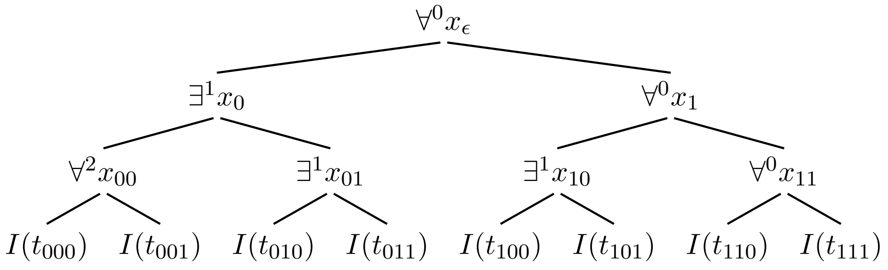
and that $\psi_3(w)$ is

$$\begin{aligned}\forall x[\,&\forall y(\,\forall z(I(z) \supset I(z + 2^y)) \;\supset\; \forall z(I(z) \supset I(z + 2^{y+2^x}))\,) \;\supset \\ &\forall y(\,\forall z(I(z) \supset I(z + 2^y)) \;\supset\; \forall z(I(z) \supset I(z + 2^{y+2^{x+2^w}}))\,)\,].\end{aligned}$$

To better see the pattern, consider a "skeletal" tree representation of $\psi_3(w)$.



The skeletal tree shows the quantifier structure of ψ_3, but omits the propositional connectives to keep it simpler. The skeletal tree can be written in a more generic form as follows:



This is intended to represent the fact that ψ_3 is equivalent to the prenex formula

$$\begin{aligned}\forall x_\epsilon \forall x_1 \forall x_{11} \exists x_0 \exists x_{01} \exists x_{10} \forall x_{00}[((I(t_{000}) \supset I(t_{001})) \supset (I(t_{010}) \supset I(t_{011}))) \\ \supset ((I(t_{100}) \supset I(t_{101})) \supset (I(t_{110}) \supset I(t_{111})))].\end{aligned}$$

The superscripts on the quantifiers indicate the order in which quantifiers are pulled out when putting ψ_3 in prenex form. For example, x_{11} is in the first (outermost) block of quantifiers of ψ_3's prenex form instead of the third (innermost) block.

The subscripts on the t's and x's indicate the path in the tree to reach that node, with "0" and "1" indicating left and right respectively. For instance, the term t_{011} (which is in fact the term $x_{01} + 2^{x_0 + 2^{x_\epsilon}}$) is reached by starting at the root and descending left, then right, then right. The empty sequence is denoted by "ϵ".

The pattern for ψ_3 generalizes to form skeletal trees of ψ_i, $i \geq 1$. The formation rules are as follows. The quantified variables in ψ_i are $x_{\vec{u}}$, for $\vec{u} \in \{0,1\}^{<i}$. The level $\ell = \ell(\vec{u})$ on the quantifier $Q^\ell x_{\vec{u}}$ is equal to the number of 0's in $\vec{u}$. The variable $x_{\vec{u}}$ is universally quantified iff its level $\ell(\vec{u})$ is even. The atomic subformulas of ψ_i are of the form $I(t_{\vec{u}})$ for $\vec{u} \in \{0,1\}^i$. If $\vec{v}$ is a sequence, let $|\vec{v}|$ denote the length of $\vec{v}$. For $p \leq |\vec{v}|$, let $\vec{v} \upharpoonright p$ denote the sequence containing the first p elements of $\vec{v}$. For $t_{\vec{u}}$ a term and $\vec{u} \in \{0,1\}^i$, we define $\nu_{\vec{u}}$ to be the sequence

$$\nu_{\vec{u}} \ := \ \langle x_{\vec{u}\upharpoonright(i-1)}, x_{\vec{u}\upharpoonright(i-2)}, \ldots, x_{\vec{u}\upharpoonright 1}, x_\epsilon, w \rangle,$$

namely, the variables along the path to node $\vec{u}$ plus the free variable w: this is the sequence of variables that potentially could appear in $t_{\vec{u}}$. Then, $t_{\vec{u}}$ is the superexponential term

$$t_{\vec{u}} \ := \ 2_{\nu_{\vec{u}}\upharpoonright(r+1)}$$

where r is the number of contiguous 1's occurring at the end of $\vec{u}$. For example, in the formula trees above, for t_{011}, there are two 1's at the end of "011", so t_{011} is equal to $2_{\langle x_{01}, x_0, x_\epsilon \rangle}$, the extended superexponential function with the subscript a sequence of length $3 = r + 1$.

A variable y_ℓ in φ_i —see (2.2) above— will code a sequence containing the values of the variables $x_{\vec{u}}$ with level $\ell(\vec{u})$ equal to ℓ. Letting $\vec{y}$ be $\langle y_0, \ldots, y_{i-1}\rangle$, the entry y_ℓ is "well-formed" provided that it codes a function with domain equal to the set of sequences $x_{\vec{u}}$ with $|\vec{u}| < i$ and $\ell(\vec{u}) = \ell$. If y_ℓ is not well-formed, then by convention it codes the constant function which is equal to zero on all inputs in its domain.

For $\vec{u} \in \{0,1\}^{<i}$, we write $X(\vec{u})$ to mean the value that $\vec{u}$ is mapped to by the function encoded by $y_{\ell(\vec{u})}$. (The intuition is that $X(\vec{u})$ equals the value of the variable $x_{\vec{u}}$.) We write $t(\vec{u})$ for the value of $t_{\vec{u}}$ when the variables $x_{\vec{u}'}$ are given the values $X(\vec{u}')$. Note that, although it is suppressed in the notation, $X(\vec{u})$ depends on the vector of values $\vec{y}$. Also, $t(\vec{u})$ depends on both $\vec{y}$ and w, and we sometimes will write it as $t(\vec{u}, \vec{y}, w)$.

Let n be a power of two. Suppose $\vec{\sigma} \in \{T, F\}^n$, $\vec{\sigma} = \langle \sigma_0, \ldots, \sigma_{n-1}\rangle$, where T and F stand for "True" and "False". Define the relation $\mathit{BIT}(\vec{\sigma})$ by ("BIT" stands for "binary implication tree")

$$\mathit{BIT}(\vec{\sigma}) = \begin{cases} \sigma_0 & \text{if } |\vec{\sigma}| = 1, \\ \mathit{BIT}(\langle \sigma_0, \ldots, \sigma_{n/2-1}\rangle) \supset \mathit{BIT}(\langle \sigma_{n/2}, \ldots, \sigma_{n-1}\rangle) & \text{otherwise.} \end{cases}$$

We identify binary vectors $\vec{u}$ in $\{0,1\}^i$ with integers, and write $\mathit{nm}(\vec{u})$ for the integer with binary representation given by $\vec{u}$.

We now can define the formula $R(\vec{y}, w)$ in (2.2) to be

$$(\exists \vec{\sigma} \in \{0,1\}^{2^i})(\mathit{BIT}(\vec{\sigma}) \wedge (\forall \vec{u} \in \{0,1\}^i)[\sigma_{\mathit{nm}(\vec{u})} = 1 \leftrightarrow I(t(\vec{u}))]).$$

Note that $\leftrightarrow$ is not in our first-order language; instead $A \leftrightarrow B$ is an abbreviation for $(A \supset B) \wedge (B \supset A)$. By inspection, the depth of R equals 5.

This completes the definition (2.2) of the formulas $\varphi_i(w)$. Clearly, φ_i has depth $i + O(1)$, namely depth i plus the depth of R.

We now give a sketch of the proof that the equivalences (2.1) have T-proofs of depth $i + O(1)$. Note that the intuition behind the definition of R is that R states that a tree of implications holds. We define formulas S_0 and S_1 that express, respectively, the hypothesis and the conclusion of the implication, so that R is equivalent to $S_0 \supset S_1$. We do this in a general way so that we can do prenex quantifier operations with the formulas S_0 and S_1.

Suppose y_j codes a function f with domain the set of $\vec{u}$'s with $|\vec{u}| < i$ and $\ell(\vec{u}) = j$ for $j > 0$. We write $y_j /\!\!/ 0$ for the code of the function g that has as domain the set of strings $u_1 \cdots u_k$ such that $0u_1 \cdots u_k$ is in the domain of f and such that $g(u_1 \cdots u_k) = f(0u_1 \cdots u_k)$. Define $y_j /\!\!/ 1$ similarly. For $\vec{y} = \langle y_1, \ldots, y_{i-1}\rangle$, define t_0 so that

$$t_0(\vec{u}, \langle y_0, \ldots, y_{k-1}, y_k /\!\!/ 0, \ldots, y_{i-1} /\!\!/ 0\rangle, k)$$

is equal to $t(0\vec{u}, \langle y_0, \ldots, y_{i-1}\rangle, w)$ for all $\vec{u}$'s of length $i-1$. (Note that t_0 does not depend on w.) Likewise, define t_1 so that

$$t_1(\vec{u}, \langle y_0, \ldots, y_{k-1}, y_k /\!\!/ 1, \ldots, y_{i-1} /\!\!/ 1\rangle, w, k)$$

is equal to $t(1\vec{u}, \langle y_0, \ldots, y_{i-1}\rangle, w)$ for all $\vec{u}$'s of length $i-1$. Let $S_0(\langle y_0, \ldots y_{i-1}\rangle, k)$ be the formula

$$(\exists \vec{\sigma} \in \{0,1\}^{2^{i-1}})(\mathit{BIT}(\vec{\sigma}) \wedge (\forall \vec{u} \in \{0,1\}^{i-1})[\sigma_{\mathit{nm}(\vec{u})} = 1 \leftrightarrow I(t_0(\vec{u}, \vec{y}, k))]).$$

Let $S_1(\langle y_1, \ldots, y_{i-1}\rangle, w, k)$ be

$$(\exists \vec{\sigma} \in \{0,1\}^{2^{i-1}})(\mathit{BIT}(\vec{\sigma}) \wedge (\forall \vec{u} \in \{0,1\}^{i-1})[\sigma_{\mathit{nm}(\vec{u})} = 1 \leftrightarrow I(t_1(\vec{u}, \vec{y}, w, k))]).$$

Clearly we have that $R(\vec{y},w)$ is equivalent to $S_0(\vec{y},w,i) \supset S_1(\vec{y},w,i)$, and this has a straightforward proof in the theory T. For $k = i, i-1, \ldots, 2, 1$, consider the formulas

$$\begin{aligned}&(\forall y_0)\cdots(\exists y_{k-1})[(\exists y_k)(\forall y_{k+1})\cdots(\exists y_{i-1})S_0(\vec{y},k)\\ &\qquad\qquad\supset (\forall y_k)(\exists y_{k+1})\cdots(\exists y_{i-2})S_1(\vec{y},w,k)],\end{aligned}\tag{2.3}$$

where the notation here assumes k is even and i is odd (and the obvious changes are made when k is odd or i is even). These formulas correspond to the formulas that are obtained as $\varphi_i(w)$ is converted out of prenex form, and into a quantifier pattern that matches that of the righthand side of (2.1). These formulas can be proved equivalent to each other, using proofs of size polynomial in i and using formulas that are no more complicated than the formulas (2.3). The equivalences of the formulas (2.3) are proved straightforwardly by noting which parts of the (functions coded by the) variables y_ℓ are used by S_0 and S_1 and using prenex reasoning. Also, note that S_1 does not depend on y_{i-1}, so the quantifier $\forall y_{i-1}$ has been omitted in front of S_1. (The notation $\vec{y}$ thus variously denotes either $\langle y_0,\ldots,y_{i-2}\rangle$ or $\langle y_0,\ldots,y_{i-1}\rangle$, as appropriate.)

Thus, at $k=1$, the formula

$$(\forall y_0)[(\forall y_1)(\exists y_2)\cdots(\exists y_{i-1})S_0(\vec{y},1) \supset (\exists y_1)(\forall y_2)\cdots(\exists y_{i-2})S_1(\vec{y},w,1)]\tag{2.4}$$

is equivalent to $\varphi_i(w)$. The value y_0 codes a function with domain $1^{<i}$: y_0 can be split into two parts, the first part codes a value y_ϵ and the remaining part codes values for $f(1^j)$ for all $1 \le j < i$. Note that S_0 depends only on the y_ϵ part of y_0. Formula (2.4) is thus equivalent to

$$\begin{aligned}&(\forall y_\epsilon)[(\forall y_1)(\exists y_2)\cdots(\exists y_{i-1})S_0(\langle y_\epsilon,y_1,\ldots,y_{i-1}\rangle,1)\\ &\qquad\qquad\supset (\forall y_0)(\exists y_1)(\forall y_2)\cdots(\exists y_{i-2})S_1(\langle y_\epsilon\cup y_0,y_1,\ldots,y_{i-2}\rangle,w,1)],\end{aligned}$$

where the notation $y_\epsilon \cup y_0$ denotes the number that codes the union of the functions coded by y_ϵ and y_0.

Paying attention to the way that S_1 uses w and the value y_ϵ, and letting $y_\epsilon(x)$ denote the code of the function f with domain $\{\epsilon\}$ such that $f(\epsilon)=x$, the last formula is equivalent to

$$\begin{aligned}&(\forall x)[(\forall y_1)(\exists y_2)\cdots(\exists y_{i-1})S_0(\langle y_\epsilon(x),y_1,\ldots,y_{i-1}\rangle,1)\\ &\qquad\qquad\supset (\forall y_0)(\exists y_1)(\forall y_2)\cdots(\exists y_{i-2})S_1(\langle y_0,y_1,\ldots,y_{i-2}\rangle,w+2^x,0)].\end{aligned}$$

The hypothesis of the implication is equivalent to $\varphi_{i-1}(x)$: to prove this equivalence in T, just prove that the subformulas of the hypothesis are equivalent to the corresponding subformulas of $\varphi_{i-1}(x)$ starting with the quantifier-free part, and working out to the entire formula. Similarly, the conclusion of the implication is equivalent to $\varphi_{i-1}(w+2^x)$.

That completes the sketch of the T-proof that the formula $\varphi_i(w)$ is equivalent to $\forall x(\varphi_{i-1}(x) \supset \varphi_{i-1}(w+2^x))$. This, plus the lower bound on the size of Q_ℓ as established in Section 1, suffices to establish the following theorem.

Theorem 2.1 *There is a constant $c \in \mathbb{N}$ and proofs P_ℓ of depth $\le \ell + c$ and height $O(\ell)$ such that every cut free proof Q_ℓ with the same conclusion as P_ℓ has height at least 2^0_ℓ. Furthermore, the same holds for Q_ℓ containing cuts on only quantifier-free formulas.*

Examination of the proof of Theorem 2.1 reveals that the constant c can equal 6. To see this, note that the formulas S_0 and S_1, like the formula R, have depth equal to 5. Furthermore, the most complex formulas used in the proof P_i, such as formulas (2.3) and (2.4), have depth $i+6$.

3 Lower bounds for relational languages

The superexponential lower bound of Theorem 2.1 was obtained for a language including a number of function symbols, including symbols for exponentiation and superexponentiation. The present section shows that the use of function symbols is entirely unnecessary, and the same lower bound can be obtained for a purely relational language. In prior work, Orevkov already obtained superexponential lower bounds for cut elimination in a purely relational language, but only with $\epsilon \approx \frac{1}{4}$.

The theory T used a finite set of function and relation symbols axiomatized by a set Γ of universal axioms. By standard techniques, the theory T can be converted to a purely relational theory T^{rel} with a $\forall\exists$-axiomatization. For this, each function symbol f of T is replaced by a relation symbol G_f that defines the graph of f; that is, $G_f(\vec{x}, y)$ indicates that $f(\vec{x}) = y$. The set Γ of universal axioms can be replaced by a set of axioms $\Gamma^{\mathrm{rel}} := \Gamma_0 \cup \Gamma_1$ where Γ_0 is a set of universal axioms and Γ_1 contains the $\forall\exists$-statements asserting the totality of the functions. In particular, for each function f, the set Γ_1 contains the formula $(\forall\vec{x})(\exists y)G_f(\vec{x}, y)$. The set Γ^{rel} axiomatizes a theory T^{rel} which is equivalent to T in the sense that models of T and T^{rel} are essentially the same up to the choice of language.

Since no functions symbols are allowed, the set Γ_0 can no longer contain the axiom $(\forall x)(I(x) \supset I(S(x))$. Instead, it now contains the formula

$$(\forall x)(\forall y)(y = S(x) \wedge I(x) \supset I(y)),$$

where "$y = S(x)$" is shorthand notation for a binary relation with parameters x and y.

The construction in the previous section of the proofs P_ℓ can be modified straightforwardly to give proofs of the corresponding statements in the new language. Formulas φ_i^{rel} that express the same condition as φ_i can be defined which still have depth $i + O(1)$ (the constant hidden in the $O(1)$ will be only slightly larger than before). Furthermore, there are proofs of

$$\Gamma^{\mathrm{rel}} \longrightarrow \varphi_k^{\mathrm{rel}}(0)$$

and of

$$\Gamma^{\mathrm{rel}}, \varphi_k^{\mathrm{rel}}(a), b = 2^a \longrightarrow \varphi_{k-1}^{\mathrm{rel}}(b)$$

which have height $O(k)$ and depth $k + O(1)$. Here the formula "$b = 2^a$" does not use the exponential 2^a as a function, but instead is a binary relation on a and b. Combining these proofs for $1 \leq k \leq \ell$, we can form a proof P_ℓ^{rel} of height $O(\ell)$ and depth $\ell + O(1)$ of the sequent

$$\Gamma^{\mathrm{rel}}, a_0 = 2^0, a_1 = 2^{a_0}, a_2 = 2^{a_1}, \ldots, a_\ell = 2^{a_{\ell-1}} \longrightarrow I(a_\ell).$$

Let Q_ℓ^{rel} be a cut free proof of this sequent (or, even a proof in which all cut formulas are quantifier-free). We claim that Q_ℓ^{rel} must have size $\geq 2_\ell^0$. To prove this, we extend the lower bound argument used earlier for Q_ℓ in Section 1. This will involve (a) removing all quantifier inferences in Q_ℓ^{rel} and removing contractions on formulas that (formerly) had quantifiers, and (b) at the same time, assigning an integer value to every free variable in Q_ℓ^{rel}.

Without loss of generality, Q_ℓ^{rel} is in free variable normal form. The only free variables in the endsequent are the variables a_k, and these are assigned the integers 2_{k+1}^0. The proof Q_ℓ^{rel} is then modified iteratively by removing one quantifier inference at a time. At each stage in this process, we will have assigned integer values to all variables that occur below all quantifiers. To remove the next quantifier, choose the lowest remaining quantifier inference. If it is a $\forall$:left inference, just omit the inference, and allow the auxiliary formula in the upper sequent to remain unchanged. In addition, omit all contraction inferences on that formula and its descendants in the proof. On the other hand, suppose the lowest quantifier inference is an $\exists$:left. This will be an inference of the form

$$\frac{G_f(\vec{s}, b), \Pi \longrightarrow \Delta}{(\exists y) G_f(\vec{s}, y), \Pi \longrightarrow \Delta}$$

where $\vec{s}$ is a vector of terms and all variables in the terms in $\vec{s}$ have already been assigned integer values $\vec{n}$. Modify Q_ℓ^{rel} by omitting this $\exists$:left inference and propagating the formula $G_f(\vec{s}, b)$ down to the endsequent in place of $(\exists y) G_f(\vec{s}, y)$. The free variable b is assigned the integer value $f(\vec{n})$ so as to make $G_f(\vec{s}, b)$ true.

Once all the quantifier inferences are removed from Q_ℓ^{rel}, we obtain a proof $Q_\ell^{\mathrm{rel}*}$ in which all formulas are quantifier-free. The number of substitution instances of $y = S(x) \wedge I(x) \supset I(y)$ in the antecedent of the endsequent of $Q_\ell^{\mathrm{rel}*}$ is less than the size $|Q_\ell^{\mathrm{rel}}|$ of Q_ℓ^{rel}. By a similar argument as before, this implies that $|Q_\ell^{\mathrm{rel}}|$ is $\geq 2_\ell^0$. This gives the following lower bound for cut elimination in relational languages.

Theorem 3.1 *Theorem* 2.1 *holds in the purely relational language described above.*

By being careful with the constructions of ϕ_i^{rel}, Theorem 3.1 can be shown to hold with the constant c equal to 8.

Although our lower bounds are very close to optimal, there is still a small gap between the lower bounds of Theorems 2.1 and 3.1 and the known upper bounds discussed in the introduction. Our lower bounds have the form 2_ℓ^0. But, since P_ℓ has height $O(\ell)$ and depth $\ell + O(1)$, the upper bounds of [**16, 25, 26**] on the size of cut free proofs are equal to

$$2_{\ell+O(1)}^{O(\ell)} = 2_{\ell+\log^*(\ell)+O(1)}^0,$$

where $\log^*$ denotes the inverse superexponential function. It is open how to close the $\log^*$ gap between the height of superexponential size upper and lower bounds.

Acknowledgement

Wenhui Zhang provided helpful comments on an earlier draft of this paper. We also thank the anonymous referee for helpful comments and suggestions.

References

[1] Matthias Baaz and Alexander Leitsch. On Skolemizations and proof complexity. *Fundamenta Informaticae*, 20:353–379, 1994.

[2] Matthias Baaz and Alexander Leitsch. Cut normal forms and proof complexity. *Annals of Pure and Applied Logic*, 97:127–177, 1999.

[3] Matthias Baaz and Alexander Leitsch. Fast cut-elimination by CERES. To appear in *Proofs, Categories and Computations*, College Publications, 2010.

[4] Arnold Beckmann and Samuel R. Buss. Corrected upper bounds for free-cut elimination. *Theoretical Computer Science*, 412(39):5433–5445, 2011.

[5] Samuel R. Buss. An introduction to proof theory. In S. R. Buss, editor, *Handbook of Proof Theory*, pages 1–78. North-Holland, 1998.

[6] Samuel R. Buss. Cut elimination in situ. Typeset manuscript, 2011.

[7] Samuel R. Buss and Alan Johnson. The quantifier complexity of polynomial-size iterated definitions in first-order logic. *Mathematical Logic Quarterly*, 56(6):573–590, 2010.

[8] Jeanne Ferrante and Charles W. Rackoff. *The Computational Complexity of Logical Theories.* Lecture Notes in Mathematics #718. Springer Verlag, Berlin, 1979.

[9] Gerhard Gentzen. Beweisbarkeit und Unbeweisbarkeit von Anfangsfällen der transfiniten Induktion in der reinen Zahlentheorie. *Mathematische Annalen*, 119:140–161, 1943. English translation in [**10**], pp. 287–308.

[10] Gerhard Gentzen. *Collected Papers of Gerhard Gentzen.* North-Holland, 1969. Edited by M. E. Szabo.

[11] Philipp Gerhardy. Refined complexity analysis of cut elimination. In *Proc. 17th Workshop on Computer Science Logic (CSL)*, Lecture Notes in Computer Science #2803, pages 212–225. Springer Verlag, 2003.

[12] Philipp Gerhardy. The role of quantifier alternations in cut elimination. *Notre Dame Journal of Formal Logic*, 46(2):165–171, 2005.

[13] Jean-Yves Girard. *Proof Theory and Logical Complexity.* Humanities Press, 1987.

[14] Edward Nelson. *Predicative Arithmetic.* Princeton University Press, 1986.

[15] V. P. Orevkov. Lower bounds for lengthening of proofs after cut-elimination. *Journal of Soviet Mathematics*, 20:2337–2350, 1982. Original Russian version in Zap. Nauchn. Sem. L.O.M.I. Steklov, 88:137–162, 1979.

[16] V. P. Orevkov. Upper bound on the lengthening of proofs by cut elimination. *Journal of Soviet Mathematics*, 34:1810–1819, 1986. Original Russian version in Zap. Nauchn. Sem. L.O.M.I. Steklov, 137:87–98, 1984.

[17] V. P. Orevkov. Applications of cut elimination to obtain estimates of proof lengths. *Soviet Mathematics Doklady*, 36:292–295, 1988. Original Russian version in Dokl. Akad. Nauk., 296(3):539–542, 1987.

[18] J. B. Paris and C. Dimitracopoulos. A note on the undefinability of cuts. *Journal of Symbolic Logic*, 48(3):564–569, 1983.

[19] Pavel Pudlák. The lengths of proofs. In S. R. Buss, editor, *Handbook of Proof Theory*, pages 547–637. Elsevier North-Holland, 1998.

[20] Robert M. Solovay. Letter to P. Hájek, August 1976.

[21] Richard Statman. Lower bounds on Herbrand's theorem. *Proceedings of the American Mathematical Society*, 75(1):104–107, 1979.

[22] Richard Statman. Speed-up by theories with infinite models. *Proceedings of the American Mathematical Society*, 81:465–469, 1981.

[23] Anne S. Troelstra and Helmut Schwichtenberg. *Basic Proof Theory.* Tracts in Theoretical Computer Science #43. Cambridge University Press, Cambridge, 2nd edition, 2000.

[24] A. S. Yessenin-Volpin. The ultra-intuitionistic criticism and the antitraditional program for foundations of mathematics. In A. Kino, J. Myhill, and R. E. Vesley, editors, *Intuitionism and Proof Theory*, pages 1–45. North-Holland, 1970.

[25] Wenhui Zhang. Cut elimination and automatic proof procedures. *Theoretical Computer Science*, 91:265–284, 1991.

[26] Wenhui Zhang. Depth of proofs, depth of cut-formulas, and complexity of cut formulas. *Theoretical Computer Science*, 129:193–206, 1994.

[27] Wenhui Zhang. Structure of proofs and the complexity of cut elimination. *Theoretical Computer Science*, 353:63–70, 2006.

From almost optimal algorithms to logics for complexity classes via listings and a halting problem

Yijia Chen[*], **Jörg Flum**[†]

[*] Department of Computer Science, Shanghai Jiao Tong University, China
yijia.chen@cs.sjtu.edu.cn

[†] Mathematisches Institut, Albert-Ludwigs-Universität Freiburg, Germany
joerg.flum@math.uni-freiburg.de

Abstract. Let C denote one of the complexity classes "polynomial time", "logspace", or "nondeterministic logspace". We introduce a logic $L(C)_{\mathrm{inv}}$ and show generalizations and variants of the equivalence ($L(C)_{\mathrm{inv}}$ captures C if and only if there is an almost C-optimal algorithm in C for the set TAUT of tautologies of propositional logic). These statements are also equivalent to the existence of a listing of subsets in C of TAUT by corresponding Turing machines and equivalent to the fact that a certain parameterized halting problem is in the parameterized complexity class $\mathrm{XC}_{\mathrm{uni}}$.

Introduction

As the title already indicates, this paper relates two topics which at first glance seem to be unrelated. On the one hand we consider almost optimal algorithms. An algorithm, say deciding the class TAUT of tautologies of propositional logic, is *almost optimal* if the time it requires to accept tautologies can be polynomially bounded in terms of the corresponding time of any other algorithm deciding TAUT.[1] In their fundamental paper [**17**] Krajíček and Pudlák not only introduced the notion of almost optimality but they also derived a series of statements equivalent to the existence of an almost optimal algorithm for TAUT, among them the existence of a polynomially optimal propositional proof system. Furthermore, they stated the following conjecture:

Conjecture 1 There is no almost optimal proof algorithm for TAUT.

On the other hand, the question of whether there is a logic capturing the complexity class P (polynomial time) remains the central open problem in descriptive complexity. By a result due to Immerman and Vardi [**12, 23**], least fixed-point logic LFP captures P on *ordered* structures. There are artificial logics capturing P (on arbitrary structures), but they do not fulfill a natural requirement to logics in this context:

(0.1) There is an algorithm which decides whether $\mathcal{A}$ is a model of φ for all structures $\mathcal{A}$ and sentences φ of the logic and which, for fixed φ, has running time polynomial in the size of $\mathcal{A}$.

[*,†] This research has been partially supported by the National Nature Science Foundation of China (60970011), the Sino-German Center for Research Promotion (GZ400), and the John Templeton Foundation through Project # 13152, the Infinity Project at the Centre de Recerca Matemàtica.

[1] All notions will be defined in a precise manner later.

If this condition is fulfilled for a logic capturing polynomial time, we speak of a P-bounded logic for P. In [**10**] Gurevich states the following conjecture:

Conjecture 2 There is no P-bounded logic for P.

The conjecture is false if one waives the effectivity condition (0.1). This is shown in [**10**, Section 7, CLAIM 2] by considering a logic, *the order-invariant least fixed-point logic*, introduced by Blass and Gurevich and which we denote by $\mathrm{LFP}_{\mathrm{inv}}$.[2] For any vocabulary the sentences of $\mathrm{LFP}_{\mathrm{inv}}$ are the sentences of least fixed-point logic LFP in a vocabulary with an additional binary relation symbol for orderings. In $\mathrm{LFP}_{\mathrm{inv}}$ for a structure $\mathcal{A}$ to be a model of φ it is required that in all structures of cardinality less than or equal to that of $\mathcal{A}$, the validity of φ (as a sentence of least fixed-point logic) does not depend on the chosen ordering, and $\mathcal{A}$ with some ordering satisfies φ. As $\mathrm{LFP}_{\mathrm{inv}}$ satisfies all requirements of a P-bounded logic for P except (0.1), Gurevich implicitly states the following conjecture:

Conjecture 2a $\mathrm{LFP}_{\mathrm{inv}}$ is not a P-bounded logic for P.

We show that

$$\text{Conjecture 1 is true} \iff \text{Conjecture 2a is true.} \tag{0.2}$$

In general, the experts tend to believe Conjecture 1, as the existence of an almost optimal algorithm for TAUT would have various consequences which seem to be unlikely (see [**15, 17**]). It is worthwhile to emphasize that we show that Conjecture 1 is equivalent to Conjecture 2a and do not claim its equivalence to Conjecture 2. The situation with Conjecture 2 is quite different; no known consequences of the existence of a P-bounded logic for P seem to be implausible. Moreover, due to results showing that there are logics capturing polynomial time on always larger classes of structures, Grohe [**9**] "mildly leans towards believing" that there is a P-bounded logic for P.

We mentioned that at first glance "almost optimal algorithms for TAUT" and "logics for P" seem to be unrelated topics. However, there are reformulations of Conjecture 1 and Conjecture 2 that are alike. In fact, it is known [**22**] that TAUT has an almost optimal algorithm if and only if there is an effective enumeration or *listing* of all subsets of TAUT that are in P by means of polynomial time Turing machines that decide them. And it is not hard to see that there is a P-bounded logic for P if and only if there is a listing of all polynomial time decidable classes of graphs closed under isomorphism, again a listing in terms of polynomial time Turing machines that decide these classes. In fact the question for a logic for P was stated in this way by Chandra and Harel [**1**] in the context of an analysis of the complexity and expressiveness of query languages. Hence the equivalence (0.2) can be reformulated as follows:

$$\begin{aligned} &\mathrm{LFP}_{\mathrm{inv}} \text{ is a P-bounded logic for P} \iff \\ &\qquad\qquad \text{there is a listing of the subsets of TAUT in P.} \end{aligned} \tag{0.3}$$

And one consequence of (0.2) is:

> *If there is a listing of the subsets of* TAUT *in* P*, then there is a listing of the polynomial time decidable classes of graphs closed under isomorphism.*

[2] In [**10**] the logic is defined in a slightly different form.

The reformulation (0.3) led us to the idea underlying the proof of the implication from left to right: We express the property of a propositional formula α of being a tautology in $\mathrm{LFP}_{\mathrm{inv}}$ by reducing the second-order quantifier in "all assignments satisfy α" to the second-order quantifier "for all orderings" hidden in the definition of the satisfaction relation of $\mathrm{LFP}_{\mathrm{inv}}$; then a listing of the appropriate sentences of $\mathrm{LFP}_{\mathrm{inv}}$ yields a listing of the subsets of TAUT in P.

Later on we realized that one gets a listing of the subsets of TAUT in P if one assumes that there is a listing of its subsets in L (logarithmic space). There are standard logics DTC (deterministic transitive closure logic) and TC (transitive closure logic) capturing, on *ordered* structures, the complexity classes L and NL (nondeterministic logarithmic space), respectively [**13, 14**]. We realized that for their corresponding order-invariant logics $\mathrm{DTC}_{\mathrm{inv}}$ and $\mathrm{TC}_{\mathrm{inv}}$ the analogues of the equivalence (0.3) hold. Therefore, by the result on listings just mentioned, $\mathrm{LFP}_{\mathrm{inv}}$ captures P if $\mathrm{DTC}_{\mathrm{inv}}$ captures L. Note that it is not known whether the existence of a logic capturing P is implied by the existence of a logic capturing L.

A more general notion of listing turned out to be helpful. For complexity classes C and C' we consider listings of the C-subsets of TAUT (that are, subsets of TAUT in C) by means of Turing machines of type C'; we write $\mathrm{List}(C, \textsc{Taut}, C')$ if such a listing exists. Here, C and C' range over the complexity classes L, NL, P, and NP. For the classes P and NP such listings were already considered and put to good use by Sadowski in [**22**]. This more general notion is also meaningful in the context of logics. If we say that a logic is a P-bounded logic for P, the second "P" refers to the classes axiomatizable in the logic and the first one to the polynomial time property expressed in (0.1). This suggests the definition of a *C'-bounded logic for C.* It turns out that these general concepts of listings and logics match. In fact we get for $C \in \{\mathrm{L}, \mathrm{NL}, \mathrm{P}\}$, $C' \in \{\mathrm{L}, \mathrm{NL}, \mathrm{P}, \mathrm{NP}\}$ with $C \subseteq C'$,

$$(0.4) \qquad L(C) \textit{ is a } C'\textit{-bounded logic for } C \iff \mathrm{List}(C, \textsc{Taut}, C').$$

Here $L(C)$ is $\mathrm{DTC}_{\mathrm{inv}}$, $\mathrm{TC}_{\mathrm{inv}}$, and $\mathrm{LFP}_{\mathrm{inv}}$ if C is L, NL, and P, respectively. This relationship between listings and logics is not only fruitful for the side of the logics but also for the side of listings. For example, we get

$$(0.5) \qquad \textit{If } \mathrm{List}(\mathrm{L}, \textsc{Taut}, \mathrm{L})\textit{, then } \mathrm{List}(\mathrm{NL}, \textsc{Taut}, \mathrm{NL}).$$

As shown in [**22**], the property $\mathrm{List}(\mathrm{P}, \textsc{Taut}, \mathrm{P})$ is equivalent to the existence of an almost optimal algorithm for TAUT (by (0.2) and (0.3)) and, as already mentioned, to the existence of a polynomially optimal propositional proof system. We show the analogues of these equivalences for L instead of P and for space optimality instead of time optimality. Moreover, we do this not only for TAUT but for arbitrary problems Q with padding. In particular, by (0.5), we get the following, perhaps surprising relationship between space optimal and time optimal algorithms:

> *Assume Q has padding. If Q has an almost space optimal algorithm, then it has an almost (time) optimal algorithm.*

While listings were the main tool for proving the implication from left to right of (0.2), a halting problem plays the role of a bridge leading to the converse direction. In [**21**] Nash, Remmel, and Vianu have raised the question whether one can *prove* Conjecture 2a. They give a reformulation of this conjecture in terms of the complexity of a halting problem for nondeterministic TMs. This reformulation is best expressed in the terminology of parameterized complexity. We introduce the following parameterized halting problem p-$\textsc{Halt}_{>}$ for nondeterministic Turing machines.

p-HALT$_>$

Instance: A nondeterministic Turing machine $\mathbb{M}$ and $\underbrace{1\ldots1}_{n}$ with $n \in \mathbb{N}$.

Parameter: $|\mathbb{M}|$, the size of $\mathbb{M}$.

Problem: Does every accepting run of $\mathbb{M}$ on the empty input tape take more than n steps?

Then, by [**21**],

$$\text{LFP}_{\text{inv}} \text{ is a P-bounded logic for P} \iff p\text{-HALT}_> \in \text{XP}_{\text{uni}}. \tag{0.6}$$

Here XP$_{\text{uni}}$ denotes a parameterized complexity class (defined in Section 1). As TAUT, the classical problem underlying p-HALT$_>$ is co-NP-complete. Based on this fact, we show that the existence of an almost optimal algorithm for TAUT implies (even is equivalent to) p-HALT$_> \in$ XP$_{\text{uni}}$, and thereby we get a proof of the missing implication of (0.2).

It seems to be hard to get reasonable upper and lower bounds for the complexity of p-HALT$_>$ by showing its (non-)membership in one of the standard classes of parameterized complexity like FPT$_{\text{uni}}$, XP$_{\text{uni}}$, XL$_{\text{uni}}$, XNP$_{\text{uni}}$, FPT, XP, XNP, ... However, for each of these classes there is a natural extension of (0.6) (for the classes FPT$_{\text{uni}}$, XP$_{\text{uni}}$, FPT, and XP they were already derived in [**3**] and [**21**]). These equivalences lead to extensions of the equivalence (0.2) and, more importantly, to interesting notions of strongly and effectively almost optimal algorithms. In [**17**] Krajíček and Pudlák show that an almost optimal algorithm for TAUT exists if NE = E. We can even derive:

> *If* NE = E*, then* TAUT *has an effectively and strongly almost optimal algorithm.*

On the other hand we have:

> *If* P[TC] $\neq$ NP[TC]*, then* TAUT *has no effectively and strongly almost optimal algorithm.*

Here P[TC] $\neq$ NP[TC] is an extension of the hypothesis NP $\neq$ P introduced in [**2**].

Let us close these introductory remarks by presenting explicitly one of the results hidden in the previous exposition that relates the four concepts mentioned in the title.

> *The following are equivalent:*
>
> - TAUT *has an almost space optimal algorithm.*
> - *The logic* DTC$_{\text{inv}}$ *is* L*-bounded for* L.
> - List(L, TAUT, L).
> - p-HALT$_> \in$ XL$_{\text{uni}}$.

The content of the different sections is the following. After recalling some basic notions in Section 1, we turn to a proof of statements relating the existence of almost optimal algorithms for TAUT with the complexity of p-HALT$_>$ in Section 2. In Section 3 we relate this complexity to capturing properties of the different invariant logics. Part of the corresponding proof is postponed to Section 4, which is devoted to so-called slicewise downward monotone parameterized problems. This concept helps us to show that it is as complex to check the order-invariance of first-order sentences as it is to check that of LFP-sentences. In Section 5 we extend the known relationship between almost optimal algorithms, polynomially optimal propositional proof systems, and listings to different variants (strong, effective, logarithmic space) of these concepts. Finally, in Section 6 we prove the relationship (0.4) between logics and listings.

This paper is an extended version of the papers [**4, 5, 6**].

1 Some preliminaries

In this section we fix some notations and recall some basic definitions and concepts from parameterized complexity.

We let $\mathbb{N}[X]$ and $\mathbb{N}_d[X]$, where $d \in \mathbb{N}$, be the set of polynomials and the set of polynomials of degree $\leq d$, respectively, with natural numbers as coefficients. We denote the alphabet $\{0, 1\}$ by Σ and the length of a string $x \in \Sigma^*$ by $|x|$. Let 1^n be the string consisting of n many 1s and let λ denote the empty string. We identify problems with subsets Q of Σ^*. Sometimes statements containing a formulation like "there is a $d \in \mathbb{N}$ such that for all $x \in \Sigma^*$: $\ldots \leq |x|^d$" or containing a term $\log |x|$ can be wrong for $x \in \Sigma^*$ with $|x| \leq 1$. We trust the reader's common sense to interpret such statements reasonably.

We denote by P and L the classes of problems Q such that $x \in Q$ is solvable by a deterministic Turing machine in time polynomial in $|x|$ and in space $O(\log |x|)$, respectively. The corresponding nondeterministic classes are NP and NL. *In this paper, C, C', C_0, etc. will always denote one of the complexity classes* L, P, NL*, and* NP.

A problem $Q \subseteq \Sigma^*$ *has padding* if there is a function $pad\colon \Sigma^* \times \Sigma^* \to \Sigma^*$ computable in logarithmic space having the following properties:

(i) For any $x, y \in \Sigma^*$, $|pad(x, y)| > |x| + |y|$ and $\big(pad(x, y) \in Q \Leftrightarrow x \in Q\big)$.

(ii) There is a logspace algorithm which, given $pad(x, y)$, recovers y.

By $\langle \ldots \rangle$ we denote some standard logspace and linear time computable tupling function with logspace and linear time computable inverses. *In this paper we always assume that Q denotes a decidable and nonempty problem.*

If $\mathbb{A}$ is any (deterministic or nondeterministic) algorithm and $\mathbb{A}$ accepts $x \in \Sigma^*$, then we denote by $t_{\mathbb{A}}(x)$ the number of steps of a shortest accepting run of $\mathbb{A}$ on x; if $\mathbb{A}$ does not accept x, then $t_{\mathbb{A}}(x) := \infty$. By convention, $n < \infty$ for $n \in \mathbb{N}$. Similarly, $s_{\mathbb{A}}(x)$ is the minimum of the space used by accepting runs on x. *By default, algorithms are deterministic.* If an algorithm $\mathbb{A}$ on input x eventually halts and outputs a value, we denote it by $\mathbb{A}(x)$.

We use deterministic and nondeterministic Turing machines with Σ as alphabet as our basic computational model for algorithms (and we often use the notions "algorithm" and TM synonymously). If necessary we will not distinguish between a Turing machine and its code, a string in Σ^*. If $\mathbb{M}$ is a deterministic or nondeterministic TM, then $L(\mathbb{M})$ is the language accepted by $\mathbb{M}$. We use TMs as acceptors and transducers. Even though we use formulations like "let $\mathbb{M}_1, \mathbb{M}_2, \ldots$ be an enumeration of *all* polynomial time TMs", from the context it will be clear that we only refer to acceptors (or that we only refer to transducers). We assume that a run of a nondeterministic Turing machine is determined by the sequence of its states.

1.1 Parameterized complexity

We view *parameterized problems* as pairs (Q, κ) consisting of a classical problem $Q \subseteq \Sigma^*$ and a *parameterization* $\kappa\colon \Sigma^* \to \mathbb{N}$, which is required to be polynomial time computable. We will present parameterized problems in the form we did it for p-HALT$_>$ in the Introduction.

A parameterized problem (Q, κ) is in the class FPT$_{\text{uni}}$ (or *uniformly fixed-parameter tractable*) if $x \in Q$ is solvable by a deterministic algorithm running in time less than or equal to $f(\kappa(x)) \cdot |x|^{O(1)}$ for some $f\colon \mathbb{N} \to \mathbb{N}$.

Let $C \in \{\mathrm{L}, \mathrm{NL}, \mathrm{P}, \mathrm{NP}\}$. A parameterized problem (Q, κ) is in the class $\mathrm{XC}_{\mathrm{uni}}$ if there is a deterministic (nondeterministic) algorithm deciding (accepting) Q and witnessing for every $k \in \mathbb{N}$ that the classical problem

$$(Q,\kappa)_k := \big\{x \in Q \bigm| \kappa(x) = k\big\},$$

the kth *slice* of (Q, κ), is in C. For example, (Q, κ) is in the class $\mathrm{XP}_{\mathrm{uni}}$ if there is a deterministic algorithm $\mathbb{A}$ deciding $x \in Q$ in time $|x|^{f(\kappa(x))}$ for some function $f \colon \mathbb{N} \to \mathbb{N}$. And (Q, κ) is in the class $\mathrm{XNP}_{\mathrm{uni}}$ if there is a nondeterministic algorithm $\mathbb{A}$ accepting Q such that for some function $f \colon \mathbb{N} \to \mathbb{N}$ we have $t_{\mathbb{A}}(x) \le |x|^{f(\kappa(x))}$ for all $x \in Q$.

We have added the subscript "uni" to the names of these classes to emphasize that they are classes of the so-called uniform parameterized complexity theory. If in the definition of $\mathrm{FPT}_{\mathrm{uni}}$, $\mathrm{XP}_{\mathrm{uni}}$, and $\mathrm{XNP}_{\mathrm{uni}}$ we require the function f to be computable, then we get the corresponding classes FPT, XP, and XNP of the strongly uniform theory. Now the interested reader will have no difficulties to define the classes XL and XNL, which we do not use in this paper.

2 Almost optimal algorithms and p-HALT$_>$

In this section we relate the existence of an almost optimal algorithm for the set TAUT of tautologies of propositional logic to the complexity of the parameterized problem p-HALT$_>$.

Recall that $\mathbb{N}_d[X]$ is the set of polynomials of degree $\le d$ and that $t_{\mathbb{A}}(x)$ denotes the minimum of the running times of accepting runs of the algorithm $\mathbb{A}$ on input x.

Definition 2.1 An algorithm $\mathbb{O}$ deciding a problem $Q \subseteq \Sigma^*$ is *almost optimal* if for every algorithm $\mathbb{A}$ deciding Q there is a polynomial $p \in \mathbb{N}[X]$ such that

$$t_{\mathbb{O}}(x) \le p(t_{\mathbb{A}}(x) + |x|) \tag{2.1}$$

for all $x \in Q$. Note that nothing is required for $x \notin Q$. If there is a $d \in \mathbb{N}$ such that the polynomial p can be chosen in $\mathbb{N}_d[X]$ for all $\mathbb{A}$, then $\mathbb{O}$ is *strongly almost optimal*.

We shall need the following fact.

Lemma 2.2 *Let Q be polynomial time reducible to Q' and assume that Q' has padding. If Q' has a (strongly) almost optimal algorithm, then so does Q.*

Proof. As Q' has padding, there is a one-to-one polynomial time reduction $\mathbb{S}$ from Q to Q' with a polynomial time inverse. Let $\mathbb{O}'$ be a (strongly) almost optimal algorithm for Q'. Now it is straightforward to show that the algorithm $\mathbb{O} := \mathbb{O}' \circ \mathbb{S}$ that on input x first computes $\mathbb{S}(x)$ and then simulates $\mathbb{O}'$ on $\mathbb{S}(x)$ is (strongly) almost optimal algorithm for Q. □

We come to the main result of this section.

Theorem 2.3

(a) TAUT *has an almost optimal algorithm* $\iff$ p-HALT$_> \in \mathrm{XP}_{\mathrm{uni}}$.

(b) TAUT *has a strongly almost optimal algorithm* $\iff$ p-HALT$_> \in \mathrm{FPT}_{\mathrm{uni}}$.

Proof. We first prove the directions from left to right in the equivalences. The classical problem HALT$_>$ underlying p-HALT$_>$ is easily seen to be co-NP-complete. As TAUT has padding, for part (a), by the previous lemma, we may assume that HALT$_>$ has an almost optimal algorithm $\mathbb{O}$. Let $\mathbb{B}$ be a "brute force" algorithm that on input $\mathbb{M}$, a nondeterministic TM, by systematically going through all runs of $\mathbb{M}$ on input λ (the empty string) of length 1, of length 2, etc. computes $t_{\mathbb{M}}(\lambda)$, the least k such that there is an accepting run of $\mathbb{M}$ on λ of length k. If $\mathbb{M}$ has no such run, then $\mathbb{B}$ on input $\mathbb{M}$ does not halt.

We show that the following algorithm $\mathbb{O}^*$ simulating $\mathbb{B}$ and $\mathbb{O}$ in parallel witnesses that p-Halt$_>$ $\in \mathrm{XP}_{\mathrm{uni}}$:

```
$\mathbb{O}^*$   // $\mathbb{M}$ a nondeterministic TM, $1^n$ with $n \in \mathbb{N}$
      in parallel simulate $\mathbb{B}$ on $\mathbb{M}$ and $\mathbb{O}$ on $\langle \mathbb{M}, 1^n \rangle$
      **if** $\mathbb{O}$ halts first **then** answer accordingly
      **if** $\mathbb{B}$ halts first **then**
          **if** $n < t_{\mathbb{M}}(\lambda)$ **then** accept **else** reject.
```

Clearly, $\mathbb{O}^*$ decides p-Halt$_>$. We still have to verify that for a fixed nondeterministic Turing machine $\mathbb{M}_0$ the algorithm $\mathbb{O}^*$ on input $\langle \mathbb{M}_0, 1^n \rangle$ runs in time polynomial in n.

Case "$\langle \mathbb{M}_0, 1^\ell \rangle \notin p$-Halt$_>$ for some $\ell \in \mathbb{N}$": Then $\mathbb{B}$ will halt on input $\mathbb{M}_0$. Thus, in the worst case, $\mathbb{O}^*$ on input $\langle \mathbb{M}_0, 1^n \rangle$ has to wait till the simulation of $\mathbb{B}$ on $\mathbb{M}_0$ halts and then $\mathbb{O}^*$ must check whether the output $t_{\mathbb{M}_0}(\lambda)$ of the computation of $\mathbb{B}$ is bigger than n or not and must answer accordingly. So in the worst case $\mathbb{O}^*$ takes time $O(t_{\mathbb{B}}(\mathbb{M}_0) + n)$.

Case "$\langle \mathbb{M}_0, 1^\ell \rangle \in p$-Halt$_>$ for all $\ell \in \mathbb{N}$": In this case $\mathbb{B}$ on input $\mathbb{M}_0$ does not halt. Hence, the running time of $\mathbb{O}^*$ on input $\langle \mathbb{M}_0, 1^n \rangle$ is determined by that of $\mathbb{O}$ on this input. Here we will argue with the almost optimality of $\mathbb{O}$ and for this purpose we consider the algorithm $\mathbb{A}(\mathbb{M}_0)$ that on input $\langle \mathbb{M}, 1^n \rangle$ accepts if $\mathbb{M} = \mathbb{M}_0$; if $\mathbb{M} \neq \mathbb{M}_0$, it simulates $\mathbb{O}$ on input $\langle \mathbb{M}, 1^n \rangle$ and answers accordingly. Clearly, $\mathbb{A}(\mathbb{M}_0)$ decides Halt$_>$ and

$$t_{\mathbb{A}(\mathbb{M}_0)}\big(\langle \mathbb{M}_0, 1^n \rangle \big) \le O(|\mathbb{M}_0| + n) \tag{2.2}$$

for all $n \in \mathbb{N}$. By the almost optimality of $\mathbb{O}$ there is a polynomial $p \in \mathbb{N}[X]$ (depending on $\mathbb{A}(\mathbb{M}_0)$ and hence on $\mathbb{M}_0$) such that, for all n,

$$t_{\mathbb{O}}(\langle \mathbb{M}_0, 1^n \rangle) \le p\Big(t_{\mathbb{A}(\mathbb{M}_0)}(\langle \mathbb{M}_0, 1^n \rangle) + |\mathbb{M}_0| + n \Big). \tag{2.3}$$

Therefore, by (2.2), the algorithm $\mathbb{O}$ runs in time polynomial in n on inputs of the form $\langle \mathbb{M}_0, 1^n \rangle$.

If the algorithm $\mathbb{O}$ is strongly almost optimal, there is a $d \in \mathbb{N}$ such that in the previous argument for all nondeterministic TMs $\mathbb{M}_0$ the polynomial p can be chosen in $\mathbb{N}_d[X]$. Then, (2.2) and (2.3) show that

$$t_{\mathbb{O}}(\langle \mathbb{M}_0, 1^n \rangle) \le f(|\mathbb{M}_0|) \cdot n^d$$

for all instances $\langle \mathbb{M}_0, 1^n \rangle$ of p-Halt$_>$ and some function $f \colon \mathbb{N} \to \mathbb{N}$, that is, p-Halt$_>$ $\in$ FPT$_{\mathrm{uni}}$.

We turn to a proof of the implication from right to left in (b) (that of (a) is obtained by the obvious modifications). Assume that $\mathbb{B}$ is a TM deciding whether $\langle \mathbb{M}, 1^n \rangle \in p$-Halt$_>$ in time

$$f(|\mathbb{M}|) \cdot n^d \tag{2.4}$$

for some $d \in \mathbb{N}$ and $f \colon \mathbb{N} \to \mathbb{N}$. The idea underlying the strongly almost optimal algorithm $\mathbb{O}$ we aim at is simple: Using an effective enumeration of all TMs, for a given input string x the algorithm $\mathbb{O}$ in a diagonal fashion performs steps of the machines in the enumeration on input x; if $\mathbb{A}$, one of these machines, accepts x, then $\mathbb{O}$ using the algorithm $\mathbb{B}$ checks whether (essentially) all inputs accepted by $\mathbb{A}$ in $\leq t_{\mathbb{A}}(x)$ steps are tautologies; if so, $\mathbb{O}$ accepts x.

We come to the details. For a deterministic TM $\mathbb{A}$ we introduce machines $\mathbb{A}'$ and $\mathbb{A}''$, the first one being nondeterministic and the second deterministic. For the machine $\mathbb{A}'$ and a suitable quadratic polynomial $p_0 \in \mathbb{N}_2[X]$ we have:

(i) $L(\mathbb{A}) \subseteq \textsc{Taut} \Leftrightarrow t_{\mathbb{A}'}(\lambda) = \infty$;
(ii) If $\mathbb{A}$ accepts a string x which is not a tautology, then $t_{\mathbb{A}'}(\lambda) \leq t_{\mathbb{A}}(x) + p_0(|x|)$.

```
$\mathbb{A}'$
    guess a string $x \in \Sigma^*$
    simulate $\mathbb{A}$ on input $x$
    if $\mathbb{A}$ accepts then
        if $x$ is not a propositional formula then accept else
            guess a valuation $v$ for $x$
            if $v$ does not satisfy $x$ then accept.
```

For the algorithm $\mathbb{A}''$ and every $x \in \Sigma^*$ we have:

(iii) $\mathbb{A}''$ accepts $x \Leftrightarrow \big(\mathbb{A}$ accepts x and $t_{\mathbb{A}'}(\lambda) > t_{\mathbb{A}}(x) + p_0(|x|)\big)$.

```
$\mathbb{A}''$    // $x \in \Sigma^*$
    simulate $\mathbb{A}$ on input $x$ thereby counting the number $t_{\mathbb{A}}(x)$ of
      steps
    if $\mathbb{A}$ rejects then reject
    $u \leftarrow t_{\mathbb{A}}(x) + p_0(|x|)$
    simulate $\mathbb{B}$ on input $\langle \mathbb{A}', 1^u \rangle$    // (it is checked whether
      $\langle \mathbb{A}', 1^u \rangle \in p$-Halt$_>$)
        if $\mathbb{B}$ accepts then accept else reject.
```

Thus, we have:

(iv) $L(A'') \subseteq \textsc{Taut}$ (by (ii) and (iii));
(v) if $L(A) \subseteq \textsc{Taut}$, then $L(A'') = L(A)$ (by (i) and (iii)).

If $\mathbb{A}''$ accepts x, then

$$\begin{aligned} t_{\mathbb{A}''}(x) &\leq O\Big(t_{\mathbb{A}}(x) + |x|^2 + t_{\mathbb{B}}(\big\langle \mathbb{A}', t_{\mathbb{A}}(x) + p_0(|x|)\big\rangle)\Big) \\ &\leq O\Big(t_{\mathbb{A}}(x) + |x|^2 + f(|\mathbb{A}'|) \cdot (t_{\mathbb{A}}(x) + p_0(|x|))^d\Big) \qquad \big(\text{by } (2.4)\big). \end{aligned}$$

As $p_0 \in \mathbb{N}_2[X]$, there exists a $p \in \mathbb{N}_{2d}[X]$ such that

$$t_{\mathbb{A}''}(x) \leq p(t_{\mathbb{A}}(x) + |x|). \tag{2.5}$$

Let $\mathbb{T}$ be any algorithm deciding $\textsc{Taut}$ and $\mathbb{L}$ be an algorithm listing all Turing machines, that is, the algorithm $\mathbb{L}$, once having been started, eventually prints out exactly all TMs. For $i \in \mathbb{N}$ we denote by $\mathbb{A}_i$ the last machine printed out by $\mathbb{L}$ in i steps; in particular, $\mathbb{A}_i$ is undefined if $\mathbb{L}$ has not printed any algorithm in i steps. We define an algorithm $\mathbb{O}$ deciding $\textsc{Taut}$:

```
$\mathbb{O}$   // $x \in \Sigma^*$
    1. $\ell \leftarrow 1$
    2. simulate the $\ell$th step of $\mathbb{T}$ on input $x$
    3. **if** $\mathbb{T}$ halts **then** answer accordingly
    4. simulate the $\ell$th step of $\mathbb{L}$
    5. **for** $i = 0$ **to** $\ell$
    6.     **if** $\mathbb{A}_i$ is defined **then** simulate the $(\ell - i)$th step of $\mathbb{A}''_i$ on $x$
    7.         **if** $\mathbb{A}''_i$ accepts **then** accept
    8. $\ell \leftarrow \ell + 1$
    9. goto 2.
```

By (iv), it should be clear that $\mathbb{O}$ decides Taut. We show that $\mathbb{O}$ is strongly almost optimal. Let $\mathbb{A}$ be any TM deciding Taut. We choose i_0 such that $\mathbb{A} = \mathbb{A}_{i_0}$. By (v), the algorithm $\mathbb{A}''_{i_0}$ decides Taut, too. Therefore, $\mathbb{O}$ on input $\alpha \in$ Taut accepts α in Line 7 for $\ell := i_0 + t_{\mathbb{A}''_{i_0}}(x)$ if it was not already accepted earlier. Thus,

$$t_{\mathbb{O}}(\alpha) = \big(i_0 + t_{\mathbb{A}''_{i_0}}(\alpha)\big)^{O(1)}$$

where the constant hidden in $O(1)$ does not depend on the algorithm $\mathbb{A}$. Thus by (2.5), there exists a constant $d' \in \mathbb{N}$ such that for all TMs $\mathbb{A}$ deciding Taut there is a $p' \in \mathbb{N}_{d'}[X]$ such that for $\alpha \in$ Taut

$$t_{\mathbb{O}}(\alpha) \leq p'\big(t_{\mathbb{A}}(\alpha) + |\alpha|\big).$$

This shows that $\mathbb{O}$ is strongly almost optimal. □

Remark 2.4 For later reference we remark that in the proof of the implication from right to left we used the time bound (2.4) of the algorithm $\mathbb{B}$ only for inputs of the form $\langle \mathbb{A}'_{i_0}, \ldots \rangle$. As $L(\mathbb{A}_{i_0}) =$ Taut, by (i) we have $t_{\mathbb{A}'_{i_0}}(\lambda) = \infty$.

2.1 Variants of Theorem 2.3

There exist various variants and extensions of Theorem 2.3. In this subsection we derive a nondeterministic version, a space version, and an effective one. We leave it to the reader to combine the variants in order to get further insights.

The nondeterministic variant. The notion of almost optimal nondeterministic algorithm was introduced by Sadowski in [**22**]; we recall it, thereby introducing a strong version, too.

Definition 2.5 A nondeterministic algorithm $\mathbb{O}$ accepting a problem $Q \subseteq \Sigma^*$ is *almost optimal* if for every nondeterministic algorithm $\mathbb{A}$ accepting Q there is a polynomial $p \in \mathbb{N}[X]$ such that

$$t_{\mathbb{O}}(x) \leq p\big(t_{\mathbb{A}}(x) + |x|\big)$$

for all $x \in Q$. If there is a $d \in \mathbb{N}$ such that for all $\mathbb{A}$ the polynomial p can be chosen in $\mathbb{N}_d[X]$, then $\mathbb{O}$ is *strongly almost optimal*.

Recall that para-NP$_{\text{uni}}$ is the class of parameterized problems (Q, κ) accepted by a nondeterministic algorithm $\mathbb{A}$ such that $t_{\mathbb{A}}(x) \leq f(\kappa(x)) \cdot |x|^{O(1)}$ for all $x \in Q$ and some function $f \colon \mathbb{N} \to \mathbb{N}$. The following nondeterministic version of Theorem 2.3 is proven in the same way.

Theorem 2.6

(a) TAUT *has an almost optimal nondeterministic algorithm* $\iff$ p-HALT$_>$ $\in$ XNP$_{\text{uni}}$.

(b) TAUT *has a strongly almost optimal nondeterministic algorithm* $\iff$ p-HALT$_>$ $\in$ para-NP$_{\text{uni}}$.

As XP$_{\text{uni}} \subseteq$ XNP$_{\text{uni}}$ and FPT$_{\text{uni}} \subseteq$ para-NP$_{\text{uni}}$, we get from Theorem 2.3 and Theorem 2.6 the following corollary, whose part (a) was already proven in [**22**]:

Corollary 2.7

(a) *If* TAUT *has an almost optimal algorithm, then it also has an almost optimal nondeterministic algorithm.*

(b) *If* TAUT *has a strongly almost optimal algorithm, then it also has a strongly almost optimal nondeterministic algorithm.*

The space variant. Recall that $s_{\mathbb{A}}(x)$ denotes the minimum space used by an accepting run of the nondeterministic algorithm $\mathbb{A}$ on input x.

Definition 2.8 A deterministic (nondeterministic) algorithm $\mathbb{O}$ deciding Q is *almost space optimal* for Q if for every deterministic (nondeterministic) algorithm $\mathbb{A}$ which decides (accepts) Q there is a $d \in \mathbb{N}$ such that, for all $x \in Q$,

$$s_{\mathbb{A}}(x) \leq d \cdot \big(s_{\mathbb{B}}(x) + \log|x|\big).$$

We say that an algorithm $\mathbb{A}$ is *loop-free* if $t_{\mathbb{A}}(x) < \infty$ implies $s_{\mathbb{A}}(x) < \infty$ for all $x \in \Sigma^*$. For an algorithm $\mathbb{C}$ the loop-free algorithm $\mathbb{C}'$, claimed to exist in the following well-known lemma, on input x simulates $\mathbb{C}$ on x and thereby records the space and the time $\mathbb{C}$ uses; once the time exceeds the number of configurations using the corresponding space, $\mathbb{C}'$ rejects.

Lemma 2.9 *One can effectively assign to every algorithm $\mathbb{C}$ a loop-free algorithm $\mathbb{C}'$ such that:*

- $L(\mathbb{C}) = L(\mathbb{C}')$;
- *for all* $x \in \Sigma^*$, $(s_{\mathbb{C}}(x) < \infty \Leftrightarrow s_{\mathbb{C}'}(x) < \infty)$;
- *for all* $x \in \Sigma^*$ *with* $s_{\mathbb{C}}(x) < \infty$ *we have* $s_{\mathbb{C}}(x) \leq s_{\mathbb{C}'}(x) \leq O(s_{\mathbb{C}}(x) + \log|x|)$.

Theorem 2.10

(a) TAUT *has an almost space optimal algorithm* $\iff$ p-HALT$_>$ $\in$ XL$_{\text{uni}}$.

(b) TAUT *has an almost space optimal nondeterministic algorithm* $\iff$ p-HALT$_>$ $\in$ XNL$_{\text{uni}}$.

Proof. Note first that the space variant of Lemma 2.2 holds if we assume that Q is logspace reducible to Q'. Then one shows the direction from left to right of the statements (a) and (b) along the lines of the corresponding proof of Theorem 2.3.

We sketch the main changes needed for the implications from right to left. For a TM $\mathbb{A}$ we define $\mathbb{A}'$ and $\mathbb{A}''$ as there. For part (a), again let $\mathbb{T}$ be an algorithm deciding TAUT and $\mathbb{L}$ be a listing of all loop-free TMs. The algorithm $\mathbb{O}$ deciding TAUT is defined by:

```
$\mathbb{O}$   // $x \in \Sigma^*$
    1. $\ell \leftarrow 1$
    2. simulate the $\ell$th step of $\mathbb{T}$ on input $x$
    3. if $\mathbb{T}$ halts then answer accordingly
    4. simulate $\mathbb{L}$ using space at most $\ell \cdot \log |x|$
    5.      if $\mathbb{L}$ prints a machine $\mathbb{A}$ then simulate $\mathbb{A}''$ on $x$ using
            space at most $\ell \cdot \log |x|$
    6.           if $\mathbb{A}''$ accepts then accept
    7. $\ell \leftarrow \ell + 1$
    8. goto 2.
```

Using a listing of the *loop-free* machines ensures that the simulation of $\mathbb{A}''$ in Line 5 eventually stops as $\mathbb{A}''$ eventually will exceed space $\ell \cdot \log |x|$ if it does not stop before. One then shows that $\mathbb{O}$ is almost space optimal.

For part (b) one uses a listing $\mathbb{L}$ of all nondeterministic Turing machines and defines the nondeterministic algorithm $\mathbb{O}$ deciding TAUT by:

```
$\mathbb{O}$   // $x \in \Sigma^*$
    1. guess $i \geq 1$
    2. simulate $\mathbb{L}$ till it outputs the $i$th machine, say, $\mathbb{A}$
    3. simulate $\mathbb{A}''$ on $x$
    4. if $\mathbb{A}''$ accepts then accept.
```

□

As $\mathrm{XL}_{\mathrm{uni}} \subseteq \mathrm{XP}_{\mathrm{uni}}$ and $\mathrm{XNL}_{\mathrm{uni}} \subseteq \mathrm{XNP}_{\mathrm{uni}}$, we obtain from the previous result the following corollary by Theorem 2.3 (a) and Theorem 2.6 (a).

Corollary 2.11

(a) *If* TAUT *has an almost space optimal algorithm, then it has an almost (time) optimal algorithm.*
(b) *If* TAUT *has an almost space optimal nondeterministic algorithm, then it has an almost (time) optimal nondeterministic algorithm.*

The effective variant. In [**3**] we have shown:

Proposition 2.12

(1) *If* $\mathrm{P}[\mathrm{TC}] \neq \mathrm{NP}[\mathrm{TC}]$, *then* p-HALT$_>$ $\notin$ FPT.
(2) *If* $\mathrm{NP}[\mathrm{TC}] \not\subseteq \mathrm{P}[\mathrm{TC}^{\log \mathrm{TC}}]$, *then* p-HALT$_>$ $\notin$ XP.

Here $\mathrm{P}[\mathrm{TC}] \neq \mathrm{NP}[\mathrm{TC}]$ means that for all *time constructible* and increasing functions h the class of problems decidable in *nondeterministic polynomial time* in h and the class of problems decidable in *deterministic polynomial time* in h are distinct, that is, $\mathrm{NDTIME}(h^{O(1)}) \neq \mathrm{DTIME}(h^{O(1)})$. By taking as h the identity function on $\mathbb{N}$ and the function $n \mapsto 2^n$ we see that $\mathrm{P}[\mathrm{TC}] \neq \mathrm{NP}[\mathrm{TC}]$ implies $\mathrm{NP} \neq \mathrm{P}$ and $\mathrm{NE} \neq \mathrm{E}$, respectively. And $\mathrm{NP}[\mathrm{TC}] \not\subseteq \mathrm{P}[\mathrm{TC}^{\log \mathrm{TC}}]$ means that $\mathrm{NTIME}(h^{O(1)}) \not\subseteq \mathrm{DTIME}(h^{O(\log h)})$ for every time constructible and increasing function h. In [**2**] we have related the assumptions $\mathrm{P}[\mathrm{TC}] \neq \mathrm{NP}[\mathrm{TC}]$ and $\mathrm{NP}[\mathrm{TC}] \not\subseteq \mathrm{P}[\mathrm{TC}^{\log \mathrm{TC}}]$ to the so-called *Measure Hypothesis.*

What is the *natural* effective version of almost optimal algorithm for TAUT and is it equivalent to the statement p-HALT$_>$ $\in$ XP? If in the definition of almost optimal algorithm (Definition 2.1) the polynomial p in (2.1) can be computed from $\mathbb{A}$, then the algorithm $\mathbb{A}$ is said to be effectively almost optimal. More precisely:

Definition 2.13 An algorithm $\mathbb{O}$ deciding Q is *effectively (and strongly) almost optimal* if there is a computable function $p\colon \Sigma^* \to \mathbb{N}[X]$ such that, for every deterministic algorithm $\mathbb{A}$ deciding Q,

$$t_{\mathbb{O}}(x) \le p(\mathbb{A})\big(t_{\mathbb{B}}(x) + |x|\big)$$

for all $x \in Q$ (and there is a $d \in \mathbb{N}$ such that even $p\colon \Sigma^* \to \mathbb{N}_d[X]$).

We could not show that the existence of an effectively almost optimal algorithm for TAUT is equivalent to p-HALT$_>$ $\in$ XP; however by analyzing the proof of Theorem 2.3, we isolate a property of p-HALT$_>$ $\in$ XP equivalent to the existence of such an algorithm for TAUT.

We say that p-HALT$_>$ is FPT-*decidable* (XP-*decidable*) *on non-halting machines* if there is an algorithm $\mathbb{A}$ deciding p-HALT$_>$, a computable function $f\colon \mathbb{N} \to \mathbb{N}$, and $c \in \mathbb{N}$ such that for all nondeterministic TMs $\mathbb{M}$ with $t_{\mathbb{M}}(\lambda) = \infty$ we have, for all $n \in \mathbb{N}$,

$$t_{\mathbb{A}}(\langle \mathbb{M}, n\rangle) \le f(|\mathbb{M}|) \cdot n^c \qquad \Big(t_{\mathbb{A}}(\langle \mathbb{M}, n\rangle) \le n^{f(|\mathbb{M}|)}\Big).$$

Lemma 2.14

(a) *If* p-HALT$_>$ $\in$ FPT (p-HALT$_>$ $\in$ XP), *then* p-HALT$_>$ *is* FPT-*decidable* (XP-*decidable*) *on non-halting machines.*

(b) *If* p-HALT$_>$ *is* FPT-*decidable* (XP-*decidable*) *on non-halting machines, then* p-HALT$_>$ $\in$ FPT$_{\text{uni}}$ (p-HALT$_>$ $\in$ XP$_{\text{uni}}$).

Proof. Part (a) is clear by the definitions. For part (b) we combine an algorithm witnessing that p-HALT$_>$ is FPT-decidable (XP-decidable) on non-halting machines with an algorithm that by “brute-force” on input $\langle \mathbb{M}, 1^n\rangle$ computes $t_{\mathbb{M}}(\lambda)$. □

The effective variant of Theorem 2.3 reads as follows:

Proposition 2.15

(a) TAUT *has an effectively almost optimal algorithm* $\iff$ p-HALT$_>$ *is* XP-*decidable on non-halting machines.*

(b) TAUT *has an effectively and strongly almost optimal algorithm* $\iff$ p-HALT$_>$ *is* FPT-*decidable on non-halting machines.*

Proof. The equivalences are obtained as that of Theorem 2.3; for the implication from right to left we use Remark 2.4. □

Corollary 2.16

(1) *If* P[TC] $\neq$ NP[TC], *then* TAUT *has no effectively and strongly almost optimal algorithm.*

(2) *If* NP[TC] $\not\subseteq$ P[TC$^{\log \text{TC}}$], *then* TAUT *has no effectively almost optimal algorithm.*

Proof. A slight modification of the proof of Proposition 2.12 in **[3]** yields that p-HALT$_>$ is not FPT-decidable on non-halting machines if P[TC] $\neq$ NP[TC] holds. Now the claims follow from Proposition 2.15. The proof of (2) is similar. □

On the other hand we have:

Corollary 2.17 *If* NE $=$ E*, then* Taut *has an effectively and strongly almost optimal algorithm.*

Proof. We consider the variant of $\text{Halt}_>$, where instead of the string 1^n we have n in binary as part of the input. Of course, this problem is in NE and hence, by assumption, it is solvable by a deterministic machine in time $2^{c\cdot(\log n+|\mathbb{M}|)}$ for some $c \in \mathbb{N}$. Thus, p-$\text{Halt}_> \in \text{FPT}$ and therefore the claim follows from Proposition 2.15. □

3 Invariant logics and p-$\text{Halt}_>$

For $C \in \{\text{L}, \text{NL}, \text{P}\}$ we introduce a logic $L(C)_{\text{inv}}$. We state the main result of this section, even though we have not introduced some of the concepts so far.

Theorem 3.1 *Let* $C \in \{\text{L}, \text{NL}, \text{P}\}$*,* $C' \in \{\text{L}, \text{NL}, \text{P}, \text{NP}\}$*, and* $C \subseteq C'$*. Then*

$$L(C)_{\text{inv}} \textit{ is a } C'\textit{-bounded logic for } C \iff p\text{-}\text{Halt}_> \in \text{XC}'_{\text{uni}}.$$

We start by recalling the concepts from logic we need.

Structures. A *vocabulary* τ is a finite set of relation symbols. Each relation symbol has an *arity.* A *structure* $\mathcal{A}$ of vocabulary τ, or τ*-structure* (or, simply structure), consists of a nonempty set A called the *universe*, and an interpretation $R^{\mathcal{A}} \subseteq A^r$ of each r-ary relation symbol $R \in \tau$. *All structures in this paper are assumed to have finite universe.* Where necessary we identify structures with strings in Σ^* in a natural way.

Logics capturing complexity classes. For our purposes a *logic* L consists

- for every vocabulary τ of a set $L[\tau]$ of strings, the set of L*-sentences of vocabulary* τ, and of an algorithm that for every τ and every string ξ decides whether $\xi \in L[\tau]$ (in particular, $L[\tau]$ is decidable for every τ);
- of a *satisfaction relation* $\models_L$; if $(\mathcal{A}, \varphi) \in \models_L$ $\big($written: $\mathcal{A} \models_L \varphi\big)$, then $\mathcal{A}$ is a τ-structure and $\varphi \in L[\tau]$ for some vocabulary τ; furthermore for each $\varphi \in L[\tau]$ the class
$$\text{Mod}_L(\varphi) := \{\mathcal{A} \mid \mathcal{A} \models_L \varphi\}$$
of *models of* φ is closed under isomorphism.

Further on, if we say "let φ be an L-sentence", we mean that, in addition to φ, a vocabulary τ with $\varphi \in L[\tau]$ is given.

Recall that C and C' range over the complexity classes L, P, NL, and NP.

Definition 3.2 Let L be a logic and C a complexity class.

(a) *L is a logic for C* if for all vocabularies τ and all classes S of τ-structures closed under isomorphism we have
$$S \in C \iff S = \text{Mod}_L(\varphi) \text{ for some } \varphi \in L[\tau].$$

(b) Let C' be a deterministic (nondeterministic) complexity class. *L is a C'-bounded logic for C* if L is a logic for C and if there is a deterministic (nondeterministic) algorithm $\mathbb{A}$ deciding (accepting) $\models_L$, which for fixed φ witnesses that $\text{Mod}_L(\varphi) \in C'$.

Clearly, if L is a C'-bounded logic for C, then $C \subseteq C'$. If $C' = C$, then property (b) yields the implication from right to left in part (a). One expects from a logic L capturing the complexity class C that it is C'-bounded for C with $C' = C$. In fact, one expects that L can be viewed as a higher programming language for C, that is, that for a fixed L-sentence φ the algorithm $\mathbb{A}$ in (b) witnesses that $\mathrm{Mod}_L(\varphi) \in C$. However we use this more liberal, a little bit artificial notion as in this way we obtain some nontrivial consequences from our results. Fagin [**8**] has shown that there exist NP-bounded logics for NP (e.g., the logic consisting of the Σ^1_1-sentences of second-order logic), while for $C \in \{\mathrm{L}, \mathrm{NL}, \mathrm{P}\}$ it is open whether there is a C-bounded logic for C.

We introduce the notion of an effectively C'-bounded logic for C only for $C' = C = \mathrm{P}$, as we do not use it for other complexity classes. If L is a P-bounded logic for P, then for every L-sentence φ the algorithm $\mathbb{A}$ of Definition 3.2 (b) witnesses that $\mathrm{Mod}_L(\varphi) \in \mathrm{P}$. However, we do not necessarily know ahead of time the bounding polynomial. The logic L *is an effectively* P*-bounded logic for* P if L is a P-bounded logic for P and if in addition to the algorithm $\mathbb{A}$ as in Definition 3.2 (b) there is a computable function that assigns to every L-sentence φ a polynomial $q \in \mathbb{N}[X]$ such that $\mathbb{A}$ decides whether $\mathcal{A} \models_L \varphi$ in $\leq q(|\mathcal{A}|)$ steps.

For every vocabulary τ we let $\tau_< := \tau \cup \{<\}$, where $<$ is a binary relation symbol not in τ chosen in some canonical way. A $\tau_<$-structure $\mathcal{A}$ is *ordered* if $<^{\mathcal{A}}$ is a (total and linear) ordering of the universe A of $\mathcal{A}$.

We say that a logic L is a *C'-bounded logic for C on ordered structures* if (a) and (b) of Definition 3.2 hold for *ordered* structures and classes of *ordered* structures. In (b), for fixed $\varphi \in L[\tau_<]$ the algorithm $\mathbb{A}$ must witness that the class of ordered models of φ is in C'. It should be clear what we mean by an *effectively* P*-bounded logic for* P *on ordered structures.*

We denote by FO, DTC, TC, and LFP *first-order logic, deterministic transitive closure logic, transitive closure logic*, and *least fixed-point logic*, respectively. We assume familiarity with FO. Concerning DTC, TC, and LFP, essentially we only need the following properties:

Theorem 3.3 ([**12, 13, 14, 23**]) *On ordered structures,* DTC *is an* L*-bounded logic for* L*,* TC *is an* NL*-bounded logic for* NL*, and* LFP *is an effectively* P*-bounded logic for* P*.*

For $C = \mathrm{L}$, $C = \mathrm{NL}$, and $C = \mathrm{P}$ we let $L(C)$ be the logic DTC, TC, and LFP, respectively.

Invariant sentences and the logic L_{inv}. We start by introducing the notion of (order-)invariance.

Definition 3.4 Let L be a logic.

(1) Let φ be an $L[\tau_<]$-sentence and $n \geq 1$. We say that φ is *$\leq n$-invariant* if for all τ-structures $\mathcal{A}$ with $|A| \leq n$ and all orderings $<_1$ and $<_2$ on A we have

$$(\mathcal{A}, <_1) \models_L \varphi \iff (\mathcal{A}, <_2) \models_L \varphi.$$

(2) The *parameterized L-invariant problem* is the problem

p-L-Inv	
Instance:	An L-sentence φ and 1^n with $n \geq 1$.
Parameter:	$\lvert\varphi\rvert$.
Problem:	Is $\varphi \leq n$-invariant?

For better readability we will often write $\langle \varphi, n \rangle \in p$-$L$-Inv instead of $\langle \varphi, 1^n \rangle \in p$-$L$-Inv.

We define the logic L_{inv}. For every vocabulary τ we set

$$L_{\mathrm{inv}}[\tau] := L[\tau_<],$$

and we define the satisfaction relation by

(3.1)
$$\mathcal{A} \models_{L_{\mathrm{inv}}} \varphi \iff \Big(\langle \varphi, |A|\rangle \in p\text{-}L\text{-}\textsc{Inv} \text{ and } (\mathcal{A}, <) \models_L \varphi \text{ for some ordering } < \text{ on } A\Big).$$

Assume that for every L-sentence φ, say, of vocabulary τ the string $\neg\varphi$ is an $L[\tau]$-sentence such that $\textsc{Mod}(\neg\varphi) = \big\{\mathcal{A} \bigm| \mathcal{A} \text{ a } \tau\text{-structure and } \mathcal{A} \notin \textsc{Mod}(\varphi)\big\}$. As $\langle \varphi, n\rangle \in p\text{-}L\text{-}\textsc{Inv}$ if and only if $\langle \neg\varphi, n\rangle \in p\text{-}L\text{-}\textsc{Inv}$, we get, for every structure $\mathcal{A}$,

$$\langle \varphi, |A|\rangle \in p\text{-}L\text{-}\textsc{Inv} \iff \big(\mathcal{A} \models_{L_{\mathrm{inv}}} \varphi \text{ or } \mathcal{A} \models_{L_{\mathrm{inv}}} \neg\varphi\big). \tag{3.2}$$

Now that we have introduced all concepts needed to understand the statement of Theorem 3.1, we start with a series of lemmas that finally will yield a proof of it.

Lemma 3.5 *If* $C \in \{\mathrm{L}, \mathrm{NL}, \mathrm{P}\}$*, then* $L(C)_{\mathrm{inv}}$ *is a logic for* C*.*

Proof. For a class S of τ-structures let $S_<$ be the following class of $\tau_<$-structures:

$$S_< := \big\{(\mathcal{A}, <) \bigm| \mathcal{A} \in S \text{ and } < \text{ is an ordering on } A\big\}.$$

Clearly, $\big(S \in C \iff S_< \in C\big)$. Now, if $S \in C$, then, by Theorem 3.3, there is an $L(C)[\tau_<]$-sentence φ such that $S_< = \textsc{Mod}_{L(C)}(\varphi)$. Hence, $S = \textsc{Mod}_{L(C)_{\mathrm{inv}}}(\varphi)$.

Conversely, let φ be an $L(C)_{\mathrm{inv}}[\tau]$-sentence. If φ is not $\le n$-invariant for some $n \ge 1$, then $\textsc{Mod}_{L(C)_{\mathrm{inv}}}(\varphi)$ contains only structures with universes of less than n elements and thus only finitely many up to isomorphism; hence it is in C. Otherwise, $\textsc{Mod}_{L(C)_{\mathrm{inv}}}(\varphi)_< = \textsc{Mod}_{L(C)}(\varphi)$. As the latter class is in C (by Theorem 3.3), so is the former. □

Lemma 3.6 *Let* $C \in \{\mathrm{L}, \mathrm{NL}, \mathrm{P}\}$*,* $C' \in \{\mathrm{L}, \mathrm{NL}, \mathrm{P}, \mathrm{NP}\}$*, and* $C \subseteq C'$*. Then* $L(C)_{\mathrm{inv}}$ *is a* C'*-bounded logic for* C *if and only if* $p\text{-}L(C)\text{-}\textsc{Inv} \in \mathrm{XC}'_{\mathrm{uni}}$*.*

Proof. By Lemma 3.5, we already know that $L(C)_{\mathrm{inv}}$ is a logic for C, that is, part (a) of Definition 3.2 is fulfilled. By (3.1) and Theorem 3.3, part (b) of this definition is fulfilled if $p\text{-}L(C)\text{-}\textsc{Inv} \in \mathrm{XC}'_{\mathrm{uni}}$. Conversely, if part (b) is fulfilled, then, by (3.2), $p\text{-}L(C)\text{-}\textsc{Inv} \in \mathrm{XC}'_{\mathrm{uni}}$. □

In the following section we will prove:

Lemma 3.7 *Let* $C \in \{\mathrm{L}, \mathrm{NL}, \mathrm{P}\}$ *and* $C' \in \{\mathrm{L}, \mathrm{NL}, \mathrm{P}, \mathrm{NP}\}$*. Then*

$$\begin{aligned} p\text{-}\textsc{Halt}_> \in \mathrm{XC}'_{\mathrm{uni}} &\iff p\text{-}L(C)\text{-}\textsc{Inv} \in \mathrm{XC}'_{\mathrm{uni}} \\ &\iff p\text{-}\mathrm{FO}_{\mathrm{inv}}\text{-}\textsc{Inv} \in \mathrm{XC}'_{\mathrm{uni}}. \end{aligned}$$

Proof of Theorem 3.1: Let C and C' be as in the statement of Theorem 3.1. Then

$$\begin{aligned} L(C)_{\mathrm{inv}} \text{ is a } C'\text{-bounded logic for } C &\iff p\text{-}L(C)\text{-}\textsc{Inv} \in \mathrm{XC}'_{\mathrm{uni}} \quad \text{(by Lemma 3.6)} \\ &\iff p\text{-}\textsc{Halt}_> \in \mathrm{XC}'_{\mathrm{uni}} \qquad \text{(by Lemma 3.7).} \end{aligned}$$ □

We draw some consequences from Theorem 3.1. Even though it is not known whether the existence of an L-bounded logic for L implies the existence of a P-bounded logic for P, we get:

Corollary 3.8 *If* $\mathrm{DTC}_{\mathrm{inv}}$ *is an* L*-bounded logic for* L*, then* $\mathrm{TC}_{\mathrm{inv}}$ *is an* NL*-bounded logic for* NL *and* $\mathrm{LFP}_{\mathrm{inv}}$ *is a* P*-bounded logic for* P*.*

Proof. As $\mathrm{XL_{uni}} \subseteq \mathrm{XNL_{uni}}$ and $\mathrm{XL_{uni}} \subseteq \mathrm{XP_{uni}}$, the results follow from Theorem 3.1. □

As $\mathrm{XNL_{uni}} \not\subseteq \mathrm{XP_{uni}}$, so far we do not know whether $\mathrm{LFP_{inv}}$ is a P-bounded logic for P if $\mathrm{TC_{inv}}$ is an NL-bounded logic for NL. Nevertheless, in Corollary 4.3 of the next section we will see that this implication holds.

Corollary 3.9

(a) *If there is an algorithm deciding $\mathcal{A} \models_{\mathrm{FO_{inv}}} \varphi$ which, for fixed first-order φ, requires space logarithmic in $|\mathcal{A}|$, then* $\mathrm{DTC_{inv}}$ *is an* L*-bounded logic for* L.

(b) *If* $\mathrm{DTC_{inv}}$ *is a* P*-bounded logic for* L, *then* $\mathrm{LFP_{inv}}$ *is a* P*-bounded logic for* P.

Proof. (a) Assume that the hypothesis on $\mathrm{FO_{inv}}$ in (a) is fulfilled. Then, by (3.2), p-FO-INV $\in \mathrm{XL_{uni}}$ and thus p-HALT$_>$ $\in \mathrm{XL_{uni}}$ by Lemma 3.7. Now our claim follows from Theorem 3.1. The proof of (b) is similar, even easier. □

It is not known whether the existence of a P-bounded logic for P implies that of an *effectively* P-bounded logic for P. However, we can show:

Proposition 3.10 *If* $\mathrm{LFP_{inv}}$ *is a* P*-bounded logic for* P, *then there is an effectively* P*-bounded logic for* P.

Proof. By (3.2), for every $\mathrm{LFP_{inv}}[\tau]$-sentence and $m \geq 1$ we have

$$(3.3) \qquad \langle \varphi, m\rangle \in p\text{-}L\text{-}\mathrm{INV} \iff \big(\mathcal{A}(\tau, m) \models_{\mathrm{LFP_{inv}}} \varphi \;\text{ or }\; \mathcal{A}(\tau, m) \models_{\mathrm{LFP_{inv}}} \neg\varphi\big),$$

where $\mathcal{A}(\tau, m)$ is the τ-structure with universe $\{1, \dots, m\}$ and empty relations (that is, every relation symbol in τ is interpreted by the empty set). Assume that $\mathrm{LFP_{inv}}$ is a P-bounded logic for P and let $\mathbb{A}$ be an algorithm according to Definition 3.2 (b). Let $\mathbb{B}$ be the algorithm that decides $\langle \varphi, m\rangle \in p$-$L$-INV via the equivalence (3.3), where it checks the disjuncts on the right hand side of (3.3) applying the algorithm $\mathbb{A}$. Then, for fixed φ the algorithm $\mathbb{B}$ has running time polynomial in m. We define the logic $T(\mathrm{LFP_{inv}})$, *time-clocked* $\mathrm{LFP_{inv}}$, by:

- for every vocabulary τ,
$$T(\mathrm{LFP_{inv}})[\tau] := \big\{(\varphi, p) \bigm| \varphi \in \mathrm{LFP_{inv}}[\tau] \text{ and } p \in \mathbb{N}[X]\big\};$$
- $\mathcal{A} \models_{T(\mathrm{LFP_{inv}})} \varphi$ iff (a) and (b) hold where
  (a) $\mathbb{B}$ accepts $\langle \varphi, |A|\rangle$ in $\leq p(|A|)$ steps;
  (b) $(\mathcal{A}, <) \models_{\mathrm{LFP}} \varphi$ for some ordering $<$, say for the ordering of A induced by (the string) $\mathcal{A}$.

It is not hard to verify that $T(\mathrm{LFP_{inv}})$ is an effectively P-bounded logic for P (here we use for checking (b) an algorithm deciding $\models_{\mathrm{LFP}}$ that witnesses that LFP is an effectively P-bounded logic for P on ordered structures (see Theorem 3.3)). □

Remark 3.11 In a slightly different way but using the same idea one can define the time-clocked version $T(L)$ for any P-bounded logic L for P. However, in general, $T(L)$ is not even a logic, as the class of models of a $T(L)$-sentence must not be closed under isomorphism. In the case of $T(\mathrm{LFP_{inv}})$ this is guaranteed by the fact that condition (a) in the definition of $\mathcal{A} \models_{T(\mathrm{LFP_{inv}})} \varphi$ only refers to the cardinality of the universe of $\mathcal{A}$.

Example 3.12 There are NP-bounded logics for P. In fact, consider the logic $\mathrm{LFP_{inv}}(\mathrm{not})$ which has the same syntax as $\mathrm{LFP_{inv}}$, that is,

$$\mathrm{LFP_{inv}}(\mathrm{not})[\tau] := \mathrm{LFP_{inv}}[\tau]$$

and whose semantics is given by

$$\mathcal{A} \models_{\mathrm{LFP_{inv}(not)}} \varphi \iff \text{not } \mathcal{A} \models_{\mathrm{LFP_{inv}}} \varphi.$$

As P is closed under complements, $\mathrm{LFP_{inv}}$(not) is a logic for P. Using the definition of the semantics, one easily verifies that $\mathrm{LFP_{inv}}$(not) is even an NP-bounded logic for P.

3.1 A variant

What does p-HALT$_>$ $\in$ XP or p-HALT$_>$ $\in$ FPT mean for the invariant logics? For simplicity, we only consider $\mathrm{LFP_{inv}}$ (for which we already answered this question in [**3**]). Arguing as in the proof of Lemma 3.6 one gets

$$\mathrm{LFP_{inv}} \text{ is an effectively P-bounded logic for P} \iff p\text{-LFP-INV} \in \mathrm{XP}.$$

Using this equivalence, one can obtain:

Theorem 3.13 $\mathrm{LFP_{inv}}$ *is an effectively* P*-bounded logic for* P $\Leftrightarrow$ p-HALT$_>$ $\in$ XP.

The result for p-HALT$_>$ $\in$ FPT is more involved and we refer the reader to [**3**] for the details. There for an LFP-formula φ we introduce its *depth*, the maximum nesting depth of LFP-operators in φ, and its *width,* essentially the maximum of the number of variables in subformulas of φ. It is not hard to see that there is an algorithm deciding whether $\mathcal{A} \models_{\mathrm{LFP}} \varphi$ in time

$$|\varphi| \cdot |\mathcal{A}|^{O((1+\mathrm{depth}(\varphi))\cdot \mathrm{width}(\varphi))}.$$

We say that $\mathrm{LFP_{inv}}$ is an *(effectively) depth-width* P*-bounded logic for* P if it is a logic for P and there is a (computable) function $h\colon \mathbb{N} \to \mathbb{N}$ and an algorithm $\mathbb{A}$ deciding whether $\mathcal{A} \models_{\mathrm{LFP_{inv}}} \varphi$ in time

$$h(|\varphi|) \cdot \|\mathcal{A}\|^{O\big((1+\mathrm{depth}(\varphi))\cdot \mathrm{width}(\varphi)\big)}.$$

The following holds:

Theorem 3.14

(a) $\mathrm{LFP_{inv}}$ *is a depth-width* P*-bounded logic for* P $\Leftrightarrow$ p-HALT$_>$ $\in$ $\mathrm{FPT_{uni}}$.
(b) $\mathrm{LFP_{inv}}$ *is an effectively depth-width* P*-bounded logic for* P $\Leftrightarrow$ p-HALT$_>$ $\in$ FPT.

4 Slicewise downward monotone parameterized problems

A parameterized problem (Q, κ) is *slicewise downward monotone* if all elements of Q have the form $\langle x, 1^n\rangle$ with $x \in \Sigma^*$ and $n \in \mathbb{N}$, if $\kappa(\langle x, 1^n\rangle) = |x|$, and finally if the slices are downward monotone, that is, for all $x \in \Sigma^*$ and $n, n' \in \mathbb{N}$,

$$\langle x, 1^n\rangle \in Q \text{ and } n' < n \text{ imply } \left\langle x, 1^{n'}\right\rangle \in Q.$$

Hence, p-HALT$_>$ and p-L-INV (for any logic L) are slicewise downward monotone. As already done for p-L-INV, often for a slicewise downward monotone (Q, κ) we will write $\langle x, n\rangle \in Q$ instead of $\langle x, 1^n\rangle \in Q$.

For any logic L the following parameterized problem is downward monotone, too.

p-L-MODEL$_>$

Instance: An L-sentence φ and 1^n with $n \in \mathbb{N}$.

Parameter: $|\varphi|$.

Problem: Does the universe of every L-model of φ have more than n elements?

In this section we aim to show that all slicewise downward monotone introduced so far have the same computational complexity.

We denote by coXNL$_{\text{uni}}$ the class of parameterized problems (Q, κ) such that their complement $(\Sigma^* \setminus Q, \kappa)$ is in XNL$_{\text{uni}}$.

Proposition 4.1

(a) *If* (Q, κ) *is slicewise downward monotone, then* (Q, κ) *is in* coXNL$_{\text{uni}}$.

(b) XNL$_{\text{uni}} \cap$ coXNL$_{\text{uni}} \subseteq$ XP$_{\text{uni}}$.

Proof. (a) Assume that (Q, κ) is slicewise downward monotone and let $\mathbb{Q}$ be an algorithm enumerating the elements of $\Sigma^* \setminus Q$. Then the following algorithm shows that $(\Sigma^* \setminus Q, \kappa) \in$ XNL$_{\text{uni}}$: If the input $y \in \Sigma^*$ does not have the form $\langle x, n\rangle$ for some $x \in \Sigma^*$ and $n \in \mathbb{N}$, then it accepts. If $y = \langle x, n\rangle$, it guesses $\ell \in \mathbb{N}$; then it checks whether $\mathbb{Q}$ in its first ℓ steps enumerates some $\langle x, m\rangle$ with $m \leq n$; if so, it accepts.

(b) Let $(Q, \kappa) \in$ XNL$_{\text{uni}} \cap$ coXNL$_{\text{uni}}$. Then there is a nondeterministic TM $\mathbb{M}$ and a function $f\colon \mathbb{N} \to \mathbb{N}$ such that, for $x \in \Sigma^*$,

- if $x \in Q$, then no run of $\mathbb{M}$ on x is rejecting and there is at least one accepting run using space $\leq f(\kappa(x)) \cdot \log |x|$;
- if $x \notin Q$, then no run of $\mathbb{M}$ on x is accepting and there is at least one rejecting run using space $\leq f(\kappa(x)) \cdot \log |x|$.

We consider the following algorithm $\mathbb{A}$ deciding Q:

```
A    // $x \in \Sigma^*$
    1. $\ell \leftarrow 1$
    2. construct the graph of configurations of M on input $x$ using space
       $\leq \ell \cdot \log |x|$
    3. if this graph contains an accepting run then accept
    4. if this graph contains a rejecting run then reject
    5. $\ell \leftarrow \ell + 1$
    6. goto 2.
```

Clearly, $\mathbb{A}$ decides Q and by our assumptions on $\mathbb{M}$ the algorithm $\mathbb{A}$ on input x will halt for some $\ell \leq f(\kappa(x))$. For fixed ℓ, Lines 2–4 take

$$2^{O(\ell \cdot \log |x|)} = |x|^{O(\ell)}$$

steps. Therefore the running time of $\mathbb{A}$ on input x can be bounded by

$$\sum_{\ell=1}^{f(\kappa(x))} |x|^{O(\ell)} = |x|^{O(f(\kappa(x)))}. \qquad \square$$

Now we can show the following results extending Corollary 2.11 and complementing Corollary 3.8.

Corollary 4.2 *If* TAUT *has an almost space optimal nondeterministic algorithm, then it has an almost (time) optimal algorithm.*

Proof. By assumption and Theorem 2.10 (b) we have p-HALT$_>$ $\in$ XNL$_{\text{uni}}$ and thus p-HALT$_>$ $\in$ XNL$_{\text{uni}} \cap$ coXNL$_{\text{uni}} \subseteq$ XP$_{\text{uni}}$ by the previous proposition; now the claim follows from Theorem 2.3 (a). □

Corollary 4.3 *If* TC$_{\text{inv}}$ *is an* NL*-bounded logic for* NL*, then* LFP$_{\text{inv}}$ *is a* P*-bounded logic for* P*.*

Proof. By assumption and Theorem 3.1 we have p-HALT$_>$ $\in$ XNL$_{\text{uni}}$ and thus p-HALT$_>$ $\in$ XNL$_{\text{uni}} \cap$ coXNL$_{\text{uni}} \subseteq$ XP$_{\text{uni}}$; applying Theorem 3.1 again we get the claim. □

To compare the complexity of parameterized problems here we use the notion of xlog-reduction: Let (Q, κ) and (Q', κ') be parameterized problems. We write $(Q, \kappa) \leq^{\text{xlog}} (Q', \kappa')$ if there is an *xlog-reduction* from (Q, κ) to (Q', κ'), that is, a mapping $R\colon \Sigma^* \to \Sigma^*$ with:

(a) For all $x \in \Sigma^*$ we have $(x \in Q \Leftrightarrow R(x) \in Q')$.
(b) $R(x)$ is computable in space $f(\kappa(x)) \cdot \log|x|$ for some computable $f\colon \mathbb{N} \to \mathbb{N}$.
(c) There is a computable function $g\colon \mathbb{N} \to \mathbb{N}$ such that $\kappa'(R(x)) \leq g(\kappa(x))$ for all $x \in \Sigma^*$.

The parameterized problems (Q, κ) and (Q', κ') are *xlog-equivalent* if $(Q, \kappa) \leq^{\text{xlog}} (Q', \kappa')$ and $(Q', \kappa') \leq^{\text{xlog}} (Q, \kappa)$. These are notions of the usual (strongly uniform) parameterized complexity theory. For example, we get the corresponding notion $\leq^{\text{xlog}}_{\text{uni}}$ of uniform parameterized complexity by allowing the functions f and g (in (b) and (c), respectively) to be arbitrary and not necessarily computable.

As we have not defined XL and XNL explicitly, we mention that we shall use part (b) of the following simple observation only for $C = \text{P}$.

Lemma 4.4 *Let* $C \in \{\text{L}, \text{NL}, \text{P}, \text{NP}\}$*.*

(a) *If* $(Q, \kappa) \leq^{\text{xlog}}_{\text{uni}} (Q', \kappa')$ *and* $(Q', \kappa') \in \text{XC}_{\text{uni}}$*, then* $(Q, \kappa) \in \text{XC}_{\text{uni}}$*.*

(b) *If* $(Q, \kappa) \leq^{\text{xlog}} (Q', \kappa')$ *and* $(Q', \kappa') \in \text{XC}$*, then* $(Q, \kappa) \in \text{XC}$*.*

We turn again to slicewise downward monotone problems. The goal of this section mentioned above can be stated in a precise form:

Theorem 4.5 *Let* $C \in \{\text{L}, \text{NL}, \text{P}\}$*. Then any two of the problems*

p-FO-MODEL$_>$, p-$L(C)$-MODEL$_>$, p-FO-INV, p-$L(C)$-INV, *and* p-HALT$_>$

are xlog-equivalent.

Note that Lemma 3.7 is an immediate consequence of this theorem and Lemma 4.4 (a).

The rest of this section is devoted to a proof of Theorem 4.5. First we remark that among the slicewise downward monotone problems with underlying classical problem in co-NP the problem HALT$_>$ is of highest complexity:

Proposition 4.6 *Let* (Q, κ) *be slicewise downward monotone and* $Q \in$ co-NP*. Then*

$$(Q, \kappa) \leq^{\text{xlog}} p\text{-HALT}_>.$$

Proof. Let $\mathbb{M}$ be a nondeterministic Turing machine accepting the complement $\Sigma^* \setminus Q$ of Q. We may assume that for some $d \in \mathbb{N}$ the machine $\mathbb{M}$ on input $\langle x, n\rangle$ performs exactly $|\langle x, n\rangle|^d$ steps. For $x \in \Sigma^*$ let $\mathbb{M}_x$ be the nondeterministic Turing machine that on empty input tape, first writes x on the tape, then guesses a natural number m, and finally it simulates the computation of $\mathbb{M}$ on input $\langle x, m\rangle$ and answers accordingly. We can assume that there is a logspace computable function h such that $\mathbb{M}_x$ makes exactly $h(x, m) \in O\big(|x| + m + |\langle x, m\rangle|^d\big)$ steps if it guesses the natural number m. Furthermore we can assume that $h(x, m) < h(x, m')$ for $m < m'$.

Then $\langle x, n\rangle \mapsto \langle \mathbb{M}_x, h(x, n)\rangle$ is an xlog-reduction from (Q, κ) to p-HALT$_>$: Clearly, if $\langle \mathbb{M}_x, h(x, n)\rangle \in p$-HALT$_>$, then by construction of $\mathbb{M}_x$ we have $\langle x, n\rangle \notin \Sigma^* \setminus Q$ and hence, $\langle x, n\rangle \in Q$. Conversely, if $\langle x, n\rangle \in Q$, then by slicewise downward monotonicity $\langle x, m\rangle \notin \Sigma^* \setminus Q$ for all $m \leq n$; thus $\langle \mathbb{M}_x, h(x, n)\rangle \in p$-HALT$_>$. □

Later on we shall use the following related result.

Lemma 4.7 *Let (Q, κ) be slicewise downward monotone and assume that there is a nondeterministic algorithm $\mathbb{A}$ accepting $\Sigma^* \setminus Q$ such that $t_{\mathbb{A}}(\langle x, n\rangle) \leq n^{f(|x|)}$ for some time constructible f and all $\langle x, n\rangle \in \Sigma^* \setminus Q$. Then*

$$(Q, \kappa) \leq^{\text{xlog}} p\text{-HALT}_>.$$

Proof. Let (Q', κ') be the problem

Instance:	$x \in \Sigma^*$ and 1^m with $m \in \mathbb{N}$.
Parameter:	$\|x\|$.
Problem:	Is $\langle x, n\rangle \in Q$ for all $n \in \mathbb{N}$ with $n^{f(\|x\|)} \leq m$?

By the previous proposition, we get our claim once we have shown:

(1) (Q', κ') is slicewise downward monotone and $Q' \in$ co-NP;
(2) $(Q, \kappa) \leq^{\text{xlog}} (Q', \kappa')$.

To see (1), let $\mathbb{A}$ be as stated above and let $\mathbb{T}$ be an algorithm witnessing the time constructibility of f; that is, $\mathbb{T}$ on input 1^k with $k \in \mathbb{N}$ computes $f(k)$ in exactly $f(k)$ steps. An algorithm $\mathbb{B}$ witnessing that $\Sigma^* \setminus Q' \in$ NP runs as follows on input $\langle x, m\rangle$:

- $\mathbb{B}$ guesses $n \in \mathbb{N}$;
- if $n = 1$, the algorithm $\mathbb{B}$ rejects in case $m = 0$;
  if $n \geq 2$, the algorithm $\mathbb{B}$ simulates m steps of the computation of $\mathbb{T}$ on input $1^{|x|}$; if thereby $\mathbb{T}$ does not stop, $\mathbb{B}$ rejects; otherwise, the simulation yields $f(|x|)$ and $\mathbb{B}$ checks whether $n^{f(|x|)} > m$ (this can be detected in time $O(m)$); in the positive case $\mathbb{B}$ rejects;
- finally $\mathbb{B}$ simulates the computation of $\mathbb{A}$ on $\langle x, n\rangle$ and answers accordingly.

(2) Note that the mapping $\langle x, n\rangle \mapsto \big\langle x, n^{f(|x|)}\big\rangle$ is an xlog-reduction. □

The following lemmas will finally yield a proof of Theorem 4.5.

Lemma 4.8 *Let $L =$ FO or $L =$ L(C) with $C \in \{$L, NL, P$\}$. Then p-L-*MODEL$_>$ $\leq^{\text{xlog}}$ *p-L-*INV.

Proof. Let φ be an $L[\tau]$-sentence. We set $\tau' := \tau \cup \{P\}$ with a new unary relation symbol P and consider the $L[\tau'_<]$-sentence

$$\psi(\varphi) := \varphi \wedge \text{“}P \text{ holds for the first element of } <\text{”}.$$

If φ has no model with exactly one element, then, for $n \in \mathbb{N}$,

$$\langle \varphi, n\rangle \in p\text{-}L\text{-}\textsc{Model}_{>} \iff \langle \psi(\varphi), n\rangle \in p\text{-}L\text{-}\textsc{Inv}.$$

For such φ we define the reduction by $\langle \varphi, n\rangle \mapsto \langle \psi(\varphi), n\rangle$; if the sentence φ has a model with exactly one element, by $\langle \varphi, n\rangle \mapsto \langle$"$P$ holds for the first element of $<$", $2\rangle$. □

Lemma 4.9 *p-LFP-Inv $\leq^{\text{xlog}} p$-Halt$_{>}$ and hence p-$L(C)$-Inv $\leq^{\text{xlog}} p$-Halt$_{>}$ for $C = \mathrm{L}$ and $C = \mathrm{NL}$.*

Proof. Consider the nondeterministic algorithm $\mathbb{A}$ that on input $\langle \varphi, m\rangle$, where φ is an LFP-sentence and $m \geq 1$, guesses a structure $\mathcal{A}$ and two orderings $<_1$ and $<_2$, and accepts if $|A| \leq m$, $(\mathcal{A}, <_1) \models_{\text{LFP}} \varphi$, and $(\mathcal{A}, <_2) \models_{\text{LFP}} \neg\varphi$. It accepts the complement of p-LFP-Inv. As LFP is an effectively P-bounded logic for P on ordered structures, the algorithm $\mathbb{A}$ witnesses that p-LFP-Inv satisfies the assumptions on (Q, κ) in Lemma 4.7. This yields the claim. □

Lemma 4.10 p-Halt$_{>}$ $\leq^{\text{xlog}} p$-FO-Model$_{>}$.

Proof. Coding configurations of a nondeterministic machine and its nondeterministic choices in a standard way it is easy to assign to every nondeterministic TM $\mathbb{M}$ a first-order sentence $\varphi_{\mathbb{M}}$ such that, for all $n \in \mathbb{N}$,

$$\mathbb{M} \text{ has accepting run on empty input of length } n \iff \varphi_{\mathbb{M}} \text{ has a model with universe of cardinality } n.$$

Thus, $\langle \mathbb{M}, n\rangle \mapsto \langle \varphi_{\mathbb{M}}, n\rangle$ is the desired xlog-reduction. □

Proof of Theorem 4.5: Immediate by Lemmas 4.8, 4.9, and 4.10. □

5 Optimal proof systems, almost optimal algorithms, and listings

By Krajíček and Pudlák [**17**] the existence of an almost optimal algorithm for Taut is equivalent to the existence of a polynomially optimal proof system for Taut and by Sadowski [**22**] it is equivalent to the existence of listings of the subsets of Taut in P. These equivalences extend to the different variants (strong, effective, and space) of almost optimal algorithms we considered in Section 2. In Section 5.1 we prove the strong variant; for the other cases we present the definitions and the results in Section 5.2.

Listings. Listings (or effective enumerations) of problems by means of TMs have been used to characterize promise classes possessing complete languages (e.g., see [**11, 16**]). In the context of almost optimal algorithms, listings of subsets of Taut have been used systematically by Sadowski (see [**22**]). We introduce our general notion of listing.

Definition 5.1 Let $Q \subseteq \Sigma^*$ and $C, C' \in \{\mathrm{L}, \mathrm{NL}, \mathrm{P}, \mathrm{NP}\}$.

(1) A *C-subset of Q* is a set Y with $Y \subseteq Q$ and $Y \in C$.

(2) A *listing of the C-subsets of Q by C'-machines* is an algorithm $\mathbb{L}$ that, once having been started, eventually yields as outputs TMs $\mathbb{M}_1, \mathbb{M}_2, \ldots$ of type C' such that

$$\{L(\mathbb{M}_i) \mid i \geq 1\} = \{Y \subseteq Q \mid Y \in C\}.$$

(3) Let $C, C' \in \{\mathrm{P}, \mathrm{NP}\}$. A listing $\mathbb{L}$ of the C-subsets of Q by C'-machines is *strong* if there is a constant $d \in \mathbb{N}$ such that for every C-subset Y of Q and every C'-machine $\mathbb{M}$ deciding Y there is a machine listed by $\mathbb{L}$ deciding $x \in Y$ in time $\leq p(t_{\mathbb{M}}(x) + |x|)$, where $p \in \mathbb{N}_d[X]$ (that is, p is a polynomial of degree $\leq d$).

We write $\mathrm{List}(C, Q, C')$ and $\mathrm{SList}(C, Q, C')$ if there is a listing and a strong listing, respectively, of the C-subsets of Q by C'-machines.

Sometimes we speak of the listing $\mathbb{M}_1, \mathbb{M}_2, \ldots$ (instead of the listing $\mathbb{L}$). By systematically adding polynomial time clocks (if C' is a time class) or devices controlling the space used, we may assume that *all* runs of the machines $\mathbb{M}_i$ on *any* input satisfy the time or space bound characteristic for C'.

Optimal proof systems. Let $Q \subseteq \Sigma^*$. A *proof system for* Q in the sense of Cook and Reckhow [**7**] is a polynomial time algorithm $\mathbb{P}$ computing a function from Σ^* onto Q. If $\mathbb{P}(w) = x$, we say that w is a $\mathbb{P}$-*proof* of x. Often we introduce proof systems implicitly by defining the corresponding function; then this definition will suggest an algorithm.

Let $\mathbb{P}$ and $\mathbb{P}'$ be proof systems for Q. A *translation from* $\mathbb{P}'$ *into* $\mathbb{P}$ is a polynomial time algorithm $\mathbb{T}$ such that $\mathbb{P}(\mathbb{T}(w')) = \mathbb{P}'(w')$ for all $w' \in \Sigma^*$.

A proof system $\mathbb{P}$ for Q is *polynomially optimal* or *p-optimal* if for every proof system $\mathbb{P}'$ for Q there is a translation from $\mathbb{P}'$ into $\mathbb{P}$. If for $\mathbb{P}$ there is a constant $d \in \mathbb{N}$ such that for every proof system $\mathbb{P}'$ for Q there is a translation $\mathbb{T}$ from $\mathbb{P}'$ into $\mathbb{P}$ and a $p \in \mathbb{N}_d[X]$ such that

$$t_{\mathbb{T}}(w') \leq p\big(t_{\mathbb{P}'}(w') + |w'|\big)$$

for all $w' \in \Sigma^*$, then $\mathbb{P}$ is *strongly p-optimal.*

A proof system $\mathbb{P}$ for Q is *optimal* if for every proof system $\mathbb{P}'$ for Q and every $w' \in \Sigma^*$ there is a $w \in \Sigma^*$ such that $\mathbb{P}(w) = \mathbb{P}'(w')$ and $|w| \leq |w'|^{O(1)}$. Again, if there is a constant $d \in \mathbb{N}$ such that for all $\mathbb{P}'$ and every $w' \in \Sigma^*$ there is a $w \in \Sigma^*$ such that $\mathbb{P}(w) = \mathbb{P}'(w')$ and $|w| \leq p\big(t_{\mathbb{P}'}(w') + |w'|\big)$ for some $p \in \mathbb{N}_d[X]$, then $\mathbb{P}$ is *strongly optimal.*

In the following theorem we summarize the results that are known for $Q = \textsc{Taut}$ mentioned at the beginning of this section. The extensions to arbitrary Q with padding are mainly due to Messner [**19**] or are implicit in Sadowski [**22**].

Theorem 5.2

(1) *For every* Q *we have* (a) $\Rightarrow$ (b) *and* (b) $\Rightarrow$ (c)*; moreover* (a), (b), *and* (c) *are all equivalent if* Q *has padding. Here*
   (a) *Q has a p-optimal proof system;*
   (b) *Q has an almost optimal algorithm;*
   (c) $\mathrm{List}(\mathrm{P}, Q, \mathrm{P})$.

(2) *For every* Q *we have* (a) $\Leftrightarrow$ (b) *and* (b) $\Rightarrow$ (c)*; moreover* (a), (b), *and* (c) *are all equivalent if* Q *has padding. Here*
   (a) *Q has an optimal proof system;*
   (b) *Q has an almost optimal nondeterministic algorithm;*
   (c) $\mathrm{List}(\mathrm{NP}, Q, \mathrm{NP})$.

5.1 The strong variant

The following series of lemmas will lead to a proof of the result corresponding to Theorem 5.2 for the strong notions.

Lemma 5.3 *Let $Q \subseteq \Sigma^*$. If there is a strongly almost optimal (nondeterministic) algorithm for Q, then* $\textsc{SList}(\mathrm{P}, Q, \mathrm{P})$ $\big(\textsc{SList}(\mathrm{NP}, Q, \mathrm{NP})\big)$.

Proof. We prove the deterministic case, the nondeterministic one is obtained by obvious changes. Let $\mathbb{O}$ be a strongly almost optimal algorithm for Q. Let $d \in \mathbb{N}$ bound the degree of the polynomials p in the definition of strongly almost optimal (Definition 2.1). Let $t\colon \Sigma^* \to \{1\}^*$ be a function such that $t = t_{\mathbb{M}}$ for some TM $\mathbb{M}$. Let $\mathbb{O}(t)$ be the algorithm that on input x simulates $\mathbb{O}$ on input x but rejects if the simulation exceeds time $t(x)$ (here we identify 1^k with k). Clearly,

(i) $L(\mathbb{O}(t))$ is a P-subset of Q if t is computable in polynomial time.

Moreover, we show:

(ii) For every P-subset Y of Q and every TM $\mathbb{M}$ deciding Y there is a polynomial time TM $\mathbb{M}'$ deciding Y with $t_{\mathbb{M}'}(x) \le O(t_{\mathbb{M}}(x))$;

(iii) For every P-subset Y of Q and every polynomial time TM $\mathbb{M}$ deciding Y there is $p \in \mathbb{N}_d[X]$ with $Y \subseteq L(\mathbb{O}(p[t_{\mathbb{M}}]))$, where $p[t_{\mathbb{M}}] : \Sigma^* \to \{1\}^*$ is defined by
$$p[t_{\mathbb{M}}](x) := p\big(t_{\mathbb{M}}(x) + |x|\big).$$

We get the TM $\mathbb{M}'$ satisfying (ii) by running $\mathbb{M}$ in parallel with any polynomial time TM deciding Y.

For (iii) we argue as follows. The following algorithm $\mathbb{C}$ decides Q: On input x, in parallel it simulates $\mathbb{M}$ and $\mathbb{O}$ on input x; if $\mathbb{M}$ halts first and accepts, then $\mathbb{C}$ accepts, otherwise it answers as $\mathbb{O}$. Note that $t_{\mathbb{C}}(x) \le O(t_{\mathbb{M}}(x))$ for $x \in Y$. By the strongly almost optimality of $\mathbb{O}$ there is a $p \in \mathbb{N}_d[X]$ with
$$t_{\mathbb{O}}(x) \le p\big(t_{\mathbb{C}}(x) + |x|\big)$$
for all $x \in Q$ and thus, for all $x \in Y$,
$$t_{\mathbb{O}}(x) \le p\big(O(t_{\mathbb{M}}(x)) + |x|\big).$$
Therefore, $Y \subseteq L(\mathbb{O}(p_1[t_{\mathbb{M}}]))$ for some $p_1 \in \mathbb{N}_d[X]$.

We fix a listing $\mathbb{M}_1, \mathbb{M}_2, \ldots$ of all polynomial time deterministic TMs. By (i)–(iii), $\big(\mathbb{M}_i(\mathbb{O}(p[t_{\mathbb{M}_i}])\big)_{i \ge 1,\ p \in \mathbb{N}_d[X]}$ is a strong listing of the P-subsets of Q, where $\mathbb{M}_i(\mathbb{O}(p[t_{\mathbb{M}_i}]))$ on input x, first simulates $\mathbb{O}(p[t_{\mathbb{M}_i}])$ on x and if this algorithm accepts, then it simulates $\mathbb{M}_i$ on input x and answers accordingly. □

In a second step we deal with the transition from listings to optimal proof systems.

Lemma 5.4 *Assume that Q has a padding function.*

(a) *If* $\textsc{SList}(\mathrm{P}, Q, \mathrm{P})$*, then Q has a strongly p-optimal proof system.*

(b) *If* $\textsc{SList}(\mathrm{NP}, Q, \mathrm{NP})$*, then Q has a strongly optimal proof system.*

Proof. Again we prove (a), thereby indicating the changes that are necessary for a proof of (b). Fix $y_0 \in Q$ and let *pad*: $\Sigma^* \times \Sigma^* \to \Sigma^*$ be a padding function for Q. Let $\mathbb{L}$ be a strong listing of the P-subsets of Q by P-machines and let $d \in \mathbb{N}$ be such that polynomials in $\mathbb{N}_d[X]$ witness the strongness of $\mathbb{L}$. We say that $v \in \Sigma^*$ is a *proof string* if it has the form
$$v = \Big\langle \mathbb{M}, w, y, 1^m, \mathbb{M}', 1^\ell, 1^r, 1^t \Big\rangle,$$
where

- $\mathbb{M}$ is a TM that on input w outputs y in $\le m$ steps;
- $\mathbb{L}$ lists $\mathbb{M}'$ in $\le \ell$ steps;

- $\mathbb{M}'$ accepts $pad(y, \langle w, 1^r\rangle)$ in $\leq t$ steps. $\big($For (b) we have to add s in the tuple v, where s is the sequence of states of a run of $\mathbb{M}'$ accepting $pad(y, \langle w, 1^r\rangle)$ in $\leq t$ steps.$\big)$

Clearly, we can decide in polynomial time whether a string $v \in \Sigma^*$ is a proof string. Moreover, if v is a proof string, then $y \in Q$ (as $L(\mathbb{M}') \subseteq Q$ and $\mathbb{M}'$ accepts $pad(y, \langle w, 1^r\rangle)$). We consider the proof system $\mathbb{P}$ defined by

$$\mathbb{P}\Big(\Big\langle \mathbb{M}, w, y, 1^m, \mathbb{M}', 1^\ell, 1^r, 1^t\Big\rangle\Big) := y$$

if $\big\langle \mathbb{M}, w, y, 1^m, \mathbb{M}', 1^\ell, 1^r, 1^t\big\rangle$ is a proof string, and $\mathbb{P}(w) := y_0$ otherwise. Clearly, $\mathbb{P}$ is polynomial time and has a subset of Q as range. So, $\mathbb{P}$ is a proof system for Q if we can show that every $y \in Q$ is in its range: As $\big\{pad(y, \langle y, 1\rangle)\big\}$ is a P-subset of Q, a machine $\mathbb{M}'$ deciding $\big\{pad(y, \langle y, 1\rangle)\big\}$ is listed by $\mathbb{L}$, say, in ℓ steps. Then $\mathbb{P}(v) = y$ for $v = \big\langle \mathbb{M}_{\text{id}}, y, y, 1^m, \mathbb{M}', 1^\ell, 1, 1^t\big\rangle$, where $\mathbb{M}_{\text{id}}$ is a machine that on input x outputs x and that on input y takes m steps, and t is the number of steps of $\mathbb{M}'$ on input $pad(y, \langle y, 1\rangle)$.

It remains to show that $\mathbb{P}$ is strongly p-optimal. For this purpose let $\mathbb{P}' \colon \Sigma^* \to Q$ be a proof system for Q. Then,

$$Graph(\mathbb{P}') := \big\{pad(y, \langle w', 1^r\rangle) \bigm| y, w' \in \Sigma^*,\ \mathbb{P}'(w') = y \text{ and } r = t_{\mathbb{P}'}(w')\big\}$$

is a P-subset of Q. Let $\mathbb{B}$ be the algorithm deciding $Graph(\mathbb{P}')$, which on input x first applying property (ii) of the padding function (see page 149) computes w' and r if $x = pad(\ldots, \langle w', 1^r\rangle)$, then it computes $y := \mathbb{P}'(w')$ and checks if $r = t_{\mathbb{P}'}(w')$, both by simulating $\mathbb{P}'$ (for at most r steps), and finally it checks whether $x = pad(y, \langle w', 1^r\rangle)$. Then there is a polynomial q_0 depending on the functions pad and $\langle\ ,\ \rangle$ only, say, of degree $d_0 \geq 1$ such that

$$t_{\mathbb{B}}(x) \leq q_0(|x|) + r \leq 2 \cdot q_0(|x|).$$

Hence, by assumption, $\mathbb{L}$ lists a machine $\mathbb{M}'$, say in ℓ steps, that decides $x \in Graph(\mathbb{P}')$ in $\leq p\big(q_0(|x|)\big)$ steps for some $p \in \mathbb{N}_d[X]$ (as $t_{\mathbb{B}}(x) + |x| \leq 3 \cdot q_0(|x|)$). Furthermore note that for $w' \in \Sigma^*$ and $x := pad(\mathbb{P}'(w'), \Big\langle w', 1^{t_{\mathbb{P}'}(w')}\Big\rangle)$ we have

$$|x| \leq q_1(t_{\mathbb{P}'}(w') + |w'|),$$

where again the polynomial q_1, say of degree d_1, depends only on the padding function and the tupling function. We define the translation $\mathbb{T}$ by

$$\mathbb{T}(w') := \Big\langle \mathbb{P}', w', \mathbb{P}'(w'), 1^{t_{\mathbb{P}'}(w')}, \mathbb{M}', 1^\ell, 1^{t_{\mathbb{P}'}(w')}, 1^t\Big\rangle \tag{5.1}$$

where

$$t := p\Big(q_0\big(q_1(t_{\mathbb{P}'}(w') + |w'|)\big)\Big).$$

It is easy to check that $\mathbb{T}(w')$ is a proof string with $\mathbb{P}(\mathbb{T}(w')) = \mathbb{P}'(w')$. As $\mathbb{M}'$ and 1^ℓ do not depend on w', the algorithm $\mathbb{T}$ on w' needs $p_1\big(t_{\mathbb{P}'}(w') + |w'|\big)$ steps for a polynomial p_1 of degree $\leq d + d_0 + d_1$; this finishes the proof. For (b) we get a tuple w with the desired properties by adding to the tuple in (5.1) a sequence s of states of a run of $\mathbb{M}'$ accepting $pad(\mathbb{P}'(w'), w')$ of length $\leq t$. □

The reader will have noticed that the previous proof also shows the conclusion of part (b) under the hypothesis $\text{SL}\textsc{ist}(\text{P}, Q, \text{NP})$ as the set $Graph(\mathbb{P}')$ is in P. However, the equivalence $\big(\text{SL}\textsc{ist}(\text{P}, Q, \text{NP}) \Leftrightarrow \text{SL}\textsc{ist}(\text{NP}, Q, \text{NP})\big)$ holds for all Q with padding. This can be shown along the lines of the proof of Proposition 5.12, where the “non-strong” version of this equivalence is derived.

We turn to the last step, the transition from proof systems to almost optimal algorithms. We need the following result on inverters due to Levin [**18**].

Definition 5.5 Let $f\colon \Sigma^* \to \Sigma^*$ be a function. An algorithm $\mathbb{A}$ *inverts* f if for every x in the *range* of f the algorithm $\mathbb{A}$ computes some w with $f(w) = x$. For x not in the range of f the algorithm $\mathbb{A}$ can behave arbitrarily.

By Levin’s result, for any algorithm $\mathbb{F}$ computing a function f there is an inverter $\mathbb{O}$, which is optimal with respect to the time required by the computation of $\mathbb{F}(\mathbb{O}(y))$.

Theorem 5.6 *There is a $d_0 \in \mathbb{N}$ such that for all algorithms $\mathbb{F}$ computing a function $f\colon \Sigma^* \to \Sigma^*$ there exists an algorithm $\mathbb{O}$ such that:*

(a) *$\mathbb{O}$ inverts f and $t_{\mathbb{O}}(y) = \infty$ for every input y not in the range of f;*

(b) *for every algorithm $\mathbb{I}$ inverting f there is an $p \in \mathbb{N}_{d_0}[X]$ such that for every y in the range of f we have*

$$t_{\mathbb{O}}(y) \le p\big(|y| + t_{\mathbb{I}}(y) + t_{\mathbb{F}}(\mathbb{I}(y))\big).$$

We now state the result yielding an almost optimal algorithm from a p-optimal proof system.

Lemma 5.7

(a) *If Q has a strongly p-optimal proof system, then Q has a strongly almost optimal algorithm.*

(b) *If Q has a strongly optimal proof system, then Q has a strongly almost optimal nondeterministic algorithm.*

Proof. Again we only prove (a); then (b) is obtained by changes similar to that indicated in the previous proof for the nondeterministic case. Let $\mathbb{P}\colon \Sigma^* \to Q$ be a strongly p-optimal proof system for Q and let $d \in \mathbb{N}$ be a bound for the degrees of the polynomials according to the strongness. Fix $y_0 \in Q$. We define the function $f\colon \Sigma^* \to \Sigma^*$ by

$$f\big(\big\langle \mathbb{A}, y, 1^m, \mathbb{S}, 1^t \big\rangle\big) := y$$

if

(F1) $\mathbb{A}$ is an algorithm that accepts $y \in \Sigma^*$ in $\le m$ steps;

(F2) $\mathbb{S}$ is an algorithm that on input $\langle y, 1^m\rangle$ computes a string w with $\mathbb{P}(w) = y$ in $\le t$ steps.

Otherwise, we set $f(w) := y_0$. It is easy to verify that the range of f is Q, in particular, (F2) guarantees that it is a subset of Q. Moreover there is an algorithm $\mathbb{F}$ that computes the function f in polynomial time, say, in time $\le q_{\mathbb{F}}(n)$ with $q_{\mathbb{F}}$ of degree $d(\mathbb{F})$. We choose d_0 according to Theorem 5.6 and an optimal inverter $\mathbb{O}$, that is, an inverter with the properties (a) and (b) of Theorem 5.6; in particular, for every inverter $\mathbb{I}$ of f there is a $p \in \mathbb{N}_{d_0}[X]$ such that, for every $y \in Q$,

$$t_{\mathbb{O}}(y) \le p\big(|y| + t_{\mathbb{I}}(y) + t_{\mathbb{F}}(\mathbb{I}(y))\big) \le p_1(|y| + t_{\mathbb{I}}(y)) \tag{5.2}$$

with $p_1 \in \mathbb{N}_{d_0+d(\mathbb{F})}[X]$ as $t_{\mathbb{F}}(\mathbb{I}(y)) \le q_{\mathbb{F}}(|\mathbb{I}(y)|) \le q_{\mathbb{F}}(t_{\mathbb{I}}(y)))$.

Let $\mathbb{Q}$ be any algorithm deciding Q. We claim that the following algorithm $\mathbb{O}||\mathbb{Q}$ is a strongly almost optimal algorithm deciding Q, where $\mathbb{O}||\mathbb{Q}$ on input y simulates $\mathbb{O}$ and $\mathbb{Q}$

on input y in parallel; if $\mathbb{O}$ halts first, then it accepts, and if $\mathbb{Q}$ halts first, then it answers accordingly.

The algorithm $\mathbb{O}\|\mathbb{Q}$ decides Q (here we use Theorem 5.6 (a)) and for $y \in Q$ we have

$$t_{\mathbb{O}\|\mathbb{Q}}(y) \le O(t_{\mathbb{O}}(y)). \tag{5.3}$$

We claim that $\mathbb{O}\|\mathbb{Q}$ is strongly almost optimal. For this purpose let $\mathbb{A}$ be any algorithm that decides Q. We get a proof system $\mathbb{P}_{\mathbb{A}}$ for Q by setting

$$\mathbb{P}_{\mathbb{A}}(w') := \begin{cases} y, & \text{if } w' = \langle y, 1^m \rangle \text{ and } \mathbb{A} \text{ accepts } y \text{ in } \le m \text{ steps,} \\ y_0, & \text{otherwise.} \end{cases}$$

Then, $t_{\mathbb{P}_{\mathbb{A}}}(w') = O(|w'|)$. Since $\mathbb{P}$ is strongly p-optimal for Q, there is a translation $\mathbb{T}$ and a polynomial $p_2 \in \mathbb{N}_d[X]$ such that, for all $w' \in \Sigma^*$,

$$\mathbb{P}(\mathbb{T}(w')) = \mathbb{P}_{\mathbb{A}}(w') \quad \text{and} \quad t_{\mathbb{T}}(w') \le p_2\big(t_{\mathbb{P}_{\mathbb{A}}}(w') + |w'|\big) \le p_2(O(|w'|)). \tag{5.4}$$

Using $\mathbb{A}$ and $\mathbb{T}$ we define an inverter $\mathbb{I}$ of the function f: On input y, the algorithm $\mathbb{I}$ simulates the algorithm $\mathbb{A}$ on y; if $\mathbb{A}$ rejects y, then $\mathbb{I}$ does not halt, if $\mathbb{A}$ accepts y, then it outputs

$$\Big\langle \mathbb{A}, y, 1^{t_{\mathbb{A}}(y)}, \mathbb{T}, 1^t \Big\rangle,$$

where t is the number of steps of the computation of $\mathbb{T}$ on $\big\langle y, 1^{t_{\mathbb{A}}(y)} \big\rangle$. Thus, for $y \in Q$,

$$\begin{aligned} t_{\mathbb{I}}(y) &\le O\Big(t_{\mathbb{A}}(y) + t_{\mathbb{T}}(\Big\langle y, 1^{t_{\mathbb{A}}(y)} \Big\rangle)\Big) \\ &= O\Big(t_{\mathbb{A}}(y) + p_2(O(|y| + t_{\mathbb{A}}(y)))\Big) \qquad \big(\text{by (5.4)}\big). \end{aligned}$$

Hence by (5.2),

$$\begin{aligned} t_{\mathbb{O}}(y) &\le p_1(|y| + t_{\mathbb{I}}(y)) \le p_1\Big(|y| + O\Big(t_{\mathbb{A}}(y) + p_2(O(|y| + t_{\mathbb{A}}(y)))\Big)\Big) \\ &\le p_3(|y| + t_{\mathbb{A}}(y)) \end{aligned}$$

for some $p_3 \in \mathbb{N}_{d_0 + d(\mathbb{F}) + d}[X]$. Together with (5.3), this shows the strongly almost optimality of $\mathbb{O}\|\mathbb{Q}$. □

It is well-known that nondeterministic algorithms and propositional proof systems are more or less the same. We use this fact to show that for arbitrary Q a strongly almost optimal nondeterministic algorithm yields a strongly optimal proof system. By the way, we could have already used this fact to get a simpler proof of part (b) of the previous lemma, a proof tailored only for the nondeterministic case.

Lemma 5.8 *If Q has a strongly almost optimal nondeterministic algorithm, then it has a strongly optimal proof system.*

Proof. Let $\mathbb{A}$ be a strongly almost optimal nondeterministic algorithm for Q and let $d \ge 1$ be such that polynomials in $\mathbb{N}_d[X]$ witness the strongness of $\mathbb{A}$. We fix $y_0 \in Q$ and define the proof system $\mathbb{P}$ for Q by

$$\mathbb{P}(w) := \begin{cases} y, & \text{if } w = \langle y, s \rangle, \text{ where } s \text{ is the sequence of states of a run of } \mathbb{A} \text{ accepting } y, \\ y_0, & \text{otherwise.} \end{cases}$$

The proof system $\mathbb{P}$ is strongly optimal: Let $\mathbb{P}'$ be any proof system for Q. The nondeterministic algorithm $\mathbb{B}(\mathbb{P}')$ accepts Q, where $\mathbb{B}(\mathbb{P}')$ on input $y \in \Sigma^*$ guesses a string w' and accepts if $\mathbb{P}'(w') = y$. Hence, there is a $q \in \mathbb{N}_d[X]$ such that for all $y \in Q$

$$t_{\mathbb{A}}(y) \leq q\big(t_{\mathbb{B}(\mathbb{P}')}(y) + |y|\big). \tag{5.5}$$

Let $\mathbb{P}'(w') = y$. Then

$$t_{\mathbb{B}(\mathbb{P}')}(y) = O\big(|w'| + t_{\mathbb{P}'}(w')\big). \tag{5.6}$$

Let s be the sequence of states of a run of $\mathbb{A}$ accepting y in $t_{\mathbb{A}}(y)$ steps and set $w := \langle y, s\rangle$. Then $\mathbb{P}(w) = y$ and

$$\begin{aligned} |w| = \big|\, \langle y, s\rangle \,\big| = O(|y| + t_{\mathbb{A}}(y)) &= O\Big(t_{\mathbb{P}'}(w') + q\big(t_{\mathbb{B}(\mathbb{P}')}(y) + |y|\big)\Big) && \text{(by (5.5))}\\ &= O\Big(t_{\mathbb{P}'}(w') + q\big(O\big(|w'| + t_{\mathbb{P}'}(w')\big) + t_{\mathbb{P}'}(w')\big)\Big) && \text{(by (5.6))}. \end{aligned}$$

This shows that $|w| \leq p_1(t_{\mathbb{P}'}(w') + |w'|)$ for some $p_1 \in \mathbb{N}_d[X]$. □

Summarizing, we have shown:

Theorem 5.9

(1) *For every Q we have* (a) $\Rightarrow$ (b) *and* (b) $\Rightarrow$ (c)*; moreover* (a), (b), *and* (c) *are all equivalent if Q has padding. Here*
 (a) *Q has a strongly p-optimal proof system;*
 (b) *Q has an strongly almost optimal algorithm;*
 (c) SList(P, Q, P).

(2) *For every Q we have* (a) $\Leftrightarrow$ (b) *and* (b) $\Rightarrow$ (c)*; moreover* (a), (b), *and* (c) *are all equivalent if Q has padding. Here*
 (a) *Q has a strongly optimal proof system;*
 (b) *Q has a strongly almost optimal nondeterministic algorithm;*
 (c) SList(NP, Q, NP).

5.2 Further variants

Here we present the space variant and the effective variant of the results of Theorem 5.2.

The space variant. First we introduce the notion of space optimal logspace proof system (for simplicity, only the "deterministic optimality") and state the result afterwards.

Definition 5.10

(1) A *logspace proof system* for Q is a logspace algorithm $\mathbb{P}$ computing a function from Σ^* onto Q.

(2) A logspace proof system $\mathbb{P}$ for Q is *space optimal* if for every logspace proof system $\mathbb{P}'$ there is a translation from $\mathbb{P}'$ into $\mathbb{P}$ computable in logspace.

In **[20]** Messner and Torán introduced the notions of logspace proof system and logspace translation. They show the equivalence of the statements "p-optimal proof systems exist", "p-optimal proof systems with respect to logspace translation exist", and "optimal logspace proof systems with respect to polynomial time translation exist". Using this fact, they show that various complexity classes contain problems complete under logspace reductions if p-optimal proof systems exist.

Theorem 5.11 *Assume $Q \subseteq \Sigma^*$ has padding.*

(1) *The following are equivalent:*
  (a) *Q has a space optimal logspace proof system.*
  (b) *Q has an almost space optimal algorithm.*
  (c) $\mathrm{List}(\mathrm{L}, Q, \mathrm{L})$.

(2) *The following are equivalent:*
  (a) *Q has an almost space optimal nondeterministic algorithm.*
  (b) $\mathrm{List}(\mathrm{NL}, Q, \mathrm{NL})$.

Proof. Part (1) and the implication from (a) to (b) in part (2) can be obtained along the lines of our proof of Theorem 5.9 in Section 5.1; however, for the implication from (a) to (b) of part (1), we need a space version of Levin's result, which we state and prove in Section 5.3 as we have not found it in the literature.

We now prove (b) ⇒ (a) in (2). Let $\mathbb{L}$ be a listing witnessing that $\mathrm{List}(\mathrm{NL}, Q, \mathrm{NL})$. It should be clear that the following nondeterministic algorithm $\mathbb{O}$ accepts Q.

```
$\mathbb{O}$   // $x \in \Sigma^*$
        1. guess an $i \in \mathbb{N}$ and compute the $i$th machine $\mathbb{M}_i$ listed by $\mathbb{L}$
        2. guess a $d \in \mathbb{N}$ (in binary)
        3. simulate $\mathbb{M}_i$ on $pad(x, x01^d)$ and output accordingly.
```

We show that $\mathbb{O}$ is almost space optimal. To that end, let $\mathbb{A}$ be a nondeterministic algorithm accepting Q. We consider the subset $LOG(\mathbb{A})$ of Q, where

$$LOG(\mathbb{A}) := \big\{pad(x, x01^d) \bigm| d \in \mathbb{N} \text{ and } \mathbb{A} \text{ accepts } x \text{ using space at most } \log d\big\}.$$

By the properties of a padding function it is easy to show that $LOG(\mathbb{A}) \in \mathrm{NL}$. Therefore, there exists an $i_0 \in \mathbb{N}$ such that the i_0th machine $\mathbb{M}_{i_0}$ listed by $\mathbb{L}$ accepts $LOG(\mathbb{A})$ in space $O(\log n)$; in particular,

$$\begin{aligned} s_{\mathbb{M}_{i_0}}(pad(x, x01^d)) \le O\big(\log |pad(x, x01^d)|\big) &= O\left(\log(|x|^{O(1)} + d^{O(1)})\right) \\ &= O(\log |x| + \log d). \end{aligned} \tag{5.7}$$

The first equality holds as *pad* is computable in logspace and hence, in polynomial time.

Let $x \in Q$. We consider the run of the algorithm $\mathbb{O}$ on input x, where it guesses $i := i_0$ (in Line 1) and $d := 2^{s_{\mathbb{A}}(x)}$(in Line 2). These choices show that

$$s_{\mathbb{O}}(x) \le O\big(c + s_{\mathbb{A}}(x) + \log |x| + \log \big|pad(x, x01^d)\big| + s_{\mathbb{M}_{i_0}}(pad(x, x01^d))\big), \tag{5.8}$$

where c counts the space for guessing i_0 and for computing the machine $\mathbb{M}_{i_0}$. By (5.7) and (5.8) we conclude that

$$s_{\mathbb{O}}(x) \le O(s_{\mathbb{A}}(x) + \log |x|). \qquad \square$$

We will use the following simple observations on listings to generalize some results of the preceding sections from $Q = \textsc{Taut}$ to arbitrary Q with padding. For $Q = \textsc{Taut}$, part (c) of the next proposition has already been shown in **[22]** by completely different means. Recall that $C, C', \dots$ range over the complexity classes L, NL, P and NP.

Proposition 5.12

(a) *Let C, C', and C'' be complexity classes with $C' \subseteq C''$. If* $\mathrm{List}(C, Q, C')$, *then* $\mathrm{List}(C, Q, C'')$.

(b) *Let C, C', and C_0 be complexity classes with $C_0 \subseteq C \subseteq C'$. If* $\mathrm{List}(C, Q, C')$, *then* $\mathrm{List}(C_0, Q, C')$.

(c) *Assume Q has padding. Then*

$$\mathrm{List}(\mathrm{NP}, Q, \mathrm{NP}) \iff \mathrm{List}(\mathrm{P}, Q, \mathrm{NP}).$$

In particular, $\mathrm{List}(\mathrm{P}, Q, \mathrm{P})$ *implies* $\mathrm{List}(\mathrm{NP}, Q, \mathrm{NP})$.

Proof. Part (a) is trivial. For (b), let $\mathbb{M}'$ be a TM of type C' and $\mathbb{M}_0$ one of type C_0. Let $\mathbb{M}'(\mathbb{M}_0)$ be the TM that on input x, first, by brute force, checks whether $\mathbb{M}'$ and $\mathbb{M}_0$ accept the same strings of length $\leq \log\log|x|$; if so, then it simulates $\mathbb{M}'$ on x (and answers accordingly), otherwise it rejects. One easily verifies that $\mathbb{M}'(\mathbb{M}_0)$ is a machine of type C'; furthermore

$$L(\mathbb{M}'(\mathbb{M}_0)) = \begin{cases} L(\mathbb{M}'), & \text{if } L(\mathbb{M}_0) = L(\mathbb{M}'), \\ \text{a finite subset of } L(\mathbb{M}'), & \text{otherwise.} \end{cases}$$

Therefore, if $\mathbb{M}'_1, \mathbb{M}'_2, \ldots$ is a listing of the C-subsets of Q by C'-machines and $\mathbb{M}_1, \mathbb{M}_2, \ldots$ an enumeration of all TMs of type C_0, then the listing $\big(\mathbb{M}'_i(\mathbb{M}_j)\big)_{i\geq 1,\ j\geq 1}$ witnesses that $\mathrm{List}(C_0, Q, C')$.

For (c), let *pad* be a padding function for Q. By (b) it suffices to show the implication from right to left. Hence, we assume that $\mathrm{List}(\mathrm{P}, Q, \mathrm{NP})$. For a nondeterministic Turing machine $\mathbb{M}$, we set

$$\mathit{Comp}(\mathbb{M}) := \big\{\mathit{pad}(x, \langle x, c\rangle) \bigm| x \in \Sigma^* \text{ and } c \text{ is a computation}^3 \text{ of } \mathbb{M} \text{ accepting } x\big\}.$$

Clearly, $\mathit{Comp}(\mathbb{M}) \in \mathrm{P}$; moreover

$$\mathit{Comp}(\mathbb{M}) \subseteq Q \iff L(\mathbb{M}) \subseteq Q. \tag{5.9}$$

Hence, if $\mathbb{M}_1, \mathbb{M}_2, \ldots$ is a listing of the P-subsets of Q by NP-machines, then $\mathbb{M}^*_1, \mathbb{M}^*_2, \ldots$ is a listing of the NP-subsets of Q by NP-machines, where $\mathbb{M}^*_i$ on input x guesses a string c and simulates $\mathbb{M}_i$ on input $\mathit{pad}(x, \langle x, c\rangle)$. □

Now we can prove Corollary 2.7 (a) for arbitrary Q with padding:

Corollary 5.13 *Assume that Q has padding. If Q has an almost optimal algorithm, then it also has an almost optimal nondeterministic algorithm.*

Proof. If Q has an almost optimal algorithm, then $\mathrm{List}(\mathrm{P}, Q, \mathrm{P})$ by Theorem 5.2. Therefore $\mathrm{List}(\mathrm{NP}, Q, \mathrm{NP})$ by part (c) of the previous proposition. Applying again Theorem 5.2 we get the claim. □

We were unable to prove the logspace analogue of Proposition 5.12 (c), however, as a byproduct of our main result of Section 6 relating listings and logics (Theorem 6.1), we will get that (c) holds for $Q = \textsc{Taut}$, that is, that $\mathrm{List}(\mathrm{L}, \textsc{Taut}, \mathrm{NL})$ implies $\mathrm{List}(\mathrm{NL}, \textsc{Taut}, \mathrm{NL})$.

For the set $\mathit{Comp}(\mathbb{M})$ introduced for any nondeterministic TM in the proof of (c) of the previous proposition, we even have $\mathit{Comp}(\mathbb{M}) \in \mathrm{L}$. We use this to obtain the following observation.

[3] That is, c is the sequence of configurations of a run of $\gtrdot$ on x.

Proposition 5.14

(a) *If Q has padding and* $\mathrm{List}(\mathrm{L}, Q, \mathrm{P})$*, then* $\mathrm{List}(\mathrm{P}, Q, \mathrm{P})$*.*

(b) *If Q has padding and* $\mathrm{List}(\mathrm{NL}, Q, \mathrm{NL})$*, then* $\mathrm{List}(\mathrm{NP}, Q, \mathrm{NP})$*.*

Proof. (a) For deterministic TMs $\mathbb{M}$ and $\mathbb{M}'$ let $\mathbb{M}(\mathbb{M}')$ be the TM that on input x, first by simulating $\mathbb{M}'$ on input x stores its sequence c of configurations and then runs $\mathbb{M}$ on input $pad(x, \langle x, c\rangle)$. By (5.9), if $\mathbb{M}_1, \mathbb{M}_2, \dots$ is a listing of the L-subsets of Q by P-machines and $\mathbb{M}'_1, \mathbb{M}'_2, \dots$ an enumeration of all polynomial time TMs, then the enumeration $\big(\mathbb{M}_i(\mathbb{M}'_j)\big)_{i\geq 1,\ j\geq 1}$ witnesses that $\mathrm{List}(\mathrm{P}, Q, \mathrm{P})$.

The proof of (b) is obtained by obvious modifications. □

As a corollary of this proposition we get, using Theorem 5.2 and Theorem 5.9, the following generalization of Corollary 2.11 and Corollary 4.2.

Corollary 5.15 *Assume that Q has padding.*

(a) *If Q has an almost space optimal algorithm, then Q has an almost (time) optimal algorithm.*

(b) *If Q has an almost space optimal nondeterministic algorithm, then Q has an almost (time) optimal algorithm.*

Proof. For (b) note that $\mathrm{List}(\mathrm{NL}, Q, \mathrm{NL})$ implies $\mathrm{List}(\mathrm{L}, Q, \mathrm{P})$ by Proposition 5.12 (a) and (b), and hence it implies $\mathrm{List}(\mathrm{P}, Q, \mathrm{P})$ by part (a) of the previous proposition. □

The effective variant. We start by introducing the effective notions.

Definition 5.16 Let $Q \subseteq \Sigma^*$.

(1) A proof system $\mathbb{P}$ for Q is *effectively p-optimal* if there are two algorithms $\mathbb{T}$ and $\mathbb{B}$ such that:
 (a) $\mathbb{T}$ and $\mathbb{B}$ compute functions defined on $\Sigma^* \times \mathbb{N}[X]$; the values of $\mathbb{T}$ are in Σ^*, that of $\mathbb{B}$ in $\mathbb{N}[X]$;
 (b) for every proof system $\mathbb{P}'$ for Q and every time bound $p' \in \mathbb{N}[X]$ of $\mathbb{P}'$,

$$\mathbb{T}(\mathbb{P}', p')$$

 is (the code of) a translation from $\mathbb{P}'$ into $\mathbb{P}$ and the polynomial $\mathbb{B}(\mathbb{P}', p')$ is a time bound for it.

(2) An algorithm $\mathbb{A}$ deciding Q is *effectively almost optimal* if there is an algorithm $\mathbb{B}$ computing a function defined on Σ^* with values in $\mathbb{N}[X]$ such that for every algorithm $\mathbb{C}$ deciding Q and every $x \in Q$ we have

$$t_{\mathbb{A}}(x) \leq \mathbb{B}(\mathbb{C})\big(t_{\mathbb{C}}(x) + |x|\big).$$

(3) An effective listing of the P-subsets of Q by P-machines is a listing $\mathbb{L}$ of the P-subsets of Q by P-machines such that for some algorithms $\mathbb{I}$ and $\mathbb{B}$ we have:
 (a) $\mathbb{I}$ and $\mathbb{B}$ compute functions defined on $\Sigma^* \times \mathbb{N}[X]$; the values of $\mathbb{I}$ are in $\mathbb{N}$, that of $\mathbb{B}$ in $\mathbb{N}[X]$;
 (b) for every TM $\mathbb{M}$ deciding a P-subset X of Q and every time bound $p \in \mathbb{N}[X]$ for $\mathbb{M}$ the $\mathbb{I}(\mathbb{M}, p)$th TM listed by $\mathbb{L}$ decides X with time bound $\mathbb{B}(\mathbb{M}, p)$.

Using an effective version of Levin's Theorem one gets, essentially by adapting our proof of Theorem 5.9, the following effective analogue of Theorem 5.2:

Theorem 5.17 *For Q with padding the following are equivalent:*

(1) *Q has an effectively p-optimal proof system.*
(2) *Q has an effectively almost optimal algorithm.*
(3) *Q has an effective listing of the* P*-subsets of Q by* P*-machines.*

Let us close this part by stating a corollary of this result obtained from Corollary 2.16 and Corollary 2.17. Part (b) is the "effective" generalization of a result due to Krajíček and Pudlák [**17**].

Corollary 5.18

(a) *If* $\mathrm{NP}[\textsc{tc}] \not\subseteq \mathrm{P}[\textsc{tc}^{\log \textsc{tc}}]$*, then* Taut *has no effectively p-optimal proof system.*
(b) *If* $\mathrm{NE} = \mathrm{E}$*, then* Taut *has an effectively p-optimal proof system.*

5.3 A space version of Levin's result

Theorem 5.19 *Let $\mathbb{F}$ be an algorithm computing a function $f\colon \Sigma^* \to \Sigma^*$. There exists an algorithm $\mathbb{O}$ such that:*

(a) *$\mathbb{O}$ inverts f and $s_{\mathbb{O}}(y) = \infty$ for every input y, which is not in the range of f (in particular, the algorithm $\mathbb{O}$ does not stop on y not in the range).*
(b) *For every algorithm $\mathbb{I}$ inverting f there is an $a \in \mathbb{N}$ such that for every y in the range of f we have*

$$s_{\mathbb{O}}(y) \leq a \cdot \big(\log|y| + s_{\mathbb{I}}(y) + \log|\mathbb{I}(y)| + s_{\mathbb{F}}(\mathbb{I}(y))\big).$$

Proof. First we introduce a notation. If $\mathbb{A}$ and $\mathbb{A}'$ are algorithms computing (partial) functions g and g' from $\Sigma^* \to \Sigma^*$, then by $\mathbb{A};\mathbb{A}'$ we denote an algorithm that computes the function $g' \circ g$, i.e., $x \mapsto g'(g(x))$.

Now let $\mathbb{F}$ be an algorithm computing a function $f\colon \Sigma^* \to \Sigma^*$. Let $\mathbb{O}$ be the TM that on input x for $k = 0, 1, 2, \ldots$ and every TM $\mathbb{M} \in \Sigma^*$ with $|\mathbb{M};\mathbb{F}| \leq k$ simulates at most $|x| \cdot |\mathbb{M};\mathbb{F}| \cdot (k - |\mathbb{M};\mathbb{F}|) \cdot 2^{k-|\mathbb{M};\mathbb{F}|}$ steps of $\mathbb{M};\mathbb{F}$ on input x as long as space $\leq k - |\mathbb{M};\mathbb{F}|$ is required;[4] if the simulation outputs x (that is, $\mathbb{M}$ computes a string w with $f(w) = x$), then it simulates $\mathbb{M}$ on x (outputting the w) and halts.

Then, $\mathbb{O}$ inverts f and $s_{\mathbb{O}}(y) = \infty$ for every input y, which is not in the range of f; hence, $\mathbb{O}$ satisfies (a). We turn to (b) and let $\mathbb{I}$ be any algorithm inverting f. There is $c \in \mathbb{N}$ such that the space required to simulate, given $\mathbb{I};\mathbb{F}$ and x, the algorithm $\mathbb{I};\mathbb{F}$ on input x is less than or equal to

$$c + \log|\mathbb{I};\mathbb{F}| + \log|x| + s_{\mathbb{I};\mathbb{F}}(x). \tag{5.10}$$

We get an upper bound on the space that $\mathbb{O}$ requires on x if we assume that k gets a value such that $|\mathbb{I};\mathbb{F}| \leq k$ and the simulation of $\mathbb{I};\mathbb{F}$ on x can be performed with space at most $k - |\mathbb{I};\mathbb{F}|$. Hence, by (5.10),

$$\begin{aligned} s_{\mathbb{O}}(x) &\leq O\big(|\mathbb{I};\mathbb{F}| + c + \log|\mathbb{I};\mathbb{F}| + \log|x| + s_{\mathbb{I};\mathbb{F}}(x)\big) \\ &\leq O\big(|\mathbb{I};\mathbb{F}| + c + \log|x| + s_{\mathbb{I}}(x) + \log|\mathbb{I}(x)| + s_{\mathbb{F}}(\mathbb{I}(x))\big) \\ &= O\big(\log|x| + s_{\mathbb{I}}(x) + \log|\mathbb{I}(x)| + s_{\mathbb{F}}(\mathbb{I}(x))\big). \end{aligned}$$ □

[4] As $\ell := |\mathbb{M};\mathbb{F}| \leq k$, the algorithm $\mathbb{M};\mathbb{F}$ has at most ℓ states, so on input x we have at most $|x| \cdot \ell \cdot (k - \ell) \cdot 2^{k-\ell}$ distinct configurations using space at most $k - |\mathbb{M};\mathbb{F}|$. So, if $\mathbb{M};\mathbb{F}$ on x halts using at most this space, so will do the simulation.

6 Logics and listings

In the concepts "$L(C)$ is a C'-bounded logic for C" and $\mathrm{List}(C, Q, C')$ two complexity classes, C and C', appear. We show that they match; more precisely, we show:

Theorem 6.1 *Let $C \in \{\mathrm{L}, \mathrm{NL}, \mathrm{P}\}$, $C' \in \{\mathrm{L}, \mathrm{NL}, \mathrm{P}, \mathrm{NP}\}$, and $C \subseteq C'$. Then*

$$L(C)_{\mathrm{inv}} \textit{ is a } C'\textit{-bounded logic for } C \iff \mathrm{List}(C, \textsc{Taut}, C').$$

We remark that one can also define the notion of "almost C'-optimal C-algorithm", an algorithm of type C almost optimal with respect to all C'-algorithms, and analyze their relationship to the notions in the preceding result. For $C, C' \in \{\mathrm{P}, \mathrm{NP}\}$ and listings this has been done in [**22**].

Proof of Theorem 6.1: We assume that $C \subseteq C'$ and that $L(C)_{\mathrm{inv}}$ is a C'-bounded logic for C. Let $\textsc{Prop}$ denote the class of all formulas of propositional logic. For a suitable vocabulary τ in logarithmic space we can associate with every $\alpha \in \textsc{Prop}$ a τ-structure $\mathcal{A}(\alpha)$ such that

(i) every propositional variable X of α corresponds to distinct elements a_X, b_X of $\mathcal{A}(\alpha)$ and there is a unary $P \in \tau$ such that $P^{\mathcal{A}(\alpha)} = \{a_X \mid X \text{ variable of } \alpha\}$;

(ii) the class $\{\mathcal{B} \mid \mathcal{B} \cong \mathcal{A}(\alpha) \text{ for some } \alpha \in \textsc{Prop}\}$ of τ-structures is axiomatizable by a $\mathrm{DTC}[\tau]$-sentence and therefore by an $L(C)[\tau]$-sentence $\varphi(\textsc{Prop})$;

(iii) if $\mathcal{B} \models \varphi(\textsc{Prop})$, then one can determine the unique $\alpha \in \textsc{Prop}$ with $\mathcal{B} \cong \mathcal{A}(\alpha)$ in logarithmic space.

An ordered $\tau_<$-structure of the form $(\mathcal{A}(\alpha), <)$ induces the assignment of the variables of α that sends a variable X to TRUE if $a_X < b_X$. As in logarithmic space we can check whether this assignment satisfies α, there is a $\mathrm{DTC}[\tau_<]$-sentence and hence an $L(C)[\tau_<]$-sentence $\varphi(\mathrm{sat})$ that for every $\alpha \in \textsc{Prop}$ expresses in $(\mathcal{A}(\alpha), <)$ that the assignment given by $<$ satisfies α. We introduce the $L(C)[\tau_<]$-sentence

$$\varphi_0 := \big(\varphi(\textsc{Prop}) \to \varphi(\mathrm{sat})\big).$$

Then φ_0 is an $L(C)_{\mathrm{inv}}[\tau]$-sentence. Every assignment of α can be obtained by some ordering $<$ of $A(\alpha)$. Hence, by the definition of $\models_{L(C)_{\mathrm{inv}}}$, we see that for every $\alpha \in \textsc{Prop}$ and every $L(C)_{\mathrm{inv}}[\tau]$-sentence φ

$$\text{if } \mathcal{A}(\alpha) \models_{L(C)_{\mathrm{inv}}} (\varphi_0 \wedge \varphi), \text{ then } \alpha \in \textsc{Taut}. \tag{6.1}$$

For $\varphi \in L(C)_{\mathrm{inv}}[\tau]$ we consider the class of models of $(\varphi_0 \wedge \varphi)$, more precisely, the set

$$Q(\varphi) := \big\{\alpha \in \textsc{Prop} \bigm| \mathcal{A}(\alpha) \models_{L(C)_{\mathrm{inv}}} (\varphi_0 \wedge \varphi)\big\}.$$

Claim The class of sets $Q(\varphi)$, where φ ranges over all $L(C)_{\mathrm{inv}}$-sentences coincides with the class of C-subsets of $\textsc{Taut}$.

Proof of the claim: First let Q be a C-subset of $\textsc{Taut}$. If Q is finite, it is easy to see that $Q = Q(\varphi)$ for some $\varphi \in L(C)_{\mathrm{inv}}$. Now let Q be infinite. The class $\{\mathcal{B} \mid \mathcal{B} \cong \mathcal{A}(\alpha) \text{ for some } \alpha \in Q\}$ is in C (by (ii) and (iii) as $\mathrm{L} \subseteq C$). As we assume that $L(C)_{\mathrm{inv}}$ is a logic for C, this class is axiomatizable by an $L(C)_{\mathrm{inv}}[\tau]$-sentence φ. As the class contains arbitrarily large structures, the formula φ is invariant. We show that $Q = Q(\varphi)$.

Assume first that $\alpha \in Q(\varphi)$, i.e., $\mathcal{A}(\alpha) \models_{L(C)_{\mathrm{inv}}} (\varphi_0 \wedge \varphi)$. Then, by invariance of φ, we have $\mathcal{A}(\alpha) \models_{L(C)_{\mathrm{inv}}} \varphi$ and thus $\alpha \in Q$. Conversely, assume that $\alpha \in Q$. Then $\mathcal{A}(\alpha) \models_{L(C)_{\mathrm{inv}}} \varphi$. As $\alpha \in \textsc{Taut}$, in order to get $\mathcal{A}(\alpha) \models_{L(C)_{\mathrm{inv}}} (\varphi_0 \wedge \varphi)$ (and hence, $\alpha \in Q(\varphi)$), it suffices to show that $(\varphi_0 \wedge \varphi)$ is $\leq |A(\alpha)|$-invariant. So let $\mathcal{B}$ be a τ-structure with $|B| \leq |A(\alpha)|$. If $\mathcal{B} \not\models_{L(C)_{\mathrm{inv}}} \varphi$, then, by invariance of φ, we have $(\mathcal{B}, <^B) \not\models_{L(C)} (\varphi_0 \wedge \varphi)$ for all orderings $<^B$ on B; if $\mathcal{B} \models_{L(C)_{\mathrm{inv}}} \varphi$, then $\mathcal{B} \cong \mathcal{A}(\beta)$ for some $\beta \in Q \subseteq \textsc{Taut}$. Hence, $(\mathcal{B}, <^B) \models_{L(C)} (\varphi_0 \wedge \varphi)$ for all orderings $<^B$ on B.

We still have to show that $Q(\varphi)$ is a C-subset of $\textsc{Taut}$ for every $\varphi \in L(C)_{\mathrm{inv}}[\tau]$. So we fix $\varphi \in L(C)_{\mathrm{inv}}[\tau]$. By (6.1), $Q(\varphi) \subseteq \textsc{Taut}$. As $L(C)_{\mathrm{inv}}$ is a logic for C, we have $Q(\varphi) \in C$. $\dashv$

As $L(C)_{\mathrm{inv}}$ is a C'-bounded logic for C, the corresponding algorithm $\mathbb{A}$ for the satisfaction relation (cf. Definition 3.2 (b)) restricted to $(\varphi_0 \wedge \varphi)$ with $\varphi \in L(C)_{\mathrm{inv}}[\tau]$ yields an algorithm of type C' accepting $Q(\varphi)$. Thus, by the Claim, the classes $Q(\varphi)$ where φ ranges over all $L(C)_{\mathrm{inv}}[\tau]$-sentences witness that $\mathrm{List}(C, \textsc{Taut}, C')$.

Now let us assume that $\mathrm{List}(C, \textsc{Taut}, C')$. It suffices to show that p-$\textsc{Halt}_> \in \mathrm{XC}'_{\mathrm{uni}}$ as then Theorem 3.1 yields the claim. Since $\mathrm{L} \subseteq C$, we have $\mathrm{List}(\mathrm{L}, \textsc{Taut}, C')$ (by Proposition 5.12 (b)). First assume that C' is a space complexity class. If $C' = \mathrm{L}$, then $\mathrm{List}(\mathrm{L}, \textsc{Taut}, \mathrm{L})$ and therefore $\textsc{Taut}$ has an almost space optimal algorithm (by Theorem 5.11) and hence p-$\textsc{Halt}_> \in \mathrm{XL}_{\mathrm{uni}}$ (by Theorem 2.10).

Let $C' = \mathrm{NL}$. We fix a logspace one-to-one reduction $\langle \mathbb{M}, 1^n \rangle \mapsto \alpha(\mathbb{M}, 1^n)$ from $\textsc{Halt}_>$ (the classical problem underlying p-$\textsc{Halt}_>$) to $\textsc{Taut}$, which is logspace invertible. Furthermore, let $\mathbb{B}$ be an algorithm that on input $\mathbb{M}$, a nondeterministic TM, computes $t_{\mathbb{M}}(\lambda)$, the least k such that there is an accepting run of $\mathbb{M}$ on the empty string λ, by "brute force". Let $\mathbb{L}$ be a listing witnessing $\mathrm{List}(\mathrm{L}, \textsc{Taut}, \mathrm{NL})$. We show that the following nondeterministic algorithm $\mathbb{A}$ witnesses that p-$\textsc{Halt}_> \in \mathrm{XNL}_{\mathrm{uni}}$:

```
$\mathbb{A}$   // $\mathbb{M}$ a nondeterministic TM, $1^n$ with $n \in \mathbb{N}$
    guess $x \in \{\mathbb{B}, \mathbb{L}\}$
    **if** $x = \mathbb{B}$ **then** simulate $\mathbb{B}$ on $\mathbb{M}$
        **if** $\mathbb{B}$ halts **then**
            **if** $n < t_{\mathbb{M}}(\lambda)$ **then** accept
    **if** $x = \mathbb{L}$ **then** guess $i \geq 1$
        simulate $\mathbb{L}$ till it outputs the $i$th machine, say, $\mathbb{C}$
        simulate $\mathbb{C}$ on $\alpha(\mathbb{M}, 1^n)$
        **if** $\mathbb{C}$ accepts **then** accept.
```

Whenever during the simulation of $\mathbb{C}$ on $\alpha(\mathbb{M}, 1^n)$ a bit of the input $\alpha(\mathbb{M}, 1^n)$ is required, the algorithm $\mathbb{A}$ simulates the reduction $\langle \mathbb{M}, 1^n \rangle \mapsto \alpha(\mathbb{M}, 1^n)$ till this bit is obtained.

Clearly $\mathbb{A}$ accepts $\textsc{Halt}_>$. We still have to verify that for some function $f \colon \mathbb{N} \to \mathbb{N}$ and every $\langle \mathbb{M}, 1^n \rangle \in p$-$\textsc{Halt}_>$ we have $s_{\mathbb{A}}(\langle \mathbb{M}, 1^n \rangle) \leq f(|\mathbb{M}|) \cdot \log n$. We fix $\langle \mathbb{M}, 1^n \rangle \in p$-$\textsc{Halt}_>$ and consider the following two cases:

Case "$\langle \mathbb{M}, 1^\ell \rangle \notin p$-$\textsc{Halt}_>$ for some $\ell \in \mathbb{N}$": Then eventually $\mathbb{B}$ will halt on input $\mathbb{M}$. Thus, in the worst case, $\mathbb{A}$ on input $\langle \mathbb{M}, 1^n \rangle$ has to wait till the simulation of $\mathbb{B}$ on $\mathbb{M}$ halts and then $\mathbb{A}$ has to verify that the output $t_{\mathbb{M}}(\lambda)$ of the computation of $\mathbb{B}$ is bigger than n. So the space $s_{\mathbb{A}}(\langle \mathbb{M}, 1^n \rangle)$ can be bounded by $c_{\mathbb{M}} \cdot \log |n|$ for some constant $c_{\mathbb{M}} \in \mathbb{N}$.

Case "$\langle \mathbb{M}, 1^\ell \rangle \in p$-$\textsc{Halt}_>$ for all $\ell \in \mathbb{N}$": Then $\{\alpha(\mathbb{M}, 1^\ell) \mid \ell \in \mathbb{N}\}$ is an L-subset of $\textsc{Taut}$ and hence $\mathbb{L}$ lists a machine $\mathbb{C}$ accepting this set. As $\mathbb{C}$ is logspace, the claim follows immediately.

The cases where C' is a time complexity class are treated similarly using algorithms analogous to those used for time classes in Section 2. □

We already mentioned that we do not know any direct proof of the following result; in particular, we do not know whether we can replace TAUT by any Q with padding.

Corollary 6.2 *If* List(L, TAUT, L)*, then* List(NL, TAUT, NL).

References

[1] A. K. Chandra and D. Harel. Structure and complexity of relational queries. *Journal of Computer and System Sciences*, 25:99–128, 1982.

[2] Y. Chen and J. Flum. On the complexity of Gödel's proof predicate. *The Journal of Symbolic Logic*, 75:239–254, 2009.

[3] Y. Chen and J. Flum. A logic for PTIME and a parameterized halting problem. In *Fields of Logic and Computation*, Lecture Notes in Computer Science 6300, 251–276, 2010.

[4] Y. Chen and J. Flum. On p-optimal proof systems and logics for PTIME. In *Proceedings of the 37th International Colloquium on Automata, Languages and Programming (ICALP'10, Track B)*, Lecture Notes in Computer Science 6199, 321–332, 2010.

[5] Y. Chen and J. Flum. On slicewise monotone parameterized problems and optimal proof systems for TAUT. In *Proceedings of the 19th EACSL Annual Conference in Computer Science Logic (CSL'10)*, Lecture Notes in Computer Science 6247, 200–214, 2010.

[6] Y. Chen and J. Flum. Listings and logics. To appear in In *Proceedings of the 26th Annual IEEE Symposium on Logic in Computer Science (LICS'11)*, 2011.

[7] S. Cook and R. Reckhow. The relative efficiency of propositional proof systems. *The Journal of Symbolic Logic*, 44:36–50, 1979.

[8] R. Fagin. Generalized first-order spectra and polynomial-time recognizable sets. In R. M. Karp (ed.), *Complexity of Computation, SIAM-AMS Proceedings, vol.* 7, 43–73, 1974.

[9] M. Grohe. Fixed-point definability and polynomial time. In *Proceedings of the 23rd International Workshop on Computer Science Logic (CSL'09)*, Lecture Notes in Computer Science 5771, pages 20–23, 2009.

[10] Y. Gurevich. Logic and the challenge of computer science. In *Current Trends in Theoretical Computer Science*, Computer Science Press, 1–57, 1988.

[11] J. Hartmanis and L. Hemachandra. Complexity classes without machines: On complete languages for UP. *Theoretical Computer Science*, 58:129–142, 1988.

[12] N. Immerman. Relational queries computable in polynomial time. *Information and Control*, 68:86–104, 1986.

[13] N. Immerman. Languages that capture complexity classes. *SIAM Journal on Computing*, 16:770–778, 1987.

[14] N. Immerman. Nondeterministic space is closed under complement. *SIAM Journal on Computing*, 17:935–938, 1988.

[15] J. Köbler, J. Messner, and J. Torán. Optimal proof systems imply complete sets for promise classes. *Information and Computation*, 184:71–92, 2003.

[16] W. Kowalczyk. Some connections between presentability of complexity classes and the power of formal systems of reasonning. In *Proceedings of Mathematical Foundations of Computer Science 1984 (MFCS'84)*, Lecture Notes in Computer Science 176, 364–369, 1984.

[17] J. Krajíček and P. Pudlák. Propositional proof systems, the consistency of first order theories and the complexity of computations. *The Journal of Symbolic Logic*, 54:1063–1088, 1989.

[18] L. Levin. Universal search problems. *Problems of Information Transmission*, 9(3):265–266, 1973.

[19] J. Messner. On the simulation order of proof systems. PhD Thesis, University of Erlangen, 2000.

[20] J. Messner and J. Torán. Optimal proof systems for propositional logic and complete sets. In *Proceedings of the 15th Annual Symposium of Theoretical Aspects of Computer Science (STACS'98)*, Lecture Notes in Computer Science 1373, 477–487, 1998.

[21] A. Nash, J. Remmel, and V. Vianu. PTIME queries revisited. In *Proceedings of the 10th International Conference on Database Theory (ICDT'05)*, Lecture Notes in Computer Science 3363, 274–288, 2005.

[22] Z. Sadowski. On an optimal propositional proof system and the structure of easy subsets. *Theoretical Computer Science*, 288(1):181–193, 2002.

[23] M. Y. Vardi. The complexity of relational query languages. In *Proceedings of the 14th Annual ACM Symposium on Theory of Computing (STOC'82)*, 137–146, 1982.

On *p*-optimal proof systems and logics for PTIME

Yijia Chen[*], **Jörg Flum**[†]

[*] Department of Computer Science, Shanghai Jiao Tong University, China
yijia.chen@cs.sjtu.edu.cn

[†] Mathematisches Institut, Albert-Ludwigs-Universität Freiburg, Germany
joerg.flum@math.uni-freiburg.de

Abstract. We prove that TAUT has a p-optimal proof system if and only if a logic related to least fixed-point logic captures polynomial time on all finite structures. Furthermore, we show that TAUT has no *effective* p-optimal proof system if $\mathrm{NTIME}(h^{O(1)}) \not\subseteq \mathrm{DTIME}(h^{O(\log h)})$ for every time constructible and increasing function h.

Introduction

As the title already indicates, this paper relates two topics which at first glance seem to be unrelated. On the one hand we consider optimal proof systems. A *proof system* in the sense of Cook and Reckhow [**6**], say for the class TAUT of tautologies of propositional logic, is a polynomial time computable function defined on $\{0,1\}^*$ and with TAUT as range. A proof system is *p-optimal* if it simulates any other proof system in polynomial time.[1] In their fundamental paper [**13**], Krajíček and Pudlák derive a series of statements equivalent to the existence of a p-optimal proof system for TAUT and state the following conjecture:

Conjecture 1 There is no p-optimal proof system for TAUT.

On the other hand, the question of whether there is a logic capturing polynomial time remains the central open problem in descriptive complexity. There are artificial logics capturing polynomial time, but they do not fulfill a natural requirement to logics in this context:

(0.1) There is an algorithm that decides whether $\mathcal{A}$ is a model of φ for all structures $\mathcal{A}$ and sentences φ of the logic and that does this for fixed φ in time polynomial in the size $\|\mathcal{A}\|$ of $\mathcal{A}$.

If this condition is fulfilled for a logic capturing polynomial time, we speak of a P-bounded logic for P. In [**10**] Gurevich states the following conjecture:

Conjecture 2 There is no P-bounded logic for P.

The conjecture is false if one waives the effectivity condition (0.1). This is shown in [**10**, Section 7, CLAIM 2]) by considering a logic introduced by Blass and Gurevich and which we denote by $L_{\le}$. For any vocabulary the sentences of $L_{\le}$ are the sentences of least fixed-point logic in a vocabulary with an additional binary relation symbol for orderings. In

[*,†] This research has been partially supported by the National Nature Science Foundation of China (60970011), the Sino-German Center for Research Promotion (GZ400), and the John Templeton Foundation through Project # 13152, the Infinity Project at the Centre de Recerca Matemàtica.

[1] All notions will be defined in a precise manner in Section 1.

$L_\leq$ for a structure $\mathcal{A}$ to be a model of φ it is required that in all structures of cardinality less than or equal to that of $\mathcal{A}$, the validity of φ (as a sentence of least fixed-point logic) does not depend on the chosen ordering, and $\mathcal{A}$ with some ordering satisfies φ.

As $L_\leq$ satisfies all requirements of a P-bounded logic for P except (0.1), Gurevich implicitly states the following conjecture:

Conjecture 2a $L_\leq$ is not a P-bounded logic for P.

The main result of this paper (cf. Theorem 2.1) tells us that

$$\text{(0.2)} \qquad \text{Conjecture 1 is true} \iff \text{Conjecture 2a is true.}$$

We mentioned that at first glance "p-optimal proof systems for Taut" and "logics for P" seem to be unrelated topics. However, there are reformulations of Conjecture 1 and Conjecture 2 that are alike. In fact, it is known [**15**] that Taut has a p-optimal proof system if and only if there is a (computable) enumeration of all subsets of Taut that are in P by means of Turing machines that decide them. And it is not hard to see that there is a P-bounded logic for P if and only if there is an enumeration of all polynomial time decidable classes of graphs closed under isomorphisms, again an enumeration in terms of Turing machines that decide these classes. In fact the question for a logic for P was stated in this way by Chandra and Harel [**2**] in the context of an analysis of the complexity and expressiveness of query languages.

Hence one consequence of (0.2) (which we only mention in this Introduction) is:

Theorem 1 *If there is an enumeration of all polynomial time decidable subsets of* Taut*, then there is an enumeration of all polynomial time decidable classes of graphs closed under isomorphisms.*

Using a special feature of the semantics of the logic $L_\leq$, one can construct (cf. Proposition 2.6) a logic that is an *effectively* P-bounded logic for P, if $L_\leq$ is a P-bounded logic for P. Here this "effectively" means that in (0.1) we can *compute* from φ a polynomial bounding the time to decide whether $\mathcal{A}$ is a model of φ. In this way we can strengthen the conclusion of Theorem 1 by requiring that every Turing machine in the enumeration comes with a polynomial time clock. Apparently this is a strengthening, while from any enumeration of the polynomial time decidable subsets of Taut we obtain one with polynomial time clocks in a trivial manner, namely by systematically adding such clocks.

In general, the experts tend to believe Conjecture 1, as the existence of a p-optimal proof system for Taut would have various consequences which seem to be unlikely (see [**12, 13**]). It is worthwhile to emphasize that we show that Conjecture 1 is equivalent to Conjecture 2a and do not claim its equivalence to Conjecture 2. The situation with Conjecture 2 is quite different; no known consequences of the existence of a P-bounded logic for P seem to be implausible. Moreover, due to results showing that there are logics capturing polynomial time on always larger classes of structures, Grohe [**9**] "mildly leans towards believing" that there is a P-bounded logic for P.

In [**3**] we have shown that $L_\leq$ is not an effectively P-bounded logic for P under the assumption $\mathrm{NP}[\textsc{tc}] \not\subseteq \mathrm{P}[\textsc{tc}^{\log \textsc{tc}}]$, which means that $\mathrm{NTIME}(h^{O(1)}) \not\subseteq \mathrm{DTIME}(h^{O(\log h)})$ for every time constructible and increasing function h. Under this assumption, we get (see Theorem 3.2) that Taut has no effectively p-optimal proof system. Here a proof system P for Taut is *effectively* p-optimal if from every other proof system for Taut we can *compute* a polynomial time simulation by P.

On the other hand, Krajíček and Pudlák [**13**] showed, assuming E = NE, that TAUT has a p-optimal proof system. Using our result [**3**] that under the assumption E = NE the logic $\big(L_=$ and hence$\big)$ $L_\leq$ is an effectively P-bounded logic for P, we can derive (see Corollary 3.4) that TAUT has an *effectively* p-optimal proof system if E = NE.

In [**5**] we extract the main idea underlying the proof of (0.2), apply it to other problems, and generalize it to the "nondeterministic case", thus obtaining statements equivalent to the existence of an optimal (not necessarily p-optimal) proof system for TAUT.

1 Preliminaries

In this section we recall concepts and results from complexity theory and logic that we will use later and fix some notation.

1.1 Complexity

We denote the alphabet $\{0,1\}$ by Σ. The length of a string $x \in \Sigma^*$ is denoted by $|x|$. We identify problems with subsets Q of Σ^*. Clearly, as done mostly, we present concrete problems in a verbal, hence uncodified form. We denote by P the class of problems Q such that $x \in Q$ is solvable in polynomial time.

All Turing machines have Σ as their alphabet and are deterministic ones if not stated otherwise explicitly. If necessary we will not distinguish between a Turing machine and its code, a string in Σ^*. If $\mathbb{M}$ is a Turing machine we denote by $\|\mathbb{M}\|$ the length of its code.

By $m^{O(1)}$ we denote the class of polynomially bounded functions from $\mathbb{N}$ to $\mathbb{N}$. Sometimes statements containing a formulation like "there is $d \in \mathbb{N}$ such that for all $x \in \Sigma^*$: $\ldots \leq |x|^d$" can be wrong for $x \in \Sigma^*$ with $|x| \leq 1$. We trust the reader's common sense to interpret such statements reasonably.

1.1.1 Optimal proof systems, almost optimal algorithms and enumerations of P-easy subsets

A *proof system* for a problem $Q \subseteq \Sigma^*$ is a surjective function $P\colon \Sigma^* \to Q$ computable in polynomial time. The proof system P for Q is *polynomially optimal* or *p-optimal* if for every proof system P' for Q there is a polynomial time computable $T\colon \Sigma^* \to \Sigma^*$ such that, for all $w \in \Sigma^*$,

$$P(T(w)) = P'(w).$$

If $\mathbb{A}$ is any algorithm we denote by $t_{\mathbb{A}}(x)$ the number of steps of the run of $\mathbb{A}$ on input x; if $\mathbb{A}$ on x does not stop, then $t_{\mathbb{A}}(x)$ is not defined.

An algorithm $\mathbb{A}$ deciding Q is *almost optimal* or *optimal on positive instances of* Q if for every algorithm $\mathbb{B}$ deciding Q there is a polynomial $p \in \mathbb{N}[X]$ such that for all $x \in Q$

$$t_{\mathbb{A}}(x) \leq p(t_{\mathbb{B}}(x) + |x|)$$

(note that nothing is required of the relationship between $t_{\mathbb{A}}(x)$ and $t_{\mathbb{B}}(x)$ for $x \notin Q$).

By definition a subset Q' of Q is P-*easy* if $Q' \in$ P. An *enumeration of* P-*easy subsets of* Q is a computable function $M\colon \mathbb{N} \to \Sigma^*$ such that

- for every $i \in \mathbb{N}$ the string $M(i)$ is a polynomial time Turing machine deciding a P-easy subset of Q;
- for every P-easy subset Q' of Q there is $i \in \mathbb{N}$ such that $M(i)$ decides Q'.

We denote by Taut the class of tautologies of propositional logic. The following theorem is well-known (cf. [**13**] for the equivalence of the first two statements and [**15**] for the equivalence to the third one):

Theorem 1.1 *The following are equivalent:*

(1) Taut *has a p-optimal proof system.*
(2) Taut *has an almost optimal algorithm.*
(3) Taut *has an enumeration of the* P*-easy subsets.*

1.2 Logic

A *vocabulary* τ is a finite set of relation symbols. Each relation symbol has an *arity.* A *structure* $\mathcal{A}$ of vocabulary τ, or τ*-structure* (or, simply structure), consists of a nonempty set A called the *universe*, and an interpretation $R^{\mathcal{A}} \subseteq A^r$ of each r-ary relation symbol $R \in \tau$. *All structures in this paper are assumed to have finite universe.*

For a structure $\mathcal{A}$ we denote by $\|\mathcal{A}\|$ the size of $\mathcal{A}$, that is, the length of a reasonable encoding of $\mathcal{A}$ as a string in Σ^* (e.g., cf. [**8**] for details). We only consider properties of structures that are invariant under isomorphisms, so it suffices that from the encoding of $\mathcal{A}$ we can recover $\mathcal{A}$ up to isomorphism. We can assume that there is a computable function lgth such that for every vocabulary τ and $m \geq 1$:

- $\|\mathcal{A}\| = \text{lgth}(\tau, m)$ for every τ-structure $\mathcal{A}$ with universe of cardinality m;
- for fixed τ, the function $m \mapsto \text{lgth}(\tau, m)$ is computable in time $m^{O(1)}$;
- $\text{lgth}(\tau \cup \{R\}, m) = O(\text{lgth}(\tau, m) + m^r)$ for every r-ary relation symbol R not in τ.

We assume familiarity with first-order logic and its extension *least fixed-point logic* LFP (e.g. see [**7**]). We denote by LFP$[\tau]$ the set of sentences of vocabulary τ of LFP. As we will introduce further semantics for the formulas of least fixed-point logic, we write $\mathcal{A} \models_{\text{LFP}} \varphi$ if the structure $\mathcal{A}$ is a model of the LFP-sentence φ. An algorithm based on the inductive definition of the satisfaction relation for LFP shows (see [**17**]):

Proposition 1.2 *The model-checking problem $\mathcal{A} \models_{\text{LFP}} \varphi$ for structures $\mathcal{A}$ and* LFP*-sentences φ can be solved in time*

$$\|\mathcal{A}\|^{O(|\varphi|)}.$$

1.2.1 Logics capturing polynomial time

For our purposes, a *logic* L consists of

- an algorithm that for every vocabulary τ and every string ξ decides whether ξ is in the set $L[\tau]$, the set of L*-sentences of vocabulary* τ;
- a *satisfaction relation* $\models_L$; if $(\mathcal{A}, \varphi) \in \models_L$, then $\mathcal{A}$ is a τ-structure and $\varphi \in L[\tau]$ for some vocabulary τ; furthermore for each τ and $\varphi \in L[\tau]$ the class of structures $\mathcal{A}$ with $\mathcal{A} \models_L \varphi$ is closed under isomorphisms.

We say that $\mathcal{A}$ is a *model* of φ if $\mathcal{A} \models_L \varphi$ $\big($that is, if $(\mathcal{A}, \varphi) \in \models_L\big)$. We set $\text{Mod}_L(\varphi) := \big\{\mathcal{A} \mid \mathcal{A} \models_L \varphi\big\}$ and say that φ *axiomatizes* the class $\text{Mod}_L(\varphi)$.

Definition 1.3 Let L be a logic.

(a) *L is a logic for* P if for all vocabularies τ and all classes C (of encodings) of τ-structures closed under isomorphisms we have

$$C \in \text{P} \iff C = \text{Mod}_L(\varphi) \text{ for some } \varphi \in L[\tau].$$

(b) L *is a* P-*bounded logic for* P if (a) holds and if there is an algorithm $\mathbb{A}$ deciding $\models_L$ (that is, for every structure $\mathcal{A}$ and L-sentence φ the algorithm $\mathbb{A}$ decides whether $\mathcal{A} \models_L \varphi$) and if moreover $\mathbb{A}$, for every fixed φ, polynomial in $\|\mathcal{A}\|$.

Hence, if L is a P-bounded logic for P, then for every L-sentence φ the algorithm $\mathbb{A}$ witnesses that $\mathrm{Mod}_L(\varphi) \in \mathrm{P}$. However, we do not necessarily know ahead of time a bounding polynomial.

(c) L *is an effectively* P-*bounded logic for* P if L is a P-bounded logic for P and if in addition to the algorithm $\mathbb{A}$ as in (b) there is a computable function that assigns to every L-sentence φ a polynomial $q \in \mathbb{N}[X]$ such that $\mathbb{A}$ decides whether $\mathcal{A} \models_L \varphi$ in $\leq q(\|\mathcal{A}\|)$ steps.

1.2.2 The logic $L_\leq$ and invariant sentences

In this section we introduce the logic $L_\leq$, a variant of least fixed-point logic.

For every vocabulary τ we let $\tau_< := \tau \cup \{<\}$, where $<$ is a binary relation symbol not in τ chosen in some canonical way. We set

$$L_\leq[\tau] = \mathrm{LFP}[\tau_<]$$

for every vocabulary τ. Before we define the satisfaction relation for $L_\leq$ we introduce the notion of $\leq m$-invariant sentence.

Definition 1.4 Let φ be an $L_\leq[\tau]$-sentence.

- For $m \geq 1$ we say that φ is $\leq m$*-invariant* if for all structures $\mathcal{A}$ with $|A| \leq m$ and all orderings $<_1$ and $<_2$ on A we have
$$(\mathcal{A}, <_1) \models_{\mathrm{LFP}} \varphi \iff (\mathcal{A}, <_2) \models_{\mathrm{LFP}} \varphi.$$
- φ is *invariant* if it is $\leq m$-invariant for all $m \geq 1$.

Finally we introduce the semantics for the logic $L_\leq$ by

$$\mathcal{A} \models_{L_\leq} \varphi \iff \Big(\varphi \text{ is} \leq |A|\text{-invariant and } (\mathcal{A}, <) \models_{\mathrm{LFP}} \varphi \text{ for some ordering} < \text{on } A\Big).$$

Immerman [**11**] and Vardi [**16**] have shown that LFP is an effectively P-bounded logic for P *on the class of ordered structures*, a result we will not need in the proof of our main theorem. However, using it one can easily show that $L_\leq$ is a logic for P.

For later purposes we remark that for every $L_\leq[\tau]$-sentence φ and $m \geq 1$ we have

$$\varphi \text{ is} \leq m\text{-invariant} \iff \neg\varphi \text{ is} \leq m\text{-invariant},$$

and thus, for every τ-structure $\mathcal{A}$,

$$\varphi \text{ is} \leq |A|\text{-invariant} \iff \big(\mathcal{A} \models_{L_\leq} \varphi \text{ or } \mathcal{A} \models_{L_\leq} \neg\varphi\big).$$

In particular,

$$\varphi \text{ is} \leq m\text{-invariant} \iff \big(\mathcal{A}(\tau, m) \models_{L_\leq} \varphi \text{ or } \mathcal{A}(\tau, m) \models_{L_\leq} \neg\varphi\big),$$

where $\mathcal{A}(\tau, m)$ is the τ-structure with universe $\{1, \ldots, m\}$, where every relation symbol in τ is interpreted by the empty relation of the corresponding arity.

Finally we remark that it can happen for $L_\leq$-sentences φ and ψ and a structure $\mathcal{A}$ that $\mathcal{A} \models_{L_\leq} (\varphi \wedge \psi)$ but neither $\mathcal{A} \models_{L_\leq} \varphi$ nor $\mathcal{A} \models_{L_\leq} \psi$.

2 Main theorem

In this section we want to show:

Theorem 2.1 TAUT *has a p-optimal proof system iff* $L_{\leq}$ *is a* P*-bounded logic for* P.

In view of Theorem 1.1 we get one direction of Theorem 2.1 with the following lemma.

Lemma 2.2 *If* $L_{\leq}$ *is a* P*-bounded logic for* P*, then there is an enumeration of the* P*-easy subsets of* TAUT.

Proof. It is easy to introduce a vocabulary τ such that in polynomial time we can associate with every propositional formula α a τ-structure $\mathcal{A}(\alpha)$ such that

- every propositional variable X of α corresponds to two distinct elements a_X, b_X of $\mathcal{A}(\alpha)$ and there is a unary relation symbol $P \in \tau$ such that $P^{\mathcal{A}(\alpha)} = \{a_X \mid X \text{ variable of } \alpha\}$;
- there is an LFP-sentence $\varphi(\text{PROP})$ of vocabulary τ axiomatizing the class
$$\{\mathcal{B} \mid \mathcal{B} \cong \mathcal{A}(\alpha) \text{ for some } \alpha \in \text{PROP}\}$$
(by PROP we denote the class of formulas of propositional logic);
- if $\mathcal{B} \models \varphi(\text{PROP})$, then one can determine the unique $\alpha \in \text{PROP}$ with $\mathcal{B} \cong \mathcal{A}(\alpha)$ in polynomial time.

Again let $\tau_< := \tau \cup \{<\}$ with a new binary $<$. Note that a $\tau_<$-structure of the form $(\mathcal{A}(\alpha), <)$ yields an assignment of the variables of α, namely the assignment sending a variable X to TRUE if and only if $a_X < b_X$. There is an LFP$[\tau_<]$-formula $\varphi(\text{sat})$ that for every $\alpha \in \text{PROP}$ expresses in $(\mathcal{A}(\alpha), <)$ that the assignment given by $<$ satisfies α.

We introduce the $L_{\leq}[\tau]$-sentence
$$\varphi_0 := \big(\varphi(\text{PROP}) \rightarrow \varphi(\text{sat})\big).$$
By the definition of $\models_{L_{\leq}}$ we see that for every $\alpha \in \text{PROP}$ and every $L_{\leq}[\tau]$-sentence φ
$$\text{(2.1)} \qquad \text{if } \mathcal{A}(\alpha) \models_{L_{\leq}} (\varphi_0 \wedge \varphi), \text{ then } \alpha \in \text{TAUT}.$$
We claim that the class of models of $(\varphi_0 \wedge \varphi)$, more precisely,
$$Q(\varphi) := \big\{\alpha \in \text{PROP} \mid \mathcal{A}(\alpha) \models_{L_{\leq}} (\varphi_0 \wedge \varphi)\big\},$$
where φ ranges over all $L_{\leq}[\tau]$-sentences, yields the desired enumeration of P-easy subsets of TAUT. By (2.1), we have $Q(\varphi) \subseteq \text{TAUT}$.

For $\varphi \in L_{\leq}[\tau]$ let the Turing machine $\mathbb{M}_\varphi$, given an input $\alpha \in \text{PROP}$, first construct $\mathcal{A}(\alpha)$ and then check whether $\mathcal{A}(\alpha) \models_{L_{\leq}} (\varphi_0 \wedge \varphi)$. Clearly, $\mathbb{M}_\varphi$ decides $Q(\varphi)$ and does this in polynomial time, as $L_{\leq}$ is a P-bounded logic for P.

Conversely, let Q be a P-easy subset of TAUT. If Q is finite, it is easy to see that $Q = Q(\varphi)$ for some $\varphi \in L_{\leq}[\tau]$. Now let Q be infinite. The class
$$\big\{\mathcal{B} \mid \mathcal{B} \cong \mathcal{A}(\alpha) \text{ for some } \alpha \in Q\big\}$$
is in P, and therefore it is axiomatizable by an $L_{\leq}[\tau]$-sentence φ. As the class contains arbitrarily large structures, the formula φ is invariant. We show that $Q = Q(\varphi)$.

Assume first that $\alpha \in Q(\varphi)$, i.e., $\mathcal{A}(\alpha) \models_{L_{\leq}} (\varphi_0 \wedge \varphi)$. Then, by invariance of φ, we have $\mathcal{A}(\alpha) \models_{L_{\leq}} \varphi$ and thus $\alpha \in Q$. Conversely, assume that $\alpha \in Q$. Then $\mathcal{A}(\alpha) \models_{L_{\leq}} \varphi$. As $\alpha \in \text{TAUT}$, in order to get $\mathcal{A}(\alpha) \models_{L_{\leq}} (\varphi_0 \wedge \varphi)$ $\big($and hence, $\alpha \in Q(\varphi)\big)$ it suffices to show that $(\varphi_0 \wedge \varphi)$ is $\leq |A(\alpha)|$-invariant. So let $\mathcal{B}$ be a τ-structure with $|B| \leq |A(\alpha)|$. If

$\mathcal{B} \not\models_{L_\leq} \varphi$, then, by invariance of φ, we have $(\mathcal{B}, <^B) \not\models_{\text{LFP}} (\varphi_0 \wedge \varphi)$ for all orderings $<^B$ on B; if $\mathcal{B} \models_{L_\leq} \varphi$, then $\mathcal{B} \cong \mathcal{A}(\beta)$ for some $\beta \in Q \subseteq$ TAUT. Hence, $(\mathcal{B}, <^B) \models_{\text{LFP}} (\varphi_0 \wedge \varphi)$ for all orderings $<^B$ on B. □

Remark 2.3 In the previous proof we have used the definition of the satisfaction relation $\models_{L_\leq}$ in order to express the universal second-order quantifier in the statement "all assignments satisfy α". Similarly, we can do with every Π^1_1-sentence $\forall R\varphi$, where φ is a first-order formula or (equivalently) LFP-formula and show in this way that there is an enumeration of the P-easy subsets closed under isomorphisms of the class of models of $\forall R\varphi$, if $L_\leq$ is a P-bounded logic for P. In fact, let k be the arity of R. If a structure $\mathcal{A}$ has n elements, we consider a structure $\mathcal{B}$ with additional disjoint unary relations $U^{\mathcal{B}}, P_0^{\mathcal{B}}, P_1^{\mathcal{B}}$ such that

$$B = U^{\mathcal{B}} \cup P_0^{\mathcal{B}} \cup P_1^{\mathcal{B}}, \quad U^{\mathcal{B}} = A, \quad |P_0^{\mathcal{B}}| = n^k \quad |P_1^{\mathcal{B}}| = n^k$$

and with an ordering $<^{\mathcal{B}}$.

With the elements in $P_0^{\mathcal{B}}$ interpreted as 0s and the elements in $P_1^{\mathcal{B}}$ interpreted as 1s, the first n^k-elements of the ordering in $P_0^{\mathcal{B}} \cup P_1^{\mathcal{B}}$ represent a natural number $< 2^{n^k}$ and thus a k-ary relation R on A, which we can compute in polynomial time (polynomial in n); hence we can define R by an LFP-formula. As in this way, by changing the ordering, we have access to all such k-ary relations R on A, we can express the quantifier $\forall R$ using the invariance requirement of $\models_{L_\leq}$.

For example, let C be the class of all pairs $(\mathcal{G}, \mathcal{H})$ of graphs such that $\mathcal{H}$ is *not* a homomorphic image of $\mathcal{G}$. By the previous observation, we see that there is an enumeration of the P-easy subclasses of C closed under isomorphisms if $L_\leq$ is a P-bounded logic for P. Of course, a subclass D of C is closed under isomorphisms if

$$\mathcal{G} \cong \mathcal{G}', \ \ \mathcal{H} \cong \mathcal{H}' \ \text{ and } \ (\mathcal{G}, \mathcal{H}) \in D \ \text{ imply } \ (\mathcal{G}', \mathcal{H}') \in D.$$

As the models of such a Π^1_1-sentence correspond to a problem Q in co-NP, a simple complexity-theoretic argument shows that there is an enumeration of the P-easy subsets of Q provided there is one for the P-easy subsets of TAUT (see also [**1**]). However, in this way, in the previous example we would not get an enumeration of those P-easy subclasses that are *closed under isomorphisms.*

In view of Theorem 1.1 the remaining direction in Theorem 2.1 is provided by the following result.

Lemma 2.4 *If* TAUT *has an almost optimal algorithm, then $L_\leq$ is a* P*-bounded logic for* P.

Proof. We assume that TAUT has an almost optimal algorithm $\mathbb{O}$ and have to show that there is an algorithm that decides $\mathcal{B} \models_{L_\leq} \varphi$ and does this for fixed φ in time $\|\mathcal{B}\|^{O(1)}$.

By the definition of $\mathcal{B} \models_{L_\leq} \varphi$ and Proposition 1.2 it suffices to show the existence of an algorithm $\mathbb{A}$ that for every $L_\leq$-sentence φ and every $m \in \mathbb{N}$ decides whether φ is $\leq m$-invariant and does this for fixed φ in time $m^{O(1)}$.

We set

$$Q := \Big\{ \Big(\chi, \ell, \text{lgth}(\tau, \ell)^{|\chi|}\Big) \ \Big| \ \tau \text{ a vocabulary, } \chi \in \text{LFP}[\tau], \ \ell \geq 1, \ \text{lgth}(\tau, \ell)^{|\chi|} \\ \text{in unary, there is a } \tau\text{-structure } \mathcal{B} \text{ with } \big(|B| \leq \ell \text{ and } \mathcal{B} \models_{\text{LFP}} \chi\big) \Big\}$$

(compare Section 1.2 for the definition of the function lgth). By Proposition 1.2, $Q \in \mathrm{NP}$. Thus there is a polynomial time reduction $R : Q \leq^p \textsc{Sat}$. We can assume that from $R(x)$ we can recover x in polynomial time. Let φ be an $L_{\leq}[\tau]$-sentence. Then

$$\begin{aligned}\varphi \text{ is not } \leq m\text{-invariant} \iff & \text{ there is a } \tau\text{-structure } \mathcal{B} \text{ and orderings } <_1, <_2 \text{ with} \\ & \big(|B| \leq m \text{ and } (\mathcal{B}, <_1, <_2) \models_{\mathrm{LFP}} \underbrace{(\varphi(<_1) \wedge \neg\varphi(<_2))}_{\varphi^*}\big) \\ \iff & \Big(\varphi^*, m, \mathrm{lgth}(\tau \cup \{<_1, <_2\}, m)^{|\varphi^*|}\Big) \in Q \\ \iff & R\Big(\varphi^*, m, \mathrm{lgth}(\tau \cup \{<_1, <_2\}, m)^{|\varphi^*|}\Big) \in \textsc{Sat}.\end{aligned}$$

We set $\alpha(\varphi, m) := R\left(\varphi^*, m, \mathrm{lgth}(\tau \cup \{<_1, <_2\}, m)^{|\varphi^*|}\right)$. Hence

$$\varphi \text{ is } \leq m\text{-invariant} \iff \neg\alpha(\varphi, m) \in \textsc{Taut}. \tag{2.2}$$

It is clear that there is an algorithm that on input (φ, m) computes $\alpha(\varphi, m)$ and for fixed φ

$$\text{it computes } \alpha(\varphi, m) \text{ in time } m^{O(1)}, \text{ in particular, } |\alpha(\varphi, m)| \leq m^{O(1)}, \tag{2.3}$$

as, for fixed τ, the function $m \mapsto \mathrm{lgth}(\tau, m)$ is polynomial in m.

Let $\mathbb{S}$ be the algorithm that on input φ by systematically going through all τ-structures with universe $\{1\}$, all with universe $\{1, 2\}, \ldots$ and all orderings of these universes computes $m(\varphi) :=$ the least m such that φ is not $\leq m$-invariant. If φ is invariant, then $m(\varphi)$ is not defined and $\mathbb{S}$ does not stop.

We show that the following algorithm $\mathbb{A}$ has the desired properties.

```
$\mathbb{A}(\varphi, m)$
// $\varphi$ an $L_{\leq}$-sentence, $m \in \mathbb{N}$
        Compute $\alpha(\varphi, m)$
        In parallel simulate $\mathbb{S}$ on input $\varphi$ and $\mathbb{O}$ on input $\neg\alpha(\varphi, m)$
        if $\mathbb{O}$ stops first, then output its answer
        if $\mathbb{S}$ stops first, then
            if $m < m(\varphi)$ then accept else reject.
```

By our assumptions on $\mathbb{O}$ and $\mathbb{S}$ and by (2.2), it should be clear that $\mathbb{A}$ on input (φ, m) decides if φ is $\leq m$-invariant. We have to show that for fixed φ it does it in time $m^{O(1)}$.

Case "φ is invariant": Then for all m we have $\neg\alpha(\varphi, m) \in \textsc{Taut}$. Thus the following algorithm $\mathbb{O}_\varphi$ decides $\textsc{Taut}$: on input $\beta \in \textsc{Prop}$ the algorithm $\mathbb{O}_\varphi$ checks whether $\beta = \neg\alpha(\varphi, m)$ for some $m \geq 1$. If so, it accepts and otherwise it runs $\mathbb{O}$ on input β and answers accordingly. By (2.3), we have

$$t_{\mathbb{O}_\varphi}(\neg\alpha(\varphi, m)) \leq m^{O(1)}. \tag{2.4}$$

As $\mathbb{O}$ is optimal, we know that there is a constant d such that for all $\beta \in \textsc{Taut}$

$$t_{\mathbb{O}}(\beta) \leq \big(|\beta| + t_{\mathbb{O}_\varphi}(\beta)\big)^d. \tag{2.5}$$

In particular, we have

$$t_{\mathbb{O}}(\neg\alpha(\varphi, m)) \leq \big(|\neg\alpha(\varphi, m)| + t_{\mathbb{O}_\varphi}(\neg\alpha(\varphi, m))\big)^d \leq m^{O(1)}.$$

By this inequality, (2.3) and (2.4), we see that for invariant φ we have $t_{\mathbb{A}}(\varphi, m) \leq m^{O(1)}$.

Case "φ is not invariant": Then $\mathbb{S}$ will stop on input φ. Thus, in the worst case, $\mathbb{A}$ on input (φ, m) has to wait till the simulation of $\mathbb{S}$ on φ stops and then must check whether the result $m(\varphi)$ of the computation of $\mathbb{S}$ is bigger than m or not and answer accordingly. So the algorithm $\mathbb{A}$ at most takes time $m^{O(1)} + O(t_{\mathbb{S}}(\varphi) + m) \leq m^{O(1)}$ (note that we fix φ, so that $t_{\mathbb{S}}(\varphi)$ is a constant). □

Corollary 2.5 *If* Taut *has a p-optimal proof system, then there is an effectively* P*-bounded logic for* P*.*

This result follows from Theorem 2.1 using the following proposition:

Proposition 2.6 *If $L_{\leq}$ is a* P*-bounded logic for* P*, then there is an effectively* P*-bounded logic for* P*.*

Proof. In Section 1.2 we have seen that for every $L_{\leq}$-sentence φ and $m \geq 1$ it holds that

(2.6) $$\varphi \text{ is } \leq m\text{-invariant} \iff \big(\mathcal{A}(\tau, m) \models_{L_{\leq}} \varphi \text{ or } \mathcal{A}(\tau, m) \models_{L_{\leq}} \neg\varphi\big),$$

where $\mathcal{A}(\tau, m)$ denotes the "empty structure" of vocabulary τ with universe $\{1, \ldots, m\}$.

Now assume that $L_{\leq}$ is a P-bounded logic for P and let $\mathbb{A}$ be an algorithm witnessing that $L_{\leq}$ is a P-bounded logic for P. By (2.6), there is a function h assigning to every $L_{\leq}$-sentence φ a polynomial $h(\varphi) \in \mathbb{N}[X]$ such that $\mathbb{A}$ decides whether φ is $\leq m$-invariant in time $h(\varphi)(m)$.

We consider the logic $T(L_{\leq})$, *time-clocked* $L_{\leq}$, defined as follows:

- for every vocabulary τ
$$T(L_{\leq})[\tau] := \big\{(\varphi, p) \mid \varphi \in L_{\leq}[\tau] \text{ and } p \in \mathbb{N}[X]\big\};$$
- $\mathcal{A} \models_{T(L_{\leq})} (\varphi, p)$ iff (a) and (b) are fulfilled, where
  (a) $\mathbb{A}$ shows via (2.6) in $\leq p(|A|)$ steps that φ is $\leq |A|$-invariant;
  (b) $(\mathcal{A}, <) \models_{\text{LFP}} \varphi$ for some ordering $<$, say with the ordering of A given by the encoding of $\mathcal{A}$.

It is not hard to verify that $T(L_{\leq})$ is an effectively P-bounded logic for P. □

Remark 2.7 In a slightly different way but using the same idea one can define the time-clocked version $T(L)$ for any P-bounded logic L for P. However, in general, $T(L)$ is not even a logic, as it can happen that the class of models of a $T(L)$-sentence is not closed under isomorphisms. In the case of $T(L_{\leq})$ this is guaranteed by the fact that condition (a) in the definition of $\mathcal{A} \models_{T(L_{\leq})} (\varphi, p)$ only refers to the cardinality of the universe of $\mathcal{A}$.

There is a further consequence of Theorem 2.1. By a reformulation of the statement "$L_{\leq}$ is a P-bounded logic for P" due to Nash et al. [**14**] (see [**3**] for a proof), we get:

Theorem 2.8 *The following are equivalent:*

(a) Taut *has a p-optimal proof system.*

(b) *There is an algorithm deciding for every nondeterministic Turing machine $\mathbb{M}$ and every natural number m whether $\mathbb{M}$ accepts the empty input tape in $\leq m$ steps and the algorithm does this for every fixed $\mathbb{M}$ in time $m^{O(1)}$.*

3 Effective versions

Let $\mathrm{NP}[\textsc{tc}] \not\subseteq \mathrm{P}[\textsc{tc}^{\log \textsc{tc}}]$ mean that $\mathrm{NTIME}(h^{O(1)}) \not\subseteq \mathrm{DTIME}(h^{O(\log h)})$ for every time constructible and increasing function h. In [**3**] we have shown:

Proposition 3.1 *Assume that* $\mathrm{NP}[\textsc{tc}] \not\subseteq \mathrm{P}[\textsc{tc}^{\log \textsc{tc}}]$. *Then* $L_{\leq}$ *is not an effectively* P*-bounded logic for* P.

Are there *natural* effective versions of the properties of Taut listed in Theorem 1.1 equivalent to the statement "$L_{\leq}$ is not an effectively P-bounded logic for P" and which therefore, by Proposition 3.1, could not hold under the assumption $\mathrm{NP}[\textsc{tc}] \not\subseteq \mathrm{P}[\textsc{tc}^{\log \textsc{tc}}]$? We did not find them. However, by analyzing the proof of Proposition 3.1, we isolate a property of an effective P-bounded logic for P that cannot be fulfilled if $\mathrm{NP}[\textsc{tc}] \not\subseteq \mathrm{P}[\textsc{tc}^{\log \textsc{tc}}]$. It turns out that this is equivalent to natural effective versions of the properties on Taut under consideration. We already state the result we aim at and then define the concepts appearing in it and present the generalization of Theorem 1.1 on which its proof is based. Due to space limitations all proofs of results in this section will be given in the full version of the paper.

Theorem 3.2 *If* $\mathrm{NP}[\textsc{tc}] \not\subseteq \mathrm{P}[\textsc{tc}^{\log \textsc{tc}}]$*, then* Taut *has no effectively p-optimal proof system.*

Let $Q \subseteq \Sigma^*$. A proof system P for Q is *effectively p-optimal* if there are two computable functions $S\colon \Sigma^* \times \mathbb{N}[X] \to \Sigma^*$ and $b\colon \Sigma^* \times \mathbb{N}[X] \to \mathbb{N}[X]$ such that for every proof system P' for Q with time bound $p \in \mathbb{N}[X]$ and every $w' \in \Sigma^*$, we have

$$P'(w') = P\big(S(P', p)(w')\big),$$

where $S(P', p)$ is (the code of) a Turing machine with time bound $b(P', p)$ and $S(P', p)(w')$ denotes the output of $S(P', p)$ on input w'.

An algorithm $\mathbb{A}$ deciding Q is *effectively almost optimal* if there is a computable function $b\colon \Sigma^* \to \mathbb{N}[X]$ such that for every algorithm $\mathbb{B}$ deciding Q we have for every $x \in Q$ we have

$$t_{\mathbb{A}}(x) \leq b(\mathbb{B})\big(t_{\mathbb{B}}(x) + |x|\big).$$

We say that Q has an *effective enumeration of* P*-easy subsets*, if it has an enumeration $M\colon \mathbb{N} \to \Sigma^*$ of P-easy subsets of Q such that there are functions $I\colon \Sigma^* \times \mathbb{N}[X] \to \mathbb{N}$ and $b\colon \Sigma^* \times \mathbb{N}[X] \to \mathbb{N}[X]$ such that for every Turing machine $\mathbb{M}$ and polynomial $p \in \mathbb{N}[X]$,

> if the Turing machine $\mathbb{M}$ recognizes a subset $Q' \subseteq Q$ with time bound p,
> then the machine $M(I(\mathbb{M}, p))$ recognizes Q' with time bound $b(\mathbb{M}, p)$.

We can prove the effective analogue of Theorem 1.1:

Theorem 3.3 *The following are equivalent:*

(1) Taut *has an effectively p-optimal proof system.*
(2) Taut *has an effectively almost optimal algorithm.*
(3) Taut *has an effective enumeration of the P-easy subsets.*

In [**3**] we have shown that if $\mathrm{E} = \mathrm{NE}$, then (the logic $L_{=}$ and hence) $L_{\leq}$ are effectively P-bounded logics for P. The proof of the previous result shows that Taut has an effectively p-optimal proof system if $L_{\leq}$ is an effectively P-bounded logic for P. Therefore we obtain the following "effective version" of a result due to Krajíček and Pudlák.

Corollary 3.4 *If* $\mathrm{E} = \mathrm{NE}$*, then* Taut *has an effectively p-optimal proof system.*

References

[1] O. Beyersdorff and Z. Sadowski. Characterizing the existence of optimal proof systems and complete sets for promise classes. In *Proceedings of the 4th Computer Science Symposium in Russia (CSR'09)*, Lecture Notes in Computer Science 5675, 47–58, 2009.

[2] A. K. Chandra and D. Harel. Structure and complexity of relational queries. *Journal of Computer and System Sciences*, 25:99–128, 1982.

[3] Y. Chen and J. Flum. A logic for PTIME and a parameterized halting problem. In *Proceedings of the 24th IEEE Symposium on Logic in Computer Science*, pages 397–406, 2009.

[4] Y. Chen and J. Flum. On the complexity of Gödel's proof predicate. *The Journal of Symbolic Logic*, 75(1): 239–254, 2010.

[5] Y. Chen and J. Flum. On slicewise monotone parameterized problems and optimal proof systems for TAUT. Available at `http://basics.sjtu.edu.cn/~chen/papers`, 2010.

[6] S. Cook and R. Reckhow. The relative efficiency of propositional proof systems. *The Journal of Symbolic Logic*, 44:36–50, 1979.

[7] H.-D. Ebbinghaus and J. Flum. *Finite Model Theory*, 2nd edition, Springer, 1999.

[8] J. Flum and M. Grohe. *Parameterized Complexity Theory*, Springer, 2006.

[9] M. Grohe. Fixed-point definability and polynomial time. In *Proceedings of the 23rd International Workshop on Computer Science Logic (CSL'09)*, Lecture Notes in Computer Science 5771, pages 20–23, 2009.

[10] Y. Gurevich. Logic and the challenge of computer science. In *Current Trends in Theoretical Computer Science*, Computer Science Press, 1–57, 1988.

[11] N. Immerman. Relational queries computable in polynomial time. *Information and Control*, 68:86–104, 1986.

[12] J. Köbler, J. Messner, and J. Torán. Optimal proof systems imply complete sets for promise classes. *Information and Computation*, 184:71–92, 2003.

[13] J. Krajíček and P. Pudlák. Propositional proof systems, the consistency of first order theories and the complexity of computations. *The Journal of Symbolic Logic*, 54:1063–1088, 1989.

[14] A. Nash, J. Remmel, and V. Vianu. PTIME queries revisited. In *Proceedings of the 10th International Conference on Database Theory (ICDT'05)*, T. Eiter and L. Libkin (eds.), Lecture Notes in Computer Science 3363, 274–288, 2005.

[15] Z. Sadowski. On an optimal propositionl proof system and the structure of easy subsets. *Theoretical Computer Science*, 288(1):181–193, 2002.

[16] M. Y. Vardi. The complexity of relational query languages. In *Proceedings of the 14th ACM Symposium on Theory of Computing (STOC'82)*, pages 137–146, 1982.

[17] M. Y. Vardi. On the complexity of bounded-variable queries. In *Proceedings of the 14th ACM Symposium on Principles of Database Systems (PODS'95)*, pages 266–276, 1995.

On slicewise monotone parameterized problems and optimal proof systems for TAUT

Yijia Chen*, **Jörg Flum**†

* Department of Computer Science, Shanghai Jiao Tong University, China
yijia.chen@cs.sjtu.edu.cn

† Mathematisches Institut, Albert-Ludwigs-Universität Freiburg, Germany
joerg.flum@math.uni-freiburg.de

Abstract. For a reasonable sound and complete proof calculus for first-order logic consider the problem to decide, given a sentence φ of first-order logic and a natural number n, whether φ has no proof of length $\leq n$. We show that there is a nondeterministic algorithm accepting this problem which, for fixed φ, has running time bounded by a polynomial in n if and only if there is an optimal proof system for the set TAUT of tautologies of propositional logic. This equivalence is an instance of a general result linking the complexity of so-called slicewise monotone parameterized problems with the existence of an optimal proof system for TAUT.

Introduction

In this paper we relate the existence of optimal proof systems for the class TAUT of tautologies of propositional logic with the complexity of slicewise monotone parameterized problems. A *proof system* in the sense of Cook and Reckhow [**4**], say for the class TAUT, is a polynomial time computable function defined on $\{0,1\}^*$ and with TAUT as range. A proof system P is *optimal* if for any other proof system P' for TAUT there is a polynomial $p \in \mathbb{N}[X]$ such that for every tautology α, if α has a proof of length n in P', then α has a proof of length $\leq p(n)$ in P.[1] In their fundamental paper [**9**] Krajíček and Pudlák showed that an optimal proof system for TAUT exists if NE = co-NE and they derived a series of statements equivalent to the existence of such an optimal proof system; however they conjectured that there is no optimal proof system for TAUT.

On the other hand, Gödel in a letter to von Neumann of 1956 (see [**6**]) asked for the complexity of the problem to decide, given a sentence φ of first-order logic and a natural number n, whether φ has a proof of length $\leq n$. In our study [**2**] of this problem we introduced the parameterized problem

p-GÖDEL	
Instance:	A first-order sentence φ and $n \in \mathbb{N}$ in unary.
Parameter:	$|\varphi|$.
Problem:	Does φ have a proof of length $\leq n$?

Here we refer to any reasonable sound and complete proof calculus for first-order logic. We do not allow proof calculi, which, for example, admit all first-order instances of

*,†This research has been partially supported by the National Nature Science Foundation of China (60970011), the Sino-German Center for Research Promotion (GZ400), and the John Templeton Foundation through Project # 13152, the Infinity Project at the Centre de Recerca Matemàtica.

[1] All notions will be defined in a precise manner in later sections.

propositional tautologies as axioms (as then it would be difficult to recognize correct proofs if $\mathrm{P} \neq \mathrm{NP}$).

In a different context, namely when trying to show that a certain logic $L_{\leq}$ for PTIME (introduced in [**7**]) does not satisfy some effectivity condition, Nash et al. introduced implicitly [**12**] (and this was done explicitly in [**1**]) the *parameterized acceptance problem* p-Acc$_{\leq}$ *for nondeterministic Turing machines*:

p-Acc$_{\leq}$	
Instance:	A nondeterministic Turing machine $\mathbb{M}$ and $n \in \mathbb{N}$ in unary.
Parameter:	$\|\|\mathbb{M}\|\|$, the size of $\mathbb{M}$.
Problem:	Does $\mathbb{M}$ accept the empty input tape in $\leq n$ steps?

Both problems, p-Gödel and p-Acc$_{\leq}$, are *slicewise monotone*, that is, their instances have the form (x, n), where $x \in \{0,1\}^*$ and $n \in \mathbb{N}$ is given in unary,[2] the parameter is $|x|$, and finally for all $x \in \{0,1\}^*$ and $n, n' \in \mathbb{N}$ we have

if (x, n) is a positive instance and $n < n'$, then (x, n') is a positive instance.

A slicewise monotone problem is in the complexity class $\mathrm{XNP}_{\mathrm{uni}}$ if there is a nondeterministic algorithm that accepts it in time $n^{f(|x|)}$ for some function $f \colon \mathbb{N} \to \mathbb{N}$. And co-$\mathrm{XNP}_{\mathrm{uni}}$ contains the complements of problems in $\mathrm{XNP}_{\mathrm{uni}}$. We show:

Theorem 0.1 Taut *has an optimal proof system if and only if every slicewise monotone problem in* NP *is in* co-$\mathrm{XNP}_{\mathrm{uni}}$.

There are trivial slicewise monotone problems which are fixed-parameter tractable. However, for the slicewise monotone problems mentioned above we can show:

Theorem 0.2 Taut *has an optimal proof system* $\iff$ p-Acc$_{\leq} \in$ co-$\mathrm{XNP}_{\mathrm{uni}}$
$\iff$ p-Gödel $\in$ co-$\mathrm{XNP}_{\mathrm{uni}}$.

In [**3**] we showed that Taut has a *p-optimal* proof system if and only if a certain logic $L_{\leq}$ is a P-bounded logic for P (=PTIME). The equivalence in the first line of Theorem 0.2 is the nondeterministic version of this result; in fact, an immediate consequence of it states that Taut has an optimal proof system if and only if $L_{\leq}$ is an NP-bounded logic for P (a concept that we will introduce in Section 5). It turns out that a slight variant of $L_{\leq}$ is an NP-bounded logic for P (without any assumption).

The content of the different sections is the following. In Section 1 and Section 2 we recall the concepts and results of parameterized complexity and on optimal proof systems, respectively, we need in Section 3 to derive the equivalence in the first line of Theorem 0.2. Furthermore, in Section 2 we claim that every problem hard for 2EXP under polynomial time reductions has no optimal proof system. In Section 4 we derive some basic properties of slicewise monotone problems, show that p-Acc$_{\leq}$ is of highest parameterized complexity among the slicewise monotone problems with classical complexity in NP, and finally show that all the slicewise monotone problems we consider in a certain sense have the same complexity (see Proposition 4.6 for the precise statement). This yields Theorem 0.1 and the remaining equivalence of Theorem 0.2. As already mentioned, in Section 5 we analyze the relationship of the existence of an optimal proof system for Taut and the properties of the logic $L_{\leq}$.

[2] The requirement that n is given in unary notation ensures that the classical complexity of most slicewise monotone problems we consider is in NP.

1 Some preliminaries

In this section we recall some basic definitions and concepts from parameterized complexity and introduce the concept of slicewise monotone parameterized problem.

We denote the alphabet $\{0,1\}$ by Σ. The length of a string $x \in \Sigma^*$ is denoted by $|x|$. We identify problems with subsets Q of Σ^*. Clearly, as done mostly, we present concrete problems in a verbal, hence uncodified form or by using other alphabets. We denote by P the class of problems Q such that $x \in Q$ is solvable in time polynomial in $|x|$.

All deterministic and nondeterministic Turing machines have Σ as their alphabet. If necessary we will not distinguish between a Turing machine and its code, a string in Σ^*. If $\mathbb{M}$ is a Turing machine we denote by $\|\mathbb{M}\|$ the length of its code.

Sometimes statements containing a formulation like "there is a $d \in \mathbb{N}$ such that for all $x \in \Sigma^*$: $\ldots \leq |x|^d$" can be wrong for $x \in \Sigma^*$ with $|x| \leq 1$. We trust the reader's common sense to interpret such statements reasonably.

If $\mathbb{A}$ is any (deterministic or nondeterministic) algorithm and $\mathbb{A}$ accepts x, then we denote by $t_{\mathbb{A}}(x)$ the number of steps of a shortest accepting run of $\mathbb{A}$ on x; if $\mathbb{A}$ does not accept x, then $t_{\mathbb{A}}(x)$ is not defined.

1.1 Parameterized complexity

We view *parameterized problems* as pairs (Q, κ) consisting of a classical problem $Q \subseteq \Sigma^*$ and a *parameterization* $\kappa\colon \Sigma^* \to \mathbb{N}$, which is required to be polynomial time computable. We will present parameterized problems in the form we did it for p-GÖDEL and p-ACC$_{\leq}$ in the Introduction.

A parameterized problem (Q, κ) is *fixed-parameter tractable* (or, in FPT) if $x \in Q$ is solvable by an *fpt-algorithm*, that is, by a deterministic algorithm running in time $f(\kappa(x)) \cdot |x|^{O(1)}$ for some computable $f\colon \mathbb{N} \to \mathbb{N}$.

Let C be a complexity class of classical complexity theory defined in terms of deterministic (nondeterministic) algorithms. A parameterized problem (Q, κ) is in the class XC$_{\text{uni}}$ if there is a deterministic (nondeterministic) algorithm deciding (accepting) Q and witnessing for every $k \in \mathbb{N}$ that the classical problem

$$(Q, \kappa)_k := \big\{x \in Q \mid \kappa(x) = k\big\},$$

the kth *slice* of (Q, κ), is in C. For example, (Q, κ) is in the class XP$_{\text{uni}}$ if there is a deterministic algorithm $\mathbb{A}$ deciding $x \in Q$ in time $|x|^{f(\kappa(x))}$ for some function $f\colon \mathbb{N} \to \mathbb{N}$. And (Q, κ) is in the class XNP$_{\text{uni}}$ if there is a nondeterministic algorithm $\mathbb{A}$ accepting Q such that for some function $f\colon \mathbb{N} \to \mathbb{N}$ we have $t_{\mathbb{A}}(x) \leq |x|^{f(\kappa(x))}$ for all $x \in Q$. Finally, a parameterized problem (Q, κ) is in the class co-XC$_{\text{uni}}$ if its complement $(\Sigma^* \setminus Q, \kappa)$ is in XC$_{\text{uni}}$.

We have added the subscript "uni" to the names of these classes to emphasize that they are classes of the so-called uniform parameterized complexity theory. If in the definition of XP$_{\text{uni}}$ and XNP$_{\text{uni}}$ we require the function f to be computable, then we get the corresponding classes of the strongly uniform theory. For example, FPT is a class of this theory.

A parameterized problem (Q, κ) is *slicewise monotone* if its instances have the form (x, n), where $x \in \Sigma^*$ and $n \in \mathbb{N}$ is given in unary, if $\kappa((x, n)) = |x|$, and finally if the slices are monotone, that is, for all $x \in \Sigma^*$ and $n, n' \in \mathbb{N}$,

$$(x, n) \in Q \text{ and } n < n' \text{ imply } (x, n') \in Q.$$

We already remarked that the problems p-GÖDEL and p-ACC$_{\leq}$ are slicewise monotone.

Clearly, every parameterized problem (Q, κ) with $Q \in \mathrm{NP}$ is in $\mathrm{XNP}_{\mathrm{uni}}$; thus we can replace co-$\mathrm{XNP}_{\mathrm{uni}}$ by $\mathrm{XNP}_{\mathrm{uni}} \cap$ co-$\mathrm{XNP}_{\mathrm{uni}}$ everywhere in Theorem 0.1 and Theorem 0.2.

2 Optimal proof systems

Let $Q \subseteq \Sigma^*$ be a problem. A *proof system for* Q is a surjective function $P \colon \Sigma^* \to Q$ computable in polynomial time. Then, if $P(w) = x$, we say that w is a *P-proof* of x. A proof system P for Q is *optimal* if for any other proof system P' for Q there is a polynomial $p \in \mathbb{N}[X]$ such that for every $x \in Q$, if x has a P'-proof of length n, then x has a P-proof of length $\leq p(n)$. Hence, any P'-proof can be translated into a P-proof by a nondeterministic polynomial time algorithm.

The corresponding deterministic concept is the notion of p-optimality. The proof system P for Q is *polynomially optimal* or *p-optimal* if for every proof system P' for Q there is a polynomial time computable $T \colon \Sigma^* \to \Sigma^*$ such that for all $w' \in \Sigma^*$

$$P(T(w')) = P'(w').$$

We list some known results. Part (1) and (2) are immediate from the definitions.

(1) Every p-optimal proof system is optimal.

(2) Every nonempty $Q \in \mathrm{P}$ has a p-optimal proof system, every nonempty $Q \in \mathrm{NP}$ has an optimal proof system.

(3) ([**8**]) If Q is nonempty and $Q \leq^p Q'$ (that is, if Q is polynomial time reducible to Q') and Q' has a (p-)optimal proof system, then Q has a (p-)optimal proof system too.

(4) ([**10**]) Every Q hard for $\mathrm{EXP} = \mathrm{DTIME}\left(2^{n^{O(1)}}\right)$ under polynomial time reductions has no p-optimal proof system.

It is not known whether there is a problem $Q \notin \mathrm{P}$ ($Q \notin \mathrm{NP}$) with a p-optimal (an optimal) proof system. As mentioned in the Introduction, Krajíček and Pudlák [**9**] conjectured that there is no optimal proof system for the set TAUT of tautologies.

Concerning (4) we did not find a corresponding result for optimal proof systems in the literature. We can show:

Proposition 2.1 *Every Q hard for* $2\mathrm{EXP} = \mathrm{DTIME}\left(2^{2^{n^{O(1)}}}\right)$ *under polynomial time reductions has no optimal proof system.*

We do not need this result (and will prove it in the full version of the paper). However we state a consequence:

Corollary 2.2 *There is no optimal proof system for the set of valid sentences of first-order logic.*

2.1 Almost optimal algorithms and enumerations of P-easy subsets

Let $Q \subseteq \Sigma^*$ be a problem. A deterministic (nondeterministic) algorithm $\mathbb{A}$ accepting Q is *almost optimal* or *optimal on positive instances of* Q if for every deterministic (nondeterministic) algorithm $\mathbb{B}$ accepting Q there is a polynomial $p \in \mathbb{N}[X]$ such that, for all $x \in Q$,

$$t_{\mathbb{A}}(x) \leq p(t_{\mathbb{B}}(x) + |x|).$$

By definition a subset Q' of Q is P-*easy* if $Q' \in \mathrm{P}$. An *enumeration of the* P-*easy subsets of* Q *by* P-*machines (by* NP-*machines)* is a computable function $M\colon \mathbb{N} \to \Sigma^*$ such that

(i) for every $i \in \mathbb{N}$ the string $M(i)$ is a deterministic (nondeterministic) Turing machine deciding (accepting) a P-easy subset of Q in polynomial time;

(ii) for every P-easy subset Q' of Q there is $i \in \mathbb{N}$ such that $M(i)$ decides (accepts) Q'.

If in the nondeterministic case instead of (i) we only require

(i′) for every $i \in \mathbb{N}$ the string $M(i)$ is a nondeterministic Turing machine accepting a subset of Q in polynomial time,

we obtain the notion of a *weak enumeration of* P-*easy subsets of* Q *by* NP-*machines.*

We denote by Taut the class of all tautologies of propositional logic. We need the following theorem:

Theorem 2.3

(1) *The following statements are equivalent:*
   (a) Taut *has a p-optimal proof system.*
   (b) Taut *has an almost optimal deterministic algorithm.*
   (c) Taut *has an enumeration of the* P-*easy subsets by* P-*machines.*

(2) *The following statements are equivalent:*
   (a) Taut *has an optimal proof system.*
   (b) Taut *has an almost optimal nondeterministic algorithm.*
   (c) Taut *has a weak enumeration of the* P-*easy subsets by* NP-*machines.*
   (d) Taut *has an enumeration of the* P-*easy subsets by* NP-*machines.*

The equivalence of (a) and (b) in (1) and (2) is due to [**9**], the equivalence to (c) to [**13**]. The equivalence in (2) to (d) will be a by-product of the proof of Theorem 3.3; the equivalence was already claimed in [**13**] but its author was so kind to point out to us that he did not realize the difference between (c) and (d): some machines $M(i)$ of a weak enumeration might accept subsets of Q which are not P-easy (but only in NP).

3 Linking slicewise monotone problems and optimal proof systems

The following result yields a uniform bound on the complexity of slicewise monotone problems whose complements have optimal proof systems.

Theorem 3.1 *Let* (Q,κ) *be a slicewise monotone parameterized problem with decidable* Q.

(1) *If* $\Sigma^* \setminus Q$ *has a p-optimal proof system, then* $(Q,\kappa) \in \mathrm{XP}_{\mathrm{uni}}$.

(2) *If* $\Sigma^* \setminus Q$ *has an optimal proof system, then* $(Q,\kappa) \in \text{co-XNP}_{\mathrm{uni}}$.

As by (3) on page 194 every nonempty problem in co-NP has a (p-)optimal proof system if Taut has one, we immediately get:

Corollary 3.2 *Let* (Q,κ) *be a slicewise monotone parameterized problem with* Q *in* NP.

(1) *If* Taut *has a p-optimal proof system, then* $(Q,\kappa) \in \mathrm{XP}_{\mathrm{uni}}$.

(2) *If* Taut *has an optimal proof system, then* $(Q,\kappa) \in \text{co-XNP}_{\mathrm{uni}}$.

Concerning Theorem 3.1(1), we should mention that Monroe [**11**] has shown that if the complement of (the classical problem underlying) p-$\textsc{Acc}_{\le}$ has an almost optimal algorithm (which by [**9**] holds if it has a p-optimal proof system), then p-$\textsc{Acc}_{\le} \in \mathrm{XP}_{\mathrm{uni}}$.

Proof of Theorem 3.1: We present the proof for (2); the proof for (1) is obtained by the obvious modifications. Let (Q, κ) be slicewise monotone and let $\mathbb{Q}$ be a deterministic algorithm deciding Q. Assume that $\Sigma^* \setminus Q$ has an optimal proof system. It is well-known [**9**] that then $\Sigma^* \setminus Q$ has an almost optimal nondeterministic algorithm $\mathbb{O}$. We have to show that $(Q, \kappa) \in \text{co-XNP}_{\text{uni}}$.

Let $\mathbb{S}$ be the algorithm that, on $x \in \Sigma^*$, by systematically applying $\mathbb{Q}$ to the inputs $(x, 0), (x, 1), \ldots$ computes

$$n(x) := \text{ the least } n \text{ such that } (x, n) \in Q.$$

If $(x, n) \notin Q$ for all $n \in \mathbb{N}$, then $n(x)$ is not defined and $\mathbb{S}$ does not stop. We show that the following algorithm $\mathbb{A}$ witnesses that $(\Sigma^* \setminus Q, \kappa) \in \text{XNP}_{\text{uni}}$.

```
$\mathbb{A}(x, n)$    // $x \in \Sigma^*$, $n \in \mathbb{N}$ in unary
    1. In parallel simulate $\mathbb{S}$ on input $x$ and $\mathbb{O}$ on input $(x, n)$
    2.     if $\mathbb{O}$ accepts then accept
    3.     if $\mathbb{S}$ stops, then
    4.         if $n < n(x)$ then accept else reject.
```

By our assumptions on $\mathbb{O}$ and $\mathbb{S}$ and the slicewise monotonicity of Q, it should be clear that $\mathbb{A}$ accepts $\Sigma^* \setminus Q$. We have to show that $\mathbb{A}$ does it in the time required by XNP_{uni}. Hence, we have to determine the running time of $\mathbb{A}$ on inputs $(x, n) \notin Q$.

Case "$(x, \ell) \notin Q$ for all $\ell \in \mathbb{N}$": In this case $\mathbb{S}$ on input x does not stop. Hence, the running time of $\mathbb{A}$ on input (x, n) is determined by $\mathbb{O}$. The following algorithm $\mathbb{O}_x$ accepts $\Sigma^* \setminus Q$: on input (y, ℓ) the algorithm $\mathbb{O}_x$ checks whether $y = x$. If so, it accepts and otherwise it runs $\mathbb{O}$ on input (y, ℓ) and answers accordingly. Clearly, for all $\ell \in \mathbb{N}$

$$t_{\mathbb{O}_x}((x, \ell)) \leq O(|x|).$$

As $\mathbb{O}$ is almost optimal, we know that there is a constant $d_x \in \mathbb{N}$ (depending on x) such that for all $(y, \ell) \in \Sigma^* \setminus Q$

$$t_{\mathbb{O}}((y, \ell)) \leq \big(|(y, \ell)| + t_{\mathbb{O}_x}((y, \ell))\big)^{d_x}.$$

In particular, we have

$$t_{\mathbb{A}}((x, n)) = O(t_{\mathbb{O}}((x, n))) \leq O\Big(\big(|(x, n)| + O(|x|)\big)^{d_x}\Big) \leq n^{d'_x}$$

for some constant $d'_x \in \mathbb{N}$ (depending on x).

Case "$(x, \ell) \in Q$ for some $\ell \in \mathbb{N}$": Then $\mathbb{S}$ will stop on input x. Thus, in the worst case, $\mathbb{A}$ on input (x, n) has to wait till the simulation of $\mathbb{S}$ on x stops and then $\mathbb{A}$ must check whether the result $n(x)$ of the computation of $\mathbb{S}$ is bigger than n or not and answer according to Line 4. So in the worst case $\mathbb{A}$ takes time $O(t_{\mathbb{S}}(x) + O(n)) \leq n^{O(t_{\mathbb{S}}(x))}$. □

We show the equivalence in the first line of Theorem 0.2:

Theorem 3.3

(1) TAUT *has a p-optimal proof system iff* p-ACC$_{\leq} \in \text{XP}_{\text{uni}}$.
(2) TAUT *has an optimal proof system iff* p-ACC$_{\leq} \in \text{co-XNP}_{\text{uni}}$.

Proof. Again we only prove (2) and by the previous corollary it suffices to show the corresponding implication from right to left.

So assume that the complement of p-ACC$_{\leq}$ is in XNP$_{\text{uni}}$ and let $\mathbb{A}$ be a nondeterministic algorithm witnessing it; in particular, $t_{\mathbb{A}}\big((\mathbb{M}, n)\big) \leq n^{f(\|\mathbb{M}\|)}$ for some function f and all $(\mathbb{M}, n) \notin p$-ACC$_{\leq}$. We show that TAUT has an enumeration of the P-easy subsets by NP-machines (and this suffices by Theorem 2.3).

We fix a deterministic Turing machine $\mathbb{M}_0$ that given a propositional formula α and an assignment checks if this assignment satisfies α in time $|\alpha|^2$.

For a deterministic Turing machine $\mathbb{M}$ let $\mathbb{M}^*$ be the nondeterministic machine that on empty input tape

- first guesses a propositional formula α;
- then checks (by simulating $\mathbb{M}$) whether $\mathbb{M}$ accepts α and rejects if this is not the case;
- finally guesses an assignment and accepts if this assignment does not satisfy α (this is checked by simulating $\mathbb{M}_0$).

A deterministic Turing machine $\mathbb{M}$ is *clocked* if (the code of) $\mathbb{M}$ contains a natural number time$(\mathbb{M})$ such that $n^{\text{time}(\mathbb{M})}$ is a bound for the running time of $\mathbb{M}$ on inputs of length n (in particular, a clocked machine is a polynomial time one).

Finally, for a clocked Turing machine $\mathbb{M}$ let $\mathbb{M}^+$ be the nondeterministic Turing machine that on input α accepts if and only if (i) and (ii) hold:

(i) $\mathbb{M}$ accepts α;
(ii) $(\mathbb{M}^*, |\alpha|^{\text{time}(\mathbb{M})+4}) \notin p$-ACC$_{\leq}$.

The machine $\mathbb{M}^+$ checks (i) by simulating $\mathbb{M}$ and (ii) by simulating $\mathbb{A}$. Hence, if $\mathbb{M}^+$ accepts α, then

$$t_{\mathbb{M}^+}(\alpha) \leq O\left(|\alpha|^{\text{time}(\mathbb{M})} + t_{\mathbb{A}}\big((\mathbb{M}^*, |\alpha|^{\text{time}(\mathbb{M})+4})\big)\right),$$

and as $t_{\mathbb{A}}\big((\mathbb{M}^*, |\alpha|^{\text{time}(\mathbb{M})+4})\big) \leq |\alpha|^{(\text{time}(\mathbb{M})+4)\cdot f(\|\mathbb{M}^*\|)}$, the Turing machine $\mathbb{M}^+$ accepts in time polynomial in $|\alpha|$.

We show that $\mathbb{M}^+$, where $\mathbb{M}$ ranges over all clocked machines, yields an enumeration of all P-easy subsets of TAUT by NP-machines. First let $\mathbb{M}$ be a clocked machine. We prove that $\mathbb{M}^+$ accepts a P-easy subset of TAUT.

$\mathbb{M}^+$ *accepts a subset of* TAUT: If $\mathbb{M}^+$ accepts α, then, by (i), $\mathbb{M}$ accepts α and by (ii), $(\mathbb{M}^*, |\alpha|^{\text{time}(\mathbb{M})+4}) \notin p$-ACC$_{\leq}$. Therefore, by definition of $\mathbb{M}^*$, every assignment satisfies α and hence $\alpha \in$ TAUT.

$\mathbb{M}^+$ *accepts a P-easy set*: If $(\mathbb{M}^*, m) \in p$-ACC$_{\leq}$ for some m, then, by slicewise monotonicity of p-ACC$_{\leq}$, the machine $\mathbb{M}^+$ accepts a finite set and hence a P-easy set. If $(\mathbb{M}^*, m) \notin p$-ACC$_{\leq}$ for all m, then $\mathbb{M}^+$ accepts exactly those α accepted by $\mathbb{M}$; as $\mathbb{M}$ is clocked, this is a set in P.

Now let $Q \subseteq$ TAUT be a P-easy subset of TAUT and let $\mathbb{M}$ be a clocked machine deciding Q. Then $\mathbb{M}^+$ accepts Q. □

4 Slicewise monotone parameterized problems

In this section we observe that p-ACC$_{\leq}$ is a complete problem in the class of slicewise monotone parameterized problems with underlying classical problem in NP. Furthermore, we shall see that in Theorem 3.3 we can replace the problem p-Acc$_{\leq}$ by other slicewise monotone parameterized problems (among them p-Gödel) by showing for them that they are in the class XP$_{\text{uni}}$ (co-XNP$_{\text{uni}}$) if and only if p-Acc$_{\leq}$ is.

4.1 The complexity of slicewise monotone problems

We start with some remarks on the complexity of slicewise monotone problems. In [**1, 2**] we have shown that p-Acc$_{\leq}$ and p-Gödel are not fixed-parameter tractable if "P $\neq$ NP holds for all time constructible and increasing functions", that is, if DTIME$(h^{O(1)}) \neq$ NTIME$(h^{O(1)})$ for all time constructible and increasing functions $h\colon \mathbb{N} \to \mathbb{N}$. However:

Proposition 4.1

(1) ([**2**]) *Let (Q, κ) be slicewise monotone. Then (Q, κ) is nonuniformly fixed-parameter tractable, that is, there is a $c \in \mathbb{N}$, a function $f\colon \mathbb{N} \to \mathbb{N}$, and for every k an algorithm deciding the slice $(Q, \kappa)_k$ in time $f(k) \cdot n^c$.*

(2) *Let (Q, κ) be slicewise monotone with enumerable Q. Then $(Q, \kappa) \in$ XNP$_{\text{uni}}$.*

Proof. (2) Let $\mathbb{Q}$ be an algorithm enumerating Q. The following algorithm shows that $(Q, \kappa) \in$ XNP$_{\text{uni}}$: On input (x, n) it guesses $m \in \mathbb{N}$ and a string c. If c is the code of an initial segment of the run of $\mathbb{Q}$ enumerating (x, m), then it accepts if $m \leq n$. □

We remark that there are slicewise monotone problems with underlying classical problem of arbitrarily high complexity that are fixed-parameter tractable. In fact, let $Q_0 \subseteq \Sigma^*$ be decidable. Then the slicewise monotone (Q, κ) with

$$Q := \big\{(x, n) \mid x \in Q_0,\ n \in \mathbb{N}, \text{ and } |x| \leq n\big\}$$

(and $\kappa((x, n)) := |x|$) is in FPT.

To compare the complexity of parameterized problems we use the standard notions of reduction that we recall first. Let (Q, κ) and (Q', κ') be parameterized problems. We write $(Q, \kappa) \leq^{\text{fpt}} (Q', \kappa')$ if there is an *fpt-reduction* from (Q, κ) to (Q', κ'), that is, a mapping $R\colon \Sigma^* \to \Sigma^*$ with:

(1) For all $x \in \Sigma^*$ we have $(x \in Q \iff R(x) \in Q')$.

(2) $R(x)$ is computable in time $f(\kappa(x)) \cdot |x|^{O(1)}$ for some computable $f\colon \mathbb{N} \to \mathbb{N}$.

(3) There is a computable function $g\colon \mathbb{N} \to \mathbb{N}$ such that $\kappa'(R(x)) \leq g(\kappa(x))$ for all $x \in \Sigma^*$.

We write $(Q, \kappa) \leq^{\text{xp}} (Q', \kappa')$ if there is an *xp-reduction* from (Q, κ) to (Q', κ'), which is defined as $(Q, \kappa) \leq^{\text{fpt}} (Q', \kappa')$ except that instead of (2) it is only required that $R(x)$ is computable in time $|x|^{f(\kappa(x))}$ for some computable $f\colon \mathbb{N} \to \mathbb{N}$.

These are notions of reductions of the usual (strongly uniform) parameterized complexity theory. We get the corresponding notions $\leq^{\text{fpt}}_{\text{uni}}$ and $\leq^{\text{xp}}_{\text{uni}}$ by allowing the functions f and g to be arbitrary (and not necessarily computable).

We shall use the following simple observation.

Lemma 4.2 *If $(Q, \kappa) \leq^{\text{xp}}_{\text{uni}} (Q', \kappa')$ and $(Q', \kappa') \in$ XP$_{\text{uni}}$, then $(Q, \kappa) \in$ XP$_{\text{uni}}$. The same holds for* XNP$_{\text{uni}}$ *instead of* XP$_{\text{uni}}$.

We turn again to slicewise monotone problems. Among these problems with underlying classical problem in NP the problem p-Acc$_{\leq}$ is of highest complexity.

Proposition 4.3 *Let (Q, κ) be slicewise monotone and $Q \in$ NP. Then*

$$(Q, \kappa) \leq^{\text{fpt}} p\text{-Acc}_{\leq}.$$

Note that this result together with Theorem 3.3(2) yields Theorem 0.1.

Proof of Proposition 4.3: Let $\mathbb{M}$ be a nondeterministic Turing machine accepting Q. We may assume that for some $d \in \mathbb{N}$ the machine $\mathbb{M}$ on input (x, n) performs exactly $|(x, n)|^d$ steps. For $x \in \Sigma^*$ let $\mathbb{M}_x$ be the nondeterministic Turing machine that on empty input tape, first writes x on the tape, then guesses a natural number m, and finally simulates the computation of $\mathbb{M}$ on input (x, m). We can assume that there is a polynomial time computable function h such that $\mathbb{M}_x$ makes exactly $h(x, m) \in O\big(|x| + m + |(x, m)|^d\big)$ steps if it chooses the natural number m. Furthermore we can assume that $h(x, m) < h(x, m')$ for $m < m'$.

Then $(x, n) \mapsto \big(\mathbb{M}_x, h(x, n)\big)$ is an fpt-reduction from (Q, κ) to p-Acc$_{\le}$: Clearly, if $(x, n) \in Q$ then $\big(\mathbb{M}_x, h(x, n)\big) \in p$-Acc$_{\le}$ by construction of $\mathbb{M}_x$. Conversely, if $\big(\mathbb{M}_x, h(x, n)\big) \in p$-Acc$_{\le}$, then by the properties of h we see that $\mathbb{M}$ accepts (x, m) for some $m \le n$. Thus, $(x, m) \in Q$ and therefore $(x, n) \in Q$ by slicewise monotonicity. □

Later on we shall use the following related result.

Proposition 4.4 *Let (Q, κ) be slicewise monotone and assume that there is a nondeterministic algorithm $\mathbb{A}$ accepting Q such that $t_{\mathbb{A}}(x, n) \le n^{f(|x|)}$ for some time constructible f and all $(x, n) \in Q$. Then*

$$(Q, \kappa) \le^{\text{xp}} p\text{-Acc}_{\le}.$$

Proof. Let (Q', κ') be the problem

Instance: $x \in \Sigma^*$ and $m \in \mathbb{N}$ in unary.
Parameter: $|x|$.
Problem: Is there an $n \in \mathbb{N}$ such that $n^{f(|x|)} \le m$ and $(x, n) \in Q$?

By the previous proposition we get our claim once we have shown:

(1) (Q', κ') is slicewise monotone and $Q' \in \text{NP}$.
(2) $(Q, \kappa) \le^{\text{xp}} (Q', \kappa')$.

To see (1), let $\mathbb{A}$ be as stated above and let $\mathbb{T}$ an algorithm witnessing the time constructibility of f; that is, $\mathbb{T}$ on input $k \in \mathbb{N}$ computes $f(k)$ in exactly $f(k)$ steps. An algorithm $\mathbb{B}$ witnessing that $Q' \in \text{NP}$ runs as follows on input (x, m):

- $\mathbb{B}$ guesses $n \in \mathbb{N}$;
- if $n = 1$, the algorithm $\mathbb{B}$ rejects in case $m = 0$;
- if $n \ge 2$, the algorithm $\mathbb{B}$ simulates m steps of the computation of $\mathbb{T}$ on input $|x|$; if thereby $\mathbb{T}$ does not stop, $\mathbb{B}$ rejects; otherwise, the simulation yields $f(|x|)$ and $\mathbb{B}$ checks whether $n^{f(|x|)} > m$ (this can be detected in time $O(m)$); in the positive case $\mathbb{B}$ rejects;
- finally $\mathbb{B}$ simulates the computation of $\mathbb{A}$ on (x, n) and answers accordingly.

As for (2), note that the mapping $(x, n) \mapsto \big(x, n^{f(|x|)}\big)$ is an xp-reduction. □

4.2 Slicewise monotone problems related to logic

In the next section we will use some further slicewise monotone problems related to first-order logic and least fixed-point logic that we introduce now.

We assume familiarity with *first-order logic* FO and its extension *least fixed-point logic* LFP (e.g., see [**5**]). We denote by FO[τ] and LFP[τ] the set of sentences of vocabulary τ

of FO and of LFP, respectively. In this paper all vocabularies are finite sets of relational symbols.

If the structure $\mathcal{A}$ is a model of the LFP-sentence φ we write $\mathcal{A} \models \varphi$. We only consider structures $\mathcal{A}$ with finite universe A. The size $\|\mathcal{A}\|$ of the structure $\mathcal{A}$ is the length of a reasonable encoding of $\mathcal{A}$ as string in Σ^*. An algorithm based on the inductive definition of the satisfaction relation for LFP shows (see [**14**]):

Proposition 4.5 *The model-checking problem $\mathcal{A} \models \varphi$ for structures $\mathcal{A}$ and* LFP*-sentences φ can be solved in time $\|\mathcal{A}\|^{O(|\varphi|)}$.*

Let $L = \text{FO}$ or $L = \text{LFP}$. First we introduce the parameterized problem

p-L-Model
Instance: An L-sentence φ and $n \in \mathbb{N}$ in unary.
Parameter: $|\varphi|$.
Problem: Is there a structure $\mathcal{A}$ with $\mathcal{A} \models \varphi$ and $|A| \leq n$?

Here, $|A|$ denotes the size of the universe A of $\mathcal{A}$. For every vocabulary τ we let $\tau_< := \tau \cup \{<\}$, where $<$ is a binary relation symbol not in τ. For $m \geq 1$ we say that an $L[\tau_<]$-sentence φ is $\leq m$*-invariant* if for all τ-structures $\mathcal{A}$ with $|A| \leq m$ we have

$$(\mathcal{A}, <_1) \models \varphi \iff (\mathcal{A}, <_2) \models \varphi$$

for all orderings $<_1$ and $<_2$ on A.

Finally we introduce the slicewise monotone parameterized problem

p-L-Not-Inv
Instance: A vocabulary τ, an $L[\tau_<]$-sentence φ and $m \geq 1$ in unary.
Parameter: $|\varphi|$.
Problem: Is φ not $\leq m$-invariant?

4.3 Membership in XP_{uni} and co-XNP_{uni}

Concerning membership in the classes XP_{uni} and co-XNP_{uni} all the slicewise monotone problems we have introduced behave in the same way:

Proposition 4.6 *Consider the parameterized problems*

p-Gödel*,* p-FO-Model*,* p-LFP-Model*,* p-FO-Not-Inv*,*
p-LFP-Not-Inv*, and* p-$\text{Acc}_{\leq}$*.*

If one of the problems is in XP_{uni}*, then all are; if one of the problems is in* co-XNP_{uni}*, then all are.*

By Theorem 3.3 this result yields Theorem 0.2. We prove it with Lemmas 4.7–4.10.

Lemma 4.7 ([**2**]) p-Gödel $\leq^{\text{fpt}}$ p-FO-Model.

Lemma 4.8 *Let $L = \text{FO}$ or $L = \text{LFP}$. Then p-L-*Model $\leq^{\text{fpt}}$ p-L-Not-Inv.

Proof. Let φ be a sentence of vocabulary τ. We set $\tau' := \tau \cup \{P\}$ with a new unary relation symbol P and consider the sentence of vocabulary $\tau'_<$

$$\psi(\varphi) := \varphi \wedge \text{“}P \text{ holds for the first element of } <\text{”}.$$

Clearly, for every $n \geq 2$

$$(\varphi, n) \in p\text{-FO-}\textsc{Model} \iff \big(\psi(\varphi), n\big) \in p\text{-FO-}\textsc{Not-Inv}$$

and the same equivalence holds for p-LFP-Model and p-LFP-Not-Inv. Thus $(\varphi, n) \mapsto \big(\psi(\varphi), n\big)$ is the desired reduction in both cases. □

Lemma 4.9 p-LFP-Not-Inv $\leq^{\mathrm{xp}}$ p-Acc$_{\leq}$.

Proof. Consider the algorithm $\mathbb{A}$ that on input (φ, m), where φ is an LFP-sentence and $m \geq 1$, guesses a structure $\mathcal{A}$ and two orderings $<_1$ and $<_2$ and accepts if $|A| \leq m$, $(\mathcal{A}, <_1) \models \varphi$, and $(\mathcal{A}, <_2) \models \neg\varphi$. Then, by Proposition 4.5, the algorithm $\mathbb{A}$ witnesses that p-LFP-Not-Inv satisfies the assumptions on (Q, κ) in Proposition 4.4. This yields the claim. □

Lemma 4.10

(1) *If* p-Gödel $\in$ XP$_{\mathrm{uni}}$, *then* p-Acc$_{\leq}$ $\in$ XP$_{\mathrm{uni}}$.
(2) *If* p-Gödel $\in$ co-XNP$_{\mathrm{uni}}$, *then* p-Acc$_{\leq}$ $\in$ co-XNP$_{\mathrm{uni}}$.

Proof. We give the proof of (2). By standard means we showed in [**2**, Lemma 7] that there exists a $d \in \mathbb{N}$ and a polynomial time algorithm that assigns to every nondeterministic Turing machine $\mathbb{M}$ a first-order sentence $\varphi_{\mathbb{M}}$ such that for $n \in \mathbb{N}$

$$(\mathbb{M}, n) \in p\text{-}\textsc{Acc}_{\leq} \implies (\varphi_{\mathbb{M}}, n^d) \in p\text{-}\textsc{Gödel}. \tag{4.1}$$

Moreover,

$$\varphi_{\mathbb{M}} \text{ has a proof} \implies \mathbb{M} \text{ accepts the empty input tape.} \tag{4.2}$$

Now assume that $\mathbb{A}$ is an algorithm that witnesses that the complement of p-Gödel is in XNP$_{\mathrm{uni}}$. We may assume that every run of $\mathbb{A}$ either accepts its input or is infinitely long. Let $d \in \mathbb{N}$ be as above. We present an algorithm $\mathbb{B}$ showing that the complement of p-Acc$_{\leq}$ is in XNP$_{\mathrm{uni}}$. On input $(\mathbb{M}, n)$ the algorithm $\mathbb{B}$ first computes $\varphi_{\mathbb{M}}$ and then runs two algorithms in parallel:

- a brute force algorithm that on input $\mathbb{M}$ searches for the least $n_{\mathbb{M}}$ such that $\mathbb{M}$ on empty input tape has an accepting run of length $n_{\mathbb{M}}$;
- the algorithm $\mathbb{A}$ on input $(\varphi_{\mathbb{M}}, n^d)$.

If the brute force algorithm halts first and outputs $n_{\mathbb{M}}$, then $\mathbb{B}$ checks whether $n_{\mathbb{M}} \leq n$ and answers accordingly.

Assume now that $\mathbb{A}$ halts first. Then $\mathbb{A}$ accepts $(\varphi_{\mathbb{M}}, n^d)$ and $\big((\varphi_{\mathbb{M}}, n^d) \notin p$-Gödel and hence $(\mathbb{M}, n) \notin p$-Acc$_{\leq}$ by (4.1) and therefore$\big)$ $\mathbb{B}$ accepts.

The algorithm $\mathbb{B}$ accepts the complement of p-Acc$_{\leq}$; note that if no run of $\mathbb{A}$ accepts $(\varphi_{\mathbb{M}}, n^d)$, then $(\varphi_{\mathbb{M}}, n^d) \in p$-Gödel and therefore $\mathbb{M}$ accepts the empty input tape by (4.2), so that in this case the computation of the brute force algorithm eventually will stop.

It remains to see that $\mathbb{B}$ accepts the complement of p-Acc$_{\leq}$ in the time required by XNP$_{\mathrm{uni}}$. We consider two cases.

$\mathbb{M}$ *halts on empty input tape*: Then an upper bound for the running time is given by the time that the brute force algorithm needs to compute $n_{\mathbb{M}}$ (and the time for the check whether $n_{\mathbb{M}} \leq n$); hence we have an upper bound of the form $n^{c_{\mathbb{M}}}$.

$\mathbb{M}$ *does not halt on empty input tape*: Then, by (4.2), we have $(\varphi_{\mathbb{M}}, n^d) \notin p$-GÖDEL; hence an upper bound is given by the running time of $\mathbb{A}$ on input $(\varphi_{\mathbb{M}}, n^d)$. □

It should be clear that Lemmas 4.7–4.10 together with Lemma 4.2 yield a proof of Proposition 4.6.

5 Optimal algorithms and the logic $L_{\leq}$

In this section we interpret Theorem 0.2 in terms of the expressive power of a certain logic.

For our purposes a *logic* L consists

- for every vocabulary τ of a set $L[\tau]$ of strings, the set of *L-sentences of vocabulary* τ and of an algorithm that for every vocabulary τ and every string ξ decides whether $\xi \in L[\tau]$ (in particular, $L[\tau]$ is decidable for every τ);
- of a *satisfaction relation* $\models_L$; if $(\mathcal{A}, \varphi) \in \models_L$, written $\mathcal{A} \models_L \varphi$, then $\mathcal{A}$ is a τ-structure and $\varphi \in L[\tau]$ for some vocabulary τ; furthermore for each $\varphi \in L[\tau]$ the class $\mathrm{Mod}_L(\varphi) := \big\{\mathcal{A} \mid \mathcal{A} \models_L \varphi\big\}$ of *models of* φ is closed under isomorphisms.

Definition 5.1 Let L be a logic.

(a) *L is a logic for* P if for all vocabularies τ and all classes C (of encodings) of τ-structures closed under isomorphisms we have

$$C \in \mathrm{P} \iff C = \mathrm{Mod}_L(\varphi) \text{ for some } \varphi \in L[\tau].$$

(b) *L is a* P*-bounded logic for* P if (a) holds and if there is an algorithm $\mathbb{A}$ deciding $\models_L$ $\big($that is, for every structure $\mathcal{A}$ and L-sentence φ the algorithm $\mathbb{A}$ decides whether $\mathcal{A} \models_L \varphi\big)$ and if moreover, for fixed φ the algorithm $\mathbb{A}$ runs in time polynomial in $\|\mathcal{A}\|$.

The relationship of these concepts with topics of this paper is already exemplified by the following simple observation.

Proposition 5.2 *Let L be a logic for* P *and define p-$\models_L$ by*

p-$\models_L$
Instance: A structure $\mathcal{A}$ and an L-sentence φ.
Parameter: $|\varphi|$.
Problem: Is $\mathcal{A} \models_L \varphi$

Then L is a P*-bounded logic for* P *if and only if* p-$\models_L \in \mathrm{XP}_{\mathrm{uni}}$.

This relationship suggests the following definition.

Definition 5.3 *L is an* NP*-bounded logic for* P if it is a logic for P and p-$\models_L \in \mathrm{XNP}_{\mathrm{uni}}$.

We introduce the logic $L_{\leq}$, a variant of LFP.[3] For every vocabulary τ we set

$$L_{\leq}[\tau] = \mathrm{LFP}[\tau_<]$$

[3] In this section, if the structure $\mathcal{B}$ is a model of an LFP-sentence φ we write $\mathcal{A} \models_{\mathrm{LFP}} \varphi$ instead of $\mathcal{A} \models \varphi$.

(recall that $\tau_< := \tau \cup \{<\}$, with a new binary $<$) and define the semantics by

$$\mathcal{A} \models_{L_\leq} \varphi \iff \Big(\varphi \text{ is } \leq |A|\text{-invariant and}$$
$$(\mathcal{A}, <) \models_{\mathrm{LFP}} \varphi \text{ for some ordering } < \text{ on } A\Big).$$

Hence, by the previous proposition and the definition of $\models_{L_\leq}$, we get:

Proposition 5.4

(1) *The following statements are equivalent:*
   (a) $L_\leq$ *is a* P*-bounded logic for* P.
   (b) p-$\models_{L_\leq} \in \mathrm{XP}_{\mathrm{uni}}$.
   (c) p-LFP-NOT-INV $\in \mathrm{XP}_{\mathrm{uni}}$.

(2) *The following statements are equivalent:*
   (a) $L_\leq$ *is an* NP*-bounded logic for* P.
   (b) p-$\models_{L_\leq} \in \mathrm{XNP}_{\mathrm{uni}}$.
   (c) p-LFP-NOT-INV $\in$ co-$\mathrm{XNP}_{\mathrm{uni}}$.

By Theorem 0.2 and Proposition 4.6 we get:

Theorem 5.5 TAUT *has an optimal proof system if and only if* $L_\leq$ *is an* NP*-bounded logic for* P.

Hence, if TAUT has an optimal proof system, then there is an NP-enumeration of P-easy classes of graphs closed under isomorphisms. We do not define the concept of NP-enumeration explicitly, however the enumeration obtained by applying the algorithm in $\mathrm{XNP}_{\mathrm{uni}}$ for p- $\models_{L_\leq}$ to the classes $\mathrm{Mod}_{L_\leq}(\varphi(\text{GRAPH}) \wedge \psi)$, where $\varphi(\text{GRAPH})$ axiomatizes the class of graphs and ψ ranges over all sentences of $L_\leq$ in the language of graphs, is such an NP-enumeration. Note that even without the assumption that TAUT has an optimal proof system we know that there is such an NP-enumeration of P-easy classes of graphs closed under isomorphisms, as the following variant $L_\leq$(not) of $L_\leq$ is an NP-bounded logic for P. The logic $L_\leq$(not) has the same syntax as $L_\leq$ and the semantics is given by the following clause:

$$\mathcal{A} \models_{L_\leq(\mathrm{not})} \varphi \quad \iff \quad \text{not } \mathcal{A} \models_{L_\leq} \varphi.$$

As the class P is closed under complements, $L_\leq$(not) is a logic for P. And $L_\leq$(not) is an NP-bounded logic for P, as p-LFP-NOT-INV $\in \mathrm{XNP}_{\mathrm{uni}}$.

References

[1] Y. Chen and J. Flum. A logic for PTIME and a parameterized halting problem. In *Proceedings of the 24th IEEE Symposium on Logic in Computer Science (LICS'09)*, pages 397–406, 2009.

[2] Y. Chen and J. Flum. On the complexity of Gödel's proof predicate. *The Journal of Symbolic Logic*, 75(1): 239–254, 2010.

[3] Y. Chen and J. Flum. On p-optimal proof systems and logics for PTIME. In *Proceedings of the 37th International Colloquium on Automata, Languages and Programming (ICALP'10, Track B)*, Lecture Notes in Computer Science 6199, pages 321–332, 2010.

[4] S. Cook and R. Reckhow. The relative efficiency of propositional proof systems. *The Journal of Symbolic Logic*, 44:36–50, 1979.

[5] H.-D. Ebbinghaus and J. Flum. *Finite Model Theory*, 2nd edition, Springer, 1999.

[6] K. Gödel. *Collected Works*, vol. VI, 372–376, Clarendon Press, 2003.

[7] Y. Gurevich. Logic and the challenge of computer science. In *Current Trends in Theoretical Computer Science*, Computer Science Press, 1–57, 1988.

[8] J. Köbler and J. Messner. Complete problems for promise classes by optimal proof systems for test sets. In *Proceedings of the 13th IEEE Conference on Computational Complexity (CCC' 98)*, 132–140, 1998.

[9] J. Krajíček and P. Pudlák. Propositional proof systems, the consistency of first order theories and the complexity of computations. *The Journal of Symbolic Logic*, 54:1063–1088, 1989.

[10] J. Messner. On optimal algorithms and optimal proof systems. In *Proceedings of the 16th Symposium on Theoretical Aspects of Computer Science (STACS'99)*, Lecture Notes in Computer Science 1563, 361–372, 1999.

[11] H. Monroe. Speedup for natural problems. *Electronic Colloquium on Computational Complexity*, Report TR09-056, 2009.

[12] A. Nash, J. Remmel, and V. Vianu. PTIME queries revisited. In *Proceedings of the 10th International Conference on Database Theory (ICDT'05)*, T. Eiter and L. Libkin (eds.), Lecture Notes in Computer Science 3363, 274–288, 2005.

[13] Z. Sadowski. On an optimal propositional proof system and the structure of easy subsets. *Theoretical Computer Science*, 288(1):181–193, 2002.

[14] M. Y. Vardi. On the complexity of bounded-variable queries. In *Proceedings of the 14th ACM Symposium on Principles of Database Systems (PODS'95)*, pages 266–276, 1995.

Hard instances of algorithms and proof systems

Yijia Chen[†], **Jörg Flum**[‡], **Moritz Müller**[§]

[†] Department of Computer Science, Shanghai Jiao Tong University, China
yijia.chen@cs.sjtu.edu.cn

[‡] Mathematisches Institut, Albert-Ludwigs-Universität Freiburg, Germany
joerg.flum@math.uni-freiburg.de

[§] Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
moritz.mueller@univie.ac.at

Abstract. Assuming that the class TAUT of tautologies of propositional logic has no almost optimal algorithm, we show that every algorithm $\mathbb{A}$ deciding TAUT has a polynomial time computable sequence witnessing that $\mathbb{A}$ is not almost optimal. The result extends to every Π_t^p-complete problem with $t \geq 1$; however, we show that assuming the Measure Hypothesis there is a problem which has no almost optimal algorithm but has an algorithm without hard sequences.

Introduction

Let $\mathbb{A}$ be an algorithm deciding a problem Q. A sequence $(x_s)_{s\in\mathbb{N}}$ of strings in Q is *hard for* $\mathbb{A}$ if it is computable in polynomial time and the sequence $(t_{\mathbb{A}}(x_s)_{s\in\mathbb{N}})$ is not polynomially bounded in s.[1] Here, $t_{\mathbb{A}}(x)$ denotes the number of steps the algorithm $\mathbb{A}$ takes on input x. Clearly, if $\mathbb{A}$ is polynomial time, then $\mathbb{A}$ has no hard sequences. Furthermore, an almost optimal algorithm for Q has no hard sequences either. Recall that an algorithm $\mathbb{A}$ is *almost optimal for* Q if for every input $x \in Q$ the running time $t_{\mathbb{A}}(x)$ is polynomially bounded in $t_{\mathbb{B}}(x)$ for any other algorithm $\mathbb{B}$ deciding Q. In fact, if $(x_s)_{s\in\mathbb{N}}$ is a hard sequence for an algorithm, then one can polynomially speed up it on $\{x_s \mid s \in \mathbb{N}\}$, so it cannot be almost optimal.

Central to this paper is the question: To what extent can we show that algorithms which are not almost optimal have hard sequences? Our main result states:

(a) *If a* co-NP-*complete problem Q has no almost optimal algorithm, then every algorithm deciding Q has hard sequences.*

Perhaps one would expect that one can strengthen (a) and show that even if a co-NP-complete problem Q has an almost optimal algorithm, then every algorithm, which is not almost optimal and decides Q, has a hard sequence. However, we show:

> *If the Measure Hypothesis holds, then every* co-NP-*complete problem with padding and with an almost optimal algorithm has an algorithm which is not almost optimal but has no hard sequences.*

Even though we can extend the result (a) to Π_t^p-complete problems (with $t \geq 1$), apparently there are some limitations as we derive the following result:

> *If the Measure Hypothesis holds, then there is a problem Q which has no almost optimal algorithm but has an algorithm without hard sequences.*

This article appeared in *Electronic Colloquium on Computational Complexity*, TR11-085, 2011.

[1] All notions will be defined in a precise manner later.

In particular, there are algorithms deciding such a Q and polynomially speeding up a given algorithm. That is, this notion of speeding up (e.g., considered in [**11, 13**]) differs from our notion of the existence of a hard sequence.

Assume that a co-NP-complete problem Q has no almost optimal algorithm. Can we even effectively assign to every algorithm deciding Q a hard sequence? We believe that under reasonable complexity-theoretic assumptions one should be able to show that such an effective procedure or at least a polynomial time procedure does not exist, but we were not able to show it. However, recall that by a result due to McCreight and Meyer [**11**] and rediscovered by Messner [**10**] we know:

> *For every* EXP*-hard problem Q there is a polynomial time effective procedure assigning to every algorithm solving Q a sequence hard for it.*

Hence, if EXP = NP, then for every NP-hard (and hence for every co-NP-hard) problem Q there is a polynomial time effective procedure assigning a hard sequence to every algorithm deciding Q.

Our proof of (a) generalizes to nondeterministic algorithms. This "nondeterministic statement" yields a version of a result due to Krajíček which he derived for non-optimal propositional proof systems: If TAUT, the set of tautologies of propositional logic, has no optimal proof system, then for every propositional proof system $\mathbb{P}$ there is a polynomial time computable sequence $(\alpha_s)_{s\in\mathbb{N}}$ of propositional tautologies α_s with $s \leq |\alpha_s|$ which only have superpolynomial $\mathbb{P}$-proofs. While it is well-known that nondeterministic algorithms for TAUT and propositional proof systems are more or less the same (so that the nondeterministic version of (a) essentially is Krajíček's result), the relationship between deterministic algorithms deciding TAUT and propositional proof systems is more subtle. Nevertheless, we are able to use (a) to derive a statement on hard sequences for propositional proof systems in case that TAUT has no *polynomially* optimal proof system.

As a byproduct, we obtain results in "classical terms" for which we do not know proofs avoiding the machinery we develop here; for example, we get:

> *Let Q be* co-NP*-complete. Then, Q has an almost optimal algorithm if and only if Q has a polynomially optimal proof system.*
>
> *If* TAUT *has no almost optimal algorithm, then every* co-NP*-hard problem has no almost optimal algorithm.*

It is still open whether there exist problems outside of NP with optimal proof systems. We show their existence (in NE) assuming the Measure Hypothesis. Krajíček and Pudlák [**7**] proved that E = NE implies that TAUT has an optimal proof system.

If for an algorithm $\mathbb{A}$ deciding a problem Q we have a hard sequence $(x_s)_{s\in\mathbb{N}}$ satisfying $s \leq |x_s|$, then $\{x_s \mid s \in \mathbb{N}\}$ is a *hard set for* $\mathbb{A}$, that is, a polynomial time decidable subset of Q on which $\mathbb{A}$ is not polynomial time. Messner [**10**] has shown for any Q with padding that all algorithms deciding Q have hard sets if and only if Q has no polynomially optimal proof system. We show for arbitrary Q that the existence of hard sets for all algorithms is equivalent to the existence of an effective enumeration of all polynomial time decidable subsets of Q, a property which has turned out to be useful in various contexts (cf. [**2, 3, 12**]). We analyze what Messner's result means for proof systems.

The content of the sections is the following. In Section 1 we recall some concepts. We deal with hard sequences for algorithms in Section 2 and for proof systems in Section 3. Section 4 is devoted to hard sets and Section 5 contains the results and the examples of problems with special properties obtained assuming that the Measure Hypothesis holds.

Finally Section 6 gives an effective procedure yielding hard sequences for nondeterministic algorithms for coNEXP-hard problems.

1 Preliminaries

We denote by Σ the alphabet $\{0,1\}$ and by $|x|$ the length of a string $x \in \Sigma^*$. We identify problems with subsets of Σ^*. *In this paper we always assume that Q denotes a decidable and nonempty problem.*

We denote by P (NP) the class of problems Q such that $x \in Q$ is solvable by a deterministic (nondeterministic) Turing machine in $|x|^{O(1)}$ steps (formally, $n^{O(1)}$ denotes the class of polynomially bounded functions on the natural numbers). A problem $Q \subseteq \Sigma^*$ *has padding* if there is a function *pad*: $\Sigma^* \times \Sigma^* \to \Sigma^*$ computable in logarithmic space having the following properties:

- For any $x, y \in \Sigma^*$, $|pad(x,y)| > |x| + |y|$ and $\big(pad(x,y) \in Q \Leftrightarrow x \in Q\big)$.
- There is a logspace algorithm which, given $pad(x,y)$, recovers y.

By $\langle \ldots, \ldots \rangle$ we denote some standard logspace computable tupling function with logspace computable inverses.

If $\mathbb{A}$ is a deterministic or nondeterministic algorithm and $\mathbb{A}$ accepts the string x, then we denote by $t_{\mathbb{A}}(x)$ the minimum number of steps of an accepting run of $\mathbb{A}$ on x; if $\mathbb{A}$ does not accept x, then $t_{\mathbb{A}}(x)$ is not defined. By $L(\mathbb{A})$ we denote the language accepted by $\mathbb{A}$. We use deterministic and nondeterministic Turing machines with Σ as alphabet as our basic computational model for algorithms (and we often use the notions "algorithm" and "Turing machine" synonymously). If necessary we will not distinguish between a Turing machine and its code, a string in Σ^*. *By default, algorithms are deterministic.* If an algorithm $\mathbb{A}$ on input x eventually halts and outputs a value, we denote it by $\mathbb{A}(x)$.

2 Hard sequences for algorithms

In this section we derive the results concerning the existence of hard sequences for co-NP-complete problems.

Let $Q \subseteq \Sigma^*$. A deterministic (nondeterministic) algorithm $\mathbb{A}$ deciding (accepting) Q is *almost optimal* if for every deterministic (nondeterministic) algorithm $\mathbb{B}$ deciding (accepting) Q we have

$$t_{\mathbb{A}}(x) \le \big(t_{\mathbb{B}}(x) + |x|\big)^{O(1)}$$

for all $x \in Q$. Note that nothing is required for $x \notin Q$.

Clearly, every problem in P has an almost optimal algorithm and every problem in NP has an almost optimal nondeterministic algorithm. There are problems outside P with an almost optimal algorithm (see Messner [**10**, Corollary 3.33]; we slightly improve his result in Section 5). However, it is not known whether there are problems outside NP having an almost optimal nondeterministic algorithm and it is not known whether there are problems with padding outside P having an almost optimal algorithm. We show in Section 5 that the former is true if the Measure Hypothesis holds.

We introduce the notion of hard sequence.

Definition 2.1 Let $Q \subseteq \Sigma^*$.

(1) Let $\mathbb{A}$ be a deterministic (nondeterministic) algorithm deciding (accepting) Q. A sequence $(x_s)_{s\in\mathbb{N}}$ is *hard for* $\mathbb{A}$ if $\{x_s \mid s \in \mathbb{N}\} \subseteq Q$, the function $1^s \mapsto x_s$ is computable in polynomial time, and $t_{\mathbb{A}}(x_s)$ is not polynomially bounded in s.

(2) The problem Q *has hard sequences for algorithms* if every algorithm deciding Q has a hard sequence.

(3) The problem Q *has hard sequences for nondeterministic algorithms* if every nondeterministic algorithm accepting Q has a hard sequence.

The proof of the following lemma is straightforward; it shows that if $(x_s)_{s\in\mathbb{N}}$ is hard for an algorithm $\mathbb{A}$, then $\mathbb{A}$ can be polynomially speeded up on $\{x_s \mid s \in \mathbb{N}\}$; thus $\mathbb{A}$ cannot be almost optimal.

Lemma 2.2 *Let $\mathbb{A}$ be a deterministic (nondeterministic) algorithm deciding (accepting) Q. If $\mathbb{A}$ has a hard sequence, then $\mathbb{A}$ is not almost optimal.*

Proof. We prove the deterministic case, the nondeterministic case is obtained by the obvious modifications. So assume that the algorithm $\mathbb{A}$ decides Q and has a hard sequence $(x_s)_{s\in\mathbb{N}}$; in particular,

$$t_{\mathbb{A}}(x_s) \text{ is not polynomially bounded in } s. \tag{2.1}$$

Let $\mathbb{G}$ be a polynomial time algorithm computing the function $1^s \mapsto x_s$. The following algorithm $\mathbb{G}^*$ accepts the set $\{x_s \mid s \in \mathbb{N}\}$ and for $x = x_s$ runs in time polynomial in s.

```
$\mathbb{G}^*$    // $x \in \Sigma^*$
    1. $\ell \leftarrow 0$
    2. for $s = 0$ to $\ell$
    3.         simulate the $(\ell - s)$th step of $\mathbb{G}$ on $1^s$
    4.         if this simulation outputs $y$ and $y = x$
                 then accept and halt
    5. $\ell \leftarrow \ell + 1$
    6. goto 2.
```

We consider the algorithm $\mathbb{A}\|\mathbb{G}^*$ that on input x runs $\mathbb{A}$ and $\mathbb{G}^*$ in parallel, both on input x, and halts, when the first of these algorithms halts, then answering in the same way. Hence, $\mathbb{A}\|\mathbb{G}^*$ accepts Q and $t_{\mathbb{A}\|\mathbb{G}^*}(x_s)$ is polynomially bounded in s. As $|x_s| \le s^{O(1)}$, by (2.1) we see that $t_{\mathbb{A}}(x_s)$ is not polynomially bounded in $t_{\mathbb{A}\|\mathbb{G}^*}(x_s) + |x_s|$; thus $\mathbb{A}\|\mathbb{G}^*$ witnesses that $\mathbb{A}$ is not an almost optimal algorithm. □

We state the main result of this section (Remark 2.7 contains extensions of the result to further classes of problems Q). As already remarked in the Introduction part (b) of this theorem is a straightforward consequence of the corresponding result for propositional proof systems due to Krajíček.

Theorem 2.3 *Let Q be a* co-NP*-complete problem. Then:*

(a) *Q has no almost optimal algorithm $\iff$ Q has hard sequences for algorithms.*

(b) *Q has no almost optimal nondeterministic algorithm $\iff$ Q has hard sequences for nondeterministic algorithms.*

The proofs of the implications from right to left are clear by the previous lemma. The following considerations will yield a proof of the converse direction. For a nondeterministic algorithm $\mathbb{A}$ and $s \in \mathbb{N}$, let $\mathbb{A}^s$ be the algorithm that rejects all $x \in \Sigma^*$ with $|x| > s$. If $|x| \leq s$, then it simulates s steps of $\mathbb{A}$ on input x. If this simulation halts and accepts, then $\mathbb{A}^s$ accepts; otherwise it rejects.

Recall that by $L(\mathbb{A})$ we denote the language accepted by $\mathbb{A}$. For $Q \subseteq \Sigma^*$ we consider the *deterministic (nondeterministic) algorithm subset problem* $\textsc{Das}(Q)$ ($\textsc{Nas}(Q)$):

$\textsc{Das}(Q)$
Instance: A deterministic algorithm $\mathbb{A}$ and 1^s with $s \in \mathbb{N}$.
Problem: $L(\mathbb{A}^s) \subseteq Q$?

$\textsc{Nas}(Q)$
Instance: A nondeterministic algorithm $\mathbb{A}$ and 1^s with $s \in \mathbb{N}$.
Problem: $L(\mathbb{A}^s) \subseteq Q$?

The following two lemmas relate the equivalent statements in Theorem 2.3(a) (in Theorem 2.3(b)) to a statement concerning the complexity of $\textsc{Das}(Q)$ (of $\textsc{Nas}(Q)$).

Lemma 2.4

(a) *If* $\langle \mathbb{A}, 1^s \rangle \in \textsc{Das}(Q)$ *is solvable in time* $s^{f(\mathbb{A})}$ *for some function* f*, then* Q *has an almost optimal algorithm.*

(b) *If there is a nondeterministic algorithm* $\mathbb{V}$ *accepting* $\textsc{Nas}(Q)$ *such that for all* $\langle \mathbb{A}, 1^s \rangle \in \textsc{Nas}(Q)$ *we have* $t_{\mathbb{V}}(\langle \mathbb{A}, 1^s \rangle) \leq s^{f(\mathbb{A})}$ *for some function* f*, then* Q *has an almost optimal nondeterministic algorithm.*

Proof. Again we only prove (a). Let $\mathbb{V}$ be an algorithm deciding $\langle \mathbb{A}, 1^s \rangle \in \textsc{Das}(Q)$ in time $s^{f(\mathbb{A})}$ for some function f. Further let $\mathbb{Q}$ be an algorithm deciding Q and let $\mathbb{A}_0, \mathbb{A}_1, \ldots$ be an effective enumeration of all algorithms. Consider the following algorithm $\mathbb{A}$ deciding Q.

```
$\mathbb{A}$    // $x \in \Sigma^*$
    1. simulate $\mathbb{Q}$ on $x$ and in parallel do the following
    2. for $i = 0$ to $|x|$ do in parallel
    3.     simulate $\mathbb{A}_i$ on $x$
    4.     if $\mathbb{A}_i$ accepts then
    5.         $s \leftarrow \max\{|x|, \text{length of the run accepting } x\}$
    6.         if $\mathbb{V}$ accepts $\langle \mathbb{A}_i, 1^s \rangle$ then accept and halt
    7.         else never halt
    8.     else never halt
    9. if $\mathbb{Q}$ stops first then answer accordingly and halt.
```

It is easy to see that $\mathbb{A}$ decides Q. We show it is almost optimal. Let $\mathbb{B}$ be any algorithm deciding Q. We choose $i_{\mathbb{B}} \in \mathbb{N}$ such that $\mathbb{B} = \mathbb{A}_{i_{\mathbb{B}}}$. Note that $\mathbb{V}$ accepts $\langle \mathbb{B}, 1^s \rangle$ for all s. Hence for inputs $x \in Q$ with $|x| \geq i_{\mathbb{B}}$ the algorithm $\mathbb{A}$, for $i = i_{\mathbb{B}}$, accepts x in Line 6 if it was not already accepted earlier. Thus, $t_{\mathbb{A}}(x)$ is polynomially bounded in

$$|x| + t_{\mathbb{B}}(x) + t_{\mathbb{V}}\left(\left\langle \mathbb{B}, 1^{\max\{|x|, t_{\mathbb{B}}(x)\}} \right\rangle\right),$$

where the term $t_{\mathbb{B}}(x)$ takes care of line 3. Hence, by assumption, it is polynomially bounded in $|x| + \max\left\{|x|, t_{\mathbb{B}}(x)\right\}^{f(\mathbb{B})}$. Altogether, $t_{\mathbb{A}}(x) \leq \left(|x| + t_{\mathbb{B}}(x)\right)^{O(1)}$. □

If Q is co-NP-complete, then the problem $\textsc{Nas}(Q)$ and hence the problem $\textsc{Das}(Q)$ are in co-NP, too (this is the reason why 1^s and not just s is part of the input of $\textsc{Nas}(Q)$ and of $\textsc{Das}(Q)$). Thus, together with Lemma 2.4 the following lemma yields the remaining claims of Theorem 2.3.

Lemma 2.5

(a) *Assume that* $\textsc{Das}(Q) \leq_p Q$*, that is, that* $\textsc{Das}(Q)$ *is polynomial time reducible to* Q*. If* $\langle \mathbb{A}, 1^s \rangle \in \textsc{Das}(Q)$ *is not solvable in time* $s^{f(\mathbb{A})}$ *for some function* f*, then* Q *has hard sequences for algorithms.*

(b) *Assume that* $\textsc{Nas}(Q) \leq_p Q$*. If there is no nondeterministic algorithm* $\mathbb{V}$ *accepting* $\textsc{Nas}(Q)$ *such that for all* $\langle \mathbb{A}, 1^s \rangle \in \textsc{Nas}(Q)$ *we have* $t_{\mathbb{V}}(\langle \mathbb{A}, 1^s \rangle) \leq s^{f(\mathbb{A})}$ *for some function* f*, then* Q *has hard sequences for nondeterministic algorithms.*

Proof. Again we only prove part (a).

Claim Assume that $\langle \mathbb{A}, 1^s \rangle \in \textsc{Das}(Q)$ is not solvable in time $s^{f(\mathbb{A})}$ for some function f. Then there is no algorithm $\mathbb{W}$ deciding $\textsc{Das}(Q)$ such that for all algorithms $\mathbb{A}$ with $L(\mathbb{A}) \subseteq Q$ there is a $c_{\mathbb{A}} \in \mathbb{N}$ such that for all $s \in \mathbb{N}$ we have $t_{\mathbb{W}}(\langle \mathbb{A}, 1^s \rangle) \leq s^{c_{\mathbb{A}}}$.

Proof of the claim. By contradiction, assume that such a $\mathbb{W}$ exists. Let $\mathbb{V}$ be the algorithm that, on an arbitrary input $\langle \mathbb{A}, 1^s \rangle$, in parallel runs $\mathbb{W}$ on $\langle \mathbb{A}, 1^s \rangle$ and computes

$$r_{\mathbb{A}} := \text{the least } r \text{ such that } L(\mathbb{A}^r) \not\subseteq Q$$

by systematically checking for $r = 0, 1, \ldots$ whether $L(\mathbb{A}^r) \not\subseteq Q$ (this is done by running for all x with $|x| \leq r$ the algorithm $\mathbb{A}$ at most r steps on input x and a decision procedure for Q on x). Note that $r_{\mathbb{A}}$ is not defined if $L(\mathbb{A}) \subseteq Q$. If $\mathbb{W}$ stops first, $\mathbb{V}$ answers accordingly; if $r_{\mathbb{A}}$ is obtained first, then $\mathbb{V}$ accepts if $s < r_{\mathbb{A}}$ and otherwise it rejects. It should be clear that the algorithm $\mathbb{V}$ decides $\langle \mathbb{A}, 1^s \rangle) \in \textsc{Das}(Q)$ in $\leq s^{f(\mathbb{A})}$ steps for some function f. ⊣

By assumption, there is a polynomial time reduction $\mathbb{S}$ from $\textsc{Das}(Q)$ to Q. Let $\mathbb{B}$ be an arbitrary algorithm deciding Q. Then the algorithm $\mathbb{B} \circ \mathbb{S}$, which on input x first simulates $\mathbb{S}$ on x and then $\mathbb{B}$ on $\mathbb{S}(x)$, decides $\textsc{Das}(Q)$. Hence, by the Claim, there exists an algorithm $\mathbb{A}$ with $L(\mathbb{A}) \subseteq Q$ such that $t_{\mathbb{B} \circ \mathbb{S}}(\langle \mathbb{A}, 1^s \rangle)$ is not polynomially bounded in s. For $s \in \mathbb{N}$ we set $x_s := \mathbb{S}(\langle \mathbb{A}, 1^s \rangle)$. Then $x_s \in Q$ for all s and the function $1^s \mapsto x_s$ is polynomial time computable. Furthermore,

$$t_{\mathbb{B} \circ \mathbb{S}}(\langle \mathbb{A}, 1^s \rangle) \leq O\Big(t_{\mathbb{S}}(\langle \mathbb{A}, 1^s \rangle) + t_{\mathbb{B}}(\mathbb{S}(\langle \mathbb{A}, 1^s \rangle))\Big) \leq s^{O(1)} + O\big(t_{\mathbb{B}}(x_s)\big).$$

As the left hand side is not polynomially bounded in s, neither is $t_{\mathbb{B}}(x_s)$. Hence $(x_s)_{s \in \mathbb{N}}$ is hard for $\mathbb{B}$. □

Remark 2.6 Assume that Q is co-NP-complete and has padding (the set $\textsc{Taut}$ is an example of such a Q). If Q has no almost optimal algorithm, then every algorithm $\mathbb{B}$ deciding Q has a hard sequence $(x_s)_{s \in \mathbb{N}}$ with $s \leq |x_s|$. Then, in particular

$$\{x_s \mid s \in \mathbb{N}\} \in \mathrm{P} \text{ and } \mathbb{B} \text{ is not polynomial time on } \{x_s \mid s \in \mathbb{N}\}.$$

In fact, it is well-known that for Q with padding we can replace any polynomial time reduction to Q by a length-increasing one. Hence, in the previous proof we may assume that $\mathbb{S}$ is length-increasing and therefore $s \leq |x_s|$.

Remark 2.7 In the proof of Theorem 2.3 we used the assumption that Q is co-NP-complete only to ensure that $\textsc{Nas}(Q) \leq_p Q$ (cf. Lemma 2.5). This condition is also fulfilled for every Q complete, say, in one of the classes Π_t^p with $t \geq 1$, E or $\textsc{Pspace}$. Thus the statements of Theorem 2.3 hold for such Q.

The argument in the last part of Lemma 2.5 shows (an instance of) the following simple lemma. Nevertheless, note that it is important that we do not require $s \leq |x_s|$ in our definition of hard sequence.

Lemma 2.8 *Assume that $\mathbb{S}$ is a polynomial time reduction from Q to Q' and let $\mathbb{B}$ be a (nondeterministic) algorithm deciding (accepting) Q'. If $(x_s)_{s\in\mathbb{N}}$ is a hard sequence for $\mathbb{B} \circ \mathbb{S}$, then $(\mathbb{S}(x_s))_{s\in\mathbb{N}}$ is a hard sequence for $\mathbb{B}$.*

Therefore, if $Q \leq_p Q'$ and Q has hard sequences for (nondeterministic) algorithms then so does Q'.

We do not know proofs of the following results not using the machinery developed here.

Theorem 2.9 *Let Q be* co-NP*-complete. Then,* TAUT *has an almost optimal algorithm if and only if Q has an almost optimal algorithm.*

Proof. Immediate by the previous lemma and Theorem 2.3. □

We remark that the implication from left to right in the previous result was already known [**7**] (see also Theorem 3.2 below).

Theorem 2.10 *Assume that* TAUT *has no almost optimal algorithm. Then every* co-NP*-hard problem has no almost optimal algorithm.*

Proof. By assumption and Theorem 2.3, TAUT has hard sequences for algorithms and so does every co-NP-hard Q by Lemma 2.8. Now the claim follows from Lemma 2.2. □

3 Hard sequences for proof systems

In this section we translate the results on hard sequences from algorithms to proof systems. We first recall some basic definitions.

A *proof system for Q* is a polynomial time algorithm $\mathbb{P}$ computing a function from Σ^* onto Q. If $\mathbb{P}(w) = x$, we say that w is a $\mathbb{P}$-*proof* of x. Often we introduce proof systems implicitly by defining the corresponding function; then this definition will suggest a corresponding algorithm.

Definition 3.1 Let $\mathbb{P}$ and $\mathbb{P}'$ be proof systems for Q. An algorithm $\mathbb{T}$ is a *translation from $\mathbb{P}'$ into $\mathbb{P}$* if $\mathbb{P}(\mathbb{T}(w')) = \mathbb{P}'(w')$ for every $w' \in \Sigma^*$. Note that translations always exist. A translation is *polynomial* if it runs in polynomial time.

A proof system $\mathbb{P}$ for Q is *p-optimal* or *polynomially optimal* if for every proof system $\mathbb{P}'$ for Q there is a polynomial translation from $\mathbb{P}'$ into $\mathbb{P}$. A proof system $\mathbb{P}$ for Q is *optimal* if for every proof system $\mathbb{P}'$ for Q and every $w' \in \Sigma^*$ there is a $w \in \Sigma^*$ such that $\mathbb{P}(w) = \mathbb{P}'(w')$ and $|w| \leq |w'|^{O(1)}$. Clearly, every p-optimal proof system is optimal.

We will often make use of the following relationship between the optimality notions for algorithms and that for proof systems (see [**7, 10**]).

Theorem 3.2

(1) *For every Q we have* (a) ⇒ (b) *and* (b) ⇒ (c)*; moreover* (a), (b), *and* (c) *are all equivalent if Q has padding. Here*

(a) *Q has a p-optimal proof system;*

(b) *Q has an almost optimal algorithm;*

(c) *There is an algorithm that decides Q and runs in polynomial time on every subset X of Q with $X \in \mathrm{P}$.*

(2) *For every Q we have* (a) $\Leftrightarrow$ (b), (b) $\Rightarrow$ (c), *and* (c) $\Rightarrow$ (d)*; moreover* (a)–(d) *are all equivalent if Q has padding. Here*

(a) *Q has an optimal proof system;*

(b) *Q has an almost optimal nondeterministic algorithm;*

(c) *There is a nondeterministic algorithm that accepts Q and runs in polynomial time on every subset X of Q with $X \in \mathrm{NP}$;*

(d) *There is a nondeterministic algorithm that accepts Q and runs in polynomial time on every subset X of Q with $X \in \mathrm{P}$.*

We use our results of Section 2 to extend the equivalence between (a) and (b) of part (1) to arbitrary co-NP-complete problems:

Theorem 3.3 *Let Q be* co-NP*-complete. Then: Q has a p-optimal proof system if and only if Q has an almost optimal algorithm.*

Proof. By Theorem 3.2(1) the left side implies the right side. Now assume that Q has an almost optimal algorithm. As $Q \times \Sigma^*$ is co-NP-complete too, it has an almost optimal algorithm (by Theorem 2.9). As $Q \times \Sigma^*$ has padding, it has a p-optimal proof system $\mathbb{P}$ (cf. Theorem 3.2(1)). Now it is routine to show that the algorithm $\mathbb{P}'$ that on input w computes $\mathbb{P}(w)$ and outputs its first component is a p-optimal proof system for Q. □

We already mentioned that for every $Q \subseteq \Sigma^*$ there is a well-known and straightforward correspondence between proof systems and nondeterministic algorithms preserving the optimality notions, so that the proof of the equivalence between (a) and (b) in Theorem 3.2(2) is immediate, In fact, if $\mathbb{P}$ is a proof system for Q, then the nondeterministic algorithm $\mathbb{A}(\mathbb{P})$ accepts Q, where $\mathbb{A}(\mathbb{P})$ on input $x \in \Sigma^*$ guesses a string w and accepts if $\mathbb{P}(w) = x$. Conversely, if $\mathbb{A}$ is a nondeterministic algorithm accepting Q, then for every fixed $x_0 \in Q$ a proof system $\mathbb{P}_\mathbb{A}$ for Q is defined by

$$\mathbb{P}_\mathbb{A}(w) := \begin{cases} x, & \text{if } w \text{ is a computation of } \mathbb{A} \text{ accepting } x, \\ x_0, & \text{otherwise.} \end{cases}$$

The proof of the corresponding equivalence in Theorem 3.2(1) is more involved and mostly, more or less explicitly, it is based on a theorem due to Levin on inverters. As we need this result, too, we recall it.

Let $\mathbb{F}$ be an algorithm computing a function from Σ^* to Σ^*. An *inverter of* $\mathbb{F}$ is an algorithm $\mathbb{I}$ that given y in the range of $\mathbb{F}$ halts with some output $\mathbb{I}(y)$ such that $\mathbb{F}(\mathbb{I}(y)) = y$. On inputs not in the range of $\mathbb{F}$, the algorithm $\mathbb{I}$ may do whatever it wants. Levin [**8**] proved the following result.

Theorem 3.4 *Let $\mathbb{F}$ be an algorithm computing a function from Σ^* into Σ^*. Then there is an optimal inverter that is, an inverter $\mathbb{O}_\mathbb{F}$ of $\mathbb{F}$ such that for every inverter $\mathbb{I}$ of $\mathbb{F}$ and all y in the range of $\mathbb{F}$ we have*

$$t_{\mathbb{O}_\mathbb{F}}(y) \leq \big(t_\mathbb{I}(y) + t_\mathbb{F}(\mathbb{I}(y)) + |y|\big)^{O(1)}.$$

Furthermore, $\mathbb{O}_\mathbb{F}$ does not halt on inputs y not in the range of $\mathbb{F}$.

We turn to hard sequences for proof systems.

Definition 3.5 Let $\mathbb{P}$ be a proof system for Q. A sequence $(x_s)_{s\in\mathbb{N}}$ is *hard (length-hard) for* $\mathbb{P}$ if $\{x_s \mid s \in \mathbb{N}\} \subseteq Q$, the function $1^s \mapsto x_s$ is computable in polynomial time, and there is no polynomial time (nondeterministic) algorithm $\mathbb{W}$ with $\mathbb{P}(\mathbb{W}(1^s)) = x_s$ for all $s \in \mathbb{N}$.

For nondeterministic $\mathbb{W}$ by the unusual notation $\mathbb{P}(\mathbb{W}(1^s)) = x_s$ we mean that for every run of $\mathbb{W}$ on 1^s outputting a string w we have $\mathbb{P}(w) = x_s$ and that there is at least one run that outputs a string. In more conventional terms, instead of "there is no polynomial time nondeterministic algorithm $\mathbb{W}$ with $\mathbb{P}(\mathbb{W}(1^s)) = x_s$", we equivalently could require that the function mapping 1^s to the minimum length in unary of a $\mathbb{P}$-proof of x_s is not polynomially bounded.

Definition 3.6 The problem Q *has hard (length-hard) sequences for proof systems* if every proof system for Q has a hard (length-hard) sequence.

As already remarked in the Introduction part (b) of the following result is due to Krajíček [**6**] who proved it by quite different means. Part (a) is already known for $Q = \text{Taut}$ (see e.g. the survey [**1**, Section 11]). We give a new proof that works for any, not necessarily paddable coNP-complete problem Q.

Theorem 3.7 *Let Q be a* co-NP*-complete problem. Then:*

(a) *Q has no p-optimal proof system iff Q has hard sequences for proof systems.*
(b) *Q has no optimal proof system iff Q has a length-hard sequence for proof systems.*

Proof. First we present a proof of the directions from right to left. Let $\mathbb{P}$ be any proof system for Q. By our assumption on Q there is a hard (length-hard) sequence $(x_s)_{s\in\mathbb{N}}$ for $\mathbb{P}$. We consider the proof system $\mathbb{P}'$ for Q by

$$\mathbb{P}'(w') := \mathbb{P}(w), \text{ if } w' = 0w; \qquad \mathbb{P}'(w') := x_s, \text{ if } w' = 1^s;$$

and $\mathbb{P}'(w') := z_0$ for some fixed element z_0 of Q otherwise. By hardness (length-hardness) no translation from $\mathbb{P}'$ into $\mathbb{P}$ is polynomial (polynomially bounded). In fact, assume that $(x_s)_{s\in\mathbb{N}}$ is, say, length-hard for $\mathbb{P}$ and by contradiction that the translation $\mathbb{T}$ from $\mathbb{P}'$ into $\mathbb{P}$ is polynomially bounded. Let q be a polynomial such that $|\mathbb{T}(w')| \leq q(|w'|)$ for all w'. Then, the nondeterministic algorithm $\mathbb{W}$ that on input 1^s guesses a string w of length $\leq q(s)$ and outputs it in case $\mathbb{P}(w) = x_s$ runs in polynomial time.

Now we present a proof of the direction from left to right; we do that only for (a) as that for (b) follows immediately from the result for algorithms by the simple correspondence between proof systems and nondeterministic algorithms mentioned above. So, assume that Q has no p-optimal proof system. By Theorem 3.3, Q has no almost optimal algorithm and hence has hard sequences for algorithms by Theorem 2.3.

Let $\mathbb{P}$ be any proof system for Q. By Theorem 3.4, we have an inverter $\mathbb{O}_\mathbb{P}$ of $\mathbb{P}$ which is optimal, that is, for every inverter $\mathbb{I}$ of $\mathbb{P}$ and $x \in Q$ we have

$$t_{\mathbb{O}_\mathbb{P}}(x) \leq \big(t_\mathbb{I}(x) + t_\mathbb{P}(\mathbb{I}(x)) + |x|\big)^{O(1)} \leq (t_\mathbb{I}(x) + |x|)^{O(1)}, \tag{3.1}$$

where the second inequality holds as $t_\mathbb{P}(w) \leq |w|^{O(1)}$ and hence $t_\mathbb{P}(\mathbb{I}(x)) \leq |\mathbb{I}(x))|^{O(1)} \leq t_\mathbb{I}(x)^{O(1)}$. Moreover, for $x \notin Q$ the algorithm $\mathbb{O}_\mathbb{P}$ will not halt on input x.

We choose an arbitrary algorithm $\mathbb{Q}$ that decides Q and consider the algorithm $\mathbb{S}$ that on input x in parallel simulates $\mathbb{Q}$ and $\mathbb{O}_\mathbb{P}$, both on input x. If $\mathbb{Q}$ halts first, then it answers accordingly and if $\mathbb{Q}_\mathbb{P}$ halts first, then it accepts. Obviously $\mathbb{S}$ decides Q and for every $x \in Q$ we have

$$t_{\mathbb{S}}(x) \leq O\big(t_{\mathbb{O}_{\mathbb{P}}}(x)\big). \tag{3.2}$$

As Q has hard sequences for algorithms, there is a polynomial time computable algorithm $\mathbb{G}$ generating a hard sequence for $\mathbb{S}$, that is, $\mathbb{G}$ on input 1^s computes $x_s \in Q$ in polynomial time such that

$$t_{\mathbb{S}}(x_s) \text{ is not polynomially bounded in } s. \tag{3.3}$$

Let $\mathbb{G}^+$ be the variant of the algorithm $\mathbb{G}^*$ in the proof of Lemma 2.2 obtained by replacing Line 4 by

if this simulation outputs y and $y = x$ **then** output 1^s and halt.

Of course, on input $x = x_s$ the algorithm $\mathbb{G}^+$ runs in time polynomial in s. We show that $(x_s)_{s\in\mathbb{N}}$ is a hard sequence for $\mathbb{P}$. So by contradiction, assume that $\mathbb{W}$ is a polynomial time algorithm with $\mathbb{P}(\mathbb{W}(1^s)) = x_s$ for all $s \in \mathbb{N}$. We consider the inverter $\mathbb{I}$ of $\mathbb{P}$ that on input x in parallel simulates $\mathbb{O}_{\mathbb{P}}$ and $\mathbb{G}^+$, both on input x. If $\mathbb{O}_{\mathbb{P}}$ halts, then it outputs the output of $\mathbb{O}_{\mathbb{P}}$ and halts; if $\mathbb{G}^+$ halts, then it simulates $\mathbb{W}$ on $\mathbb{G}^+(x)$, outputs $\mathbb{W}(\mathbb{G}^+(x))$, and halts.

By definition of $\mathbb{G}^+$ the algorithm $\mathbb{I}$ runs on input x_s in time polynomial in s, hence so does $\mathbb{O}_{\mathbb{P}}$ by (3.1) as $|x_s| \leq s^{O(1)}$. But then by (3.2), the same holds for the algorithm $\mathbb{S}$ contradicting (3.3). □

In the previous proof the hard (length-hard) sequence $(x_s)_{s\in\mathbb{N}}$ constructed for a proof system for Q was the hard sequence of a suitable (nondeterministic) algorithm for Q. Hence, by Remark 2.6, for Q with padding, we can require in Theorem 3.7 that for the claimed hard sequence $(x_s)_{s\in\mathbb{N}}$ we have $s \leq |x_s|$.

4 Hard subsets

As already remarked in the Introduction, if for an algorithm $\mathbb{A}$ deciding a problem Q we have a hard sequence $(x_s)_{s\in\mathbb{N}}$ satisfying $s \leq |x_s|$, then $\{x_s \mid s \in \mathbb{N}\}$ is a polynomial time decidable subset of Q on which $\mathbb{A}$ is not polynomial time. We then speak of a hard set for $\mathbb{A}$ even if its elements cannot be generated in polynomial time. More precisely:

Definition 4.1 Let $Q \subseteq \Sigma^*$.

(1) Let $\mathbb{A}$ be a deterministic or nondeterministic algorithm accepting Q. A subset X of Q is *hard for* $\mathbb{A}$ if $X \in \mathrm{P}$ and $\mathbb{A}$ is not polynomial time on X.

(2) The problem Q *has hard sets for algorithms* if every algorithm deciding Q has a hard set.

(3) The problem Q *has hard sets for nondeterministic algorithms* if every nondeterministic algorithm accepting Q has a hard set.

Using these notions the equivalences (a) ⇔ (c) in Theorem 3.2 can be expressed in the following way:

Assume that Q has padding. Then:

(1) *Q has no almost optimal algorithm $\iff$ Q has hard sets for algorithms.*

(2) *Q has no almost optimal nondeterministic algorithm $\iff$ Q has hard sets for nondeterministic algorithms.*

Hence, we get (we leave the nondeterministic variant to the reader):

Corollary 4.2 *Assume Q has padding.*

(a) *If Q has hard sequences for algorithms, then Q has hard sets for algorithms.*

(b) *If in addition Q is* co-NP*-complete, then Q has hard sequences for algorithms if and only if Q has hard sets for algorithms.*

Proof. (a) If Q has hard sequences for algorithms, then, by Lemma 2.2, Q has no almost optimal algorithm and thus, by the previous remark, Q has hard sets for algorithms.

Again the previous remark together with Theorem 2.3 yields (b). □

Assume that Q has an almost optimal algorithm. Then, in general, one cannot show that every algorithm deciding Q, which is not almost optimal, has a hard set. In fact, Messner [**10**, Corollary 3.33] has presented a P-immune Q_0 with an almost optimal algorithm. Of course, no algorithm deciding Q_0 has a hard set.

For an arbitrary problem Q the existence of hard subsets is equivalent to a (non-) listing property. We introduce this property.

Let C be the complexity class P or NP. A set X is a C-subset of Q if $X \subseteq Q$ and $X \in C$. Let C' be also one of the classes P or NP. We write $\mathrm{List}(C, Q, C')$ and say that there is a *listing of the C-subsets of Q by C'-machines* if there is an algorithm that, once having been started, lists Turing machines $\mathbb{M}_1, \mathbb{M}_2, \dots$ of type C' such that

$$\big\{L(\mathbb{M}_i) \mid i \geq 1\big\} = \big\{X \subseteq Q \mid X \in C\big\}.$$

For Q with padding the equivalences in the following proposition were known [**12**].

Proposition 4.3

(1) *Q has hard sets for algorithms $\iff$ not* $\mathrm{List}(\mathrm{P}, Q, \mathrm{P})$.

(2) *Every nondeterministic algorithm $\mathbb{A}$ accepting Q is not polynomial on at least one subset X of Q with $X \in$* NP $\iff$ *not* $\mathrm{List}(\mathrm{NP}, Q, \mathrm{NP})$.

Proof. We only prove the first claim as the second one can be obtained along the same lines. First we assume that not $\mathrm{List}(\mathrm{P}, Q, \mathrm{P})$. Let $\mathbb{A}$ be an algorithm deciding Q. For $d \in \mathbb{N}$, by $\mathbb{A}(d)$ we denote the algorithm that on input x simulates $\mathbb{A}$ on input x but rejects if the simulation exceeds time $|x|^d$.

We show that there is a P-subset X of Q such that, for all d,

$$X \not\subseteq \mathbb{A}(d).$$

Of course, then this X is hard for $\mathbb{A}$.

Otherwise, we fix an effective enumeration $\mathbb{D}_1, \mathbb{D}_2, \dots$ of all polynomial time Turing machines. Then $(\mathbb{D}_i(\mathbb{A}(j))_{i,j\geq 1}$ is a listing of the P-subsets of Q, where $\mathbb{D}_i(\mathbb{A}(j))$ on input x, first simulates $\mathbb{A}(j)$ on x and if this algorithm accepts, then it simulates $\mathbb{D}_i$ on input x and answers accordingly. In fact, as $\mathbb{A}(j)$ has to accept x, we have $L(\mathbb{D}_i(\mathbb{A}(j))) \subseteq Q$. And if X is a P-subset of Q accepted by $\mathbb{D}_i$, we choose a d such that $X \subseteq \mathbb{A}(d)$. Then $L(\mathbb{D}_i(\mathbb{A}(d))) = X$.

Conversely, assume that Q has hard sets for algorithms. By contradiction assume that $\mathbb{L}$ is a listing witnessing $\mathrm{List}(\mathrm{P}, Q, \mathrm{P})$. Let $\mathbb{Q}$ be an algorithm deciding Q. Consider the algorithm $\mathbb{A}$ that on input x simulates $\mathbb{Q}$ on x and in parallel for $i = 1, 2, \dots$ does the following:

- performs the ith step of $\mathbb{L}$;
- if $\mathbb{M}_1, \dots, \mathbb{M}_s$ are the machines listed by $\mathbb{L}$ so far, it performs an additional step of each of the $\mathbb{M}_j$s on x; if one of these accepts it accepts.

If $\mathbb{Q}$ halts first, it answers accordingly.

It should be clear that $\mathbb{A}$ accepts Q. By assumption, there is a set X hard for $\mathbb{A}$. Let $\mathbb{M}_{i_0}$ accept X. By definition of $\mathbb{A}$ it should be clear that $\mathbb{A}$ is polynomial on X, a contradiction. □

We close this section by introducing hard subsets for proof systems and stating the corresponding result.

Definition 4.4

(1) Let $\mathbb{P}$ be a proof system for Q. A subset X of Q is *hard (length-hard) for* $\mathbb{P}$ if $X \in \mathrm{P}$ and there is no polynomial time (nondeterministic) algorithm $\mathbb{W}$ such that $\mathbb{P}(\mathbb{W}(x)) = x$ for all $x \in X$ (cf. the remark after Definition 3.5 for the precise meaning of this last condition in the nondeterministic case).

(2) *Q has hard (length-hard) sets for proof systems* if every proof system for Q has a hard (length-hard) set.

The following result can be obtained along the lines of the proof of Theorem 3.7. Again, due to the close relationship between nondeterministic algorithms and proof systems, part (b) can be viewed as a reformulation of the result for algorithms.

Theorem 4.5 *Let Q be a problem with padding. Then:*

(a) *Q has no p-optimal proof system $\iff$ Q has hard sets for proof systems.*

(b) *Q has no optimal proof system $\iff$ Q has length-hard sets for proof systems.*

5 Assuming the Measure Hypothesis

In this section we present some examples of problems with special properties; some yield limitations to possible extensions of results mentioned in this paper. Most are proven assuming the Measure Hypothesis.

5.1 Complex sets with optimal algorithms and with optimal proof systems

For every $Q \in \mathrm{NP}$, say, accepted by the polynomial time nondeterministic algorithm $\mathbb{A}$, the proof system $\mathbb{P}$ is optimal, where $\mathbb{P}(w) := x$ if w is an accepting computation of $\mathbb{A}$ on input x; and otherwise, $\mathbb{P}(w) := z_0$ for some fixed element z_0 of Q. The question whether there are sets outside of NP with optimal proof systems was stated by Krajíček and Pudlák [**7**] and is still open. As already mentioned they proved that Taut has an optimal proof system if E = NE.

We prove that there are problems in NE and outside of NP with optimal proof systems if the Measure Hypothesis holds. As a byproduct we get that there exist problems in E and outside of P with optimal algorithms (thereby we do not need the Measure Hypothesis). Here an algorithm $\mathbb{A}$ deciding Q is *optimal* if for every algorithm $\mathbb{B}$ deciding Q we have

$$t_{\mathbb{A}}(x) \leq (t_{\mathbb{B}}(x) + |x|)^{O(1)}$$

for *all* $x \in \Sigma^*$. Clearly, every problem in P has an optimal algorithm.

Let C be a class of problems. Recall that a problem Q is *C-immune* if no infinite subset of Q is in C; and it is *C-bi-immune* if Q and its complement $\Sigma^* \setminus Q$ are C-immune. For a function $t\colon \mathbb{N} \to \mathbb{N}$ we denote by $\mathrm{D{\small TIME}}_0(t)$ and $\mathrm{DTIME}(t)$ the class of problems decidable by a Turing machine $\mathbb{M}$ with $t_{\mathbb{M}}(x) \leq t(x)$ for all $x \in \Sigma^*$ and

$t_{\mathbb{M}}(x) \le c \cdot t(x)$ for all $x \in \Sigma^*$ and some constant $c \in \mathbb{N}$. The nondeterministic classes $\textsc{Ntime}_0(t)$ and $\textsc{Ntime}(t)$ are defined accordingly. Hence $\mathrm{E} = \bigcup_{d\in\mathbb{N}} \mathrm{DTIME}(2^{d\cdot n})$ and $\mathrm{NE} = \bigcup_{d\in\mathbb{N}} \textsc{Ntime}(2^{d\cdot n})$.

Lemma 5.1 *Let $\ell \in \mathbb{N}$ with $\ell \ge 1$.*

(a) *If $Q \in \mathrm{E}$ is a $\textsc{Dtime}_0(2^{\ell\cdot n})$-bi-immune problem, then Q has an optimal algorithm.*

(b) *If $Q \in \mathrm{NE}$ is a $\textsc{Ntime}_0(2^{\ell\cdot n})$-immune problem, then Q has an almost optimal nondeterministic algorithm.*

Proof. We prove (a); part (b) is obtained by the obvious modifications. Assume that the Turing machine $\mathbb{M}$ decides the $\textsc{Dtime}_0(2^{\ell\cdot n})$-bi-immune problem Q in time $c \cdot 2^{d\cdot n}$ for some $c, d \in \mathbb{N}$. We claim that $\mathbb{M}$ is optimal.

Assume otherwise, then there is a machine $\mathbb{M}'$ deciding Q and witnessing that $\mathbb{M}$ is not optimal. Then for every $i \in \mathbb{N}$ there exists an x_i such that

$$t_{\mathbb{M}}(x_i) > \big(t_{\mathbb{M}'}(x_i) + |x_i|\big)^i.$$

It follows that for every $i \in \mathbb{N}$

$$c \cdot 2^{d\cdot|x_i|} \ge t_{\mathbb{M}}(x_i) > t_{\mathbb{M}'}(x_i)^i.$$

Thus $t_{\mathbb{M}'}(x_i) \le 2^{\ell\cdot|x_i|/2}$ for all sufficiently large $i \in \mathbb{N}$. Of course, infinitely many of these x_i's are in Q, or they are in $\Sigma^* \setminus Q$. In the first case consider the following machine:

```
M''        // $x \in \Sigma^*$
    1. simulate $\mathbb{M}'$ on $x$ for at most $2^{\ell\cdot|x|/2}$ steps
    2. if the simulation halts and accepts then accept else reject.
```

It accepts an infinite subset of Q in time $2^{\ell\cdot n}$. This contradicts our immunity assumption. The second case is handled similarly. □

We use the following result due to Mayordomo [**9**]. Statement (b) of it uses the *Measure Hypothesis* [**5**], that is, the assumption

NP does not have measure 0 in E.

For the corresponding notion of measure we refer to [**9**]. This hypothesis is sometimes used in the theory of resource bounded measures.

Theorem 5.2 *Let $\ell \ge 1$.*

(a) *The class of $\textsc{Dtime}_0(2^{\ell\cdot n})$-bi-immune problems has measure 1 in E. In particular, the class E contains $\textsc{Dtime}_0(2^{\ell\cdot n})$-bi-immune problems.*

(b) *If the Measure Hypothesis holds, then $\mathrm{NP} \cap \mathrm{E}$ contains $\textsc{Dtime}_0(2^{\ell\cdot n})$-bi-immune problems.*

From the previous lemma and theorem we get:

Corollary 5.3

(1) *There exist problems in $\mathrm{E} \setminus \mathrm{P}$ with optimal algorithms.*

(2) *If the Measure Hypothesis holds, then there exist problems in $\mathrm{NP}\setminus\mathrm{P}$ with optimal algorithms.*

We already remarked that Messner [**10**] showed the existence of problems in E $\setminus$ P with *almost* optimal algorithms.

Theorem 5.4 *If the Measure Hypothesis holds, then there exist problems in* NE $\setminus$ NP *with optimal proof systems.*

Proof. It suffices to show that there is a $Q \in \mathrm{NE}$ which is $\textsc{Ntime}_0(2^n)$-immune. Then, by Lemma 5.1, such a Q has an almost optimal nondeterministic algorithm and hence, an optimal proof system by Theorem 3.2.

By Theorem 5.2(b) there is a $Q_0 \in \mathrm{NP}$ which is $\textsc{Dtime}_0(2^{2n})$-bi-immune problem. We choose $d \geq 1$ such that $Q_0 \in \textsc{Ntime}(n^d)$. We set

$$Q := \left\{1^m \mid m \in \mathbb{N} \text{ and } 1^{2^m} \in Q_0\right\}.$$

Then $Q \in \mathrm{NE}$. Moreover, Q is infinite as otherwise the set $\{1^{2^m} \mid m \in \mathbb{N} \text{ and } 1^{2^m} \notin Q_0\}$ would be an infinite subset of $\Sigma^* \setminus Q_0$ in P contradicting the bi-immunity property of Q_0. Finally we show that Q is $\textsc{Ntime}_0(2^n)$-immune. By contradiction assume that there is an infinite $S \subseteq Q$ accepted by a nondeterministic algorithm $\mathbb{S}$ in time 2^n. Then the set

$$S^* := \left\{1^n \mid n = 2^m \text{ for some } m \in \mathbb{N} \text{ and } 1^m \in S\right\}$$

is an infinite subset of Q_0. The algorithm that first computes m from 1^n and then deterministically simulates all possible runs of $\mathbb{S}$ on 1^m runs in time

$$n^{O(1)} + O(2^{2^m}) = n^{O(1)} + O(2^n) \leq 2^{2n}$$

for sufficiently large n. This contradicts the $\textsc{Dtime}_0(2^{2n})$-immunity of Q_0. □

5.2 Non-optimal algorithms without hard sequences

In this final part we show that, assuming the Measure Hypothesis,

- every problem with padding and with an almost optimal algorithm has an algorithm which is not almost optimal but has no hard sequence;
- there is a problem without almost optimal algorithm having an algorithm without hard sequence.

Our proofs are based on the following proposition.

Proposition 5.5 *If the Measure Hypothesis holds, then there is a problem* $Q_0 \in \mathrm{P}$ *such that*

(a) *there is an algorithm* $\mathbb{B}$ *deciding* Q_0 *which is not almost optimal (or, equivalently, is not polynomial time) but has no hard sequences;*

(b) *every algorithm* $\mathbb{A}$ *deciding* Q_0 *with*

$$t_{\mathbb{A}}(x) \leq 2^{e \cdot (\log |x|)^2}$$

for every $x \in \Sigma^*$ *and some constant* $e \geq 1$ *has no hard sequences;*

(c) *there is a proof system for* Q_0 *which is not optimal but has no hard sequences.*

In the proof we shall use:

Lemma 5.6 *Let* $\mathbb{A}$ *be an algorithm deciding a problem* Q_0 *with*

$$t_{\mathbb{A}}(x) \leq 2^{e \cdot (\log |x|)^2} \tag{5.1}$$

for all $x \in \Sigma^*$ and some $e \geq 1$. Assume that $(x_s)_{s\in\mathbb{N}}$ is a hard sequence for $\mathbb{A}$. Then there is a sequence $s_0 < s_1 < s_2 < \dots$ such that

$$\lim_{i\to\infty} \frac{\log s_i}{(\log|x_{s_i}|)^2} = 0, \quad \textit{i.e.,} \quad s_i = 2^{o\left((\log|x_{s_i}|)^2\right)}.$$

In particular, the set $\{x_{s_i} \mid i \in \mathbb{N}\}$ is infinite.

Proof. Assume otherwise that, for some $\varepsilon > 0$ and some $n \in \mathbb{N}$ and all $s \geq n$,

$$\frac{\log s}{(\log|x_s|)^2} \geq \varepsilon,$$

or equivalently, $s \geq 2^{\varepsilon\cdot(\log|x_s|)^2}$; then $s \geq t_{\mathbb{A}}(x_s)^{\varepsilon/e}$ by assumption. This contradicts the hardness of $(x_s)_{s\in\mathbb{N}}$. □

Proof of Proposition 5.5. (a) and (b): By the Measure Hypothesis there is a $\text{DTIME}_0(2^n)$-bi-immune $Q_1 \in \text{NP}$. In particular, there exists a nondeterministic Turing machine $\mathbb{M}$ with binary nondeterminism and a $d \in \mathbb{N}$ such that for all $y \in \Sigma^*$ (with $|y| \geq 2$) the machine $\mathbb{M}$ decides whether $y \in Q_1$ in $\leq |y|^d$ steps. Thus for $y \in \Sigma^*$ every string $x \in \{0,1\}^{|y|^d}$ determines a unique run of $\mathbb{M}$ on y. We set

$$Q_0 := \big\{x \in \{0,1\}^* \mid \text{for some } n \in \mathbb{N} \text{ we have } |x| = n^d \text{ and} \\ x \text{ determines an accepting run of } \mathbb{M} \text{ on input } 1^n\big\}.$$

Then Q_0 is infinite, as otherwise the set $\{1^n \in Q_1 \mid n \in \mathbb{N}\}$ would be finite contradicting the $\text{DTIME}_0(2^n)$-bi-immunity of Q_1. Clearly $Q_0 \in \text{P}$. Let $\mathbb{A}_0$ be an algorithm deciding Q_0 in polynomial time and let $\mathbb{B}$ be the algorithm deciding Q_0 by first simulating $\mathbb{A}$, and then making an appropriate number of dummy steps such that, for some $e \geq 1$ and all $y \in \Sigma^*$,

$$t_{\mathbb{B}}(y) = 2^{e\cdot(\log|y|)^2}. \tag{5.2}$$

Then $\mathbb{A}_0$ witnesses that $\mathbb{B}$ is not almost optimal.

We finish our proof by showing that for every algorithm $\mathbb{A}$ deciding Q_0 such that, for some $e \geq 1$ and all $y \in \Sigma^*$,

$$t_{\mathbb{A}}(y) \leq 2^{e\cdot(\log|y|)^2},$$

has no hard sequences. Towards a contradiction assume $\mathbb{A}$ has a hard sequence $(x_s)_{s\in\mathbb{N}}$. We set

$$L_0 := \big\{1^n \mid \text{for some } s \in \mathbb{N},\ |x_s| = n^d \text{ and } x_s \text{ determines an accepting run of } \mathbb{M} \text{ on } 1^n\big\}.$$

Clearly, $L_0 \subseteq Q_1$. We choose a polynomial time algorithm $\mathbb{G}$ computing the function $1^s \mapsto x_s$. The following algorithm $\mathbb{C}$ accepts L_0.

```
C       // $y \in \Sigma^*$
    1. $n \leftarrow |y|$
    2. if $y \neq 1^n$ then reject
    3. $\ell \leftarrow 0$
    4. for $s = 0$ to $\ell$
    5.         simulate the $(\ell - s)$th step of $\mathbb{G}$ on $1^s$
    6.         if the simulation outputs $x$ with $|x| = n^d$ then accept
    7. $\ell \leftarrow \ell + 1$
    8. goto 3.
```

By (5.2) we can apply Lemma 5.6 to $\mathbb{A}$ and get a sequence $s_0 < s_1 < s_2 < \dots$. For $i \in \mathbb{N}$ we let

$$n_i := \sqrt[d]{|x_{s_i}|}. \tag{5.3}$$

Hence, x_{s_i} is an accepting run of $\mathbb{M}$ on input 1^{n_i}. We show that

$$t_{\mathbb{C}}(1^{n_i}) = 2^{o\left((\log n_i)^2\right)}. \tag{5.4}$$

In fact, as $\mathbb{G}$ runs in polynomial time, we have $|x_{s_i}| \le |s_i|^{O(1)}$, and by (5.3) therefore, $|n_i| \le |s_i|^{O(1)}$. Now one easily sees that $\mathbb{C}$ accepts 1^{n_i} in time polynomial in s_i, too. By Lemma 5.6,

$$s_i = 2^{o\left((\log |x_{s_i}|)^2\right)}.$$

Thus (5.3) implies that

$$s_i = 2^{o\left((\log n_i)^2\right)}.$$

Hence, we get (5.4).

Finally, we consider the algorithm $\mathbb{C}^*$ that on input y simulates $\mathbb{C}$ for $2^{|y|}$ steps and accepts if the simulation accepts. By (5.4), $\mathbb{C}^*$ accepts an infinite subset of L_0. As $L_0 \subseteq Q_1$, this contradicts the $\mathrm{DTIME}_0(2^n)$-bi-immunity of Q_1.

To prove (c), let Q_0 and $\mathbb{B}$ be as in part (a). We leave it to the reader to show that the following proof system $\mathbb{P}$ for Q_0 is not optimal but has no hard sequence. For $w \in \Sigma^*$, let

$$\mathbb{P}(w) := x \quad \text{if } w \text{ is a computation of } \mathbb{B} \text{ accepting } x,$$

and $\mathbb{P}(w) := z_0$ for some fixed $z_0 \in Q_0$ otherwise. □

Theorem 5.7 *Let Q be a problem with padding and with an almost optimal algorithm. If the Measure Hypothesis holds, then there is an algorithm deciding Q, which is not optimal, has hard sets but does not have hard sequences.*

Proof. Let *pad* and $\mathbb{O}$ be a padding function and an almost optimal algorithm for Q, respectively. With Proposition 5.5(a) choose a $Q_0 \in \mathrm{P}$ and an algorithm $\mathbb{B}$ deciding Q_0 which is not almost optimal but has no hard sequences. Fix $z_0 \in Q$ and let $\mathbb{A}$ be the algorithm deciding Q that on input x first checks (in polynomial time) whether $x = \mathit{pad}(z_0, y)$ with $y \in Q_0$ (using the properties of the padding function and a polynomial time algorithm deciding Q_0); if so, it simulates $\mathbb{B}$ on y; otherwise it simulates $\mathbb{O}$ on x.

Clearly, $\mathbb{A}$ is not almost optimal as it can be speeded up on $\{\mathit{pad}(z_0, y) \mid y \in Q_0\}$, a hard set of $\mathbb{A}$. By contradiction, assume $(x_s)_{s\in\mathbb{N}}$ is a hard sequence for $\mathbb{A}$ and let $y_0 \in Q_0$. For $s \ge 1$ we set

$$y_s := \begin{cases} y, & \text{if } x_s = \mathit{pad}(z_0, y) \text{ with } y \in Q_0, \\ y_{s-1}, & \text{otherwise} \end{cases}$$

and

$$z_s := \begin{cases} z_{s-1}, & \text{if } x_s = \mathit{pad}(z_0, y) \text{ with } y \in Q_0, \\ x_s, & \text{otherwise.} \end{cases}$$

Then either $(y_s)_{s\in\mathbb{N}}$ is a hard sequence for $\mathbb{B}$ or $(z_s)_{s\in\mathbb{N}}$ is a hard sequence for $\mathbb{O}$, in both cases a contradiction. □

Corollary 5.8 *If the Measure Hypothesis holds, then the following are equivalent:*

(i) *Every* co-NP-*complete problem has no almost optimal algorithm.*

(ii) *Every non-almost optimal algorithm deciding a* co-NP-*complete problem has hard sequences.*

Proof. We already know that (i) implies (ii) by Theorem 2.3(a). Assume (ii) and by contradiction, suppose that Q is a co-NP-complete problem with an almost optimal algorithm. By Theorem 2.9, we may assume that Q has padding. Then, by the previous theorem, there is a non-almost optimal algorithm deciding Q without hard sequences, contradicting (ii). □

The following example shows that the padding hypothesis is necessary in Theorem 5.7.

Example 5.9 Let $Q := \{1^n \mid n \in \mathbb{N}\}$. As $Q \in \mathrm{P}$, it has an almost algorithm. However, the set Q itself is a hard set and $(1^s)_{s\in\mathbb{N}}$ a hard sequence for every non-optimal (that is, for every superpolynomial) algorithm deciding Q.

Finally, we show that also problems *without* almost optimal algorithm may have algorithms without hard sequences:

Theorem 5.10 *If the Measure Hypothesis holds, there is a problem which has hard sets for algorithms (and hence has no almost optimal algorithm) but has algorithms without hard sequences.*

Proof. Let $Q_0 \in \mathrm{P}$ be a problem with the properties stated in Proposition 5.5. We fix an effective enumeration

$$\mathbb{A}_0, \mathbb{A}_1, \dots \tag{5.5}$$

of all algorithms such that there is an universal algorithm $\mathbb{U}$ which on every input $\langle 1^i, x\rangle$ simulates the algorithm $\mathbb{A}_i$ on input $\langle 1^i, x\rangle$ in such a way that

$$t_{\mathbb{U}}\big(\langle 1^i, x\rangle\big) \leq (i+1)\cdot t_{\mathbb{A}_i}(\langle i, x\rangle)^2. \tag{5.6}$$

For every $i \in \mathbb{N}$ we let

$$S_i := \Big\{ \langle 1^i, x\rangle \;\Big|\; x \in Q_0 \text{ and } \mathbb{A}_i \text{ does } \textit{not} \text{ accept } \langle 1^i, x\rangle \text{ in } \leq 2^{(\log|x|)^2} \text{ steps}\Big\}. \tag{5.7}$$

Finally, we set

$$Q := \bigcup_{i\in\mathbb{N}} S_i$$

and show that Q is a problem with the properties mentioned in the theorem.

Claim 1 Let $k \in \mathbb{N}$. If $\mathbb{A}_k$ $\big($see (5.5)$\big)$ decides Q, then $S_k = \big\{ \langle 1^k, x\rangle \;\big|\; x \in Q_0\big\}$.

Proof of Claim 1. Otherwise, there exists an $x_0 \in Q_0$ with $\langle 1^k, x_0\rangle \notin S_k$. It follows that

$$\begin{aligned} x_0 \in Q_0 \quad & \text{with } \langle 1^k, x_0\rangle \notin S_k \\ & \Longrightarrow \mathbb{A}_k \text{ accepts } \langle 1^k, x_0\rangle \text{ in } \leq 2^{(\log|x|)^2} \text{ steps} && \text{(by (5.7))} \\ & \Longrightarrow \mathbb{A}_k \text{ accepts } \langle 1^k, x_0\rangle \\ & \Longrightarrow \langle 1^k, x_0\rangle \in Q && \text{(as } \mathbb{A}_k \text{ decides } Q) \\ & \Longrightarrow \langle 1^k, x_0\rangle \in S_k && \text{(since all } S_i\text{'s are disjoint).} \end{aligned}$$

This is a contradiction. ⊣

Claim 2 Q has hard sets for algorithms.

Proof of Claim 2. Assume that $\mathbb{A}_k$ decides Q. By Claim 1, $S_k = \{ \langle 1^k, x \rangle \mid x \in Q_0 \}$ and by (5.7), for every $x \in Q_0$,

$$t_{\mathbb{A}_k}\big(\big\langle 1^k, x \big\rangle \big) > 2^{(\log |x|)^2}.$$

As $Q_0 \in \mathrm{P}$, S_k is thus a hard set for $\mathbb{A}_k$. $\dashv$

Claim 3 For all sufficiently large $d \in \mathbb{N}$ there is an algorithm $\mathbb{Q}_d$ deciding Q such that

$$t_{\mathbb{Q}_d}\big(\langle 1^i, x \rangle \big) = (i+1) \cdot 2^{d \cdot (\log |x|)^2}$$

for every $i \in \mathbb{N}$ and $x \in \Sigma^*$.

Proof of Claim 3. By (5.6) and (5.7) as $Q_0 \in \mathrm{P}$. $\dashv$

Now we choose a sufficiently large $d \in \mathbb{N}$ and consider the algorithm $\mathbb{Q}_d$ of Claim 3. Assume that $\mathbb{Q}_d$ has a hard sequence

$$\big(\langle 1^{i_s}, x_s \rangle \big)_{s \in \mathbb{N}}.$$

By (5.7) every x_s is in Q_0 and by hardness,

$$t_{\mathbb{Q}_d}\big(\langle 1^{i_s}, x_s \rangle \big) = (i_s + 1) \cdot 2^{d \cdot (\log |x_s|)^2}$$

is superpolynomial in s. Since the mapping $1^s \mapsto \langle 1^{i_s}, x_s \rangle$ is computable in polynomial time, we have $|i_s| \leq |s|^{O(1)}$. Therefore,

$$(5.8) \qquad 2^{d \cdot (\log |x_s|)^2} \text{ is superpolynomial in } s.$$

As Q_0 is decidable in polynomial time and d is sufficiently large, we have an algorithm $\mathbb{A}$ deciding Q_0 in time $2^{d \cdot (\log |x|)^2}$ on every instance $x \in \Sigma^*$. Then (5.8) implies that $(x_s)_{s \in \mathbb{N}}$ is a hard sequence for $\mathbb{A}$, which contradicts Proposition 5.5(b). □

6 Getting hard sequences in an effective way

We have mentioned in the Introduction that McCreight and Meyer [**11**] have shown that for every EXP-hard problem Q there is a polynomial time procedure assigning to every algorithm deciding Q a hard sequence. Based on their proof we derive a "nondeterministic" version.

Theorem 6.1 *Let Q be a* coNEXP*-hard problem. Then there is a polynomial time computable function $g \colon \Sigma^* \times \{1\}^* \to \Sigma^*$ such that for every nondeterministic algorithm $\mathbb{A}$ accepting Q the sequence $\big(g(\mathbb{A}, 1^s)\big)_{s \in \mathbb{N}}$ is hard for $\mathbb{A}$.*

Proof. Consider the problem

Q_0
Instance: A nondeterministic algorithm $\mathbb{A}$.
Problem: Is it true that $\mathbb{A}$ does not accept $\mathbb{A}$ in at most $2^{|\mathbb{A}|}$ steps?

Claim 4 If $\mathbb{B}$ is a nondeterministic algorithm accepting Q_0, then $\mathbb{B} \in Q_0$ and therefore, $t_{\mathbb{B}}(\mathbb{B}) > 2^{|\mathbb{B}|}$.

Proof of Claim 4. Assume that $\mathbb{B} \notin Q_0$. Therefore, $\mathbb{B}$ does not accept $\mathbb{B}$. Then, by the definition of Q_0, we have $\mathbb{B} \in Q_0$, a contradiction. $\dashv$

To every nondeterministic algorithm $\mathbb{A}$ and every $s \in \mathbb{N}$ we can assign in time polynomial in $\mathbb{A}$ and s a nondeterministic algorithm $\mathbb{A}_s$ with

$$(6.1) \qquad |\mathbb{A}_s| \geq s, \quad L(\mathbb{A}_s) = L(\mathbb{A}), \quad \text{and} \quad t_{\mathbb{A}_s} = t_{\mathbb{A}}$$

(say, by adding s new "dummy" states).

Claim 5 If $\mathbb{A}$ is a nondeterministic algorithm accepting Q_0, then $(\mathbb{A}_s)_{s\in\mathbb{N}}$ is a hard sequence for $\mathbb{A}$.

Proof of Claim 5. It suffices to verify that, for all $s \in \mathbb{N}$,

$$(6.2) \qquad \mathbb{A}_s \in Q_0,$$

$$(6.3) \qquad t_{\mathbb{A}}(\mathbb{A}_s) > 2^s.$$

By (6.1) we know that $L(\mathbb{A}_s) = L(\mathbb{A})$. Hence, (6.2) holds by Claim 4, which also shows the first inequality in

$$t_{\mathbb{A}}(\mathbb{A}_s) = t_{\mathbb{A}_s}(\mathbb{A}_s) > 2^{|\mathbb{A}_s|} \geq 2^s,$$

the second one and the equality holding by (6.1). ⊣

Now let Q be coNEXP-hard. Since $Q_0 \in$ coNEXP there is a polynomial time reduction $\mathbb{S}$ from Q_0 to Q. Again, for a nondeterministic algorithm $\mathbb{A}$ let $\mathbb{A} \circ \mathbb{S}$ be the nondeterministic algorithm that on input $x \in \Sigma^*$ first runs $\mathbb{S}$ on x and then runs $\mathbb{A}$ on $\mathbb{S}(x)$.

For a nondeterministic algorithm $\mathbb{A}$ and $s \in \mathbb{N}$, we define

$$g(\mathbb{A}, 1^s) := \mathbb{S}((\mathbb{A} \circ \mathbb{S})_s).$$

Clearly, g is polynomial time computable. If $\mathbb{A}$ decides Q, then $\mathbb{A}\circ\mathbb{S}$ decides Q_0; therefore, $((\mathbb{A} \circ \mathbb{S})_s)_{s\in\mathbb{N}}$ is a hard sequence for $\mathbb{A} \circ \mathbb{S}$ by Claim 5. Hence, $\big(g(\mathbb{A}, 1^s)\big)_{s\in\mathbb{N}}$ is a hard sequence for $\mathbb{A}$ by Lemma 2.8. □

Acknowledgements

The authors wish to thank the John Templeton Foundation for its support under Grant #13152, *The Myriad Aspects of Infinity.* This research has also been partially supported by the National Nature Science Foundation of China (60970011) and the Sino-German Center for Research Promotion (GZ584). Yijia Chen is affiliated with BASICS and MOE-MS Key Laboratory for Intelligent Computing and Intelligent Systems which is supported by National Nature Science Foundation of China (61033002).

References

[1] O. Beyersdorff. On the correspondence between arithmetic theories and propositional proof systems —a survey. *Mathematical Logic Quarterly*, 55(2):116–137, 2009.

[2] Y. Chen and J. Flum. On p-optimal proof systems and logics for PTIME. In *Proceedings of the 37th International Colloquium on Automata, Languages and Programming (ICALP'10, Track B)*, volume 6199 of *Lecture Notes in Computer Science*, pp. 321–322, 2010.

[3] Y. Chen and J. Flum. Listings and logics. Electronic Colloquium on Computational Complexity (ECCC), TR11-020, 2011.

[4] S. Cook and R. Reckhow. The relative efficiency of propositional proof systems. *The Journal of Symbolic Logic*, 44:36–50, 1979.

[5] J. M. Hitchcock and A. Pavan. Hardness hypotheses, derandomization, and circuit complexity. In *Proceedings of the 24th Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS'04)*, 336–347, 2004.

[6] J. Krajíček. *Bounded arithmetic, propositional logic, and complexity theory.* Cambridge University Press, 1995.

[7] J. Krajíček and P. Pudlák. Propositional proof systems, the consistency of first order theories and the complexity of computations. *The Journal of Symbolic Logic*, 54:1063–1088, 1989.

[8] L. Levin. Universal search problems. *Problems of Information Transmission*, 9(3):265–266, 1973.

[9] E. Mayordomo. Almost every set in exponential time is P-bi-immune. *Theoretical Computer Science*, 136(2): 487–506, 1994.

[10] J. Messner. On the simulation order of proof systems. PhD Thesis, University of Erlangen, 2000.

[11] A. Meyer. A supervisor's reminiscence what we were thinking. Talk at the Stockmeyer-Symposium, 2005.

[12] Z. Sadowski. On an optimal propositional proof system and the structure of easy subsets of TAUT. *Theoretical Computer Science*, 288(1):181–193, 2002.

[13] L. Stockmeyer. The complexity of decision problems in automata theory. PhD Thesis, MIT, 1974.

On optimal probabilistic algorithms for Sat

Yijia Chen[†], **Jörg Flum**[‡], **Moritz Müller**[§]

[†] Department of Computer Science, Shanghai Jiao Tong University, China
yijia.chen@cs.sjtu.edu.cn

[‡] Mathematisches Institut, Albert-Ludwigs-Universität Freiburg, Germany
joerg.flum@math.uni-freiburg.de

[§] Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
moritz.mueller@univie.ac.at

Abstract. Assuming the existence of one-way functions we show that Sat does not have in certain sense optimal probabilistic algorithms.

Introduction

A major aim in the development of algorithms for hard problems is to decrease the running time. In particular one asks for algorithms that are optimal: A deterministic algorithm $\mathbb{A}$ deciding a language $L \subseteq \Sigma^*$ is *optimal* (or *(polynomially) optimal* or *p-optimal*) if for any other algorithm $\mathbb{B}$ deciding L there is a polynomial p such that

$$t_{\mathbb{A}}(x) \leq p(t_{\mathbb{B}}(x) + |x|) \tag{0.1}$$

for all $x \in \Sigma^*$. Here $t_{\mathbb{A}}(x)$ denotes the running time of $\mathbb{A}$ on input x. If (0.1) is only required for all $x \in L$, then $\mathbb{A}$ is said to be an *almost optimal algorithm for L* (or to be *optimal on positive instances of L*).

Various recent papers address the question whether such optimal algorithms exist for NP-complete or coNP-complete problems (cf. [**1**]), even though the problem has already been considered in the seventies when Levin [**4**] observed that there exists an optimal algorithm that finds a witness for every satisfiable propositional formula. Furthermore the relationship between the existence of almost optimal algorithms for a language L and the existence of "optimal" proof systems for L has been studied [**3, 5**].

Here we present a result (see Theorem 1.1) that can be interpreted as stating that (under the assumption of the existence of one-way functions) there is no optimal *probabilistic* algorithm for Sat.

1 Probabilistic speed-up

For a propositional formula α we denote by $\|\alpha\|$ the number of literals in it, counting repetitions. Hence, the actual length of any reasonable encoding of α is polynomially related to $\|\alpha\|$.

The main result of this short note reads as follows:

Presented at the conference *Logical Approaches to Barriers in Computing and Complexity*, Greifswald. Preprint of the Department of Mathematics and Computer Science at the University Greifswald No. 6, 2010. The authors thank the John Templeton Foundation for its support under Grant #13152, *The Myriad Aspects of Infinity.*

Theorem 1.1 *Assume one-way functions exist. Then for every probabilistic algorithm $\mathbb{A}$ deciding* Sat *there exists a probabilistic algorithm $\mathbb{B}$ deciding* Sat *such that, for all $d \in \mathbb{N}$ and sufficiently large $n \in \mathbb{N}$,*

$$\Pr\Big[\begin{array}{l}\textit{there is a satisfiable } \alpha \textit{ with } \|\alpha\| = n \textit{ such that}\\ \mathbb{A} \textit{ does not accept } \alpha \textit{ in at most } (t_{\mathbb{B}}(\alpha) + \|\alpha\|)^d \textit{ steps}\end{array}\Big] \geq \frac{1}{5}.$$

Note that $t_{\mathbb{A}}(\alpha)$ and $t_{\mathbb{B}}(\alpha)$ are random variables, and the probability is taken over the coin tosses of $\mathbb{A}$ and $\mathbb{B}$ on α.

Here we say that a probabilistic algorithm $\mathbb{A}$ decides Sat if it decides Sat as a nondeterministic algorithm, that is

$$\begin{aligned}\alpha \in \text{Sat} &\implies \Pr[\mathbb{A} \text{ accepts } \alpha] > 0,\\ \alpha \notin \text{Sat} &\implies \Pr[\mathbb{A} \text{ accepts } \alpha] = 0.\end{aligned}$$

In particular, $\mathbb{A}$ can only err on 'yes'-instances.

Note that in the first condition the error probability is not demanded to be bounded away from 0, say by a constant $\epsilon > 0$. As a more usual notion of probabilistic decision, say $\mathbb{A}$ *decides* Sat *with one-sided error ϵ* if

$$\begin{aligned}\alpha \in \text{Sat} &\implies \Pr[\mathbb{A} \text{ accepts } \alpha] > 1 - \epsilon,\\ \alpha \notin \text{Sat} &\implies \Pr[\mathbb{A} \text{ accepts } \alpha] = 0.\end{aligned}$$

For this concept we get

Corollary 1.2 *Assume one-way functions exist and let $\epsilon > 0$. Then for every probabilistic algorithm $\mathbb{A}$ deciding* Sat *with one-sided error ϵ there exists a probabilistic algorithm $\mathbb{B}$ deciding* Sat *with one-sided error ϵ such that, for all $d \in \mathbb{N}$ and sufficiently large $n \in \mathbb{N}$,*

$$\Pr\Big[\begin{array}{l}\textit{there is a satisfiable } \alpha \textit{ with } \|\alpha\| = n \textit{ such that}\\ \mathbb{A} \textit{ does not accept } \alpha \textit{ in at most } (t_{\mathbb{B}}(\alpha) + \|\alpha\|)^d \textit{ steps}\end{array}\Big] \geq \frac{1}{5}.$$

This follows from the fact that in the proof of Theorem 1.1 we choose the algorithm $\mathbb{B}$ in such way that on any input α the error probability of $\mathbb{B}$ on α is not worse than the error probability of $\mathbb{A}$ on α.

2 Witnessing failure

The proof of Theorem 1.1 is based on the following result.

Theorem 2.1 *Assume that one-way functions exist. Then there is a probabilistic polynomial time algorithm $\mathbb{C}$ satisfying the following conditions.*

(1) *On input $n \in \mathbb{N}$ in unary the algorithm $\mathbb{C}$ outputs with probability one a satisfiable formula β with $\|\beta\| = n$.*

(2) *For every $d \in \mathbb{N}$ and every probabilistic algorithm $\mathbb{A}$ deciding* Sat *and sufficiently large $n \in \mathbb{N}$,*

$$\Pr\big[\mathbb{A} \textit{ does not accept } \mathbb{C}(n) \textit{ in } n^d \textit{ steps}\big] \geq \frac{1}{3}.$$

In the terminology of fixed-parameter tractability this theorem tells us that the parameterized construction problem associated with the following parameterized decision problem p-CounterExample-Sat is in a suitably defined class of randomized nonuniform fixed-parameter tractable problems.

Instance: An algorithm $\mathbb{A}$ *deciding* Sat and $d, n \in \mathbb{N}$ in unary.
Parameter: $\|\mathbb{A}\| + d$.
Problem: Does there exist a satisfiable CNF-formula α with $\|\alpha\| = n$ such that $\mathbb{A}$ does not accept α in n^d many steps?

Note that this problem is a promise problem. We can show:

Theorem 2.2 *Assume that one-way functions exist. Then the problem*

p-CounterExample-Sat

is nonuniformly fixed-parameter tractable.[1]

This result is an immediate consequence of the following

Theorem 2.3 *Assume that one-way functions exist. For every infinite set $I \subseteq \mathbb{N}$ the problem*

Sat$_I$
Instance: A CNF-formula α with $\|\alpha\| \in I$.
Problem: Is α satisfiable?

is not *in* PTIME.

The decision problem p-CounterExample-Sat has the following associated construction problem:

Instance: An algorithm $\mathbb{A}$ *deciding* Sat and $d, n \in \mathbb{N}$ in unary.
Parameter: $\|\mathbb{A}\| + d$.
Problem: Construct a satisfiable CNF-formula α with $\|\alpha\| = n$ such that $\mathbb{A}$ does not accept α in n^d many steps, if one exists.

We do not know anything on its (deterministic) complexity; its nonuniform fixed-parameter tractability would rule out the existence of strongly almost optimal algorithms for Sat. By definition, an algorithm $\mathbb{A}$ deciding Sat is a *strongly almost optimal algorithm for* Sat if there is a polynomial p such that, for any other algorithm $\mathbb{B}$ deciding Sat,

$$t_{\mathbb{A}}(\alpha) \leq p(t_{\mathbb{B}}(\alpha) + |\alpha|)$$

for all $\alpha \in$ Sat. Then the precise statement of the result just mentioned reads as follows:

Proposition 2.4 *Assume that* P $\neq$ NP. *If the construction problem associated with* p-CounterExample-Sat *is nonuniformly fixed-parameter tractable, then there is no strongly almost optimal algorithms for* Sat.

[1] This means that there is a $c \in \mathbb{N}$ such that for every algorithm $\mathbb{A}$ deciding Sat and every $d \in \mathbb{N}$ there is an algorithm that decides for every $n \in \mathbb{N}$ whether $(\mathbb{A}, d, n)$ is a positive instance of p-CounterExample-Sat in time $O(n^c)$; here the constant hidden in $O(\ )$ may depend on $\mathbb{A}$ and d.

3 Some proofs

We now show how to use an algorithm $\mathbb{C}$ as in Theorem 2.1 to prove Theorem 1.1.

Proof of Theorem 1.1 *from Theorem* 2.1: Let $\mathbb{A}$ be an algorithm deciding Sat. We choose $a \in \mathbb{N}$ such that for every $n \geq 2$ the running time of the algorithm $\mathbb{C}$ (provided by Theorem 2.1) on input n is bounded by n^a. We define the algorithm $\mathbb{B}$ as follows:

```
$\mathbb{B}(\alpha)$        // $\alpha \in$ CNF
    1. $\beta \leftarrow \mathbb{C}(\|\alpha\|)$
    2. if $\alpha = \beta$ then accept and halt
    3. else Simulate $\mathbb{A}$ on $\alpha$.
```

Let $d \in \mathbb{N}$ be arbitrary. Set $e := d \cdot (a+2)+1$ and fix a sufficiently large $n \in \mathbb{N}$. Let S_n denote the range of $\mathbb{C}(n)$. Furthermore, let $T_{n,\beta,e}$ denote the set of all strings $r \in \{0,1\}^{n^e}$ that do not determine a (complete) accepting run of $\mathbb{A}$ on β that consists in at most n^e many steps. Observe that a (random) run of $\mathbb{A}$ does not accept β in at most n^e steps if and only if $\mathbb{A}$ on β *uses* $T_{n,\beta,e}$, that is, its first at most n^e many coin tosses on input β are described by some $r \in T_{n,\beta,e}$. Hence by (2) of Theorem 2.1 we conclude

$$\sum_{\beta \in S_n} \big(\Pr[\beta = \mathbb{C}(n)] \cdot \Pr_{r \in \{0,1\}^{n^e}} [r \in T_{n,\beta,e}] \big) \geq \frac{1}{3}. \tag{3.1}$$

Let $\alpha \in S_n$ and apply $\mathbb{B}$ to α. If the execution of $\beta \leftarrow \mathbb{C}(\|\alpha\|)$ in Line 1 yields $\beta = \alpha$, then the overall running time of the algorithm $\mathbb{B}$ is bounded by $O\big(n^2 + t_{\mathbb{C}}(n)\big) = O(n^{a+1}) \leq n^{a+2}$ for n is sufficiently large. If in such a case a run of the algorithm $\mathbb{A}$ on input α uses an $r \in T_{n,\alpha,e}$, then it does not accept α in time $n^e = n^{(a+2)\cdot d+1}$ and hence not in time $(t_{\mathbb{B}}(\alpha) + \|\alpha\|)^d$. Therefore,

$$\begin{aligned}
&\Pr\Big[\text{there is a satisfiable } \alpha \text{ with } \|\alpha\| = n \text{ such that} \\
&\qquad \mathbb{A} \text{ does not accept } \alpha \text{ in at most } (t_{\mathbb{B}}(\alpha) + \|\alpha\|)^d \text{ steps}\Big] \\
&\geq 1 - \Pr\big[\text{for every input } \alpha \in S_n \text{ the algorithm } \mathbb{B} \text{ does not generate } \alpha \\
&\qquad\qquad \text{in Line 3, or } \mathbb{A} \text{ does not use } T_{n,\alpha,e}\big] \\
&= 1 - \prod_{\alpha \in S_n} \big((1 - \Pr[\alpha = \mathbb{C}(n)]) + \Pr[\alpha = \mathbb{C}(n)] \cdot \Pr_{r \in \{0,1\}^{n^e}} [r \notin T_{n,\alpha,e}] \big) \\
&= 1 - \prod_{\alpha \in S_n} \big(1 - \Pr[\alpha = \mathbb{C}(n)] \cdot \Pr_{r \in \{0,1\}^{n^e}} [r \in T_{n,\alpha,e}] \big) \\
&\geq 1 - \left(\frac{\sum_{\alpha \in S_n} \big(1 - \Pr[\alpha = \mathbb{C}(n)] \cdot \Pr_{r \in \{0,1\}^{n^e}}[r \in T_{n,\alpha,e}]\big)}{|S_n|} \right)^{|S_n|} \\
&= 1 - \left(1 - \frac{\sum_{\alpha \in S_n} \Pr[\alpha = \mathbb{C}(n)] \cdot \Pr_{r \in \{0,1\}^{n^e}}[r \in T_{n,\alpha,e}]}{|S_n|} \right)^{|S_n|} \\
&\geq 1 - \left(1 - \frac{1}{3 \cdot |S_n|} \right)^{|S_n|} \geq \frac{1}{5}.
\end{aligned}$$

□

Theorem 2.1 immediately follows from the next lemma.

Lemma 3.1 *Assume that one-way functions exist. Then there is a randomized polynomial time algorithm $\mathbb{H}$ satisfying the following conditions:*

(H1) *Given $n \in \mathbb{N}$ in unary the algorithm $\mathbb{H}$ computes with probability one a* satisfiable *CNF α of size $\|\alpha\| = n$.*

(H2) *For every* probabilistic *algorithm $\mathbb{A}$ deciding* Sat *and every $d, p \in \mathbb{N}$ there exists an $n_{\mathbb{A},d,p} \in \mathbb{N}$ such that, for all $n \geq n_{\mathbb{A},d,p}$,*

$$\Pr\left[\mathbb{A} \textit{ accepts } \mathbb{H}(n) \textit{ in time } n^d\right] \leq \frac{1}{2} + \frac{1}{n^p},$$

where the probability is taken uniformly over all possible outcomes of the internal coin tosses of the algorithms $\mathbb{A}$ and $\mathbb{H}$.

(H3) *The cardinality of the range of (the random variable) $\mathbb{H}(n)$ is superpolynomial in n.*

Sketch of proof: We present the construction of the algorithm $\mathbb{H}$. By the assumption that one-way functions exist, we know that there is a pseudorandom generator (e.g., see **[2]**), that is, there is an algorithm $\mathbb{G}$ such that:

(G1) For every $s \in \{0,1\}^*$ the algorithm $\mathbb{G}$ computes a string $\mathbb{G}(s)$ with $|\mathbb{G}(s)| = |s|+1$ in time polynomial in $|s|$.

(G2) For every probabilistic polynomial time algorithm $\mathbb{D}$, every $p \in \mathbb{N}$, and all sufficiently large $\ell \in \mathbb{N}$ we have

$$\left| \Pr_{s \in \{0,1\}^\ell}\left[\mathbb{D}(\mathbb{G}(s)) = 1\right] - \Pr_{r \in \{0,1\}^{\ell+1}}\left[\mathbb{D}(r) = 1\right] \right| \leq \frac{1}{\ell^p}.$$

(In the above terms, the probability is also taken over the internal coin toss of $\mathbb{D}$.)

Let the language Q be the range of $\mathbb{G}$,

$$Q := \left\{\mathbb{G}(s) \;\middle|\; s \in \{0,1\}^*\right\}.$$

Q is in NP by (G1). Hence, there is a polynomial time reduction $\mathbb{S}$ from Q to Sat, which we can assume to be injective. We choose a constant $c \in \mathbb{N}$ such that $\|\mathbb{S}(r)\| \leq |r|^c$ for every $r \in \{0,1\}^*$. For every propositional formula β and every $n \in \mathbb{N}$ with $n \geq \|\beta\|$ let $\beta(n)$ be an equivalent propositional formula with $\|\beta(n)\| = n$. We may assume that $\beta(n)$ is computed in time polynomial in n.

One can check that the following algorithm $\mathbb{H}$ has the properties claimed in the lemma.

```
$\mathbb{H}(n)$ // $n \in \mathbb{N}$
    $m \leftarrow \lfloor \sqrt[c]{n-1} \rfloor - 1$
    Choose an $s \in \{0,1\}^m$ uniformly at random
    $\beta \leftarrow \mathbb{S}(\mathbb{G}(s))$
    Output $\beta(n)$.
```

□

References

[1] O. Beyersdorff and Z. Sadowski. Characterizing the existence of optimal proof systems and complete sets for promise classes. Electronic Colloquium on Computational Complexity, Report TR09-081, 2009.

[2] O. Goldreich. *Foundations of Cryptography, vol.* 1 *(Basic Tools).* Cambridge University Press, 2001.

[3] J. Krajíček and P. Pudlák. Propositional proof systems, the consistency of first order theories and the complexity of computations. *The Journal of Symbolic Logic*, 54(3):1063–1079, 1989.

[4] L. Levin. Universal search problems (in Russian). *Problemy Peredachi Informatsii*, 9(3):115–116, 1973.

[5] J. Messner. On optimal algorithms and optimal proof systems. STACS'99, Lecture Notes in Computer Science 1563, 541–550, 1999.

Consistency, optimality, and incompleteness

Yijia Chen[†], Jörg Flum[‡], Moritz Müller[§]

[†] Department of Computer Science, Shanghai Jiao Tong University, China
yijia.chen@cs.sjtu.edu.cn

[‡] Mathematisches Institut, Albert-Ludwigs-Universität Freiburg, Germany
joerg.flum@math.uni-freiburg.de

[§] Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
moritz.mueller@univie.ac.at

Abstract. Assume that the problem P_0 is not solvable in polynomial time. Let T be a first-order theory containing a sufficiently rich part of true arithmetic. We characterize $T \cup \{Con_T\}$ as the minimal extension of T proving for some algorithm that it decides P_0 as fast as any algorithm $\mathbb{B}$ with the property that T proves that $\mathbb{B}$ decides P_0. Here, Con_T claims the consistency of T. As a byproduct, we obtain a version of Gödel's Second Incompleteness Theorem. Moreover, we characterize problems with an optimal algorithm in terms of arithmetical theories.

Introduction

By Gödel's Second Incompleteness Theorem, a consistent, computably enumerable and sufficiently strong first-order theory T cannot prove its own consistency Con_T. In other words, $T \cup \{Con_T\}$ is a proper extension of T.

In Bounded Arithmetic one studies the complexity of proofs in terms of the computational complexity of the concepts involved in the proofs (see e.g. [**1**, Introduction]). Stronger theories allow reasoning with more complicated concepts. For example, a computational problem may be solvable by an algorithm whose proof of correctness needs tools not available in the given theory; moreover, stronger theories may know of faster algorithms solving the problem. When discussing these issues with the authors, Sy-David Friedman asked whether $T \cup \{Con_T\}$ can be characterized in this context as a minimal extension of T. We could prove the following (all terms will be defined in the paper).

Theorem 1 *Let P_0 be a decidable problem which is not decidable in polynomial time. Then there is a finite true arithmetical theory T_0 and a computable function F assigning to every computably enumerable theory T with $T \supseteq T_0$ an algorithm $F(T)$ such that* (a) *and* (b) *hold.*

(a) *T_0 proves that $F(T)$ is as fast as any algorithm T-provably deciding P_0.*

(b) *For every theory T^* with $T^* \supseteq T$ the following are equivalent:*

(i) *T^* proves Con_T.*

(ii) *The algorithm $F(T)$ T^*-provably decides P_0.*

(iii) *There is an algorithm such that T^* proves that it decides P_0 and that it is as fast as any algorithm T-provably deciding P_0.*

A preliminary version of this article appeared in *Proceedings of the 7th Computability in Europe, Mathematical Theory and Computational Practice* (CiE'11), Lecture Notes in Computer Science 6735, pp. 61–70, 2011.

Hence, by merely knowing the extension T of T_0 we are able to compute the algorithm $F(T)$, which is, provably in T_0, as fast as any algorithm T-provably deciding P_0; however, in order to prove that $F(T)$ decides P_0 we need the full strength of $T \cup \{Con_T\}$. In this sense, $T \cup \{Con_T\}$ is a minimal extension of T.

It is known [**8**] that there are problems P_0 such that one can effectively assign to every algorithm $\mathbb{A}$ deciding P_0 a further algorithm $\mathbb{B}$ deciding P_0 such that $\mathbb{A}$ is not as fast as $\mathbb{B}$. Based on this fact, from our considerations yielding a proof of Theorem 1 we obtain a version of Gödel's Second Incompleteness Theorem.

The content of the different sections is the following. In Section 2, by a standard diagonalization technique we derive a result showing for every computably enumerable set D of algorithms the existence of an algorithm that on every input behaves as some algorithm in D and that is as fast as every algorithm in D (see Lemma 2.1). In Theorem 3.4 of Section 3 we characterize problems with an optimal algorithm in terms of arithmetical theories. Section 4 contains a proof of Theorem 1. Finally, we derive the Second Incompleteness Theorem in Section 5.

Many papers in computational complexity, older and recent ones, address the question whether hard problems have *optimal* or *almost optimal* algorithms. Although Levin [**5**] observed that there exists an optimal algorithm that finds a satisfying assignment for every satisfiable propositional formula, it is not known whether the class of satisfiable propositional formulas or the class of tautologies have an almost optimal algorithm.

Krajíček and Pudlák [**4**] showed for the latter class that an almost optimal algorithm exists if and only if "there exists a finitely axiomatized fragment T of the true arithmetic such that, for every finitely axiomatized consistent theory S, there exists a deterministic Turing machine $\mathbb{M}$ and a polynomial p such that for any given n, in time $\leq p(n)$ the machine $\mathbb{M}$ constructs a proof in T of $\mathrm{Con}_S(\underline{n})$". Here $\mathrm{Con}_S(\underline{n})$ claims that no contradiction can be derived from S by proofs of lengths at most n.

Hartmanis [**2**] and Hutter [**3**] considered 'provable' algorithms, where 'provable' refers to a computably enumerable, more or less specified true theory T. Hartmanis compares the class of problems decidable within a given time bound with the class of problems T-provably decidable within this time bound and he studies time hierarchy theorems in this context. Hutter constructs an algorithm "which is the fastest and the shortest" deciding a given problem. As Hutter says, Peter van Emde Boas pointed out to him that it is not provable that his algorithm decides the given problem and that his proof is a "meta-proof which cannot be formalized within the considered proof system" and he adds that "a formal proof of its correctness would prove the consistency of the proof system, which is impossible by Gödel's Second Incompleteness Theorem".

Unlike these papers we do not assume in Theorem 1 that T is a true theory.

1 Some preliminaries

First we fix some notations and introduce some basic concepts. We consider problems as subsets of Σ^*, the set of strings over the alphabet $\Sigma := \{0, 1\}$. For an algorithm $\mathbb{A}$ and a string $x \in \Sigma^*$ we let $t_{\mathbb{A}}(x)$ denote the running time of $\mathbb{A}$ on x. In case $\mathbb{A}$ does not halt on x, we set $t_{\mathbb{A}}(x) := \infty$. If $t_{\mathbb{A}}(x)$ is finite, we denote by $\mathbb{A}(x)$ the output of $\mathbb{A}$ on x.

If $\mathbb{A}$ and $\mathbb{B}$ are algorithms, then $\mathbb{A}$ *is as fast as* $\mathbb{B}$ if there is a polynomial p such that

$$t_{\mathbb{A}}(x) \leq p\big(t_{\mathbb{B}}(x) + |x|\big) \tag{1.1}$$

for every $x \in \Sigma^*$. Note that here we do not require that $\mathbb{A}$ and $\mathbb{B}$ decide the same $P \subseteq \Sigma^*$.

An algorithm deciding a problem P is *optimal* if it is as fast as every other algorithm deciding P, that is, if it has no superpolynomial speedup on an infinite subset of Σ^*. An algorithm $\mathbb{A}$ deciding P is *almost optimal* if (1.1) holds for every other algorithm $\mathbb{B}$ deciding P and every $x \in P$ (hence nothing is required of the relationship between $t_{\mathbb{A}}(x)$ and $t_{\mathbb{B}}(x)$ for $x \notin P$).

We do not distinguish algorithms from their codes by strings and we do not distinguish strings from their codes by natural numbers. However, we do not fix a computation model (Turing machines, random access machines...) for algorithms. We state the results in such a way that they hold for every standard computation model.

2 Diagonalizing over algorithms

In computability theory diagonalization techniques are used in various contexts. We will make use of the following result.

Lemma 2.1 (Diagonalization Lemma) *Let D be a computably enumerable and nonempty set of algorithms. Then there is an algorithm $\mathbb{A}$ such that* (a) *and* (b) *hold.*

(a) *The algorithm $\mathbb{A}$ halts precisely on those inputs on which at least one algorithm in D halts, and in that case it outputs the same as some algorithm in D; more formally, for all $x \in \Sigma^*$,*

- $t_{\mathbb{A}}(x) < \infty \iff t_{\mathbb{D}}(x) < \infty$ *for some* $\mathbb{D} \in D$;
- *if* $t_{\mathbb{A}}(x) < \infty$, *then there is* $\mathbb{D} \in D$ *with* $\mathbb{A}(x) = \mathbb{D}(x)$.

(b) *There is a $d \in \mathbb{N}$ such that* [1] *for all $\mathbb{D} \in D$ there is a $c_{\mathbb{D}} \in \mathbb{N}$ such that, for all $x \in \Sigma^*$,*

$$t_{\mathbb{A}}(x) \leq c_{\mathbb{D}} \cdot \big(t_{\mathbb{D}}(x) + |x|\big)^d.$$

Moreover, there is a computable function that maps any algorithm $\mathbb{E}$ enumerating the set D of algorithms to an algorithm $\mathbb{A}$ satisfying (a) *and* (b).

In particular, if all algorithms in D decide $P \subseteq \Sigma^*$, then $\mathbb{A}$ is an algorithm deciding P as fast as every $\mathbb{D} \in D$.

Proof. Let the algorithm $\mathbb{E}$ enumerate the set D of algorithms, that is, $\mathbb{E}$, once having been started, eventually prints out exactly the algorithms in D. For each $i \in \mathbb{N}$ we denote by $\mathbb{E}_i$ the last algorithm printed out by $\mathbb{E}$ in i steps; in particular, $\mathbb{E}_i$ is undefined if $\mathbb{E}$ hasn't printed any algorithm in i steps. Algorithm $\mathbb{A}$ is defined as follows.

```
$\mathbb{A}(x)$    // $x \in \Sigma^*$
    1. $\ell \leftarrow 0$
    2. for $i = 0$ to $\ell$
    3.         if $\mathbb{E}_i$ is defined then simulate the $(\ell - i)$th step
    4.           of $\mathbb{E}_i$ on $x$
    5.         if the simulation halts then halt and output
    6.           accordingly
    7. $\ell \leftarrow \ell + 1$
    8. goto 2.
```

[1] As the proof shows, the constant $d \in \mathbb{N}$ does not even depend on D but it depends on the concrete machine model one uses.

Of course (the code of) $\mathbb{A}$ can be computed from (the code of) $\mathbb{E}$. It is easy to see that $\mathbb{A}$ satisfies (a). Furthermore, there are constants $c_0, d_0 \in \mathbb{N}$ such that for all $x \in \Sigma^*$ and every $\ell \in \mathbb{N}$, lines 2–6 take time at most

$$c_0 \cdot (\ell + |x|)^{d_0}. \tag{2.1}$$

To verify (b), let $\mathbb{D} \in D$ and $i_{\mathbb{D}}$ be the minimum $i \in \mathbb{N}$ with $\mathbb{E}_i = \mathbb{D}$. Fix an input $x \in \Sigma^*$. For

$$\ell = i_{\mathbb{D}} + t_{\mathbb{E}_{i_{\mathbb{D}}}}(x) \quad \text{and} \quad i = i_{\mathbb{D}}$$

the simulation in line 3 halts if it did not halt before. Therefore

$$\begin{aligned} t_{\mathbb{A}}(x) &\le O\left(\sum_{\ell=0}^{i_{\mathbb{D}}+t_{\mathbb{D}}(x)} (\ell + |x|)^{d_0}\right) \qquad \big(\text{by (2.1)}\big) \\ &\le O\Big(\big(i_{\mathbb{D}} + t_{\mathbb{D}}(x) + |x|\big)^{d_0+1}\Big) \le c_{\mathbb{D}} \cdot \big(t_{\mathbb{D}}(x) + |x|\big)^{d_0+1} \end{aligned}$$

for an appropriate constant $c_{\mathbb{D}} \in \mathbb{N}$ only depending on $\mathbb{D}$. □

The preceding proof uses the idea underlying standard proofs of a result due to Levin [**5**]. Even more, Levin's result is also a consequence of Lemma 2.1.

Example 2.2 (Levin [**5**]) Let $F\colon \Sigma^* \to \Sigma^*$ be computable. An *inverter of* F is an algorithm $\mathbb{I}$ that given y in the image of F halts with some output $\mathbb{I}(y)$ such that $F(\mathbb{I}(y)) = y$. On inputs not in the image of F, the algorithm $\mathbb{I}$ may do whatever it wants.

Let $\mathbb{F}$ be an algorithm computing F. For an arbitrary algorithm $\mathbb{B}$ define $\mathbb{B}^*$ as follows. On input y the algorithm $\mathbb{B}^*$ simulates $\mathbb{B}$ on y; if the simulation halts, then by simulating $\mathbb{F}$ it computes $F(\mathbb{B}(y))$; if $F(\mathbb{B}(y)) = y$, then it outputs $\mathbb{B}(y)$, otherwise it does not stop. Thus if $\mathbb{B}^*$ halts on $y \in \Sigma^*$, then it outputs a preimage of y and

$$t_{\mathbb{B}^*}(y) \le O\big(t_{\mathbb{B}}(y) + t_{\mathbb{F}}(\mathbb{B}(y)) + |y|\big). \tag{2.2}$$

Furthermore, if $\mathbb{B}$ is an inverter of F, then so is $\mathbb{B}^*$.

Let $D := \big\{\mathbb{B}^* \mid \mathbb{B} \text{ is an algorithm}\big\}$. Denote by $\mathbb{I}_{\text{opt}}$ an algorithm having for this D the properties of the algorithm $\mathbb{A}$ in Lemma 2.1. By the previous remarks it is easy to see that $\mathbb{I}_{\text{opt}}$ is an inverter of F. Moreover, by Lemma 2.1(b) and (2.2), we see that for any other inverter $\mathbb{B}$ of F there exists a constant $c_{\mathbb{B}}$ such that for all y in the image of F

$$t_{\mathbb{I}_{\text{opt}}}(y) \le c_{\mathbb{B}} \cdot \big(t_{\mathbb{B}}(y) + t_{\mathbb{F}}(\mathbb{B}(y)) + |y|\big)^d.$$

In this sense $\mathbb{I}_{\text{opt}}$ is an optimal inverter of F.

3 Algorithms and arithmetical theories

To talk about algorithms and strings we use *arithmetical formulas*, that is, first-order formulas in the language $L_{\text{PA}} := \{+, \cdot, 0, 1, <\}$ of Peano Arithmetic. Arithmetical sentences are *true* (*false*) if they hold (do not hold) in the standard L_{PA}-model. For a natural number n let $\dot{n}$ denote the natural L_{PA}-term without variables denoting n (in the standard model).

Recall that an arithmetical formula is Δ_0 if all quantifiers are bounded and it is Σ_1 if it has the form $\exists x_1 \dots \exists x_m \psi$ where ψ is Δ_0.

We shall use a Δ_0-formula

$$Run(u, x, y, z)$$

that defines (in the standard model) the set of tuples (u, x, y, z) such that u is an algorithm that on input x outputs y by the (code of a complete finite) run z; recall that we do not distinguish algorithms from their codes by strings and strings from their codes by natural numbers.

> *For the rest of this paper we fix a decidable $P_0 \subseteq \Sigma^*$ and an algorithm $\mathbb{A}_0$ deciding P_0.*

The formula

$$\mathit{Dec}_{P_0}(u) := \forall x \exists y \exists z \mathit{Run}(u, x, y, z) \wedge \\ \forall x \forall y \forall y' \forall z \forall z' \big((\mathit{Run}(\dot{\mathbb{A}}_0, x, y, z) \wedge \mathit{Run}(u, x, y', z')) \to y = y' \big)$$

defines the set of algorithms deciding P_0.

Let L_{all} with $L_{\text{PA}} \subset L_{\text{all}}$ be a language containing countably many function and relation symbols of every arity ≥ 1 and countably many constants. A *theory* is a set T of first-order L_{all}-sentences. We write $T \vdash \varphi$ if the theory T proves the sentence φ.

Definition 3.1 Let T be a theory.

(a) An algorithm $\mathbb{A}$ *T-provably decides* P_0 if $T \vdash \mathit{Dec}_{P_0}(\dot{\mathbb{A}})$.

(b) T is *sound for P_0-decision* means that for every algorithm $\mathbb{A}$

if $T \vdash \mathit{Dec}_{P_0}(\dot{\mathbb{A}})$, then $\mathbb{A}$ decides P_0.

(c) T is *complete for P_0-decision* means that for every algorithm $\mathbb{A}$

if $\mathbb{A}$ decides P_0, then $T \vdash \mathit{Dec}_{P_0}(\dot{\mathbb{A}})$.

For a computably enumerable sound theory T that proves $\mathit{Dec}_{P_0}(\dot{\mathbb{A}}_0)$ the set

$$D(T) := \big\{ \mathbb{D} \mid T \vdash \mathit{Dec}_{P_0}(\dot{\mathbb{D}}) \big\} \tag{3.1}$$

is a computably enumerable and nonempty set of algorithms deciding P_0. Thus, by Lemma 2.1 for $D = D(T)$ we get an algorithm $\mathbb{A}$ deciding P_0 as fast as every algorithm in $D(T)$. If in addition T is complete for P_0-decision, then $D(T)$ would be the set of all algorithms deciding P_0 and thus $\mathbb{A}$ would be an optimal algorithm for P_0. So, the problem P_0 would have an optimal algorithm if we can find a computably enumerable theory that is both sound and complete for P_0-decision. Unfortunately, there is no such theory as shown by the following proposition. We relax these properties in Definition 3.3 and show in Theorem 3.4 that the new ones are appropriate to characterize problems with optimal algorithms.

Proposition 3.2 *There is no computably enumerable theory that is sound and complete for P_0-decision.*

Proof. We assume that there is a computably enumerable theory T that is sound and complete for P_0-decision and derive a contradiction by showing that then the halting problem for Turing machines would be decidable.

For every Turing machine $\mathbb{M}$ we consider two algorithms. On every input $x \in \Sigma^*$ the first algorithm $\mathbb{B}_1(\mathbb{M})$ first checks whether x codes a run of $\mathbb{M}$ accepting the empty input tape and then it simulates $\mathbb{A}_0$ on x (recall $\mathbb{A}_0$ is the fixed algorithm deciding P_0). If x codes an accepting run, then $\mathbb{B}_1(\mathbb{M})$ reverses the answer $\mathbb{A}_0(x)$ of $\mathbb{A}_0$ on x, otherwise it outputs exactly $\mathbb{A}_0(x)$. Clearly $\mathbb{B}_1(\mathbb{M})$ decides P_0 if and only if $\mathbb{M}$ does not halt on the empty input tape.

The second algorithm $\mathbb{B}_2(\mathbb{M})$, on every input $x \in \Sigma^*$ first checks exhaustively whether $\mathbb{M}$ halts on the empty input tape; if eventually it finds an accepting run, then it simulates $\mathbb{A}_0$ on x and outputs accordingly. It is easy to verify that $\mathbb{B}_2(\mathbb{M})$ decides P_0 if and only if $\mathbb{M}$ halts on the empty input tape.

As T is sound for P_0-decision, it proves at most one of $\mathit{Dec}_{P_0}(\dot{\overline{\mathbb{B}_1(\mathbb{M})}})$ and $\mathit{Dec}_{P_0}(\dot{\overline{\mathbb{B}_2(\mathbb{M})}})$, and as it is complete for P_0-decision it proves at least one of these sentences. Hence, given $\mathbb{M}$, by enumerating the T-provable sentences we can decide whether $\mathbb{M}$ halts on the empty input tape. □

Definition 3.3 A theory T is *almost complete for P_0-decision* if for every algorithm $\mathbb{A}$ deciding P_0 there is an algorithm T-provably deciding P_0 that is as fast as $\mathbb{A}$.

Theorem 3.4 *The following are equivalent for decidable $P_0 \subseteq \Sigma^*$:*

(i) *P_0 has an optimal algorithm;*

(ii) *There is a computably enumerable and arithmetical theory T that is sound and almost complete for P_0-decision.*

Proof. (i) ⇒ (ii): We set $T := \{\mathit{Dec}_{P_0}(\dot{\mathbb{A}})\}$ where $\mathbb{A}$ is an optimal algorithm for P_0. Then T is a computably enumerable true arithmetical theory. Truth implies soundness and almost completeness follows from the optimality of $\mathbb{A}$.

(ii) ⇒ (i): Let T be as in (ii). Then the set $D(T)$ defined by (3.1) is nonempty by almost completeness of T and, by soundness, it is a computably enumerable set of algorithms deciding P_0. By Lemma 2.1 for $D = D(T)$ we get an algorithm $\mathbb{A}$ deciding P_0 as fast as every algorithm in $D(T)$ and hence by almost completeness as fast as any algorithm deciding P_0. Thus, $\mathbb{A}$ is an optimal algorithm for P_0. □

A result related to the implication (ii) ⇒ (i) is shown by Sadowski in [**7**]. He shows assuming that there does not exist an almost optimal algorithm for the set TAUT of all propositional tautologies, that for every theory T there exists a subset of TAUT in P which is not T-provably in PTIME (cf. [**7**, Definition 7.5]).

4 Proof of Theorem 1

Recall that $P_0 \subseteq \Sigma^*$ and that $\mathbb{A}_0$ is an algorithm deciding P_0. A theory T is Σ_1-*complete* if every true arithmetical Σ_1-sentence is provable in T. The following result is a consequence of Lemma 2.1.

Lemma 4.1 *Assume that P_0 is not decidable in polynomial time. Let T be a computably enumerable Σ_1-complete theory with $T \vdash \mathit{Dec}_{P_0}(\dot{\mathbb{A}}_0)$. Then there is an algorithm $\mathbb{A}$ such that:*

(a) *The algorithm $\mathbb{A}$ is total (i.e., $t_{\mathbb{A}}(x) < \infty$ for all $x \in \Sigma^*$) and as fast as every algorithm T-provably deciding P_0;*

(b) *T is consistent if and only if $\mathbb{A}$ decides P_0.*

Moreover, there is a computable function diag *that maps any algorithm $\mathbb{E}$ enumerating some Σ_1-complete theory T with $T \vdash \mathit{Dec}_{P_0}(\dot{\mathbb{A}}_0)$ to an algorithm $\mathbb{A}$ with* (a) *and* (b)*.*

Proof. For an algorithm $\mathbb{B}$ let $\mathbb{B}\|\mathbb{A}_0$ be the algorithm that on input $x \in \Sigma^*$ runs $\mathbb{B}$ and $\mathbb{A}_0$ on x in parallel and returns the first answer obtained. Then

$$t_{\mathbb{B}\|\mathbb{A}_0} \leq O\Big(\min\big\{t_{\mathbb{B}}, t_{\mathbb{A}_0}\big\}\Big). \tag{4.1}$$

Claim 1 If T is consistent and $T \vdash \mathit{Dec}_{P_0}(\dot{\mathbb{B}})$, then $\mathbb{B}\|\mathbb{A}_0$ decides P_0.

Proof of Claim 1: By contradiction, assume that T is consistent and $T \vdash \mathit{Dec}_{P_0}(\dot{\mathbb{B}})$ but $\mathbb{B}\|\mathbb{A}_0$ does not decide P. Then $\mathbb{B}\|\mathbb{A}_0$ and $\mathbb{A}_0$ differ on some input $x \in \Sigma^*$. Thus $t_{\mathbb{B}}(x) \leq t_{\mathbb{A}_0}(x)$ and in particular $\mathbb{B}$ halts on x. Therefore, the following Σ_1-sentence φ is true:

$$\varphi := \exists x \exists y \exists y' \exists z \exists z' \big(\mathit{Run}(\dot{\mathbb{A}}_0, x, y, z) \wedge \mathit{Run}(\dot{\mathbb{B}}, x, y', z') \wedge \neg y = y'\big).$$

By Σ_1-completeness, $T \vdash \varphi$. However, φ logically implies $\neg \mathit{Dec}_{P_0}(\dot{\mathbb{B}})$ and thus T is inconsistent, a contradiction. $\dashv$

The set

$$D_1(T) := \Big\{\mathbb{B}\|\mathbb{A}_0 \;\Big|\; T \vdash \mathit{Dec}_{P_0}(\dot{\mathbb{B}})\Big\}$$

is nonempty as $\mathbb{A}_0\|\mathbb{A}_0 \in D_1(T)$ by assumption. Let $\mathbb{A}$ be the algorithm obtained for $D = D_1(T)$ by Lemma 2.1. We show that statement (a) holds. By Lemma 2.1(b), there is a $d \in \mathbb{N}$ such that for all $\mathbb{B}$ with $T \vdash \mathit{Dec}_{P_0}(\dot{\mathbb{B}})$ there is a $c_{\mathbb{B}}$ such that for all $x \in \Sigma^*$ we have $t_{\mathbb{A}}(x) \leq c_{\mathbb{B}} \cdot \big(t_{\mathbb{B}\|\mathbb{A}_0}(x) + |x|\big)^d$. Now (a) follows from (4.1).

For consistent T, by Claim 1 the set $D_1(T)$ only contains algorithms deciding P_0, thus $\mathbb{A}$ decides P_0 by Lemma 2.1.

If T is inconsistent, let $\mathbb{B}_{\text{bad}}$ be an algorithm that accepts every input in the first step. Then $\mathbb{B}_{\text{bad}}\|\mathbb{A}_0 \in D_1(T)$ by inconsistency of T. Thus, by Lemma 2.1(b), the algorithm runs in polynomial time and thus does not decide P_0.

As from an algorithm enumerating T we effectively get an algorithm enumerating $D_1(T)$, by Lemma 2.1 it should be clear that a computable function *diag* as claimed exists. □

Remark 4.2 As the preceding proof shows we only need the assumption that P_0 is not decidable in polynomial time in the proof of the implication from right to left in (b).

Proof of Theorem 1: Recall that Robinson introduced a finite, Σ_1-complete, and true arithmetical theory Q. Let P_0 be a decidable problem which is not decidable in polynomial time. Among others, the finite true arithmetical theory T_0 claimed to exist in Theorem 1 will extend Q and contain a formalization of Lemma 4.1.

We choose a Σ_1-formula $\mathit{Prov}(x, y)$ defining (in the standard model) the set of pairs (m, n) such that algorithm m enumerates a theory[2] that proves the sentence n. We let

$$\mathit{Con}(x) := \neg \mathit{Prov}\big(x, \ulcorner \neg 0 \dot{=} 0 \urcorner\big)$$

(here $\ulcorner \varphi \urcorner$ denotes the Gödel number of φ). If $\mathbb{E}$ enumerates a theory T, we write Con_T for $\mathit{Con}(\dot{\mathbb{E}})$.[3]

Let $f\colon \mathbb{N} \to \mathbb{N}$ be the function given by

$$f(m) := \ulcorner \mathit{Dec}_{P_0}(\dot{m}) \urcorner.$$

Both this function f and the function *diag* from Lemma 4.1 are computable and hence Σ_1-definable in Q. For the sake of completeness we recall what this means, say, for f. There is an arithmetical Σ_1-formula $\varphi_f(x, y)$ such that, for all $m, k \in \mathbb{N}$,

[2] We may assume that every enumeration algorithm enumerates a theory by deleting those printed strings that are not sentences.

[3] The notation is ambiguous, as the definition depends on the choice of $\mathbb{E}$, however not the arguments to follow.

- if $f(m) = k$, then $Q \vdash \varphi_f(\dot{m}, \dot{k})$;
- if $f(m) \neq k$, then $Q \vdash \neg\, \varphi_f(\dot{m}, \dot{k})$;
- $Q \vdash \exists^{=1} y\, \varphi_f(\dot{m}, y)$.

For better readability we write arithmetical formulas using f and *diag* as function symbols.

Further, let the arithmetical formula $\mathit{As\text{-}fast\text{-}as}(x, y)$ define the pairs (n, m) such that algorithm n is as fast as algorithm m and let $\mathit{Ptime}(x)$ define the set of polynomial time algorithms. Finally, we set

$$\mathit{Afap}(x, y) := \forall z\big(\mathit{Prov}(x, f(z)) \to \mathit{As\text{-}fast\text{-}as}(y, z)\big).$$

Then for an algorithm $\mathbb{E}$ enumerating a theory T the statement "the algorithm $F(T)$ is as fast as any algorithm T-provably deciding P_0", that is, the statement (a) in Theorem 1 is formalized by the sentence

$$\mathit{Afap}\big(\dot{\mathbb{E}}, \dot{F(T)}\big). \tag{4.2}$$

Let $\mathit{e\text{-}Rob}(x)$ be a Σ_1-formula expressing that the algorithm x enumerates a theory extending $Q \cup \{\mathit{Dec}_{P_0}(\dot{\mathbb{A}}_0)\}$.

We now define the theory T_0. It extends $Q \cup \{\mathit{Dec}_{P_0}(\dot{\mathbb{A}}_0)\}$ by the following sentences (s1)–(s5):

(s1) $\forall x\big(\mathit{e\text{-}Rob}(x) \to \mathit{Afap}(x, \mathit{diag}(x))\big)$,
(a formalization of Lemma 4.1(a))

(s2) $\forall x\big((\mathit{Con}(x) \wedge \mathit{e\text{-}Rob}(x)) \to \mathit{Dec}_{P_0}(\mathit{diag}(x))\big)$,
(a formalization of part of Lemma 4.1(b))

(s3) $\forall x(\mathit{Ptime}(x) \to \neg \mathit{Dec}_{P_0}(x))$,
(P_0 is not in P)

(s4) $\forall x\big(\neg \mathit{Con}(x) \to \forall y(\mathit{Sent}(y) \to \mathit{Prov}(x, y))\big)$
(every inconsistent theory proves every sentence; here $\mathit{Sent}(y)$ is a Δ_0-formula defining the first-order L_{all}-sentences)

(s5) $\forall x \forall y\big((\mathit{As\text{-}fast\text{-}as}(x, y) \wedge \mathit{Ptime}(y)) \to \mathit{Ptime}(x)\big)$
(if algorithm x is as fast as the polynomial algorithm y, then it is polynomial too).

Let T be a computably enumerable extension of T_0 and let $\mathbb{E}$ be an algorithm enumerating T. We claim that for the algorithm

$$F(T) := \mathit{diag}(\mathbb{E})$$

(see Lemma 4.1) the statements (a) and (b) of Theorem 1 hold.

The arithmetical sentence $\dot{F(T)} = \mathit{diag}(\dot{\mathbb{E}})$ is Σ_1 and true, so T_0 proves it by Σ_1-completeness (as $T_0 \supseteq Q$). By the same reason, $T_0 \vdash \mathit{e\text{-}Rob}(\dot{\mathbb{E}})$. As T_0 contains (s1), $T_0 \vdash \mathit{Afap}(\dot{\mathbb{E}}, \dot{F(T)})$; that is, T_0 proves that $F(T)$ is as fast as any algorithm T-provably deciding P_0. Thus (a) in Theorem 1 holds.

We turn to (b). Let T^* be a theory with $T^* \supseteq T$.

(i) $\Rightarrow$ (ii): So, we assume that $T^* \vdash \mathit{Con}_T$. We already know that T_0, and hence T^*, proves $\mathit{e\text{-}Rob}(\dot{\mathbb{E}})$. As T^* contains (s2), for $x = \dot{\mathbb{E}}$ we see that $T^* \vdash \mathit{Dec}_{P_0}(\mathit{diag}(\dot{\mathbb{E}}))$ and thus $T^* \vdash \mathit{Dec}_{P_0}(\dot{F(T)})$; that is, $F(T)$ T^*-provably decides P_0.

(ii) $\Rightarrow$ (iii): Immediate by part (a) of the theorem.

(iii) $\Rightarrow$ (i): Let $\mathbb{A}$ be an algorithm such that $T^* \vdash \mathit{Dec}_{P_0}(\dot{\mathbb{A}})$ and $T^* \vdash \mathit{Afap}(\dot{\mathbb{E}}, \dot{\mathbb{A}})$; the latter means that

$$T^* \vdash \forall z(\mathit{Prov}(\dot{\mathbb{E}}, f(z)) \to \mathit{As\text{-}fast\text{-}as}(\dot{\mathbb{A}}, z)). \tag{4.3}$$

Let $\mathbb{B}$ be an algorithm such that

$$T^* \vdash \mathit{Ptime}(\dot{\mathbb{B}}). \tag{4.4}$$

Then T^* proves the following implications:

$$\begin{aligned}
&\neg \mathit{Con}_T \to \mathit{Prov}(\dot{\mathbb{E}}, f(\dot{\mathbb{B}})) && \text{(by (s4) and as } \mathit{Sent}(f(\dot{\mathbb{B}})) \text{ is } \Sigma_1)\\
&\neg \mathit{Con}_T \to \mathit{As\text{-}fast\text{-}as}(\dot{\mathbb{A}}, \dot{\mathbb{B}}) && \text{(by (4.3))}\\
&\neg \mathit{Con}_T \to \mathit{Ptime}(\dot{\mathbb{A}}) && \text{(by (4.4) and (s5))}\\
&\neg \mathit{Con}_T \to \neg \mathit{Dec}_{P_0}(\dot{\mathbb{A}}) && \text{(by (s3)).}
\end{aligned}$$

As $T^* \vdash \mathit{Dec}_{P_0}(\dot{\mathbb{A}})$, we see that $T^* \vdash \mathit{Con}_T$. □

5 Gödel's Second Incompleteness Theorem

Let P_{exp} be the following problem:

P_{exp}
Instance: An algorithm $\mathbb{A}$.
Problem: Is it true that $\mathbb{A}$ does not accept $\mathbb{A}$ in at most $2^{|\mathbb{A}|}$ steps?

Theorem 5.1 ([8]) *There is a polynomial time computable function g that maps any algorithm $\mathbb{A}$ deciding P_{exp} to an algorithm $g(\mathbb{A})$ deciding P_{exp} such that $\mathbb{A}$ is not as fast as $g(\mathbb{A})$.*

Proof. We fix a polynomial time computable function which assigns to every algorithm $\mathbb{A}$ and $n \geq 1$ an algorithm $\mathbb{A}_n$ where $\mathbb{A}_n$ is "the same as $\mathbb{A}$ but padded with n useless instructions". The properties of $\mathbb{A}_n$ we need are

$$|\mathbb{A}_n| \geq n, \; t_{\mathbb{A}_n} = t_{\mathbb{A}}, \text{ and } \mathbb{A}_n \text{ and } \mathbb{A} \text{ accept the same language.} \tag{5.1}$$

Note that any algorithm $\mathbb{A}$ deciding P_{exp} does not reject $\mathbb{A}$. Hence, for such an $\mathbb{A}$ we have $\mathbb{A} \in P_{\text{exp}}$ and $t_{\mathbb{A}}(\mathbb{A}) > 2^{|\mathbb{A}|}$. Moreover, by (5.1), we have

$$t_{\mathbb{A}}(\mathbb{A}_n) = t_{\mathbb{A}_n}(\mathbb{A}_n) > 2^{|\mathbb{A}_n|} \geq 2^n \tag{5.2}$$

(the strict inequality holding as $\mathbb{A}_n$ decides P_{exp}, too).

The function g computes for any algorithm $\mathbb{A}$ the following algorithm $\mathbb{B} := g(\mathbb{A})$: On input x the algorithm $\mathbb{B}$ first checks whether $x \in \{\mathbb{A}_1, \mathbb{A}_2, \ldots\}$ (this can be done in time polynomial in $|x|$); if so, $\mathbb{B}$ immediately accepts, otherwise it simulates $\mathbb{A}$ on x and answers accordingly. Clearly, if $\mathbb{A}$ decides P_{exp}, then $\mathbb{B}$ decides P_{exp} and superpolynomially speeds up $\mathbb{A}$ on $\{\mathbb{A}_1, \mathbb{A}_2, \ldots\}$ by (5.2). □

Using this result and results of preceding sections we derive the following version of Gödel's Second Incompleteness Theorem:

Theorem 5.2 *There is a finite true arithmetical theory T_1 such that for every computably enumerable theory $T \supseteq T_1$,*

$$\textit{if } T \textit{ is consistent, then } T \textit{ does not prove } \mathit{Con}_T.$$

Proof. We take as P_0 the problem P_{exp} of the preceding theorem and let g be the function defined there. We know that P_0 is not decidable in polynomial time. Furthermore, as in the previous sections, we fix an algorithm $\mathbb{A}_0$ deciding P_0. Let T_0 be the true arithmetical, finite, and Σ_1-complete theory defined in the previous section satisfying Theorem 1.

Being computable, g is Σ_1-definable; for simplicity of notation we use g like a function symbol in arithmetical formulas. This is to be understood as explained in the previous proof. The theory T_1 is obtained from T_0 by adding the sentence

(s6) $\forall x(\mathit{Dec}_{P_0}(x) \to \mathit{Dec}_{P_0}(g(x)))$.

Let $T \supseteq T_1$ be a theory enumerated by the algorithm $\mathbb{E}$. Assume that T is consistent. Then, by Lemma 4.1(b),

$$\mathit{diag}(\mathbb{E}) \text{ decides } P_0$$

and thus, by Theorem 5.1,

(5.3) $$\mathit{diag}(\mathbb{E}) \text{ is not as fast as } g(\mathit{diag}(\mathbb{E})).$$

Observe that $T \vdash \textit{e-Rob}(\dot{\mathbb{E}})$ being a true Σ_1-sentence. By contradiction, suppose that $T \vdash \mathit{Con}_T$, that is, $T \vdash \mathit{Con}(\dot{\mathbb{E}})$. Then, $T \vdash \mathit{Dec}_{P_0}(\mathit{diag}(\dot{\mathbb{E}}))$ by (s2) and hence, $T \vdash \mathit{Dec}_{P_0}(g(\mathit{diag}(\dot{\mathbb{E}}))$ by (s6). Setting $\mathbb{B} := g(\mathit{diag}(\mathbb{E}))$ the sentence $\dot{\mathbb{B}} = g(\mathit{diag}(\dot{\mathbb{E}}))$ is a true Σ_1-sentence; so T proves it. Then, $T \vdash \mathit{Dec}_{P_0}(\dot{\mathbb{B}})$. This means that $\mathbb{B} = g(\mathit{diag}(\mathbb{E}))$ T-provably decides P_0. By Lemma 4.1(a), $\mathit{diag}(\mathbb{E})$ is as fast as $g(\mathit{diag}(\mathbb{E}))$ contradicting (5.3). □

Let T_1 be the theory just defined. We show that for every true and computably enumerable arithmetical theory $T \supseteq T_1$, the extension $T \cup \{\mathit{Con}_T\}$ knows of strictly faster algorithms deciding P_{exp} than T:

Corollary 5.3 *Let T_1 be the theory defined in the previous proof. Then for every true and computably enumerable arithmetical theory $T \supseteq T_1$ there is an algorithm $\mathbb{A}$ such that:*

(a) *The algorithm $\mathbb{A}$ $T \cup \{\mathit{Con}_T\}$-provably decides P_{exp} and is as fast as every algorithm that T-provably decides P_{exp}.*

(b) *No algorithm that T-provably decides P_{exp} is as fast as $\mathbb{A}$.*

Proof. Let T be as stated. By Theorem 1 for $P_0 := P_{\text{exp}}$ and $T^* := T \cup \{\mathit{Con}_T\}$, we get that the algorithm $\mathbb{A} := F(T)$ $T \cup \{\mathit{Con}_T\}$-provably decides P_0. Furthermore,

(5.4) $$\mathbb{A} \text{ is as fast as any algorithm that } T\text{-provably decides } P_0.$$

This shows (a). For (b), let $\mathbb{B}$ be an arbitrary algorithm that T-provably decides P_0. Then, by (s6),

(5.5) $$\text{the algorithm } g(\mathbb{B}) \ T\text{-provably decides } P_0.$$

As T is a true arithmetical theory, the algorithms $\mathbb{B}$ (and $g(\mathbb{B})$) decide P_0. Hence, by Theorem 5.1,

(5.6) $$\mathbb{B} \text{ is not as fast as } g(\mathbb{B}).$$

From (5.4)–(5.6) we conclude that $\mathbb{B}$ is not as fast as $\mathbb{A}$. □

One can get rid of the assumption that T must be a *true* arithmetical theory in the previous result by adding to T_1 a further true arithmetical sentence:

Corollary 5.4 *There is a finite true arithmetical theory T_2 such that for every consistent, computably enumerable theory $T \supseteq T_2$ there is an algorithm $\mathbb{A}$ such that:*

(a) *The algorithm $\mathbb{A}$ $T \cup \{Con_T\}$-provably decides P_{exp} and is as fast as every algorithm that T-provably decides P_{exp}.*

(b) *No algorithm that T-provably decides P_{exp} is as fast as $\mathbb{A}$.*

Proof. Again, we take as P_0 the problem P_{exp} and let $\mathbb{A}_0$ be an algorithm deciding it. Let h be the function that maps an algorithm $\mathbb{B}$ to $\mathbb{B}||\mathbb{A}_0$ (as in the proof of Lemma 4.1 the algorithm $\mathbb{B}||\mathbb{A}_0$ on input $x \in \Sigma^*$ runs $\mathbb{B}$ and $\mathbb{A}_0$ in parallel and returns the first answer obtained).

The theory T_2 is obtained from T_1 by adding the true arithmetical sentence

(s7) $\forall x(Dec_{P_0}(x) \to Dec_{P_0}(h(x)))$.

Let T be as stated and again let $\mathbb{A} := F(T)$. As in the previous proof, we see that statement (a) holds true.

For (b), let $\mathbb{B}$ be an algorithm with $T \vdash Dec_{P_0}(\dot{\mathbb{B}})$. Using first (s7) and then (s6) we get

$$\text{the algorithm } g(h(\mathbb{B}))\ T\text{-provably decides } P_0.$$

By (a), therefore it suffices to show that $\mathbb{B}$ is not as fast as $g(h(\mathbb{B}))$. As by definition of h, the algorithm $h(\mathbb{B})$ is as fast as $\mathbb{B}$ (see (4.1)), it already suffices to show that $h(\mathbb{B})$ is not as fast as $g(h(\mathbb{B}))$. By Claim 1 in the proof of Lemma 4.1, we know that $h(\mathbb{B})$ decides P_0. Then Theorem 5.1 indeed proves that $h(\mathbb{B})$ is not as fast as $g(h(\mathbb{B}))$. □

Acknowledgements

The authors thank the John Templeton Foundation for its support under Grant #13152, *The Myriad Aspects of Infinity.* Yijia Chen is affiliated with BASICS and MOE-MS Key Laboratory for Intelligent Computing and Intelligent Systems, which is supported by National Nature Science Foundation of China (61033002).

References

[1] S. A. Cook and P. Nguyen. *Logical Foundations of Proof Complexity*. Cambridge University Press, 2010.

[2] J. Hartmanis. Relations between diagonalization, proof systems, and complexity gaps. *Theoretical Computer Science*, 8:239–253, 1979.

[3] M. Hutter. The fastest and shortest algorithm for all well-defined problems. *International Journal of Foundations of Computer Science*, 13:431–443, 2002.

[4] J. Krajíček and P. Pudlák. Propositional proof systems, the consistency of first order theories and the complexity of computations. *The Journal of Symbolic Logic*, 54:1063–1079, 1989.

[5] L. Levin. Universal search problems (in Russian). *Problemy Peredachi Informatsii*, 9:115–116, 1973.

[6] J. Messner. On optimal algorithms and optimal proof systems. In *Proceedings of the 16th Symposium on Theoretical Aspects of Computer Science (STACS'99)*, Lecture Notes in Computer Science 1563, 361–372, 1999.

[7] Z. Sadowski. On an optimal propositional proof system and the structure of easy subsets. *Theoretical Computer Science*, 288:181–193, 2002.

[8] L. Stockmeyer. *The complexity of decision problems in automata theory and logic.* PhD thesis, MIT, 1974.

Some definitorial suggestions for parameterized proof complexity

Jörg Flum*, **Moritz Müller**†

[‡] Mathematisches Institut, Albert-Ludwigs-Universität Freiburg, Germany
joerg.flum@math.uni-freiburg.de

[§] Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
moritz.mueller@univie.ac.at

Abstract. We introduce a (new) notion of parameterized proof system. For parameterized versions of standard proof systems such as Extended Frege and Substitution Frege, we compare their complexity with respect to parameterized simulations.

Introduction

Consider the following problems for graphs: the vertex cover problem VC, the clique problem CLIQUE, and the dominating set problem DS; they ask, given a graph G and a natural number k, whether G contains a cardinality k vertex cover, clique, and dominating set, respectively. All three problems are NP-complete and hence, from the point of view of polynomial reductions any two of them have the same computational complexity.

Taking in each case the natural number k as the parameter of an instance we get the parameterized problems p-VC, p-CLIQUE, and p-DS. In parameterized complexity there is not only a new notion of tractability, namely fixed-parameter tractability, but also the notion of reducibility has been adapted so that it preserves fixed-parameter tractability; the new notion being that of fpt-reduction. One knows that p-VC $\leq_{\mathrm{fpt}}$ p-CLIQUE (that is, p-VC is fpt-reducible to p-CLIQUE) and p-CLIQUE $\leq_{\mathrm{fpt}}$ p-DS. However, accepting the hypotheses FPT $\neq$ W[1] and W[1] $\neq$ W[2] (which are fundamental hypotheses of parameterized complexity and each of them implies P $\neq$ NP) neither p-CLIQUE $\leq_{\mathrm{fpt}}$ p-VC nor p-DS $\leq_{\mathrm{fpt}}$ p-CLIQUE. As Downey and Fellows write in [**7**]:

> *Parameterized reductions tend to be much more* structure preserving *than classical reductions, and certainly most classical reductions . . . are definitely not parameterized reductions. . . . Parameterized reductions are sufficientlly refined that instead of one large class of naturally intractable problems all of the same complexity, there seem to be many sets of natural combinatorial problems, all intractable in the parameterized sense, and yet of differing parameterized complexity.*

In proof theory among the proof systems best studied there are Frege systems, Extended Frege systems, and Substitution Frege systems. Classically, they are compared via polynomial simulations. It is known that there are polynomial simulations between any Extended Frege system and any Substitution Frege system, while it is not known whether Extended Frege systems and Substitution Frege systems may be simulated by Frege systems. The question arises whether also in this context parameterized complexity yields new insights or even allows a more fine-grained analysis. In this note we want to lay down the conceptual framework for such an analysis. Furthermore, we give some positive and some negative answers and state some open problems.

What are natural parameterizations of proof systems? Recall that the definitions of parameterized complexity are tailored to address complexity issues in situations where we know that the parameter is relatively small. We believe that for Extended Frege systems the number of extension axioms used in a proof could be a natural parameter. At least, if we start with an arbitrary, say, random tautology it does not seem plausible that many extension axioms can be used in a proof with advantage. We should emphasize the word "random" here. For example, in a standard example often mentioned to motivate the use of extension axioms, namely the formalization of the pigeon-principle in propositional logic, the number of extension axioms used to derive the n pigeonhole principle by a straightforward induction on n is $\Omega(n^3)$ and hence, by no means, relatively small.[1] Similarly the number of applications of the substitution rule seems to be a natural parameter for Substitution Frege Systems.

As proof systems are functions, simulations between them should be value-preserving functions (as are the standard polynomial simulations). We believe that this fact has not been taken into account appropriately in the approaches to proof theory using parameterized complexity. Taking this fact seriously, we define the notion of fpt-simulation. When we realized that our notion coincides with the notion of parsimonious reduction between parameterized counting functions, we were confirmed in our belief that this is the appropriate definition.

We show that under fpt-simulations the parameterized versions of Extended Frege and Substitution Frege are both equivalent to Frege. In this sense, the notion of fpt-simulation does not offer a more fine-grained complexity analysis of these proof systems; or, expressing it in positive terms, we gain the insight that there is a simulation, say, of an Extended Frege system in a Frege system whose superpolynomial running time is confined to a factor depending only on the number of extension axioms used in the original proof. Similarly, we see that there is a simulation of Substitution Frege in Extended Frege where the number of extension axioms is bounded in terms of the number of applications of the substitution rule.

Having in mind the goal of a more refined analysis, we propose to study the relationship between these proofs systems under parameterized polynomial simulations, a notion that in some sense refines both, polynomial simulations and fpt-simulations: such a simulation is a polynomial simulation with the additional property that it increases the parameter at most polynomially. We do not see any way to simulate Substitution Frege in Extended Frege in this sense (while conversely this is easy). However, we construct a parameterized polynomial simulation of treelike Substitution Frege in treelike Extended Frege.

Related work

A different approach to introduce parameterizations into proof complexity has been initiated by Dantchev et al. [**6**]. They introduced parameterized proof systems for *parameterized* problems. They considered the following parameterized problem: given a pair (α, k) of a CNF α and $k \in \mathbb{N}$, where k is the parameter, decide whether α has no satisfying assignment of Hamming weight at most k. The proof systems they had in mind are classical refutation systems such as Resolution that may freely use additional clauses expressing

[1] It is well-known that Buss [**3**] gave polynomial proofs of the pigeon-principle in Frege systems.

the constraint on the Hamming weight. The goal of this approach is to strengthen lower bounds of classical refutation systems by showing that their parameterized counterparts are not *fpt bounded.*[2] It can be understood as a parameterized analogue of Cook's program, here trying to prove coW[2] $\not\subseteq$ paraNP. For this approach Beyersdorff et al. [**1**] lack an interpretation of the parameterization of the proof system and argue that it can be dispensed with.

1 Preliminaries

In this section we fix some notations and recall some definitions and results, in the first part of parameterized complexity theory and in the second part of proof theory.

1.1 Parameterized complexity

Formally, a *parameterized problem* is a pair (Q, κ) consisting of a (classical) problem $Q \subseteq \{0,1\}^*$ and a polynomial time computable *parameterization* $\kappa\colon \{0,1\}^* \to \mathbb{N}$ that maps any input $x \in \{0,1\}^*$ to its *parameter* $\kappa(x) \in \mathbb{N}$. A parameterized problem (Q, κ) is *fixed-parameter tractable*, that is, tractable from the point of view of parameterized complexity, if there is an algorithm solving $x \in Q$ in $\leq f(\kappa(x)) \cdot |x|^{O(1)}$ steps for some computable $f\colon \mathbb{N} \to \mathbb{N}$.

A function $R\colon \{0,1\}^* \to \{0,1\}^*$ is fpt-*computable* with respect to a parameterization κ if $R(x)$ can be computed in time $f(\kappa(x)) \cdot |x|^{O(1)}$, where again $f\colon \mathbb{N} \to \mathbb{N}$ is computable.

Also the notion of polynomial reduction, that is, the natural notion of reduction preserving classical tractability, has to be adapted so that it preserves fixed-parameter tractability. An fpt-*reduction* R from a parameterized problem (Q, κ) to another (Q', κ') is an fpt-computable (with respect to κ) reduction from Q to Q' such that $\kappa'(R(x)) \leq g(\kappa(x))$ for some computable $g\colon \mathbb{N} \to \mathbb{N}$ and all $x \in \{0,1\}^*$. We write $(Q, \kappa) \leq_{\text{fpt}} (Q', \kappa')$ if there is an fpt-reduction from (Q, κ) to (Q', κ').

1.2 Proof theory

A *proof system* for a problem $Q \subseteq \{0,1\}^*$ is a polynomial time computable surjection P from $\{0,1\}^*$ onto Q. If $P(w) = x$, then w is a P-*proof* of x. In case $Q = \textsc{Taut}$, we call P *propositional.* A proof system P is *p-bounded* if any $x \in Q$ has a P-proof of size $|x|^{O(1)}$. Cook and Reckhow [**5**] observed that a p-bounded propositional proof system exists if and only if NP = coNP. Cook's program asks to prove that natural propositional proof systems are not p-bounded.

Proof systems for a problem Q are compared in strength via p-simulations: a *p-simulation* of a proof system P' in a proof system P is a polynomial time computable function R such that $P(R(w')) = P'(w')$ for all $w' \in \{0,1\}^*$; in case such an R exists, we say P *p-simulates* P' and write $P' \leq_{\text{pol}} P$; if additionally, P' p-simulates P, we call P and P' *p-equivalent.*

A *Frege system* F is a propositional proof system given by finitely many axiom schemes (in the de Morgan language) and finitely many rules including, for simplicity, modus ponens. An F-*proof* of a (propositional) formula α from a set of formulas Γ is a sequence of formulas such that each of them is either a member of Γ or a substitution instance of an axiom scheme or follows from earlier formulas in the sequence by one of

[2] As pointed out in [**1**] one should restrict attention to instances (α, k) with contradictory α.

the rules of F; furthermore, the last formula of the sequence is α. An F-proof of α is an F-proof of α from the empty set of formulas. Frege systems are assumed to be *implicationally complete*, that is, whenever a set of formulas Γ logically implies α, then there exists an F-proof of α from Γ.

For a Frege system F we denote by F^* the proof system *treelike* F: an F-proof π is *treelike* if every occurrence of a formula in π is used as an hypothesis in an application of a rule at most once; equivalently, π is treelike if it can be written as a tree labeled by the formulas in π such that the leaves are labeled by the substitution instances of the axiom schemes and the labels of inner nodes are obtained by one of the rules from their immediate predecessors.

The following are well-known [**5, 10**].

Theorem 1.1

(1) (Cook, Reckhoff) *Any two Frege systems are p-equivalent.*
(2) (Krajíček) *F and F^* are p-equivalent for every Frege system F.*

By part (1) of this theorem we get that, instead of (2), we could claim

$$F_1 \text{ and } F_2^* \text{ are } p\text{-equivalent for Frege systems } F_1 \text{ and } F_2.$$

The same observation applies to all equivalences mentioned in this paper (not only to p-equivalences but also to fpt-equivalences and *pp*-equivalences introduced later).

There are two well-studied extensions of a Frege system F:

Extension Frege. Let F be a Frege system. The *Extension Frege system EF* adds to F the *extension rule*: It allows to add in a proof of α (without any hypotheses) an *extension axiom* $(r \leftrightarrow \sigma)$ where σ is a propositional formula and the *extension variable* r neither occurs in σ nor in α nor in any earlier line of the proof.

Equivalently, an *EF*-proof of α is an F-proof of α from an extension sequence whose extension variables do not occur in α. Here, an *extension sequence* (for α) of length k is a sequence of the form

$$(r_1 \leftrightarrow \sigma_1), \dots, (r_k \leftrightarrow \sigma_k)$$

with pairwise distinct *extension variables* $r_1, \dots, r_k$ such that r_i does not occur in σ_j for $1 \leq j \leq i$.

By EF^* we denote the treelike version of *EF*.

Substitution Frege. Let F be a Frege system. The *Substitution Frege system SF* adds to F the *substitution rule* that allows to derive from the formula α the formula $\alpha[x/\sigma]$ where $\alpha[x/\sigma]$ is obtained from α by substituting the variable x by the formula σ. By SF^* we denote the treelike version of *SF*.

In [**2**] Buss introduces two restrictions of *SF*:

- *Boolean Substitution Frege BSF* requires that in any application of the substitution rule the formula σ to be the Boolean constant $\top$ (TRUE) or $\bot$ (FALSE).
- *Renaming Frege RF* requires σ to be a variable.

Again, BSF^* and RF^* denote the treelike versions of these systems.

Natural simulations of *EF* and *SF* in F roughly proceed as follows:

- Let π be an *EF*-proof. To delete the first extension axiom $(r \leftrightarrow \sigma)$ substitute everywhere in π the formula σ for r; this transforms the extension axiom into the tautology $(\sigma \leftrightarrow \sigma)$ for which we add a linear size F-proof. Proceed like this with the second extension axiom and so on. If π contains k extension axioms, the resulting F-proof has size $|\pi|^{O(k)}$.

- Let π be an *SF*-proof. Let the first application in π of the substitution rule yield $\alpha[x/\sigma]$ from α. Replace it by a proof of $\alpha[x/\sigma]$ obtained by applying the substitution x/σ to the initial segment of π up to α. If π contains k substitution inferences, the resulting F-proof has size $|\pi|^{O(k)}$.

Hence, both simulations are not polynomial ones. In fact, it is open whether $EF \leq_{\text{pol}} F$ and whether $SF \leq_{\text{pol}} F$. However, the following is known [**2, 12**]:

Theorem 1.2

(1) *EF, EF*, SF, SF*, RF, BSF are p-equivalent for every Frege system F.*
(2) *RF*, BSF* and F are p-equivalent for every Frege system F.*

Comparing their status with that of RF^* and of BSF^* we see that perhaps *RF* and *BSF* are proof systems where the ability to reuse already derived lines adds power. We shall see a similar phenomenon for *SF* in the parameterized setting.

2 Parameterized proof systems and fpt-simulations

In this section we introduce the main new concepts of this paper, parameterized proof systems and simulations between them.

Definition 2.1 A *parameterized proof system for Q* is a pair (P, κ) such that P is a proof system for Q and κ a parameterization.

Having in mind, as we do, to compare Frege systems, Extended Frege systems, and Substitution Frege systems, it seems not natural to consider a more general notion of parameterized proof systems where P is only required to be an fpt-computable (with respect to κ) function from $\{0,1\}^*$ onto Q instead of a polynomial time computable one.

We identify a (classical) proof system P for Q with the parameterized proof system $(P, 0)$, i.e., P with the parameterization that is constantly 0.

For an Extended Frege systems *EF* we denote by κ_{EF} the parameterization

$$\kappa_{EF}(w) := \text{number of extension axioms in } w.$$

Similarly, for a Substitution Frege systems *SF* we denote by κ_{SF} the parameterization

$$\kappa_{SF}(w) := \text{number of applications of the substitution rule in } w.$$

We consider the restriction EF^* of *EF* with the parameterization κ_{EF} and the restrictions SF^*, $BSF^{(*)}$, and $RF^{(*)}$ of *SF* with the parameterization κ_{SF}. We denote the resulting parameterized proof systems by *p-EF*, *p-EF**, *p-SF*, *p-RF*, *p-BSF*, *p-SF**, *p-RF** and *p-BSF**.

In order to compare parameterized proof systems in strength we use the following notion of simulation. We already mentioned that for parameterized counting problems the notion coincides with that of fpt parsimonious reduction introduced in [**8**, Definition 14.10].

Definition 2.2 Let (P, κ) and (P', κ') be parameterized proof systems for $Q \subseteq \{0,1\}^*$. An fpt-*simulation* of (P', κ') in (P, κ) is a function $R\colon \{0,1\}^* \to \{0,1\}^*$ such that

(a) R is fpt-computable with repect to κ';
(b) $P'(w') = P(R(w'))$ for all $w' \in \{0,1\}^*$;

(c) $\kappa(R(w')) \leq g(\kappa'(w'))$ for some computable $g\colon \mathbb{N} \to \mathbb{N}$ and all $w' \in \{0,1\}^*$.

In case such an R exists, we say that (P,κ) fpt-*simulates* (P',κ') and write $(P',\kappa') \leq_{\text{fpt}} (P,\kappa)$. The problems (P,κ) and (P',κ') are fpt-*equivalent*, written $(P,\kappa) \equiv_{\text{fpt}} (P,\kappa)$, if $(P,\kappa) \leq_{\text{fpt}} (P',\kappa')$ and $(P',\kappa') \leq_{\text{fpt}} (P,\kappa)$.

Note that if P and P' are classical proof systems for a problem Q, then P fpt-simulates P' if and only if P p-simulates P'. However, in general, neither $(P,\kappa) \leq_{\text{fpt}} (P',\kappa')$ implies $P \leq_{\text{pol}} P'$ nor $P \leq_{\text{pol}} P'$ implies $(P,\kappa) \leq_{\text{fpt}} (P',\kappa')$.

Lemma 2.3 *If $(P,\kappa) \leq_{\text{fpt}} (P',\kappa')$ and $(P',\kappa') \leq_{\text{fpt}} (P'',\kappa'')$, then $(P,\kappa) \leq_{\text{fpt}} (P'',\kappa'')$.*

3 Comparing proof systems via fpt-simulations

By the following result all parameterized proof systems introduced so far are fpt-equivalent.

Theorem 3.1 *p-EF, p-SF, and F are pairwise* fpt-*equivalent.*[3]

As $F \leq_{\text{fpt}}$ *p-EF*, the theorem follows from the following three propositions showing (among others):

$$p\text{-}EF \leq_{\text{fpt}} p\text{-}SF \leq_{\text{fpt}} p\text{-}BSF \leq_{\text{fpt}} F.$$

In Proposition 3.2 and Proposition 3.3 we obtain the first two 'inequalities' by merely observing that known p-simulations already are fpt-simulations.

Proposition 3.2 *p-EF $\leq_{\text{fpt}}$ p-SF and p-EF* $\leq_{\text{fpt}}$ p-SF*.*

Proof. Cook and Reckhow's original p-simulation [**5**] of *EF* in *SF* is an fpt-simulation of *p-EF* in *p-SF*; this yields the first assertion.

We turn to the second claim. An *EF**-proof π of α is an F^*-proof of α from an extension sequence $(r_1 \leftrightarrow \sigma_1), \dots, (r_k \leftrightarrow \sigma_k)$ (recall that the r_i have to be pairwise distinct and that r_i neither occurs in σ_j for $1 \leq j \leq i$ nor in α). By the Deduction Theorem for F (see [**11**, Lemma 4.4.10]) there is an F-proof π' of

$$(r_k \leftrightarrow \sigma_k) \to (r_{k-1} \leftrightarrow \sigma_{k-1}) \to \cdots \to (r_1 \leftrightarrow \sigma_1) \to \alpha \tag{3.1}$$

(where the iterated implications are associated to the right) of size $|\pi|^{O(1)}$. By part (2) of Theorem 1.1 we can assume that π' is treelike.

By our assumption on the extension variables, the variable r_k occurs exactly once in (3.1). We apply the substitution rule and substitute σ_k for r_k in (3.1); hence we get the formula obtained from (3.1) by replacing the equivalence $(r_k \leftrightarrow \sigma_k)$ by $(\sigma_k \leftrightarrow \sigma_k)$. We add a short F^*-proof of $(\sigma_k \leftrightarrow \sigma_k)$ and apply modus ponens to arrive at formula (3.1) with $k-1$ instead of k. Repeating this process gives an *SF**-proof of α of size $O(k \cdot |\pi'|)$. We observe that in this simulation k extension axioms are simulated in *SF** by k applications of the substitution rule. Therefore, this is an fpt-simulation. □

[3] The second author gave a talk at the workshop *Proof Complexity* (11w5103, Banff International Research Station) on this subject mentioning that at that time we did not know whether *p-EF* $\leq_{\text{fpt}}$ F. Kaveh Ghasemloo pointed out that he was convinced that such a simulation could be constructed via the system G_1^* (cf. [**4**, p. 179]).

Proposition 3.3 *p-SF* $\leq_{\mathrm{fpt}}$ *p-BSF.*

Proof. Buss [2] simulates an application of the substitution rule $\frac{\alpha}{\alpha[x/\sigma]}$ as follows: first, he applies twice the *BSF*-substitution rule to get

$$\alpha[x/\top] \text{ and } \alpha[x/\bot]$$

from α; then he adds short proofs of

$$((\sigma \wedge \alpha[x/\top]) \rightarrow \alpha[x/\sigma]) \text{ and } ((\neg\sigma \wedge \alpha[x/\bot]) \rightarrow \alpha[x/\sigma]).$$

Finally, he derives $\alpha[x/\sigma]$ from these four formulas.

In this way, an *SF*-proof with k applications of the substitution rule is transformed in polynomial time into an *BSF*-proof with $2k$ applications of the *BSF*-substitution rule. Hence, this is an fpt-simulation. □

Proposition 3.4 *p-BSF* $\leq_{\mathrm{fpt}}$ *F.*

Proof. Let π be an *BSF*-proof of β with k applications of the *BSF*-substitution rule. Let π_1 be the initial segment of π that ends in the premise α of the first application $\frac{\alpha}{\alpha[x/\sigma]}$with $\sigma \in \{\top, \bot\}$ of this rule. We obtain the *F*-proof π_1' of $\alpha[x/\sigma]$ by applying the substitution x/σ to every line of π_1. Furthermore, delete all occurrences of $\alpha[x/\sigma]$in π, thus getting π'. Then π_1', π' is a *BSF*-proof of β with $(k-1)$ applications of the *BSF*-substitution rule and of size at most $2|\pi|$. Repeating this process we finally obtain an *F*-proof of β of size $2^k \cdot |\pi|$. □

Standard *p*-simulations of *SF* in *EF* (e.g., see [12]) map an *SF*-proof π of a formula $\alpha(\overline{x})$ (where $\overline{x}$ are the propositional variables in α) with k applications of the substitution rule and ℓ lines to an *EF*-proof with $\ell \cdot |\overline{x}|$ extension axioms. They are not fpt-simulations. By the previous theorem there is an fpt-simulation of *p-SF* in *p-EF*. We encourage the reader to give a 'direct' one.

4 Comparing proof systems via parameterized polynomial simulations

In the previous section we have seen that fpt-simulations are too coarse in the sense that they do not distinguish any two of the parameterized proof system considered so far. In this section therefore we analyze these proof systems under a notion of simulation which strengthens both, the notion of *p*-simulation and that of fpt-simulation. For parameterized decision problems this concept was introduced in [9].

Definition 4.1 Let (P, κ) and (P', κ') be parameterized proof systems for $Q \subseteq \{0,1\}^*$. A *pp-simulation* (or *parameterized polynomial simulation*) of (P', κ') in (P, κ) is a *p*-simulation R of P' in P such that

$$\kappa'(R(w')) \leq q(\kappa(w')) \text{ for some polynomial } q \text{ and all } w' \in \{0,1\}^*.$$

In case such an R exists, we say that (P, κ) *pp-simulates* (P', κ') and write $(P', \kappa') \leq_{\mathrm{pp}} (P, \kappa)$. The problems (P, κ) and (P', κ') are *pp-equivalent*, written $(P, \kappa) \equiv_{\mathrm{pp}} (P, \kappa)$, if $(P, \kappa) \leq_{\mathrm{pp}} (P', \kappa')$ and $(P', \kappa') \leq_{\mathrm{pp}} (P, \kappa)$.

Clearly, if $(P', \kappa') \leq_{\mathrm{pp}} (P, \kappa)$, then $P' \leq_{\mathrm{pol}} P$ and $(P', \kappa') \leq_{\mathrm{fpt}} (P, \kappa)$.

As the proofs of Proposition 3.2 and of Proposition 3.3 show, we get:

Proposition 4.2 *p-$EF \leq_{\rm pp} p$-SF, p-$EF^* \leq_{\rm pp} p$-SF^*, and p-$SF \leq_{\rm pp} p$-BSF.*

Example 4.3 The p-simulation of *BSF* in *RF* from **[2]** maps a *BSF*-proof with k substitution inferences of a formula with m variables to an *RF*-proof with $k \cdot (m-1)$ substitution inferences. This is not a *pp*-simulation (not even an fpt-simulation).

By the results of the previous section there is an fpt-simulation of p-*SF* in p-*EF* even though (as mentioned at the end of that section) standard p-simulations of *SF* in *EF* are not fpt-simulations. We do not know whether p-$SF \leq_{\rm pp} p$-EF. However, this holds for the tree-like versions of these proof systems:

Theorem 4.4 *p-$SF^* \leq_{\rm pp} p$-EF^*.*

Proof. We say that an SF^*-proof of β from an extension sequence (for β) is an ESF^*-proof of β if every application of the substitution rule has the form

$$\frac{\alpha}{\alpha[x/\sigma]}$$

where the formula $x \wedge \sigma$ does not contain any extension variable.

Clearly, an EF^*-proof of β is an ESF^*-proof of β without applications of the substitution rules.

We now describe how to stepwise eliminate applications of the substitution rule in ESF^*-proofs. So, let π be an ESF^*-proof of β with k applications of the substitution rule. We depict π as a labeled tree T with β at the root; for any node t of T labeled by γ the subtree T_t rooted at this node (and consisting of the predecessors of this node) constitutes an ESF^*-proof of γ. Consider a node t such that

- t is labeled by a formula $\alpha[x/\sigma]$ obtained from its predecessor t^- labeled by α by an application of the substitution rule (via the substitution x/σ);
- no further applications of the substitution rule occur in T_t.

Let r be a variable not occuring in π and obtain $T_{t^-}(x/r)$ by substituting x by r in all formulas of T_{t^-}. By the proviso on the applications of the substitution rule in an ESF^*-proof, the variable x is not a substitution variable and hence extension axioms of T are transformed into extension axioms in $T_{t^-}(x/r)$. Hence, $T_{t^-}(x/r)$ is an F^*-proof of $\alpha[x/r]$ from a set of extension axioms.

Let π' be a short F^*-proof of

$$(\alpha[x/r] \to ((r \leftrightarrow \sigma) \to \underbrace{\alpha[x/r][r/\sigma]}_{=\alpha[x/\sigma]})).$$

Using the new extension axiom $(r \leftrightarrow \sigma)$ (and adding some applications of modus ponens) we merge this F^*-proof with $T_{t^-}(x/r)$ to get a F^*-proof of $\alpha[x/\sigma]$ from an extension sequence.

$$\dfrac{\dfrac{\dfrac{\vdots\ T_{t^-}(x/r)}{\alpha[x/r]} \qquad \dfrac{\vdots\ \pi'}{(\alpha[x/r] \to ((r \leftrightarrow \sigma) \to \alpha[x/\sigma])}}{((r \leftrightarrow \sigma) \to \alpha[x/\sigma])} \qquad (r \leftrightarrow \sigma)}{\alpha[x/\sigma]}$$

Replace in the original proof π the subtree $T_t(x/r)$ by this new proof, thus obtaining a proof π''. It should be clear that π'' is an ESF^*-proof of β with $k-1$ applications of the substitution rule.

Iterating this process k times we finally get an F^*-proof π^* of β from an extension sequence (for β) consisting of k extension axioms. As π^* is obtained from π in polynomial time the mapping $\pi \mapsto \pi^*$ is the desired *pp*-simulation of *p-SF** in *p-EF**. □

Note that in the previous proof we have used that the *SF*-proof we start with is treelike: the simulation replaces all predecessors of a formula obtained by a substitution rule. In an arbitrary *SF*-proof some later inferences may be based on some formulas not further available.

We prove the following result by standard means:

Proposition 4.5 *p-EF* $\leq_{\mathrm{pp}}$ *p-EF**.

Proof. Let $\pi = \alpha_1, \dots, \alpha_s$ be an *EF*-proof with k extension axioms. For $1 \leq i \leq s$ we set $\gamma_i := \bigwedge_{j=1}^{i} \alpha_j$. We construct for $i = 1, \dots, s$ successively *EF**-proofs π_i of γ_i such that the variables in π_i are precisely those in $\alpha_1, \dots, \alpha_i$ and the extension axioms in π_i are the same as in $\alpha_1, \dots, \alpha_i$.

The tree π_1 just consists of the root labeled by α_1. Assume that we have already constructed the *EF**-proof π_i of γ_i. To construct π_{i+1} we first consider the case where α_{i+1} is an extension axiom or a substitution instance of an axiom of F. Let π^1 be a short F^*-proof of $(u \to (v \to (u \wedge v)))$. Then $\pi^1[u/\gamma_i, v/\alpha_{i+1}]$ is an F^*-proof of $(\gamma_i \to (\alpha_{i+1} \to \gamma_{i+1}))$ of size $O(|\gamma_{i+1}|)$. As an intermediate step we get an F^*-proof π^2 of $(\alpha_{i+1} \to \gamma_{i+1})$ from the F^*-proofs π_i and $\pi^1[u/\gamma_i, v/\alpha_{i+1}]$ by an application of modus ponens. A further modus ponens inference yields from π^2 and the 'leaf' α_{i+1} the desired F^*-proof π_{i+1} of γ_{i+1}.

$$\cfrac{\cfrac{\begin{array}{c}\vdots\ \pi^1[u/\gamma_i, v/\alpha_{i+1}]\\ (\gamma_i \to (\alpha_{i+1} \to \gamma_{i+1}))\end{array} \qquad \begin{array}{c}\vdots\ \pi_i\\ \gamma_i\end{array}}{(\alpha_{i+1} \to \gamma_{i+1})} \qquad \alpha_{i+1}}{\gamma_{i+1}}$$

Now assume that α_{i+1} is obtained by one of the rules of F. The general case being analogous, we treat the case where this rule is modus ponens. So assume α_{i+1} is obtained from α_k and α_ℓ (where $1 \leq k, \ell \leq i$) by modus ponens. Let π^1 be an F^*-proof of $(\bigwedge_{j=1}^{i} u_j \to (u_k \wedge u_\ell))$ of size polynomial in i. Substituting in π^1 the u_js by the α_js yields an F^*-proof π^2 of $(\gamma_i \to (\alpha_k \wedge \alpha_\ell))$ of size polynomial in $|\gamma_i|$.

To a short F^*-proof of $((u \to v) \to ((v \to w) \to (u \to (u \wedge w))))$ we apply the substitution $[u/\gamma_i, v/(\alpha_k \wedge \alpha_\ell), w/\alpha_{i+1}]$ obtaining an F^*-proof π^3 of size $O(|\gamma_{i+1}|)$ of

$$((\gamma_i \to (\alpha_k \wedge \alpha_\ell)) \to ((\alpha_k \wedge \alpha_\ell) \to \alpha_{i+1}) \to (\gamma_i \to \gamma_{i+1}))).$$

Finally, let π^4 be an F^*-proof of $((\alpha_k \wedge \alpha_\ell) \to \alpha_{i+1})$ of size $O(|\alpha_k| + |\alpha_\ell| + |\alpha_{i+1}|)$ (recall that α_{i+1} was obtained from α_k and α_ℓ by modus ponens). Now it is easy to merge π_i, π^1, π^2, π^3, and π^4 to an F^*-proof π_{i+1} of γ_{i+1}.

It is easy to construct a treelike proof π^* of α_s from π_s. It is clear that π^* can be computed from π in polynomial time. □

Theorem 4.6 *$F \equiv_{\mathrm{pp}}$ p-BSF* $\equiv_{\mathrm{pp}}$ p-RF* $\leq_{\mathrm{pp}}$ p-EF $\equiv_{\mathrm{pp}}$ p-EF* $\equiv_{\mathrm{pp}}$ SF* $\leq_{\mathrm{pp}}$ p-SF $\equiv_{\mathrm{pp}}$ p-BSF.*

Proof. The first two equivalences are easy to see. The third equivalence follows from the preceding proposition. The equivalence *p-EF** $\equiv_{\mathrm{pp}}$ *p-SF** follows from Proposition 4.2 and Theorem 4.4. The last equivalence follows from Proposition 4.2, too. □

Hence, the proof systems mentioned in the previous theorem belong to at most three distinct *pp*-degrees. Are these degrees distinct? Note that this theorem does not mention *p-RF*. Does it belong to any of these degrees? Of course, $F \leq_{\mathrm{pp}}$ *p-RF* $\leq_{\mathrm{pp}}$ *p-SF*. Furthermore, we can show the following:

Proposition 4.7 *If p-RF* $\leq_{\mathrm{pp}}$ *p-EF, then p-SF* $\leq_{\mathrm{pp}}$ *p-EF.*

Proof. Assume *p-RF* $\leq_{\mathrm{pp}}$ *p-EF*. By Proposition 4.2 it suffices to show *p-BSF* $\leq_{\mathrm{pp}}$ *p-EF*. So let $\pi = \alpha_1, \dots, \alpha_s$ be a *BSF*-proof with k substitution inferences (substituting a variable by $\bot$ or by $\top$). Let $y_1, \dots, y_k$ and $z_1, \dots, z_k$ be new variables (not occurring in π) and let

$$\delta := \bigwedge_{i=1}^{k} \neg y_i \wedge \bigwedge_{i=1}^{k} z_i.$$

Consider the sequence

$$(\delta \to \alpha_1), \dots, (\delta \to \alpha_s).$$

This sequence can be "filled up" to an *RF*-proof with k substitution inferences (substituting a variable by another variable): if α_i in π is a substitution instance of an axiom, replace $(\delta \to \alpha_i)$ by a short F-proof of $(\delta \to \alpha_i)$. If α_i is obtained by modus ponens from $\alpha_j, \alpha_{j'}$ with $j, j' < i$, then replace $(\delta \to \alpha_i)$ by a short F-proof of $(\delta \to \alpha_i)$ from $(\delta \to \alpha_j)$ and $(\delta \to \alpha_{j'})$. Finally, if α_i is obtained by a substitution inference, then there is $j < i$ such that $\alpha_i = \alpha_j[x/\bot]$ or $\alpha_i = \alpha_j[x/\top]$ for some variable x. Assume this is the ℓth substitution inference $(1 \leq \ell \leq k)$ in π and that $\alpha_i = \alpha_j[x/\bot]$ (the other case $\alpha_i = \alpha_j[x/\top]$ is similar). Replace $(\delta \to \alpha_i)$ by the following *RF*-proof: give a short F-proof of $(\delta \wedge \alpha_j[x/y_\ell] \to \alpha_i)$ (note that $\neg y_\ell$ is a conjunct of δ) and derive $\alpha_j[x/y_\ell]$ from α_j by an *RF* substitution inference; from these two formulas it is easy to derive $(\delta \to \alpha_i)$.

Clearly, this *RF*-proof can be computed from π in polynomial time. By assumption we can in polynomial time compute from this *RF*-proof an *EF*-proof π' of $(\delta \to \alpha_s)$ with $k^{O(1)}$ extension axioms. Since the y_i's and the z_i's occur in δ, they are not used as extension variables in π'. Let π'' result from π' by substituting $\bot$ for all occurrences of the y_i's and $\top$ for all occurrences of the z_i's. Then (note the y_i's and the z_i's do not occur in α_s) π'' is an *EF*-proof of $(\delta' \to \alpha_s)$ where δ' is a true Boolean sentence (a true formula without variables). Adding a short proof of δ' and an application of modus ponens gives an *EF*-proof of α_s. □

Acknowledgements

The authors thank the John Templeton Foundation for its support through Grant #13152, *The Myriad Aspects of Infinity.* The second author thanks the FWF (Austrian Research Fund) for its support through Grant P 23989 - N13.

References

[1] O. Beyersdorff, N. Galesi, M. Lauria and A. Razborov. Parameterized bounded-depth Frege is not optimal. Proceedings of the 38th International Colloquium on Automata, Languages and Programming (ICALP), pp. 630–641, Springer-Verlag, 2011.

[2] S. Buss. Some remarks on the lengths of propositional proofs. *Archive for Mathematical Logic*, 34:377–394, 1995.

[3] S. Buss. Polynomial size proofs of the propositional pigeon principle. *The Journal of Symbolic Logic*, 52:916–927, 1987.

[4] S. Cook and P. Nguyen. *Logical Foundations of Proof Complexity.* Cambridge University Press, 2010.

[5] S. Cook and R. Reckhow. The relative efficiency of propositional proof systems. *The Journal of Symbolic Logic*, 44:36–50, 1979.

[6] S. S. Dantchev, B. Martin, and S. Szeider. Parameterized proof complexity. *Computational Complexity*, 20(1):51–85, 2011.

[7] R. G. Downey and M. R. Fellows. *Parameterized Complexity.* Springer-Verlag, 1999.

[8] J. Flum and M. Grohe. *Parameterized Complexity Theory.* Springer-Verlag, 2006.

[9] L. Fortnow and R. Santhanam. Infeasibility of instance compression and succinct PCPs for NP. *Journal of Computer and System Sciences*, 77(1):91–106, 2011.

[10] J. Krajíček. On the number of steps in proofs. *Annals of Pure and Applied Logic*, 41:153–178, 1989.

[11] J. Krajíček. *Bounded arithmetic, propositional logic, and complexity theory.* Cambridge University Press, 1995.

[12] J. Krajíček and P. Pudlák. Propositional proof systems, the consistency of first order theories and the complexity of computations. *The Journal of Symbolic Logic*, 54:1063–1088, 1989.

On the structure of honest elementary degrees

Lars Kristiansen[*], **Robert S. Lubarsky**[†], **Jan-Christoph Schlage-Puchta**[‡], **Andreas Weiermann**[‡]

[*] Department of Mathematics, Universitetet i Oslo, Norway
`larsk@math.uio.no`

[†] Department of Mathematical Sciences, Florida Atlantic University, USA
`Robert.Lubarsky@alum.mit.edu`

[‡] Department of Mathematics, Universiteit Gent, Belgium
`jcsp@cage.ugent.be, weiermann@cage.ugent.be`

Abstract. We present some new results, and survey old results, on the structure of honest elementary degrees. This paper should be a suitable first introduction to the honest elementary degrees.

Introduction

This paper is devoted to the study of the structure of the honest elementary degrees. We present some new results, but this is also a kind of introduction and survey paper. The new material is found in Sections 6 and 7. In the remaining sections, we survey the same material as we do in Part I of [**9**], but we give more detailed proofs and more elaborated explanations. This should be the most thorough and readable introduction to the honest elementary degrees available so far. But be aware that we are talking about a technical introduction, and it is beyond the scope of this paper to motivate our study of the honest elementary degrees.

The roots of our subject can be found in subrecursion theory from the 1970s. Some relevant papers are Meyer & Ritchie [**13**] and Machtey [**10, 11, 12**]. The theory of honest elementary degrees, in the form presented here, was developed by Kristiansen in a series of papers (and a thesis) [**4, 6, 7, 8**] ([**5**]) from the 1990s. A considerable number of the results surveyed in Sections 2, 3, 4 and 5 was initially published in these papers.

A recent paper by Kristiansen, Schlage-Puchta and Weiermann [**9**] shows how to generalise honest elementary degree theory to so-called honest α-elementary degree theory. This generalisation connects honest degree theory with proof theory and provability of Π^0_2-statements in formal systems for mathematics, e.g. Peano Arithmetic. Such a connection yields a strong motivation for further research in honest degree theory.

1 Preliminaries

We assume that the reader is familiar with the most basic concepts of classical computability theory; see e.g. [**14**] or [**16**]. We also assume acquaintance with subrecursion

The first, second and fourth author gratefully acknowledge partial support by grants from the John Templeton Foundation —grant no. 13396 and grant no. 13152.

theory and, in particular, with the elementary functions. An introduction to this subject can be found in [**15**] or [**17**]. Here we just state some important basic facts and definitions; see [**15**] and [**17**] for proofs.

The *initial elementary functions* are the projection functions ($\mathcal{I}^n_i$), the constants 0 and 1, addition (+) and modified subtraction ($\dot{-}$). The *elementary definition schemes* are *composition*, that is, $f(\vec{x}) = h(g_1(\vec{x}), \ldots, g_m(\vec{x}))$ and *bounded sum* and *bounded product*, that is, respectively $f(\vec{x}, y) = \sum_{i<y} g(\vec{x}, i)$ and $f(\vec{x}, y) = \prod_{i<y} g(\vec{x}, i)$. A *function is elementary* if it can be generated from the initial elementary functions by the elementary definition schemes. A *relation* $R(\vec{x})$ *is elementary* when there exists an elementary function f with range $\{0, 1\}$ such that $f(\vec{x}) = 0$ iff $R(\vec{x})$ holds. Relations may also be called *predicates*, and we will use the two words interchangeably. A function f has *elementary graph* if the relation $f(\vec{x}) = y$ is elementary. When we can define a function g from the function f plus the initial elementary functions by the elementary schemes, we will say that g *is elementary in* f.

The definition scheme $(\mu z \le x)[\ldots]$ is called the *bounded μ-operator*, and

$$(\mu z \le y)[R(\vec{x}, z)]$$

denotes the least $z \le y$ such that the relation $R(\vec{x}, z)$ holds. Let $(\mu z \le y)[R(\vec{x}, z)] = 0$ if no such z exists. The elementary functions are closed under the bounded μ-operator. If f is defined by a primitive recursion over g and h and $f(\vec{x}, y) \le j(\vec{x}, y)$, then f is defined by *bounded primitive recursion* over g, h and j. The elementary functions are closed under bounded primitive recursion, but not under primitive recursion. Moreover, the elementary relations are closed under the operations of the propositional calculus and under bounded quantification, i.e., $(\forall x \le y)[R(x)]$ and $(\exists x \le y)[R(x)]$.

Let $2^x_0 = x$ and $2^x_{n+1} = 2^{2^x_n}$, and let $\mathcal{S}$ denote the successor function. The class of elementary functions equals the closure of $\{0, \mathcal{S}, \mathcal{I}^n_i, 2^x, \max\}$ under composition and bounded primitive recursion. Given this characterisation of the elementary functions, it is easy to see that for any elementary function f, we have $f(\vec{x}) \le 2^{\max(\vec{x})}_k$ for some fixed k. It is also easy to see that the class of functions elementary in f is the closure of $\{0, \mathcal{S}, \mathcal{I}^n_i, 2^x, \max, f\}$ under composition and bounded primitive recursion. As remarked above, the elementary functions are not closed under primitive recursion, but the elementary predicates will be closed under (unbounded) primitive recursion, that is, when a predicate $P(\vec{x}, y)$ is defined by $P(\vec{x}, 0) \Leftrightarrow \phi(\vec{x})$ and $P(\vec{x}, y+1) \Leftrightarrow \psi(\vec{x}, P(\vec{x}, y), y)$, then P will be elementary if ϕ and ψ are elementary.

Uniform systems for coding finite sequences of natural numbers are available inside the class of elementary functions. Let $\overline{f}(x)$ be the code number for the sequence $\langle f(0), f(1), \ldots, f(x)\rangle$. Then $\overline{f}$ belongs to the elementary functions if f does. We will be quite informal and indicate the use of coding functions with the notations $\langle \ldots \rangle$ and $(x)_i$ where $(\langle x_0, \ldots, x_i, \ldots, x_n\rangle)_i = x_i$. (So $(x, i) \mapsto (x)_i$ is an elementary function.) Our coding system is monotone, i.e., $\langle x_0, \ldots, x_n\rangle < \langle x_0, \ldots, x_n, y\rangle$ holds for any y, and $\langle x_0, \ldots, x_i, \ldots, x_n\rangle < \langle x_0, \ldots, x_i{+}1, \ldots, x_n\rangle$. All the closure properties of the elementary functions can be proved by using Gödel numbering and coding techniques.

For unary functions f, g, we use $f \le g$ to denote $\forall x \in \mathbb{N}[f(x) \le g(x)]$, and we use f^k to denote the k-th iterate of the function f, that is, $f^0(x) = x$ and $f^{k+1}(x) = ff^k(x)$.

2 Honest elementary degrees and the growth theorem

Definition 2.1 A function $f\colon \mathbb{N} \to \mathbb{N}$ is *honest* if it is monotone ($f(x) \leq f(x+1)$), dominates 2^x ($f(x) \geq 2^x$) and has elementary graph.

Note that, when f is honest, we have $f^{y+1}(x) > f^y(x)$, but we do not necessarily have $f(x+y) > f(x)$. From now on, we reserve the letters $f, g, h, \ldots$ to denote honest functions. Small Greek letters like $\phi, \psi, \xi, \ldots$ will denote number-theoretic functions not necessarily being honest.

Definition 2.2 A function ϕ is *elementary* in a function ψ, written $\phi \leq_E \psi$, if ϕ can be generated from the initial functions ψ, 2^x, max, 0, $\mathcal{S}$ (successor), $\mathcal{I}_i^n$ (projections) by composition and bounded primitive recursion.

We define the relation $\equiv_E$ by $f \equiv_E g \Leftrightarrow f \leq_E g \wedge g \leq_E f$. Now, $\equiv_E$ is an equivalence relation on the honest functions, and we will use $\mathcal{H}$ denote the set of $\equiv_E$-equivalence classes of honest functions. The elements of $\mathcal{H}$ are the *honest elementary degrees.* Honest elementary degrees will normally just be called *degrees*, and following the tradition of classical computability theory, we use boldface lowercase Latin letters $\mathbf{a}, \mathbf{b}, \mathbf{c}, \ldots$ to denote our degrees.

We will use $\deg(f)$ to denote the degree of the honest function f, that is,

$$\deg(f) = \{g \mid g \equiv_E f\}.$$

We define the relation $<_E$ by $f <_E g \Leftrightarrow f \leq_E g \wedge g \not\leq_E f$; and the relation $|_E$ by $f \mid_E g \Leftrightarrow f \not\leq_E g \wedge g \not\leq_E f$. We will use $<$, $\leq$, $|$ to denote the relations induced on the degrees by $<_E, \leq_E, |_E$ respectively. We use standard, and presumably very familiar, language with respect to these ordering relations, and we will, e.g., say that f *lies below* g if $f \leq_E g$; that g *is strictly above* f if $f <_E g$; that $\mathbf{c}$ *lies strictly between* $\mathbf{a}$ *and* $\mathbf{b}$ if $\mathbf{a} < \mathbf{c} < \mathbf{b}$; that $\mathbf{a}$ and $\mathbf{b}$ *are incomparable* if $\mathbf{a} \mid \mathbf{b}$; and so on.

Theorem 2.3 (Growth Theorem) *Let f and g be honest functions. Then,*

$$g \leq_E f \iff g \leq f^k \text{ for some fixed } k.$$

Proof. Recall that f is monotone and dominates 2^x. By induction on the build-up of a function ψ, form the initial functions 0, $\mathcal{S}$, $\mathcal{I}_i^n$, 2^x, max, f by composition and bounded primitive recursion, it is easy to prove that there exists $k \in \mathbb{N}$ such that $\psi(\vec{x}) \leq f^k(\max(\vec{x}))$. Hence, if $g \leq_E f$, we have $g \leq f^k$ for some fixed k.

Now, suppose that $g \leq f^k$. Since g is honest, the relation $g(x) = y$ is elementary. We have $g(x) = (\mu y \leq f^k(x))[g(x) = y]$. Hence $g \leq_E f$, since the functions elementary in f are closed under composition and the bounded μ-operator. □

The structure of honest elementary degrees is comparable to a classical computability-theoretic degree structure, e.g., the structure of Turing degrees, but the Growth Theorem makes it possible to abandon classical computability-theoretic proof methods and investigate this structure by asymptotic analysis and methods of number theoretic nature. To prove that $g \leq_E f$, it is sufficient to provide a fixed k such that $g(x) \leq f^k(x)$; to prove that $g \not\leq_E f$, it is sufficient to prove that such a k does not exist. Thus, there is no need[1] for the standard computability-theoretic machinery involving enumerations, diagonalisations and constructions with requirements to be satisfied. This makes the proofs concise and transparent.

[1] Well, at least we can achieve a lot without resorting to such a machinery; see Section 6.

3 The lattice of honest elementary degrees

Definition 3.1 *Least upper bounds* and *greatest lower bounds* are defined the usual way, and a partially ordered structure where each pair of elements has both a least upper bound and a greatest lower bound is called *a lattice.*

We define the *join* of the honest functions f and g, written $\max[f,g]$, by

$$\max[f,g](x) = \max(f(x), g(x)).$$

We define the *meet* of the honest functions f and g, written $\min[f,g]$, by

$$\min[f,g](x) = \min(f(x), g(x)).$$

Lemma 3.2 *Let f and g be honest functions. Then,* $\max[f,g]$ *and* $\min[f,g]$ *are honest functions.*

Proof. It is trivial that $\max[f,g]$ and $\min[f,g]$ are monotone and dominate 2^x. To verify that $\max[f,g]$ and $\min[f,g]$ have elementary graphs, observe that $\max[f,g](x) = y$ holds iff

$$(f(x) = y \ \wedge\ (\exists i \leq y)[g(x) = i]) \ \vee\ (g(x) = y \ \wedge\ (\exists i < y)[f(x) = i])$$

and that $\min[f,g](x) = y$ holds iff

$$(f(x) = y \ \wedge\ (\forall i \leq y)[g(x) \neq i]) \ \vee\ (g(x) = y \ \wedge\ (\forall i < y)[f(x) \neq i]).$$

The relations $f(x) = y$ and $g(x) = y$ are elementary. Furthermore, the elementary relations are closed under bounded quantification and the operations of the propositional calculus. Hence, both $\max[f,g](x) = y$ and $\min[f,g](x) = y$ are elementary relations. □

Lemma 3.3 *Let f and g be honest functions. Then, we have*

$$\min(f^m(x), g^n(x)) \leq \min[f,g]^{m+n}(x).$$

Proof. We prove this lemma by induction on $m + n$. The lemma holds trivially when $m = 0$ or $n = 0$. Now, assume that $m > 0$ and $n > 0$. Then, without loss of generality, we may assume that $\min[f,g](x) = f(x)$. Together with the induction hypothesis this yields

$$\begin{aligned}\min(f^m(x), g^n(x)) &\leq \min(f^{m-1}(f(x)), g^n(f(x))) \\ &\leq \min[f,g]^{m-1+n}(f(x)) = \min[f,g]^{m+n}(x).\end{aligned}$$

□

Lemma 3.4 *Let f, g, h be honest functions.*

(i) $\min[f,g] \leq_E f$ *and* $\min[f,g] \leq_E g$.

(ii) *If $h \leq_E f$ and $h \leq_E g$, then $h \leq_E$* $\min[f,g]$.

Proof. We prove (ii). Assume $h \leq_E f$ and $h \leq_E g$. By the Growth Theorem we have m, n such that $h(x) \leq f^m(x)$ and $h(x) \leq g^n(x)$. By Lemma 3.3, we have

$$h(x) \leq \min(f^m(x), g^n(x)) \leq \min[f,g]^{n+m}(x).$$

By another application of the Growth Theorem, we have $h \leq_E \min[f,g]$. This proves (i). The proof of (ii) is straightforward by the Growth Theorem. □

Lemma 3.5 *Let f, g, h be honest functions.*

(i) $f \leq_E \max[f,g]$ *and* $g \leq_E \max[f,g]$.

(ii) *If $f \leq_E h$ and $g \leq_E h$, then* $\max[f,g] \leq_E h$.

Proof. Both (i) and (ii) follow straightforwardly from the Growth Theorem. □

Lemma 3.6 *For any honest functions f, f_1, g, g_1 such that $f \leq_E f_1$ and $g \leq_E g_1$,*

(i) $\min[f,g] \leq_E \min[f_1,g_1]$;
(ii) $\max[f,g] \leq_E \max[f_1,g_1]$.

Proof. Now $\leq_E$ is transitive, and thus (i) follows immediately from Lemma 3.4, and (ii) follows immediately from Lemma 3.5. □

Our previous lemma entails that

$$(f \equiv_E f_1 \;\wedge\; g \equiv_E g_1) \;\Rightarrow\; (\max[f,g] \equiv_E \max[f_1,g_1] \;\wedge\; \min[f,g] \equiv_\alpha \min[f_1,g_1])$$

when f, f_1, g, g_1 are honest functions. By Lemma 3.2, we know that $\max[f,g]$ and $\min[f,g]$ are honest functions whenever f and g are. Hence, the next definition makes sense.

Definition 3.7 Let f and g be honest functions such that $\deg(f) = \mathbf{a}$ and $\deg(g) = \mathbf{b}$. We define the *join of* $\mathbf{a}$ *and* $\mathbf{b}$, written $\mathbf{a} \cup \mathbf{b}$, by $\mathbf{a} \cup \mathbf{b} = \deg(\max[f,g])$. We define the *meet of* $\mathbf{a}$ *and* $\mathbf{b}$, written $\mathbf{a} \cap \mathbf{b}$, by $\mathbf{a} \cap \mathbf{b} = \deg(\min[f,g])$.

Theorem 3.8 (Distributive Lattice) *The structure $\langle \mathcal{H}, \leq, \cup, \cap \rangle$ is a distributive lattice, that is, for any $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathcal{H}$, we have*

(i) $\mathbf{a} \cap \mathbf{b}$ *is the greatest lower bound of* $\mathbf{a}$ *and* $\mathbf{b}$ *under the ordering* $\leq$;
(ii) $\mathbf{a} \cup \mathbf{b}$ *is the least upper bound of* $\mathbf{a}$ *and* $\mathbf{b}$ *under the ordering* $\leq$;
(iii) $\mathbf{a} \cup (\mathbf{b} \cap \mathbf{c}) = (\mathbf{a} \cup \mathbf{b}) \cap (\mathbf{a} \cup \mathbf{c})$ *and* $\mathbf{a} \cap (\mathbf{b} \cup \mathbf{c}) = (\mathbf{a} \cap \mathbf{b}) \cup (\mathbf{a} \cap \mathbf{c})$.

Proof. It follows from Lemma 3.4 (i) that $\mathbf{a} \cap \mathbf{b}$ is a lower bound of $\mathbf{a}$ and $\mathbf{b}$, and by Lemma 3.4 (ii), $\mathbf{a} \cap \mathbf{b}$ is indeed the greatest lower bound of $\mathbf{a}$ and $\mathbf{b}$. This proves (i).

The proof of (ii) is symmetric, using Lemma 3.5 in place of Lemma 3.4. Finally, (iii) holds since $\max(x, \min(y,z)) = \min(\max(x,y), \max(x,z))$ and $\min(x, \max(y,z)) = \max(\min(x,y), \min(x,z))$. □

Let $\mathbf{a}$ and $\mathbf{b}$ be two degrees such that $\mathbf{a} \leq \mathbf{b}$. Now, we do not necessarily have $f \leq g$ for any $f \in \mathbf{a}$ and $g \in \mathbf{b}$. But there will always be some $f \in \mathbf{a}$ and some $g \in \mathbf{b}$ such that we have $f(x) \leq g(x)$, or even $f(x) < g(x)$, for all x. This is consequence of the lemmas above: Pick an arbitrary $f_1 \in \mathbf{a}$ and an arbitrary $g_1 \in \mathbf{b}$, and let $f = \min[f_1, g_1]$ and $g = \max[f_1, g_1]$. Now we obviously have and $f(x) \leq g(x) < g^2(x)$ for all x, but we also have $f \in \mathbf{a}$ and $g, g^2 \in \mathbf{b}$.

Theorem 3.9 (Density-Splitting) *Let* $\mathbf{a}$ *and* $\mathbf{b}$ *be degrees such that* $\mathbf{a} < \mathbf{b}$. *Then, there exist incomparable degrees* $\mathbf{c}_0$ *and* $\mathbf{c}_1$ *such that* $\mathbf{a} = \mathbf{c}_0 \cap \mathbf{c}_1$ *and* $\mathbf{b} = \mathbf{c}_0 \cup \mathbf{c}_1$.

Proof. Pick honest functions f and g such that $\deg(g) = \mathbf{a} < \mathbf{b} = \deg(f)$ and $g(x) < f(x)$. We define the sequence $d_0 < d_1 < d_2 < \ldots$. Let $d_0 = 0$, let d_{2i+1} be the least y such that

$$(\exists z \leq y)\,[\, f(z) \leq y \;\wedge\; (\exists w \leq z)\,[\, d_{2i} \leq w \;\wedge\; g^i(w) < f(w)\,]\,] \tag{3.1}$$

and let $d_{2i+2} = f(d_{2i+1})$. Next we define the functions h_0 and h_1. For $\jmath \in \{0,1\}$, let $h_\jmath(x) = \max(H_\jmath(x), g(x))$, where

$$H_\jmath(x) = \begin{cases} f(x) & \text{if } d_{4i+2\jmath} \leq x \leq d_{4i+2\jmath+1} \text{ for some } i, \\ H_\jmath(x-1) & \text{otherwise.} \end{cases}$$

Since $f \not\leq_E g$, there will for each i exist infinitely many z such that $g^i(z) < f(z)$. Thus, there will always be a number satisfying the definition of d_{2i+1}, and thus the sequence $d_0 < d_1 < d_2 < \dots$ is well defined.

We will now prove that h_1 and h_2 are honest functions. First, we will argue that the relation $d_i = y$ is elementary. This is not obvious as a relation like $g^i(w) < f(w)$ is not necessarily elementary even if f and g are honest functions. However, the relation $g^i(w) < f(w) \leq y$ will be elementary (in i, w and y) whenever g and f are honest. Now, the statement (3.1) involved in the definition of $d_i = y$ is equivalent to

$$(\exists z \leq y)\,[\,f(z) \leq y \ \wedge\ (\exists w \leq z)\,[\,d_{2i} \leq w \ \wedge\ g^i(w) < f(w) \leq y\,]\,].$$

Moreover, the elementary relations are closed under primitive recursion, bounded quantifiers and propositional operations. Thus, $d_i = y$ is indeed an elementary relation. When we know that $d_i = y$ is elementary, it becomes easy to see that h_0 and h_1 have elementary graphs. Furthermore, it is obvious that h_0 and h_1 are monotone and dominate 2^x, and thus, we are dealing with two honest functions.

Next, we will prove that $\min[h_0, h_1] \equiv_E g$, that $\max[h_0, h_1] \equiv_E f$, and that $h_0 \mid_E h_1$. The theorem follows.

We start by proving $\min[h_0, h_1] \equiv_E g$. By the Growth Theorem it suffices to prove that $\min[h_0, h_1](x) = g(x)$. Assume we have $d_{4i+2} \leq x < d_{4i+4}$. Then

$$\begin{aligned}
h_0(x) &= \max(H_0(x), g(x)) && \text{def. of } h_0\\
&= \max(H_0(d_{4i+1}), g(x)) && \text{def. of } H_0\\
&= \max(f(d_{4i+1}), g(x)) && \text{def. of } H_0\\
&= \max(d_{4i+2}, g(x)) && \text{def. of } d_{4i+2}\\
&= \max(x, g(x)) && \text{as } d_{4i+2} \leq x\\
&= g(x) && \text{as } g(x) \geq 2^x.
\end{aligned}$$

A symmetric argument shows that $h_1(x) = g(x)$ when there exists i such that $d_{4i} \leq x < d_{4i+2}$. Hence, for any x, we either have $h_0(x) = g(x)$ or $h_1(x) = g(x)$, and since $\min[h_0, h_1](x) \geq g(x)$, we can conclude that $\min[h_0, h_1](x) = g(x)$. This proves that $\min[h_0, h_1] \equiv_E g$.

Our next task is to prove that $\max[h_0, h_1] \equiv_E f$. It follows straightaway from our definitions that we have $\max[h_0, h_1](x) \leq f(x)$. We will prove that $f(x) \leq \max[h_0, h_1]^2(x)$, and thus, we have $\max[h_0, h_1] \equiv_E f$ by the Growth Theorem. The proof of $f(x) \leq \max[h_0, h_1]^2(x)$ splits into two cases. *Case* (i): When x is in the interval $d_{2i} \dots d_{2i+1} - 1$ for some i, we have $f(x) \leq \max[h_0, h_1]^2(x)$ as either h_0 or h_1 will equal f in this interval. *Case* (ii): Assume x is in the interval $d_{2i+1} \dots d_{2i+2} - 1$ for some i, and note that

$$h_0(d_j) = f(d_j) \text{ or } h_1(d_j) = f(d_j) \tag{3.2}$$

holds for any j. We have

$$\begin{aligned}
f(x) &\leq f(d_{2i+2}) && f \text{ is monotone}\\
&= \max[h_0, h_1](d_{2i+2}) && (3.2)\\
&= \max[h_0, h_1](f(d_{2i+1})) && \text{def. of } d_{2i+2}\\
&= \max[h_0, h_1]^2(d_{2i+1}) && (3.2)\\
&\leq \max[h_0, h_1]^2(x) && \max[h_0, h_1] \text{ is monotone.}
\end{aligned}$$

This completes the proof of $\max[h_0, h_1] \equiv_E f$.

Finally, we prove $h_0 \mid_E h_1$. Fix an arbitrary $m \in \mathbb{N}$. We will argue that there exists x such that $h_0^m(x) < h_1(x)$. Let $k \geq 2m$. By the definition of d_{4k+3} there exists a number x_k in the interval $d_{4k+2}, \dots, d_{4k+3}$ such that

$$d_{4k+2} \leq g^m(x_k) \leq g^k(x_k) < f(x_k) \leq d_{4k+3}. \tag{3.3}$$

Now, since $d_{4k+2} \leq x_k \leq g^k(x_k) < d_{4k+3}$, it follows from the definitions of h_0 and H_0 that

$$\begin{aligned} h_0(g^\ell(x_k)) &= \max(H_0(g^\ell(x_k)), gg^\ell(x_k)) = \max(H_0(d_{4k+1}), g^{\ell+1}(x_k)) \\ &= \max(d_{4k+2}, g^{\ell+1}(x_k)) = \max(x_k, g^{\ell+1}(x_k)) = g^{\ell+1}(x_k) \end{aligned} \tag{3.4}$$

holds for any $\ell < k$. When we combine (3.3), (3.4) and the definition of h_1, we get $h_0^m(x_k) = g^m(x_k) \leq g^k(x_k) < f(x_k) = h_1(x_k)$. This proves that, for any m, we can find x such that $h_0^m(x) < h_1(x)$. By the Growth Theorem, we have $h_1 \not\leq_E h_0$. The proof that $h_0 \not\leq_E h_1$ is symmetric. Hence, $h_0 \mid_E h_1$. □

Results being obviously equivalent to Theorem 3.8 and Theorem 3.9 are proved by Machtey [**11, 12**] by traditional computability-theoretic methods.

4 A jump operator on honest elementary degrees

We will now define an operator $(\cdot)'$ transforming an honest function f into a faster increasing honest function f'. This operator will be called the *jump operator*.

Definition 4.1 For any honest function f, we define *the jump of f*, written f', by $f'(x) = f^{x+1}(x)$.

Lemma 4.2 *Let f be an honest function. Then, f' is an honest function.*

Proof. It is obvious that f' is monotone and dominates 2^x. Let $\psi(x, y)$ be an elementary function that places a bound on the code number for the sequence $\langle y, y, \dots, y\rangle$ of length $x + 1$. Then, $f'(x) = y$ is equivalent to

$$(\exists s \leq \psi(x, y))[(s)_0 = f(x) \ \wedge \ (\forall i < x)[(s)_{i+1} = f((s)_i)] \ \wedge \ (s)_x = y].$$

Thus, the relation $f'(x) = y$ is elementary since all the functions, relations and operations involved in this expression are elementary. This proves that f' has elementary graph. □

Lemma 4.3 (Monotonicity of the Jump Operator) *Let f and g be honest functions. Then, we have*

$$g \leq_E f \ \Rightarrow \ g' \leq_E f'.$$

Proof. Suppose that $g \leq_E f$. By the Growth Theorem, we have a fixed k such that $g(x) \leq f^k(x)$. Now

$$g'(x) = g^{x+1}(x) \leq (f^k)^{x+1}(x) \leq f^{(kx+k)+1}(kx + k) = f'(kx + k) \leq (f')^{2k}(x)$$

and $g' \leq_E f'$ follows by another application of the Growth Theorem. □

Lemma 4.3 entails that $f' \equiv_E g'$ whenever f and g are honest functions such that $f \equiv_E g$. Hence, the jump operator on the honest functions induce an operator on the honest elementary degrees.

Definition 4.4 For any honest elementary degree $\mathbf{a}$, we define *the jump of* $\mathbf{a}$, written $\mathbf{a}'$, by $\mathbf{a}' = \deg(f')$ where f is some honest function such that $\mathbf{a} = \deg(f)$. Furthermore, we define the *zero degree*, written $\mathbf{0}$, by $\mathbf{0} = \deg(2^x)$.

The proof of the next theorem is straightforward. See Kristiansen [**6**] for the details.

Theorem 4.5 (Canonical Degrees) *We have* $\mathbf{0} < \mathbf{0}' < \mathbf{0}'' < \dots$. *Furthermore,* $\mathbf{0}$ *is the least degree, that is,* $\mathbf{0} \leq \mathbf{a}$ *holds for any degree* $\mathbf{a}$.

The jump operators of classical computability theory are defined by enumerating all the functions reducible to an oracle function f, e.g., the Turing jump $\mathcal{J}(f)$ of the function f is defined by $\mathcal{J}(f)(\langle e, x\rangle) = \{e\}^f(x)$ where $\{e\}^f$ denotes the e-th function Turing computable in f and $\langle \cdot, \cdot \rangle$ is a computable bijection from $\mathbb{N} \times \mathbb{N}$ into $\mathbb{N}$. Jump operators based on enumerations are considered to be natural. The reader should note that our jump operator is equivalent to such a natural jump operator of classical computability theory: Let $\{[i]^f\}_{i\in\mathbb{N}}$ be an elementary enumeration of the functions elementary in the honest functions f, and let $\mathcal{J}(f)(\langle e, x\rangle) = [e]^f(x)$ where $\langle \cdot, \cdot \rangle$ is an elementary bijection from $\mathbb{N} \times \mathbb{N}$ into $\mathbb{N}$. Then, we indeed have $f' \equiv_E \mathcal{J}(f)$. For a proof and further details, see [**6**] and [**5**].

However, in our context, the advantage of defining f' as an iteration of f is obvious: The Growth Theorem is very well suited for dealing with a jump operator based on iterations; we can introduce the canonical degrees $\mathbf{0}, \mathbf{0}', \dots,$ and proceed to develop our degree theory without resorting to enumerations and the apparatus of classical computability theory.

Definition 4.6 We define the *n-th jump of an honest degree* $\mathbf{a}$ *(function f)*, written $\mathbf{a}^{[n]}$ ($f^{[n]}$), by $\mathbf{a}^{[0]} = \mathbf{a}$ and $\mathbf{a}^{[n+1]} = \mathbf{a}^{[n]\prime}$ ($f^{[0]} = f$ and $f^{[n+1]} = f^{[n]\prime}$). A degree $\mathbf{a}$ strictly below $\mathbf{0}'$ is *low*$_n$ if $\mathbf{a}^{[n]} = \mathbf{0}^{[n]}$, and *high*$_n$ if $\mathbf{a}^{[n]} = \mathbf{0}^{[n+1]}$.

Our strategy for proving the existence of low$_n$ and high$_n$ degrees, will be as follows: First we provide degrees $\mathbf{a}_\ell$ and $\mathbf{a}_h$ strictly between $\mathbf{0}^{[n]}$ and $\mathbf{0}^{[n+1]}$ such that $\mathbf{a}'_\ell = \mathbf{0}^{[n+1]}$ $\mathbf{a}'_h = \mathbf{0}^{[n+2]}$. Thereafter we prove that for any degree $\mathbf{b}$ strictly between $\mathbf{0}^{[k+1]}$ and $\mathbf{0}^{[k+2]}$, we can find a degree $\mathbf{c}$ strictly between $\mathbf{0}^{[k]}$ and $\mathbf{0}^{[k+1]}$ such that $\mathbf{c}' = \mathbf{b}$.

Theorem 4.7 *Let f be a strictly monotone and honest function. Then, there exists an honest function g such that $f <_E g$ and $g' \equiv_E f'$.*

Proof. Let $g(x) = f'f(f')^{-1}(x)$ where $(f')^{-1}$ denotes the inverse of f' given by

$$(f')^{-1}(x) = (\mu i)[f'(i) \geq x].$$

Since f' is strictly monotone, we have $(f')^{-1}f'(x) = x$ and $f'(f')^{-1}(x) \geq x$. Furthermore, we have $g(x) = y$ iff

$$(\exists u, v < y)\,[\,(\forall w < u)[f'(w) < x] \;\wedge\; f'(u) \geq x \;\wedge\; f(u) = v \;\wedge\; f'(v) = y\,]$$

and thus it is easy to see that the graph of g is elementary. It is also easy to see that g is monotone and dominates 2^x. Hence, g is an honest function.

Now, $f(x) \leq ff'(f')^{-1}(x) \leq f'f(f')^{-1}(x) = g(x)$, and for any fixed k and sufficiently large x, we have

$$\begin{aligned}
f^k(x) &\leq f^k f'(f')^{-1}(x) \\
&= f^k f^{(f')^{-1}(x)+1}((f')^{-1}(x)) && \text{def. of } f' \\
&\leq f^{k+(f')^{-1}(x)+1}(k + (f')^{-1}(x)) \\
&= f'(k + (f')^{-1}(x)) && \text{def. of } f' \\
&< f'(f(f')^{-1}(x)) && f(x) \geq 2^x \text{ and } x \text{ is large} \\
&= g(x) && \text{def. of } g.
\end{aligned}$$

Hence, we have $f <_E g$ by the Growth Theorem.

Next, we observe that $g^k(x) = f'g^k(f')^{-1}(x)$ for any $k > 0$. This is trivially true when $k = 1$, and, by an induction hypothesis, we have

$$g^{k+1}(x) = gg^k(x) = gf'f^k(f')^{-1}(x) = f'f(f')^{-1}f'f^k(f')^{-1}(x) = f'f^{k+1}(f')^{-1}(x).$$

Thereby, $g'(x) = g^{x+1}(x) = f'f^{x+1}(f')^{-1}(x) \leq f'f^{x+1}(x) = f'f'(x)$, and then we have $g' \leq_E f'$ by the Growth Theorem. Since $f <_E g$, we also have $g' \equiv_E f'$ by the monotonicity of the jump operator. □

Theorem 4.8 *Let f be an honest function. Then, there exists an honest function g such that $g <_E f'$ and $g' \equiv_E f''$.*

Proof. For any $i \in \mathbb{N}$, let $d_{3i+1} = f''(d_{3i})$, let $d_{3i+2} = f'(d_{3i+1})$, and let $d_{3i+3} = f'(d_{3i+2})$. Let $d_0 = 0$. Furthermore, let

$$G(x) = \begin{cases} f'(x) & \text{if } d_{3i} \leq x \leq d_{3i+1} \text{ for some } i, \\ G(x-1) & \text{otherwise,} \end{cases}$$

and let $g(x) = \max(G(x), f(x))$. It is easy to check that g is honest.

First we prove that $f'' \equiv_E g'$. Observe that for any $j \leq d_{3i+1} + 1$, we have $d_{3i} \leq (f')^j(d_{3i}) \leq (f')^{d_{3i}+1}(d_{3i}) = f''(d_{3i}) = d_{3i+1}$. Hence, by the definition of g, we have

$$f''(d_{3i}) = (f')^{d_{3i}+1}(d_{3i}) = g^{d_{3i}+1}(d_{3i}) = g'(d_{3i}) \tag{4.1}$$

for any $i \in \mathbb{N}$. Now, let x be arbitrary and let i be the unique number such that $d_{3i} \leq x < d_{3i+3}$. Then

$$\begin{aligned}
(g')^4(x) &\geq (g')^4(d_{3i}) && \text{as } g' \text{ is monotone} \\
&= (g')^3 f''(d_{3i}) && (4.1) \\
&= (g')^3(d_{3i+1}) && \text{def. of } d_{3i+1} \\
&\geq (g')(f')^2(d_{3i+1}) && \text{as } f(x) \leq g(x) \\
&\geq (g')(d_{3i+3}) && \text{def. of } d_{3i+3} \\
&= f''(d_{3i+3}) && (4.1) \\
&\geq f''(x) && \text{as } f'' \text{ is monotone.}
\end{aligned}$$

This proves $f'' \leq (g')^4$, and $f'' \leq_E g'$ follows by the Growth Theorem. Moreover, since $g \leq f'$, we have $g \leq_E f'$, and thus also $g' \leq_E f''$ by the monotonicity of the jump operator. This proves that $f'' \equiv_E g'$.

Next we prove that $g <_E f'$. It is obvious that $g \leq_E f'$ since $g(x) \leq f'(x)$. Hence, we are left to prove that $f' \not\leq_E g$. Assume $d_{3i+2} \leq x < d_{3i+3}$. Then, straightaway from the definition of g and the sequence $\{d_j\}_{j\in\mathbb{N}}$, we have

$$\begin{aligned} g(x) &= \max(G(x), f(x)) = \max(G(d_{3i+1}), f(x)) \\ &= \max(f'(d_{3i+1}), f(x)) = \max(d_{3i+2}, f(x)) = \max(x, f(x)) = f(x), \end{aligned}$$

that is, $g(x) = f(x)$ holds for any x in the interval $d_{3i+2}, \dots, d_{3i+3} - 1$. Let k be an arbitrary fixed number, and pick any i such that $d_{3i+2} + 1 > k$. Then,

$$d_{3i+3} = f'(d_{3i+2}) = f^{d_{3i+2}+1}(d_{3i+2}) > f^k(d_{3i+2}) = g^k(d_{3i+2}).$$

The last equality holds since we have $d_{3i+2} \leq f^{\ell}(d_{3i+2}) < d_{3i+3}$ when $\ell \leq k$. This, proves that for any fixed k there exists x such that $f'(x) > g^k(x)$, and thus the Growth Theorem yields $f' \not\leq_E g$. □

Corollary 4.9 *For any n, there exists degrees $\mathbf{a}_\ell$ and $\mathbf{a}_h$ strictly between $\mathbf{0}^{[n]}$ and $\mathbf{0}^{[n+1]}$ such that $\mathbf{a}'_\ell = \mathbf{0}^{[n+1]}$ and $\mathbf{a}'_h = \mathbf{0}^{[n+2]}$.*

Proof. Let f be an honest function such that $\deg(f) = \mathbf{0}^{[n]}$. By Theorem 4.7, we have an honest function g_0 such that $f <_E g_0$ and $f' \equiv_E g'_0$. Let $\mathbf{a}_\ell = \deg(g_0)$. Then we have $\mathbf{0}^{[n]} < \mathbf{a}_\ell < \mathbf{0}^{[n+1]} = \mathbf{a}'_\ell$. By Theorem 4.8, we have an honest function g_1 such that $g_1 <_E f'$ and $f'' \equiv_E g'_1$. Let $\mathbf{a}_h = \deg(g_1)$. Then we have $\mathbf{0}^{[n]} < \mathbf{a}_h < \mathbf{0}^{[n+1]}$ and $\mathbf{0}^{[n+2]} = \mathbf{a}'_h$. (The monotonicity of the jump operator assures that $\mathbf{a}_\ell < \mathbf{0}^{[n+1]}$ and that $\mathbf{0}^{[n]} < \mathbf{a}_h$.) □

Theorem 4.10 (Jump Inversion) *Let f and g_0 be honest functions such that*

$$f' \leq_E g_0 \leq_E f''.$$

Then, there exists an honest function h such that $h \leq_E f'$ and $h' \equiv_E g_0$.

Proof. Since f is honest, we have $f'(x+1) \geq 2^{f'(x)}$ (and we also have $f'(x) \geq 2^x_{x+1}$). We can assume without loss of generality that we also have $g_0(x+1) \geq 2^{g_0(x)}$. Otherwise, let $g_1(0) = g_0(0)$ and $g_1(x+1) = \max(2^{g_1(x)}, \max[g_0, f'](x+1))$. Then we obviously have $\max[g_0, f'] \leq g_1$. Furthermore, for some $u, v \leq x$ we have

$$g_1(x) = 2_u^{\max[g_0,f'](v)} \leq 2_{\max[g_0,f'](x)+1}^{\max[g_0,f'](x)} \leq \max[g_0, f'] \max[g_0, f'](x).$$

Thus, we have $\max[g_0, f'] \equiv_E g_1$ by the Growth Theorem, moreover, since $f' \leq_E g_0$, we have $g_0 \equiv_E g_1$. This shows that we may replace g_0 by g_1 to ensure that $g_0(x+1) \geq 2^{g_0(x)}$.

We define the function g by recursion on its argument x. Let $g(0) = g_0(0)$ and let

$$g(x+1) = \begin{cases} f''(y+1) & \text{where } y \text{ is the least number such that} \\ & g(x) \leq f''(y) < f''(y+1) < g_0(x+1), \\ & \\ g_0(x+1) & \text{if such } y \text{ does not exist.} \end{cases}$$

Claim I The function g is honest and

(a) $g \equiv_E g_0$;
(b) $g(x) \leq f''(y) \Rightarrow g(x+1) \leq f''(y+1)$ for any $x, y \in \mathbb{N}$.

It is easy to see that g is an honest function, and Clause (b) of the claim is a straightforward consequence of the definition of g. We will now argue that $g(x+1) \leq g_0(x+1) \leq g(2x+1)$, and thus, Clause (a) follows by the Growth Theorem. It is obvious that $g(x+1) \leq g_0(x+1)$. In order to verify that $g_0(x+1) \leq g(2x+1)$, we observe that there might, or might not, exist $\ell > 0$ and a sequence $y_0, \dots, y_\ell$ such that

$$g(x) \leq f''(y_0) \leq f''(y_1) \leq \dots \leq f''(y_\ell) \leq g_0(x+1) \leq f''(y_\ell + 1).$$

If such a sequence does not exist, we have $g(x+1) = g_0(x+1)$ and thus also $g_0(x+1) \leq g(2x+1)$. If such a sequence exists, we have $g(x+i) = f''(y_i)$ for $y = 1, \dots, \ell$ and $g(x+\ell+1) \geq g_0(x+1)$. Moreover, since $g_0(z) \leq f''(z)$ holds for any z, the sequence $y_0, \dots, y_\ell$ cannot be very long; indeed, $\ell \leq x$. Hence $g_0(x+1) \leq g(x+\ell+1) \leq g(x+x+1)$. This completes the proof of Claim I.

For any injection ϕ, we define the function $\mathcal{I}_\phi$ by $\mathcal{I}_\phi(x) = \max(\mathcal{S}_\phi(x), 2^x)$ where $\mathcal{S}_\phi(0) = 0$ and

$$\mathcal{S}_\phi(x) = \begin{cases} \phi(i+1) & \text{if } x = \phi(i) \text{ for some } i, \\ \mathcal{S}_\phi(x-1) & \text{otherwise,} \end{cases}$$

when $x > 0$. The straightforward proof that $\mathcal{I}_\phi$ is an honest function whenever ϕ is an honest function, is left to the reader. We will prove that $\mathcal{I}'_g \equiv_E g$ and $\mathcal{I}_g \leq_E \mathcal{I}_{f''}$ and $\mathcal{I}_{f''} \leq_E f'$. Our theorem follows from these facts as we have $g_0 \equiv_E g$ by Claim I (a).

Claim II For any honest function h where $h(x+1) \geq 2^{h(x)}$, we have

(a) $h(x+1) = \mathcal{I}_h(h(x))$;
(b) $h(i) \leq x < h(i+1) \;\Rightarrow\; \mathcal{I}_h(x) = \max(h(i+1), 2^x)$.

Clause (a) of this claim holds since

$$\mathcal{I}_h(h(x)) = \max(\mathcal{S}_h(h(x)), 2^{h(x)}) = \max(h(x+1), 2^{h(x)}) = h(x+1)$$

and Clause (b) follows easily from Clause (a) and the definition of $\mathcal{I}_h$.

We will now prove that $\mathcal{I}'_g \equiv_E g$. Since $g(x+1) \geq 2^{g(x)}$, we have $g(x) = \mathcal{I}^x_g(g(0))$ by Claim II(a). Hence, it is easy to see that there exist fixed m, n such that $(\mathcal{I}'_g)^m(x) \geq g(x)$ and $g^n(x) \geq \mathcal{I}'_g(x)$ (recall that $\mathcal{I}'_g(x) = \mathcal{I}^{x+1}_g(x)$), and thus, the Growth Theorem yields $\mathcal{I}'_g \equiv_E g$.

Next we prove that $\mathcal{I}_g \leq_E \mathcal{I}_{f''}$. By the Growth Theorem, it suffices to prove $\mathcal{I}_g \leq \mathcal{I}^2_{f''}$. Pick an arbitrary x. If $\mathcal{I}_g(x) = 2^x$, we have $\mathcal{I}_g(x) \leq \mathcal{I}^2_{f''}(x)$ as f'' grows sufficiently fast. Now, assume $\mathcal{I}_g(x) \neq 2^x$. Fix the unique i and j such that $g(i) \leq x < g(i+1)$ and $f''(j) \leq g(i) < f''(j+1)$. Now

$$\begin{aligned} \mathcal{I}^2_{f''}(x) &\geq \mathcal{I}^2_{f''}(f''(j)) && \text{as } \mathcal{I}^2_{f''} \text{ is monotone and } x \geq f''(j) \\ &= f''(j+2) && \text{Claim II (a)} \\ &\geq g(i+1) && \text{Claim I (b) and } g(i) < f''(j+1) \\ &= \mathcal{I}_g(x) && \text{Claim II (b).} \end{aligned}$$

This proves $\mathcal{I}_g \leq_E \mathcal{I}_{f''}$.

Finally, we will prove that $\mathcal{I}_{f''} \leq_E f'$. Indeed, we will prove something stronger (given the Growth Theorem), namely that we have $\mathcal{I}_{h'} \leq h^2$ for any honest function h where $h(x+1) \geq 2^{h(x)}$. For such an h, we have

$$\mathcal{I}_{h'}(h'(x)) = h'(x+1) = h^{x+2}(x+1) \leq h^2 h^{x+1}(x) = h^2(h'(x)). \tag{4.2}$$

Claim II assures that the first equality of (4.2) holds. The remaining relations of (4.2) hold trivially. Now, pick any x and fix the unique i such that $h'(i) \leq x < h'(i+1)$. If $\mathcal{I}_{h'}(x) = 2^x$, then $\mathcal{I}_{h'}(x) \leq h^2(x)$ holds trivially. If $\mathcal{I}_{h'}(x) \neq 2^x$, we have $\mathcal{I}_{h'}(x) = h'(i+1)$ by Claim II (b), and thus

$$\begin{aligned}
\mathcal{I}_{h'}(x) &= h'(i+1) \\
&= \mathcal{I}_{h'}(h'(i)) && \text{Claim II (a)} \\
&\leq h^2(h'(i)) && (4.2) \\
&\leq h^2(x) && \text{as } x \geq h'(i).
\end{aligned}$$

This completes the proof of the theorem. □

Corollary 4.11 *Let* $\mathbf{a}$ *be a degree strictly between* $\mathbf{0}^{[n+1]}$ *and* $\mathbf{0}^{[n+2]}$*. Then, there exists a degree* $\mathbf{b}$ *strictly between* $\mathbf{0}^{[n]}$ *and* $\mathbf{0}^{[n+1]}$ *such that* $\mathbf{b}' = \mathbf{a}$*.*

Proof. Let f, g be honest function such that $\deg(f) = \mathbf{0}^{[n+1]}$ and $\deg(g) = \mathbf{a}$. We can without loss of generality assume that $f(x) \geq 2^x_{x+1}$. Now, Theorem 4.10 yields an honest function h such that $h \leq f$ and $h' \equiv g$. Let $\mathbf{b} = \deg(h)$. Then, we have $\mathbf{b}' = \mathbf{a}$, and by the monotonicity of the jump operator we also have $\mathbf{0}^{[n]} < \mathbf{b} < \mathbf{0}^{[n+1]}$. □

The next corollary follows straightforwardly from Corollary 4.9 and Corollary 4.11.

Corollary 4.12 (Low and High Degrees) *For any* $n \in \mathbb{N}$*, there exists a degree which is low*$_n$*, and there exists a degree which is high*$_n$*.*

Clause (i) of the next theorem is also proved in [**6**], whereas (ii) is stated as an open problem in [**6**].

Theorem 4.13

(i) *For any degrees* $\mathbf{a}$ *and* $\mathbf{b}$*, we have* $\mathbf{a}' \cup \mathbf{b}' \leq (\mathbf{a} \cup \mathbf{b})'$*. Moreover, there exist* $\mathbf{a}$ *and* $\mathbf{b}$ *such that* $\mathbf{a}' \cup \mathbf{b}' = (\mathbf{a} \cup \mathbf{b})'$*, and there exist* $\mathbf{a}$ *and* $\mathbf{b}$ *such that* $\mathbf{a}' \cup \mathbf{b}' < (\mathbf{a} \cup \mathbf{b})'$*.*

(ii) *For any degrees* $\mathbf{a}$ *and* $\mathbf{b}$*, we have* $\mathbf{a}' \cap \mathbf{b}' = (\mathbf{a} \cap \mathbf{b})'$*.*

Proof. We start by proving (ii). Now, $\mathbf{a} \geq \mathbf{a} \cap \mathbf{b}$ holds in any lattice, and thus, by the monotonicity of the jump operator, we also have $\mathbf{a}' \geq (\mathbf{a} \cap \mathbf{b})'$. By the same token, we have $\mathbf{b}' \geq (\mathbf{a} \cap \mathbf{b})'$. Hence, as $\mathbf{a}' \cap \mathbf{b}'$ is the greatest lower bound of $\mathbf{a}'$ and $\mathbf{b}'$, we have $\mathbf{a}' \cap \mathbf{b}' \geq (\mathbf{a} \cap \mathbf{b})'$. We will now prove that $\mathbf{a}' \cap \mathbf{b}' \leq (\mathbf{a} \cap \mathbf{b})'$ also holds. Let f, g be honest functions such that $\mathbf{a} = \deg(f)$ and $\mathbf{b} = \deg(g)$. We have

$$\begin{aligned}
\min[f', g'](x) &= \min(f'(x), g'(x)) \\
&= \min(f^{x+1}(x), g^{x+1}(x)) && \text{def. of the jump} \\
&\leq \min[f, g]^{2(x+1)}(x) && \text{Lemma 3.3} \\
&\leq \min[f, g]^{\min[f,g]'(x)+1+x+1}(x) \\
&= \min[f, g]^{\min[f,g]'(x)+1} \min[f, g]^{x+1}(x) \\
&= \min[f, g]^{\min[f,g]'(x)+1}(\min[f, g]'(x)) \\
&= \min[f, g]' \min[f, g]'(x),
\end{aligned}$$

and thus, by the Growth Theorem, we infer that $\min[f', g'] \leq_E \min[f, g]'$. This proves that $\mathbf{a}' \cap \mathbf{b}' \leq (\mathbf{a} \cap \mathbf{b})'$, and (ii) follows.

We turn to the proof of (i). The proof of $\mathbf{a}' \cup \mathbf{b}' \leq (\mathbf{a} \cup \mathbf{b})'$ (for any degrees $\mathbf{a}, \mathbf{b}$) is symmetric to the proof of $\mathbf{a}' \cap \mathbf{b}' \geq (\mathbf{a} \cap \mathbf{b})'$ given above. Furthermore, it is obvious that there exist degrees $\mathbf{a}, \mathbf{b}$ such that $\mathbf{a}' \cup \mathbf{b}' = (\mathbf{a} \cup \mathbf{b})'$. The existence of $\mathbf{a}$ and $\mathbf{b}$ such that $\mathbf{a}' \cup \mathbf{b}' < (\mathbf{a} \cup \mathbf{b})'$ is a consequence of the following claim.

Claim For any degree $\mathbf{c} \geq \mathbf{0}'$, there exist degrees $\mathbf{a}$ and $\mathbf{b}$ such that $\mathbf{c} = \mathbf{a} \cup \mathbf{b} = \mathbf{a}' = \mathbf{b}'$.

By this claim, we have degrees $\mathbf{a}, \mathbf{b}, \mathbf{c}$ such that

$$\mathbf{a}' \cup \mathbf{b}' = \mathbf{c} \cup \mathbf{c} = \mathbf{c} < \mathbf{c}' = (\mathbf{a} \cup \mathbf{b})'.$$

To prove the claim, let $\mathbf{c}$ be a degree above $\mathbf{0}'$, and let f be an honest function such that $\deg(f') = \mathbf{c}$. Such a f exists by Theorem 4.10. Define the sequence $\{d_i\}_{i\in\mathbb{N}}$ by $d_0 = 0$ and $d_{i+1} = f'(d_i)$; define the functions G and H by $G(0) = H(0) = 0$ and, for $x > 0$, by

$$G(x) = \begin{cases} f'(x) & \text{if } x = d_{2i} \text{ for some } i, \\ G(x-1) & \text{otherwise,} \end{cases}$$

$$H(x) = \begin{cases} f'(x) & \text{if } x = d_{2i+1} \text{ for some } i, \\ H(x-1) & \text{otherwise,} \end{cases}$$

and, finally, let $g(x) = \max(f(x), G(x))$ and $h(x) = \max(f(x), H(x))$. It turns out that the claim holds when $\mathbf{a} = \deg(g)$ and $\mathbf{b} = \deg(h)$. The proof that this is indeed the case is nontrivial, and the details can be found in [**6**]. □

An *intermediate degree* is a degree below $\mathbf{0}'$ which, for any $n \in \mathbb{N}$, is neither low$_n$ nor high$_n$. We conclude this section by a theorem stating the existence of an intermediate degree.

Theorem 4.14 *There exists a degree* $\mathbf{a}$ *such that, for any* $n \in \mathbb{N}$, $\mathbf{0}^{[n]} < \mathbf{a}^{[n]} < \mathbf{0}^{[n+1]}$.

Proof. Let $f(x) = 2^x$. Define the sequence $\{d_i\}_{i\in\mathbb{N}}$ by $d_0 = 0$ and $d_{i+1} = f^{[d_i]}(d_i)$; define the function G by $G(0) = 0$ and, for $x > 0$, by

$$G(x) = \begin{cases} f'(x) & \text{if } d_{3i} \leq x < d_{3i+1} \text{ for some } i, \\ G(x-1) & \text{otherwise,} \end{cases}$$

and let $g(x) = \max(f(x), G(x))$. Now, g is an honest function, and $f \leq g \leq f'$. By the Growth Theorem, we have

$$\mathbf{0} = \deg(f) \leq \deg(g) \leq \deg(f') = \mathbf{0}'.$$

By the monotonicity of the jump operator, we have $\mathbf{0}^{[n]} \leq \deg(g)^{[n]} \leq \mathbf{0}^{[n+1]}$ for any $n \in \mathbb{N}$. It remains to prove that $\deg(g)^{[n]} \not\leq \mathbf{0}^{[n]}$ and $\mathbf{0}^{[n+1]} \not\leq \deg(g)^{[n]}$. The details can be found in [**7**]. □

5 On cupability and capability

Definition 5.1 A degree $\mathbf{a}$ *cups (up) to* a degree $\mathbf{b}$ if there exists $\mathbf{c}$ such that $\mathbf{c} < \mathbf{b}$ and $\mathbf{a} \cup \mathbf{c} = \mathbf{b}$. A degree $\mathbf{a}$ *caps (down) to* a degree $\mathbf{b}$ if there exists $\mathbf{c}$ such that $\mathbf{b} < \mathbf{c}$ and $\mathbf{a} \cap \mathbf{c} = \mathbf{b}$.

Next we define the binary relation $\ll$ on honest functions. A function $\rho\colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ is *universal* for an honest degree $\mathbf{a} = \deg(f)$ if for every $\xi\colon \mathbb{N} \to \mathbb{N}$ such that $\xi \leq_E f$, we have $\xi(x) = \rho(n, x)$ for some $n \in \mathbb{N}$. The relation $f \ll g$ holds if there exists a universal function ρ for the degree $\deg(f)$ such that $\rho \leq_E g$. We will also use $\ll$ to denote the corresponding relation on honest degrees.

The situation $\mathbf{a} \ll \mathbf{b}$ implies that $\mathbf{a} < \mathbf{b}$, but there exist degrees $\mathbf{a}, \mathbf{b}$ such that $\mathbf{a} < \mathbf{b}$ and $\mathbf{a} \not\ll \mathbf{b}$. The next lemma gives a characterisation of the $\ll$-relation.

Lemma 5.2 *Let g and f be honest functions. Then* (1) *and* (2) *are equivalent:*

(1) *$g \ll f$.*

(2) *There exists m such that, for any k, we have we have $g^k(x) < f^m(x)$ for all but finitely many x.*

Proof. To prove this lemma, we need a refined version of the Kleene Normal Form Theorem. We assume the reader is familiar with the *computable functions*, *indexes* for computable functions, *computation trees* and other well-known concepts in computability theory. When e is an index for the computable function f, we adopt the traditional abuse of notation and write $\{e\}(\vec{x})$ both for (i) the computation of $f(\vec{x})$ associated with e and for (ii) the eventual result of the computation. Let $\mathcal{U}$ be a function such that $\mathcal{U}(\langle x_1, \ldots, x_m\rangle) = x_m$, i.e., a function giving the last coordinate of a sequence number. Let $\mathcal{T}$ be the Kleene predicate, i.e., the predicate $\mathcal{T}(e, \langle x_1, \ldots, x_n\rangle, t)$ holds iff t is a computation tree for $\{e\}(x_1, \ldots, x_n)$. The relation $\mathcal{T}$ is elementary, so is the function $\mathcal{U}$, and for each total computable function ϕ we have

$$\phi(x_1, \ldots, x_n) = \{e\}(x_1, \ldots, x_n) = \mathcal{U}(\mu z[\mathcal{T}(e, \langle x_1, \ldots, x_n\rangle, z)])$$

when e is a computable index for ϕ.

> **Claim** (Normal Form Theorem) An n-ary function ψ is elementary in an honest function f iff there exist a recursive index e for ψ and a fixed number k such that
>
> $$\{e\}(x_1, \ldots, x_n) = \mathcal{U}(\mu y \leq f^k(\max(x_1, \ldots, x_n))[\mathcal{T}(e, \langle x_1, \ldots, x_n\rangle, y)]).$$

We sketch a proof of this claim: Assume

$$\psi(\vec{x}) = \{e\}(\vec{x}) = \mathcal{U}(\mu y \leq f^k(\max(\vec{x}))[\mathcal{T}(e, \langle\vec{x}\rangle, y)]).$$

The predicate $\mathcal{T}$ is elementary, and $\mathcal{U}$ and max are elementary functions. The elementary functions are closed under composition and the bounded μ-operator. Thus, ψ is elementary in f. To prove the other direction of the equivalence, assume that ψ is elementary in the honest function f. Then, ψ can be build from the functions $0, \mathcal{S}, \mathcal{I}^n_i$, max and f by composition and bounded primitive recursion. Complete the proof of the claim by induction on such a build-up of ψ. (The details can be found in [**5**].)

We will now turn to the proof of the lemma. Fix m such that, for any k, we have $g^k(x) < f^m(x)$ for all but finitely many x. Then, for every k, there exists $n_k \in \mathbb{N}$ such that

$$g^k(x) < n_k + f^m(x) \tag{5.1}$$

holds for all x. Let ξ be any unary function elementary in g. By the claim we have an index e for ξ, an elementary predicate $\mathcal{T}_1$, an elementary function $\mathcal{U}$ and a fixed $\ell \in \mathbb{N}$ such that

$$\xi(x) = \mathcal{U}((\mu t \leq g^\ell(x))[\mathcal{T}_1(e, x, t)]).$$

By (5.1), we have $n_\ell \in \mathbb{N}$ such that

$$\xi(x) = \mathcal{U}((\mu t \leq g^\ell(x))[\mathcal{T}_1(e, x, t)]) = \mathcal{U}((\mu t \leq n_\ell + f^m(x))[\mathcal{T}_1(e, x, t)]).$$

Let $\rho(\langle e,n\rangle,x)=\mathcal{U}((\mu t<n+f^m(x))[\mathcal{T}_1(e,x,t)])$. Then, we have $\rho\leq_E f$, and for every unary function ξ elementary in g, there exists n such that $\xi(x)=\rho(n,x)$. This proves that (1) implies (2).

Assume $g\ll f$. Then, there exists a function ρ such that ρ is a universal function for $\deg(g)$ and $\rho\leq_E f$. Let $\psi(x)=(\max_{i\leq x}\max_{j\leq x}\rho(i,j))+1$. Then, we have $\psi\leq_E f$, and hence, there exists m such that $\psi(x)\leq f^m(x)$. It is easy to see that for any unary function ϕ elementary in g, we have $\phi(x)<\psi(x)\leq f^m(x)$ for all but finitely many x. Thus, for any k, as $g^k\leq_E g$, we have $g^k(x)<f^m(x)$ for all but finitely many x. This proves that (2) implies (1). □

The next theorem was proved for the first time in [**8**].

Theorem 5.3 *If* $\mathbf{0}\ll\mathbf{a}<\mathbf{b}$*, then* $\mathbf{a}$ *cups to* $\mathbf{b}$.

Proof. Let f and g be honest functions such that $\deg(f)=\mathbf{a}$, and $\deg(g)=\mathbf{b}$, and $f\leq g$. Define the sequence $\{d_i\}_{i\in\mathbb{N}}$ by $d_0=0$; $d_{2i+1}=g(d_{2i})$; and $d_{2i+2}=f(d_{2i+1})$. Furthermore, define the function h by $h(x)=\max(H(x),2^x)$ where $H(0)=0$ and, for $x>0$

$$H(x)=\begin{cases} g(x) & \text{if } x=d_{2i} \text{ for some } i,\\ H(x-1) & \text{otherwise.}\end{cases}$$

It is possible to prove that h is an honest function such that $\max[f,h]\equiv_E g$ and $g\not\leq_E h$. The details can be found in [**8**]. □

We have tried hard to strengthen Theorem 5.3 by proving that $\mathbf{a}$ cups up to $\mathbf{b}$ whenever $\mathbf{0}<\mathbf{a}<\mathbf{b}$. We have not succeeded, and thus it remains an open problem if there exist degrees other than $\mathbf{0}$ that do not cup up to degrees above them. However, with a possible exceptions of some degrees not being $\ll$-above $\mathbf{0}$, any degree cups up to any degree above it, and thus, "cups up to" is a not a very restrictive relation. We will see that the relation "caps down to" is far more restrictive.

Lemma 5.4 *Let g,f be honest functions such that f caps to g and $g\leq f$. Then, there exist a fixed $c\in\mathbb{N}$ such that for each k, we have $f^k(x)\leq g^{ck}(x)$ for infinitely many x.*

Proof. Since f caps to g we have an honest h such that $\min[f,h]\leq_E g$. By the Growth Theorem, we can fix a $c\in\mathbb{N}$ such that $\min[f,h]\leq g^c$. Now, as $\min[f,h]$ and g are monotone, we also have $\min[f,h]^k\leq g^{ck}$ (for any k). Moreover, as we have assumed $g\leq f$, we have $\min[f,h]^k\leq g^{ck}\leq f^{ck}$ (for any k). As f caps to g by h, we have $h\not\leq_E f$, and thus, by the Growth Theorem, for any $c,k\in\mathbb{N}$ we have infinitely many values $x_0,x_1,x_2,\ldots$ such that $f^{ck}(x_i)<h(x_i)$. For each x_i of these values, we have

$$\min[f,h]^k(x_i)\leq g^{ck}(x_i)\leq f^{ck}(x_i)<h(x_i). \tag{5.2}$$

This entails that $\min[f,h]^k(x_i)=f^k(x_i)$. If not, (5.2) yields a contradiction. Thus, (5.2) entails that $f^k(x_i)\leq g^{ck}(x_i)$ for each x_i in the sequence $x_0,x_1,x_2,\ldots$. □

Theorem 5.5 *If* $\mathbf{a}\ll\mathbf{b}$*, then* $\mathbf{b}$ *does not cap to* $\mathbf{a}$.

Proof. Assume that $\deg(g)=\mathbf{a}\ll\mathbf{b}=\deg(f)$ and that $\mathbf{b}$ caps to $\mathbf{a}$. We can without loss of generality assume $g\leq f$. Since $\mathbf{a}\ll\mathbf{b}$, Lemma 5.2 yields a fixed m such that for any k, we have $g^k(x)<f^m(x)$ for all but finitely many x. Since $\mathbf{b}$ caps to $\mathbf{a}$, Lemma 5.4 yields a fixed c such that for each k, we have $f^k(x)\leq g^{ck}(x)$ for infinitely many x. This is a contradiction. □

It is natural to ask whether the converse of Theorem 5.5 also holds, that is, do we have $\mathbf{a} \ll \mathbf{b}$ if, and only if, $\mathbf{b}$ does not cap to $\mathbf{a}$? (This was stated as an open problem in [**7**].) The next theorem gives a negative answer to this question.

Theorem 5.6 *There exist degrees* $\mathbf{a} < \mathbf{b}$ *such that* $\mathbf{b}$ *does not cap to* $\mathbf{a}$ *even if we have* $\mathbf{a} \not\ll \mathbf{b}$.

Proof. Let f be an honest function such that $f(x) \geq 2^x_x$. We will construct an honest function g and prove the two following claims.

> **Claim I** For any m, we have $g^{m^2}(x) = f^m(x)$ for infinitely many x.
>
> **Claim II** For any m, we have $g^{m^2}(x) < f^{3m+1}(x)$ for all but finitely many x.

Let $\nu(k)$ equal 1 plus the exponent of 2 in the prime factorisation of $k+2$. Thus, ν is an elementary function. (Any elementary function ϕ such that the set $\{x \mid \phi(x) = n\}$ is infinite for all $n > 0$, could replace ν in this proof.) For each $k \in \mathbb{N}$, we will define a sequence $d_{k,0} < d_{k,1} < \ldots < d_{k,\nu(k)^2}$. Moreover, for each k, we will have $d_{k,\nu(k)^2} < d_{k+1,0}$. Let $d_{0,0} = 0$. For each $j \in \{1, \ldots, \nu(k)^2\}$, let

$$d_{k,j} = \begin{cases} f(d_{k,j-\nu(k)}) & \text{if } \nu(k) \text{ divides } j, \\ 2^{d_{k,j-1}} & \text{otherwise,} \end{cases}$$

and let $d_{k+1,0} = f'(d_{k,\nu(k)^2})$. Furthermore, let

$$G(x) = \begin{cases} d_{k,i+1} & \text{if } d_{k,i} \leq x < d_{k,i+1} \text{ for some } k, i, \\ d_{k,\nu(k)^2} & \text{if } d_{k,\nu(k)^2} \leq x < d_{k+1,0} \text{ for some } k, i, \end{cases}$$

and let $g(x) = \max(2^x, G(x))$. This completes the construction of g. The reader should note the following properties of g (and f):

(P1) $g(d_{k,i}) = d_{k,i+1}$ for any k and any $i < \nu(k)^2$;
(P2) for any k and any $i < \nu(k)^2$, we have $g(d_{k,i}) = f(d_{k,i})$ if $\nu(k)$ divides i;
(P3) for any k and any $i < \nu(k)^2$, we have $g(d_{k,i}) = 2^{d_{k,i}}$ if $\nu(k)$ does not divide i;
(P4) $g^{\nu(k)^2}(d_{k,0}) = d_{k,\nu(k)^2} = f^{\nu(k)}(d_{k,0})$ for any k;
(P5) for any m, we have $g^m(d_{k,\nu(k)^2}) = 2_m^{d_{k,\nu(k)^2}} < d_{k+1,0}$ for all but finitely many k.

These five properties are more or less straightforward consequences of the construction of g; in particular, to see that (P5) holds, note that $d_{k+1,0} = f'(d_{k,\nu(k)^2})$ and $f(x) \geq 2^x_x$.

Claim I follows straightaway from (P4). For any m we have $g^{m^2}(d_{k,0}) = f^m(d_{k,0})$ for each of the infinitely many k's such that $\nu(k) = m$. We turn to the proof of Claim II. The proof splits into two cases, namely the case when x lies in an interval of the form $d_{k,0}, \ldots, d_{k,\nu(k)}-1$, and the case when x lies in an interval of the form $d_{k,\nu(k)}, \ldots, d_{k+1,0}-1$.

We will first prove that we have $g^{m^2}(x) < f^{3m+1}(x)$ when x is sufficiently large and lies in an interval of the form $d_{k,0}, \ldots, d_{k,\nu(k)} - 1$. The proofs splits into the two sub-cases

$m \geq \nu(k)$ and $m < \nu(k)$. First, assume that $m \geq \nu(k)$. We have

$$\begin{aligned}
f^{3m+1}(x) &= f^{(3m+1)-\nu(k)} f^{\nu(k)}(x) && \\
&\geq f^{(3m+1)-\nu(k)} f^{\nu(k)}(d_{k,0}) && \text{f is monotone} \\
&= f^{(3m+1)-\nu(k)}(d_{k,\nu(k)^2}) && \text{(P4)} \\
&> f(d_{k,\nu(k)^2}) && \text{as $m \geq \nu(k)$} \\
&\geq 2^{d_{k,\nu(k)^2}}_{d_{k,\nu(k)^2}} && \text{as $f(x) \geq 2^x_x$} \\
&\geq 2^{d_{k,\nu(k)^2}}_{m^2} && \text{x is large} \\
&= g^{m^2}(d_{k,\nu(k)^2}) && \text{(P5) and x is large} \\
&\geq g^{m^2}(x) && \text{g is monotone.}
\end{aligned}$$

Next, assume that $m < \nu(k)$. Fix the unique i such that $d_{k,i} \leq x < d_{k,i+1}$. Since $m < \nu(k)$, there will be at most one number j in the interval $i, \ldots, \min(i+m, \nu(k)^2)$ such that $\nu(k)$ divides j. Hence, by (P2), (P3) and (P5), there exist m_0, m_1 such that

$$g^m(x) \;\leq\; 2^{f(2^x_{m_1})}_{m_0} \;\leq\; f^3(x). \tag{5.3}$$

Furthermore, g is monotone and $x \leq d_{k,\nu(k)^2}$, and then, by (P5), we have

$$g^{m^2}(x) \;\leq\; g^{m^2}(d_{k,\nu(k)^2}) \;< d_{k+1,0} \tag{5.4}$$

for all but finitely many x. It follows from (5.3) and (5.4) that $g^{m^2}(x) < f^{3m+1}(x)$ for all sufficiently large x.

The reader is invited to verify that we also have $g^{m^2}(x) < f^{3m+1}(x)$ for sufficiently large x lying in intervals of the form $d_{k,\nu(k)}, \ldots, d_{k+1,0} - 1$. To verify this, note that for any x in such an interval we have $g(x) = 2^x$ whereas $f(x) \geq 2^x_x$. This completes the proof of Claim II.

We will briefly now argue that g is honest an honest function. The function f is honest by assumption. First we argue that $d_{k,j} = x$ is an elementary relation in k, j, x. Let $a \mid b$ denote the relation "a divides b". This relation is elementary. We have

$$\begin{aligned}
&d_{k,j} = x \iff \\
&\qquad \big(j \neq 0 \;\wedge\; \nu(k) \mid j \;\wedge\; \exists x_0 < x \,[\, d_{k,j-\nu(k)} = x_0 \;\wedge\; f(x_0) = x \,] \big) \\
&\qquad\quad \vee \big(j \neq 0 \;\wedge\; \neg\, \nu(k) \mid j \;\wedge\; \exists x_0 < x \,[\, d_{k,j-1} = x_0 \;\wedge\; 2^{x_0} = x \,] \big) \\
&\qquad\qquad \vee \big(j = 0 \;\wedge\; \exists x_0 < x \,[\, d_{k,\nu(k)^2} = x_0 \;\wedge\; 2^{x_0} = x \,] \big) \\
&\qquad\qquad\qquad\qquad\qquad \vee \big(k = 0 \;\wedge\; j = 0 \;\wedge\; x = 0 \big).
\end{aligned}$$

This can be viewed as a recursive definition of $d_{k,j} = x$. All the functions, relations and operations involved are elementary. Thus, we have defined the relation $d_{k,j} = x$ by a recursion scheme of the form

$$R(k,j,x) \iff \phi(R(k_0,j_0,x_0), R(k_1,j_1,x_1), R(k_2,j_2,x_2))$$

where ϕ is an elementary predicate and $k_0, k_1, k_2 \leq k$; $j_0, j_1, j_2 \leq k$; and $x_0, x_1, x_2 \leq x$. The elementary predicates are closed under such a recursion scheme, and hence, $d_{k,j} = x$ is an elementary relation. Thus, $\exists k, j \leq x[d_{k,j} = x]$ is an elementary predicate. Once we have realised that this predicate is elementary, it becomes easy to see that g has elementary graph. Obviously, g is monotone and dominates 2^x. Thereby, g is honest.

We will now prove the theorem. We have $g \leq_E f$ by the Growth Theorem since $g \leq f$. Let m be any number. Pick x such that $x > m$ and $x = d_{k,\nu(k)^2}$ for some k. By (P5), we have $g^m(x) = 2^x_m < 2^x_x \leq f(x)$. Hence, we have $f \not\leq_E g$ by the Growth Theorem. This proves $g <_E f$. Claim I says that for any m there exist infinitely many x such that $g^{m^2}(x) = f^m(x)$. This entails that there cannot exist a fixed number n such that we for any m have $g^m(x) < f^n(x)$ for all but finitely many x. Thus, we have $g \not\lll f$ by Lemma 5.2. Finally, Claim II and Lemma 5.4 entail that f does not cup to g, and then, our theorem holds when $\mathbf{a} = \deg(g)$ and $\mathbf{b} = \deg(f)$. □

6 Controllable irreducibility and the pendulum theorem

Definition 6.1 A sequence of natural numbers $\{d_i\}_{i\in\mathbb{N}}$ is *elementary* if the relation $d_i = y$ is elementary. An honest function f is *controllably irreducible* to an honest function g if there exists an elementary sequence $d_0 < d_1 < d_2 < \ldots$ such that for any k we have $g^k(d_i) < f(d_i)$ for all but finitely many i.

In the next theorem we assume that a function f is controllably irreducible to a function h. We do not know how to prove this theorem if we only assume that f is irreducible to h.

Theorem 6.2 (Pendulum) *Let f, g and h be honest functions such that f is controllably irreducible to h and $g <_E f \leq_E g'$. Then there exists an honest function g_0 such that*

(i) $g <_E g_0 <_E f$ *(and f is controllably irreducible to g_0),*
(ii) $g_0 \not\leq_E h$, *and*
(iii) $g_0' \equiv_E g'$.

Proof. Let $e_0 < e_1 < e_2 < \ldots$ be an elementary sequence such that for any k we have $h^k(e_i) < f(e_i)$ for all sufficiently large e_i. Such a sequence exists since f is controllably irreducible to h. We construct the sequence $d_0 < d_1 < d_2 < \ldots$ by letting $d_0 = 0$ and $d_{i+1} = e_j$ where where e_j is the least element in the sequence $e_0 < e_1 < e_2 < \ldots$ such that

$$g'g'g'(d_i) < e_j \ \wedge\ \exists y \leq e_j \exists x \leq y\, [\, f(x) = y \ \wedge\ g^i(x) < y\,].$$

The sequence $\{d_i\}_{i\in\mathbb{N}}$ is well defined as $f \not\leq_E g$, and, by the Growth Theorem, for each i there exists infinitely many x such that $g^i(x) < f(x)$. Moreover, the sequence is elementary as d_{i+1} is defined from d_i by elementary operations.

Let $g_0(x) = \max(\mathcal{S}_f(x), g(x))$ where $\mathcal{S}_f(0) = 0$ and

$$\mathcal{S}_f(x) = \begin{cases} f(x) & \text{if } x = d_i \text{ for some } i, \\ \mathcal{S}_f(x-1) & \text{otherwise,} \end{cases}$$

when $x > 0$. Since that f and g are honest and $\{d_i\}_{i\in\mathbb{N}}$ is elementary, it is straightforward to verify that that g_0 is an honest function.

We will first prove that Clause (i) of the Theorem holds. Since $g <_E f$, we can without loss of generality assume that $g(x) \leq f(x)$. This entails that we also have $g_0(x) \leq f(x)$, and thus, $g_0 \leq_E f$ follows by the Growth Theorem. Moreover, we have constructed g_0 such that we for each k have infinitely many x such that $g_0^k(x) < f(x)$, and thus, again by the Growth Theorem, we have $f \not\leq_E g_0$. This proves that $g_0 <_E f$. Obviously, we also have $g <_E g_0$. Thus, (i) holds.

It is easy to prove that (ii) holds. In order to see that $g_0 \not\leq_E h$, just observe that for any k we have $g_0(d_i) = f(d_i) > h^k(d_i)$ for all but finitely many d_i, and then, use the Growth Theorem. This completes the proof of (ii).

Claim Let $g'(d_i) \leq x \leq g'g'(d_i)$. Then, $g_0^y(x) = g^y(x)$ whenever $y \leq x$.

It should not be hard to see that this claim holds: Observe that

(a) $g_0(z) = g(z)$ for any z in the interval $g'(d_i), \ldots, d_{i+1} - 1$;
(b) $g^y(x) < g'(x) < g'(g'g'(d_i)) \leq d_{i+1}$.

The claim follows easily from (a) and (b).

Next we prove that $g_0'(x) \leq g'g'g'(x)$. Pick an arbitrary x and fix i such that $d_i \leq x < d_{i+1}$. There exists a maximal number z such that $z \leq x+1$ and

$$g_0'(x) = g_0^{x+1}(x) = g_0^{(x+1)-z}g^z(x).$$

If $z = x+1$, then $g_0'(x) \leq g'g'g'(x)$ holds trivially. Assume $z < x+1$. Now, $z < x+1$ implies that $d_{i+1} \leq g^z(x)$. This is easily verified by inspecting the definition of g_0. Furthermore, note that we can assume that $f(x) \leq g'(x)$. There will be no loss of generality to assume this as $f \leq_E g'$. We have

$$\begin{aligned}
g_0'(x) &= g_0^{(x+1)-z}g^z(x) & \\
&= g_0^{x-z}g_0g^z(x) & \\
&= g_0^{x-z}\max(\mathcal{S}_f(g^z(x)), gg^z(x)) & \text{def. of } g_0 \\
&\leq g_0^{x-z}\max(f(g^z(x)), gg^z(x)) & \text{def. of } \mathcal{S}_f \\
&\leq g_0^{x-z}\max(f(g'(x)), g'(x)) & \text{def. of } g' \text{ and } z \leq x \\
&\leq g_0^{x-z}g'g'(x) & \text{since } f(x) \leq g'(x).
\end{aligned}$$

This proves that $g_0'(x) \leq g_0^{x-z}g'g'(x)$ for some $z \leq x$ such that $d_{i+1} \leq g^z(x)$. We also have $g'(d_{i+1}) \leq g'g^z(x) \leq g'g'(x) \leq g'g'(d_{i+1})$, and hence, $g_0'(x) \leq g'g'g'(x)$ follows by Claim.

This proves that $g_0'(x) \leq g'g'g'(x)$ holds for any x. By the Growth Theorem, we have $g_0' \leq_E g'$. Furthermore, it is easy to see that $g \leq_E g_0$, and hence, we have $g' \leq_E g_0'$ by the monotonicity of the jump operator. Thus, $g_0 \equiv_E g$. This completes the proof of (iii). □

Before we investigate the notion of controllable irreducibility further, we will discuss what it should mean for a degree to be controllably irreducible to another degree: The Growth Theorem entails that if f is controllably irreducible to g, then f is controllably irreducible to any h elementary in g. So we can say that f is controllably irreducible to deg(g) if f is controllably irreducible to some, or equivalently all, representative(s) in deg(g). The same cannot be said when replacing f by its degree. This motivates the next definition.

Definition 6.3 A degree **a** is *controllably irreducible* to a degree **b** when some function in **a** is controllably irreducible to some, or equivalently all, function(s) in **b**. A degree **a** is *not controllably irreducible* to a degree **b** when no function in **a** is controllably irreducible to some, or equivalently all, function(s) in **b**. A degree **b** is *slightly above* a degree **a** when $\mathbf{a} < \mathbf{b}$ and **b** is not controllably irreducible to **a**.

The next theorem entails that if there exists one degree that is slightly above a degree **a**, then there will be a lot of degrees slightly above **a**.

Theorem 6.4 *Let* $\mathbf{b}$ *be slightly above* $\mathbf{a}$*, and let* $\mathbf{a} \leq \mathbf{c}_i \leq \mathbf{b}$ *for* $i = 1, 2$*. Then,* $\mathbf{c}_2$ *cannot be controllably irreducible to* $\mathbf{c}_1$*.*

Proof. Assume that $\mathbf{c}_2$ is controllably irreducible to $\mathbf{c}_1 = \deg(g)$. Then, there exist $f \in \mathbf{c}_2$ and and elementary sequence $d_0 < d_1 < d_2 < \dots$ such that for any k we have $g^k(d_i) < f(d_i)$ for all but finitely many i. Let $\mathbf{a} = \deg(h_1)$ and $\mathbf{b} = \deg(h_2)$. We can without loss of generality assume that $h_1 \leq g$ and $f \leq h_2$, and then, for any k, we have $h_1^k(d_i) < h_2(d_i)$ for all but finitely many i. This contradicts that $\mathbf{b}$ is slightly above $\mathbf{a}$. □

The next theorem requires proof techniques based on enumerations and diagonalisations. This is the first result we prove on the structure of honest elementary degrees that requires such techniques.

Theorem 6.5 *There exists a degree that is slightly above* $\mathbf{0}$*.*

Proof. We will construct an honest function f such that $\deg(f)$ is not controllably irreducible to $\mathbf{0} = \deg(2^x)$. We have to prove that no function in $\deg(f)$ is controllably irreducible to 2^x. By the Growth Theorem, it is sufficient to prove that no finite iterate of f is controllably irreducible to 2^x. Besides, we have to prove that f is not elementary, that is, we have to prove that no fixed iterate of 2^x dominates f.

Thus, on the one hand, f will have to grow somewhat fast: at some point it must be greater than any given iterate of 2^x. On the other hand, we must make certain that no elementary sequence $d_0 < d_1 < d_2 < \dots$ is a witness to the undesired controlled irreducibility. That involves diagonalising against all such possible sequences. Furthermore, this diagonalisation must work for all finite iterations of f.

To improve readability, we will throughout this proof use the notation $2_x(y)$ in place of 2_x^y.

We need a master list of sequences $d_0 < d_1 < d_2 < \dots$. There is no good elementary listing of all such total sequences, but there is one if we allow for partial (finite) sequences, as follows. Let $t_0, t_1, t_2 \dots$ be a listing of all elementary functions in two variables induced by using some primitive recursive coding of the base functions and operations allowed in the definition of elementarity. There is no universal elementary function for this listing; that is, the relation $t_i(x, y) = z$ is not elementary. However, because of the simplicity of the coding, one can code a particular computation as an integer and use that the relation

$$q \text{ bounds a witness that } t_i(x, y) = z$$

is elementary. For every elementary sequence $d_0 < d_1 < d_2 < \dots$ there is an i such that $t_i(x, y)$ is the characteristic function of the relation $d_x = y$. In the other direction, given i and q, it is elementary to see whether t_i looks like the characteristic function of such a sequence when considering only witnesses beneath q. If t_i is not the characteristic function of such a sequence, then eventually there will be a witness beneath q showing that. Let T_i be the sequence so induced by t_i, either an infinite sequence $d_0 < d_1 < d_2 < \dots$ if t_i is a good characteristic function, or a finite sequence if not. We will have to diagonalise against T_i if it is total without knowing whether it is total.

We now define a function f as follows. At stage n we will define f on the n-th interval $I_n = [x_n, x_{n+1})$. To start, put $I_0 = \{0\}$, and $f(0) = 2$. We use an auxiliary function $L(n) \subseteq n$, which tells us at stage $n+1$ which T_i's (for $i < n$) do *not* need to be attended to. (One problem is that some T_i might always demand attention. Once it gets attended to, it gets put on the list L, allowing other requirements to be met. It will eventually be taken off the list and, if it remains active, will then be attended to again.) To start, put

$L(0) = \emptyset$. Suppose inductively that we have defined the set $L(n-1)$ and the function f up to x_n. We will define I_n (i.e., determine x_{n+1}), and f on I_n, and $L(n)$, in several steps. First consider $J_{n,0} = [x_n, 2_n(x_n)]$ (the first sub-interval of I_n). We would like to pick a T_i to work on, if possible. So consider all $j < n$ not in $L(n-1)$ for which some $z \in J_{n,0}$ is in the range of T_j. Choose the pair j, z for which $y = 2_j(z)$ is less than $2_{n+1}(x_n)$, bounds a witness that z is in the range of T_j, and is the minimal such number; if there are several choices giving the same value, pick the one with j minimal. We call this value of j the active index for the interval I_n. Then we put $f(x) = \max(y, 2^x)$ on $[x_n, 2_n(x_n)]$. The outcome of this action is that f grows reasonably fast (at least as fast as 2_j) from x_n to that z, and no faster than that afterwards for a while. We set $L(n) = (L(n-1) \cup \{j\}) \setminus \{0, \ldots, j-1\}$: since j just got attended to, it can be ignored for a while, yet allows smaller requirements to receive attention. If no such pair j, z exists, we put $f(x) = 2_{n+1}(x_n)$ on $J_{n,0}$.

Now we need to consider iterations of f, and make sure that they grow slowly. We will define $J_{n,k}$ and X_k inductively on k. $J_{n,0}$ is already defined; let $X_0 = \{0, \ldots, n-1\}$. Suppose we have already defined the interval $J_{n,k-1} = [x_{n,k-1}, x_{n,k})$ and f on $J_{n,k-1}$. Then we put $J_{n,k} = [x_{n,k}, 2_n(x_{n,k}))$ and set $f(x) = 2^x$ on this interval. If there exists some $i \in X_{k-1}$ such that T_i has a value in $J_{n,k-1}$, then we put $X_k = X_{k-1} \setminus \{i\}$. For some k there will be so such i (as X_0 is finite and the X-sequence is monotonically shrinking). When that happens, put $x_{n+1} = x_{n,k+1}$. That completes stage n.

This completes the definition of f. To complete the proof of the theorem, we will prove that

(1) f is honest;
(2) f is not elementary;
(3) no function in $\deg(f)$ is controllably reducible to a function in $\mathbf{0}$.

First we prove (1). We obviously have $f(x) \geq 2^x$ for every x. Furthermore, each interval I_n contains one subinterval $[x_n, q]$ (namely for $q = 2_{j-1}(z)$), on which f is constant and equal to 2^q, and one subinterval $[z+1, x_{n+1}]$, on which f equals 2^x. Hence in the interior of each I_n f is non-decreasing. Finally $f(x_{n+1}-1) = 2^{x_{n+1}-1} < 2^{x_{n+1}} \leq f(x_{n+1})$, hence f is globally non-decreasing. It remains to show that the graph of f is elementary. The auxiliary function L can be encoded into integers up to 2^n, so for a given x we can decide what kind of interval x is in, and which values $j \in \{1, \ldots, n\} \setminus L(n-1)$ are possible. In particular for each x we can compute the value x_n for which $x_n \leq x < x_{n+1}$, and it suffices to compute $f(x_n)$ from these data. This is possible because we have $f(x_n) = y$ iff

$$\begin{aligned}(\exists j \leq n)(\exists \xi, \zeta < y)[\, j \notin L(n-1) \;\wedge\; t_j(\xi, \zeta) = 1 \;\wedge\; 2_j(\zeta) = y \,]\\ \wedge\, (\forall y' < y)\neg(\exists j \leq n)(\exists \xi, \zeta < y)[\, j \notin L(n-1) \;\wedge\; t_j(\xi, \zeta) = 1 \;\wedge\; 2_j(\zeta) = y \,].\end{aligned}$$

Hence, the graph of f is elementary. This proves that f is an honest function.

We turn to the proof of (2). We have to show that for every k there exists some x, such that $f(x) > 2_k(x)$. For this it is sufficient to show that for every k there exists some $\ell > k$ such that ℓ is active in some interval I_n. There are infinitely many simple ways to describe the function $x \mapsto 2^x$, so choose some term t_ℓ describing this function with $\ell > k$ such that $2_\ell(x)$ bounds a witness that $T_\ell(x) = 2^x$. The range of T_ℓ intersects each of the intervals $J_{n,0}$. Hence, if neither ℓ nor any $j > \ell$ is active for any n, then for every n some $j < \ell$ is active. Then in each step some integer is added to $L(n)$, while some smaller integers are removed. Eventually every integer less than ℓ is either in $L(n)$ or never active. (In some detail, if $\ell - 1$ is ever active, it will be put onto $L(n)$ and never removed, while if $\ell - 1$ is never active then it is fine too. Once $\ell - 1$ is settled, continue to the stage, if any, when $\ell - 2$ is active. Iterate. Since ℓ is finite, this eventually halts.) At that point there is nothing stopping ℓ from being active, which is what we wanted to show. This proves that f is not an elementary function.

We will now prove (3). By the Growth Theorem, it suffices to show the following claim:

> **(*)** Let $\ell \in \mathbb{N}$. Then there does not exist any elementary sequence $d_0 < d_1 < d_2 < \dots$ such that for any k we have $f^{\ell}(d_m) > 2_k(d_m)$ for all but finitely many m.

Now, for every elementary sequence $d_0 < d_1 < d_2 < \dots$, we have $T_i(\jmath) = d_\jmath$ for some i. Thus, by (*), it suffices to show the following claim:

> **(**)** Let $\ell \in \mathbb{N}$, and let T_i be total. Then there exists a k such that we have $f^{(\ell)}(T_i(m)) \leq 2_k(T_i(m))$ for infinitely many m.

The proof of (**) splits into two cases.

Case I: $T_i(m) \in J_{n,\jmath}$ with $\jmath > 0$ for infinitely many m. Then, for $n > i$ we have that such an interval $J_{n,\jmath}$ is not the last interval in the chain $J_{n,0}, \dots, J_{n,k}$. Hence $f(x) = 2^x$ holds true on $[T_i(m), 2_n(T_i(m))]$, and for $n > \ell$ we have $f^{\ell}(T_i(m)) = 2_\ell(T_i(m))$.

Case II: not Case I. Then, $T_i(m) \in J_{n,0}$ for all but finitely many m. If i is active infinitely often, then for the witness $z = T_i(m)$ to this we have $f(T_i(m)) = 2_i(T_i(m))$, and $f^{(\ell)}(T_i(m)) = 2_{i+\ell-1}(T_i(m))$ for $i + \ell \leq n$, which suffices. If not, then i is active only finitely often. Once i is no longer active, it is never added to $L(n)$, but it is eventually removed from $L(n)$ (by the proof that f is not elementary). Once that happens, for each interval $J_{n,0}$ containing a value $T_i(m)$, i was not active because of some pair $\jmath, z$ with $2_\jmath(z) \leq 2_i(T_i(m))$. But then we have again $f^{(\ell)}(T_i(m)) \leq 2_{i+\ell-1}(T_i(m))$ for $i + \ell \leq n$.

This completes the proof that no function in deg(f) is controllably reducible to a function in $\mathbf{0}$. □

Corollary 6.6

(i) *There exist degrees* $\mathbf{a}$ *and* $\mathbf{b}$ *such that* $\mathbf{a}$ *is not controllably irreducible to* $\mathbf{b}$ *and vice versa.*

(ii) *Any countable partial ordering can be embedded in the degrees slightly above* $\mathbf{0}$.

Proof. By Theorem 6.5 and the Density-Splitting Theorem, we have a degree $\mathbf{a}$ slightly above $\mathbf{0}$ and two incomparable degrees $\mathbf{b}_1, \mathbf{b}_2$ such that $\mathbf{0} < \mathbf{b}_i < \mathbf{a}$ (for $i = 1, 2$). By Theorem 6.4, $\mathbf{b}_1$ will not be controllably irreducible to $\mathbf{b}_2$, and $\mathbf{b}_2$ will not be controllably irreducible to $\mathbf{b}_1$. This proves (i). Furthermore, we know that any countable partial ordering can be embedded between two degrees $\mathbf{a}$ and $\mathbf{b}$ whenever $\mathbf{a} < \mathbf{b}$. Thus, (ii) follows from Theorem 6.5 and Theorem 6.4. □

7 A Σ_1-complete first-order theory

In this section we give a first-order theory for deriving theorems on honest elementary degrees. We will prove that this theory is powerful enough to derive any true Σ_1-statement, that is, any true statement in the form $\exists x_1, \dots, x_n A$ where A is a quantifier-free and does not contain other variables than $x_1, \dots, x_n$. The reader should be aware that the proofs in this section may be a bit sketchy.

Definition 7.1 Let

$$a \cup b = c \equiv a \leq c \ \wedge \ b \leq c \ \wedge \ \forall d\,[\,a \leq d \ \wedge \ b \leq d \ \rightarrow \ c \leq d\,]$$

and let

$$a \cap b = c \equiv a \geq c \ \wedge \ b \geq c \ \wedge \ \forall d\,[\,a \geq d \ \wedge \ b \geq d \ \rightarrow \ c \geq d\,].$$

Furthermore, let $a \mid b \equiv a \not\leq b \ \wedge \ b \not\leq a$ and $a < b \equiv a \leq b \ \wedge \ a \neq b$.

Let $\mathcal{L}$ be the first-order language $\{\leq, \cdot', 0\}$, and let T be an $\mathcal{L}$-theory which, in addition to standard axioms stating that $\leq$ is a partial ordering, contains the following axioms:

- $\forall a\,[\,0 \leq a\,]$ (Bottom Element)
- $\forall a, b\,[\,a \leq b \rightarrow a' \leq b'\,]$ (Monotonicity)
- $\forall a\,[\,a \neq a'\,]$ (Strictness)
- $\forall a, b \exists c\,[\,a \cup b = c\,]$ and $\forall a, b \exists c\,[\,a \cap b = c\,]$ (Lattice)
- $\forall a, b, c\,[\,a \cup (b \cap c) = (a \cup b) \cap (a \cup b)\,]$ (Distributivity)
- $\forall a, b\,[\,a < b \rightarrow \exists c_1, c_2\,[\,c_1 \mid c_2 \ \wedge \ c_1 \cap c_2 = a \ \wedge \ c_1 \cup c_2 = b\,]\,]$ (Density)
- $\forall a \exists b\,[\,a < b \ \wedge \ b' = a'\,]$ (Low Degrees)
- $\forall a \exists b\,[\,b < a' \ \wedge \ b' = a''\,]$ (High Degrees)
- $\forall a, b\,[\,a' \leq b \leq a'' \ \rightarrow \ \exists c\,[\,c \leq a \ \wedge \ c' = b\,]]$ (Jump Inversion)
- $\forall a, b, c\,[\,a < b \leq a' \ \wedge \ b \not\leq c \rightarrow \exists d\,[\,a < d < b \ \wedge \ d' = a' \ \wedge \ d \not\leq c\,]]$ (Pendulum)
- $\forall a, b[a' \cap b' = (a \cap b)']$.

Note that $\cap$ and $\cup$ are not symbols of the language $\mathcal{L}$, but all the axioms can be reduced to first-order statements over $\mathcal{L}$ in an obvious way. That $\cap$ distributes over $\cup$, that is $a \cap (b \cup c) = (a \cap b) \cup (a \cap b)$, follows from the axioms; see Birkhoff [**1**].

Definition 7.2 A sublattice L of a jump lattice is *complete* when

$$a, b \in L \ \wedge \ a' < b \ \Rightarrow \ a' \in L.$$

A lattice L is *connected* if for any two elements $x, y \in L$ there exists a sequence of elements $z_1, \ldots, z_k \in L$ such that

- $x = z_1$ and $y = z_k$;
- $z_i < z_{i+1}$ or $z_i > z_{i+1}$ (for $i \in \{1, \ldots, k-1\}$).

Lemma 7.3 *Let L be a finite complete and connected sublattice of a jump lattice which is a model of T. There exists a homomorphism $L \rightarrow \mathbb{N}$ where the jump in $\mathbb{N}$ is the successor function.*

Proof. If L is a complete connected lattice containing n points, then we can enumerate the points of L as $\ell_1, \ldots, \ell_n$ such that $\{\ell_1, \ldots, \ell_k\}$ is a complete and connected lattice for all $k \leq n$: choose ℓ_1 arbitrarily, and choose jumps or jump inverses of existing elements whenever this is possible.

We prove the lemma by induction on the number of elements in L. Suppose that L is a complete sublattice together with a homomorphism $\varphi\colon L \rightarrow \mathbb{N}$, and let ℓ be some point not occurring in L. If ℓ is neither the jump of an element in L, nor is $\ell' \in L$, then we define $\varphi(\ell)$ to be the maximum of $\{\varphi(x) : x < \ell\}$. Thus we have $\varphi(\ell) \geq \varphi(x)$ for all $x < \ell$. Since $<$ is transitive, this also implies $\varphi(\ell) \leq \varphi(x)$ for all $x > \ell$.

If $\ell' \in L$, we put $\varphi(\ell) = \varphi(\ell') - 1$. If this happens to be negative, we just increase all values of φ by 1. As L is a complete lattice, there are no elements $x \in L$ with $x < \ell$. Suppose that $x > \ell$. Then $x' > \ell'$, hence $\varphi(x') \geq \varphi(\ell')$, and therefore $\varphi(x) \geq \varphi(\ell)$. A similar argument applies if there is some $x \in L$ with $x' = \ell$. □

Lemma 7.4 *Let $\mathfrak{L}$ be any model of T. Let L be a finite lattice, and let $a, b, c_1, \dots, c_n$ elements of $\mathfrak{L}$ such that $a < b \leq a'$ and $b \mid c_i$ for $i = 1, \dots, n$. Then, there exists an embedding $\psi\colon L \to \mathfrak{L}$ such that for any $x \in \psi(L)$ we have*

- *$a < x < b$;*
- *$x \mid c_i$ for $i = 1, \dots, n$;*
- *$x' = a'$.*

Proof. To prove this lemma, we must use that $\mathfrak{L}$ satisfies the Pendulum Axiom and the Density Axiom. We omit the details. □

Lemma 7.5 *Let L be a finite jump lattice which is contained in a model of T, and let $\mathfrak{L}$ be an arbitrary model of T. Then there exists an embedding $\psi\colon L \to \mathfrak{L}$.*

Proof. We can without loss of generality assume that the lattice L is complete and connected. Let φ be the homomorphism given by Lemma 7.3, and assume that $n = \max\{\varphi(a) \mid a \in L\}$. Furthermore, let $L(k) = \{a \in L \mid \varphi(a) = k\}$. We will call $L(k)$ the *k-th level of L*. We can without loss of generality assume that there is only one element of level $L(n)$ and that each element of level k jumps to an element of level $k+1$, that is, for each $a \in L(k)$ there exists $b \in L(k+1)$ such that $a' = b$.

We will construct the embedding $\psi\colon L \to \mathfrak{L}$ level by level. First we construct $\psi\colon L(n) \to \mathfrak{L}$, then we construct $\psi\colon L(n-1) \to \mathfrak{L}$, and so on. There is only one degree a at level n, let $\psi(a)$ be an arbitrary degree strictly between $\mathbf{0}^{[n]}$ and $\mathbf{0}^{[n+1]}$.

Assume we have constructed $\psi\colon L(k+1) \to \mathfrak{L}$. We will now construct $\psi\colon L(k) \to \mathfrak{L}$. Let $m_0, m_1, \dots, m_{n_k}$ be an enumeration of the elements in $L(k+1)$ such that m_i is a maximal element in the set $\{m_i, \dots, m_{n_k}\}$, and let

$$\mathrm{inv}(a) = \{\, b \mid b \in L(k) \ \wedge\ b' = a \,\}.$$

Now, $\mathrm{inv}(m_0), \mathrm{inv}(m_1), \dots, \mathrm{inv}(m_{n_k})$ are disjunct sets, and

$$L(k) = \mathrm{inv}(m_0) \ \cup\ \mathrm{inv}(m_1) \ \cup\ \dots\ \cup\ \mathrm{inv}(m_{n_k}).$$

We construct the embedding $\psi\colon L(k) \to \mathfrak{L}$ by constructing first the embedding $\psi\colon \mathrm{inv}(m_0) \to \mathfrak{L}$, then the embedding $\psi\colon \mathrm{inv}(m_1) \to \mathfrak{L}$, and so on.

Here is how to construct $\psi\colon \mathrm{inv}(m_i) \to \mathfrak{L}$ (for any $i \in \{0, \dots, n_k\}$). Pick a maximal element $a \in \mathrm{inv}(m_i)$. The embedding ψ is now defined for all $b \in L$ such that $b > a$. Let $\alpha \in \mathcal{L}$ be given by $\alpha = \bigcap\{\psi(b) \mid b > a\}$. Now we have $\alpha' \geq \psi(a') \geq \alpha$ as $\mathcal{L}$ satisfies the axiom $\forall a, b[a' \cap b' = (a \cap b)']$. As $\mathfrak{L}$ satisfies the Jump Inversion Axiom, the Low Degree Axiom and the Pendulum Axiom, we can now find a suitable interval where we can, by Lemma 7.4, embed all elements in $\mathrm{inv}(m_i)$ that cannot be distinguished from a by comparing them to elements already embedded. Next we consider a maximal element in $\mathrm{inv}(m_i)$ not yet treated, and construct ψ on the set of elements equivalent to this element as we did for the elements equivalent to a. Continuing downwards in this way we construct ψ for all elements in $\mathrm{inv}(m_i)$. □

Theorem 7.6 (Σ_1-completeness) *Let $\mathfrak{H}$ denote the $\mathcal{L}$-structure of honest elementary degrees (our standard model for T), and let A be a Σ_1-statement in the language $\mathcal{L}$. Then*

$$\mathfrak{H} \models A \iff T \vdash A.$$

Proof. By Theorem 7.5, we know that if a finite jump lattice does not embed into an arbitrary model for T, then it will not embed into any model of T. Thus, a Σ_1-statement

A will be satisfied in all models for T if, and only if, A is satisfied in some model for T. By the Completeness Theorem for first-order logic, we have

$$\mathfrak{H} \models A \iff T \models A \iff T \vdash A. \qquad \square$$

References

[1] Birkhoff, G.: *Lattice Theory.* American Mathematical Society Colloquium Publications, Volume XXV, 1967.

[2] Blankertz, B. and Weiermann, A.: How to characterize provably total functions, in: *Gödel '96: Logical Foundations of Mathematics, Computer Science and Physics (ed. Hajek)*, Springer Lecture Notes in Logic, 6, Springer, 1996, 205–213.

[3] Buchholz, W., Cichon, A. and Weiermann, A.: A uniform approach to fundamental sequences and hierarchies, *Mathematical Logic Quarterly* **40**(2) (1994), 273–286.

[4] Kristiansen, L.: Information content and computational complexity of recursive sets, in: *Gödel '96: Logical Foundations of Mathematics, Computer Science and Physics (ed. Hajek)*, Springer Lecture Notes in Logic, 6, Springer, 1996, 235–246.

[5] Kristiansen, L.: *Papers on Subrecursion Theory*, Dr Scient Thesis, ISSN 0806-3036, ISBN 82-7368-130-0, Research report 217, Department of Informatics, University of Oslo, 1996.

[6] Kristiansen, L.: A jump operator on honest subrecursive degrees, *Archive for Mathematical Logic* **37** (1998), 105–125.

[7] Kristiansen, L.: Low$_n$, high$_n$, and intermediate subrecursive degrees, in: *Combinatorics, Computation and Logic (eds. Calude and Dinneen)*, Australian Computer Science Communications 21(3), Springer, Singapore, 1999, 286–300.

[8] Kristiansen, L.: Subrecursive degrees and fragments of Peano Arithmetic, *Archive for Mathematical Logic* **40** (2001), 365–397.

[9] Kristiansen, L., Schlage-Puchta, J.-C. and Weiermann, A.: Streamlined subrecursive degree theory, *Annals of Pure and Applied Logic* (2011), `doi:10.1016/j.apal.2011.11.004`.

[10] Machtey, M.: Augmented loop languages and classes of computable functions, *Journal of Computer and System Sciences* **6** (1972), 603–624.

[11] Machtey, M.: The honest subrecursive classes are a lattice, *Information and Control* **24** (1974), 247–263.

[12] Machtey, M.: On the density of honest subrecursive classes, *Journal of Computer and System Sciences* **10** (1975), 183–199.

[13] Meyer A. R. and Ritchie D. M: A classification of the recursive functions, *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik* **18** (1972), 71–82.

[14] Odifreddi, P: *Classical Recursion Theory*, North-Holland, 1989.

[15] Péter, R.: *Rekursive Funktionen*, Verlag der Ungarischen Akademie der Wissenschaften, Budapest, 1957. [English translation: Academic Press, New York, 1967.]

[16] Rogers, H.: *Theory of Recursive Functions and Effective Computability*, McGraw Hill, 1967.

[17] Rose, H. E.: *Subrecursion. Functions and Hierarchies*, Clarendon Press, Oxford, 1984.

Part III

Computations and Sets

Partially definable forcing and bounded arithmetic

Albert Atserias*, **Moritz Müller**†

* Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya, Barcelona, Spain
atserias@lsi.upc.edu

† Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
moritz.mueller@univie.ac.at

Abstract. We present a general forcing framework to prove independence results in bounded arithmetic and, closely related, lower bounds in propositional proof complexity.

Introduction

Various independence results in bounded arithmetic have been obtained using forcing type arguments. By bounded arithmetic we mean a first-order theory of arithmetic where induction is restricted to formulas of some particular syntactic form, typically formulas with bounded quantifiers. We describe a frame for forcing that can be seen as a common generalization of these arguments and Cohen forcing in set theory.

This introduction informally gives some general motivation, describes the connection to propositional proof complexity, reviews the mentioned forcing type arguments, compares them with Cohen forcing and then describes in some more detail the contents of this paper. More precise information can be found following the references, mainly pointing to surveys. All results are stated and proved in a generally accessible language. Some of their links to bounded arithmetic and propositional proof complexity are made explicit by remarks intended for the informed reader.

Foundational questions and complexity

Basic questions concerning the foundations of mathematics quickly lead to fundamental open problems from computational complexity theory such as P vs. NP or NP vs. co-NP. Indeed, Krajíček argues that these questions can be understood as "quantitative versions" [**27**, Section 5] of the central questions of mathematical logic a century ago, namely for the consistency and the decidability of first-order theories. Also Krajíček and Pudlák [**29**] tie the viability of versions of Hilbert's program to the nondeterministic time complexity of co-NP.

Pudlák argues that our understanding of independence is unsatisfactory in that "except for Gödel's theorem which gives only special formulas, no general method is known to prove independence of (arithmetical) Π_1 sentences" [**34**, Section 3]. Here progress is braked by the fact that already weak arithmetical theories like those in Buss' hierarchy correspond in a certain precise sense to the complexity classes in the polynomial hierarchy; [**11, 23**] are monographs, [**9, 10**] surveys on the subject.

[†] The second author thanks the John Templeton Foundation for its support under Grant #13152, *The Myriad Aspects of Infinity.*

Furthermore, establishing independence from bounded arithmetics is roughly equivalent to establishing proof-size lower bounds for propositional logic.

Proof complexity

For a sufficiently general notion of propositional proof system, the conjecture NP $\neq$ co-NP means that no propositional proof system has short proofs of all tautologies (i.e., of size polynomial in the length of the tautology) [**12**]. But today this is open even for the usual textbook systems, called *Frege systems*: Hilbert style calculi given by finitely many inference rules; [**4, 35, 37, 41, 48**] survey known lower bounds for weaker systems with partly different emphases.

Now, arithmetical theories are simulated by (often natural) propositional proof systems in the sense that theorems of the theory translate to sequences of tautologies with short proofs in the system (see [**29**] for a general treatment, [**7**] for a recent survey).

Example 0.1 (Paris–Wilkie translation) The theory $I\Delta_0(R)$ is Peano arithmetic where the induction scheme is adopted only for bounded formulas but in the language augmented by some new, say, binary relation symbol R. If $I\Delta_0(R)$ proves a Δ_0-formula $\varphi(R, x)$, then $\forall R \forall x \varphi(R, x)$ is true (in the standard model). This translates to a sequence of tautologies $\langle\varphi(R, x)\rangle_m, m \in \mathbb{N}$: insert m for x in $\varphi(R, x)$, replace bounded quantifiers by conjunctions or disjunctions, replace atoms not mentioning R by their truth values and keep atoms of the form $Rk\ell$ as propositional variables.

Paris and Wilkie [**32**] construct from a proof of $\varphi(R, x)$ in $I\Delta_0(R)$ and $m \in \mathbb{N}$ a short (length $m^{O(1)}$) proof of $\langle\varphi(R, x)\rangle_m$ in a *bounded depth* Frege system. This is a Frege system where only formulas of at most some fixed $\wedge/\vee$-alternation rank are allowed. ⌟

This way, independence can be inferred from proof-size lower bounds. A weak converse holds too. It is based on a type of argument invented by Ajtai [**1**], and it is here where forcing comes in.

Forcing in bounded arithmetic

Cohen's method of forcing cannot be used to prove independence of arithmetical statements because V_ω is not changed in generic extensions. In an informal sense however, forcing has been used to prove independence from weak arithmetics.

Paris and Wilkie used "a simple forcing argument" [**32**, p. 333] to show that the least number principle for existential formulas mentioning R does not suffice to prove the (bijective) pigeonhole principle PHP(R, x): "R is not a bijection from $\{y \mid y \leq x\}$ onto $\{y \mid y < x\}$". Riis [**38**] "proved by forcing" [**38**, p. 1] that even the least number principle for formulas with a certain amount of universal quantification does not suffice (Buss' theory $T^1_2(R)$). Furthermore, Riis generalized this to other principles.

Ajtai [**1**] proved that $I\Delta_0(R)$ does not prove PHP(R, x). In fact, he proved that the tautologies $\langle\text{PHP}(R, x)\rangle_n, n \geq 1$, do not have short proofs in bounded depth Frege systems; [**30, 33**] improved this to an exponential proof-size lower bound, implying independence from Buss' $T_2(R)$.

Ajtai's argument

Ajtai constructs an expansion (M, R) of a model M of true arithmetic where PHP(R, n) fails for some $n \in M$. Assume that M contains a size n^{100} depth 17 Frege proof π of $\langle\text{PHP}(R, x)\rangle_n$. But this formula is 'false' in (M, R) under the assignment corresponding to R. The art is to construct R in such a way that (M, R) satisfies the least number principle up to n^{100} for the property of being a 'false' line in π. Then π contains a first 'false' line. One argues that this contradicts the soundness of the system and concludes that π cannot exist.

The construction of R is "done according to the general ideas of Cohen's method of forcing " [**1**, p. 348]. However, the argument is "mostly combinatorial and probabilistic" [**1**, p. 347] relying on specialized and difficult versions of so-called switching lemmas in circuit complexity. As Ben-Sasson and Harsha put it, it is "extremely difficult to understand and explain" [**6**, § 1]. Lots of efforts have been made to simplify and reinterpret Ajtai's argument e.g. as a construction of valuations in Boolean algebras [**35**] or partial Boolean algebras [**23, 24**] or recently in terms of Buss–Pudlák games [**6, 36**]. In [**26**] Krajíček gives some general account, motivated "to understand the combinatorics behind constructions" [**26**, p. 437] like Ajtai's. Conceptually, later improvements [**5, 30, 33, 49**] of Ajtai's result "eliminate the non-standard model theory" [**5**, p. 367] and the forcing mode of speech. And technically, the mentioned switching lemmas have been improved and simplified (see [**3, 47**] for surveys).

Despite these efforts, not much is known on how to apply Ajtai's argument to stronger systems or other principles (cf. [**23**, Chapter 12] for known results). Perhaps one can say the abovementioned efforts did not lead to an understanding of Ajtai's argument as instantiating some general method as Pudlák asks for.

Comparison with Cohen forcing

This sorry state of affairs clearly contrasts with Cohen forcing in set theory. We recall briefly and informally its set-up. With a model M of ZF and a 'generic' set G external to M one associates a model $M[G]$ containing G. Intuitively, G being 'generic' means being 'random' with respect to possible partial information about it. Forcing is a way to reason about $M[G]$ using partial information about G. A piece of partial information p *forces* φ if any generic G 'satisfying' p leads to a model $M[G]$ satisfying φ. Such pieces can be extended in various, possibly incompatible ways, so we think of them as being partially ordered (the *forcing frame*).

The Extension Lemma states that extension preserves forcing. Reasoning about forcing rests on this and, following Shoenfield [**42**], two more central lemmas: the Truth Lemma asserts that every sentence true in $M[G]$ is already forced by some partial information p about G; the Definability Lemma states that forcing, as a binary relation, is definable in M. In turn, these forcing lemmas rest on the Forcing Completeness Theorem, a characterization of the 'semantic' forcing notion above by a handier 'syntactic' notion that is defined via recursion on logical syntax. This understanding of forcing underlies the "Principal Theorem" [**42**] stating that $M[G]$ models ZF. This way an independence question is reduced to a combinatorial task of designing an appropriate forcing frame.

In contrast, the mentioned forcing type arguments [**1, 32, 38**] in bounded arithmetic are not based on some more general background theory of forcing. Ajtai writes "Our terminology will be similar to the terminology of forcing but we actually do not use any result from it" [**1**, p. 348]. Insofar it is not completely clear why one should refer to these arguments as forcing arguments. Technically, the crucial difference is that the Definability Lemma fails. Forcing Completeness is proved neither in the original arguments nor in later presentations [**23**, § 12.7], [**50**], that emphasize the forcing mode of speech. In [**1, 32**] no 'syntactic' notion is defined, in [**38**] it is, but one for which Forcing Completeness fails.

This work

We propose a general background theory of forcing as a unifying way to understand the arguments of Paris, Wilkie, Riis and Ajtai [**1, 32, 38**]. In Section 1 we develop forcing generally as a method to construct generic "associates" that may happen to be extensions or expansions or neither, and without a Definability Lemma. It is general enough to naturally accommodate the mentioned forcing arguments [**1, 32, 38**] as well as Cohen forcing and many others.

In the context of bounded arithmetics, a Principal Theorem would state that generic expansions satisfy the least number principle for a certain fragment of formulas. In Section 2 we show this holds true when using a forcing that is in an appropriate sense 'definable' for the fragment in question. Thereby again, independence questions reduce to a combinatorial task of designing forcing frames.

In Section 3 we prove the independence results in [**1, 32, 38**] by this method. The aim is to understand the progress as being constituted by inventing forcings that are 'definable' for larger and larger fragments.

Related work

Forcing has been developed outside set theory in many different settings ([**2, 19**] survey some), and the development here follows these known lines. We refer to the examples throughout the text for a comparison with some other works. Forcing against bounded arithmetic has been developed by Takeuti and Yasumoto [**45, 46**] following not Cohen's original method but its reformulation by Scott and Solovay [**40**] as a method to construct Boolean valued models (see Remark 1.7). Scott [**40**] describes such a model for a 3rd order theory of the reals, by interpreting the language over real valued random variables. In his recent book [**28**] Krajíček develops such *forcing with random variables* in full detail as a method to study bounded arithmetics by using algorithmically restricted random variables. Ajtai's result can be proved using this method.

1 Forcing in general

This section develops a general frame for forcing arguments. In 1.1 we fix notation and establish basic facts concerning 'syntactic' forcing relations. In principle, countless 'syntactic" forcing relations $\Vdash$ may be defined, depending on how $\Vdash$ interacts with the logical symbols. Throughout this paper we assume (first-order) formulas to be written in the logical symbols $\{\forall, \exists, \wedge, \vee, \neg\}$ and we shall restrict attention to two kinds of forcings only, namely, *universal* and *existential forcings.*[1] Roughly, the choice depends on whether $\{\forall, \wedge, \neg\}$ or $\{\exists, \vee, \neg\}$ is taken as primitive while the other logical symbols are defined using the usual classical dualities. Existential forcing is often used, but we shall see that it has some disadvantages over universal forcing (Remark 1.34). In 1.2 we define a notion of genericity that is sufficiently general for all our purposes. In 1.3 we define *generic associates* and prove the Truth Lemma and the Forcing Completeness Theorem. Section 1.4 considers an important type of forcing that we call *conservative.* Section 1.5 gives examples and, finally, Section 1.6 discusses *weak forcing.*

In this section we fix

- a countable *forcing frame* $(P, \leq, D_0, D_1, \ldots)$ (defined below);

[1] See [**2, 18**] for examples of forcings that are neither universal nor existential.

- a countable structure M interpreting a countable language L;
- a countable language $L^* \supseteq L$.

The *forcing language* is $L^*(M)$, that is L^* together with all $a \in M$ as constants (we do not distinguish between M and its universe notationally). If not explicitly specified otherwise we let $\varphi, \psi, \ldots$ range over $L^*(M)$-sentences.

1.1 Forcing relations

We recall some elementary forcing terminology. A *forcing frame* is a structure $(P, \leq, D_0, D_1, \ldots)$ such that $\leq$ partially orders P and $D_0, D_1, \ldots$ are subsets of P. We use $p, q, r, \ldots$ to range over elements of P, called *conditions.* If $p \leq q$ we say p *extends* q and call p an *extension* of q. If p, q have a common extension, then they are *compatible*, symbolically $p \| q$; otherwise they are *incompatible*, symbolically $p \perp q$.

A set of conditions $X \subseteq P$ is *downward-closed* if it contains all extensions of its elements; being *upward-closed* is similarly explained. The set X is *consistent* if it contains a common extension of any two of its elements. If X is both upward-closed and consistent, then it is a *filter.* Further, X is *dense below* p if for every $q \leq p$ there is $r \leq q$ such that $r \in X$. Finally, X is *dense* if it is dense below all conditions, or equivalently, if every condition has an extension in X.

Definition 1.1 A *pre-forcing* is a binary relation $\Vdash$ between conditions and $L^*(M)$-sentences. If $p \Vdash \varphi$, we say *p forces φ.*

We use the notation

$$[\varphi] := \{p \mid p \Vdash \varphi\}.$$

Definition 1.2 A pre-forcing $\Vdash$ is *universal* or *existential* if it satisfies the following conditions of *universal* respectively *existential forcing recurrence*:

	universal	*existential*
$p \Vdash \neg\varphi$	iff $\forall q \leq p : q \nVdash \varphi$	iff $\forall q \leq p : q \nVdash \varphi$
$p \Vdash (\varphi \wedge \psi)$	iff $p \Vdash \varphi$ and $p \Vdash \psi$	iff $p \Vdash \neg(\neg\varphi \vee \neg\psi)$
$p \Vdash (\varphi \vee \psi)$	iff $p \Vdash \neg(\neg\varphi \wedge \neg\psi)$	iff $p \Vdash \varphi$ or $p \Vdash \psi$
$p \Vdash \forall x \chi(x)$	iff $\forall a \in M : p \Vdash \chi(a)$	iff $p \Vdash \neg\exists x \neg\chi(x)$
$p \Vdash \exists x \chi(x)$	iff $p \Vdash \neg\forall x \neg\chi(x)$	iff $\exists a \in M : p \Vdash \chi(a)$

Observe that a universal or existential pre-forcing is uniquely determined by its restriction to the atomic sentences of the forcing language.

Solving the recurrence one sees, for universal pre-forcings, that $p \Vdash \exists x \chi(x)$ if and only if $\bigcup_{a \in M}[\chi(a)]$ is dense below p. For existential pre-forcings one sees $p \Vdash \forall x \chi(x)$ if and only if $[\chi(a)]$ is dense below p for all $a \in M$. We collect some further direct consequences:

Lemma 1.3 *If $\Vdash$ is a universal or an existential pre-forcing, then*

(1) $p \Vdash \neg\neg\varphi$ *if and only if* $[\varphi]$ *is dense below* p.
(2) (Consistency) $[\varphi] \cap [\neg\varphi] = \emptyset$.
(3) $[\varphi] \cup [\neg\varphi]$ *is dense.*

Definition 1.4 Let $\Vdash$ be a pre-forcing and Φ be a set of $L^*(M)$-formulas.

(a) $\Vdash$ *satisfies Extension for* Φ if for every $\varphi \in \Phi$, the set $[\varphi]$ is downward-closed.

(b) $\Vdash$ *satisfies Stability for* Φ if for every $\varphi \in \Phi$ and $p \in P$, we have that p forces φ whenever $[\varphi]$ is dense below p.

For $\Phi = L^*(M)$ we omit the reference to it.

(c) $\Vdash$ is a *forcing* if it satisfies Extension and Stability for $L^*(M)$-atoms.

Lemma 1.5

(1) (Extension) *Universal and existential forcings satisfy Extension.*

(2) (Stability) *Universal forcings satisfy Stability.*

(3) *For a universal forcing* $\Vdash$ *it holds that* $p \Vdash \varphi$ *if and only if* $[\varphi]$ *is dense below* p.

(4) *For a universal forcing* $\Vdash$ *it holds that* $p \nVdash \varphi$ *if and only if* $q \Vdash \neg\varphi$ *for some* $q \leq p$.

Proof. Extension can be shown by a straightforward induction using forcing recurrence. We prove Stability by induction on (the number of logical symbols) in φ. Having a universal forcing we can assume that φ is written in the logical base $\{\wedge, \neg, \forall\}$.

– For atomic φ Stability is part of the definition of being a forcing.

– For the $\neg$-step argue indirectly: if $p \nVdash \neg\varphi$, then by forcing recurrence some $q \leq p$ forces φ, so by Extension and Consistency no extension of q forces $\neg\varphi$. Hence $[\neg\varphi]$ is not dense below p.

– For the $\wedge$-step, note $[(\varphi \wedge \psi)] = [\varphi] \cap [\psi]$ by universal recurrence. If this set is dense below p then so are both $[\varphi]$ and $[\psi]$. By induction p forces both φ and ψ, and hence $p \Vdash (\varphi \wedge \psi)$ by universal recurrence.

– The $\forall$-step is similar.

Part (3) is immediate by (1) and (2), and (4) follows from (3): $p \nVdash \varphi$ if and only if $[\varphi]$ is not dense below p if and only if there is $q \leq p$ such that for all $r \leq q$, $r \nVdash \varphi$, if and only if (by forcing recurrence) there is $q \leq p$ such that $q \Vdash \neg\varphi$. □

Example 1.6 Let $\Vdash$ be a universal forcing. A pre-forcing of obvious interest is:

$$p\|\varphi \text{ if and only if } p \nVdash \neg\varphi, \text{ that is, } q \Vdash \varphi \text{ for some } q \leq p.$$

We have $\Vdash \subseteq \|$ by Consistency. Stability of $\Vdash$ implies: $p\|\neg\neg\varphi$ if and only if $p\|\varphi$. Further

$$\begin{array}{r|ll}
p\|\neg\varphi & \text{iff} \quad \exists q \leq p : q \not\|\varphi & \\
p\|(\varphi \vee \psi) & \text{iff} \quad p\|\varphi \text{ or } p\|\psi & \text{(as existential pre-forcing)} \\
p\|\exists x \chi(x) & \text{iff} \quad \exists a \in M : p\|\chi(a) & \text{(as existential pre-forcing).}
\end{array}$$

Remark 1.7 (Boolean valued models) The last lemma has a natural topological reading. Namely, $(P, \leq)$ carries the topology whose open sets are the downward-closed sets. A set $X \subseteq P$ has interior $\overset{\circ}{X} = \{p \mid \forall q \leq p : q \in X\}$ and closure $\overline{X} = \{p \mid \exists q \in X : q \leq p\}$. For example, $\{p \mid p\|\varphi\} = \overline{[\varphi]}$. The sets equal to the interior of their closure are the regularly open ones. Note $\overset{\circ}{\overline{X}} = \{p \mid X \text{ is dense below } p\}$.

Thus Extension means that the sets $[\varphi]$ are open and Stability means that they are even regularly open. The regularly open sets form a complete Boolean algebra in such a way that, for universal forcings, the map $\varphi \mapsto [\varphi]$ is a Boolean valuation of $L^*(M)$ in this algebra.

1.2 Genericity

Let $\Vdash$ be an existential or universal forcing. Ideally, one would like to call a set generic if it intersects every dense set. As in general such sets do not exist, one has to restrict attention to those dense sets coming from a certain 'sufficiently rich' but countable Boolean algebra $\mathcal{B}(\Vdash)$.

In set theory usually the forcing frame is a set in M and one simply takes the algebra of its M-definable subsets (cf. Example 1.25). As M models ZF it is not surprising that this algebra is sufficiently rich. For some purposes (cf. Examples 1.27, 1.29, 1.30) already the algebra generated by the $[\varphi]$s is sufficiently rich, but not so in forcing against bounded arithmetic. One needs the family to contain certain sets as e.g. $\bigcup_{a\in M}\bigcap_{b\in M}[\varphi(a,b)]$ that we do construct in proofs. In [**1, 32, 38**] suitable algebras are defined ad hoc for their respective situations and there seems to be no canonical choice. That is why we padded the forcing frame by the sets $D_0, D_1, \ldots$: these sets will determine an algebra $\mathcal{B}(\Vdash)$ defined below (Definition 1.10).

Definition 1.8 A set $G \subseteq P$ is *generic* if it is a filter and intersects every dense (in P) set in $\mathcal{B}(\Vdash)$.

Our definition of $\mathcal{B}(\Vdash)$ follows Stern [**44**]: consider the two-sorted first-order structure (P, M) consisting of one sort carrying the forcing frame $(P, \leq, D_0, D_1, \ldots)$ and a second sort carrying the structure M. We let individual variables $\mu, \nu, \xi, \ldots$ range over the first sort and $x, y, z, \ldots$ range over the second sort.

For each L^*-atom $\varphi = \varphi(x_1, \ldots, x_r)$ let R_φ be an $r+1$-ary relation symbol of sort $P \times M^r$. The structure $(P, M)^{\Vdash}$ expands (P, M) by interpreting such a symbol R_φ by $\{p\overline{a} \in P \times M^r \mid p \Vdash \varphi(\overline{a})\}$.

We call the two-sorted first-order language of $(P, M)^{\Vdash}$ the *Stern formalism.* By forcing recurrence it is straightforward to show:

Lemma 1.9 *For every L^*-formula $\varphi(\overline{x})$ the set $\{p\overline{a} \mid p \Vdash \varphi(\overline{a})\}$ is definable in $(P, M)^{\Vdash}$.*

Given an L^*-formula $\varphi(\overline{x})$ we write

$$\xi \vdash \varphi(\overline{x})$$

for a formula of the Stern formalism defining $\{p\overline{a} \mid p \Vdash \varphi(\overline{a})\}$ in $(P, M)^{\Vdash}$. Here and in the following, *definable* (in a certain structure) always means definable *with parameters* (from the structure).

Definition 1.10 The *forcing algebra* $\mathcal{B}(\Vdash)$ is the set of all subsets of P definable in $(P, M)^{\Vdash}$ by a formula $\varphi(\xi)$ of the Stern formalism.

Clearly, the forcing algebra $\mathcal{B}(\Vdash)$ is countable. Thus, by a well-known argument:

Lemma 1.11 *Every condition is contained in some generic set.*

Sketch of Proof. Given $p \in P$, choose $p_1 \leq p$ in the first dense set, then $p_2 \leq p_1$ in the second dense set and so on. The filter generated by the sequence $p, p_1, p_2, \ldots$ is generic. □

Lemma 1.12 *If G is generic and $D \in \mathcal{B}(\Vdash)$ is dense below $p \in G$, then there is $q \in G \cap D$ with $q \leq p$.*

Proof. Let $D \in \mathcal{B}(\Vdash)$ be dense below p. It is routine to verify that

$$D(p) := (D \cap \{q \mid q \leq p\}) \cup \{q \mid p \perp q\}$$

is dense. Further $D(p) \in \mathcal{B}(\Vdash)$: if D is defined by $\varphi_D(\xi)$, then $D(p)$ is defined by

$$(\varphi_D(\xi) \wedge \xi \leq p) \vee \neg\exists\nu(\nu \leq \xi \wedge \nu \leq p),$$

a formula (with parameters) of the Stern formalism. By genericity there exists an $r \in G \cap D(p)$. As $p \in G$ and G is consistent, $r \notin \{q \mid p \perp q\}$, so $r \in D \cap \{q \mid q \leq p\}$. □

1.3 Generic associates

Let $\Vdash$ be a universal or existential forcing. The aim is to define for suitable $G \subseteq P$ (and our fixed structure M) an $L^*(M)$-structure $M[G]$ in such a way, that it models the following theory in the forcing language $L^*(M)$:

$$\mathrm{Th}(G) := \big\{\varphi \in L^*(M) \mid \exists p \in G : p \Vdash \varphi\big\}.$$

Obviously this cannot work in general, e.g. $\mathrm{Th}(G)$ may contradict usual first-order equality axioms. But we shall see that this is the only obstacle provided we stick to the idea that the constants from M "name" all the elements of $M[G]$.

First observe that for generic G, the theory $\mathrm{Th}(G)$ is complete and formally consistent in the following sense:

Lemma 1.13 *Let G be generic. For every $L^*(M)$-sentence φ either $\varphi \in \mathrm{Th}(G)$ or $\neg\varphi \in \mathrm{Th}(G)$, but not both.*

Proof. By Lemmas 1.3(3) and 1.9, G intersects $[\varphi] \cup [\neg\varphi] \in \mathcal{B}(\Vdash)$. Hence $\varphi \in \mathrm{Th}(G)$ or $\neg\varphi \in \mathrm{Th}(G)$ – but not both: assume there would be $p \in G$ forcing φ and $q \in G$ forcing $\neg\varphi$. Since G is a filter and filters are consistent, there would exist r extending both p and q; by Extension, r would force both φ and $\neg\varphi$ contradicting Consistency (of forcing). □

To define $M[G]$ we rely on some elementary facts about factorizations: for a theory T in a language L containing some constant symbol, the *Herbrand term structure* $\mathfrak{T}(T)$ for T has as universe all closed L-terms, interprets a function symbol $f \in L$ by $\bar{t} \mapsto f(\bar{t})$ and interprets a relation symbol $R \in L$ by $\{\bar{t} \mid R\bar{t} \in T\}$. Note that in $\mathfrak{T}(T)$ every closed term denotes itself. A *congruence* $\sim$ on $\mathfrak{T}(T)$ is an equivalence relation on $\mathfrak{T}(T)$ such that functions in $\mathfrak{T}(T)$ (i.e., interpretations of function symbols of L) map equivalent arguments (i.e., componentwise equivalent argument tuples) to equivalent values and every relation of $\mathfrak{T}(T)$ is a union of equivalence classes of tuples. In this case, let $\mathfrak{T}(T)/{\sim}$ denote the L-structure induced by $\mathfrak{T}(T)$ on the $\sim$-classes in the natural way. In $\mathfrak{T}(T)/{\sim}$ every closed term t denotes its $\sim$-class $t/{\sim}$.

Fact 1.14 If $\sim_T := \{(s,t) \mid s = t \in T\}$ is a congruence on $\mathfrak{T}(T)$, then the atomic sentences true in $\mathfrak{T}(T)/{\sim_T}$ are precisely those contained in T.

Definition 1.15 Let $G \subseteq P$. If $\sim_{\mathrm{Th}(G)}$ is a congruence on $\mathfrak{T}(\mathrm{Th}(G))$ and every closed term of the forcing language is $\sim_{\mathrm{Th}(G)}$-congruent to a constant $a \in M$, then we say $M[G]$ *is defined* and set

$$M[G] := \mathfrak{T}(\mathrm{Th}(G))/{\sim_{\mathrm{Th}(G)}} \,.$$

If G is generic and $M[G]$ defined, then $M[G]$ is a *generic associate of* M.

We call $M[G]$ a *generic extension of* M, if $L = L^*$ and there is an embedding of M into $M[G]$.

We call $M[G]$ a *generic expansion of* M, if

$$a \mapsto a/{\sim_{\mathrm{Th}(G)}} \colon M \cong M[G] \restriction L,$$

that is, if the map that sends each $a \in M$ to its $\sim_{\mathrm{Th}(G)}$-congruence class $a/{\sim_{\mathrm{Th}(G)}}$ is an isomorphism of M onto the restriction of $M[G]$ to L.

Remark 1.16 Sometimes we shall need the assumption that $M[G]$ is defined for every generic G. Because this assumption is trivially satisfied in all applications we are aware of, we consider it as a mere technicality and make no efforts to avoid it.

Lemma 1.17 *Let G be generic.*

(1) *$M[G]$ is defined if for all closed $L^*(M)$-terms t, t', all $L^*(M)$-atoms $\varphi(x)$ and all $p \in P$,*
  (a) *if $p \Vdash t = t'$, then $q \Vdash t' = t$ for some $q \leq p$,*
  (b) *if $p \Vdash \varphi(t)$ and $p \Vdash t = t'$, then $q \Vdash \varphi(t')$ for some $q \leq p$,*
  (c) *$q \Vdash t = a$ for some $q \leq p$ and $a \in M$.*

(2) *If $M[G]$ is defined, then it has universe $\{a/{\sim_{\mathrm{Th}(G)}} \mid a \in M\}$.*

We omit the proof.

Theorem 1.18 (Truth Lemma) *Let G be generic. If $M[G]$ is defined, then*

$$\mathrm{Th}(M[G]) = \mathrm{Th}(G).$$

Proof. We have to show: $M[G] \models \varphi$ if and only if $p \Vdash \varphi$ for some $p \in G$. We have two cases depending of whether $\Vdash$ is universal or existential. In both cases we proceed by induction on φ.

The case where $\Vdash$ is existential is easy. The base case follows by construction (Fact 1.14). Both the $\vee$-step and the $\exists$-step are trivial. Finally, $\neg\varphi \in \mathrm{Th}(M[G])$, that is, $\varphi \notin \mathrm{Th}(M[G])$, is equivalent to $\varphi \notin \mathrm{Th}(G)$ by induction and thus to $\neg\varphi \in \mathrm{Th}(G)$ by Lemma 1.13.

The case where $\Vdash$ is universal is more complicated. The base case and the $\neg$-step follow exactly as in the existential case. The $\wedge$-step is straightforward using the consistency of G. For the $\forall$-step, first assume that some $p \in G$ forces $\forall x\varphi(x)$, i.e., $p \Vdash \varphi(a)$ for every $a \in M$ by universal recurrence. By induction $M[G] \models \varphi(a)$ for every $a \in M$. Hence $M[G] \models \forall x\varphi(x)$ by Lemma 1.17(2). Conversely, assume $\forall x\varphi(x) \notin \mathrm{Th}(G)$. We aim to show $\varphi(a) \notin \mathrm{Th}(M[G])$ for some $a \in M$. By Lemma 1.13, $\neg\forall x\varphi(x) \in \mathrm{Th}(G)$, i.e., some $p \in G$ forces $\neg\forall x\varphi(x)$. By universal recurrence this means that for every $q \leq p$ there is $a \in M$ such that $q \nVdash \varphi(a)$. By Lemma 1.5(4) this means: for every $q \leq p$ there is $a \in M$ and there is $r \leq q$ such that $r \Vdash \neg\varphi(a)$. In other words, the set

$$D := \bigcup_{a \in M} [\neg\varphi(a)]$$

is dense below p. Clearly, $D \in \mathcal{B}(\Vdash)$: it is defined by $\exists x(\xi \vdash \neg\varphi(x))$, a formula (with parameters) of the Stern formalism (cf. Lemma 1.9). As $p \in G$, G intersects D by Lemma 1.12, i.e., there is some $a \in M$ such that $\neg\varphi(a) \in \mathrm{Th}(G)$. Then $\varphi(a) \notin \mathrm{Th}(G)$ by Lemma 1.13, so $\varphi(a) \notin \mathrm{Th}(M[G])$ by induction. □

Corollary 1.19 *Assume $M[G]$ is defined for every generic G. Then:*

(1) *If $\Vdash$ is existential, then $p \Vdash \varphi$ implies $M[G] \models \varphi$ for every generic G containing p.*

(2) (Forcing Completeness) *If $\Vdash$ is universal, then $p \Vdash \varphi$ if and only if $M[G] \models \varphi$ for every generic G containing p.*

Proof. By the Truth Lemma $p \Vdash \varphi$ implies $M[G] \models \varphi$ for every generic G containing p. This shows (1) and the forward direction of (2). The backward direction of (2) relies on Lemma 1.5(4) for universal forcings: if $p \nVdash \varphi$, there is $q \leq p$ such that $q \Vdash \neg\varphi$. By Lemma 1.11 there is a generic G containing q. By the Truth Lemma $M[G] \models \neg\varphi$, i.e., $M[G] \not\models \varphi$. Being a filter, G contains p. □

Corollary 1.20 *Assume that $M[G]$ is defined for every generic G and that $\Vdash$ is universal. Then for every condition $p \in P$ the set $\{\varphi \mid p \Vdash \varphi\}$ is closed under logical consequence.*

Proof. For every $p \in P$, the set of φ satisfying the right hand side of Forcing Completeness is obviously closed under logical consequence. □

Example 1.21 Let $\Vdash$ be a universal forcing and recall Example 1.6. Assume $M[G]$ is defined for every generic G. Then $p \| \varphi$ if and only if $M[G] \models \varphi$ for some generic G containing p. Further, $\{\varphi \mid p \| \varphi\}$ is closed under logical consequence. ⌟

We have the following preservation result.

Theorem 1.22 *Let T be a universal L^*-theory. If both*

(i) *for every condition p, the theory T is consistent with*

$$\mathrm{Lit}(p) := \{\varphi \mid p \Vdash \varphi, \varphi \textit{ is an } L^*(M)\textit{-Literal}\},$$

(ii) *and for every closed $L^*(M)$-term t, the set $\bigcup_{a \in M}[t = a]$ is dense,*

then $M[G]$ is defined for every generic G and satisfies T.

Proof. Let G be generic. To show $M[G]$ is defined we verify the three conditions (a), (b), (c) in Lemma 1.17(1). For (a), if $p \Vdash t = t'$ but $q \nVdash t' = t$ for every $q \leq p$, then $p \Vdash \neg t' = t$ by forcing recurrence. But then $\mathrm{Lit}(p)$ and hence $\mathrm{Lit}(p) \cup T$ is inconsistent, contradicting (i). Condition (b) is similarly verfied and (c) is the same as (ii).

To show $M[G] \models T$ is suffices to show that $M[G]$ embeds into a model of T (since T is universal). For this it suffices to show that $T \cup \mathrm{Diag}(M[G])$ is consistent. So let Δ be a finite subset of $\mathrm{Diag}(M[G])$. Then $\Delta \subseteq \mathrm{Th}(G)$ by the Truth Lemma, that is, every literal $\lambda \in \Delta$ is forced by some $p_\lambda \in G$. Since G is consistent it contains a common extension p of all the p_λ's. Then $\Delta \subseteq \mathrm{Lit}(p)$ by Extension and $T \cup \Delta$ is consistent by (i). □

1.4 Conservative forcing

Let $\Vdash$ be an existential or universal forcing. Which forcings produce generic expansions? We characterize these as follows.

Definition 1.23 The forcing $\Vdash$ is *conservative* if for every condition p and every atomic $L(M)$-sentence φ (i.e., without a symbol from $L^* \setminus L$)

$$p \Vdash \varphi \text{ if and only if } M \models \varphi.$$

Proposition 1.24 *If $\Vdash$ is conservative, then every generic associates is a generic expansion. The converse holds true in case $\Vdash$ is universal and $M[G]$ is defined for every generic G.*

Proof. For the first statement, let $M[G]$ be a generic associate of M. By Lemma 1.17 the map $a \mapsto a/{\sim_{\mathrm{Th}(G)}}\colon M \to M[G] \upharpoonright L$ is surjective. If it is not an isomorphism, then $\mathrm{Th}(M)$ and $\mathrm{Th}(M[G])$ disagree on some atomic $L(M)$-sentence. As $\mathrm{Th}(M[G]) = \mathrm{Th}(G)$ by the Truth Lemma, this contradicts conservativity.

For the second statement, argue indirectly and assume $\Vdash$ is not conservative. Choose an atomic $L(M)$-sentence φ and a condition p witnessing this. Then $p \Vdash \varphi$ if and only if $M \not\models \varphi$. By Forcing Completeness we find a generic associate $M[G]$ of M such that $M[G] \models \varphi$ if and only if $p \Vdash \varphi$. Therefore $\varphi \in \mathrm{Th}(M) \triangle \mathrm{Th}(M[G])$, so we infer that $\mathrm{Th}(M) \neq \mathrm{Th}(M[G] \upharpoonright L)$ and $M[G]$ cannot be an expansion of M. □

1.5 Some examples

Cohen forcing from set theory can be viewed as a special case of our general set-up:

Example 1.25 (Cohen forcing) Cohen forcing starts with a countable transitive standard model M of, say, ZF + GCH and wants $M[G]$ to be an extension of M. In particular $L^* = L = \{\in\}$. Different forcing extensions are obtained by different choices of $(P, \leq)$, typically a set in M, while the forcing $\Vdash_{\mathrm{Co}}$ is kept fix.

Following e.g. **[16]** one can define this forcing by universal forcing recurrence stipulating for atoms:

$$\begin{aligned} p \Vdash_{\mathrm{Co}} a \in b &\iff \big\{q \mid \exists r \exists c\, ((c,r) \in b \wedge q \leq r \wedge q \Vdash_{\mathrm{Co}} a = c)\big\} \text{ is dense below } p, \\ p \Vdash_{\mathrm{Co}} a = b &\iff \forall c \in \mathrm{dom}(a \cup b) : p \Vdash_{\mathrm{Co}} (c \in a \leftrightarrow c \in b). \end{aligned}$$

It is not hard to show that this uniquely determines a universal pre-forcing. The technicality of the definition is to ensure that it is a forcing. Genericity is defined to mean: intersect every dense set that is definable in M. This coincides with our notion for $\emptyset = D_0 = D_1 = \ldots$.

In set theory one defines $M[G]$ as follows: the membership symbol $\in$ is interpreted by membership and the constants $a \in M$ are interpreted by $a_G := \{b_G \mid \exists p \in G : (b,p) \in a\}$. It is easily seen that $M[G]$ is an extension of M for every generic G.

Under this definition of $M[G]$, one can show the Truth Lemma for atoms, that is: for every generic G, $a_G = b_G$ if and only if $a \sim_{\mathrm{Th}(G)} b$ and $a_G \in b_G$ if and only if $\mathrm{Th}(G)$ contains the atom $a \in b$. It follows that $M[G]$ in our sense is defined for every generic G. Second it follows that $M[G]$ in our sense is isomorphic to $M[G]$ in the sense of set theory. Indeed, $\{(a/{\sim_{\mathrm{Th}(G)}}, a_G) \mid a \in M\}$ is such an isomorphism. ⌟

Feferman was the first explicitly using forcing outside set-theory, namely to adress questions in computability theory. But already Cantor's back and forth method can be seen as a forcing argument. Both are examples of conservative forcing:

Example 1.26 (Cantor's Theorem) We give this simple example in some detail, because it reappears in similar form in Section 3.

Let $M = (A, A')$ be a countable two-sorted structure where the two sorts A and A' carry dense linear orders without endpoints $\preceq$ and $\preceq'$ respectively (i.e., $L = \{\preceq, \preceq'\}$). Set $L^* := L \cup \{R\}$ for a new binary relation symbol R.

Define the forcing frame $(P, \leq, D_0, D_1, \ldots)$ as follows: P is the set of all finite partial isomorphisms between A and A'; take $p \leq q$ to mean $p \supseteq q$; finally the sets $D_0, D_1, \ldots$ enumerate the sets $\{p \mid a \in \mathrm{dom}(p)\}, \{p \mid a' \in \mathrm{im}(p)\}$ for $a \in A, a' \in A'$. Each of these sets is dense.

To define a conservative universal pre-forcing $\Vdash_{\mathrm{Ca}}$ it suffices to define $p \Vdash_{\mathrm{Ca}} \varphi$ for φ an atom of the form Rab. Take this to mean $(a, b) \in p$.

Then $\Vdash_{\mathrm{Ca}}$ is a forcing: that $\Vdash_{\mathrm{Ca}}$ satisfies Extension for atoms is obvious. Because $\Vdash_{\mathrm{Ca}}$ is conservative we only have to show that it satifies Stability for atoms of the form Raa' for $a \in A, a' \in A'$. Argue indirectly: if $p \not\Vdash_{\mathrm{Ca}} Raa'$, then $(a, a') \notin p$. Choose $b' \neq a'$ such that $q := p \cup \{(a, b')\}$ is a condition (im(p) is finite). Then $q \leq p$ and no extension of q contains (a, a'), so no extension of q forces Raa'. Hence $[Raa']$ is not dense below p.

It is easy to see that $M[G]$ is defined for every generic G (e.g. by Lemma 1.17(1)). By Proposition 1.24 every generic associate $M[G]$ is a generic expansion of M, that is, $a \mapsto a/{\sim_{\mathrm{Th}(G)}}$: $M \cong M[G] \restriction L$. By definition, $M[G]$ interprets R by

$$\Big\{(a/{\sim_{\mathrm{Th}(G)}}, b/{\sim_{\mathrm{Th}(G)}}) \mid \exists p \in G : p \Vdash_{\mathrm{Ca}} Rab\Big\} = \Big\{(a/{\sim_{\mathrm{Th}(G)}}, b/{\sim_{\mathrm{Th}(G)}}) \mid (a, b) \in \bigcup G\Big\}.$$

Thus $a \mapsto a/{\sim_{\mathrm{Th}(G)}}$: $(M, \bigcup G) \cong M[G]$. From this and the fact that G intersects all the sets $D_0, D_1, \ldots$, it easily follows that $\bigcup G$ is an isomorphism from $(A, \preceq)$ onto $(A', \preceq')$. ⌟

Example 1.27 (Feferman forcing) In [**14**] Feferman considers $M = \mathbb{N}$ interpreting the language L that has relation symbols for the graphs of successor, addition and multiplication. L^* expands L by at most countably many unary predicate symbols. A condition p is a finite consistent set of literals in the new predicates $L^* \setminus L$ and constants from $\mathbb{N}$. A condition p extends another q if $p \supseteq q$. For the sets $D_0, D_1, \ldots$ choose, say, always $\emptyset$. Feferman defines a conservative existential pre-forcing $\Vdash_{\mathrm{Fe}}$ by letting p force an atom involving a new predicate if and only if the atom belongs to p. It is not hard to see that $\Vdash_{\mathrm{Fe}}$ is a forcing and that $M[G]$ is defined for every generic G. Applications of Feferman forcing in computability theory are surveyed in [**31**]. ⌟

Variations and generalizations of Feferman forcing have been studied in complexity theory:

Example 1.28 (Generic oracles) In [**15**] Fenner et al. generalize Feferman forcing for the case where $L^* = L \cup \{R\}$ for one new unary predicate R. View a Feferman condition p as the set of functions in $\{0, 1\}^{\mathbb{N}}$ that map $n \in \mathbb{N}$ to 0 or 1 whenever $Rn \in p$ or $\neg Rn \in p$ respectively. Now, instead of using these basic clopen sets as conditions, [**15**] use perfect sets in $\{0, 1\}^{\mathbb{N}}$. Forcing frames considered in [**15**] are certain subframes of this forcing frame (cf. [**15**, Definition 3.3]). Straightforwardly, Fenner et al. let a perfect set p force an atom Rn if and only if every function in p maps n to 1. This determines a conservative existential pre-forcing, that is actually a forcing on the frames considered. For various frames, Fenner et al. study complexity classes relativized by R in generic expansions. ⌟

Finally Robinson developed forcing in model theory:

Example 1.29 (Finite Robinson forcing) We degrade M to a set of constants, i.e., we let $L = \emptyset$. Further let L^* be a countable language and T be a consistent L^*-theory; $T_\forall$ is the set of universal consequences of T.

Define the following forcing frame $(P, \leq, D_0, D_1, \ldots)$. A condition p is a finite set of $L^*(M)$-literals such that $T_\forall \cup p$ is consistent. Define $p \leq q$ to mean $p \supseteq q$. Finally, let $D_0, D_1, \ldots$ enumerate the sets $\bigcup_{a \in M}[t = a]$ for closed $L^*(M)$-terms t. It is easy to see that these sets are dense in P.

To define an existential pre-forcing $\Vdash_{\mathrm{Ro}}$, it suffices to define $p \Vdash_{\mathrm{Ro}} \varphi$ for atomic φ. Take this to mean $T_\forall \cup p \vdash \varphi$.

Then $\Vdash_{\mathrm{Ro}}$ is a forcing: Extension for atoms is obvious. To verify Stability for atoms, argue indirectly and assume $p \nVdash_{\mathrm{Ro}} \varphi$ where φ is an atom. Then $q := p \cup \{\neg\varphi\}$ is a condition. Clearly no extension of q forces φ, so $[\varphi]$ is not dense below p.

By Theorem 1.22, $M[G]$ is defined for every generic G and satisfies $T_\forall$. Note $\bigcup G$ is roughly the same as $\mathrm{Diag}(M[G])$. Hence the Truth Lemma essentially[2] says, that generic associates are *finitely generic for* T, so in particular such structures exist ([**18**, Theorem 5.11]). Their theory can be seen as a generalized model-companion for T. We refer to the book [**18**] for more information. Keisler [**21**] gives some more model-theoretic and algebraic applications of Robinson forcing and some variants of it. $\lrcorner$

1.6 Weak forcing

This section is not needed in the following. In [**42**] Shoenfield develops Cohen forcing in an indirect way: as an intermediate step he introduces an existential forcing $\Vdash$ and verifies the Truth Lemma for it. The actual forcing wanted, namely one satisfying Forcing Completeness, is then obtained as the *weak forcing* $\Vdash^*$:

$$p \Vdash^* \varphi \text{ if and only if } p \Vdash \neg\neg\varphi.$$

Cohen forcing, and more generally, any universal forcing coincides with its weak version (Lemmas 1.3(1) and 1.5(3)). In other contexts weak forcings play a more important role [**2, 15, 18, 21, 31, 44**]. Often, starting with a particular existential forcing one verifies certain desired properties for the corresponding weak forcing. We use the opportunity of having a more general set-up and include a short discussion of the two notions.

Example 1.30 (Keisler forcing) In [**21**] Keisler studies generally existential pre-forcings that satisfy Extension for atoms and the conditions in Lemma 1.17(1), and proves Forcing Completeness for $\Vdash^*$ [**21**, Corollary 1.6]. In a similar context, Stern notes universal recurrence for $\Vdash^*$ [**44**, Proposition 1-1]. $\lrcorner$

Proposition 1.31 *Assume* $\Vdash$ *is a universal or existential pre-forcing satisfying Extension for atoms. Then:*

(1) $\Vdash \subseteq \Vdash^*$.

(2) $\Vdash^*$ *satisfies Stability, i.e.,* $(\Vdash^*)^* = \Vdash^*$.

(3) *If* $M[G]$ *is defined for every generic* G*, then* $\Vdash^*$ *is a universal forcing.*

Proof. Recall Remark 1.7. By Lemma 1.3(1)

$$\{p \mid p \Vdash^* \varphi\} = \{p \mid [\varphi] \text{ is dense below } p\} = \overset{\circ}{\overline{[\varphi]}}.$$

Part (1) is clear as the sets $[\varphi]$ for $\varphi \in L^*(M)$, are open and trivially $X \subseteq \overset{\circ}{\overline{X}}$ for open X. Part (2) then follows from

$$\overset{\circ}{\overline{\overset{\circ}{\overline{X}}}} \subseteq \overset{\circ}{\overline{X}}.$$

For (3) we have to show that $\Vdash^*$ satisfies Extension and Stability for atoms and satisfies universal forcing recurrence. But $\Vdash^*$ satisfies Extension (for all sentences of the forcing language) as sets of the form $\overset{\circ}{\overline{X}}$ are open. Further $\Vdash^*$ satisfies Stability by (2). So $\Vdash^*$ is a forcing. To show $\Vdash^*$ satisfies universal recurrence we first observe:

Claim 1 $\Vdash^*$ satisfies Forcing Completeness, i.e., $p \Vdash^* \varphi$ if and only if $M[G] \models \varphi$ for every generic G containing p.

[2] The forcing used in [**18**] is slightly different from $\Vdash_{\mathrm{Ro}}$ as defined here.

Proof of Claim 1*:* We infer from (2) that $\Vdash^*$ satisfies Lemma 1.5(4) as seen in the proof there. The claim then follows as in the proof of Corollary 1.19(2). $\square$

The claim implies that $\Vdash^*$ satisfies the $\forall$-clause and the $\wedge$-clause of universal recurrence. The $\neg$-clause for $\Vdash^*$ follows immediately from the $\neg$-clause for $\Vdash$. $\square$

Corollary 1.32 *Assume $\Vdash$ is a universal forcing and $\Vvdash$ is an existential forcing such that $\Vdash$ and $\Vvdash$ agree on atoms of the forcing language. If further $M[G]$ is defined for every generic G, then $\Vdash = \Vvdash^*$.*

Proof. As $\Vvdash$ is a forcing, we have for atomic φ: $\{p \mid p \Vvdash \varphi\} = \{p \mid \{q \mid q \Vvdash \varphi\}$ is dense below $p\} = \{p \mid p \Vvdash \neg\neg\varphi\} = \{p \mid p \Vvdash^* \varphi\}$. Thus $\Vvdash$ and $\Vvdash^*$ agree on atoms and hence so do $\Vvdash^*$ and $\Vdash$. By Proposition 1.31(3), $\Vvdash^*$ satisfies universal recurrence, so $\Vvdash^* = \Vdash$. $\square$

Proposition 1.33 *Assume $\Vdash$ is an existential pre-forcing such that there are p_0, φ_0 such that $p_0 \nVdash \varphi_0$ and $p_0 \nVdash \neg\varphi_0$. Then $\{\varphi \mid p_0 \Vdash \varphi\}$ is not closed under logical consequence. If $\Vdash$ satisfies Extension for atoms and $M[G]$ is defined for every generic G, then $\Vdash$ does not satisfy Stability.*

Proof. By assumption, p_0 does not force $(\varphi_0 \vee \neg\varphi_0)$. This is valid, so $\{\varphi \mid p_0 \Vdash \varphi\}$ is not closed under logical consequence. If $\Vdash$ satisfies Extension for atoms and $M[G]$ is defined for every generic G, then $\Vdash^*$ is a universal forcing by Proposition 1.31(3), so by Corollary 1.20 every valid sentence is weakly forced by every condition. Hence $\Vdash \neq \Vdash^*$ and $\Vdash$ does not satisfy Stability . $\square$

Remark 1.34 (Universal versus existential forcing) Intuitively, Corollary 1.20 says that universal forcing refers to the meaning of a sentence, not to its syntax. In contrast existential forcing is syntax sensible, if not trivial (Proposition 1.33), and Forcing Completeness fails. Informally, existential forcing has defects and these defects may be repaired when moving to the weak forcing (Proposition 1.31).

1.7 Summary

To sum up, given an L-model M and $L^* \supseteq L$, one specifies a forcing frame $(P, \leq, D_0, D_1, \ldots)$, a relation $\Vdash$ between conditions and *atoms* of the forcing language that satisfies Extension and Stability (for atoms, cf. Definition 1.4).

Then (universal or existential) forcing recurrence determines a (universal or existential) forcing $\Vdash$. For every generic G the generic associate $M[G]$, if defined, satisfies the Truth Lemma, i.e., in $M[G]$ is true exactly what is forced by some condition in G.

Moreover, to get a conservative forcing (Definition 1.23) it suffices to specify $\Vdash$ only for $L^*(M) \setminus L(M)$-atoms. In this case, $M[G]$ is isomorphic to an L^*-expansion of M.

2 Principal theorems

In set theory one usually considers the case where $M[G]$ is an extension of a model M of ZF (Example 1.25). Independence results are based on the “Principal Theorem” **[42]** stating that every generic extension $M[G]$ models ZF.

In weak theories of arithmetic one is often interested in constructing generic expansions of a countable nonstandard model M of true arithmetic (cf. Introduction). To get relativized independence results one needs the generic expansions to model some weak arithmetic. This boils down to the question of when generic expansions satisfy certain least number principles.

In this section we fix

- a countable forcing frame $(P, \leq, D_0, D_1, \ldots)$;
- a conservative universal forcing $\Vdash$;
- an ordered countable L-structure M satisfying the least number principle (defined below);
- a countable language $L^* \supset L$.

A model is *ordered* if it interprets the symbol $<$ by some linear order on its universe. Given an ordered model N and $b_0 \in N$, the quantifiers $\forall x < b_0$ and $\exists x < b_0$ are called b_0*-bounded.*

Remark 2.1 Due to conservativity, forcing recurrence works for bounded quantifiers as it does for unbounded quantifiers:

$$\begin{array}{r|l} p \Vdash \forall x < b_0 \chi(x) & \text{iff} \quad \forall a <^M b_0 : p \Vdash \chi(a) \\ p \Vdash \exists x < b_0 \chi(x) & \text{iff} \quad p \Vdash \neg\forall x < b_0 \neg\chi(x) \end{array}$$

Note, $p \Vdash \exists x < b_0 \chi(x)$ if and only if $\bigcup_{a <^M b_0} [\chi(a)]$ is dense below p.

Definition 2.2 Let N be an ordered model, $b_0 \in N$ and Φ be a set of formulas in the language of N with parameters from N.

(a) N satisfies the *least number principle for* Φ if every nonempty subset of its universe that is definable by a formula in Φ has a $<^N$-least element.

(b) N satisfies the least number principle for Φ *up to* b_0 if it satisfies the least number principle for $\{(\varphi(x) \wedge x < b_0) \mid \varphi(x) \in \Phi\}$.

We omit reference to Φ, if it is the set of all formulas in the language of N with parameters from N.

2.1 Partial definability

Recall Examples 1.6, 1.21: a condition p is compatible with φ, written $p \| \varphi$, if p does not force $\neg\varphi$, or equivalently, if some extension of p forces φ.

Definition 2.3 Let $b_0 \in M$ and $\varphi = \varphi(\overline{x})$ be an $L^*(M)$-formula.

(a) $\Vdash$ is *definable for* φ if for every $p \in P$ the set $\{\overline{a} \mid p \| \varphi(\overline{a})\}$ is definable in M.

(b) $\Vdash$ is *densely definable for* φ *up to* b_0 if for every $p \in P$ there is $q \leq p$ such that $\{\overline{c} <^M b_0 \mid q \| \varphi(\overline{c})\}$ is definable in M.

We say $\Vdash$ is (densely) definable (up to b_0) for a set Φ of $L^*(M)$-formulas if $\Vdash$ is (densely) definable (up to b_0) for every $\varphi \in \Phi$.

Here, for $\overline{c} = c_1 \cdots c_k$ by $\overline{c} <^M b_0$ we mean $c_i <^M b_0$ for every $1 \leq i \leq k$.

Lemma 2.4 *Let $b_0 \in M$ and Φ be a set of $L^*(M)$-formulas that is closed under negations. Then*

(1) *$\Vdash$ is definable for Φ if and only if for every $\varphi(x) \in \Phi$ and $p \in P$ the set $\{\overline{c} \mid p \Vdash \varphi(\overline{c})\}$ is definable in M.*

(2) *$\Vdash$ is densely definable for Φ up to b_0 if and only if for every $\varphi(x) \in \Phi$ and $p \in P$ there is $q \leq p$ such that $\{\overline{c} <^M b_0 \mid q \Vdash \varphi(\overline{c})\}$ is definable in M.*

Proof. For the forward directions note $p \Vdash \varphi$ if and only if $p \Vdash \neg\neg\varphi$ (by Stability) if and only if $p \not\| \neg\varphi$. For the backward directions note $p \| \varphi$ if and only if $p \not\Vdash \neg\varphi$. □

Recall that, by conservativity, every generic associate is a generic expansion (Proposition 1.24).

Theorem 2.5 (Principal) *Let $b_0 \in M$ and Φ be a set of $L^*(M)$-formulas. If $\Vdash$ is densely definable for Φ up to b_0, then every generic expansion of M satisfies the least number principle for Φ up to b_0.*

In particular, if $\Vdash$ is definable for Φ, then every generic expansion of M satisfies the least number principle for Φ.

Proof. The second statement follows from the first noting that definability implies dense definability up to any $b_0 \in M$. To prove the first, let $M[G]$ be a generic expansion of M and $\varphi(x) \in \Phi$ be such that $M[G] \models \exists x < b_0\varphi(x)$. We look for a least element in the set defined by $\varphi(x)$ in $M[G]$. It suffices to find $a <^M b_0$ such that $M[G] \models \varphi(a)$ and $M[G] \not\models \varphi(b)$ for every $b <^M a$. Define

$$D_\varphi := \bigcup_{a<^M b_0} \bigcap_{b<^M a} [(\varphi(a) \wedge \neg\varphi(b))].$$

Claim 2 D_φ is dense below every condition forcing $\exists x < b_0\varphi(x)$.

Proof of Claim 2. Given p forcing $\exists x < b_0\varphi(x)$ we are looking for some $q \leq p$ in D_φ. By universal recurrence $\bigcup_{a\in M}[a < b_0 \wedge \varphi(a)]$ is dense below p. By conservativity each set $[a < b_0 \wedge \varphi(a)]$ equals $[\varphi(a)]$ or $\emptyset$ depending on whether $a <^M b_0$ or not. Hence $\bigcup_{a<^M b_0}[\varphi(a)]$ is dense below p, so for some $b <^M b_0$ there is an extension $q_b \leq p$ forcing $\varphi(b)$.

By dense definability applied to $\varphi \in \Phi$ and $q_b \in P$ we find some $\tilde{q} \leq q_b$ such that

$$C := \{c <^M b_0 \mid \tilde{q} \nVdash \neg\varphi(c)\}$$

is definable in M. By Extension $\tilde{q} \Vdash \varphi(b)$, so $\tilde{q} \nVdash \neg\varphi(b)$ by Consistency. Hence $b \in C$, so $C \neq \emptyset$. Because M satisfies the least number principle, C has a least element $a \leq^M b <^M b_0$. As $a \in C$ we have $\tilde{q} \nVdash \neg\varphi(a)$, so by forcing recurrence we find $q_a \leq \tilde{q}$ forcing $\varphi(a)$. Then $q_a \leq \tilde{q} \leq q_b \leq p$. To show $q_a \in D_\varphi$, it suffices to show $q_a \Vdash \neg\varphi(b')$ for every $b' <^M a$. But any $b' <^M a \leq^M b <^M b_0$ is not in C by minimality of a, so $\tilde{q} \Vdash \neg\varphi(b')$ and hence also $q_a \Vdash \neg\varphi(b')$ by Extension. □

Choose $p_0 \in G$ forcing $\exists x < b_0\varphi(x)$ by the Truth Lemma. Note that $D_\varphi \in \mathcal{B}(\Vdash)$ as it is defined by the following formula (with parameters) of the Stern formalism (cf. Lemma 1.9):

$$\exists x\big(x < b_0 \wedge \forall y\big(y < x \to \big(\xi \vdash (\varphi(x) \wedge \neg\varphi(y))\big)\big)\big).$$

The claim and Lemma 1.12 imply that there is a condition $p \in G \cap D_\varphi$. Hence there is $a <^M b_0$ such that for every $b <^M a$ we have $p \Vdash (\varphi(a) \wedge \neg\varphi(b))$. By the Truth Lemma $M[G] \models \varphi(a)$ and $M[G] \models \neg\varphi(b)$ for every $b <^M a$. Thus a is a least element as we are looking for. □

Here is a dual formulation of the Principal Theorem:

Corollary 2.6 *Let $b_0 \in M$ and Φ be a set of $L^*(M)$-formulas. If for every $\varphi(\overline{x}) \in \Phi$ and $p \in P$ there is $q \leq p$ such that*

$$\{\overline{c} <^M b_0 \mid q \Vdash \varphi(\overline{c})\}$$

is definable in M, then every generic expansion of M satisfies transfinite induction for Φ up to b_0*, that is, for every $\varphi(x) \in \Phi$ the sentence*

$$\forall y < b_0(\forall z < y\varphi(z) \to \varphi(y)) \to \forall x < b_0\varphi(x).$$

Proof. The assumption implies that $\Vdash$ is densely definable for $\neg\Phi$ up to b_0 (see the proof of Lemma 2.4). Now observe that the least number principle for $\neg\Phi$ up to b_0 is equivalent to transfinite induction for Φ up to b_0. □

Lemma 2.7

(1) *Let Ψ be the set of $L^*(M)$-formulas φ such that $\Vdash$ is definable for φ. Then Ψ is closed under disjunctions and existential quantification.*

(2) *Let $b_0 \in M$ and Ψ be the set of $L^*(M)$-formulas φ such that $\Vdash$ is densely definable for φ up to b_0. Then Ψ is closed under disjunctions and b_0-bounded existential quantification.*

Proof. (1) and closure under disjunction in (2) follow easily from the recurrence in Example 1.6. We show closure under b_0-bounded existential quantification in (2).

Let $\varphi(y\overline{x}) \in \Psi$ and $p \in P$. We are looking for $q \leq p$ such that $\{\overline{a} <^M b_0 \mid q \| \exists y < b_0 \varphi(y\overline{a})\}$ is definable in M. Because $\varphi \in \Psi$ we find $q \leq p$ such that $\{a\overline{a} <^M b_0 \mid q \| \varphi(a\overline{a})\}$ is definable in M. Then also

$$\{\overline{a} <^M b_0 \mid \exists a <^M b_0 : q \| \varphi(a\overline{a})\}$$

is definable in M. By conservativity $a <^M b_0$ is equivalent with $s \Vdash a < b_0$ for any condition s. Hence the above set equals

$$\{\overline{a} <^M b_0 \mid \exists a \in M : q \| (a < b_0 \wedge \varphi(a\overline{a}))\},$$

and this is the set we want (see the recurrence in Example 1.6). □

2.2 Definable antichains

We sketch a method to establish dense definability. We are going to apply it in the next section. The method is intended for the typical situation where P is an (in general undefinable) subset of M and there are $L(M)$-formulas $\varphi(x,y), \psi(x,y)$ such that, for all $p, q \in P$,

$$(p \leq q \iff M \models \varphi(p,q)) \quad \text{and} \quad (p \| q \iff M \models \psi(p,q)).$$

In this case, the following two lemmas reduce dense definability of forcing to the definability of predense antichains refining given definable antichains.

We recall some standard forcing terminology: an *antichain* is a set of pairwise incompatible conditions. An antichain A is *maximal in* $X \subseteq P$ if $A \subseteq X$ and every $p \in X$ is compatible with some element of A. A set $X \subseteq P$ is *predense* (*below* p) if every condition (extending p) is compatible with some condition in X. E.g. an antichain is predense if and only if it is maximal in P. We write

$$X \downarrow q := \{p \in X \mid p \leq q\} \quad \text{and} \quad X \downarrow Y := \textstyle\bigcup_{q \in Y} X \downarrow q.$$

The method is based on the simple observation that in order to define the forcing for some φ it suffices to define a maximal antichain in $[\varphi]$:

Lemma 2.8 *If $p \leq q$ and X is a maximal antichain in $[\varphi] \downarrow q$, then $p \| \varphi$ if and only if p is compatible with some condition in X.*

Proof. If $p \| \varphi$, then there is $r \in [\varphi]$ extending p. Then $r \in [\varphi] \downarrow q$ since $r \leq p \leq q$. By maximality of X, r is compatible with some condition in X, and hence, as $r \leq p$, so is p. The converse is immediate by Extension. □

To find maximal antichains we intend to proceed by induction on φ. How to get, say, a maximal antichain in $[\neg\varphi]$ from a maximal antichain X in $[\varphi]$? The next lemma shows that this can be done via a predense antichain *refining* X in the following sense:

Definition 2.9 For $X, Y \subseteq P$ we say X *refines* Y if every condition in X that is compatible with some condition in Y already extends some condition in Y.

Lemma 2.10 *Let φ, ψ be $L^*(M)$-sentences, $\chi(x)$ an $L^*(M)$-formula, $b_0 \in M$ and $p \in P$.*

(1) *If X is a maximal antichain in $[\varphi] \downarrow p$, and $A \subseteq P \downarrow p$ is an antichain that is predense below p and refines X, then $A \setminus (A \downarrow X)$ is a maximal antichain in $[\neg\varphi] \downarrow p$.*

(2) *If X and Y are maximal antichains in $[\neg\varphi] \downarrow p$ and $[\neg\psi] \downarrow p$ respectively, and $A \subseteq P \downarrow p$ is an antichain that is predense below p and refines $X \cup Y$, then $A \setminus (A \downarrow (X \cup Y))$ is a maximal antichain in $[\varphi \wedge \psi] \downarrow p$.*

(3) *If for every $a <^M b_0$, the set X_a is a maximal antichain in $[\neg\chi(a)] \downarrow p$, and $A \subseteq P \downarrow p$ is an antichain that is predense below p and refines $\bigcup_{a <^M b_0} X_a$, then $A \setminus (A \downarrow \bigcup_{a <^M b_0} X_a)$ is a maximal antichain in $[\forall x < b_0 \chi(x)] \downarrow p$.*

Proof. We only show (3). Obviously $A' := A \setminus (A \downarrow \bigcup_{a <^M b_0} X_a)$ is an antichain in $P \downarrow p$. To see $A' \subseteq [\forall x < b_0 \chi(x)]$, let $q \notin [\forall x < b_0 \chi(x)]$ be given. We claim $q \notin A'$. If $q \notin A$, there is nothing to show, so we assume $q \in A$ and claim $q \in A \downarrow \bigcup_{a <^M b_0} X_a$. Since $q \not\Vdash \forall x < b_0 \chi(x)$ there is $a_0 <^M b_0$ such that $q \not\Vdash \chi(a_0)$ (Remark 2.1). By Lemma 1.5(4) some extension $r \leq q$ forces $\neg\chi(a_0)$. By maximality of X_{a_0}, the condition r, and hence also q, is compatible with some condition in $X_{a_0} \subseteq \bigcup_{a <^M b_0} X_a$. Since $q \in A$ and A refines $\bigcup_{a <^M b_0} X_a$, we get $q \in A \downarrow \bigcup_{a <^M b_0} X_a$.

To see that A' is maximal, let $q \leq p$ force $\forall x < b_0 \chi(x)$. Then q is compatible with some $r \in A$ since A is predense below p. We claim $r \in A'$, i.e., $r \notin A \downarrow \bigcup_{a <^M b_0} X_a$. Otherwise r forces $\neg\chi(a_0)$ for some $a_0 <^M b_0$ by Extension, and thus also $\neg\forall x < b_0 \chi(x)$ (Corollary 1.20). Hence r cannot be compatible with q by Extension and Consistency, a contradiction. □

Corollary 2.11 *Let Φ be a set of $L^*(M)$-sentences and assume P has a maximum 1_P. Assume further that A is a predense antichain such that $A \subseteq [\varphi] \cup [\neg\varphi]$ for every $\varphi \in \Phi$. If ψ, χ are Boolean combinations of sentences from Φ, then*

(1) *$A \cap [\psi]$ is a maximal antichain in $[\psi]$;*

(2) *$A \cap [\neg\psi] = A \setminus (A \cap [\psi])$ is a maximal antichain in $[\neg\psi]$;*

(3) *$A \cap [\psi \wedge \chi] = (A \cap [\psi]) \cap (A \cap [\chi])$ is a maximal antichain in $[\psi \wedge \chi]$.*

Proof. First show by a straightforward induction that $A \subseteq [\psi] \cup [\neg\psi]$ for every Boolean combination ψ of sentences from Φ. This implies (1): to see maximality, observe that any $p \in [\psi]$ must be compatible with some condition in A by predensity, and since such a condition cannot be in $[\neg\psi]$ by Extension and Consistency, it must be in $[\psi]$.

Knowing (1) for ψ and χ, we get (2) and (3) applying Lemma 2.10 for $p := 1_P$: note that, in general, if A is an antichain and $X \subseteq A$, then A refines X, and $A \downarrow X = X$. □

2.3 Full definability

The forcing frame P is *definable in M* if there is a first-order interpretation of $(P, \leq)$ in M.

Examples 2.12 In set theory, Cohen forcing (Example 1.25) uses definable forcing frames. Easton forcing extends Cohen forcing in that it allows the forcing frame to be a proper class in M, i.e., instead of being a set in M it is only assumed to be definable in M. In case the class frame is in a certain sense approximable by set frames, one can define a forcing that satisfies the forcing lemmas and the Principal Theorem (cf. Introduction).

In arithmetic, Feferman forcing (Example 1.27) uses definable forcing frames. This is due to the fact that it starts with the standard model. Simpson [**43**] gives an example of a definable forcing frame starting with a nonstandard model of arithmetic. ⌟

An easy induction shows (as Lemma 1.9):

Lemma 2.13 *Assume that the forcing frame is definable in M and $\Vdash$ is definable for $L^*(M)$-atoms. Then $\Vdash$ is definable for all $L^*(M)$-formulas.*

Then the Principal Theorem implies:

Corollary 2.14 *Assume the forcing frame is definable in M and $\Vdash$ is definable for $L^*(M)$-atoms. Then every generic expansion of M satisfies the least number principle.*

Example 2.15 (Knight's trick) In [**22**] Knight uses a forcing frame $(P, \leq, D_0, D_1, \ldots)$ with $D_0 = D_1 = \cdots = \emptyset$ and pads M with some other sorts such that $(P, \leq)$ becomes interpretable in the padded structure. Then this structure interprets the Stern formalism. Knight uses a conservative existential forcing (on the padded structure).[3] Lemma 1.9 then gives full definability ([**22**, Lemma 2.2]).

To sample one of Knight's applications, her padding becomes superfluous when M is an ω-model of ZFC and the forcing becomes definable in M. Knight shows that any elementary end extension of M by another ω-model has a generic expansion interpreting a universal choice function that preserves the elementary embedding. ⌟

2.4 Forcing and propositional proofs

We give an intuitive summary of the development sofar as a method to establish lower bounds on the size of propositional proofs.

Recall Example 0.1. Let φ be a relational first-order sentence that is true in all finite models and let $\varphi^{<x}$ result from φ by replacing every quantifier Qy by $Qy < x$. Then every $\langle \varphi^{<x} \rangle_m, m \geq 1$, is a tautology. We would like to establish a lower bound on the length of proofs of these tautologies in a given propositional proof system. Assume proofs in our system are sequences of 'lines' with the last line being the formula proved.

Let M be elementary equivalent to some 'standard' L-model $(\mathbb{N}, <, \ldots)$ and contain nonstandard elements. Let L^* extend L by te language of φ. Design a forcing such that $\varphi^{<x}$ is falsified by some nonstandard $n \in M$ in some generic expansion $M[G]$ of M. Define an L^*-formula 'line y in proof z is false' such that any (code of a) proof π of $\langle \varphi^{<x} \rangle_n$ has a 'false' last line. Show in $M[G]$ that the system is sound: if line y in proof z is false, then so is some line $y' < y$.

The art is to construct the forcing frame such that the forcing is densely definable up to b_0 for all sentences 'line i in proof π is false' where $b_0 \in M$ is as large as possible and π is any (code of a) size $<^M b_0$ proof of $\langle \varphi^{<x} \rangle_n$. Typically, the logical complexity of the

[3] To be correct, Knight uses a conservative existential pre-forcing that satisfies Extension but not necessarily Stability for atoms.

formula 'line i in proof π is false' will reflect the logical complexity of the propositional formulas the system operates with as well as the bound b_0.

For every L-term $s(x)$ such that $s^M(n) <^M b_0$, one can then conclude that the function $s^{\mathbb{N}} : \mathbb{N} \to \mathbb{N}$ cannot upper bound the sizes of proofs of the tautologies $\langle \varphi^{<x} \rangle_m, m \geq 1$.

3 Forcing against bounded arithmetic

We define Paris–Wilkie forcing, Riis forcing and Ajtai forcing, prove a definability result for each and give the corresponding independence results.

In this section we fix

- a countable language L containing $\{+, \cdot, 0, 1, <\}$;
- a countable L-structure M that is elementarily equivalent to an L-expansion of $(\mathbb{N}, +, \cdot, 0, 1, <)$;
- $L^* := L \cup \{R\}$ for a new binary relation symbol $R \notin L$.

We fix some notation. For $n \in M$ we write

$$[n] := \{a \in M \mid a <^M n\}.$$

A relation R over M is *bounded (in M)* if there is $b \in M$ such that any component of any tuple in R is $<^M b$. As $\mathbb{N}$ codes every bounded (in $\mathbb{N}$) relation by an element, M codes every definable bounded (in M) relation by an element. If $m \in \mathbb{N}$ is such a code we let

$$\|m\|$$

denote the cardinality of the coded relation. This is not to be confused with

$$|m|$$

denoting $\log(m+1)$ (rounded down). Using the definitions of these functions in the standard model $(\mathbb{N}, +, \cdot, 0, 1, <)$, we get corresponding functions $\|\cdot\|^M$ and $|\cdot|^M$ in M and we shall omit the superscripts.

For arbitrary $n, m \in M$,

$$n <^M m^{o(1)}$$

means that $n^\ell <^M m$ for every $\ell \in \mathbb{N}$.

3.1 Paris–Wilkie forcing

Paris and Wilkie [**32**] gave "the first forcing argument in the context of weak arithmetic" [**23**, p. 278] establishing independence of the pigeonhole principle $\forall x \mathrm{PHP}(R, x)$ from the least number principle for existential formulas. Recall $\mathrm{PHP}(R, x)$ expresses "R is not a bijection from $\{y \mid y \leq x\}$ onto $\{y \mid y < x\}$".

Theorem 3.1 (Paris–Wilkie, 1985) *Let $n \in M$ be such that $[n]$ is infinite. Then M has an L^*-expansion satisfying both $\neg\mathrm{PHP}(R, n)$ and the least number principle for existential $L^*(M)$-formulas.*

Let $n \in M$ with infinite $[n]$. We define a forcing frame

$$(P, \leq, D_0, D_1, \ldots).$$

Note that every finite bijection from a subset of $[n] \cup \{n\}$ onto a subset of $[n]$ is coded by an element in M. We let P be the set of all these codes. Note that P is not definable in M. As partial order we use $p \leq q$ if and only if $p \supseteq q$. Here, and below, we blur the distinction between p and the bijection coded.

The family $D_0, D_1, \dots$ enumerates the (countably many) sets

$$\{p \mid b \in \mathrm{dom}(p)\}, \{p \mid c \in \mathrm{im}(p)\} \quad \text{for } b \leq^M n, c <^M n.$$

To determine a universal pre-forcing $\Vdash_{\mathrm{PW}}$ it suffices to define $p \Vdash_{\mathrm{PW}} \varphi$ for atoms φ. Furthermore, we want a conservative forcing, so it suffices to define $p \Vdash_{\mathrm{PW}} \varphi$ for φ an $L^*(M)$-atom that is not an $L(M)$-atom. Such an atom has the form Rst for closed $L(M)$-terms s, t. We set

$$p \Vdash_{\mathrm{PW}} Rst \iff (s^M, t^M) \in p.$$

It is easy to check (cf. Example 1.26):

Lemma 3.2

(1) $\Vdash_{\mathrm{PW}}$ *is a forcing.*
(2) $M[G]$ *is defined and a generic expansion of* M *for every generic* G.
(3) $M[G]$ *violates* $\mathrm{PHP}(R, n)$ *for every generic* G.

Lemma 3.3 $\Vdash_{\mathrm{PW}}$ *is definable for quantifier free* $L^*(M)$*-formulas.*

We give the proof exemplifying the method of definable antichains from Section 2.2. However, a direct proof would be equally easy. Note that we are in the "typical situation" that we have $L(M)$-formulas $\varphi(x, y), \psi(x, y)$ such that, for all $p, q \in P$,

$$(p \leq q \iff M \models \varphi(p, q)) \quad \text{and} \quad (p \| q \iff M \models \psi(p, q)).$$

E.g. $\psi(x, y)$ is a formula expressing that both x and y code partial bijections that agree on arguments on which they are both defined.

Proof of Lemma 3.3. Let $\varphi = \varphi(\overline{x})$ be a quantifier free $L^*(M)$-formula. For $\overline{c}$ from M let $T(\overline{c})$ be the set of those $a \in M$ that are denoted by some closed term in $\varphi(\overline{c})$. Further let $A_{\overline{c}}$ be the set of all inclusively minimal partial bijections p such that both $\mathrm{dom}(p)$ contains $T(\overline{c}) \cap [n+1]$ and $\mathrm{im}(p)$ contains $T(\overline{c}) \cap [n]$. As $T(\overline{c})$ is finite, $A_{\overline{c}} \subseteq P$. It is routine to verify that $A_{\overline{c}}$ is a predense antichain in P and equal to $\alpha(y, \overline{c})(M)$ for a suitable $L(M)$-formula $\alpha(y, \overline{x})$.

For an atom $\psi = \psi(\overline{x})$ occuring in $\varphi(\overline{x})$ we have $A_{\overline{c}} \subseteq [\psi(\overline{c})] \cup [\neg\psi(\overline{c})]$. Further there is an $L(M)$-formula $\xi^\psi(y, \overline{x})$ such that $\xi^\psi(y, \overline{c})(M)$ defines $A_{\overline{c}} \cap [\psi(\overline{c})]$. We find such a formula $\xi^\chi(z, \overline{x})$ for every Boolean combination χ of such atoms following the recursion in Corollary 2.11(2), (3). In particular, we find an $L(M)$-formula $\xi^\varphi(y, \overline{x})$ such that $\xi^\varphi(y, \overline{c})$ defines a maximal antichain in $[\varphi(\overline{c})]$.

Lemma 2.8 (for $q = \emptyset$) implies that $\Vdash_{\mathrm{PW}}$ is definable for $\varphi(\overline{x})$. □

Proof of Theorem 3.1. Choose a generic G (Lemma 1.11). Up to isomorphism, then $M[G]$ expands M and violates $\mathrm{PHP}(R, n)$ (Lemma 3.2). By Lemmas 3.3 and 2.7, $\Vdash_{\mathrm{PW}}$ is definable for existential $L^*(M)$-formulas. By the Principal Theorem 2.5, $M[G]$ satisfies the least number principle for these formulas. □

3.2 Riis forcing

One may wonder what in the above proof is special about the pigeonhole principle. Riis pointed out that essentially what is needed is that the principle fails in the infinite

[38, 39]. He uses existential forcing and allows for certain infinite conditions. The point is that the new forcing frame allows to define the forcing for more formulas.

For an $L(M)$-formula $\varphi_0(x, y)$, let

$$(R : \hat{y}\varphi_0(x, y) \sim [x])$$

be a formula expressing "R is a bijection from $\{y \mid \varphi_0(x, y)\}$ onto $[x]$". This is an $L^*(M)$-formula with free variable x.

Definition 3.4 An $L(M)$-formula $\varphi_0(xy)$ *defines an* $n^{\Omega(1)}$ *family in* M if there are $\ell \in \mathbb{N}$ and an $L(M)$-formula $\sigma(yz, x)$ such that for every $n \in M$, $\sigma(yz, n)(M)$ is a surjection from $(\varphi_0(n, y)(M))^\ell$ onto $[n]$ provided $\varphi_0(n, y)(M)$ is nonempty.

If every $\varphi(ny)(M), n \in M$, is bounded and, say, coded by $c_n \in M$, then defining a $n^{\Omega(1)}$-family means that there is an $\ell \in \mathbb{N}$ such that $n \leq^M \|c_n\|^\ell$ for every $n \in M$. For example, $b_0 <^M n^{o(1)}$ if and only if $(x = n \wedge y < b_0)$ does not define a $n^{\Omega(1)}$-family in M.

Examples 3.5 If we choose for $\varphi_0(x, y)$ the formula $y \leq x$, then $(R : \hat{y}\varphi_0(n, y) \sim [n])$ becomes $\neg\mathrm{PHP}(R, n)$. If we choose $y < x{\cdot}x$, then our formula negates the weak pigeonhole principle with n^2 pigeons and n holes. Choosing $y = y$ we negate the cardinal principle (cf. **[20]**).

If we assume that the L-structure $(\mathbb{N}, +, \ \cdot\ , 0, 1, <, \ldots)$ additionally interprets an infinite unary predicate U and a binary relation $E \subseteq U \times U$, then $(R : \hat{y}\varphi_0(n, y) \sim [n])$ with $\varphi_0(x, y) := Uy$ expresses that R copies the infinite directed graph (U, E) to the new universe $[n]$ ("Finitization" **[38]**).

These examples define $n^{\Omega(1)}$ families in M. ⌟

Let $\Sigma_1^{b_0}(R)$ denote the closure of the set of quantifier-free $L^*(M)$-formulas by existential and b_0-bounded quantification (i.e., quantifiers of the form $\exists x < b_0$ and $\forall x < b_0$, cf. page 297).

Theorem 3.6 (Riis, 1993) *Let* $\varphi_0(xy)$ *define an* $n^{\Omega(1)}$ *family in* M *and let* $b_0, n \in M$ *be such that* $b_0 <^M n^{o(1)}$. *Then* M *has an* L^**-expansion satisfying both* $(R : \hat{y}\varphi_0(ny) \sim [n])$ *and the least number principle for* $\Sigma_1^{b_0}(R)$.

Remark 3.7 The reader familiar with bounded arithmetic will notice the following. Use Buss' language for L and choose n and b_0 such that both $b_0 <^M n^{o(1)}$ and $|n| <^M b_0^{o(1)}$. By the second inequality $M \models |t(n)| < b_0$ for every (parameter free) L-term $t(x)$ and hence $\Sigma_1^{b_0}(R)$ includes all $\Sigma_1^b(R)$ formulas with parameters bounded by some L-term in n. Thus the restriction of the expansion to the corresponding cut is a model of $T_2^1(R)$ and $(R : \hat{y}\varphi_0(ny) \sim [n])$.

Let $\varphi_0(xy)$ and $b_0, n \in M$ accord the assumption of Theorem 3.6. We prove the theorem only for the case where $[b_0]$ is infinite. In case $[b_0]$ is finite, b_0-bounded quantifiers can be eliminated and one can argue as in the last section.

Definition 3.8 A relation R over M is ℓ*-small* if it is empty or there are $\ell \in \mathbb{N}$ and an $L(M)$-definable surjection from $[b_0]^\ell$ onto R. R is *small* if it is ℓ-small for some $\ell \in \mathbb{N}$.

Then $[n]$ is not small and neither is

$$A_0 := \varphi_0(ny)(M).$$

Here, and only here, we use the assumption that $\varphi_0(xy)$ defines a $n^{\Omega(1)}$-family in M.

We define the forcing. An ℓ-small bijection from a subset of A_0 onto a subset of $[n]$ is $L(M)$-definable and bounded in M, and hence coded by an element of M. Let $P_\ell \subseteq M$ be the set of all these codes. The set of conditions is

$$P := \bigcup_{\ell \in \mathbb{N}} P_\ell .$$

Again we set $p \leq q$ if $p \supseteq q$, and let the family $D_0, D_1, \ldots$ enumerates the sets $\{p \mid a \in \operatorname{dom}(p)\}, \{p \mid c \in \operatorname{im}(p)\}$ for $a \in A_0, c \in [n]$.

The forcing relation is defined as in the previous section: $p \Vdash_{\mathrm{Ri}} Rst$ if and only if $(s^M, t^M) \in p$. This uniquely determines a conservative universal pre-forcing, and in fact a forcing (cf. Example 1.26).

Lemma 3.9 *Let $\ell \in \mathbb{N}$.*

(1) *$P_\ell \subseteq P_{\ell+1} \subseteq P \subseteq M$.*
(2) *P_ℓ is $L(M)$-definable.*
(3) *If $p, q \in P_\ell$, then $p \cup q \in P_{\ell+1}$ and $p \cap q \in P_\ell$.*
(4) *The sets $D_0, D_1, \ldots$ are dense.*

Proof. We only show (4). Observe that both the domain and range of a condition $p \in P$ are small. As neither A_0 nor $[n]$ is small, both $(A_0 \setminus \operatorname{dom}(p))$ and $([n] \setminus \operatorname{im}(p))$ are infinite. Then (4) follows easily. □

Lemma 3.10 $\Vdash_{\mathrm{Ri}}$ *is definable for $\Sigma_1^{b_0}(R)$.*

This implies the theorem:

Proof of Theorem 3.6. Clearly, $M[G]$ is defined for every generic G. By Proposition 1.24 all generic associates of M are generic expansions and it should be clear that they all satisfy $(R : \hat{y}\varphi_0(ny) \sim [n])$. Thus Theorem 3.6 follows from Lemma 3.10 and the Principal Theorem. □

To prove Lemma 3.10 we rely on the following lemma. It can be shown following the proof of Lemma 3.3 in the previous section.

Lemma 3.11 *For every quantifier free $L^*(M)$-formula $\varphi(\overline{x})$ there is an $L(M)$-formula $\Diamond^\varphi(y, \overline{x})$ such that for every $p \in P$*

$$\Diamond^\varphi(p, \overline{x})(M) = \{\overline{c} \mid p \| \varphi(\overline{c})\}.$$

Proof of Lemma 3.10. For variable tuples $\overline{x} = x_1 \cdots x_\ell, \overline{y} = y_1 \cdots y_\ell$ let $Q\overline{xy}$ abbreviate the quantifier prefix

$$\forall x_1 < b_0 \exists y_1 \ \forall x_2 < b_0 \exists y_2 \ \cdots \ \forall x_\ell < b_0 \exists y_\ell .$$

It suffices to show that $\Vdash_{\mathrm{Ri}}$ is definable for every formula of the form $Q\overline{xy}\varphi$ where φ is a quantifier free $L^*(M)$-formula (by Corollary 1.20 since $M[G]$ is defined for every generic G). Fix a quantifier free $L^*(M)$-formula $\varphi(\overline{z})$. Define the formula

$$\Box^\varphi(y, \overline{z}) := \neg \Diamond^{\neg\varphi}(y, \overline{z}).$$

By Lemma 3.11 and Stability, we have for every condition $p \in P$

$$\Box^\varphi(p, \overline{z})(M) = \{\overline{c} \mid p \Vdash_{\mathrm{Ri}} \varphi(\overline{c})\}.$$

For a tuple $\overline{c}$ from M let $A_{\overline{c}}$ be the predense antichain as defined in the proof of Lemma 3.3. In particular, $A_{\overline{c}} \subseteq [\varphi(\overline{c})] \cup [\neg\varphi(\overline{c})]$ and $A_{\overline{c}} \cap [\varphi(\overline{c})]$ is a maximal antichain in $[\varphi(\overline{c})]$. Further $A_{\overline{c}} \subseteq P_1$ since every condition in $A_{\overline{c}}$ is finite and we assumed that $[b_0]$ is infinite.

For any two tuples $\overline{x}, \overline{y}$ of variables of the same length ℓ we show the following: for all $p \in P$ and all $\overline{c}'$ there is $q \in P_{\ell+1}, q \| p$ such that

$$\text{(3.1)} \qquad \text{if } p \| Q\overline{xy}\ \varphi(\overline{x}, \overline{y}, \overline{c}'), \text{ then } M \models Q\overline{xy}\ \Box^{\varphi}(p \cup q, \overline{x}, \overline{y}, \overline{c}'),$$

where $\overline{c}'$ ranges over assignments to the free variables $\overline{z}'$ of $Q\overline{xy}\ \varphi$. It is not hard to see that then

$$\exists u(u \in P_{\ell+1} \wedge u \| p \wedge Q\overline{xy}\ \Box^{\varphi}(p \cup u, \overline{x}, \overline{y}, \overline{z}'))$$

defines $\{\overline{c}' \mid p \| Q\overline{xy}\ \varphi(\overline{x}, \overline{y}, \overline{c}')\}$ in M. Here "$u \in P_{\ell+1}$" is an $L(M)$-formula according Lemma 3.9(2) and "$x \| y$" is an $L(M)$-formula expressing compatibility as in the previous section.

We prove (3.1) by induction on ℓ. The base case, $\ell = 0$, is easy: if $p \| \varphi(\overline{c})$, then p is compatible with some $q \in A_{\overline{c}} \cap [\varphi(\overline{c})]$ by Lemma 2.8. But $A_{\overline{c}} \subseteq P_1$ and $p \cup q \Vdash_{\mathrm{Ri}} \varphi(\overline{c})$.

For the inductive step, let $x\overline{x}, y\overline{y}$ be length $\ell + 1$ tuples of variables, let $\overline{c}'$ range over assigments to the free variables $\overline{z}'$ in $Qx\overline{x}y\overline{y}\ \varphi$ and write $\varphi = \varphi(x\overline{x}, y\overline{y}, \overline{z}')$.

Let $p \in P$ and $\overline{c}'$ be such that $p \| Qx\overline{x}y\overline{y}\ \varphi(x\overline{x}, y\overline{y}, \overline{c}')$, i.e., there is $\tilde{p} \leq p$ forcing $\forall x < b_0 \exists y\ Q\overline{xy}\ \varphi(x\overline{x}, y\overline{y}, \overline{c}')$. Using universal recurrence and Remark 2.1, this is easily seen to be equivalent to: for every $a <^M b_0$ and every $q \leq \tilde{p}$ there is $b \in M$ such that $q \| Q\overline{xy}\ \varphi(a\overline{x}, b\overline{y}, \overline{c}')$. By induction we get for every $a <^M b_0$ and every $q \leq \tilde{p}$

$$M \models \exists u(u \in P_{\ell+1} \wedge u \| q \wedge \exists y\ Q\overline{xy}\ \Box^{\varphi}(q \cup u, a\overline{x}, y\overline{y}, \overline{c}')).$$

Let $\psi(z)$ be the formula

$$\exists u\big((u = \emptyset \vee \exists v\text{“}v \text{ is a surjection from } [z] \times [b_0]^{\ell+1} \text{ onto } u\text{”}) \\ \wedge u \| \tilde{p} \wedge \forall x < z \exists y\ Q\overline{xy}\ \Box^{\varphi}(\tilde{p} \cup u, x\overline{x}, y\overline{y}, \overline{c}')\big).$$

Claim 3 $M \models \psi(b_0)$.

Proof of Claim 3. It is straightforward to verify $M \models \psi(0)$ and $M \models (\psi(a) \rightarrow \psi(a+1))$ for every $a <^M b_0$. □

By the claim there is $\tilde{q} \in P$ (even in $P_{\ell+2}$) compatible with $\tilde{p}$ such that

$$M \models Qx\overline{x}y\overline{y}\ \Box^{\varphi}(\tilde{p} \cup \tilde{q}, x\overline{x}, y\overline{y}, \overline{c}').$$

As M satisfies the least number principle it defines Skolem functions: there are $L(M)$-definable functions $f\overline{f} = f, f_1, f_2, f_3, \ldots$ such that

$$M \models \forall x\overline{x} < b_0 \Box^{\varphi}\big(\tilde{p} \cup \tilde{q}, x\overline{x}, f(x)\overline{f}(x\overline{x}), \overline{c}'\big).$$

Here, $\overline{f}(x\overline{x})$ is shorthand for $f_1(xx_1)f_2(xx_1x_2)f_3(xx_1x_2x_3)\cdots$. Recall how the antichains $A_{\overline{c}}$ are defined: they consist in all $\subseteq$-minimal conditions whose domain contains $T(\overline{c}) \cap [n+1]$ and whose image contains $T(\overline{c}) \cap [n]$, where $T(\overline{c})$ is the set of things named by closed terms in $\varphi(\overline{c})$. Write $T(\overline{z}) = T(x\overline{x}, y\overline{y}, \overline{z}')$ and set

$$S(\overline{c}') := \bigcup_{a\overline{a} <^M b_0} T(a\overline{a}, f(a)\overline{f}(a\overline{a}), \overline{c}').$$

Let $B_{\overline{c}'}$ be defined for $S(\overline{c}')$ as $A_{\overline{c}}$ is for $T(\overline{c})$. Then $B_{\overline{c}'}$ is an $L(M)$-definable set of $(\ell+2)$-small conditions. Furthermore $B_{\overline{c}'}$ is a predense antichain that refines every $A_{\overline{c}}$, where $\overline{c}$ is of the form $a\overline{a}f(a)\overline{f}(a\overline{a})\overline{c}'$ for some $a\overline{a} <^M b_0$.

Claim 4 If $a\overline{a} <^M b_0$ and $\overline{c} = a\overline{a}f(a)\overline{f}(a\overline{a})\overline{c}'$, then $B_{\overline{c}'} \subseteq [\varphi(\overline{c})] \cup [\neg\varphi(\overline{c})]$.

Proof of Claim 4. Let $a\overline{a} <^M b_0$ and $\overline{c} = a\overline{a}f(a)\overline{f}(a\overline{a})\overline{c}'$. Every $r \in B_{\overline{c}'}$ is compatible with some condition in $A_{\overline{c}}$ by predensity. Since $B_{\overline{c}'}$ refines $A_{\overline{c}}$, r extends some condition in $A_{\overline{c}}$. Since $A_{\overline{c}} \subseteq [\varphi(\overline{c})] \cup [\neg\varphi(\overline{c})]$, also $r \in [\varphi(\overline{c})] \cup [\neg\varphi(\overline{c})]$ by Extension. □

By predensity there is $r \in B_{\overline{c}'}$ such that $(\tilde{p} \cup \tilde{q}) \| r$. Then $r \| p$ and $r \in P_{\ell+2}$, so we are left to check $M \models \forall x\overline{x} < b_0 \Box^{\varphi}\big(p \cup r, x\overline{x}, f(x)\overline{f}(x\overline{x}), \overline{c}'\big)$. By Extension it suffices to verify $M \models \forall x\overline{x} < b_0 \Box^{\varphi}\big(r, x\overline{x}, f(x)\overline{f}(x\overline{x}), \overline{c}'\big)$. But otherwise there is $a\overline{a} <^M b_0$ such that $r \not\Vdash_{\mathrm{Ri}} \varphi(a\overline{a}, f(a)\overline{f}(a\overline{a}), \overline{c}')$. By the last claim, then $r \Vdash_{\mathrm{Ri}} \neg\varphi(a\overline{a}, f(a)\overline{f}(a\overline{a}), \overline{c}')$. But $\tilde{p} \cup \tilde{q} \Vdash_{\mathrm{Ri}} \varphi(a\overline{a}, f(a)\overline{f}(a\overline{a}), \overline{c}')$ (by the choice of $\tilde{q}$), so r cannot be compatible with $\tilde{p} \cup \tilde{q}$ by Extension and Consistency, a contradiction. □

3.3 Ajtai forcing

We prove Ajtai's result [**1**] including its improvements from [**30, 33**]. Compared to Riis' Theorem 3.6 it embodies an exponential improvement concerning the bound b_0, but only concerns b_0-bounded formulas. Citing Zambella [**50**, p. 403], any techniques that can allow to handle $\Sigma_1^{b_0}(R)$ for such big b_0 would be extremely interesting.

Let $\Delta_0^{b_0}(R)$ denote the closure of the set of quantifier-free $L^*(M)$-formulas by b_0-bounded quantification (cf. page 297).

Theorem 3.12 (Ajtai, 1988) *Let $b_0, n \in M$ be such that $|b_0| <^M n^{o(1)}$. Then M has an L^*-expansion satisfying both $\neg\mathrm{PHP}(R, n)$ and the least number principle for $\Delta_0^{b_0}(R)$ up to b_0.*

Remark 3.13 The reader familiar with bounded arithmetic will notice the following. Use Buss' language for L and choose $b_0, n \in M$ such that both $|b_0| < n^{o(1)}$ and $|n| < |b_0|^{o(1)}$. By the second inequality b_0 bounds $t^M(n)$ for every L-term $t(x)$. Thus the restriction of the expansion to the corresponding cut is a model of $T_2(R)$ that violates $\mathrm{PHP}(R, n)$.

Fix some $d \in \mathbb{N}$. Following Section 2.4 it is not hard to infer from Theorem 3.12 that proofs of $\langle \mathrm{PHP}(R, x)\rangle_m, m \in \mathbb{N}$, in depth d Frege systems must have size at least $2^{m^{\epsilon}}$ for some $\epsilon > 0$ (depending on d).

For $m \in \mathbb{N}$ consider the following finite forcing frame $(P(m), \leq)$ (without a family $D_0, D_1, \ldots$): the conditions are all finite partial bijections from $[m+1]$ to $[m]$ and $p \leq q$ means $p \supseteq q$. Again, we blur the distinction of the bijection and its code in $\mathbb{N}$. The *size* $\|p\|$ of a condition p is its cardinality, i.e., the number of pigeons mapped. The *rank* of a set $X \subseteq P(m)$ is the maximal size of a condition in X (and, say, 0 if X is empty).

Now fix M and $n, b_0 \in M$ according the assumptions of Theorem 3.12. Observe that there are uniform definitions of $P(m)$ in the standard model $(\mathbb{N}, +, \cdot, 0, 1, <)$ meaning that there is a $\{+, \cdot, 0, 1, <\}$-formula $\varphi(x, y)$ such that $P(m) = \varphi(m, y)(\mathbb{N})$ for every $m \in \mathbb{N}$. Applied in M, these definitions give forcing frames $(P(m), \leq)$ with size function $\|\cdot\|$ also for (nonstandard) $m \in M$. Further note that M defines the function $m \mapsto m^{\epsilon}$ (rounded up) for any (standard) rational $0 < \epsilon < 1$.

We now define the forcing frame P. It is going to be an undefinable subframe of the definable frame $P(n)$. The set $\{p \in P(n) \mid \|p\| <^M n - n^{\epsilon}\}$ is definable in M for every standard rational $0 < \epsilon < 1$. We let P be the union of all these sets. As usual $p \leq q$ means $p \supseteq q$, and the family $D_0, D_1, \ldots$ enumerates the sets $\{p \in P \mid b \in \mathrm{dom}(p)\}$ and $\{p \in P \mid c \in \mathrm{im}(p)\}$ for $b \leq^M n, c <^M n$. It is easy to see that these sets are dense (in P).

We define the forcing as in the previous two sections: we let $p \in P$ force an atom Rst if $(s^M, t^M) \in p$ and denote by $\Vdash_{\mathrm{Aj}}$ the resulting conservative universal pre-forcing. It is easy to see that $\Vdash_{\mathrm{Aj}}$ is a forcing and that $M[G]$ is defined for every generic G (cf. Section 3.1).

Lemma 3.14 $\Vdash_{\mathrm{Aj}}$ *is densely definable for* $\Delta_0^{b_0}(R)$ *up to* b_0.

Proof of Theorem 3.12. It is clear that every generic associate of M violates $\mathrm{PHP}(R, n)$ and by conservativity it is a generic expansion of M (Proposition 1.24). Thus Theorem 3.12 follows from the above lemma by the Principal Theorem. □

To prove Lemma 3.14 we follow the method of definable antichains from Section 2.2. Note that Lemma 3.14 follows easily from Lemma 2.8 and:

Lemma 3.15 *Let* $p \in P$. *For every* $\varphi(\overline{x}) \in \Delta_0^{b_0}(R)$ *there is* $r \in P, r \leq p$ *and a sequence of sets* $(X_{\overline{a}})_{\overline{a}<^M b_0}$ *in* M *such that for every* $\overline{a} <^M b_0$, *the set* $X_{\overline{a}}$ *is a maximal antichain in* $[\varphi(\overline{a})] \downarrow r$ *of rank at most* $\|r\| + |b_0|$.

Here, by saying that a sequence $(X_{\overline{a}})_{\overline{a}<^M b_0}$ of subsets of M is *in* M we mean that the set $\{(\overline{a}, c) \mid \overline{a} <^M b_0, c \in X_{\overline{a}}\}$ is coded in M.

We are thus left to prove this lemma. To do so we intend to use Lemma 2.10. Therefore we need to define predense antichains refining given sets and it is here where the finite combinatorics enter the argument. The idea is to show that suitable antichains exist in $P(m)$ for $m \in \mathbb{N}$ sufficiently large. Then M codes these antichains for the infinite $P(n)$.

As a first problem, predensity does not make much sense in finite frames nor in $P(n)$. Therefore we shall calibrate the notion in the definition below. Second, suitable antichains need not to exist, but they do exist after restricting attention to conditions that extend a suitably chosen condition r. This choice is done according to the Switching Lemma 3.18 below, the combinatorial core of the argument.

Details follow.

Definition 3.16 Let $m, k \in \mathbb{N}$, $q \in P(m)$ and $X \subseteq P(m)$. Then X is *k-predense (in $P(m)$) below* q if every condition that extends q and has size at most $m - k$ is compatible with a condition in X.

For $m \in M$, $p \in P(m)$ and $X \subseteq P(m)$ write

$$X^p := \{q \setminus p \mid q \in X, p \| q\} \quad \text{and} \quad X \cup p := \{q \cup p \mid q \in X, p \| q\}.$$

Note that $P(m)^p \cong P(m - \|p\|)$ via a size preserving isomorphism. By saying that an antichain is k-predense in $P(m)^p$ we mean that its image under this isomorphism is k-predense in $P(m - \|p\|)$. In the same way k-predensity is explained in $P(n)^p$.

Lemma 3.17 *Let* $X \subseteq P$, $p, q \in P, q \leq p$ *and let* φ *be an* $L^*(M)$*-sentence. If* X *is a maximal antichain in* $[\varphi] \downarrow p$ *and has rank at most* $\|p\| + |b_0|$, *then* $X \cup q$ *is a maximal antichain in* $[\varphi] \downarrow q$.

Proof. As $X \subseteq P \downarrow p$, X^p has rank at most $|b_0|$. Then $X \cup q = X^p \cup q$ has rank at most $\|q\| + |b_0|$, so $X \cup q \subseteq P$. Clearly, $X \cup q$ is an antichain.

To show containment in $[\varphi] \downarrow q$, let $r \in X \cup q$ and choose $s \in X, q\|s$ such that $r = s \cup q$. Since $X \subseteq [\varphi]$, we have $s \cup q \in [\varphi] \downarrow q$ by Extension.

To show maximality, let $r \in [\varphi] \downarrow q$. By maximality of X, r is compatible with some $s \in X$. As $r \leq q$, q is compatible with s. Then $s \cup q \in X \cup q$ is compatible with r. □

Lemma 3.18 (Switching) *Let $\ell, m, k, N \in \mathbb{N}, k \leq \ell$ be sufficiently large and $X_1, \dots, X_N$ be subsets of $P(m)$ of rank at most k. Assume*

$$N^{2/k}\ell^{100} < m. \tag{3.2}$$

Then there is $q \in P(m)$ of size at most $m - \ell$ such that for every $1 \leq i \leq N$ there is an antichain $A_i \subseteq P(m)^q$ refining X_i^q that is k-predense in $P(m)^q$ and has rank at most k.

This lemma can be proved by the probabilistic method or a direct (involved) counting argument. Details can be found in [**23**, Lemma 12.3.10] or in the references pointed out in the Introduction.

Applied in M the Switching Lemma provides us with suitable antichains in restrictions of $P(n)$. The following easy lemma allows to move these antichains to P.

Lemma 3.19 *Let $p \in P$ and $X, Y \subseteq P(n)^p$ have rank at most $|b_0|$.*

(1) *If X is an antichain in $P(n)^p$, then $X \cup p$ is an antichain in P.*
(2) *If X is $|b_0|$-predense in $P(n)^p$, then $X \cup p$ is predense in P below p.*
(3) *If X refines Y in $P(n)^p$, then $X \cup p$ refines $Y \cup p$ in P.*

Proof. We only show (2). Note $X \cup p \subseteq P$ because it has rank at most $\|p\| + |b_0|$. Let $q \in P, q \leq p$ and choose $0 < \epsilon < 1$ such that $\|q\| <^M n - n^\epsilon$. Then $\|q \setminus p\| = \|q\| - \|p\| <^M n - n^\epsilon - \|p\| <^M (n - \|p\|) - |b_0|$. Since $(q \setminus p) \in P(n)^p$ and X is $|b_0|$-predense in $P(n)^p$, there is $r \in X$ such that $q \setminus p$ is compatible with r in $P(n)^p$. Then $q \cup r = q \cup (r \cup p)$ extends both q and $r \cup p \in X \cup p$. As $q \cup r$ has size $<^M \|q\| + |b_0|$ it is in P, so q and $r \cup p$ are compatible in P. □

The rest of the argument is straightforward:

Proof of Lemma 3.15. Let $p \in P$. Call a $L^*(M)$-formula *good* if the lemma holds for it. It is easy to verify that atomic formulas are good: for an atom $\varphi(\overline{x})$ of the form *Rst* with $L(M)$-terms $t = t(\overline{x}), s = s(\overline{x})$ take $r := p$ and define $X_{\overline{a}} := \{r \cup \{(s^M(\overline{a}), t^M(\overline{a}))\}\}$ or $X_{\overline{a}} := \emptyset$ depending on whether $r \cup \{(s^M(\overline{a}), t^M(\overline{a}))\}$ is a partial bijection from $[n+1]$ to $[n]$ or not. Similarly, for an $L(M)$-atom $\varphi(\overline{x})$ set $r := p$ and $X_{\overline{a}} := \{r\}$ or $X_{\overline{a}} := \emptyset$ depending on whether $M \models \varphi(\overline{a})$ or not.

We leave it to the reader to verify that the set of good formulas is closed under conjunctions and negations. As the set of good formulas is closed under logical equivalence (Corollary 1.20), we are thus left to show it is closed under b_0-bounded universal quantification.

So assume $\varphi(x\overline{x})$ is good. Then $\neg\varphi(x\overline{x})$ is good and we can choose $r \in P, r \leq p$ and antichains $(X_{a\overline{a}})_{a\overline{a} <^M b_0}$ that satisfy the claim for $\neg\varphi(x\overline{x})$. Since every antichain $X_{a\overline{a}}$ is in $P \downarrow r$ and has rank at most $\|r\| + |b_0|$, we know that $X^r_{a\overline{a}}$ has rank at most $|b_0|$.

Choose $0 < \epsilon < 1$ such that $\|r\| <^M n - n^\epsilon$. Then $n^\epsilon <^M n - \|r\| =: m$. Observe that as partial orders

$$P(n)^r \cong P(m),$$

via an isomorphism that is definable in M and preserves the size $\|\cdot\|$. For $\overline{a} <^M b_0$ let

$$Y_{\overline{a}} := \bigcup_{a <^M b_0} X^r_{a\overline{a}}.$$

Note $Y_{\overline{a}}$ has rank at most $|b_0|$, and the sequence $(Y_{\overline{a}})_{\overline{a} <^M b_0}$ is in M.

We intend to apply the Switching Lemma in M to get refining antichains for the sequence $(Y_{\overline{a}})_{\overline{a} <^M b_0}$. We check its assumptions: calculated in M, the sequence has length

$N := b_0^{\ell_0}$ for ℓ_0 the length of $\overline{x}$. Especially $N^{2/|b_0|}$ (calculated in M) is bounded by a standard number in M (i.e., by a closed $\{+,1\}$-term). Therefore we can choose a rational $0 < \epsilon' < 1$ (e.g. $\epsilon' := 1/101$) such that the inequality (3.2) of the Switching Lemma is satisfied for $\ell := m^{\epsilon'}$ and $k := |b_0|$ (and m, N as defined above). Further $k = |b_0| <^M (n^\epsilon)^{\epsilon'} <^M m^{\epsilon'} = \ell$.

Thus the lemma applies: we find $r' \in P(n)^r$ of size at most $m - m^{\epsilon'}$ such that, writing $s := (r \cup r')$, the following holds: for every $\overline{a} <^M b_0$ there is an $A_{\overline{a}} \subseteq (P(n)^r)^{r'} = P(n)^s$ coded in M such that, in $P(n)^s$,

(i) $A_{\overline{a}}$ is an antichain that is $|b_0|$-predense,
(ii) $A_{\overline{a}}$ refines $Y_{\overline{a}}^{r'} \subseteq P(n)^s$,
(iii) $A_{\overline{a}}$ has rank at most $|b_0|$.

Note that s has size $\|r\| + \|r'\| \leq^M n - m + m - m^{\epsilon'} <^M n - n^{\epsilon\epsilon'}$, so $s \in P$. Further note that with $Y_{\overline{a}}$ also $Y_{\overline{a}}^{r'}$ has rank at most $|b_0|$. By Lemma 3.19 we get in P:

(iv) $(A_{\overline{a}} \cup s)$ is an antichain that is predense below s,
(v) $(A_{\overline{a}} \cup s)$ refines $Y_{\overline{a}}^{r'} \cup s = \bigcup_{a <^M b_0} (X_{a\overline{a}} \cup s)$,
(vi) $(A_{\overline{a}} \cup s)$ has rank at most $\|s\| + |b_0|$.

By Lemma 3.17 $(X_{a\overline{a}} \cup s)$ is a maximal antichain in $[\neg\varphi(a\overline{a})] \downarrow s$. By $(A_{\overline{a}} \cup s) \subseteq P \downarrow s$, (iv) and (v) the assumptions of Lemma 2.10(3) are satisfied. Thus we get a maximal antichain in $[\forall x < b_0 \varphi(x\overline{a})] \downarrow s$ setting

$$Z_{\overline{a}} := (A_{\overline{a}} \cup s) \setminus \big((A_{\overline{a}} \cup s) \downarrow \bigcup_{a <^M b_0} (X_{\overline{a}} \cup s)\big).$$

Then $(Z_{\overline{a}})_{\overline{a} <^M b_0}$ is in M and has rank at most $\|s\| + |b_0|$ by (vi). □

3.4 Notes

Compared to Riis' original argument [**38**] our proof of Theorem 3.6 relies on the stability of universal forcing (and Corollary 1.20) while Riis uses an existential forcing, and it is simpler in that it sidesteps the analysis of an auxiliary pre-forcing in [**38**].

Compared with other proofs of Theorem 3.12, roughly, the predense antichains in our argument correspond to the complete systems in [**30**] and in [**50**], to branches in shallow decision trees in [**28, 49**] or to the small covers in [**1**].

Forcing type arguments for Ajtai's result have been given in [**1, 50**] and [**23**, Section 12.7] and recently in [**28**]. In [**23**, Section 12.7] Krajíček presents the method of k-evaluations of propositional formulas [**30**] as a forcing type argument. Our proof constructs for certain φ a predense antichain together with its maximal part in $[\varphi]$. These pairs of sets give rise to a modified notion of $|b_0|$-evaluation. As Zambella's [**50**] our argument sidesteps a detour through propositional logic like in [**1, 23, 30, 49**]. Further it avoids the restriction to "internal" generics in [**50**].

Krajíček's recent proof in [**28**] (cf. Introduction) uses forcing with random variables, developed in [**28**] as a general method to construct Boolean valued models of bounded arithmetics. This recent argument, the argument given here and in fact all known arguments for Ajtai's result use the Switching Lemma in one or another form. The main obstacle to generalize Ajtai's argument to other principles is the difficulty to find analogues of this lemma. Our interpretation of the role of this lemma is roughly as follows: it states the existence of refining antichains. The rest of the argument can be taken over by the general machinery, the method of definable antichains and the Principal Theorem as described in Section 2.

Acknowledgements

We thank Achim Blumensath, Sam Buss, Jörg Flum, Sy-David Friedman and Juan-Carlos Martínez for their comments and encouragement at earlier stages of this work.

References

[1] M. Ajtai, The complexity of the pigeonhole principle, *Proceedings of the 29th Annual Symposion on the Foundations of Computer Science* (1988), 346–355.

[2] J. Avigad, Forcing in proof theory, *The Bulletin of Symbolic Logic* **10(3)** (2004), 305–333.

[3] P. Beame, A switching lemma primer, Technical Report UW-CSE-95-07-01, University of Washington, 1994.

[4] P. Beame and T. Pitassi, Propositional Proof Complexity: Past, Present, and Future, *Bulletin of the European Association for Theoretical Computer Science, The Computational Complexity Column*, 65, E. Allender (ed.), 1998,66–89.

[5] S. Bellantoni and T. Pitassi and A. Urquhart, Approximation and small-depth Frege proofs, *SIAM Journal on Computing* **21(6)** (1992).

[6] E. Ben-Sasson and P. Harsha, Lower bounds for bounded depth Frege proofs via Buss–Pudlák games, *ACM Transactions on Computational Logic* **11(3)** (2010).

[7] O. Beyersdorff, On the correspondence between arithmetic theories and propositional proof systems —a survey, *Mathematical Logic Quarterly* **55(2)** (2009), 116–137.

[8] S. R. Buss, *Handbook of Proof Theory*, Studies in Logic and the Foundations of Mathematics 137, Elsevier, 1998.

[9] S. R. Buss, Bounded Arithmetic and Propositional Proof Complexity, *Logic of Computation*, Springer, 1995, 67–122.

[10] S. R. Buss, First-order proof theory of arithmetic, chapter II in *Handbook of Proof Theory*, S. R. Buss (ed.), 1998, 79–147.

[11] S. A. Cook and P. Nguyen, *Logical Foundations of Proof Complexity*, Cambridge University Press, 2010.

[12] S. A. Cook and A. R. Reckhow, The relative efficiency of propositional proof systems, *The Journal of Symbolic Logic* **44(1)** (1979), 36–50.

[13] H.-D. Ebbinghaus and J. Flum, *Finite Model Theory*, Perspectives in Mathematical Logic, Springer, 2nd edition, 1999.

[14] S. Feferman, Some applications of forcing and generic sets, *Fundamentae Mathematicae* **56** (1965), 325–345.

[15] S. Fenner, L. Fortnow, S. A. Kurtz and L. Li, An oracle builder's toolkit, *Information and Computation* **182** (2003), 95–136.

[16] S. D. Friedman, Topics in Set Theory, Course Notes.

[17] J. Hastad, Almost optimal lower bounds for small depth circuits, *Proceedings of the 18th Annual ACM Symposium on Theory of Computing* 1986, 6–20.

[18] J. Hirschfeld and W. H. Wheeler, *Forcing, Arithmetic, Division Rings*, Lecture Notes in Mathematics 454, 1975.

[19] W. Hodges, *Building Models by Games*, Cambridge University Press, 1985.

[20] R. Kaye, The theory of κ-like models of arithmetic, *Notre Dame Journal of Formal Logic* **36(4)**, 1995, 547–559.

[21] H. J. Keisler, Forcing and the omitting types theorem, in: *Studies in Model Theory*, Morley (ed.), Studies in Mathematics, 8, The Mathematical Association of America, 1973, 96–133.

[22] J. F. Knight, Generic expansions of structures, *The Journal of Symbolic Logic* **38(4)** (1973), 561–570.

[23] J. Krajíček, *Bounded Arithmetic, Propositional Logic, and Complexity Theory*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, 1995.

[24] J. Krajíček, On Frege and extended Frege proof systems, in: *Feasible Mathematics II*, P. Clote and J. Remmel (eds.), Birkhäuser, 1995, 284–319.

[25] J. Krajíček, Tautologies from pseudo-random generators, *The Bulletin of Symbolic Logic* **7(2)** (2001), 197–212.

[26] J. Krajíček, Combinatorics of first order structures and propositional proof systems, *Archive of Mathematical Logic* **43** (2004), 427–441.

[27] J. Krajíček, Proof complexity, in: *European Congress of Mathematics (ECM)*, A. Laptev (ed.), Stockholm, Sweden, Zurich: European Mathematical Society, 2005, 221–231.

[28] J. Krajíček, *Forcing with Random Variables and Proof Complexity*, London Mathematical Society Lecture Note Series, Cambridge University Press 382, 2011.

[29] J. Krajíček and P. Pudlák, Propositional Proof systems, the consistency of first order theories and the complexity of computations, *The Journal of Symbolic Logic* **54(3)** (1989), 1063–1079.

[30] J. Krajíček, P. Pudlák and A. Woods, An exponential lower bound to the size of bounded eepth Frege proofs of the pigeonhole principle, *Random Structures and Algorithms* **7(1)** (1995), 15–39.

[31] P. Odifreddi, Forcing and reducibilities, *The Journal of Symbolic Logic* **48(2)** (1983), 288–310.

[32] J. Paris and A. J. Wilkie, Counting problems in bounded arithmetic, *Methods in Mathematical Logic* **1130** (1985), 317–340.

[33] T. Pitassi, P. Beame and R. Impagliazzo, Exponential lower bounds for the pigeonhole principle, *Computational Complexity* **3** (1993), 97–108.

[34] P. Pudlák, A bottom-up approach to foundations of mathematics, *Proceedings Gödel'96, Logical Foundations of Mathematics*, Computer Science and Physics —Kurt Gödel's Legacy, Springer Lecture Notes in Logic **6**, 1996, 81–97.

[35] P. Pudlák, The lengths of proofs, chapter VIII of *Handbook of Proof Theory*, S. R. Buss (ed.), 1998, 547–637.

[36] P. Pudlák and S. R. Buss, How to lie without being (easily) convicted and the lengths of proofs in propositional calculus, *Computer Science Logic'94*, Pacholski and Tiuryn (eds.), Springer Lecture Notes in Computer Science **933**, 1994, 151–162.

[37] A. Razborov, Lower bounds for propositional proofs and independence results in Bounded Arithmetic, *Proceedings of the 23rd ICALP*, Lecture Notes in Computer Science **1099**, 1996, 48–62.

[38] S. Riis, Finitisation in bounded arithmetic, *BRICS Report Series* RS-94-23, 1994.

[39] S. Riis, Making infinite structures finite in models of second order bounded arithmetic, in: *Arithmetic, Proof Theory and Computational Complexity*, Oxford University Press, 1993, 289–319.

[40] D. Scott, A proof of the independence of the continuum hypothesis, *Mathematical Systems Theory* **1(2)** (1967), 89–111.

[41] N. Segerlind, The complexity of propositional proofs, *The Bulletin of Symbolic Logic* **13(4)** (2007), 417–481.

[42] J. R. Shoenfield, Unramified forcing, *Axiomatic Set Theory*, Proceedings of Symposia in Pure Mathematics VIII, American Mathematical Society, 1971, 357–381.

[43] S. G. Simpson, Forcing and models of arithmetic, *Proceedings of the American Mathematical Society* **43(1)** (1974), 193–194.

[44] J. Stern, A new look at the interpolation problem, *The Journal of Symbolic Logic* **40(1)** (1975), 1–13.

[45] G. Takeuti and M. Yasumoto, Forcing on bounded arithmetic, in *Gödel'96*, P. Hájek (ed.), Lecture Notes in Logic **6**, 1996, 120–138.

[46] G. Takeuti and M. Yasumoto, Forcing on bounded arithmetic 2, *The Journal of Symbolic Logic* **63(3)** (1998).

[47] N. Thapen, Notes on switching lemmas, manuscript, 2009.

[48] A. Urquhart, The complexity of propositional proofs, *The Bulletin of Symbolic Logic* **1(4)** (1995), 425–467.

[49] A. Urquhart and X. Fu, Simplified lower bounds for propositional proofs, *Notre Dame Journal of Formal Logic* **73(4)** (1996), 523–544.

[50] D. Zambella, Forcing in finite structures, *Mathematical Logic Quarterly* **43(3)** (1997), 401–412.

Safe recursive set functions

Arnold Beckmann[*], **Samuel R. Buss**[†], **Sy-David Friedman**[‡]

[*] Department of Computer Science, College of Science, Swansea University, UK
a.beckmann@swansea.ac.uk

[†] Department of Mathematics, University of California, San Diego, USA
sbuss@math.ucsd.edu

[‡] Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
sdf@logic.univie.ac.at

Abstract. We introduce the *safe recursive set functions* based on a Bellantoni–Cook style subclass of the primitive recursive set functions. We show that under a natural encoding of finite strings by hereditarily finite sets, the functions computed by safe recursive set functions are exactly the functions computed by alternating exponential time Turing machines with polynomially many alternations.

We characterise the safe recursive set functions on arbitrary sets in definability-theoretic terms. In its strongest form, we show that a function on arbitrary sets is safe recursive if, and only if, it is uniformly definable in some polynomial level of a refinement of Jensen's J-hierarchy, relativised to the transitive closure of the function's arguments.

We observe that safe-recursive functions on infinite binary strings are equivalent to functions computed by so-called infinite-time Turing machines in time less than ω^ω. We also give a machine model for safe recursion which is based on set-indexed parallel processors and the natural bound on running times.

1 Safe recursive set functions

We consider a subclass of the primitive recursive set functions [**10**]. Inspired by Bellantoni and Cook's characterization of the polynomial time computable functions [**1**], we divide arguments of set functions into normal and safe ones. By writing $f(\vec{x} \,/\, \vec{a})$ we indicate that $\vec{x}$ are f's normal arguments, and $\vec{a}$ its safe arguments. Bellantoni and Cook use the notation $f(\vec{x};\vec{a})$ instead of $f(\vec{x} \,/\, \vec{a})$, using semicolon (;) instead of slash (/). We use the slash instead, as we find it improves readability. Set functions whose arguments are typed in this way will be denoted *safe set functions.*

1.1 Safe rudimentary set functions

We first define *safe rudimentary set functions* based on rudimentary set functions [**9**].

Definition 1.1 The set of *safe rudimentary set functions (sRud)* is the smallest class of safe set functions that contains the initial functions (i)–(iii) and is closed under *bounded union* (iv) and *safe composition* (v):

All three authors thank the John Templeton Foundation, Project #13152, for supporting their participation in the CRM Infinity Project at the Centre de Recerca Matemàtica, Barcelona, Catalonia, Spain, during which this project was instigated.

[*] This research was partially done while the author was a visiting fellow at the Isaac Newton Institute for the Mathematical Sciences in the programme “Semantics & Syntax”.

[†] Supported in part by NSF grants DMS-0700533 and DMS-1101228, and by a grant from the Simons Foundation (#208717 to Sam Buss).

[‡] Supported in part by the FWF (Austrian Science Fund) through FWF project number P 22430-N13.

(i) (Projection) $\pi_j^{n,m}(x_1, \ldots, x_n \,/\, x_{n+1}, \ldots, x_{n+m}) = x_j$, for $1 \leq j \leq n+m$, is in *sRud*.
(ii) (Difference) $\mathrm{d}(/\, a, b) = a \setminus b$ is in *sRud*.
(iii) (Pairing) $\mathrm{p}(/\, a, b) = \{a, b\}$ is in *sRud*.
(iv) (Bounded Union) If g is in *sRud*, then

$$f(\vec{x} \,/\, \vec{a}, b) = \bigcup_{z \in b} g(\vec{x} \,/\, \vec{a}, z)$$

is in *sRud*.
(v) (Safe Composition) If $h, \vec{r}, \vec{t}$ are in *sRud*, then

$$f(\vec{x} \,/\, \vec{a}) = h(\vec{r}(\vec{x} \,/) \,/\, \vec{t}(\vec{x} \,/\, \vec{a}))$$

is in *sRud*.

We list a few functions which are definable in *sRud*. Details of the definitions of some of these can also be found in [**9**]. Let (a, b) denote Kuratowski's ordered pair $\{\{a\}, \{a, b\}\}$. The functions pr_ℓ and pr_r extract the first and second element from an ordered pair.

- $\mathrm{Union}(/\, a) = \cup a$ and $\mathrm{Intersec}(/\, a, b) = a \cap b$.

  $\mathrm{Union}(/\, a) = \bigcup_{z \in a} \pi_1^{0,1}(/\, z)$ and $\mathrm{Intersec}(/\, a, b) = c \setminus ((c \setminus a) \cup (c \setminus b))$ for $c = a \cup b$.

- $\mathrm{Succ}(/\, a) = a \cup \{a\}$, $\mathrm{kop}(/\, a, b) = (a, b)$, $\mathrm{pr}_\ell(/\, (a, b)) = a$, $\mathrm{pr}_r(/\, (a, b)) = b$.

  $f(/\, c) = \bigcup_{z \in c} \bigcup_{y \in c} (z \setminus y)$ satisfies $f(/\, (a, b)) = \begin{cases} \{b\} & \text{if } a \neq b \\ \emptyset & \text{otherwise.} \end{cases}$
  Thus $\mathrm{pr}_\ell(/\, c) = \cup(\cup c \setminus f(/\, c))$.
  $g(/\, c) = \cup(c \setminus \{\cup c\})$ satisfies $g(/\, (a, b)) = \begin{cases} \{a\} & \text{if } a \neq b \\ \emptyset & \text{otherwise.} \end{cases}$
  Thus $\mathrm{pr}_r(/\, c) = \cup(\cup c \setminus g(/\, c))$.

- $\mathrm{Cond}_=(/\, a, b, c, d) = \begin{cases} a & \text{if } c = d \\ b & \text{otherwise.} \end{cases}$

  Let $\overline{g}(\,/\, a, c, z) = \bigcup \{a \colon u \in c \backslash z \cup z \backslash c\}$ and $g(\,/\, a, c, z) = a \setminus \overline{g}(/\, a, c, z)$.
  Then $\overline{g}(\,/\, a, c, z) = \begin{cases} a & \text{if } z \neq c \\ \emptyset & \text{otherwise} \end{cases}$ and $g(\,/\, a, c, z) = \begin{cases} a & \text{if } z = c \\ \emptyset & \text{otherwise.} \end{cases}$
  Thus $\mathrm{Cond}_=(/\, a, b, c, d) = g(/\, a, c, d) \cup \overline{g}(/\, b, c, d)$.

- $\mathrm{Cond}_\in(/\, a, b, c, d) = \begin{cases} a & \text{if } c \in d \\ b & \text{otherwise.} \end{cases}$

  Let $h(\,/\, a, c, d) = \bigcup \{g(/\, a, c, z) \colon z \in d\}$ (g as defined for $\mathrm{Cond}_=$) and $\overline{h}(\,/\, b, c, d) = b \backslash h(/\, b, c, d)$.
  Then $h(\,/\, a, c, d) = \begin{cases} a & \text{if } c \in d \\ \emptyset & \text{otherwise} \end{cases}$ and $\overline{h}(\,/\, b, c, d) = \begin{cases} b & \text{if } c \notin d \\ \emptyset & \text{otherwise.} \end{cases}$
  Thus $\mathrm{Cond}_\in(/\, a, b, c, d) = h(/\, a, c, d) \cup \overline{h}(/\, b, c, d)$.

- $\mathrm{Appl}(/\, a, b) = \{y \colon (\exists x \in b)(x, y) \in a\}$.

  Let $g(/\, b, c) = \begin{cases} \{\mathrm{pr}_r(c)\} & \text{if } \mathrm{pr}_\ell(c) \in b \\ \emptyset & \text{otherwise.} \end{cases}$ Then $\mathrm{Appl}(/\, a, b) = \bigcup \{g(/\, b, c) \colon c \in a\}$.

- $\mathrm{Prod}(/\, a, b) = \{(x, y) \colon x \in a, y \in b\} =: a \times b$, by first observing that

  $$f(/\, x, b) = \{(x, y) \colon y \in b\} = \bigcup \big\{\{(x, y)\} \colon y \in b\big\}$$

  is in *sRud*, and then that $\mathrm{Prod}(/\, a, b) = \bigcup \big\{f(x, b) \colon x \in a\big\}$.

1.2 Predicative set recursion

We extend the safe rudimentary set function by a predicative set recursion scheme.

Definition 1.2 The set of *safe recursive set functions (SRSF)* is the smallest class which contains the safe rudimentary set functions and is closed under safe composition, bounded union and the following scheme:

(Predicative Set Recursion) If h is in *SRSF*, then

$$f(x, \vec{y} \,/\, \vec{a}) = h(x, \vec{y} \,/\, \vec{a}, \{f(z, \vec{y} \,/\, \vec{a})\colon z \in x\})$$

is in *SRSF*. Observe that according to our convention for denoting functions, x is a normal argument of f, and $\{f(z, \vec{y} \,/\, \vec{a})\colon z \in x\}$ is substituted at a safe argument of h.

We show that ordinal addition and multiplication are in *SRSF*. We will see later that ordinal exponentiation cannot be defined in *SRSF*. In a set context, let $0, 1, 2, \dots$ denote ordinals in the usual sense, e.g., $0 = \emptyset$ and $1 = \{\emptyset\}$.

- $\mathrm{Add}(x \,/\, a) = \begin{cases} a & \text{if } x = 0 \\ \mathrm{Succ}(/ \bigcup\{\mathrm{Add}(z \,/\, a)\colon z \in x\}) & \text{if } x = \mathrm{Succ}(/ \bigcup x) \\ \bigcup\{\mathrm{Add}(z \,/\, a)\colon z \in x\} & \text{otherwise.} \end{cases}$

  $\alpha + \beta := \mathrm{Add}(\beta \,/\, \alpha)$ satisfies the usual recursive equations for ordinal addition. Observe that for $\alpha + \beta$, β is a normal argument and α a safe argument.

- $\mathrm{Mult}(x, y \,/) = \begin{cases} 0 & \text{if } x = 0 \\ \mathrm{Add}(y \,/ \bigcup\{\mathrm{Mult}(z, y \,/)\colon z \in x\}) & \text{if } x = \mathrm{Succ}(/ \bigcup x) \\ \bigcup\{\mathrm{Mult}(z, y \,/)\colon z \in x\} & \text{otherwise.} \end{cases}$

  $\alpha \cdot \beta := \mathrm{Mult}(\beta, \alpha \,/)$ satisfies the usual recursive equations for ordinal multiplication. Observe that for $\alpha \cdot \beta$, both α and β are normal.

It should be pointed out here that as Mult has no safe arguments we cannot similarly define exponentiation via predicative set recursion, as we did for Add and Mult.

In many situations it will be convenient to define predicates instead of functions. In the following we provide the necessary background for this.

Definition 1.3 (Predicates) A predicate $R(\vec{x} \,/\, \vec{a})$ is in *SRSF* (resp. in *sRud*) if the function

$$\chi_R(\vec{x} \,/\, \vec{a}) = \begin{cases} 1 & \text{if } R(\vec{x} \,/\, \vec{a}) \\ 0 & \text{otherwise} \end{cases}$$

is in *SRSF* (resp. in *sRud*). Recall that 0 and 1 in a set theoretic context denote ordinals.

Examples of predicates in *sRud* are $a \in b$, $a \notin b$, $a = b$, and $a \neq b$ for safe a, b, which can be seen using the safe rudimentary functions $\mathrm{Cond}_\in$ and $\mathrm{Cond}_=$ as provided before.

Predicates can be used to define functions by separation in the usual way. E.g., assume $R(\vec{x} \,/\, \vec{a}, b)$ is a predicate in *SRSF*, and $B(\vec{x} \,/\, \vec{a})$ a function in *SRSF*. Then $f(\vec{x} \,/\, \vec{a}) = \{b \in B(\vec{x} \,/\, \vec{a})\colon R(\vec{x} \,/\, \vec{a}, b)\}$ is a function in *SRSF*. To see this, let

$$sel(\vec{x} \,/\, \vec{a}, b) = \begin{cases} \{b\} & \text{if } R(\vec{x} \,/\, \vec{a}, b) \\ \emptyset & \text{otherwise} \end{cases} = \mathrm{Cond}_=(/\, \emptyset, \{b\}, \chi_R(\vec{x} \,/\, \vec{a}, b), 0).$$

Then $f(\vec{x} \,/\, \vec{a})$ can be defined by bounded union as $\bigcup_{b \in B(\vec{x} \,/\, \vec{a})} sel(\vec{x} \,/\, \vec{a}, b)$.

Proposition 1.4 (Closure Properties of Predicates) *Predicates in SRSF (in sRud, resp.) are closed under Boolean operations and bounded quantification over safe arguments.*

Proof. Let Q, Q_1 and Q_2 be predicates in *SRSF* (in *sRud*, resp.). Then $\neg Q_1(\vec{x} \,/\, \vec{a})$, $Q_1(\vec{x} \,/\, \vec{a}) \vee Q_2(\vec{x} \,/\, \vec{a})$ and $(\exists c \in a_1) Q(\vec{x} \,/\, \vec{a}, c)$ are predicates in *SRSF* (in *sRud*, resp.):

- $P(\vec{x} \,/\, \vec{a}) \Leftrightarrow \neg Q_1(\vec{x} \,/\, \vec{a})$ can be defined as
$$\chi_P(\vec{x} \,/\, \vec{a}) = \{\emptyset\} \setminus \chi_{Q_1}(\vec{x} \,/\, \vec{a}).$$
- $P(\vec{x} \,/\, \vec{a}) \Leftrightarrow Q_1(\vec{x} \,/\, \vec{a}) \vee Q_2(\vec{x} \,/\, \vec{a})$ can be defined as
$$\chi_P(\vec{x} \,/\, \vec{a}) = \mathrm{Cond}_\in \Big(/\, 1, 0, 1, \big\{ \chi_{Q_1}(\vec{x} \,/\, \vec{a}), \chi_{Q_2}(\vec{x} \,/\, \vec{a}) \big\} \Big).$$
- $P(\vec{x} \,/\, \vec{a}) \Leftrightarrow (\exists c \in a_1) Q(\vec{x} \,/\, \vec{a}, c)$ can be defined as
$$\chi_P(\vec{x} \,/\, \vec{a}) = \mathrm{Cond}_\in \Big(/\, 1, 0, 0, \bigcup_{c \in a_1} \chi_Q(\vec{x} \,/\, \vec{a}, c) \Big). \qquad \square$$

Further examples of predicates in *sRud* are $\mathrm{trans}(/\, a)$ (a is transitive) and $\mathrm{Ord}(/\, a)$ (a is an ordinal). This can be seen using the previous proposition:

$$\begin{aligned} \mathrm{trans}(/\, a) &\iff \forall b \in a \; \forall c \in b \; c \in a \\ \mathrm{Ord}(/\, a) &\iff \mathrm{trans}(/\, a) \wedge \forall b \in a \; \mathrm{trans}(/\, b). \end{aligned}$$

1.3 Bounding ranks

A very important property of safe recursive set functions is that they increase ranks only polynomially. This can be proven similarly to the corresponding Lemma 4.1 in [**1**]. Let

$$\mathrm{rk}(x) = \bigcup \{ \mathrm{rk}(y) + 1 \colon y \in x \}$$

denote the rank of x. Observe that $\mathrm{rk}(x \,/)$ is in *SRSF*. It should be stressed that the next theorem is *not* restricted to sets of finite rank.

Theorem 1.5 *Let f be a function in SRSF. There is a polynomial q_f such that*

$$\mathrm{rk}(f(\vec{x} \,/\, \vec{a})) \leq \max_i \mathrm{rk}(a_i) + q_f(\mathrm{rk}(\vec{x}))$$

for all *sets $\vec{x}$, $\vec{a}$.*

Proof. The proof is by induction on the definition of f in *SRSF*. Our construction will ensure that q_f will always be a multi-variable polynomial with coefficients given by natural numbers. This implies that it will be a *monotone* polynomial on ordinals, i.e., if any of its arguments will be increased, leaving the other arguments the same, its value does not decrease.

We will only consider the case that f is defined by predicative set recursion, the other cases (base cases, bounded union, safe composition) are left to the reader.

If $f(x, \vec{y} \,/\, \vec{a})$ is defined by predicative set recursion from h, then by induction hypothesis we have q_h bounding h. Define q_f such that

$$q_f(\alpha, \vec{\beta}) = (1 + q_h(\alpha, \vec{\beta})) \cdot (1 + \alpha).$$

We show that $\mathrm{rk}(f(x,\vec{y}\,/\,\vec{a})) \leq \max\{\mathrm{rk}(\vec{a})\} + q_f(\mathrm{rk}(x),\mathrm{rk}(\vec{y}))$ by $\in$-induction on x:

$$\begin{aligned}
&\mathrm{rk}(f(x,\vec{y}\,/\,\vec{a}))\\
&\quad= \mathrm{rk}\left(h(x,\vec{y}\,/\,\vec{a},\{f(z,\vec{y}\,/\,\vec{a})\colon z\in x\})\right)\\
&\quad\leq \max\left\{\,\mathrm{rk}(\vec{a}),\mathrm{rk}\left(\{f(z,\vec{y}\,/\,\vec{a})\colon z\in x\}\right)\right\} + q_h(\mathrm{rk}(x),\mathrm{rk}(\vec{y}))\\
&\quad= \max\Big\{\,\mathrm{rk}(\vec{a}),\bigcup\big\{\,\mathrm{rk}(f(z,\vec{y}\,/\,\vec{a}))+1\colon z\in x\big\}\Big\} + q_h(\mathrm{rk}(x),\mathrm{rk}(\vec{y}))\\
&\quad\leq \max\Big\{\,\mathrm{rk}(\vec{a}),\bigcup\big\{\,\max\{\mathrm{rk}(\vec{a})\} + q_f(\mathrm{rk}(z),\mathrm{rk}(\vec{y}))+1\colon z\in x\big\}\Big\} + q_h(\mathrm{rk}(x),\mathrm{rk}(\vec{y}))\\
&\quad= \max\{\mathrm{rk}(\vec{a})\} + \bigcup\big\{q_f(\mathrm{rk}(z),\mathrm{rk}(\vec{y}))+1\colon z\in x\big\} + q_h(\mathrm{rk}(x),\mathrm{rk}(\vec{y}))\\
&\quad= \max\{\mathrm{rk}(\vec{a})\} + \bigcup\big\{q_f(\mathrm{rk}(z),\mathrm{rk}(\vec{y}))+1+q_h(\mathrm{rk}(x),\mathrm{rk}(\vec{y}))\colon z\in x\big\},
\end{aligned}$$

where for the second "$\leq$" we used the $\in$-induction hypothesis. Let α be $\mathrm{rk}(x)$, β_i be $\mathrm{rk}(y_i)$, and γ be $\mathrm{rk}(z)$. Assume $\gamma<\alpha$; then we will show that

$$q_f(\gamma,\vec{\beta})+1+q_h(\alpha,\vec{\beta}) \leq q_f(\alpha,\vec{\beta}). \tag{1.1}$$

Using this we can continue our calculation showing

$$\mathrm{rk}(f(x,\vec{y}\,/\,\vec{a})) \leq \max\{\mathrm{rk}(\vec{a})\} + q_f(\mathrm{rk}(x),\mathrm{rk}(\vec{y})).$$

We finish by proving (1.1):

$$\begin{aligned}
q_f(\gamma,\vec{\beta})+1+q_h(\alpha,\vec{\beta}) &= (1+q_h(\gamma,\vec{\beta}))\cdot(1+\gamma)+1+q_h(\alpha,\vec{\beta})\\
&\leq (1+q_h(\alpha,\vec{\beta}))\cdot(1+\gamma)+1+q_h(\alpha,\vec{\beta})\\
&= (1+q_h(\alpha,\vec{\beta}))\cdot(1+\gamma+1)\\
&\leq (1+q_h(\alpha,\vec{\beta}))\cdot(1+\alpha)\\
&= q_f(\alpha,\vec{\beta}).
\end{aligned}$$

□

Corollary 1.6 *Ordinal exponentiation cannot be computed by a safe recursive set function.*

2 Computing on hereditarily finite sets

For this section, we *restrict our attention to the set $\mathbb{H}F$ of hereditarily finite sets only.* Our main result for $\mathbb{H}F$ is that the *SRSF* functions acting on $\mathbb{H}F$ can be characterized in terms of $\mathrm{ATIME}\left(2^{n^{O(1)}},n^{O(1)}\right)$; namely, the class of predicates computable by an alternating Turing machine which runs in time $2^{n^{O(1)}}$ with up to $n^{O(1)}$ many alternations. It is interesting to note that this complexity class is known to characterize the decision problem for the theory of the reals with addition. In particular, the theory of the reals with addition is many-one complete for $\mathrm{ATIME}\left(2^{n^{O(1)}},n^{O(1)}\right)$ under polynomial time reductions [**2, 3, 5**].

On $\mathbb{H}F$ we will often drive a recursion by some special sets which we denote *skinny drivers*. We define the *skinny drivers of rank n*, sd_n, by induction on n as follows: $\mathrm{sd}_0=\emptyset$ and $\mathrm{sd}_{n+1}=\{\mathrm{sd}_n\}$. Turning our attention to skinny drivers on $\mathbb{H}F$ is not a restriction, as the function $\mathrm{sd}(x\,/) = \mathrm{sd}_{\mathrm{rk}(x)}$ is in *SRSF*, which can be seen as follows:

$$\mathrm{sd}(x\,/) = \overline{\mathrm{sd}}(\mathrm{rk}(x\,/)\,/); \qquad \overline{\mathrm{sd}}(\alpha\,/) = h(/\ \{\overline{\mathrm{sd}}(\beta)\colon \beta \in \alpha\});$$

$$h(/\,b) = \bigcup_{z \in b} g(/\,z, \bigcup b); \qquad g(/\,z, c) = \begin{cases} \emptyset & \text{if } z \in c \\ \{z\} & \text{otherwise.} \end{cases}$$

Predicative set recursion based on skinny drivers can be written in a simplified way.

Proposition 2.1 (Skinny Predicative Set Recursion) *Let g, h be in SRSF of appropriate arities. Then there exists some f in SRSF which satisfies*

$$\begin{aligned} f(\emptyset, \vec{y}\,/\,\vec{a}) &= g(\vec{y}\,/\,\vec{a}); \\ f(\{d\}, \vec{y}\,/\,\vec{a}) &= h(\{d\}, \vec{y}\,/\,\vec{a}, f(d, \vec{y}\,/\,\vec{a})). \end{aligned}$$

Proof. Let

$$H(x, \vec{y}\,/\,\vec{a}, b) = \begin{cases} g(\vec{y}\,/\,\vec{a}) & \text{if } x = \emptyset \\ h(x, \vec{y}\,/\,\vec{a}, \bigcup b) & \text{otherwise.} \end{cases}$$

Then f defined by predicative set recursion on x in H satisfies the required equations. □

In the previous subsection we have seen one important property of *SRSF* that ranks of sets grow polynomially only. Another important property deals with sizes of sets, in particular their growth rate. Since there are super-exponentially many sets of rank n, Theorem 1.5 implies a super-exponential bound on the size of the transitive closure of $f(\vec{x}\,/\,\vec{a})$ for $f \in$ *SRSF*. The following Theorem 2.3 will give a substantial improvement over this by showing a double exponential upper bound. Functions which satisfy such a double exponential size upper bound will be called *dietary* —the following definition will make this notion precise.

Let $|a|$ denote the cardinality of a set a, and $\mathrm{tc}(a)$ its transitive closure.

Definition 2.2 A function $f(\vec{x}\,/\,\vec{a})$ in *SRSF* is called *dietary* if, for some polynomial p,

$$|\,\mathrm{tc}(f(\vec{x}\,/\,\vec{a}))| \leq |\,\mathrm{tc}(\{\vec{x}, \vec{a}\})|^{2^{p(\mathrm{rk}(\vec{x}))}}$$

for all $\vec{x}, \vec{a} \in \mathbb{HF}$.

Theorem 2.3 *All functions in SRSF are dietary.*

Proof. The proof is by induction on the definition of f in *SRSF*. We will construct *monotone* polynomials q_f, and show that they can serve as the polynomial p in the bound of the assertion that f is dietary. We will only consider the case that f is defined by predicative set recursion, the other cases (base cases, bounded union, safe composition) are left to the reader.

If f is defined by predicative set recursion from h, then by induction hypothesis we have h dietary with bounding polynomial q_h. Define q_f such that

$$q_f(\alpha, \vec{\beta}) = (1 + q_h(\alpha, \vec{\beta})) \cdot (1 + \alpha).$$

We will show that $|f(x,\vec{y}\,/\,\vec{a})| \le |\operatorname{tc}(x,\vec{y},\vec{a})|^{2^{q_f(\mathrm{rk}(x),\mathrm{rk}(\vec{y}))}}$ by $\in$-induction on x. We have

$$\begin{aligned}
|f(x,\vec{y}\,/\,\vec{a})| &= |h(x,\vec{y}\,/\,\vec{a},\{f(z,\vec{y}\,/\,\vec{a})\colon z\in x\})| \\
&\le |\operatorname{tc}\left(\{x,\vec{y},\vec{a},\{f(z,\vec{y}\,/\,\vec{a})\colon z\in x\}\}\right)|^{2^{q_h(\mathrm{rk}(x),\mathrm{rk}(\vec{y}))}} \\
&\le \Big(|\operatorname{tc}(\{x,\vec{y},\vec{a}\})| + \sum_{z\in x} |\operatorname{tc}(f(z,\vec{y},\vec{a}))| + |x| + 1\Big)^{2^{q_h(\mathrm{rk}(x),\mathrm{rk}(\vec{y}))}}.
\end{aligned}$$

Let α be $\mathrm{rk}(x)$ and β_i be $\mathrm{rk}(y_i)$. For $z\in x$ we compute, using $\in$-induction hypothesis,

$$|\operatorname{tc}(f(z,\vec{y}\,/\,\vec{a}))| \;\le\; |\operatorname{tc}(\{x,\vec{y},\vec{a}\})|^{2^{q_f(\mathrm{rk}(z),\vec{\beta})}} \;\le\; |\operatorname{tc}(\{x,\vec{y},\vec{a}\})|^{2^{q_f(\alpha-1,\vec{\beta})}}.$$

We continue our computation from above:

$$\begin{aligned}
|f(x,\vec{y}\,/\,\vec{a})| &\le \Big(|\operatorname{tc}(\{x,\vec{y},\vec{a}\})| + |x|\cdot|\operatorname{tc}(\{x,\vec{y},\vec{a}\})|^{2^{q_f(\alpha-1,\vec{\beta})}} + |x| + 1\Big)^{2^{q_h(\alpha,\vec{\beta})}} \\
&\le \Big((|x|+1)\cdot|\operatorname{tc}(\{x,\vec{y},\vec{a}\})|^{2^{q_f(\alpha-1,\vec{\beta})}}\Big)^{2^{q_h(\alpha,\vec{\beta})}} \\
&\le |\operatorname{tc}(\{x,\vec{y},\vec{a}\})|^{2^{q_f(\alpha-1,\vec{\beta})+1}\cdot 2^{q_h(\alpha,\vec{\beta})}} \\
&\le |\operatorname{tc}(\{x,\vec{y},\vec{a}\})|^{2^{q_f(\alpha,\vec{\beta})}}.
\end{aligned}$$

For the last inequality we observe:

$$\begin{aligned}
2^{q_f(\alpha-1,\vec{\beta})+1}\cdot 2^{q_h(\alpha,\vec{\beta})} &= 2^{q_f(\alpha-1,\vec{\beta})+1+q_h(\alpha,\vec{\beta})} \\
&= 2^{(1+q_h(\alpha-1,\vec{\beta}))\cdot(1+\alpha-1)+1+q_h(\alpha,\vec{\beta})} \\
&\le 2^{(1+q_h(\alpha,\vec{\beta}))\cdot(1+\alpha-1)+1+q_h(\alpha,\vec{\beta})} \\
&= 2^{(1+q_h(\alpha,\vec{\beta}))\cdot(1+\alpha)} = 2^{q_f(\alpha,\vec{\beta})}. \qquad \square
\end{aligned}$$

That the bounds given in the definition of "dietary" are sharp, can be seen in the following way. Let $\mathrm{Sq}(/\,a) = \mathrm{Prod}(/\,a,a)$. Define f by skinny predicative set recursion as follows: $f(\emptyset\,/\,a) = a$ and $f(\{d\}\,/\,a) = \mathrm{Sq}(/\,f(d\,/\,a))$. Then f is in *SRSF*, and satisfies $|f(\mathrm{sd}_n\,/\,a)| = |a|^{2^n}$.

2.1 Simulating alternating Turing machines

We will describe a way in which alternating Turing machine computations can be simulated in *SRSF*. An *alternating Turing machine (ATM)* is given by an 8-tuple $(Q,\Sigma,\Gamma,\delta,q_0,q_{\mathrm{accept}},q_{\mathrm{reject}},g)$ where the first 7 components form the ingredients of a non-deterministic Turing machine in the usual way, that is, Q is a finite set of states which includes three designated states: the start state q_0, the accepting state q_{accept}, and the rejecting state q_{reject}, Σ is the input alphabet, Γ the work tape alphabet which includes Σ and an additional symbol $\sqcup$ denoting a blank tape cell, and $\delta \subset Q\times\Gamma\times Q\times\Gamma\times\{L,R\}$ is the transition relation. In addition to this, $g\colon Q\to\{\vee,\wedge\}$ divides the set of states into universal ($\wedge$) and existential ($\vee$) states. A *configuration* is given by a 3-tuple (u,q,v) where q is a state in Q, u and v are words over Γ , which indicates the configuration where the current state is q, the tape content is uv, and the head position is the first symbol of v (the tape contains only blanks following the last symbol of v), and the label of this configuration is given by $g(q)$.

We will not define the behavior of our ATMs in full detail, these will be obvious from the context. We do use two special conventions, however, that might lead to confusion if not stated explicitly. First, we assume that the tape is open only to the right, initially the input word is written as the first entries from the left with the head positioned at the first symbol. Second, when we mention a time bound for an ATM, then we assume that the ATM is equipped with a counter, and enters the reject state should the time bound be exceeded.

We are interested in a complexity class of alternating time with a bounded number of alternation. Given functions $t(n)$ and $q(n)$ we define the set $\mathrm{ATIME}(t(n), q(n))$ to consist of all languages which can be decided by some alternating Turing machine which runs, on inputs of length n, in time bounded by $O(t(n))$, such that the number of alternations on each computation path is bounded by $O(q(n))$.

For our simulation of an ATM $(Q, \Sigma, \Gamma, \delta, q_0, q_{\text{accept}}, q_{\text{reject}}, g)$ within *SRSF*, we assume that the alphabet Γ consists of sets only, and that $\emptyset \notin \Gamma$. Taking into considerations that functions in *SRSF* are dietary, and increase ranks of sets only polynomially, we will represent configurations as sets in the following way: The tape content will be encoded as a full binary tree (the *tape tree*) whose leaves are labeled with elements from Γ; and the head position will be encoded as a binary sequence (the *head path*) of length corresponding to the height of the tape tree. For this, we define the empty sequence by $\emptyset$, and in general the binary sequence $\langle i_1, \ldots, i_n\rangle$ of length n by $(i_1, (i_2, \ldots, (i_n, \emptyset) \ldots))$. Let $\mathcal{T}_n^\Gamma$ be the set of all tape trees of height n, and $\mathcal{P}_n$ be the set of all head paths of length n. Observe that a tape tree of height n stores tapes of length 2^n. The set of all configurations of size 2^n is now given as $\mathcal{C}_n^M = Q \times \mathcal{P}_n \times \mathcal{T}_n^\Gamma$. All these sets can be defined by functions in *SRSF*: Choose $\mathcal{P}$ in *SRSF* satisfying $\mathcal{P}(\emptyset\,/) = \{\emptyset\}$ and $\mathcal{P}(\{d\}\,/) = \mathrm{Prod}(/\,\{0,1\}, \mathcal{P}(d\,/))$, then $\mathcal{P}(\mathrm{sd}_n\,/) = \mathcal{P}_n$. Choose $\mathcal{T}^M$ in *SRSF* such that $\mathcal{T}^M(\emptyset\,/) = \Gamma$ and $\mathcal{T}^M(\{d\}\,/) = \mathrm{Sq}(/\,\mathcal{T}^M(d\,/))$, then $\mathcal{T}^M(\mathrm{sd}_n\,/) = \mathcal{T}_n^\Gamma$. Define $C^M(d\,/)$ as $\mathrm{Prod}(/\,Q, \mathrm{Prod}(/\,\mathcal{P}(d\,/), \mathcal{T}^M(d\,/)))$ then $\mathcal{C}^M(\mathrm{sd}_n\,/) = \mathcal{C}_n^M$.

We define a predicate Next^M describing successor configurations according to M. $\mathrm{Next}^M(\mathrm{sd}_n\,/\,c, c')$ will be true if $c, c' \in \mathcal{C}_n^M$ and c' is a possible next configuration from c. It can be defined as a predicate in *SRSF* in the following way:

$$\mathrm{Next}^M(d\,/\,(q,p,t),(q',p',t')) \iff$$
$$\bigvee_{(q,s,q',s',o)\in\delta} \big[\mathit{Read}(d\,/\,p,t) = s \;\wedge\; \mathit{Move}_o(d\,/\,p) = p' \;\wedge\; \mathit{Prt}(d\,/\,p,t,s') = t'\big],$$

where $\mathit{Read}(d\,/\,p,t)$ outputs the symbol on tape t at position p:

$$\mathit{Read}(\emptyset\,/\,p,t) = t;$$
$$\mathit{Read}(\{d\}\,/\,(i,p),(t_0,t_1)) = \mathit{Read}(d\,/\,p,t_i).$$

$\mathit{Move}_o(d\,/\,p)$ computes the head position obtained by moving from position p in direction $o \in \{L, R\}$, where $\langle 0,\ldots,0\rangle$ denotes the very left position (see Figure 1):

$\langle 0,0,0\rangle$	$\langle 1,0,0\rangle$	$\langle 0,1,0\rangle$	$\langle 1,1,0\rangle$	$\langle 0,0,1\rangle$	$\langle 1,0,1\rangle$	$\langle 0,1,1\rangle$	$\langle 1,1,1\rangle$

FIGURE 1. A tape of length 8 with pointers. Note that the cells are indexed by binary strings in reversed bit order.

$$\begin{aligned}
Move_o(\emptyset \,/\, p) &= 0;\\
Move_L(\{d\} \,/\, (i,p)) &= \begin{cases} \langle 0,\ldots,0\rangle & \text{if } (i,p) = \langle 0,\ldots,0\rangle\\ (0,p) & \text{if } i = 1\\ (1, Move_L(d \,/\, p)) & \text{if } i = 0; \end{cases}\\
Move_R(\{d\} \,/\, (i,p)) &= \begin{cases} \langle 1,\ldots,1\rangle & \text{if } (i,p) = \langle 1,\ldots,1\rangle\\ (1,p) & \text{if } i = 0\\ (0, Move_R(d \,/\, p)) & \text{if } i = 1; \end{cases}
\end{aligned}$$

and $Prt(d \,/\, p, t, s')$ computes the tape obtained by printing the symbol s' on tape t at position p:

$$\begin{aligned}
Prt(\emptyset \,/\, p, t, s) &= s;\\
Prt(\{d\} \,/\, (0,p), (t_0, t_1), s) &= (Prt(d \,/\, p, t_o, s), t_1);\\
Prt(\{d\} \,/\, (1,p), (t_0, t_1), s) &= (t_0, Prt(d \,/\, p, t_1, s)).
\end{aligned}$$

We also need a predicate $\overline{\text{Next}}^M$ describing successor configurations according to M for which the label according to g does not change. Let the labeling function g be extended to configurations in the obvious way: $g((q,p,t)) := g(q)$. $\overline{\text{Next}}^M(d \,/\, c, c')$ can be defined as $\text{Next}^M(d \,/\, c, c')$ and $g(c) = g(c')$.

Our next aim is to define a binary relation $\mathcal{N}_n^M$ on $\mathcal{C}_n^M$ which represents the iteration of $\overline{\text{Next}}^M$. The situation of iterating a binary relation R on a set A will occur at various places, therefore we will first explain how to achieve this as a function in *SRSF*.

Given two sets r and s (think of $r \subseteq A \times B$ and $s \subseteq B \times C$) we define their composition $r \circ s$ to be the set $\{(x,z) \in A \times C \colon (\exists y \in B)(x,y) \in r \wedge (y,z) \in s\}$. This can be defined in *sRud* as $Comp(/\, r, s) = r \circ s$ because *sRud* is closed under Boolean connectives and bounded quantification. Let A and R be sets (think of R being a binary relation on A). We define the iteration of R on A as

$$\text{Iter}(d_n \,/\, R, A) = \{(x,y) \in A \times A : \text{there is a path in } R \text{ from } x \text{ to } y \text{ of length} \leq 2^n\},$$

which can be defined by skinny recursion in *SRSF* as follows:

$$\begin{aligned}
\text{Iter}(\emptyset \,/\, R, A) &= R \cup \{(x,x) : x \in A\};\\
\text{Iter}(\{d\} \,/\, R, A) &= Comp(/\ \text{Iter}(d \,/\, R, A), \text{Iter}(d \,/\, R, A)).
\end{aligned}$$

Let us return to our task of iterating $\overline{\text{Next}}^M$. We define

$$\begin{aligned}
\overline{\mathcal{N}}^M(d \,/) &= \left\{(c,c') \in \mathcal{C}^M(d \,/) \ : \ \overline{\text{Next}}^M(d \,/\, c, c')\right\};\\
\mathcal{N}^M(d \,/) &= \text{Iter}(d \,/\, \overline{\mathcal{N}}^M(d \,/), \mathcal{C}^M(d \,/)).
\end{aligned}$$

Thus $\overline{\mathcal{N}}_n^M = \overline{\mathcal{N}}^M(\text{sd}_n \,/)$ and $\mathcal{N}_n^M = \mathcal{N}^M(\text{sd}_n \,/)$.

Let $\text{NEXT}^M(\text{sd}_n \,/\, c, c')$ denote the predicate on configurations $c, c' \in \mathcal{C}_n^M$ which is true iff c' follows from c according to M such that either c, c' and all intermediate configurations have the same label and c' is an accepting or rejecting configuration, or c and all intermediate configurations have the same label, and c' is the first with a different label:

$$\begin{aligned}&\mathrm{NEXT}^M(d\,/\,c,(q',p',t'))\\ &\qquad\iff (\exists c''\in\mathcal{C}^M(d\,/))\big[(c,c'')\in\mathcal{N}^M(d\,/)\ \wedge\ \mathrm{Next}^M(d\,/\,c'',(q',p',t'))\\ &\qquad\qquad\wedge\ [g(c)\neq g(q')\vee q'\in\{q_{\mathrm{accept}},q_{\mathrm{reject}}\}]\big];\\ &\mathcal{NEXT}^M(d\,/)=\big\{(c,c')\in\mathcal{C}^M(d\,/)\ :\ \mathrm{NEXT}^M(d\,/\,c,c')\big\}\,.\end{aligned}$$

Finally, we define the accepting states of an alternating computation according to M. Let C be a set (the set of configurations) and N a binary relation on C (taking configurations to a next alternating configuration). $\mathrm{Accept}^M(\mathrm{sd}_n\,/\,c,C,N)$ will be true if c has an accepting computation of at most n alternations:

$$\begin{aligned}&\mathrm{Accept}^M(\emptyset\,/\,c,C,N)\iff c\in C\ \wedge\ \mathit{state}(c)=q_{\mathrm{accept}};\\ &\mathrm{Accept}^M(\{d\}\,/\,c,C,N)\iff\\ &\qquad\mathrm{Accept}^M(d\,/\,c,C,N)\\ &\qquad\vee\ [g(c)=\text{“}\wedge\text{”}\ \wedge\ (\forall c'\in C)((c,c')\in N\ \rightarrow\ \mathrm{Accept}^M(d\,/\,c',C,N))]\\ &\qquad\vee\ [g(c)=\text{“}\vee\text{”}\ \wedge\ (\exists c'\in C)((c,c')\in N\ \wedge\ \mathrm{Accept}^M(d\,/\,c',C,N))];\end{aligned}$$

$$\mathrm{Accept}^M(d\,/c)\iff\mathrm{Accept}^M(d\,/\,c,\mathcal{C}^M(d\,/),\mathcal{NEXT}^M(d\,/)).$$

Now that we have described how accepting configurations of ATMs can be computed in *SRSF*, we turn to the missing bit of initializing the tape with an input word. This initialization part depends on how words are coded in $\mathbb{H}F$, a topic we will discuss next.

2.2 Encoding words in $\mathbb{H}F$

Any encoding $\nu\colon\Sigma^*\to\mathbb{H}F$ of finite words into $\mathbb{H}F$ gives rise to a class of computable functions over Σ^* which we will denote by SRSF_ν.

Definition 2.4 A function $f\colon\Sigma^*\to\Sigma^*$ is in SRSF_ν, if there exists some $F\in\mathit{SRSF}$ such that the following diagram commutes:

$$\begin{tikzcd}\mathbb{H}F \arrow[r, "F"] & \mathbb{H}F\\ \Sigma^* \arrow[u, "\nu"] \arrow[r, "f"] & \Sigma^*. \arrow[u, "\nu"]\end{tikzcd}$$

In general, the function $f\colon(\Sigma^*)^k\to\Sigma^*$ is in SRSF_ν if

$$\forall w_1,\dots,w_k\in\Sigma^*\quad\nu(f(w_1,\dots,w_k))=F(\nu(w_1),\dots,\nu(w_k)\,/)$$

for some $F\in\mathit{SRSF}$.

Definition 2.5 We call two encodings ν and ν' equivalent if they can be transformed in each other with functions from *SRSF*. That is, there exist $f,g\in\mathit{SRSF}$ such that

$$\forall w\in\Sigma^*\quad\big(f(\nu(w)\,/)=\nu'(w)\ \text{ and }\ g(\nu'(w)\,/)=\nu(w)\big).$$

Lemma 2.6 *If ν and ν' are equivalent, then $\mathit{SRSF}_\nu=\mathit{SRSF}_{\nu'}$.*

Several encodings of Σ^* in $\mathbb{H}F$ are possible, but not all will be suitable. We will discuss a few encodings mentioned in the literature.

2.2.1 The Ackerman encoding

The Ackerman encoding (cf. [**11**]) Ack: $\mathbb{N} \to \mathbb{H}F$ is given by

$$\mathrm{Ack}(2^{n_1} + 2^{n_2} + \cdots + 2^{n_k}) = \{\mathrm{Ack}(n_1), \mathrm{Ack}(n_2), \ldots, \mathrm{Ack}(n_k)\}$$

for $n_1 > n_2 > \cdots > n_k \geq 0$, $k \geq 0$. This encoding does not give rise to a nice class $\mathit{SRSF}_{\mathrm{Ack}}$ of functions. For example, $\mathit{SRSF}_{\mathrm{Ack}}$ does not include the function $n \mapsto n \dot{-} 1$: Let 2_n denote the exponentiation tower to base 2 of height n, then $\mathrm{Ack}(2_n) = \mathrm{sd}_n$. It is then easy to see that a function F which represents $n \mapsto n \dot{-} 1$ on $\mathbb{H}F$ with respect to Ack cannot be dietary, by considering F's behavior on $\mathrm{Ack}(2_n)$.

2.2.2 Two feasible encodings

We will now define two feasible encodings ν_{l} and ν_{m}. We call them feasible, because the rank of the encoded word will be of order the length of the word. Actually, both encodings will be equivalent and thus give rise to the same class of functions.

The first encoding, ν_{l}, encodes words as a list using ordered pairs. Let $\nu_{\mathrm{l}}(\lambda) = \emptyset$ and $\nu_{\mathrm{l}}(wx) = (x, \nu_{\mathrm{l}}(w))$, then $\mathrm{rk}(\nu_{\mathrm{l}}(w)) = 2|w| + O(1)$ (the constant term comes from the ranks contributed by elements in Σ).

The second encoding, ν_{m}, of a word is given as a map from the position of a letter (coded by the rank of a skinny driver) to the letter. Let $x_n \ldots x_1$ denote a word over Σ of length n. We define

$$\nu_{\mathrm{m}}(x_n \ldots x_1) = \{(\mathrm{sd}_j, x_j) \colon j = 1, \ldots, n\}.$$

Then $\mathrm{rk}(\nu_{\mathrm{m}}(w)) = |w| + O(1)$.

We leave it to the reader to verify that ν_{l} and ν_{m} are equivalent.

The main result of this section is the characterization of $\mathit{SRSF}_{\nu_{\mathrm{m}}}$ as the functions computable by an ATM in exponential time with polynomial many alternations.

Theorem 2.7 *A function $f(x)$ is in SRSF_{ν_m} if, and only if, f can be computed by some machine in* $\mathrm{ATIME}(2^{n^c}, n^c)$ *for some constant c.*

Here we prove one part of this result, that all functions computable by ATM's in exponential time with polynomial many alternations are in $\mathit{SRSF}_{\nu_{\mathrm{m}}}$. The other part will be the subject of the next section.

Theorem 2.8 *Let f be a function computable in* $\mathrm{ATIME}(2^{n^k}, n^k)$ *for some constant k. Then f is in SRSF_{ν_m}.*

Proof. Let $L \subseteq \Sigma^*$ be a language computable by some $O(2^{n^k})$-time ATM M with $O(n^k)$ many alternations on each computation path. We will define some predicate $P \in \mathit{SRSF}$ such that

$$w \in L \iff P(\nu_{\mathrm{m}}(w))$$

for all $w \in \Sigma^*$. In the following, we will use w to range over words in Σ^*, and m to range over codes of words $\nu_{\mathrm{m}}(\Sigma^*)$.

First, we define a function cwl in *sRud* which computes the length of a coded word as a skinny driver:

$$\mathrm{cwl}(/\, m) = \bigcup_{x \in \bigcup m} h(/\, x, m); \quad h(/\, x, m) = \begin{cases} \{x\} & \text{if } \mathrm{Appl}(m, x) \neq \emptyset \text{ and } \mathrm{Appl}(m, \{x\}) = \emptyset \\ \emptyset & \text{otherwise;} \end{cases}$$

$$\mathrm{cwl}(m\, /) = \mathrm{cwl}(/\, m).$$

Then $\mathrm{cwl}(/\,\nu_{\mathrm{m}}(w)) = \mathrm{sd}_{|w|}$ for $w \in \Sigma^*$.

Second, we have seen that ordinal multiplication is in *SRSF*, so is $f_1(\alpha\,/) = \alpha^k$ for ordinals α. Let

$$f_2(m\,/) = \mathrm{sd}(f_1(\mathrm{rk}(\mathrm{cwl}(m\,/)\,/)\,/)\,/).$$

Fix $w \in \Sigma^*$. Let l denote the ordinal representing $|w|$. We observe that $\mathrm{rk}(\mathrm{sd}_{|w|}\ /) = l$. Thus

$$\begin{aligned} f_2(\nu_{\mathrm{m}}(w)\,/) &= \mathrm{sd}(f_1(\mathrm{rk}(\mathrm{sd}_{|w|}\ /)\,/)\,/) = \mathrm{sd}(f_1(l\,/)\,/) \\ &= \mathrm{sd}(l^k\,/) = \mathrm{sd}_{\mathrm{rk}(l^k)} = \mathrm{sd}_{|w|^k}\,. \end{aligned}$$

Third, we define some functions suitable to produce the initial configuration based on an input word. $\mathrm{null}(\mathrm{sd}_n\ /) = \langle 0, \ldots, 0\rangle$ points to the first position of the tape:

$$\mathrm{null}(\emptyset\,/) = \emptyset; \qquad \mathrm{null}(\{d\}\,/) = (0, \mathrm{null}(d\,/)).$$

$\mathrm{blank}(\mathrm{sd}_n\ /)$ computes the blank tape of length 2^n:

$$\mathrm{blank}(\emptyset\,/) = \sqcup; \qquad \mathrm{blank}(\{d\}\,/) = (\mathrm{blank}(d\,/), \mathrm{blank}(d,\ /)).$$

The next function, moveR, computes the movement of the head position to the right. $\mathrm{moveR}(\mathrm{sd}_k, \mathrm{sd}_n\ /\,p)$ computes the head position after moving k steps to the right from position p, assuming that p is of length n:

$$\mathrm{moveR}(\emptyset, e\,/\,p) = p; \qquad \mathrm{moveR}(\{d\}, e\,/\,p) = \mathit{Move}_R(e\,/\ \mathrm{moveR}(d, e\,/\,p)).$$

$\mathrm{moveR}(\mathrm{sd}_k, e\,/) = \mathrm{moveR}(\mathrm{sd}_k, e\,/\ \mathrm{null}(e\,/))$ then computes the head position after moving k steps to the right from the first position.

Finally, we can compute initial configurations. $\mathrm{Init}^M(\nu_{\mathrm{m}}(w)\,/)$ computes the initial tape of length $2^{|w|^k}$ with $\nu_{\mathrm{m}}(w)$ standing at the very left end of the tape:

$$\begin{aligned} \mathrm{Init}(m\,/) &= \mathrm{Init}(\mathrm{cwl}(m\,/), f_2(m\,/)\,/\,m); \\ \mathrm{Init}(\emptyset, e\,/\,m) &= \mathrm{blank}(e\,/); \\ \mathrm{Init}(\{d\}, e\,/\,m) &= \mathit{Prt}(e\,/\ \mathrm{moveR}(d, e\,/), \mathrm{Init}(d, e\,/\,m), \mathrm{Appl}(/\,m, d)). \end{aligned}$$

Now we can put things together. Define P as

$$P(m\,/\,) \iff \mathrm{Accept}^M(f_2(m\,/)\,/\ \mathrm{Init}^M(\,/)).$$

Then $P \in \mathit{SRSF}$ has the desired property that $w \in L$ if and only if $P(\nu_{\mathrm{m}}(w))$ for all $w \in \Sigma^*$. □

2.3 The converse of Theorem 2.8

For the converse of Theorem 2.8, we shall prove the following theorem.

Theorem 2.9 *Let $f(x)$ be an $\mathit{SRSF}_{\nu_{\mathrm{m}}}$ function. Then f can be computed y some machine in* ATIME$(2^{n^c}, n^c)$*, for some constant c.*

The proof of Theorem 2.9 will use induction on the formation of $\mathit{SRSF}_{\nu_{\mathrm{m}}}$ functions, with the main induction step being the definition by safe recursion. However, the definition of an $\mathit{SRSF}_{\nu_{\mathrm{m}}}$ function may use intermediate *SRSF* functions which may not be $\mathit{SRSF}_{\nu_{\mathrm{m}}}$ functions. Even worse, these intermediate functions may output sets which have double exponential size $2^{2^{n^c}}$. For instance, the set $\mathcal{C}_n^M$ defined above is an example of a set with double exponential size. For this reason it is necessary to state and prove a generalized form of Theorem 2.9 that will apply to all *SRSF* functions, not just $SRSF_{\nu_{\mathrm{m}}}$ functions.

Definition A set A has *local cardinality* N provided A and every member of $\mathrm{tc}(A)$ has cardinality $\le N$.

Definition An *indexed* tree T is a finite rooted tree in which, for a given node x in T, the children of x are indexed by non-negative integers. That is, for each $i \ge 0$, there is at most one node y which is the child of x of index i. We call y the i-th child of x, however it should be noted that some children may be missing; for example, x might have a third child, but no second child.

Definition An indexed tree T has *local index rank* N provided that all nodes in T have their children indexed by numbers $< N$.

Definition Let A be a set with local cardinality N and rank $\le R$. A can be (non-uniquely) represented by an indexed finite tree T as follows. The subtree of T rooted at the i-th child of the root of T is called the i-th subtree of T. If A is empty, then T is the tree with a single node, namely its root node has no children. For A a general set, T *represents* A is defined by the condition that the elements of A are precisely the sets B for which there is some $i < N$ such that the i-th subtree of T represents B. That is, T represents A provided:

$$A = \{B : \text{for some } i, \text{ the } i\text{-th subtree of } T \text{ exists and represents } B\}.$$

Definition Let $\langle i_1, \ldots, i_\ell \rangle$ be a sequence of integers and T be a tree. This sequence denotes a path in T that starts at the root, and proceeds to the i_1-st node of T if it exists, and continues along the path represented by $\langle i_2, \ldots, i_\ell \rangle$ in the i_1-st subtree of T (if it exists). For $I = \langle i_1, \ldots, i_\ell \rangle$, we write T_I for the subtree of T rooted at the end of the path I in T.

Let the *rank of an indexed tree* T be defined by assigning the tree with a single node rank 0, and inductively assigning a general tree rank the supremum of the successors of ranks of children of T's root, i.e., $\max\big\{(\text{rank of } T_{\langle i \rangle}) + 1 : i < N\big\}$ where N is the local index rank of T. We observe that a set of local cardinality N and rank R can be represented by an indexed tree of local index rank N and rank R. Conversely, an indexed tree of local index rank N and rank R represents a set of local cardinality N and rank R.

Definition An algorithm M *recognizes* a tree T provided that on input $\langle i_1, \ldots, i_\ell \rangle$, M returns a Boolean value indicating whether the path $\langle i_1, \ldots, i_\ell \rangle$ exists in T.

When working with an algorithm M that recognizes a tree T of local index rank N, we shall often have N equal to the value 2^{2^p} for some $p \ge 0$. Note that if the rank of T is bounded by R, then any path $\langle i_1, \ldots, i_\ell \rangle$ is bounded by N^R, and hence is coded by a bit string of length $O(R \log N) = O(R \cdot 2^p)$.

More generally, we may have $N = q^{2^p}$ for some value q, at least for the intermediate parts of some of our proofs.

In our applications, we will have both p and R equal to $n^{O(1)}$, and we usually have $q = 2$. Logarithms are always base 2.

Lemma 2.10 *There are algorithms $M_=$ and $M_\in$ which take as input values $p, R > 0$ and oracles for trees S and T both with local index rank $\le N = 2^{2^p}$ and rank $\le R$, and which output Boolean values indicating whether $A = B$ and $A \in B$, respectively, where A and B*

are the sets represented by S and T, respectively. Furthermore, the algorithms $M_=$ and $M_\in$ run in time $2^p \cdot R^{O(1)}$ using $O(R)$ many alternations.

Proof. We define slightly more general algorithms $M_=^{S,T}(p, R, I, J)$ and $M_\in^{S,T}(p, R, I, J)$ which decide whether $A_I = B_I$ and $A_I \in B_I$, where A_I and B_J are the sets represented by S_I and T_J.

Here $M_=^{S,T}(p, R, I, J)$ universally calls two algorithms for checking $A_I \subseteq B_J$ and $A_I \supseteq B_J$. The algorithm for $A_I \subseteq B_J$ first universally chooses $i < N$ and checks whether path $I * \langle i \rangle$ exists in S. If not, it accepts. Otherwise, it then existentially chooses $j < N$, checks that $J * \langle j \rangle$ in T exists and rejects if not. Otherwise, it verifies whether $M_=^{S,T}(p, R, I * \langle i \rangle, J * \langle j \rangle)$. This determines whether $A_I \subseteq B_J$.

The same algorithm is used to determine whether $A_I \supseteq B_J$.

$M_\in^{S,T}(p, R, I, J)$ existentially chooses $j < N$, and checks whether $J * \langle j \rangle$ is in T. If not, it rejects, otherwise it determines whether $M_=^{S,T}(p, R, I, J * \langle j \rangle)$. □

The proof of Lemma 2.10 actually proves a better bound on the number of alternations used by the two algorithms. Namely,

Lemma 2.11 *Lemma* 2.10 *still holds if M is required to use $O(\min\{R_S, R_T\})$ alternations, where R_S and R_T are the ranks of S and T, respectively.*

Proof. The algorithms as described already have alternations bounded in this way. □

Definition A safe set function $f(\vec{x} \,/\, \vec{a})$ is AEP-computable (where "AEP" stands for "ATIME(Exp, Poly)") provided there are polynomials p, q and r, and an ATM M, such that the following holds. Let $\vec{X}$ and $\vec{A}$ be trees which represent sets $\vec{x}$ and $\vec{a}$. Let the local index rank of $\vec{X}$ and $\vec{A}$ be bounded by N_x and N_a, respectively, and their ranks be bounded by R_x and R_a, respectively. Let $N_{xa} = \max\{N_x, N_a, 2\}$ and $H_a = \max\{R_a, 1\}$. Then $M^{\vec{X},\vec{A}}$ recognizes a tree T which represents the set $f(\vec{x} \,/\, \vec{a})$ such that T has local index rank $\leq N = N_{xa}^{2^{p(R_x)}}$ and rank $\leq R = R_a + r(R_x)$. Furthermore, $M^{\vec{X},\vec{A}}$ runs in time $(H_a \cdot \log N)^{O(1)}$ with $\leq Q = H_a \cdot q(R_x)$ many alternations.

Note that Q depends on R_a multiplicatively, and N depends on only R_x.

Lemma 2.12 *The set equality relation, the set membership relation, the projection functions, the difference function and the pairing function are AEP-computable.*

Proof. For set equality and set membership, use the algorithm from the proof of Lemma 2.10 above. The theorem is obvious for the projection functions since M just computes the same function as one of its oracles. Next consider the pairing function $p(/\, a, b) = \{a, b\}$. If A and B are trees representing the sets a and b, then the tree representing the pair $\{a, b\}$ is

$$\{\langle i \rangle * I : I \text{ is a path in } A \text{ if } i = 0, \text{ or a path in } B \text{ if } i = 1\}.$$

The property "I is a path in A" (resp, "in B") is computed by invoking one of the oracle inputs. Finally, consider the set difference function $d(/\, a, b) = a \setminus b$. The tree representing the set difference $a \setminus b$ consists of the following paths:

$$\{I = \langle i_1, i_2, ..., i_\ell \rangle : I \text{ is a path in } A, \text{ and for all } j,\ A_{i_1} \text{ is not equal to } B_j\}.$$

M computes this property by universally branching to verify both (a) check that $I \in A$ using the oracle for A, and (b) universally choosing j (this takes $\log N_b$ time where

N_b bounds the local index rank of tree B) and invoking $M_=$ to verify that A_{i_1} is not equal to B_j. □

Theorem 2.13 *Every SRSF function is AEP-computable.*

The proof of Theorem 2.13 will show that the formation methods of bounded union, safe composition, and safe recursion preserve the property of being AEP computable. An important ingredient in the construction is how one composes algorithms that use alternation without losing control of the number of alternations. Specifically, suppose that f and g are algorithms that use run times t_f and t_g, and have number of alternations bounded by q_f and q_g. Then, loosely speaking, their composition $f \circ g$ can be computed in time approximately $t_f + t_g$ with $q_f + q_g + O(1)$ many alternations. The basic idea for the algorithm for $f \circ g$ is as follows. Run the algorithm for f; but whenever it needs to query its input (namely, the value of g), it *existentially guesses* the needed input value, and branches universally to both (a) verify the correctness of its guess by executing the algorithm for g, and (b) continue the computation of f. (Alternately, it could branch universally and then existentially.) Note that the algorithm for g is run only once in any given execution path, so contributes only additively to the run time. However, this "basic idea" can increase the number of alternations by the number of times f reads its input (which is more than we can allow); and a better construction is needed. The better construction is as follows:

Algorithm for $f \circ g$: Simulate f by splitting the computation up into existential portions and universal portions. There are at most q_f such portions by assumption. When starting an existential portion, initially guess all input values provided by g that will be needed throughout the computation for this portion. In addition, existentially guess (or, non-deterministically execute) the entire computation for this existential portion using the guessed input values. Then branch universally to either (a) check any one of the guessed input values, by running the algorithm for g and accepting iff it gives the guessed input value, or (b) proceed to the next universal portion. Universal portions of the computation of f are handled dually.

The run time for the algorithm is clearly $O(t_f + t_g)$. And, the number of alternations is at most $q_f + q_g + O(1)$. (The $+O(1)$ is needed for an alternation that may occur as g is invoked; it is also needed to handle the case where f is deterministic and $q_f = 0$.)

Clearly, this construction can be iterated for repeated compositions and this will allow us to handle safe recursion.

Proof of Theorem 2.13. The argument splits into cases of bounded union, safe composition, and safe recursion. The basic idea is to use the method described above for nesting calls to functions, along with the bounds established in the proofs of Theorems 1.5 and 2.3.

Case: Bounded union. $f(\vec{x} \,/\, \vec{a}, b) = \cup_{z \in b}\, g(\vec{x} \,/\, \vec{a}, z)$. The induction hypothesis that g is AEP-computable gives polynomials p_g, q_g and r_g, and an ATM M_g. Let $\vec{X}$, $\vec{A}$, B be trees representing sets $\vec{x}$, $\vec{a}$, b, with local index ranks bounded by N_x, N_a and N_b, respectively, and ranks bounded by R_x, R_a and R_b, respectively. Without loss of generality, $N_x, N_a, N_b \geq 2$ and $R_a, R_b \geq 1$. Let $N_{xab} = \max\{N_x, N_a, N_b\}$, and $R_{ab} = \max\{R_a, R_b\}$.

We describe the behavior of $M^{\vec{X},\vec{A},B}$ on input $\langle i\rangle * I$: M treats i as a pair (j_1, j_2), and universally (a) checks that $\langle j_1\rangle$ is a path in B, and (b) runs $M_g^{\vec{X},\vec{A},B_{j_1}}$ on input $\langle j_2\rangle * I$. Clearly, $M^{\vec{X},\vec{A},B}$ computes a tree T representing $f(\vec{x} \,/\, \vec{a})$. Let N_g be an upper bound to the local index rank of the tree computed by $M_g^{\vec{X},\vec{A},B_{j_1}}$, and R_g an upper bound to its rank. Let Q_g bound the number of alternations for $M_g^{\vec{X},\vec{A},B_{j_1}}$.

T has local index rank bounded by $O(N_g \cdot N_b) = O(N_{xab}^{2^{p_g(R_x)}} \cdot N_b) = N_{xab}^{2^{p_g(R_x)+O(1)}}$ and rank bounded by $R_g \le R_{ab} + r_g(R_x)$. The algorithms runs in time bounded by

$$(R_{ab} \log N_{xab})^{O(1)} + (R_{ab} \log N_g)^{O(1)} \le (R_{ab} \log N_{xab}^{2^{p_g(R_x)}})^{O(1)}$$

with $Q_g + 1 \le R_{ab} \cdot (q_g(R_x) + 1)$ many alternations.

Case: Safe composition. $f(\vec{x} \,/\, \vec{a}) = h(s(\vec{x}/)/t(\vec{x}/\vec{a}))$. Here s and t may be vectors of functions, but we omit this for simplicity (nothing essential is changed in the proof). The induction hypotheses give polynomials p_h, p_s, p_t, q_h, q_s, q_t, r_h, r_s, and r_t, and machines M_h, M_s, and M_t. Let $\vec{X}$ and $\vec{A}$ be trees representing sets $\vec{x}$ and $\vec{a}$, repectively, with local index ranks bounded by N_x and N_a, respectively, and ranks bounded by R_x and R_a respectively. Without loss of generality, $N_x, N_a \ge 2$ and $R_a \ge 1$. Let $N_{xa} = \max(N_x, N_a)$, $N_{st} = \max(N_s, N_t)$, $p_{st} = p_s + p_t$, and $q_{st} = q_s + q_t$. We have that $N_{st} \le N_{xa}^{2^{p_{st}(R_x)}}$.

Let M be the straightforward algorithm for f, based on composing the algorithms for h, s and t. $M^{\vec{X},\vec{A}}$ will recognize a tree T whose rank is bounded by

$$R_t + r_h(R_s) \le R_a + r_t(R_x) + r_h(r_s(R_x))$$

so we can choose $r_f = r_t + r_h \circ r_s$. The local index rank of T is bounded by

$$N_{st}^{2^{p_h(R_s)}} \le \big(N_{xa}^{2^{p_{st}(R_x)}}\big)^{2^{p_h(r_s(R_x))}} = N_{xa}^{2^{p_{st}(R_x)+p_h(r_s(R_x))}}.$$

The run time of M is bounded by, for some $c = O(1)$,

$$\begin{aligned}
&O(\max\{(\mathrm{runtime}(s), \mathrm{runtime}(t)\} + \mathrm{runtime}(h))\\
&\quad \le O\Big(\max\{(\log N_x) \cdot 2^{p_s(R_x)}, R_a \cdot (\log N_{xa}) \cdot 2^{p_t(R_x)}\}^c + \big(R_t \cdot (\log N_{st}) \cdot 2^{p_h(R_s)}\big)^c\Big)\\
&\quad \le O\Big(\big(R_a \cdot (\log N_{xa}) \cdot 2^{p_{st}(R_x)}\big)^c + \big((R_a + r_t(R_x)) \cdot (\log N_{xa}) \cdot 2^{p_{st}(R_x)+p_h(r_s(R_x))}\big)^c\Big)\\
&\quad \le \big(R_a \cdot (\log N_{xa}) \cdot 2^{p_f(R_x)}\big)^c
\end{aligned}$$

for an appropriately chosen polynomial p_f. Say, $p_f = p_s + p_t + r_t + p_h \circ r_s + O(1)$.

The number of alternations of this algorithm is bounded by

$$\begin{aligned}
&\max\{\mathrm{alternations}(s), \mathrm{alternations}(t)\} + \mathrm{alternations}(h) + O(1)\\
&\quad \le \max\{q_s(R_x), R_a \cdot q_t(R_x)\} + R_t \cdot q_h(R_s) + O(1)\\
&\quad \le R_a \cdot q_{st}(R_x) + (R_a + r_t(R_x)) \cdot q_h(r_s(R_x)) + O(1)\\
&\quad \le R_a \cdot q_f(R_x)
\end{aligned}$$

for an appropriate polynomial q_f.

Case: Safe recursion. $f(x, \vec{y} \,/\, \vec{a}) = h(x, \vec{y} \,/\, \vec{a}, \{f(z, \vec{y} \,/\, \vec{a})\colon z \in x\})$. The induction hypothesis gives polynomials p_h, q_h, r_h, and a machine M_h. Let X, $\vec{Y}$ and $\vec{A}$ be trees representing sets x, $\vec{y}$ and $\vec{a}$, respectively, with local index ranks bounded by N_x, N_y and N_a, respectively, and ranks bounded by R_x, R_y and R_a, respectively. Without loss of generality, $N_x, N_y, N_a \ge 2$ and $R_a \ge 1$. Let $N_{xya} = \max(N_x, N_y, N_a)$, and $R_{xy} = \max(R_x, R_y)$. With M we denote the (yet to be defined) algorithm for computing f.

Let $\overline{f}(x, \vec{y} \,/\, \vec{a})$ be the set $\{f(z, \vec{y} \,/\, \vec{a})\colon z \in x\}$. Then $\overline{f}$ can be computed by a machine $M_{\overline{f}}^{X,\vec{Y},\vec{A}}$ which on input $\langle i\rangle * I$ first tests whether $\langle i\rangle$ is a path in X, and if so calls $M^{X_i,\vec{Y},\vec{A}}$ on input I. Then, $M^{X,\vec{Y},\vec{A}}$ computes the composition of h with $\overline{f}$ using the above algorithm. Let $R_{\overline{f}}$ ($N_{\overline{f}}$, respectively) denote a bound to the rank (local index rank, respectively) of the tree computed by $M_{\overline{f}}^{X,\vec{Y},\vec{A}}$.

Clearly, $M^{X,\vec{Y},\vec{A}}$ computes a tree T which represents $f(x, \vec{y} \,/\, \vec{a})$. To obtain a bound for the rank of T we can choose r_f similar to the proof of Theorem 1.5: Let $r_f(z) = r'_f(z, z)$ with

$$r'_f(z, z') = (1 + r_h(z'))(1 + z).$$

The same calculation done in that proof carries over here to show by induction on R_x that the rank is $\leq R_a + r'_f(R_x, R_{xy})$.

In order to bound the local index rank of T we choose p_f similar to the proof of Theorem 2.3. Let $p_f(z) = p'_f(z, z)$ for

$$p'_f(z, z') = (p_h(z') + r_f(z') + O(1)) \cdot (1 + z).$$

We show by induction on R_x that the local index rank of T is $\leq N = N_{xya}^{2^{p'_f(R_x,R_{xy})}}$ and that the run time is $\leq (R_a \log N)^{O(1)}$. In case $R_x = 0$ both assertions follow easily. For $R_x > 0$, we calculate as a bound for the local index rank of T

$$\begin{aligned}
\max\{N_{xya}, N_{\overline{f}}\}^{2^{p_h(R_{xy})}} &\leq \max\{N_{xya}, N_x, N_{xya}^{2^{p'_f(R_x-1,R_{xy})}}\}^{2^{p_h(R_{xy})}} \\
&\leq N_{xya}^{2^{p'_f(R_x-1,R_{xy})+p_h(R_{xy})}} \leq N_{xya}^{2^{p'_f(R_x,R_{xy})}}.
\end{aligned}$$

The run time of M can be bounded by, for some $c = O(1)$,

$$\begin{aligned}
&O(\text{runtime}(M_h) + \text{runtime}(M_{\overline{f}})) \\
&\quad \leq O\Big(\big(\max\{R_a, R_{\overline{f}}\} \cdot (\log \max\{N_{xya}, N_{\overline{f}}\}) \cdot 2^{p_h(R_{xy})}\big)^c \\
&\qquad\quad + \big(R_a \cdot (\log N_{\overline{f}}) + R_a \cdot (\log N_{xya}) \cdot 2^{p'_f(R_x-1,R_{xy})}\big)^c\Big) \\
&\quad \leq O\Big(\big((R_a + r_f(R_{xy})) \cdot (\log N_{xya}) \cdot 2^{p'_f(R_x-1,R_{xy})} \cdot 2^{p_h(R_{xy})}\big)^c \\
&\qquad\quad + \big(R_a \cdot (\log N_{xya}) \cdot 2^{p'_f(R_x-1,R_{xy})} + R_a \cdot (\log N_{xya}) \cdot 2^{p'_f(R_x-1,R_{xy})}\big)^c\Big) \\
&\quad \leq \big(\log(N_{xya}) \cdot 2^{p_h(R_{xy})+r_f(R_{xy})+O(1)+p'_f(R_x-1,R_{xy})}\big)^c \\
&\quad = \big(\log(N_{xya}) \cdot 2^{p'_f(R_x,R_{xy})}\big)^c.
\end{aligned}$$

Let $q'_f(z, z') = (r_f(z') + O(1)) \cdot q_h(z') \cdot (1 + z)$. We will show that the overall number of alterations of $M^{X,\vec{Y},\vec{A}}$ is bounded by $R_a \cdot q'_f(R_x, R_{xy})$ by induction on R_x. Then choosing $q_f(z) = q'_f(z, z)$ gives the desired bound. If $R_x = 0$, the overall number of alterations can be calculated as

$$\text{alternations}(M_h) \leq R_a \cdot q_h(R_{xy}) \leq R_a \cdot q'_f(0, R_{xy}).$$

If $R_x > 0$ we obtain

$$\begin{aligned}
&\text{alternations}(M_h) + \text{alternations}(M_{\overline{f}}) + O(1)\\
&\qquad \le \max\{R_a, R_{\overline{f}}\} \cdot q_h(R_{xy}) + R_a \cdot q'_f(R_x - 1, R_{xy}) + O(1)\\
&\qquad \le (R_a + r_f(R_{xy})) \cdot q_h(R_{xy}) + R_a \cdot q'_f(R_x - 1, R_{xy}) + O(1)\\
&\qquad \le R_a \cdot ((r_f(R_{xy}) + O(1)) \cdot q_h(R_{xy}) + q'_f(R_x - 1, R_{xy}))\\
&\qquad = R_a \cdot q'_f(R_x, R_{xy}). \qquad \Box
\end{aligned}$$

It is easy to verify that Theorem 2.9 is a corollary of Theorem 2.13.

3 Computing on arbitrary sets

Our goal in this section is to characterise the safe-recursive functions (i.e., the functions in *SRSF*) in definability-theoretic terms. To achieve this we will use a relativisation of Gödel's L-hierarchy. Our result breaks into two parts: an upper bound result, showing that every safe-recursive function satisfies our definability criterion, and a lower bound result, showing that any function satisfying our definability criterion is in fact safe-recursive. First we introduce:

3.1 The relativised Gödel hierarchy

For a transitive set T, define the *L^T-hierarchy* as follows:

$$\begin{aligned}
L^T_0 &= T\\
L^T_{\alpha+1} &= \mathrm{Def}(L^T_\alpha)\\
L^T_\lambda &= \cup_{\alpha<\lambda} L^T_\alpha \quad \text{for limit } \lambda,
\end{aligned}$$

where for any set X, $\mathrm{Def}(X)$ denotes the set of all subsets of X which are first-order definable over the structure $(X, \in)$ with parameters. The following facts are easily verified:

Lemma 3.1 *For any transitive set T,*

(1) *T is an element of L^T_1;*
(2) *each L^T_α is transitive and $\alpha \le \beta$ implies $L^T_\alpha \subseteq L^T_\beta$;*
(3) $\mathrm{Ord}(L^T_\alpha) = \mathrm{Ord}(T) + \alpha$, *where* $\mathrm{Ord}(X)$ *denotes* $\mathrm{Ord} \cap X$ *for any set X.*

Gödel demonstrated the following definability result for the L-hierarchy: For limit α, the sequence $(L_\beta : \beta < \alpha)$ is definable over $(L_\alpha, \in)$ and the definition is independent of α. (See for example [**4**, Chapter II, Lemma 2.8].) His argument readily yields the following refinement, which will be needed for our upper bound result.

Lemma 3.2 *Let $k < \omega$ be sufficiently large, and let T be transitive, α an ordinal and $\varphi(\vec{x}, \vec{y})$ a formula. Let $\mathcal{D}$ consist of all triples $(U, \beta, \vec{p})$ such that for some $\gamma < \alpha$: U is a transitive element of $L^T_{\gamma+1}$, $\gamma + \beta + k < \alpha$ and $\vec{p}$ is a sequence (with the same length as $\vec{y}$) of elements of L^U_β. Then the function with domain $\mathcal{D}$ sending $(U, \beta, \vec{p})$ to $(L^U_\beta, \{\vec{x}\colon L^U_\beta \vDash \varphi(\vec{x}, \vec{p})\})$ is definable over L^T_α via a definition independent of T, α.*

For our lower bound result we will need the following (see [**8**, Corollary 13.8]):

Lemma 3.3 (Gödel) *There exists a list of functions* $G_1(x,y),\dots,G_{10}(x,y)$ *such that for transitive* T, $T \cup \bigcup_{1\le i\le 10} \mathrm{range}(G_i \restriction T\times T)$ *is transitive and* $\mathrm{Def}(T)$ *consists of those subsets of* T *which belong to the closure of* $T\cup\{T\}$ *under the* G_i*'s. Moreover, for each* i *the associated function* G_i^* *defined by* $G_i^*(/\,x,y) = G_i(x,y)$ *belongs to sRud.*

3.2 The upper bound result

Recall that we identify finite sequences $\vec{x}$ of sets with individual sets, using Kuratowski pairing. For any set x let $\mathrm{tc}(x)$ denote the transitive closure of x. The rank of $\mathrm{tc}(x)$ (in the von Neumann hierarchy of V_α's) is the same as $\mathrm{rk}(x)$, the rank of x. Given two finite sequences $\vec{x}$, $\vec{y}$, we write $\vec{x}*\vec{y}$ for their concatenation.

Definition 3.4 For sequences $\vec{x}$, $\vec{y}$ and $0<n\le\omega$ we define $\mathrm{SR}_n(\vec{x}\,/\,\vec{y})$ as $L^{\mathrm{tc}(\vec{x}*\vec{y})}_{n+\mathrm{rk}(\vec{x})^n}$.

Our upper bound result is the following refinement of Theorem 1.5:

Theorem 3.5 *If* $f(\vec{x}\,/\,\vec{y})$ *is safe-recursive then for some finite* n, $f(\vec{x}\,/\,\vec{y})$ *is uniformly definable in* $\mathrm{SR}_n(\vec{x}\,/\,\vec{y})$*, i.e., for some formula* $\varphi(\vec{x},\vec{y},z)$ *we have:*

(1) $f(\vec{x}\,/\,\vec{y})$ *belongs to* $\mathrm{SR}_n(\vec{x}\,/\,\vec{y})$ *for all* $\vec{x}$, $\vec{y}$*;*
(2) $f(\vec{x}\,/\,\vec{y}) = z$ *if and only if* $(\mathrm{SR}_n(\vec{x}\,/\,\vec{y}),\in)\models\varphi(\vec{x},\vec{y},z)$.

To see that this implies Theorem 1.5, note that all elements of $\mathrm{SR}_n(\vec{x}\,/\,\vec{y})$ have rank at most $\mathrm{rk}(\vec{x}*\vec{y})+n+\mathrm{rk}(\vec{x})^n \le \max(\mathrm{rk}(\vec{x}),\mathrm{rk}(\vec{y}))+k+\mathrm{rk}(\vec{x})^n$ for some finite k, which is bounded by $\max_i \mathrm{rk}(y_i)$ + a polynomial in $\mathrm{rk}(\vec{x})$.

Proof of Theorem 3.5. As in the proof of Theorem 1.5 we proceed by induction on the clauses that generate f as a safe-recursive function. The base cases of *Projection, Difference* and *Pairing* are left to the reader. For *Bounded Union*, we have:

$$f(\vec{x}\,/\,\vec{y},z) = \bigcup_{w\in z} g(\vec{x}\,/\,\vec{y},w)$$

and by induction there is a finite n such that $g(\vec{x}\,/\,\vec{y},w)$ is uniformly definable in $\mathrm{SR}_n(\vec{x}\,/\,\vec{y},w)$. By the definability of union, it then follows from Lemma 3.2 that $f(\vec{x}\,/\,\vec{y},z)$ is uniformly definable in $\mathrm{SR}_{n+k}(\vec{x}\,/\,\vec{y},z)$ for sufficiently large k.

For *Safe Composition*, we have:

$$f(\vec{x}\,/\,\vec{y}) = h(\vec{r}(\vec{x}\,/)\,/\,\vec{t}(\vec{x}\,/\,\vec{y}))$$

and by induction n_h, n_{r_i} and n_{t_j} witnessing the theorem for the functions h, r_i for each i and t_j for each j, respectively. By Lemma 3.2 we can choose a large n and combine the uniform definitions of the $r_i(\vec{x}\,/)$'s in the $\mathrm{SR}_{n_{r_i}}(\vec{x}\,/)$'s, of the $t_j(\vec{x}\,/\,\vec{y})$'s in the $\mathrm{SR}_{n_{t_j}}(\vec{x}\,/\,\vec{y})$'s and of $h(\vec{r}(\vec{x}\,/)\,/\,\vec{t}(\vec{x}\,/\,\vec{y}))$ in $\mathrm{SR}_{n_h}(\vec{r}(\vec{x}\,/)\,/\,\vec{t}(\vec{x}\,/\,\vec{y}))$ to produce a uniform definition of $f(\vec{x}\,/\,\vec{y})$ inside $\mathrm{SR}_n(\vec{x}\,/\,\vec{y})$.

For *Predicative Set Recursion*, we have:

$$f(x,\vec{y}\,/\,\vec{z}) = h(x,\vec{y}\,/\,\vec{z},\{f(w,\vec{y}\,/\,\vec{z})\colon w\in x\}).$$

Choose $n>1$ to witness the Theorem for h, i.e., so that $h(x,\vec{y}\,/\,\vec{z},u)$ is uniformly definable in $\mathrm{SR}_n(x,\vec{y}\,/\,\vec{z},u)$. Fix $\vec{y}$ and $\vec{z}$. By induction on $\mathrm{rk}(x)$ we show that $f(x,\vec{y}\,/\,\vec{z})$ is uniformly definable in $L^{\mathrm{tc}(\langle x\rangle*\vec{y}*\vec{z})}_{n+\mathrm{rk}(\langle x\rangle*\vec{y})^n\cdot k\cdot(\mathrm{rk}(x)+1)}$ (where $k>n$ is fixed as in Lemma 3.2). If $\mathrm{rk}(x)$ is 0 then we want to show that $f(0,\vec{y}\,/\,\vec{z}) = h(0,\vec{y}\,/\,\vec{z},0)$ is an element of $L^{\mathrm{tc}(\langle 0\rangle*\vec{y}*\vec{z})}_{n+\mathrm{rk}(\langle 0\rangle*\vec{y})^n\cdot k}$, which is true by the choice of n. If $\operatorname{rk}(x) > 0$ then by induction we know that for $w \in x$, $f(w, \vec{y} \,/\, \vec{z})$ is uniformly definable in

$$L^{\operatorname{tc}(\langle w\rangle * \vec{y} * \vec{z})}_{n+\operatorname{rk}(\langle w\rangle * \vec{y})^n \cdot k \cdot (\operatorname{rk}(w)+1)};$$

it follows that $\{f(w, \vec{y} \,/\, \vec{z})\colon w \in x\}$ is uniformly definable over $L^{\operatorname{tc}(\langle x\rangle * \vec{y} * \vec{z})}_{n+\operatorname{rk}(\langle x\rangle * \vec{y})^n \cdot k \cdot \operatorname{rk}(x)}$.

By choice of n, $f(x, \vec{y} \,/\, \vec{z}) = h(x, \vec{y} \,/\, \vec{z}, \{f(w, \vec{y} \,/\, \vec{z})\colon w \in x\})$ is uniformly definable in

$$L^{\operatorname{tc}(\langle x\rangle * \vec{y} * \vec{z} * \langle \{f(w, \vec{y} \,/\, \vec{z})\colon w \in x\}\rangle)}_{n+\operatorname{rk}(\langle x\rangle * \vec{y})^n}$$

and hence in $L^{\operatorname{tc}(\langle x\rangle * \vec{y} * \vec{z})}_{n+\operatorname{rk}(\langle x\rangle * \vec{y})^n \cdot k \cdot (\operatorname{rk}(x)+1)}$ by Lemma 3.2. This completes the induction step.

Now by choosing m large enough so that $n + \operatorname{rk}(\langle x\rangle * \vec{y})^n \cdot k \cdot (\operatorname{rk}(x) + 1)$ is less than $m + \operatorname{rk}(\langle x\rangle * \vec{y})^m$ we have that $f(x, \vec{y} \,/\, \vec{z})$ is uniformly definable in $\mathrm{SR}_m(x, \vec{y} \,/\, \vec{z})$, as desired. □

Note that if there are no safe arguments then $\mathrm{SR}_n(\vec{x} \,/\,)$ takes a particularly nice form and we have:

Corollary 3.6 *Suppose that $f(\vec{x} \,/\,)$ is safe-recursive. Then for some finite n and some formula φ we have, for $\vec{x} \neq \langle 0\rangle$, that*

(1) *$f(\vec{x} \,/\,)$ belongs to $L^{\operatorname{tc}(\vec{x})}_{n+\operatorname{rk}(\vec{x})^n}$;*

(2) *$f(\vec{x} \,/\,) = y$ if and only if $L^{\operatorname{tc}(\vec{x})}_{n+\operatorname{rk}(\vec{x})^n} \models \varphi(\vec{x}, y)$.*

For any transitive set T let $\mathrm{SR}(T)$ denote $L^T_{(2+\operatorname{rk}(T))^\omega}$.

Corollary 3.7 *For transitive T, $\mathrm{SR}(T)$ contains $T \cup \{T\}$ and is closed under SRSF functions (i.e., T contains $f(\vec{x} \,/\, \vec{y})$ whenever f is safe-recursive and T contains the components of $\vec{x}$, $\vec{y}$).*

We shall soon see that $\mathrm{SR}(T)$ is in fact the smallest such set.

3.3 The lower bound result

Now we aim for a converse of Theorem 3.5. We begin by showing that a certain initial segment of the L^T-hierarchy can be generated by iteration of a safe-recursive function.

Lemma 3.8 *Suppose that $f(x \,/\,)$ is safe-recursive with ordinal values and $g(\,/\, x)$ is safe-recursive with the property that $x \subseteq g(\,/\, x)$ for all x. By induction on α define $g^\alpha(\,/\, x)$ by: $g^0(\,/\, x) = x$, $g^{\alpha+1}(\,/\, x) = g(\,/\, g^\alpha(\,/\, x))$, $g^\lambda(\,/\, x) = \cup_{\alpha<\lambda} g^\alpha(\,/\, x)$ for limit λ. Then the function $h(x \,/\,) = g^{f(x \,/\,)}(\,/\, x)$ is safe-recursive.*

Proof. Imitating the proof that multiplication can be defined from addition via a safe recursion, first define the function $k(x, y \,/\,)$ via a safe recursion as follows:

$$k(x, y \,/\,) = \begin{cases} y & \text{if } x = 0 \\ g(\,/\, \cup \{k(z, y \,/\,)\colon z \in x\}) & \text{if } x = \operatorname{Succ}(\,/\, \cup x) \\ \cup \{k(z, y \,/\,)\colon z \in x\} & \text{otherwise.} \end{cases}$$

Then k is safe-recursive and note that for each ordinal α, $k(\alpha, y \,/\,) = g^\alpha(\,/\, y)$. It follows from safe composition that $h(x \,/\,) = k(f(x \,/\,), x \,/\,)$ is also safe-recursive. □

Recall that the rank function $\mathrm{rk}(x\,/)$ is safe-recursive. We say that a function $f(\vec{x}\,/\,\vec{y})$ is *safe-recursive with parameter* p iff for some safe-recursive function $g(\vec{x}, z\,/\,\vec{y})$, we have $f(\vec{x}\,/\,\vec{y}) = g(\vec{x}, p\,/\,\vec{y})$ for all $\vec{x}$, $\vec{y}$.

Corollary 3.9

(1) *The function* $\mathrm{tc}(x\,/)$ *computing the transitive closure of* x*, is safe-recursive.*
(2) *The function* $L(x, T\,/) = L^T_{\mathrm{rk}(x)}$ *is safe-recursive with parameter* ω*.*
(3) *For each finite* n*, the function* $\mathrm{SR}_n(\vec{x}\,/)$ *is safe-recursive with parameter* ω*.*

Proof.

(1) The transitive closure of x is obtained by iterating the sRud function $g(/\,x) = (x \cup (\cup x))$ $\mathrm{rk}(x)$ times. So the result follows from the previous lemma.
(2) The function $g(/\,x) = x\cup$ the union of the ranges of the Gödel functions on x (see Lemma 3.3) belongs to *sRud*. It follows from the previous lemma that the function $g^*(T\,/) = \mathrm{Def}(T) =$ the closure of $T \cup \{T\}$ under g^* (restricted to transitive T) is safe-recursive with parameter ω, as $\mathrm{Def}(T)$ is obtained by iterating g ω times. Similarly, as the function $\mathrm{rk}(x\,/)$ is safe-recursive, an application of the previous lemma gives the safe-recursiveness of $L(x, T\,/)$.
(3) This follows from 1 and 2, using the fact that ordinal multiplication is safe-recursive. □

We therefore get the following partial converse to Theorem 3.5.

Theorem 3.10 *Suppose that for some finite* n*,* $f(\vec{x}\,/\,\vec{y})$ *is uniformly definable in* $\mathrm{SR}_n(\vec{x}\,/\,\vec{y})$*. Then* $f(\vec{x}\,/\,\vec{y})$ *is safe-recursive with parameter* ω*. Moreover there is a safe-recursive function* $g(\vec{x}\,/\,\vec{y})$ *such that* $f(\vec{x}\,/\,\vec{y}) = g(\vec{x}\,/\,\vec{y})$ *whenever* $\vec{x}$ *has a component of infinite rank (i.e., whenever* $\mathrm{rk}(\vec{x})$ *is infinite).*

Proof. By Corollary 3.9, 3.9, the function $\mathrm{SR}_n(\vec{x}\,/\,\vec{y})$ is safe-recursive with parameter ω. For any formula $\varphi(\vec{x}, \vec{y}, z)$, the function $g(/\,T, p) = \{(\vec{x}, \vec{y}) : T \vDash \varphi(\vec{x}, \vec{y}, p)\}$ is in *sRud* (see for example [**4**, Chapter VI, Lemma 1.17]). It follows that any function which is uniformly definable in $\mathrm{SR}_n(\vec{x}\,/\,\vec{y})$ is also safe-recursive with parameter ω. For the "moreover" clause, note that there is a safe-recursive function $f(x\,/)$ whose value is ω for x of infinite rank, and therefore ω can be eliminated as a parameter when $\vec{x}$ has a component of infinite rank. □

Corollary 3.11 *The safe-recursive functions with parameter* ω *are exactly the functions* $f(\vec{x}\,/\,\vec{y})$ *which are uniformly definable in* $\mathrm{SR}_n(\vec{x} * \langle\omega\rangle\,/\,\vec{y})$ *for some finite* n*.*

Note that the closure of $\{0\}$ under safe-recursive functions is L_ω, the set of hereditarily finite sets and when T is transitive of infinite rank then ω belongs to the safe-recursive closure of T. Therefore we have:

Corollary 3.12 *For transitive* T*,* $\mathrm{SR}(T) = L^T_{(2+\mathrm{rk}(T))^\omega}$ *is the smallest set which contains* $T \cup \{T\}$ *as a subset and is closed under safe-recursive functions.*

We therefore obtain the following hierarchy of iterated safe-recursive closures. Define:

$$\begin{aligned} \mathrm{SR}_0 &= \emptyset \\ \mathrm{SR}_{\alpha+1} &= \mathrm{SR}(\mathrm{SR}_\alpha) \\ \mathrm{SR}_\lambda &= \cup_{\alpha<\lambda} \mathrm{SR}_\alpha \qquad \text{for limit } \lambda. \end{aligned}$$

Corollary 3.13 *For every α,* $\mathrm{SR}_{1+\alpha} = L_{\omega^{\omega^\alpha}}$.

To eliminate the parameter ω from Corollary 3.11 we redefine SR_n slightly, using a slower hierarchy for L^T. Define M^T_α inductively as follows:

$$\begin{aligned} M^T_0 &= T \\ M^T_{\alpha+1} &= M^T_\alpha \cup \bigcup_{1\le i\le 10} \mathrm{range}(G_i \restriction ((M^T_\alpha \cup \{M^T_\alpha\}) \times (M^T_\alpha \cup \{M^T_\alpha\})) \\ M^T_\lambda &= \cup_{\alpha<\lambda} M^T_\alpha \quad \text{for limit } \lambda. \end{aligned}$$

This hierarchy is very close to Jensen's S-hierarchy, a refinement of his J-hierarchy (see [**9**, p. 244]). We have the following (see [**9**, p. 255]):

Lemma 3.14 *For any transitive set T:*

(1) *T is an element of M^T_1.*
(2) *Each M^T_α is transitive and $\alpha \le \beta$ implies $M^T_\alpha \subseteq M^T_\beta$.*
(3) $\mathrm{Ord}(M^T_\lambda) = \mathrm{Ord}(T) + \lambda$ *for limit λ.*
(4) *$M^T_\alpha = L^T_\alpha$ if α is ω or $\omega \cdot \alpha = \alpha$. In particular, $M^T_{\mathrm{rk}(x)^\omega} = L^T_{\mathrm{rk}(x)^\omega}$ if x has rank greater than* 1.

Definition 3.15 For sequences $\vec{x}$, $\vec{y}$ and $0 < n \le \omega$ we define $\mathrm{SR}^*_n(\vec{x} \,/\, \vec{y})$ to be $M^{\mathrm{tc}(\vec{x}*\vec{y})}_{n+\mathrm{rk}(\vec{x})^n}$.

Lemma 3.2 and Theorem 3.5 (the upper bound result) go through with L replaced by M and $\mathrm{SR}_n(\vec{x} \,/\, \vec{y})$ replaced by $\mathrm{SR}^*_n(\vec{x} \,/\, \vec{y})$. But now the lower bound result can be improved, as the parameter ω can be dropped in the version of Corollary 3.9 (3.9), (3.9) in which L is replaced by M and SR is replaced by SR^*: Whereas obtaining $L^T_{\alpha+1}$ from L^T_α requires a safe recursion of length ω, $M^T_{\alpha+1}$ is obtained from M^T_α by a single application of a function in *sRud*. In conclusion, we get the following characterisation:

Theorem 3.16 *The safe-recursive functions are exactly the functions $f(\vec{x} \,/\, \vec{y})$ which are uniformly definable in $\mathrm{SR}^*_n(\vec{x} \,/\, \vec{y})$ for some finite n.*

3.4 Safe recursion on binary ω-sequences

We let $\{0,1\}^\omega$ denote all ω-sequences of 0's and 1's. Note that if x belongs to $\{0,1\}^\omega$ then x has rank ω. It follows that $\mathrm{SR}_n(x\,/)$ is equal to $L^{\mathrm{tc}(x)}_{\omega^n}$ for $0 < n \le \omega$. Moreover the latter can be equivalently written as $L_{\omega^n}[x]$, where $L_\alpha[x]$ is the α-th level of the relativised Gödel's L-hierarchy in which x is introduced as a new unary predicate.

Thus the safe-recursive functions restricted to elements of $\{0,1\}^\omega$ as normal inputs take the following form:

$$f(x\,/) = y \quad \text{iff} \quad L_{\omega^n}[x] \vDash \varphi(x,y)$$

for some formula φ.

The following is implicit in the analysis of the "Theory Machine", the universal infinite-time Turing machine considered in [**6**].

Theorem 3.17 *For any function $g : \{0,1\}^\omega \to \{0,1\}^\omega$, the following are equivalent:*

(1) *g is computable by an infinite-time Turing machine (see* [**7**]*) in time β for some $\beta < \omega^\omega$.*

(2) *g is of the form*

$$g(x) = y \quad \text{\textit{iff}} \quad L_\beta[x] \vDash \varphi(x, y)$$

for some formula φ and some $\beta < \omega^\omega$.

From this we see that the safe-recursive functions restricted to normal inputs in $\{0,1\}^\omega$ with values in $\{0,1\}^\omega$ are equivalent to the functions computed by an infinite-time Turing machine in time less than ω^ω. Interestingly, these are exactly the functions which are "computable in polynomial time" on an infinite-time Turing machine in the sense of [**12**].

4 A machine model for safe recursion

We finish by briefly describing a simple machine model with parallel processors which with the natural bound on running times yields the class of safe-recursive functions.

To each set x assign a processor P_x, which computes in ordinal stages. The value computed by P_x at stage α is denoted by P_x^α. The entire machine M is determined by a function $h(/\, x)$ in sRud and a finite $n > 0$. We write $M = M_h^n$.

P_x^α is defined by induction on α as follows. for any x and α we denote $\{(y, \beta, P_y^\beta) \colon y \in x, \beta \leq \alpha\}$ by $P_{\in x}^{\leq\alpha}$ and $\{(x, \beta, P_x^\beta : \beta < \alpha\}$ by $P_x^{<\alpha}$. Now define:

$$P_x^\alpha = h(/\, P_{\in x}^{\leq\alpha} \cup P_x^{<\alpha}). \tag{4.1}$$

Thus the value computed by processor P_x at stage α is determined by the history of the values of processors P_y for $y \in x$ at stages $\leq \alpha$ together with the values of processor P_x itself at stages $< \alpha$.

The function $f(x\, /) = f^{M_h^n}(x\, /)$ computed by M_h^n is given by: $f(x\, /) = P_x^{\mathrm{rk}(x)^n}$.

Theorem 4.1 *The safe-recursive functions $f(x\, /)$ are exactly those computed by a machine M_h^n for some $h(/\, x)$ in sRud and some finite $n > 0$.*

Proof. It follows from the safe-recursion scheme that the function $g(x, y\, /) = P_x^{\mathrm{rk}(y)^n}$ is safe-recursive (where P_x^α is defined as above, using h). It follows that $f(x\, /) = g(x, x\, /)$, the function computed by M_h^n, is also safe-recursive. Conversely, in view of the improved characterisation of safe-recursive functions given by Theorem 3.16, it suffices to observe that the M-hierarchy, given by applying the Gödel functions iteratively, is obtained by iteration of a function in *sRud* and therefore is captured by Definition (4.1) above. □

References

[1] Stephen Bellantoni and Stephen Cook. A new recursion-theoretic characterization of the polytime functions. *Comput. Complexity*, 2(2):97–110, 1992.

[2] Leonard Berman. The complexity of logical theories. *Theoretical Computer Science*, 11:71–77, 1980.

[3] Anna R. Bruss and Albert R. Meyer. On the time-space classes and their relation to the theory of real addition. *Theoretical Computer Science*, 11:59–69, 1980.

[4] Keith J. Devlin. *Constructibility.* Perspectives in Mathematical Logic. Springer-Verlag, Berlin, 1984.

[5] Jeanne Ferrante and Charles W. Rackoff. A decision procedure for the first order theory of real addition with order. *SIAM Journal on Computing*, 4(1):69–76, 1975.

[6] S. D. Friedman and P. D. Welch. Two observations concerning infinite time Turing machines. Technical report, I. Dimitriou (ed.), BIWOC 2007 Report, pages 44–47. Hausdorff Centre for Mathematics, Bonn, January 2007.

[7] Joel David Hamkins and Andy Lewis. Infinite time Turing machines. *J. Symbolic Logic*, 65(2):567–604, 2000.

[8] Thomas Jech. *Set Theory.* Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2003. The third millennium edition, revised and expanded.

[9] Ronald Björn Jensen. The fine structure of the constructible hierarchy. *Ann. Math. Logic*, 4:229–308, 1972.

[10] Ronald Björn Jensen and Carol Karp. Primitive recursive set functions. In *Axiomatic Set Theory (Proc. Sympos. Pure Math., Vol. XIII, Part I, Univ. California, Los Angeles, 1967)*, pages 143–176. Amer. Math. Soc., Providence, RI, 1971.

[11] Vladimir Yu. Sazonov. On bounded set theory. In *Logic and Scientific Methods (Florence, 1995)*, volume 259 of *Synthese Lib.*, pages 85–103. Kluwer Academic Publishers, Dordrecht, 1997.

[12] Ralf Schindler. P $\neq$ NP for infinite time Turing machines. *Monatsh. Math.*, 139(4):335–340, 2003.

Strong isomorphism reductions in complexity theory

Samuel R. Buss[†], **Yijia Chen**[‡], **Jörg Flum**[§], **Sy-David Friedman**[¶], **Moritz Müller**[¶]

[†] Department of Mathematics, University of California, San Diego, USA
sbuss@math.ucsd.edu

[‡] Department of Computer Science, Shanghai Jiao Tong University, China
yijia.chen@cs.sjtu.edu.cn

[§] Mathematisches Institut, Albert-Ludwigs Universität Freiburg, Germany
joerg.flum@math.uni-freiburg.de

[¶] Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
sdf@logic.univie.ac.at,moritz.mueller@univie.ac.at

Abstract. We give the first systematic study of strong isomorphism reductions, a notion of reduction more appropriate than polynomial time reduction when, for example, comparing the computational complexity of the isomorphim problem for different classes of structures. We show that the partial ordering of its degrees is quite rich. We analyze its relationship to a further type of reduction between classes of structures based on purely comparing for every n the number of nonisomorphic structures of cardinality at most n in both classes. Furthermore, in a more general setting we address the question of the existence of a maximal element in the partial ordering of the degrees.

Introduction

In many areas of computational complexity, polynomial time reduction is the appropriate notion for comparing the complexity of problems. However, suppose that we face, for example, the problem of comparing the complexity of the isomorphism problem for two classes C and D of graphs. Here

$$\textsc{Iso}(C) := \big\{(\mathcal{A},\mathcal{B}) \mid \mathcal{A},\mathcal{B} \in C \text{ and } \mathcal{A} \cong \mathcal{B}\big\}$$

is the isomorphism problem for C (more precisely, the set of positive instances of this problem) and $\textsc{Iso}(D)$ is defined analogously. Probably we would not accept a polynomial time computable function $f\colon C \times C \to D \times D$ with

$$(\mathcal{A},\mathcal{B}) \in \textsc{Iso}(C) \iff f(\mathcal{A},\mathcal{B}) \in \textsc{Iso}(D)$$

as the right notion of reduction in this context but we would seek a *strong isomorphism reduction*, that is, a polynomial time computable function $f\colon C \to D$ with

$$\mathcal{A} \cong \mathcal{B} \iff f(\mathcal{A}) \cong f(\mathcal{B}). \tag{0.1}$$

This paper is devoted to the study of this type of reduction. For us the motivation for this study came from various areas:

Published in *Journal of Symbolic Logic* 76(4):1381–1402, 2011. The authors thank the John Templeton Foundation for its support through Grant #13152, *The Myriad Aspects of Infinity.*

[†]The first author has been supported in part by NSF grant DMS-0700533.

Computational complexity: The isomorphism relation (on a class C) is an equivalence relation. In the context of arbitrary equivalence relations a notion of reduction defined analogously as in (0.1) (and that for the isomorphism relation coincides with our notion) has been introduced in [**7**]. However that paper is mainly devoted to other problems (see the end of Section 6 for some more details); concerning the notion of reduction only some open problems are stated in [**7**], problems we address in our paper.

Descriptive set theory: For the isomorphism relation our notion of reduction was first considered by the fourth author (see [**8**]) inspired by the analogous notion from descriptive set theory (see [**9**]). In descriptive set theory, C and D denote classes of structures with universe $\mathbb{N}$ and the function f satisfying (0.1) is required to be Borel (in the topology generated by the first-order definable classes).

Descriptive complexity: The existence of a logic capturing polynomial time remains the central open problem of descriptive complexity theory. For many classes C of graphs (or of other types of structures), one shows that a logic L captures polynomial time *on* C by defining in L an invariantization for C. From the definition of invariantization (given in Section 3), one immediately gets that if C is strongly isomorphism reducible to D, then C has an invariantization if D has one.

This paper contains the first systematic study of strong isomorphism reductions. In Section 2 and Section 3 we introduce our framework, derive some basic properties of strong isomorphism reductions, and explain via invariantizations and canonizations the relationship to logics capturing polynomial time mentioned above. At various places of our analysis, invariantizations and canonizations will be valuable tools. Their relationship and the computational complexity of problems related to these notions have been studied in [**2, 3, 7, 11, 15, 16**].

We denote by $\leq_{\text{iso}}$ the partial ordering on the set of degrees induced by strong isomorphism reductions. In Section 2 we observe that (the degree of) the class of graphs is the $\leq_{\text{iso}}$ maximum element. Furthermore, by Theorem 3.7 we see that some "basic algebraic classes of structures" all have the same strong isomorphism degree. In Section 4 we show that the structure of $\leq_{\text{iso}}$ is rich already when restricting to classes with an invariantization.

Assume that C is strongly isomorphism reducible to D. Since such reductions are computable in polynomial time we know that for some polynomial $p \in \mathbb{N}[X]$ and all $n \in \mathbb{N}$ the number of isomorphism types of structures in C with at most n elements is at most the number of isomorphism types of structures in D with at most $p(n)$ elements. If this condition is satisfied, then following [**8**] we say that C is potentially reducible to D. Already in Section 4 this concept is the main tool to demonstrate the richness of the partial ordering $\leq_{\text{iso}}$. We believe that the notions of strong isomorphism reducibility and that of potential reducibility are distinct but can only show this under the hypothesis U2EXP $\cap$ co-U2EXP $\neq$ 2EXP (see Section 5). It turns out in Section 6 that we would get P $\neq$ #P if we could separate the two notions without any complexity-theoretic assumption.

The isomorphism relation is an equivalence relation in NP. In Section 7 we study reductions (defined in analogy to (0.1)) between arbitrary equivalence relations in NP. In particular, we show that there is a maximum element in the corresponding partial ordering if and only if there is an effective enumeration of these equivalence relations by means of clocked Turing machines. Even if we restrict to equivalence relations in P (= PTIME), we cannot show that a maximum element exists; we can guarantee its existence if a p-optimal propositional proof system exists. The existence of a maximum element for equivalence relations in P was addressed in [**7**, Open Question 4.14].

The authors wish to acknowledge the generous support of the John Templeton Foundation and the Centre de Recerca Matemàtica through the CRM Infinity Project. Sam Buss' work was supported in part by NSF grant DMS-0700533. This article will appear in the December 2011 issue of the Journal of Symbolic Logic.

1 Some preliminaries

Throughout the paper Σ denotes the alphabet $\{0,1\}$, and Σ^* is the set of strings over this alphabet. For $n \in \mathbb{N}$ we denote by 1^n the string $11\dots1$ of length n. An ordered pair (x,y) of strings $x = x_1\dots x_k$, $y = y_1\dots y_\ell$ with $x_1,\dots,y_\ell \in \Sigma$ is coded (identified) with the string $x_1x_1\dots x_kx_k01y_1y_1\dots y_\ell y_\ell$. We do similarly for tuples of arbitrary length. Sometimes statements containing a formulation like "there is a $d \in \mathbb{N}$ such that for all $x \in \Sigma^*$: $\dots \leq |x|^d$" can be wrong for $x \in \Sigma^*$ with $|x| \leq 1$ (here $|x|$ denotes the length of the string x). We trust the reader's common sense to interpret such statements reasonably.

1.1 Structures and classes of structures

A *vocabulary* τ is a finite set of relation symbols, function symbols, and constant symbols. The universe of a τ-structure $\mathcal{A}$ will be denoted by the corresponding Latin letter A and the interpretation of a symbol $s \in \tau$ in $\mathcal{A}$ by $s^{\mathcal{A}}$.

> *All structures in this paper are assumed to be finite and to have* $[n] := \{1,2,\dots,n\}$ *as universe for some* $n \in \mathbb{N}$.

Therefore, in a canonical way we can identify structures with nonempty strings over Σ. In particular, $|\mathcal{A}|$ for a structure $\mathcal{A}$ is the length of the string $\mathcal{A}$. Furthermore, we may assume that for every vocabulary τ there is a polynomial $q_\tau \in \mathbb{N}[X]$ such that $|A| \leq |\mathcal{A}| \leq q_\tau(|A|)$ for every τ-structure $\mathcal{A}$, where for a set M we denote by $|M|$ its cardinality.

A class C of τ-structures is *closed under isomorphism* if, for all structures $\mathcal{A}$ and $\mathcal{B}$,

$$\mathcal{A} \in C \text{ and } \mathcal{A} \cong \mathcal{B} \text{ imply } \mathcal{B} \in C$$

(recall that we restrict to structures with universe $[n]$ for some $n \in \mathbb{N}$).

> *In the rest of the paper C (and D) will always denote a class of structures which is in* P*, is closed under isomorphism, and contains arbitrarily large (finite) structures. Moreover, all structures in a fixed class will have the same vocabulary.*

Examples of such classes are:

- The classes Set, Boole, Field, Group, Abelian, and Cyclic of sets (structures of empty vocabulary), Boolean algebras, fields, groups, abelian groups, and cyclic groups, respectively.
- The class Graph of (undirected and simple) graphs. We view graphs as $\tau_{\textsc{Graph}}$-structures, where $\tau_{\textsc{Graph}} := \{E\}$ for a binary relation symbol E.
- The class Ord of linear orderings. Here we use the vocabulary $\tau_{\textsc{Ord}} := \{<\}$ with a binary relation symbol $<$.
- The class Lop of *L*inear *O*rderings with a distinguished *P*oint and the class Lou of *L*inear *O*rderings with a *U*nary relation. Let $\tau_{\textsc{Lop}} := \tau_{\textsc{Ord}} \cup \{c\}$ with a constant symbol c and $\tau_{\textsc{Lou}} := \tau_{\textsc{Ord}} \cup \{P\}$ with a unary relation symbol P.

Then Lop (Lou) is the class of all τ_{Lop}-structures (τ_{Lou}-structures) $\mathcal{A}$ such that $(A, <^{\mathcal{A}}) \in$ Ord.

There is a natural one-to-one correspondence between strings in Σ^* and structures in Lou, namely the function which assigns to a string $x = x_1 \dots x_n \in \Sigma^*$ the structure $\mathcal{A} \in$ Lou with universe $[n]$, where $<^{\mathcal{A}}$ is the natural ordering on $[n]$ and $P^{\mathcal{A}} := \{i \in [n] \mid x_i = 1\}$.

2 Strong isomorphism reductions

We define the notion of strong isomorphism reduction already indicated in the Introduction and present first examples.

Definition 2.1 Let C and D be classes. We say that C is *strongly isomorphism reducible to* D and write $C \le_{\mathrm{iso}} D$, if there is a function $f\colon C \to D$ computable in polynomial time such that, for all $\mathcal{A}, \mathcal{B} \in C$,

$$\mathcal{A} \cong \mathcal{B} \iff f(\mathcal{A}) \cong f(\mathcal{B}).$$

We then say that f is a *strong isomorphism reduction* from C to D and write $f\colon C \le_{\mathrm{iso}} D$. If $C \le_{\mathrm{iso}} D$ and $D \le_{\mathrm{iso}} C$, denoted by $C \equiv_{\mathrm{iso}} D$, then C and D *have the same strong isomorphism degree.*

Examples 2.2

(a) The map sending a field to its multiplicative group shows that

$$\textsc{Field} \le_{\mathrm{iso}} \textsc{Cyclic}.$$

(b) Cyclic $\le_{\mathrm{iso}}$ Abelian $\le_{\mathrm{iso}}$ Group; more generally, if $C \subseteq D$, then $\mathrm{id}_C\colon C \le_{\mathrm{iso}} D$ for the identity function id_C on C.

(c) Set $\equiv_{\mathrm{iso}}$ Ord $\equiv_{\mathrm{iso}}$ Cyclic.

Remark 2.3 We can reduce the notion of strong isomorphism reduction to the notion of polynomial time reduction. For this, we introduce the problem

Iso(C)
Instance: $\mathcal{A}, \mathcal{B} \in C$.
Problem: Is $\mathcal{A} \cong \mathcal{B}$?

A function $f\colon C \to D$ induces the function $\hat{f}\colon C \times C \to D \times D$ with $\hat{f}(\mathcal{A}, \mathcal{B}) := \big(f(\mathcal{A}), f(\mathcal{B})\big)$. Then

$$f\colon C \le_{\mathrm{iso}} D \iff \hat{f}\colon \textsc{Iso}(C) \le_p \textsc{Iso}(D),$$

where $\hat{f}\colon \textsc{Iso}(C) \le_p \textsc{Iso}(D)$ means that $\hat{f}$ is a polynomial time reduction from Iso(C) to Iso(D).

Of course, it is easy to construct polynomial time reductions from Iso(C) to Iso(D) that are not of the form $\hat{f}$ for some $f\colon C \le_{\mathrm{iso}} D$. Moreover, in Remark 4.2 we shall present classes C and D such that

$$\textsc{Iso}(C) \le_p \textsc{Iso}(D) \text{ but not } C \le_{\mathrm{iso}} D.$$

This answers [**7**, Open Question 4.13].

As already mentioned in the Introduction one of our goals is to study the relation $\le_{\mathrm{iso}}$. First we see that this relation has a maximum element:

Proposition 2.4 *$C \leq_{\text{iso}}$ GRAPH for all classes C.*

Proof. Let τ be a vocabulary and S be the class of all τ-structures. It is well-known that there is a strong isomorphism reduction from S to GRAPH (even a first-order interpretation, e.g. see [**6**, Proposition 11.2.5(i)]). In particular, its restriction to a class C of τ-structures shows that $C \leq_{\text{iso}}$ GRAPH. □

3 Invariantizations and canonizations

One of the central aims of algebra and of model theory is to describe the isomorphism type of a structure by means of an invariant. The underlying notion of invariantization is also relevant in our context. We use it (and the related notion of canonization) to show that most classes of structures mentioned in Section 1.1 have the same strong isomorphism degree (cf. Corollary 3.8).

Definition 3.1 An *invariantization for C* is a polynomial time computable function $\text{Inv}\colon C \to \Sigma^*$ such that, for all $\mathcal{A}, \mathcal{B} \in C$,

$$\mathcal{A} \cong \mathcal{B} \iff \text{Inv}(\mathcal{A}) = \text{Inv}(\mathcal{B}).$$

Lemma 3.2 *If $C \leq_{\text{iso}} D$ and D has an invariantization, then also C has an invariantization.*

Proof. If Inv is an invariantization for D and $f\colon C \leq_{\text{iso}} D$, then $\text{Inv} \circ f$ is an invariantization for C. □

LOU is a maximum class among those with an invariantization:

Proposition 3.3 *For a class C, the following are equivalent:*

(1) *C has an invariantization.*
(2) *$C \leq_{\text{iso}}$ LOU.*
(3) *There is a class D of ordered structures such that $C \leq_{\text{iso}} D$.*

Here, a class D is a class of ordered structures *if its vocabulary contains a binary relation symbol which in all structures of D is interpreted as a linear ordering of the universe.*

Proof. (1) implies (2) by the natural correspondence between strings in Σ^* and structures in LOU. That (2) implies (3) is trivial. To see that (3) implies (1) assume that there is a class D of ordered structures such that $C \leq_{\text{iso}} D$. As ordered structures have no nontrivial automorphisms, every ordered structure $\mathcal{A}$ is isomorphic to a unique structure $\mathcal{A}'$ whose ordering $<^{\mathcal{A}'}$ is the natural linear ordering on its universe $\{1, \ldots, |A'|\}$. Thus the mapping on D defined by $\mathcal{A} \mapsto \mathcal{A}'$ is an invariantization of D. Now we apply Lemma 3.2. □

It is open whether the class GRAPH has an invariantization or equivalently (by Proposition 2.4 and Proposition 3.3) whether LOU is a maximum element of $\leq_{\text{iso}}$. Moreover, it is known [**11, 15**] that an invariantization for GRAPH yields a canonization.

Definition 3.4 A function $\text{Can}\colon C \to C$ computable in polynomial time is a *canonization for C* if

(1) for all $\mathcal{A}, \mathcal{B} \in C$, $\big(\mathcal{A} \cong \mathcal{B} \iff \text{Can}(\mathcal{A}) = \text{Can}(\mathcal{B})\big)$;
(2) for all $\mathcal{A} \in C$, $\mathcal{A} \cong \text{Can}(\mathcal{A})$.

Every class C of ordered structures, in particular LOU, has a canonization. In fact, the mapping $\mathcal{A} \mapsto \mathcal{A}'$ defined for all ordered structures in the previous proof is a canonization for C.

We do not define the notion of a *logic capturing* P *on a class* C (e.g., see [**6**]). However we mention that canonizations and invariantizations are important in descriptive complexity theory as:

Proposition 3.5

(1) *If C has a canonization, then there is a logic capturing* P *on C.*
(2) *If* GRAPH *has an invariantization, then there is a logic capturing* P *(on all finite structures).*

Clearly, every canonization is an invariantization. Often the invariantizations we encounter in mathematics yield canonizations. For example, consider the class FIELD of fields. Then an invariant for a field $\mathcal{K}$ is the pair $(p_{\mathcal{K}}, n_{\mathcal{K}})$, where $p_{\mathcal{K}}$ is its characteristic and $n_{\mathcal{K}}$ its dimension over the prime field. As for every invariant (p, n) one can explicitly construct a canonical field $\mathcal{F}_{p^n}$ of this invariant, we see that the mapping $\mathcal{K} \mapsto \mathcal{F}_{p_{\mathcal{K}}^{n_{\mathcal{K}}}}$ is a canonization. This canonization has a further property, it is a canonization that has a polynomial time enumeration:

Definition 3.6 Let Can be a canonization for the class C. The *enumeration induced by* Can is the enumeration

$$\mathcal{A}_1, \mathcal{A}_2, \ldots$$

of the image $\mathrm{Can}(C)$ of C such that $\mathcal{A}_i <_{\mathrm{lex}} \mathcal{A}_j$[1] for $i < j$. If the mappings $\mathcal{A}_n \mapsto 1^n$ and $1^n \mapsto \mathcal{A}_n$ are computable in polynomial time, then Can *has a polynomial time enumeration.*

Note that the mapping $\mathcal{A}_n \mapsto 1^n$ is computable in polynomial time if and only if we get an invariantization Inv of C by setting

$$\mathrm{Inv}(\mathcal{A}) := 1^n \iff \mathrm{Can}(\mathcal{A}) = \mathcal{A}_n.$$

The classes SET, FIELD, ABELIAN, CYCLIC, ORD, and LOP have canonizations with polynomial time enumerations (for ABELIAN see [**13**], for example). The classes BOOLE and LOU have canonizations but none with a polynomial time enumeration: For BOOLE the function $1^n \mapsto \mathcal{A}_n$ will not be computable in polynomial time, as there are, up to equivalence, "too few" Boolean algebras of cardinality $\leq n$, namely $\lfloor \log n \rfloor$; for LOU the function $\mathcal{A}_n \mapsto 1^n$ won't be computable in polynomial time, as there are "too many" structures in LOU of cardinality $\leq n$, namely $2^{n+1} - 1$.

Theorem 3.7 *Assume that the classes C and D have canonizations with polynomial time enumerations. Then $C \equiv_{\mathrm{iso}} D$.*

Corollary 3.8 *The classes* SET, FIELD, ABELIAN, CYCLIC, ORD, *and* LOP *all have the same strong isomorphism degree.*

Proof of Theorem 3.7. Let C and D be classes with canonizations Can_C and Can_D which have polynomial time enumerations $\mathcal{A}_1, \mathcal{A}_2, \ldots$ and $\mathcal{B}_1, \mathcal{B}_2, \ldots$ respectively. We define a strong isomorphism reduction f from C to D by:

$$f(\mathcal{A}) = \mathcal{B}_n \iff \mathrm{Can}_C(\mathcal{A}) = \mathcal{A}_n.$$

[1] By $<_{\mathrm{lex}}$ we denote the standard (length-)lexicographic ordering on Σ^*.

Hence, $C \leq_{\text{iso}} D$; by symmetry we get $D \leq_{\text{iso}} C$. □

An analysis of the previous proof shows that we already obtain $C \leq_{\text{iso}} D$ if the mappings $\mathcal{A}_n \mapsto 1^n$ and $1^n \mapsto \mathcal{B}_n$ are computable in polynomial time. By this, we get, for example, Boole$\leq_{\text{iso}}$Cyclic.

4 On $\leq_{\text{iso}}$ below Lop

As we have seen that the structure of $\leq_{\text{iso}}$ between Lou and Graph is linked with central open problems of descriptive complexity, we turn our attention to the structure below Lou. In this section we show that there, in fact even below Lop, the structure is quite rich. In fact, this section is devoted to a proof of the following result:[2]

Theorem 4.1 *The partial ordering of the countable atomless Boolean algebra is embeddable into the partial ordering induced by $\leq_{\text{iso}}$ on the degrees of strong isomorphism reducibility below* Lop. *More precisely, let $\mathcal{B}$ be a countable atomless Boolean algebra. Then there is a one-to-one function $b \mapsto C_b$ defined on B such that, for all $b, b' \in B$,*

- *C_b is a subclass of* Lop*;*
- *$b \leq b' \Leftrightarrow C_b \leq_{\text{iso}} C_{b'}$.*

Recall that the partial ordering of an atomless Boolean algebra has infinite antichains and infinite chains, even chains of ordertype the rationals.

Remark 4.2 By the preceding result, for example we see that there exist an infinite $\leq_{\text{iso}}$-antichain of classes C below Lop, whose problems Iso(C) are pairwise equivalent under usual polynomial time reductions. Indeed, even Iso$(C) \in$ P for all $C \subseteq$ Lop.

The reader not interested in the details of the proof of Theorem 4.1 should read until Lemma 4.5 and can then skip the rest of this section. We obtain Theorem 4.1 by comparing the number of isomorphism types of structures with universe of bounded cardinality in different classes. First we introduce the relevant notations and concepts.

For a class C we let $C(n)$ be the subclass consisting of all structures in C with universe of cardinality $\leq n$ and we let $\#C(n)$ be the number of isomorphism types of structures in $C(n)$, more formally,

$$C(n) := \{\mathcal{A} \in C \mid |A| \leq n\} \quad \text{and} \quad \#C(n) := |C(n)/_{\cong}|.$$

Here, for a class of structures S we denote by $S/_{\cong}$ the set of isomorphism classes in S.

Examples 4.3

(1) #Boole$(n) = \lfloor \log n \rfloor$, #Cyclic$(n) = n$ and #Set$(n) =$ #Ord$(n) = n + 1$.
(2) #Lop$(n) = \sum_{i=1}^{n} i = (n+1) \cdot n/2$ and #Lou$(n) = \sum_{i=0}^{n} 2^i = 2^{n+1} - 1$.
(3) For every vocabulary τ there is a polynomial $p_\tau \in \mathbb{N}[X]$ such that $\#C(n) \leq 2^{p_\tau(n)}$ for all $n \in \mathbb{N}$ (see Subsection 1.1).
(4) (E.g., see [**1**]) #Group(n) is superpolynomial but subexponential (more precisely, #Group$(n) \leq n^{O(\log^2 n)}$).

Definition 4.4 A class C is *potentially reducible* to a class D, written $C \leq_{\text{pot}} D$, if there is some polynomial $p \in \mathbb{N}[X]$ such that $\#C(n) \leq \#D(p(n))$ for all $n \in \mathbb{N}$. Of course, by $C \equiv_{\text{pot}} D$ we mean $C \leq_{\text{pot}} D$ and $D \leq_{\text{pot}} C$.

[2] Recall that up to isomorphism there is a unique countable atomless Boolean algebra (e.g., see [**10**]).

The following lemma explains the term potentially reducible.

Lemma 4.5 *If $C \leq_{\mathrm{iso}} D$, then $C \leq_{\mathrm{pot}} D$.*

Proof. Let $f\colon C \leq_{\mathrm{iso}} D$. As f is computable in polynomial time, there is a polynomial p such that for all $\mathcal{A} \in C$ we have $|f(A)| \leq p(|A|)$, where $f(A)$ denotes the universe of $f(\mathcal{A})$. As f strongly preserves isomorphisms, it therefore induces a one-to-one map from $\{\mathcal{A} \in C \mid |A| \leq n\}/_{\cong}$ to $\{\mathcal{B} \in D \mid |B| \leq p(n)\}/_{\cong}$. □

We state some consequences of this simple observation:

Proposition 4.6

(1) $\textsc{Cyclic} \not\leq_{\mathrm{iso}} \textsc{Boole}$ *and* $\textsc{Lou} \not\leq_{\mathrm{iso}} \textsc{Lop}$.

(2) $C \leq_{\mathrm{pot}} \textsc{Lou}$ *for all classes C and* $\textsc{Lou} \equiv_{\mathrm{pot}} \textsc{Graph}$.

(3) *The strong isomorphism degree of* $\textsc{Group}$ *is strictly between that of* $\textsc{Lop}$ *and* $\textsc{Graph}$*, that is,* $\textsc{Lop} \leq_{\mathrm{iso}} \textsc{Group} \leq_{\mathrm{iso}} \textsc{Graph}$*, but* $\textsc{Lop} \not\equiv_{\mathrm{iso}} \textsc{Group}$ *and* $\textsc{Group} \not\equiv_{\mathrm{iso}} \textsc{Graph}$.

(4) *The potential reducibility degree of* $\textsc{Group}$ *is strictly between that of* $\textsc{Lop}$ *and* $\textsc{Lou}$*, that is,* $\textsc{Lop} \leq_{\mathrm{pot}} \textsc{Group} \leq_{\mathrm{pot}} \textsc{Lou}$*, but* $\textsc{Lop} \not\equiv_{\mathrm{pot}} \textsc{Group}$ *and* $\textsc{Group} \not\equiv_{\mathrm{pot}} \textsc{Lou}$.

Proof. Using the previous lemma we see that

- (1) follows by Examples 4.3 (1), (2);
- (2) from Examples 4.3 (2), (3) and Proposition 2.4;
- $\textsc{Group} \leq_{\mathrm{iso}} \textsc{Graph}$ holds by Proposition 2.4 and
$$\textsc{Lop} \leq_{\mathrm{iso}} \textsc{Cyclic} \leq_{\mathrm{iso}} \textsc{Group}$$
by Corollary 3.8 and Example 2.2 (b); the remaining claims in (3) follow from (4) as $\textsc{Lou} \equiv_{\mathrm{pot}} \textsc{Graph}$;
- the first claim follows from the first claim in (3) as $\textsc{Lou} \equiv_{\mathrm{pot}} \textsc{Graph}$; the remaining claims follow from Examples 4.3 (2), (4). □

The following concepts and tools will be used in the proof of Theorem 4.1. We call a function $f\colon \mathbb{N} \to \mathbb{N}$ *value-polynomial* if it is increasing and $f(n)$ can be computed in time $f(n)^{O(1)}$. Let VP be the class of all value-polynomial functions.

For $f \in \mathrm{VP}$ the set
$$C_f := \{\mathcal{A} \in \textsc{Lop} \mid |A| \in \mathrm{im}(f)\}$$
is in P and is closed under isomorphism. As there are exactly $f(k)$ pairwise nonisomorphic structures of cardinality $f(k)$ in $\textsc{Lop}$, we get
$$\#C_f(n) = \sum_{k \in \mathbb{N} \text{ with } f(k) \leq n} f(k).$$
The following proposition contains an essential idea underlying the proof of Theorem 4.1, even though it is not used explicitly. Loosely speaking, if the gaps between consecutive values of $f \in \mathrm{VP}$ "kill" every polynomial, then there are classes C and D with $C \not\leq_{\mathrm{pot}} D$.

Proposition 4.7 *Let $f \in \mathrm{VP}$ and assume that for every polynomial $p \in \mathbb{N}[X]$ there is an $n \in \mathbb{N}$ such that*
$$\sum_{k \in \mathbb{N} \textit{ with } f(2k) \leq n} f(2k) > \sum_{k \in \mathbb{N} \textit{ with } f(2k+1) \leq p(n)} f(2k+1). \tag{4.1}$$

Then C_{g_0} is not potentially reducible to C_{g_1}, where $g_0, g_1\colon \mathbb{N} \to \mathbb{N}$ are defined by $g_0(n) := f(2n)$ and $g_1(n) := f(2n+1)$.

Proof. By contradiction, assume that there is some polynomial $p \in \mathbb{N}[X]$ such that $\#C_{g_0}(n) \leq \#C_{g_1}(p(n))$ for all $n \in \mathbb{N}$. Choose n such that (4.1) holds. Then

$$\#C_{g_0}(n) = \sum_{f(2k)\leq n} f(2k) > \sum_{f(2k+1)\leq p(n)} f(2k+1) = \#C_{g_1}(p(n)),$$

a contradiction. □

Lemma 4.8 *The images of the functions in* VP *together with the finite subsets of* $\mathbb{N}$ *are the elements of a countable Boolean algebra $\mathcal{V}$ (under the usual set-theoretic operations). The factor algebra $\mathcal{V}/_\equiv$, where for $b, b' \in V$*

$$b \equiv b' \iff (b \setminus b') \cup (b' \setminus b) \text{ is finite},$$

is a countable atomless Boolean algebra.

Proof. For a function $f\colon \mathbb{N} \to \mathbb{N}$ we denote by $\mathrm{im}(f)$ the image of f. Using the definition of value-polynomial function we verify that for $f, g \in \mathrm{VP}$ the sets

$$\mathbb{N} \setminus \mathrm{im}(f), \quad \mathrm{im}(f) \cap \mathrm{im}(g), \quad \text{and} \quad \mathrm{im}(f) \cup \mathrm{im}(g)$$

are images of value-polynomial functions provided they are infinite. For example, assume that $\mathbb{N} \setminus \mathrm{im}(f)$ is infinite. We choose an algorithm $\mathbb{A}$ and a polynomial $p \in \mathbb{N}[X]$ such that for every $n \in \mathbb{N}$ the algorithm $\mathbb{A}$ computes $f(n)$ in time $p(f(n))$. Let h be the function enumerating $\mathbb{N} \setminus \mathrm{im}(f)$ in increasing order, that is, $h\colon \mathbb{N} \to (\mathbb{N} \setminus \mathrm{im}(f))$ is increasing and surjective. We show that h is value-polynomial too.

A corresponding algorithm inductively computes pairs $(h(0), m_0)$, $(h(1), m_1), \ldots$ with

$$f(m_n) < h(n) < f(m_n + 1)$$

for all $n \in \mathbb{N}$; if $f(0) > 0$ and hence $h(0) = 0$, we set $(h(0), m_0) = (0, -1)$. For $n \geq 1$ the algorithm gets $(h(n), m_n)$ from $(h(n-1), m_{n-1})$ by the following steps:

1. Let $k := h(n-1) + 1$ and $\ell := m_{n-1}$.
2. Simulate $\mathbb{A}$ on $\ell + 1$ for at most $p(k)$ steps.
3. If $\mathbb{A}$ does not halt or if it outputs $f(\ell+1)$ and $f(\ell+1) > k$, then $(h(n), m_n) = (k, \ell)$.
4. Otherwise (i.e., if $f(\ell+1) = k$), let $k := k+1$ and $\ell := \ell + 1$, and goto 2.

It should be clear that the algorithm yields $(h(n), m_n)$ (more precisely, $(h(0), m_0)$, $(h(1), m_1)$, …, $(h(n), m_n)$) in time polynomial in $h(n)$.

We leave the proof of the remaining claims to the reader. □

The lemma just proved shows that the set of images of functions in VP has a rich structure. We compose the functions in VP with a "stretching" function h, which guarantees that the gaps between consecutive values "kill" every polynomial. Then we can apply the idea of the proof of Proposition 4.7 to show that the set of the $\leq_{\mathrm{pot}}$-degrees has a rich structure too.

We define $h\colon \mathbb{N} \to \mathbb{N}$ by recursion: $h(0) := 0$ and

$$h(n+1) = (h(0) + \cdots + h(n))^n .$$

One easily verifies that h is value-polynomial.

For $f, g \in \mathrm{VP}$ set

$$f \subseteq^* g \iff \mathrm{im}(f) \setminus \mathrm{im}(g) \text{ is finite}.$$

By the homogeneity properties of atomless countable Boolean algebras, to prove Theorem 4.1 it suffices to find a corresponding embedding defined only on the nonzero elements of $\mathcal{V}/_{\equiv}$. In general $f \subseteq^* g$ and $g \subseteq^* f$ do not imply $C_{h\circ f} = C_{h\circ g}$. However, by the following lemma we get an embedding of $\mathcal{V}/_{\equiv}$ into the partial ordering of the $\leq_{\mathrm{iso}}$-degrees as required by Theorem 4.1 by defining the mapping on a set of representatives, more precisely on a set $R \subseteq \mathrm{VP}$ such that

- for every $f \in \mathrm{VP}$ there is exactly one $g \in R$ with $f \subseteq^* g$ and $g \subseteq^* f$.

Lemma 4.9 *The mapping $f \mapsto C_{h\circ f}$ from* VP *to $\{C \subseteq \textsc{Lou} \mid C \text{ a class}\}$ is one-to-one, and for all $f, g \in \mathrm{VP}$:*

(1) *if $C_{h\circ f} \leq_{\mathrm{iso}} C_{h\circ g}$, then $f \subseteq^* g$;*
(2) *if $f \subseteq^* g$ and $g \not\subseteq^* f$, then $C_{h\circ f} \leq_{\mathrm{iso}} C_{h\circ g}$.*

For the proof of Lemma 4.9 we need an appropriate way to invert increasing functions $f\colon \mathbb{N} \to \mathbb{N}$. We define $f^{-1}\colon \mathbb{N} \to \mathbb{N}$ by

$$f^{-1}(n) := \max\{i \mid f(i) \leq n\},$$

where we set $\max \emptyset := 0$. We collect some properties of this inverse in the following lemma, whose simple proof we omit. We denote by $\mathrm{id}_{\mathbb{N}}$ the identity function on $\mathbb{N}$.

Lemma 4.10

(1) *If $f\colon \mathbb{N} \to \mathbb{N}$ is increasing, then f^{-1} is nondecreasing, $f^{-1} \leq \mathrm{id}_{\mathbb{N}}$, $f^{-1} \circ f = \mathrm{id}_{\mathbb{N}}$ and $f(f^{-1}(n)) \leq n$ for all $n \geq f(0)$.*
(2) *If $f, g\colon \mathbb{N} \to \mathbb{N}$ are increasing, then $(f \circ g)^{-1} = g^{-1} \circ f^{-1}$.*
(3) *If $f \in \mathrm{VP}$, then f^{-1} is computable in polynomial time.*

A further notation is useful. For $f\colon \mathbb{N} \to \mathbb{N}$ let $f^{\Sigma}\colon \mathbb{N} \to \mathbb{N}$ be defined by

$$f^{\Sigma}(n) := \sum_{i \leq n} f(i).$$

Lemma 4.11 *Let $f, g\colon \mathbb{N} \to \mathbb{N}$ be functions and assume that the function g is increasing. Then $(f \circ g)^{\Sigma} \leq f^{\Sigma} \circ g$.*

Proof. This is seen by direct calculation:

$$(f \circ g)^{\Sigma}(n) = \sum_{i \leq n} f(g(i)) = \sum_{\substack{i \leq g(n) \\ i \in \mathrm{im}(g)}} f(i) \leq \sum_{i \leq g(n)} f(i) = f^{\Sigma} \circ g(n);$$

here the second equality uses that g is increasing. □

Furthermore observe that:

Lemma 4.12 *If $f \in \mathrm{VP}$, then for all $n \in \mathbb{N}$ we have $\#C_f(n) = (f^{\Sigma} \circ f^{-1})(n)$.*

Proof of Lemma 4.9. The mapping $f \mapsto C_{h\circ f}$ is one-to-one: Assume $C_{h\circ f} = C_{h\circ g}$. Then $\mathrm{im}(h \circ f) = \mathrm{im}(h \circ g)$ and thus, $\mathrm{im}(f) = \mathrm{im}(g)$ as h is one-to-one. Since f and g are both increasing, this yields $f = g$. We prove the remaining statements of Lemma 4.9 by the following two claims.

Claim 1 Let $f, g \in \mathrm{VP}$ and $f \subseteq^* g$ and $g \not\subseteq^* f$. Then $C_{h\circ f} \leq_{\mathrm{iso}} C_{h\circ g}$.

Proof of Claim 1: By our assumptions, the set $\mathrm{im}(h \circ f) \setminus \mathrm{im}(h \circ g)$ is finite (as $f \subseteq^* g$ implies $h \circ f \subseteq^* h \circ g$) and (by injectivity of h) the set $\mathrm{im}(h \circ g) \setminus \mathrm{im}(h \circ f)$ is infinite. Then

$C_{h\circ f} \leq_{\mathrm{iso}} C_{h\circ g}$ is witnessed by a function sending the (up to $\cong$) finitely many structures in $C_{h\circ f} \setminus C_{h\circ g}$ to $C_{h\circ g} \setminus C_{h\circ f}$ and which is the identity on all other structures in $C_{h\circ f}$.

Claim 2 Let $f, g \in \mathrm{VP}$ and $f \not\subseteq^* g$. Then $C_{h\circ f} \not\leq_{\mathrm{iso}} C_{h\circ g}$.

Proof of Claim 2: By contradiction assume $C_{h\circ f} \leq_{\mathrm{iso}} C_{h\circ g}$. Then $C_{h\circ f}$ is potentially reducible to $C_{h\circ g}$ by Lemma 4.5. Hence there is $p \in \mathbb{N}[X]$ such that $\#C_{h\circ f}(n) \leq \#C_{h\circ g}(p(n))$ for all $n \in \mathbb{N}$. We show that this is wrong for some n. For this purpose we choose k such that

$$(4.2) \qquad g(0) < f(k), \quad p(h(f(k))) < h(f(k)+1), \text{ and } f(k) \in \mathrm{im}(f) \setminus \mathrm{im}(g)$$

(by the definition of h and the assumption $f \not\subseteq^* g$ such a k exists). Then we get

$$\begin{aligned}
&\#C_{h\circ g}(p(h(f(k))))\\
&\quad = (h\circ g)^{\Sigma} \circ (h\circ g)^{-1}(p(h(f(k)))) \quad \text{(by Lemma 4.12)}\\
&\quad = (h\circ g)^{\Sigma} \circ (g^{-1} \circ h^{-1})(p(h(f(k)))) \quad \text{(by Lemma 4.10(2))}\\
&\quad \leq (h\circ g)^{\Sigma} \circ g^{-1}(f(k)) \quad (\text{by } p(h(f(k))) < h(f(k)+1) \text{ —see (4.2)—}\\
&\qquad\qquad\qquad\qquad\qquad\qquad \text{and by definition of } h^{-1})\\
&\quad = (h\circ g)^{\Sigma} \circ g^{-1}(f(k)-1) \quad (\text{as } f(k) \notin \mathrm{im}(g))\\
&\quad \leq h^{\Sigma} \circ g \circ g^{-1}(f(k)-1) \quad \text{(by Lemma 4.11)}\\
&\quad \leq h^{\Sigma}(f(k)-1) \quad (\text{by Lemma 4.10(1) as } g(0) < f(k))\\
&\quad < h(f(k)) \quad (\text{by definition of } h)\\
&\quad \leq \#C_{h\circ f}(h(f(k))) \quad (\text{by definition of } \#C_{h\circ f}). \qquad \square
\end{aligned}$$

5 Strong isomorphism reducibility and potential reducibility

We know that Graph$\leq_{\mathrm{pot}}$ Lou while Graph$\leq_{\mathrm{iso}}$ Lou is equivalent to Graph having an invariantization (cf. Propositions 4.6(2) and 3.3). However, so far in all concrete examples of classes C and D, for which we know the status of $C \leq_{\mathrm{iso}} D$ and of $C \leq_{\mathrm{pot}} D$, we had that

$$C \leq_{\mathrm{iso}} D \iff C \leq_{\mathrm{pot}} D.$$

So the question arises whether the relations of strong isomorphism reducibility and potential reducibility coincide. Recall that we require the classes C and D to be closed under isomorphism and decidable in polynomial time. Generalizing the proof idea of Theorem 3.7, we shall see in the next section that indeed the relations $\leq_{\mathrm{iso}}$ and $\leq_{\mathrm{pot}}$ coincide if $\mathrm{P} = \#\mathrm{P}$. We believe that they are distinct but could only show:

Theorem 5.1 *If* $\mathrm{U2EXP} \cap \text{co-U2EXP} \neq \mathrm{2EXP}$, *then the relations of strong isomorphism reducibility and that of potential reducibility are distinct.*

Recall that

$$\mathrm{2EXP} := \mathrm{DTIME}\Big(2^{2^{n^{O(1)}}}\Big) \quad \text{and} \quad \mathrm{N2EXP} := \mathrm{NTIME}\Big(2^{2^{n^{O(1)}}}\Big).$$

The complexity class U2EXP consists of those $Q \in \mathrm{N2EXP}$ for which there is a nondeterministic Turing machine of type N2EXP that for every $x \in Q$ has exactly one accepting run. Finally, $\text{co-U2EXP} := \{\Sigma^* \setminus Q \mid Q \in \mathrm{U2EXP}\}$.

The rest of this section is devoted to a proof of this result. We explain the underlying idea: Assume $Q \in$ U2EXP $\cap$ co-U2EXP. We construct classes C and D which contain structures in the same cardinalities and which contain exactly two nonisomorphic structures in these cardinalities. Therefore they are potentially reducible to each other. While it is trivial to exhibit two nonisomorphic structures in C of the same cardinality, from any two concrete nonisomorphic structures in D we obtain information on membership in Q for all strings of a certain length. If $C \leq_{\text{iso}} D$, we get concrete nonisomorphic structures in D (in time allowed by 2EXP) by applying the strong isomorphism reduction to two nonisomorphic structures in C and therefore obtain $Q \in$ 2EXP.

Proof of Theorem 5.1. Let $Q \in$ U2EXP $\cap$ co-U2EXP. Then there exists a nondeterministic Turing machine $\mathbb{M}$ and a constant $d \geq 2$ such that (M1)–(M5) hold:

(M1) The machine $\mathbb{M}$ has three terminal states *'yes,' 'no,'* and *'maybe'.*

(M2) For $x \in \Sigma^*$, every run of $\mathbb{M}$ on input x stops after *exactly* $2^{2^{|x|^d}}$ many steps.

(M3) For $x \in Q$ exactly one run of $\mathbb{M}$ on x stops in 'yes' and none in 'no'.

(M4) For $x \notin Q$ exactly one run of $\mathbb{M}$ on x stops in 'no' and none in 'yes'.

(M5) The machine $\mathbb{M}$ has exactly two different choices for the next step in every nonterminal state.

We say that a run of $\mathbb{M}$ *takes a decision* if it ends in 'yes' or in 'no'.

For $n \in \mathbb{N}$ we set $\ell(n) := 2^{2^{n^d}}$. For $x \in \Sigma^n$, by (M2) and (M5), every run of $\mathbb{M}$ on input x can be identified with a binary string $r \in \{0,1\}^{\ell(n)}$. Conversely, from such a string r we can determine a run of $\mathbb{M}$ on x.

Let $m(n) := 2^n$ and $x_1, x_2, \ldots, x_{m(n)}$ be the enumeration of all strings of Σ^n in the lexicographic ordering. We call a binary string s of length $m(n) \cdot \ell(n) = 2^n \cdot 2^{2^{n^d}}$ a *decision string* if for every $i \in [m(n)]$ the ith substring of s of length $\ell(n)$ corresponds to a run of $\mathbb{M}$ on x_i taking a decision; more precisely, if we have $s = s_1\hat{\ }s_2\hat{\ }\cdots\hat{\ }s_{m(n)}$ with $|s_i| = \ell(n)$ for $i \in [m(n)]$, then s_i corresponds to a run of $\mathbb{M}$ on x_i taking a decision. By our assumptions (M3) and (M4) we get:

(5.1) for every $n \in \mathbb{N}$ there is exactly one decision string of length $m(n) \cdot \ell(n)$.

We turn every string s of length $m(n) \cdot \ell(n)$ into a structure $\mathcal{A}(s)$ over the vocabulary $\tau = \{\mathit{One}, \mathit{Zero}, R\}$, where *One* and *Zero* are unary relation symbols and R is a binary relation symbol. Let

$$A(s) := [m(n) \cdot \ell(n)],$$
$$R^{\mathcal{A}(s)} := \big\{(j, j+1) \mid j \in [m(n) \cdot \ell(n) - 1]\big\}.$$

For s a decision string, let

$$\mathit{One}^{\mathcal{A}(s)} := \big\{j \mid j \in [m(n) \cdot \ell(n)] \text{ and the } j\text{th bit of } s \text{ is one}\big\},$$
$$\mathit{Zero}^{\mathcal{A}(s)} := \big\{j \mid j \in [m(n) \cdot \ell(n)] \text{ and the } j\text{th bit of } s \text{ is zero}\big\},$$

and let $\mathit{One}^{\mathcal{A}(s)} = \mathit{Zero}^{\mathcal{A}(s)} = \emptyset$ otherwise. By (5.1) for every $s, s' \in \{0,1\}^{m(n)\cdot\ell(n)}$

$$\mathcal{A}(s) \not\cong \mathcal{A}(s') \iff \text{exactly one of } s \text{ and } s' \text{ is a decision string.} \tag{5.2}$$

Let D_n be the class containing, up to isomorphism, the structures $\mathcal{A}(s)$ with $s \in \{0,1\}^{m(n)\cdot\ell(n)}$. The following is straightforward.

(D1) The universe of every structure in D_n has cardinality $m(n) \cdot \ell(n)$.
(D2) $|D_n/_{\cong}| = 2$.

We set

$$D := \bigcup_{n \in \mathbb{N}} D_n.$$

Finally, we let

$$C := \bigcup_{n \in \mathbb{N}} C_n,$$

where for $n \in \mathbb{N}$ every structure in the class C_n is isomorphic to the complete graph $K_{m(n) \cdot \ell(n)}$ on $m(n) \cdot \ell(n)$ vertices or to its complement $\overline{K}_{m(n) \cdot \ell(n)}$. Then:

(C1) The universe of every structure in C_n has cardinality $m(n) \cdot \ell(n)$.
(C2) $|C_n/_{\cong}| = 2$.

Hence, $C \leq_{\text{pot}} D$.

Claim If $f \colon C \leq_{\text{iso}} D$, then there is $n_0 \in \mathbb{N}$ such that, for all $n \geq n_0$,

$$f\left(C_n/_{\cong}\right) = D_n/_{\cong}. \tag{5.3}$$

By this equality we mean:

- $f(\mathcal{A}) \in D_n$ for every $\mathcal{A} \in C_n$;
- for every $\mathcal{B} \in D_n$ there exists an $\mathcal{A} \in C_n$ such that $f(\mathcal{A}) \cong \mathcal{B}$.

Proof of the claim: First observe that by (C2) and (D2) it suffices to show that $f(C_n) \subseteq D_n$ for all sufficiently large $n \in \mathbb{N}$. As f is computable in polynomial time there is $c \in \mathbb{N}$ such that for every $n \in \mathbb{N}$ and $\mathcal{A} \in C_n$

$$\text{the universe of } f(\mathcal{A}) \text{ has} \leq \left(2^n \cdot 2^{2^{n^d}}\right)^c \text{ elements.}$$

We choose $n_0 \in \mathbb{N}$ such that, for all $n \geq n_0$,

$$\left(2^n \cdot 2^{2^{n^d}}\right)^c < 2^{n+1} \cdot 2^{2^{(n+1)^d}}.$$

Hence, for $n \geq n_0$,

$$f\left(\bigcup_{q \leq n} C_q\right) \subseteq \bigcup_{q \leq n} D_q.$$

As $\bigcup_{q \leq n} C_q$ and $\bigcup_{q \leq n} D_q$ contain, up to isomorphism, the same number of structures, the Claim follows.

Now assume $f \colon C \leq_{\text{iso}} D$. Then the following algorithm $\mathbb{A}$ witnesses that $Q \in 2\text{EXP}$. Let n_0 be as in the Claim. For $x \in \Sigma^n$ with $n \geq n_0$ the algorithm $\mathbb{A}$ computes the structures

$$f\left(K_{m(n) \cdot \ell(n)}\right) \quad \text{and} \quad f\left(\overline{K}_{m(n) \cdot \ell(n)}\right);$$

they are nonisomorphic and in D_n by the Claim. In particular, by (5.2) we get a run of $\mathbb{M}$ on input x taking a decision; the algorithm $\mathbb{A}$ answers accordingly. □

6 If strong isomorphism reducibility and potential reducibility are distinct then $\mathrm{P} \neq \#\mathrm{P}$

In the previous section we have seen that under some complexity-theoretic assumption the two notions of reduction (strong isomorphism reducibility and potential reducibility) are distinct. One might wonder whether we can separate them without any such complexity-theoretic assumption. We show in this section that this would settle some open problem in complexity theory; more precisely, we show the statement of the title of this section.[3] In particular, by Proposition 4.6(2), if Lou is not a maximum element of $\leq_{\mathrm{iso}}$, then $\mathrm{P} \neq \#\mathrm{P}$. We prove the main result in a more general setting.

For a class C consider the equivalence relation $E(C)$ on Σ^* induced by the isomorphism relation, that is,

$$\begin{aligned} E(C) :=& \big\{(\mathcal{A},\mathcal{B}) \mid \mathcal{A},\mathcal{B} \in C \text{ and } \mathcal{A} \cong \mathcal{B}\big\} \\ & \cup \big\{(x,y) \mid x,y \in \Sigma^*,\ x \notin C \text{ and } y \notin C\big\}. \end{aligned} \tag{6.1}$$

Of course, $E(C)$ is in NP. In this section we consider arbitrary such equivalence relations on Σ^* and show that the corresponding two notions of reduction coincide if $\mathrm{P} = \#\mathrm{P}$. We start by introducing all relevant concepts; we do not restrict ourselves to equivalence relations in NP, but consider equivalence relations in an arbitrary complexity class (for an equivalence relation E on Σ^* we also write xEy for $(x,y) \in E$).

Definition 6.1

(1) Let CC be an arbitrary complexity class. Then we denote by CC(eq) the set of equivalence relations E on Σ^* with $E \in \mathrm{CC}$.

(2) Let E and E' be equivalence relations on Σ^*. We say that E is *strongly equivalence reducible to* E' and write $E \leq_{\mathrm{eq}} E'$, if there is a function $f\colon \Sigma^* \to \Sigma^*$ computable in polynomial time such that, for all $x,y \in \Sigma^*$,

$$xEy \iff f(x)E'f(y).$$

We then say that f is a *strong equivalence reduction* from E to E' and write $f\colon E \leq_{\mathrm{eq}} E'$.

Clearly, $E(C) \in \mathrm{NP(eq)}$ for every class C of structures; furthermore, $E(\mathrm{Lou}) \in \mathrm{P(eq)}$. Let Prop and Taut denote the set of all formulas of propositional logic and the set of tautologies, respectively. Note that $E_{\mathrm{equiv}} \in \text{co-NP(eq)}$, where

$$\begin{aligned} E_{\mathrm{equiv}} :=& \{(\alpha,\beta) \mid \alpha,\beta \in \mathrm{Prop} \text{ and } (\alpha \leftrightarrow \beta) \in \mathrm{Taut}\} \\ & \cup \{(x,y) \mid x,y \notin \mathrm{Prop}\}. \end{aligned}$$

Clearly, if C and D are classes of structures as in the previous sections, then

$$C \leq_{\mathrm{iso}} D \iff E(C) \leq_{\mathrm{eq}} E(D).$$

We generalize the notion of potential reducibility to equivalence relations.

[3] Recall that $\mathrm{P} = \#\mathrm{P}$ means that for every polynomial time nondeterministic Turing machine $\mathbb{M}$ the function $f_{\mathbb{M}}$ such that $f_{\mathbb{M}}(x)$ is the number of accepting runs of $\mathbb{M}$ on $x \in \Sigma^*$ is computable in polynomial time. The class $\#\mathrm{P}$ consists of all the functions $f_{\mathbb{M}}$.

Definition 6.2 Let E and E' be equivalence relations on Σ^*. We say that E is *potentially reducible* to E' and write $E \leq_{\mathrm{pot}} E'$ if there is a $p \in \mathbb{N}[X]$ such that for all $n \in \mathbb{N}$ the number $|\Sigma^{\leq n}/E|$ of E-equivalence classes containing a string in

$$\Sigma^{\leq n} := \big\{x \in \Sigma^* \mid |x| \leq n\big\}$$

is at most $\big|\Sigma^{\leq p(n)}/E'\big|$.

Due to our definition (6.1) of $E(C)$, the new notion coincides with the old one for equivalence relations of the form $E(C)$:

Proposition 6.3 *Let C and C' be classes. Then*

$$C \leq_{\mathrm{pot}} C' \iff E(C) \leq_{\mathrm{pot}} E(C').$$

Proof. Recall that the empty string is not (the encoding of) a structure. Let C be a class of τ-structures and C' a class of τ'-structures. By the assumptions made in Subsection 1.1, there are polynomials $p_\tau, p_{\tau'} \in \mathbb{N}[X]$ such that for every τ-structure $\mathcal{A}$

$$|A| \leq |\mathcal{A}| \leq p_\tau(|A|) \tag{6.2}$$

and for every τ'-structure $\mathcal{B}$

$$|B| \leq |\mathcal{B}| \leq p_{\tau'}(|B|). \tag{6.3}$$

Assume first that $C \leq_{\mathrm{pot}} C'$, say $\#C(n) \leq \#C'(p(n))$ for some polynomial p. Then

$$|\Sigma^{\leq n}/E(C)| \leq \#C(n) + 1 \leq \#C'(p(n)) + 1 \leq |\Sigma^{\leq p_{\tau'}(p(n))}/E(C')|$$

(the first inequality holds by (6.1) and (6.2), the last one by (6.1) and (6.3)). Conversely, assume that $E(C) \leq_{\mathrm{pot}} E(C')$, say $|\Sigma^{\leq n}/E(C)| \leq \big|\Sigma^{\leq p(n)}/E(C')\big|$ with $p \in \mathbb{N}[X]$. Then

$$\#C(n) + 1 \leq |\Sigma^{\leq p_\tau(n)}/E(C)| \leq \Big|\Sigma^{\leq p(p_\tau(n))}/E(C')\Big| \leq \#C'(p(p_\tau(n))) + 1. \qquad \square$$

Along the lines of the proof of Lemma 4.5, one shows that $E \leq_{\mathrm{eq}} E'$ implies $E \leq_{\mathrm{pot}} E'$. For equivalence relations we can show that $\leq_{\mathrm{eq}}$ is finer than $\leq_{\mathrm{pot}}$ under weaker assumptions than that of Theorem 5.1:

Proposition 6.4 *If* $\mathrm{NP} \neq \mathrm{P}$, *then the relations of strong equivalence reduction and that of potential reducibility do not coincide on* $\mathrm{NP(eq)}$.

Proof. Assume $Q \in \mathrm{NP} \setminus \mathrm{P}$. We define E_Q by

$$xE_Qy \iff \Big(x = y \text{ or } \big(x = b\hat{}z \text{ and } y = (1-b)\hat{}z \text{ for some } z \in Q \text{ and } b \in \Sigma\big)\Big).$$

By our assumptions on Q, we have $E_Q \in \mathrm{NP(eq)}$. We let E be the identity on Σ^*. Clearly, $E_Q \leq_{\mathrm{pot}} E$. As $Q \notin \mathrm{P}$, we get $E_Q \not\leq_{\mathrm{eq}} E$, as any $f\colon E_Q \leq_{\mathrm{eq}} E$ would yield a polynomial time decision procedure for Q. $\square$

Generalizing the proof idea of Theorem 3.7 we show:

Theorem 6.5 *If the relations of strong equivalence reduction and that of potential reducibility do not coincide on* $\mathrm{NP(eq)}$, *then* $\mathrm{P} \neq \#\mathrm{P}$.

To prove this theorem we first generalize the notions of canonization and of enumeration induced by a canonization.

Definition 6.6 Let $E \in \mathrm{CC(eq)}$. A function $\mathrm{Can}\colon \Sigma^* \to \Sigma^*$ is a *canonization for E* if it is polynomial time computable and

(1) for all $x, y \in \Sigma^*$, $\big(xEy \iff \mathrm{Can}(x) = \mathrm{Can}(y)\big)$;
(2) for all $x \in \Sigma^*$, $xE\ \mathrm{Can}(x)$.

Let Can be a canonization of E. The *enumeration induced by* Can is the enumeration

$$x_1, x_2 \dots$$

of $\mathrm{Can}(\Sigma^*)$ such that $x_i <_{\mathrm{lex}} x_j$ for $i < j$.

If E has a canonization, then $E \in \mathrm{P}$: to decide whether xEy we compute $\mathrm{Can}(x)$ and $\mathrm{Can}(y)$ and check whether $\mathrm{Can}(x) = \mathrm{Can}(y)$.

Now it is easy to explain the idea underlying the proof of Theorem 6.5. First we show that (under the assumption P = NP) every E $\in$ P(eq) has a canonization Can_E. Then, given $E, E' \in \mathrm{P(eq)}$, we define a strong equivalence reduction $f\colon \Sigma^* \to \Sigma^*$ from E to E' as follows: Let $x \in \Sigma^*$. If $\mathrm{Can}_E(x)$ is the ith element in the enumeration induced by Can_E, then we let $f(x)$ be the ith element in the enumeration induced by $\mathrm{Can}_{E'}$. By the properties of canonizations it should be clear that

$$xEy \iff f(x)E'f(y)$$

(we can even replace $f(x)E'f(y)$ by $f(x) = f(y)$). So it remains to show (under suitable assumptions) that f is computable in polynomial time and to show that every equivalence relation has a canonization.

The following lemma was already proven in [**2**].

Lemma 6.7 *If* P = NP*, then every* $E \in \mathrm{P(eq)}$ *has a canonization; in fact, then the mapping sending each* $x \in \Sigma^*$ *to the* $\leq_{\mathrm{lex}}$*-first member of the* E*-equivalence class of* x *is a canonization.*

Proof. Let $E \in \mathrm{P(eq)}$ and assume P = NP. Then we know that the polynomial hierarchy collapses, P = PH. So it suffices to show that the mapping defined in the statement of this lemma can be computed by an alternating polynomial time algorithm $\mathbb{A}$ with a constant number of alternations. This is easy: on input $x \in \Sigma^*$ the algorithm $\mathbb{A}$ guesses existentially $y \in \Sigma^*$ with $|y| \leq |x|$ and xEy; then $\mathbb{A}$ guesses universally a further $z \in \Sigma^*$ with $|z| \leq |x|$ and xEz; if $y \leq_{\mathrm{lex}} z$, then $\mathbb{A}$ outputs y otherwise it rejects. □

Lemma 6.8 *Let* $E \in \mathrm{P(eq)}$ *be an equivalence relation with a canonization* Can*. Then the following problem is in* #P*:*

Instance:	$x \in \Sigma^*$.
Problem:	Compute i (in binary) such that $\mathrm{Can}(x)$ is the ith element in the enumeration induced by Can.

Proof. Consider a nondeterministic polynomial time algorithm $\mathbb{A}$ which on input $x \in \Sigma^*$ runs as follows: It first computes the string $y := \mathrm{Can}(x)$. Then $\mathbb{A}$ guesses a string $z \in \Sigma^*$ with $|z| \leq |y|$. Finally it accepts if $\mathrm{Can}(z) = z$ and $z \leq_{\mathrm{lex}} y$. It should be clear that the number of accepting runs of $\mathbb{A}$ on x is

$$|\{z \mid z \leq_{\mathrm{lex}} \mathrm{Can}(x) \text{ and } \mathrm{Can}(z) = z\}|. \qquad \square$$

Proof of Theorem 6.5. Assume that P = #P. Let $E, E' \in \mathrm{NP(eq)}$ be equivalence relations and assume that $E \leq_{\mathrm{pot}} E'$, that is, $|\Sigma^{\leq n}/E| \leq |\Sigma^{\leq p(n)}/E'|$ for some polynomial p and all $n \in \mathbb{N}$. We show $E \leq_{\mathrm{eq}} E'$.

As P = #P, we have P = NP. Hence $E, E' \in \mathrm{P(eq)}$. Therefore, by Lemma 6.7 there are canonizations Can_E of E and $\mathrm{Can}_{E'}$ of E' and there are polynomial time algorithms

$\mathbb{A}$ and $\mathbb{A}'$ that solve the problem of the preceding lemma for E and E', respectively. The following nondeterministic polynomial time algorithm computes an $f\colon E{\leq_{\text{eq}}}E'$. On input $x \in \Sigma^*$, it computes $\text{Can}_E(x)$ and $n := |\text{Can}_E(x)|$ and guesses a string $x' \in \Sigma^{\leq p(n)}$ with $\text{Can}_{E'}(x') = x'$. Simulating $\mathbb{A}$ and $\mathbb{A}'$, it checks whether $\text{Can}_E(x)$ and x' are at the same position in the enumeration induced by Can_E and in the enumeration induced by $\text{Can}_{E'}$, respectively; in the positive case it outputs x', otherwise it rejects. As $|\Sigma^{\leq n}/E| \leq |\Sigma^{\leq p(n)}/E'|$ such an $x' \in \Sigma^{\leq p(n)}$ with $\text{Can}_{E'}(x') = x'$ at the same position as $\text{Can}_E(x)$ exists. As $\text{P} = \text{NP}$, the function f is computable in polynomial time. □

We briefly point to the papers [**2, 3, 7**] that deal with related problems. Let Inv(eq) be the class of equivalence relations having an invariantization (defined in analogy to Definition 3.1), Can(eq) the class of equivalence relations having a canonization and finally, Lexfirst(eq) the class of equivalence relations having a canonization that maps every string to the $\leq_{\text{lex}}$-first element of its equivalence class. Clearly

$$\text{Lexfirst(eq)} \subseteq \text{Can(eq)} \subseteq \text{Inv(eq)} \subseteq \text{P(eq)}. \tag{6.4}$$

Lemma 6.7 shows that $\text{Lexfirst(eq)} = \text{Can(eq)} = \text{Inv(eq)} = \text{P(eq)}$ if $\text{P} = \#\text{P}$. Blass and Gurevich [**2**], for example, prove that $\text{Lexfirst(eq)} \neq \text{Can(eq)}$ unless the polynomial hierarchy collapses, and Fortnow and Grochow [**7**] show that $\text{Can(eq)} = \text{Inv(eq)}$ would imply that integers can be factored in probabilistic polynomial time. Blass and Gurevich [**2, 3**] compare the complexity of the "problems underlying the definition of the sets in (6.4)". Finally, the book [**16**], among other things, deals with the question whether two propositional formulas are logically equivalent up to a permutation of their variables. It is not hard to see that the isomorphism problem for a class C can be rephrased in these terms; however no analogue of $\leq_{\text{iso}}$ is considered in [**16**].

7 On maximum elements in P(eq) and NP(eq)

In this section we study whether there is a maximum element with respect to strong equivalence reductions in the classes P(eq) and NP(eq), that is, in the classes of deterministic and nondeterministic polynomial time equivalence relations. We already mentioned that the existence of a maximum element in P(eq) is mentioned as [**7**, Open Question 4.14]; the notion of strong equivalence reduction was already introduced in that paper and called kernel reduction there.

Let Sat be the set of satisfiable propositional formulas. Consider the NP-equivalence relation

$$E_{\text{sat}} := \big\{(\alpha,\beta) \mid \alpha,\beta \in \textsc{Prop} \text{ and } \big(\alpha = \beta \text{ or } \alpha,\beta \in \textsc{Sat}\big)\big\};$$

more precisely, to get an equivalence relation on Σ^*, we define E_{sat} to be

$$\big\{(\alpha,\beta) \mid \alpha,\beta \in \textsc{Prop} \text{ and } \big(\alpha = \beta \text{ or } \alpha,\beta \in \textsc{Sat}\big)\big\} \cup \big\{(x,y) \mid x,y \notin \textsc{Prop}\big\}.$$

However, henceforth if we speak of an equivalence relation E whose field $\text{Fld}(E) := \{x \mid (x,x) \in E\}$ is a proper subset of Σ^*, we identify it with the equivalence relation $E \cup \big\{(x,y) \mid x,y \in \Sigma^* \setminus \text{Fld}(E)\big\}$. We use E_{sat} to show:

Proposition 7.1 *If the polynomial hierarchy* PH *does not collapse, then* $E(\textsc{Graph})$ *is not a maximum element in* $(\text{NP(eq)}, \leq_{\text{eq}})$*; in fact, then* $E_{\text{sat}} \not\leq_{\text{eq}} E(\textsc{Graph})$.

Proof. For $\alpha \in \textsc{Prop}$ and a propositional variable X, we have $(\alpha \in \textsc{Sat} \Leftrightarrow \alpha E_{\text{sat}} X)$. By contradiction, assume that $f\colon E_{\text{sat}}{\leq_{\text{eq}}}E(\textsc{Graph})$. We have $f(X) \in \textsc{Graph}$; otherwise, SAT $\in$ P, which contradicts our assumption that the polynomial hierarchy does not collapse. Then, for every $\alpha \in$ PROP,

$$\alpha \in \text{SAT} \iff f(\alpha) \cong f(X).$$

Thus E(GRAPH) would be NP-complete. It is well-known [**4**] that this fact implies that $\Sigma_2^p = \text{PH}$. □

We show that the existence of a maximum element in $(\text{NP(eq)}, \leq_{\text{eq}})$ is equivalent to the existence of an effective enumeration of NP(eq). This result is also true for P(eq) and co-NP(eq). Effective enumerations of problems have been used to characterize promise classes possessing complete languages, that is, maximum elements under polynomial time reductions (e.g., see [**12, 14**]). Even though we are dealing with a different type of reduction, our method is similar. To state our precise result we introduce some notions. A deterministic or nondeterministic Turing machine $\mathbb{M}$ is *clocked* (more precisely, *polynomially time-clocked*), if (the code of) $\mathbb{M}$ contains a natural number $\text{time}(\mathbb{M})$ such that $n^{\text{time}(\mathbb{M})}$ is a bound for the running time of $\mathbb{M}$ on inputs of length n. So, by this definition, all runs of a clocked machine are of polynomial length. Of course, the function $\mathbb{M} \mapsto \text{time}(\mathbb{M})$, defined on the set of clocked machines, is computable in polynomial time.

Definition 7.2 Let CC $\in \{\text{P}, \text{NP}, \text{co-NP}\}$. Let L be a set of languages L with $L \subseteq \Sigma^*$. We say that

$$L_0, L_1, \ldots$$

is a CC-*enumeration of* L *by clocked Turing machines*, if $\text{L} = \{L_0, L_1, \ldots\}$ and there is a computable function $\mathbb{M}$ defined on $\mathbb{N}$ such that $\mathbb{M}(i)$ for $i \in \mathbb{N}$ is (the code of) a clocked Turing machine of type CC accepting L_i.

Proposition 7.3 *Let* CC $\in \{\text{P}, \text{NP}, \text{co-NP}\}$. *Then the following are equivalent:*

(1) $(\text{CC(eq)}, \leq_{\text{eq}})$ *has a maximum element.*

(2) *There is a* CC*-enumeration* $E_0, E_1, \ldots$ *of* CC*(*eq*) by clocked Turing machines.*

Proof. (1) ⇒ (2): Assume that E is a maximum in $(\text{CC(eq)}, \leq_{\text{eq}})$ and let $\mathbb{M}_{\max}$ be a Turing machine of type CC accepting E. Of course, there is a computable function $\mathbb{M}'$ such that $\mathbb{M}'(i)$ for $i \in \mathbb{N}$ is a deterministic clocked Turing machine computing a function $f_i \colon \Sigma^* \to \Sigma^*$ such that $f_0, f_1, \ldots$ is an enumeration of all polynomial time computable functions from Σ^* to Σ^*. We define the machine $\mathbb{M}_{\max} \circ \mathbb{M}'(i)$ in a straightforward manner such that it decides

$$E_i := \big\{(x, y) \mid (f_i(x), f_i(y)) \in E\big\}.$$

We let $\mathbb{M}$ be the function defined on $\mathbb{N}$ with $\mathbb{M}(i) := \mathbb{M}_{\max} \circ \mathbb{M}'(i)$. As from a polynomial bounding $\mathbb{M}_{\max}$ and $\text{time}(\mathbb{M}'(i))$ we get a time bound for $\mathbb{M}(i)$, we can assume that $\mathbb{M}(i)$ is clocked. It should be clear that $E_0, E_1, \ldots$ has the desired properties.

(2) ⇒ (1): Let $E_0, E_1, \ldots$ be as in (2) and let $\mathbb{M}$ be a corresponding computable function. By padding if necessary, we may assume that the graph $\{(1^i, 1^{|\mathbb{M}(i)|}) \mid i \in \mathbb{N}\}$ is decidable in polynomial time and that $i \leq |\mathbb{M}(i)|$ for all $i \in \mathbb{N}$. We define the relation E as follows (for better reading we denote here, and in the proof of Lemma 7.6, the string 1^ℓ, that is the string $11\ldots1$ of length ℓ, by $\langle \ell \rangle$):

$$E := \Big\{ \Big((\mathbb{M}(i), x, \langle (2 + 2|x|)^{\text{time}(\mathbb{M}(i))} \rangle), (\mathbb{M}(i), y, \langle (2 + 2|y|)^{\text{time}(\mathbb{M}(i))} \rangle) \Big) \;\Big|\; i \in \mathbb{N} \text{ and } (x, y) \in E_i \Big\}.$$

By the effectivity properties of $\mathbb{M}$, we have $E \in \mathrm{CC(eq)}$ (more precisely $E \cup \{(x, y) \mid x, y \in \Sigma^* \setminus \mathrm{Fld}(E)\} \in \mathrm{CC(eq)}$). Clearly, for $i \in \mathbb{N}$ the mapping $x \mapsto (\mathbb{M}(i), x, \langle(2{+}2|x|)^{\mathrm{time}(\mathbb{M}(i))}\rangle)$ is a strong equivalence reduction from E_i to E, hence E is a maximum element. □

Below we will show that $(\mathrm{NP(eq)}, \leq_{\mathrm{eq}})$ has a maximum element if NP = co-NP. Note that we do not even know whether $(\mathrm{P(eq)}, \leq_{\mathrm{eq}})$ has a maximum element. The main result concerning this problem that we have reads as follows (later we recall the definition of p-optimal proof system):

Theorem 7.4 *If* Taut *has a p-optimal proof system, then* $(\mathrm{P(eq)}, \leq_{\mathrm{eq}})$ *has a maximum element.*

The following observations will lead to a proof of this result.

Definition 7.5 Let $\mathbb{M}$ be a deterministic or nondeterministic Turing machine and $n \in \mathbb{N}$. The machine $\mathbb{M}$ *defines an equivalence relation on* $\Sigma^{\leq n}$ if the set

$$\big\{(x, y) \mid x, y \in \Sigma^{\leq n} \text{ and } \mathbb{M} \text{ accepts } (x, y)\big\}$$

is an equivalence relation on $\Sigma^{\leq n}$.

An analysis of the complexity of the first of the following problems will be crucial for our purposes.

Equiv(P)
Instance: A deterministic clocked Turing machine $\mathbb{M}$ and $n \in \mathbb{N}$.
Problem: Does $\mathbb{M}$ define an equivalence relation on $\Sigma^{\leq n}$?

Equiv(NP)
Instance: A nondeterministic clocked Turing machine $\mathbb{M}$ and $n \in \mathbb{N}$.
Problem: Does $\mathbb{M}$ define an equivalence relation on $\Sigma^{\leq n}$?

Lemma 7.6

(1) *If* $(\mathbb{M}, n) \in$ Equiv(P) *is solvable by a deterministic algorithm in time* $n^{f(\|\mathbb{M}\|)}$ *for some function* $f \colon \mathbb{N} \to \mathbb{N}$*, then* P(eq) *has a maximum element.*[4]

(2) *If* $(\mathbb{M}, n) \in$ Equiv(NP) *is solvable by a nondeterministic algorithm in time* $n^{f(\|\mathbb{M}\|)}$ *for some function* $f \colon \mathbb{N} \to \mathbb{N}$*, then* NP(eq) *has a maximum element.*

Proof. Let $\mathbb{A}$ be an algorithm, deterministic for (1) and nondeterministic for (2), witnessing that $(\mathbb{M}, n) \in$ Equiv(P) in (1) and $(\mathbb{M}, n) \in$ Equiv(NP) in (2) is solvable in time $n^{f(\|\mathbb{M}\|)}$ for some $f \colon \mathbb{N} \to \mathbb{N}$. An equivalence relation E_0 on Σ^* is defined by letting uE_0v hold if and only if

$$u = v \text{ or } \Big(u = \big(\mathbb{M}, x, (2 + 2 \cdot |x|)^{\mathrm{time}(\mathbb{M})}, 1^t\big) \text{ and}$$
$$v = \big(\mathbb{M}, x', (2 + 2 \cdot |x'|)^{\mathrm{time}(\mathbb{M})}, 1^{t'}\big) \text{ and (i) – (iii) are fulfilled}\Big),$$

where

(i) $\mathbb{M}$ is a clocked Turing machine of type CC, where CC = P for (1) and CC = NP for (2);

(ii) $\mathbb{A}$ accepts $(\mathbb{M}, |x|)$ in at most t steps and $(\mathbb{M}, |x'|)$ in at most t' steps;

(iii) $\mathbb{M}$ accepts (x, x').

[4] By $\|\mathbb{M}\|$ we denote the length of a reasonable encoding of $\mathbb{M}$ by a string of Σ^*.

Clearly, $E_0 \in \mathrm{CC(eq)}$. We show that E_0 is a maximum element. Let $E \in \mathrm{CC(eq)}$ be arbitrary and let $\mathbb{M}$ be a clocked Turing machine deciding E. Then

$$x \mapsto (\mathbb{M}, x, (2 + 2 \cdot |x|)^{\mathrm{time}(\mathbb{M})}, \langle |x|^{f(\|\mathbb{M}\|)} \rangle)$$

is computable in polynomial time and hence a strong equivalence reduction from E to E_0. □

Theorem 7.7 *The following hold:*

(1) *If* $\mathrm{E} = \mathrm{NE}$*, then* $\mathrm{P(eq)}$ *has a maximum element.*

(2) *If* $\mathrm{NP} = \text{co-NP}$*, then* $\mathrm{NP(eq)}$ *has a maximum element.*

Proof. (1) We may assume that n is written in binary in the instances $(\mathbb{M}, n)$ of Equiv(P) (and that a string of length $\|\mathbb{M}\| \cdot \log n$ is given as an additional input). We consider the following nondeterministic algorithm $\mathbb{A}$ accepting the complement of Equiv(P). On input $(\mathbb{M}, n)$, it guesses one of the three axioms of an equivalence relation, say, the transitivity axiom; then $\mathbb{A}$ guesses $x, y, z \in \Sigma^{\leq n}$, it simulates $\mathbb{M}$ on input (x, y), on input (y, z), and on input (x, z) and accepts if $\mathbb{M}$ accepts the first two inputs but not the third one. As we may assume that $\|\mathbb{M}\| \geq \mathrm{time}(\mathbb{M})$, the algorithm $\mathbb{A}$ runs in time $\|\mathbb{M}\| \cdot n^{O(\mathrm{time}(\mathbb{M}))} = 2^{O(\|\mathbb{M}\| \cdot \log n)}$. By the assumption $\mathrm{E} = \mathrm{NE}$, there is a deterministic algorithm deciding the complement of Equiv(P) and hence Equiv(P) itself in time $2^{O(\|\mathbb{M}\| \cdot \log n)}$. Now our claim follows from the preceding lemma.

(2) The following alternating algorithm $\mathbb{A}$ decides the complement of Equiv(NP): On input $(\mathbb{M}, n)$ (again we may assume that $\|\mathbb{M}\| \geq \mathrm{time}(\mathbb{M})$), it existentially guesses one of the three axioms of an equivalence relation, say, the transitivity axiom; then $\mathbb{A}$ existentially guesses $x, y, z \in \Sigma^{\leq n}$ and runs of $\mathbb{M}$ accepting (x, y) and (y, z); furthermore it yields the string $\langle n^{\|\mathbb{M}\|} \rangle$. Finally $\mathbb{A}$ universally simulates $\mathbb{M}$ on input (x, z) and accepts if $\mathbb{M}$ rejects. The algorithm $\mathbb{A}$ has one alternation. By our assumption $\mathrm{NP} = \text{co-NP}$, its universal part (an algorithm of type co-NP with inputs $\mathbb{M}$, (x, z), and $\langle n^{\|\mathbb{M}\|} \rangle$) can be simulated by a nondeterministic algorithm running in time $n^{O(\|\mathbb{M}\|)}$. Altogether we get a nondeterministic algorithm accepting (the complement of) Equiv(NP) in time $n^{O(\|\mathbb{M}\|)}$. Now our claim follows from the preceding lemma. □

We consider the *acceptance problem for nondeterministic Turing machines*:

Acc$_{\leq}$	
Instance:	A nondeterministic Turing machine $\mathbb{M}$ and $n \in \mathbb{N}$.
Problem:	Does $\mathbb{M}$ accept the empty input tape in $\leq n$ steps?

Lemma 7.8 *The following are equivalent:*

(1) $(\mathbb{M}, n) \in \mathrm{Acc}_{\leq}$ *is solvable deterministically in time* $n^{f(\|\mathbb{M}\|)}$ *for some* $f \colon \mathbb{N} \to \mathbb{N}$.

(2) $(\mathbb{M}, n) \in$ Equiv(P) *is solvable deterministically in time* $n^{f(\|\mathbb{M}\|)}$ *for* $f \colon \mathbb{N} \to \mathbb{N}$.

Proof. (1) ⇒ (2): Assume that $(\mathbb{M}, n) \in \mathrm{Acc}_{\leq}$ (where $\mathbb{M}$ is a nondeterministic machine and $n \in \mathbb{N}$) can be solved by an algorithm $\mathbb{A}$ in time $n^{f(\|\mathbb{M}\|)}$ for some $f \colon \mathbb{N} \to \mathbb{N}$. Then the following algorithm $\mathbb{B}$ will witness that Equiv(P) is decidable in the time claimed in (2). Let $(\mathbb{M}, n)$ be an instance of Equiv(P), in particular $\mathbb{M}$ is a deterministic clocked Turing machine. We may assume that $\mathbb{M}$ on input (x, y) runs for exactly $|(x, y)|^{\mathrm{time}(\mathbb{M})}$ steps. Let $\widetilde{\mathbb{M}}$ be the nondeterministic Turing machine that on empty input tape, in the first phase guesses one of the three axioms of an equivalence relation, say, the transitivity axiom; then in the second phase $\widetilde{\mathbb{M}}$ guesses $x, y, z \in \Sigma^*$; finally in the third phase it simulates $\mathbb{M}$ on input (x, y), on input (y, z), and on input (x, z) and accepts if $\mathbb{M}$ accepts the first two inputs but not the third one. We can assume that $\widetilde{\mathbb{M}}$ does this simulation in such a way that it runs for exactly $(2 + 2 \cdot \max\{x, y, z\})^{\mathrm{time}(\mathbb{M})}$ steps on each of the tuples (x, y), (y, z), and (x, z).

Let k_1, $k_2(x, y, z)$, and $k_3(x, y, z)$ be the exact time $\widetilde{\mathbb{M}}$ uses for the first phase, the second phase and the third phase, respectively. As indicated for the third phase we may arrange things in such a way that there are (nonconstant) polynomials k_2', k_3' such that

$$\begin{aligned} k_2(x, y, z) &= k_2'(\max\{|x|, |y|, |z|\}) \text{ and} \\ k_3(x, y, z) &= k_3'(\max\{|x|, |y|, |z|\}) \end{aligned}$$

and such that if for example $\widetilde{\mathbb{M}}$ has chosen the symmetry axiom and $x, y \in \Sigma^*$, then $k_2'(\max\{|x|, |y|\})$ is also the exact number of steps $\widetilde{\mathbb{M}}$ uses for the second phase. As k_2' and k_3' are increasing functions, we get

$$(\mathbb{M}, n) \notin \textsc{Equiv} \iff (\widetilde{\mathbb{M}}, k + k_2'(n) + k_3'(n)) \in \textsc{Acc}_{\leq},$$

which gives the desired bound.

(2) ⇒ (1): For a nondeterministic Turing machine $\mathbb{M}$ let $\widehat{\mathbb{M}}$ be the deterministic Turing machine that on input (x, y) with $x, y \in \Sigma^*$ first checks whether $x \neq y$; if so, it accepts; if $x = y$, it simulates the $|x|$ steps of a run of $\mathbb{M}$ on empty input tape, namely the steps corresponding to (the bits in) x and rejects if in these $|x|$ steps $\mathbb{M}$ accepts; otherwise $\widehat{\mathbb{M}}$ accepts. Thus, for every $n \in \mathbb{N}$,

$$\begin{aligned} &(\mathbb{M}, n) \in \textsc{Acc}_{\leq} \iff \\ &\qquad \widehat{\mathbb{M}} \text{ does not define an equivalence relation on } \Sigma^{\leq n}. \end{aligned}$$

As from the definition of $\widehat{\mathbb{M}}$ we immediately get a polynomial time bound, we can assume that $\widehat{\mathbb{M}}$ is clocked, so that the preceding equivalence immediately gives the claim. □

A *proof system* for Taut is a surjective function $S \colon \Sigma^* \to \textsc{Taut}$ computable in polynomial time. The proof system S for Taut is *p-optimal* if for every proof system S' for Taut there is a polynomial time computable $T \colon \Sigma^* \to \Sigma^*$ such that, for all $w \in \Sigma^*$,

$$S(T(w)) = S'(w).$$

It is not known whether there is a p-optimal proof system for Taut, even though it is conjectured there is no such p-optimal proof system. In **[5]** it has been shown that:

Proposition 7.9 *The following are equivalent:*

(1) *There is a p-optimal proof system for* Taut.

(2) $(\mathbb{M}, n) \in \textsc{Acc}_{\leq}$ *is solvable in time* $n^{f(\|\mathbb{M}\|)}$ *for some function* $f \colon \mathbb{N} \to \mathbb{N}$.

Proof of Theorem 7.4. If there is a p-optimal proof system for Taut, by the previous proposition and Lemma 7.8 we see that $(\mathbb{M}, n) \in \textsc{Equiv}(\mathrm{P})$ is solvable in time $n^{f(\|\mathbb{M}\|)}$ for some function $f \colon \mathbb{N} \to \mathbb{N}$. Now the claim follows from Lemma 7.6. □

References

[1] H. U. Besche, B. Eick and E. A. O'Brien. The groups of order at most 2000, *Electronic Research Announcements of the American Mathematical Society*, 7:1–4, 2001.

[2] A. Blass and Y. Gurevich. Equivalence relations, invariants, and normal forms. *SIAM Journal of Computing*, 13:682–689, 1984.

[3] A. Blass and Y. Gurevich. Equivalence relations, invariants, and normal forms, II. Lecture Notes in Computer Science, 171:24–42, 1984.

[4] R. B. Boppana, J. Hastad and S. Zachos. Does co-NP have short interactive proofs? *Information Processing Letters*, 25(2):127–132, 1987.

[5] Y. Chen and J. Flum. On p-optimal proof systems and logics for PTIME. In *Proceedings of the 37th International Colloquium on Automata, Languages and Programming (ICALP'10)*, Lecture Notes in Computer Science 6199, pp. 321–332, Springer, 2010.

[6] H.-D. Ebbinghaus and J. Flum. *Finite Model Theory*, Second Edition. Perspectives in Mathematical Logic, Springer 1999.

[7] L. Fortnow and J. Grochow. Complexity classes of equivalence problems revisited, 2009, `arXiv:0907.4775v1` [cs.CC].

[8] S. Friedman. Descriptive set theory for finite structures, Lecture at the Kurt Gödel Research Center, 2009, Available at `http://www.logic.univie.ac.at/~sdf /papers/wien-spb.pdf`.

[9] H. Friedman and L. Stanley. A Borel reducibility theory for classes of countable structures, *Journal Symbolic Logic*, 54:894–914, 1989.

[10] S. Givant and P. Halmos. *Introduction to Boolean Algebras*, Springer, 2008.

[11] Y. Gurevich. From invariants to canonization. *Bulletin of the European Association for Theoretical Computer Science* 63, pp. 115–119, 1997.

[12] J. Hartmanis and L. Hemachandra. Complexity classes without machines: On complete languages for UP. *Theoretical Computer Science*, 58:129–142, 1988.

[13] T. Kavitha. Efficient algorithms for abelian group isomorphism and related problems. In *Proceedings of the 23rd Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS'02)*, Lecture Notes in Computer Science 2914, pp. 277–288, Springer, 2003.

[14] W. Kowalczyk. Some connections between presentability of complexity classes and the power of formal systems of reasoning. In *Proceedings of Mathematical Foundations of Computer Science, (MFCS'84)*, Lecture Notes in Computer Science 176, Springer, pp. 364–369, 1984.

[15] G. Miller. Isomorphism testing for graphs of bounded genus. In *Proceedings of the 12th Annual ACM Symposium on Theory of Computing (STOC'80)*, 225–235, 1980.

[16] T. Thierauf. *The computational complexity of equivalence and isomorphism problems.* Lecture Notes in Computer Science, 1852, Springer, 2000.

On Σ^1_1 equivalence relations over the natural numbers

Ekaterina B. Fokina[†], **Sy-David Friedman**[†]

[†] Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
efokina@logic.univie.ac.at, sdf@logic.univie.ac.at

Abstract. We study the structure of Σ^1_1 equivalence relations on hyperarithmetical subsets of ω under reducibilities given by hyperarithmetical or computable functions, called h-reducibility and FF-reducibility, respectively. We show that the structure is rich even when one fixes the number of properly Σ^1_1 (i.e., Σ^1_1 but not Δ^1_1) equivalence classes. We also show the existence of incomparable Σ^1_1 equivalence relations that are complete as subsets of $\omega \times \omega$ with respect to the corresponding reducibility on sets. We study complete Σ^1_1 equivalence relations (under both reducibilities) and show that existence of infinitely many properly Σ^1_1 equivalence classes that are complete as Σ^1_1 sets (under the corresponding reducibility on sets) is necessary but not sufficient for a relation to be complete in the context of Σ^1_1 equivalence relations.

Introduction

In [**7, 9**] the notion of hyperarithmetical and computable reducibility of Σ^1_1 equivalence relations on hyperarithmetical subsets of ω was used to study the question of completeness of natural equivalence relations on hyperarithmetical classes of computable structures as a special class of Σ^1_1 equivalence relations on ω. In this paper we use this approach to study the structure of Σ^1_1 equivalence relations on ω as a whole.

In descriptive set theory, the study of definable equivalence relations under Borel reducibility has developed into a rich area. The notion of Borel reducibility allows one to compare the complexity of equivalence relations on Polish spaces; for details see e.g. [**11, 14, 15**]. As proved by Louveau and Velickovic in [**19**], the partial order of inclusion modulo finite sets on $\mathcal{P}(\omega)$ can be embedded into the partial order of Borel equivalence relations modulo Borel reducibility. Thus, the structure of Borel equivalence relations under $\leq_B$ is shown to be very rich.

In computable model theory, equivalence relations have also been a subject of study, e.g. [**2, 4, 16**], etc. In these papers, equivalence relations of rather low complexity were studied (computable, Σ^0_1, Π^0_1, having degree in the Ershov hierarchy). In [**7**] Σ^1_1 equivalence relations on computable structures were investigated. The authors used the notions of hyperarithmetical and computable reducibility of Σ^1_1 equivalence relations on ω to estimate the complexity of natural equivalence relations on hyperarithmetical classes of computable structures.

In this paper we take up the general theory of Σ^1_1 equivalence relations on hyperarithmetical subsets of ω. We show that the general structure of Σ^1_1 equivalence relations on hyperarithmetical subsets of ω under reducibilities given by hyperarithmetical or computable functions is very rich. Namely, the structure of Σ^1_1 sets under hyperarithmetical

Appeared in Mathematical Logic Quarterly, vol. 58 (2012), no. 1–2, pp. 113–124.

[†]The authors would like to thank the FWF (Austrian Research Fund) for supporting this research through Grants number P 19898 - N18 and M 1188 - N13, as well as the John Templeton Foundation for its support through Project #13152, Myriad Aspects of Infinity.

many-one-reducibility (*hm*-reducibility) is embeddable into the structure of Σ^1_1 equivalence relations under reducibility given by a hypearithmetical function. Moreover, this embedding can be taken to have range within the class of Σ^1_1 equivalence relations with a unique properly Σ^1_1 equivalence class. Furthermore, we show that there are properly Σ^1_1 equivalence relations with only finite equivalence classes, and there are Σ^1_1 relations with exactly n properly Σ^1_1 equivalence classes, for $n \leq \omega$. We also show that a Σ^1_1 equivalence relation with infinitely many properly (moreover, *hm*-complete) Σ^1_1 classes need not be complete with respect to the hyperarithmetical reducibility.

1 Background

Here we list some definitions and facts that we will use throughout the paper. We assume familiarity with the main notions from recursion theory. The standard references are [**22, 24**].

1.1 Linear orderings

Definition 1.1 Let K be a class of structures closed under isomorphism and K^c be the set of its computable members.

(1) An *enumeration* of $K^c/_{\cong}$ is a sequence $(\mathcal{A}_n)_{n\in\omega}$ of elements of K^c representing each isomorphism type in K^c at least once.

(2) An enumeration $(\mathcal{A}_n)_{n\in\omega}$ of $K^c/_{\cong}$ is *computable (hyperarithmetical)* if there is a computable (hyperarithmetical) function f which, for every n, gives a computable index $f(n)$ for the computable structure $\mathcal{A}_n$.

As proved in [**13**]:

Proposition 1.2 *There exists a computable enumeration of all isomorphism types for computable linear orderings.*

Thus, we can consider ω as a set of effective codes for computable linear orderings. We will denote by L_n the n-th computable linear order in this enumeration. We will abbreviate the set of codes for linear orderings as LO and the set of codes for well-orderings as WO.

Theorem 1.3 (e.g. [**22**, Chapter 16, Corollary XXa]) *The set WO is a Π^1_1-complete set; moreover, there exists a computable function $f(z,x)$ such that, for every z, the Π^1_1 set with the Π^1_1 index z is 1-reducible to WO by the function $\lambda x[f(z,x)]$.*

In view of Theorem 1.3 one can think about Π^1_1 sets in the following way. Let A be a Π^1_1 set and let m be its Π^1_1 index. Then for every $x \in A$, the ordinal isomorphic to $L_{f(m,x)}$ may be considered as "the level" at which the membership of x is determined.

Theorem 1.4 (Bounding) *For each computable ordinal α, let WO_α denote the set of codes for computable well-orderings isomorphic to an ordinal less than α. Then if F is a hyperarithmetical function from a hyperarithmetical subset of ω into WO, there exists a computable α such that the range of F is contained in WO_α.*

Theorem 1.5 (Uniformization) *Every Π^1_1 binary relation on $X \times Y$, where $X, Y \subseteq \omega$ are hyperarithmetical, contains a Π^1_1 (hyperarithmetical) function with the same domain.*

1.2 Reducibilities on Σ^1_1 equivalence relations

The following definitions were introduced in [**7**]:

Definition 1.6 Let E, E' be Σ^1_1 equivalence relations on hyperarithmetical subsets $X, Y \subseteq \omega$, respectively.

(1) The relation E is *h-reducible* to E', denoted by $E \leq_h E'$, iff there exists a hyperarithmetical function f such that, for all $x, y \in X$,

$$xEy \iff f(x)E'f(y).$$

(2) The relation E is *FF-reducible* to E', denoted by $E \leq_{FF} E'$, iff there exists a partial computable function f with $X \subseteq \operatorname{dom}(f), f[X] \subseteq Y$ such that, for all $x, y \in X$,

$$xEy \iff f(x)E'f(y).$$

Remark 1.7 A definition analogous to that of FF-reducibility was introduced in [**1**] for the case of c.e. equivalence relations.

Definition 1.8 We say that equivalence relations E, F are *h-equivalent* (*FF-equivalent*), denoted by $E \equiv_h F$ ($E \equiv_{FF} F$, respectively), if $E \leq_h F$ and $F \leq_h E$ ($E \leq_{FF} F$ and $F \leq_{FF} E$, respectively).

Obviously, every Σ^1_1 equivalence relation on a hyperarithmetical subset of ω is h-equivalent to a Σ^1_1 equivalence relation on ω. For FF-reducibility the situation is different:

Fact 1.9 There exists a Σ^1_1 equivalence relation E on a hyperarithmetical subset X of ω such that for no Σ^1_1 equivalence relation E' on ω, $E \equiv_{FF} E'$.

Proof. Consider an arbitrary Σ^1_1 equivalence relation on a hyperarithmetical set X and suppose there exists a relation E' on ω such that $E \equiv_{FF} E'$. Let f be a computable function which witnesses $E' \leq_{FF} E$. Then $f(\omega)$ is a c.e. subset of X. Therefore if a Σ^1_1 equivalence relation is defined on a hyperarithmetical set without a c.e. subset, it cannot be FF-equivalent to an equivalence relation on ω. □

From [**12**], every *computable* equivalence relation on ω is FF-equivalent to one of the following:

(1) For some finite n, the equivalence relation $x \equiv y \ mod\ n$, which defines a computable equivalence relation with exactly n infinite equivalence classes and no finite classes.

(2) The equality relation, which defines a computable equivalence relation with infinitely many classes of size one, and no other classes.

Thus, the partial ordering of the computable equivalence structures, modulo the FF-reducibility, is isomorphic to $\omega + 1$.

In the current paper we are mostly interested in properly Σ^1_1 equivalence relations, i.e., equivalence relations that are Σ^1_1 but not Δ^1_1. The reason is the following:

Fact 1.10 Let id_ω denote the equality on ω.

(1) $\mathrm{id}_\omega \leq_h E$ for any Σ^1_1 equivalence relation E with infinitely many equivalence classes.

(2) Any Δ^1_1 equivalence relation on a hyperarithmetical subset of ω is h-reducible to id_ω.

Proof. Define a function $f\colon \omega \to X$, where $X = \mathrm{dom}(E)$ is hyperarithmetical, in the following way:

$$f(x) = \mu y[y \in X \& \bigwedge_{z \leq x} \neg f(z)Ey].$$

By its definition, f is a Π^1_1 function with $\mathrm{dom}(f) = \omega$, thus f is a hyperarithmetical function. Obviously, $x = y \iff f(x)Ef(y)$.

To prove the second statement, let E be a Δ^1_1 equivalence relation on a hyperarithmetical set X. Without loss of generality we assume $0 \notin X$. Consider a function $f(x)$ defined on X in the following way:

$$f(x) = \mu z[xEz].$$

For $x \notin X$ define $f(x) = 0$. Then the function f is hyperarithmetical and $xEy \iff f(x) = f(y) \neq 0$. □

Therefore all the Δ^1_1 equivalence relations on ω with infinitely many equivalence classes are h-equivalent.

The question we study in the present paper is the following:

Question 1.11 How complicated is the structure of all Σ^1_1 equivalence relations on ω under h-reducibility (or FF-reducibility)?

1.3 Hyperarithmetical many-one reducibility on Σ^1_1 sets

In what follows we use the standard notions of m-reducibility and 1-reducibility [**24**]:

Definition 1.12

(1) A set $A \subseteq \omega$ is *many-one reducible* (*m-reducible*) to a set $B \subseteq \omega$, denoted by $A \leq_m B$, if there exists a computable function f such that, for every $n \in \omega$,

$$n \in A \iff f(n) \in B.$$

(2) A set $A \subseteq \omega$ is 1-*reducible* to a set $B \subseteq \omega$, denoted by $A \leq_1 B$, if A is m-reducible to B via a 1-1 computable function.

These reducibilities will be useful for the study of the structure of Σ^1_1 equivalence relations with respect to FF-reducibility.

Consider a hyperarithmetical version of the m-reducibility on subsets of ω. It will play an important role in the investigation of complexity of the structure of Σ^1_1 equivalence relations relative to h-reducibility.

Definition 1.13 Let A, B be subsets of ω. We say that A is *hyperarithmetically m-reducible* to B, denoted by $A \leq_{hm} B$, iff there exists a hyperarithmetical function f with $A \subseteq \mathrm{dom}(f)$, such that, for every $n \in \omega$,

$$n \in A \iff f(n) \in B.$$

Every equivalence relation can also be considered as a set of pairs, thus, compared to other sets via m- or hm-reducibilities. The following is straightforward:

Fact 1.14 Let E, F be Σ^1_1 equivalence relations on hyperarithmetical subsets of ω.

(1) If $E \leq_{FF} F$ then $E \leq_m F$.
(2) If $E \leq_h F$ then $E \leq_{hm} F$.

We state that the structure of hm-degrees of Σ^1_1 subsets of ω is rather complicated.

Theorem 1.15 *The countable atomless Boolean algebra may be embedded into the hm-degrees of* Π^1_1 *subsets of* ω.

Proof. We start as in the proof of [**24**, Chapter IX, Theorem 2.1]. Let $(\alpha_i)_{i\in\omega}$ be a uniformly computable sequence of computable subsets of ω which form a dense Boolean algebra under $\cup, \cap$. For each $i \in \omega$, we are going to build a Π^1_1 set A_i such that the mapping

$$\alpha \longmapsto A_\alpha = \{\langle i, x\rangle \mid i \in \alpha,\ x \in A_i\}$$

gives the desired embedding, i.e.,

(1) $\alpha \subseteq \beta$ iff $A_\alpha \leq_{hm} A_\beta$;
(2) $\deg(A_{\alpha\cap\beta}) \leq \deg(A_\alpha), \deg(A_\beta)$;
(3) $\deg(A_{\alpha\cup\beta}) \geq \deg(A_\alpha), \deg(A_\beta)$.

Notice that the implication from left to right of the first property, as well as the second and the third properties follow from the definition of $\mathcal{A}_{\alpha_i}$. To ensure the implication from right to left of the first property, we will use the ideas of metarecursion [**23**]. We will build the Π^1_1 sets A_i's in ω_1^{CK} steps in such a way that no A_i is hm-reducible to the set $A_{\neq i} = \{\langle k, x\rangle \mid k \in \omega, k \neq i, x \in A_k\}$.

The whole construction will take now ω_1^{CK} steps, but as only the Π^1_1 subsets of ω are considered, there will be only ω-many requirements. Thus, each of them may be injured only finitely many times. This approach is used, for example, in [**23**, Chapter VI, Theorems 2.1, 2.4].

Let $(f_j)_{j\in\omega}$ be a universal Π^1_1 enumeration of all Π^1_1 functions on ω. Such an enumeration exists, e.g., by [**22**, Section 16.5]. Recall that the hyperarithmetical functions are the total Π^1_1 functions. Then our requirements are:

$$R_{i,j} : A_i \neq f_j^{-1}[A_{\neq i}] \text{ and } A_i \text{ is co-infinite.}$$

We build our sets in stages $\sigma < \omega_1^{\mathrm{CK}}$. We assign requirements to stages in such a way that each requirement is assigned to cofinally many stages. At stage 0 we do nothing.

At stage $0 < \sigma < \omega_1^{\mathrm{CK}}$, let $R_{i,j}$ be the current requirement. The strategy to satisfy $R_{i,j}$ is the following. Look for an $n > 2^j$ such that $f_j^\sigma(n) \downarrow \notin A^\sigma_{\neq i}$. Put n into A_i and restrain $f_j^\sigma(n)$ from entering $A_{\neq i}$. This may injure requirements with lower priority.

Lemma 1.16 *For all* i, j*, the requirement* $R_{i,j}$ *acts only finitely many times.*

Proof. This is because the requirements are ordered in order type omega, and between any two stages at which the $(n+1)$-st requirement acts, one of the first n requirements must have acted. It follows by induction on n that the n-th requirement only acts finitely many times. □

Lemma 1.17 *For all* $i, j \in \omega$*,* $A_i \neq f_j^{-1}[A_{\neq i}]$.

Proof. Assume the opposite, i.e., for some $i \in \omega$, $A_i \leq_{hm} A_{\neq i}$ via f_j. Choose a stage σ where requirement $R_{i,j}$ is considered and requirements of higher priority have ceased to act; also choose an $n > 2^j$ such that $f_j^\sigma(n) \downarrow$ and $f_j^\sigma(n) \notin A^\sigma_{\neq i}$. Such an n exists, as at most 2^k numbers less than 2^{k+1} are added to A_i for each k and therefore A_i is co-infinite. But then at stage σ a number was added to A_i to violate the reduction f_j, contradiction. □

The lemmas above prove the theorem. □

Corollary 1.18 *The countable atomless Boolean algebra may be embedded into the hm-degrees of Σ^1_1 subsets of ω.*

Note that there are, or course, much deeper statements about the structure of c.e. m-degrees (e.g., [**5, 18, 21**]) that one could try to lift to hm-degrees of Π^1_1 sets. However, Corollary 1.18 provides enough evidence that the structure of hm-degrees of Σ^1_1 sets is rich.

2 A complete Σ^1_1 equivalence relation

We start the section by establishing some general properties of Σ^1_1 equivalence relations.

Definition 2.1 An equivalence relation E is *complete* in a class $\mathcal{R}$ of equivalence relations (with a specified reducibility), if $E \in \mathcal{R}$ and every equivalence relation from $\mathcal{R}$ is reducible to E (with respect to the chosen reducibility).

Theorem 2.2

(1) *There exists a universal Σ^1_1 enumeration of all Σ^1_1 equivalence relations on ω.*

(2) *There exists a complete Σ^1_1 equivalence relation U (with respect to h- or FF-reducibility).*

Proof. Let $\{A_e\}_{e\in\omega}$ be the standard Σ^1_1 enumeration of all Σ^1_1 subsets of $\omega \times \omega$ (for instance, as in [**22**]). Define the equivalence relation R_e as the reflexive transitive closure of A_e, i.e.,

$$\begin{aligned} xR_e y \iff & x = y \vee (\exists z_0, \dots, z_k)[z_0 = x \& \dots \& z_k = y \& (\forall i < k)(\langle z_i, z_{i+1}\rangle) \in A_e] \\ & \vee (\exists z_0, \dots, z_k)[z_0 = y \& \dots \& z_k = x \& (\forall i < k)(\langle z_i, z_{i+1}\rangle) \in A_e]. \end{aligned}$$

Then every Σ^1_1 equivalence relation appears in this enumeration; moreover, from the properties of the enumeration $\{A_e\}_{e\in\omega}$, the enumeration $\{R_e\}_{e\in\omega}$ is universal.

Now define an equivalence relation R as follows:

$$\langle x, e\rangle R \langle y, e\rangle \iff xR_e y.$$

Then R is an h- and FF-complete Σ^1_1 equivalence relation. □

A useful and rather straightforward property of complete Σ^1_1 equivalence relations is the following:

Proposition 2.3 *An h-complete (or FF-complete) Σ^1_1 equivalence relation has infinitely many properly Σ^1_1 equivalence classes.*

Proof. Under h- or FF-reducibility properly Σ^1_1 equivalence classes are mapped to properly Σ^1_1 equivalence classes. In Theorem 6.1 below we show that there exist Σ^1_1 equivalence relations with infinitely many properly Σ^1_1 equivalence classes. Thus, a complete Σ^1_1 equivalence relation must also have this property. □

Recall the notion of hm-reducibility on subsets of ω introduced in Section 1.3. There exist Σ^1_1 equivalence relations with infinitely many hm-complete classes (e.g., as in Theorem 6.1 below). Therefore,

Corollary 2.4 *An h-complete (FF-complete) Σ^1_1 equivalence relation must have infinitely many properly Σ^1_1 equivalence classes that are hm-complete (m-complete, respectively).*

In a following section we will show that this condition is necessary but not sufficient for a relation to be h- or FF-complete among Σ^1_1 equivalence relations.

Remark 2.5 In [**7**] the authors showed that, in fact, the natural equivalence relation of bi-embeddability on the class of computable trees (here we mean the standard model-theoretic notion of embedding of structures) is FF-complete (thus, also h-complete) for the class of all Σ^1_1 equivalence relations on ω, where trees are considered in the signature with one unary function symbol interpreted as the predecessor function. Furthermore, [**9**] shows that the isomorphism relation on many natural classes of computable structures is FF-complete among Σ^1_1 equivalence relations.

By the above results, there exist h-degrees formed by Δ^1_1 equivalence relations with exactly n equivalence classes, for $n \leq \omega$, and a greatest h-degree of Σ^1_1 equivalence relations, namely, that of a complete Σ^1_1 equivalence relation. The next step is to show that the structure of h-degrees of properly Σ^1_1 equivalence relations is not trivial.

Proposition 2.6 *There exists a Σ^1_1 equivalence relation on ω which is neither Δ^1_1 nor h-complete.*

Proof. Let $(L_m)_{m\in\omega}$ be the numbering of all computable linear orderings on ω. Consider the following equivalence relation $E_{\omega_1^{\mathrm{CK}}}$:

$$mE_{\omega_1^{\mathrm{CK}}}n \iff \begin{array}{l}\text{either } L_m, L_n \text{ are not well-orders (i.e., } m,n \notin WO)\\ \text{or } L_m \cong L_n.\end{array}$$

The relation $E_{\omega_1^{\mathrm{CK}}}$ is Σ^1_1 but not Δ^1_1 as otherwise the equivalence class consisting of non-well-orderings would be a Δ^1_1 set, a contradiction. Moreover, for every computable ordinal α, the equivalence class of $E_{\omega_1^{\mathrm{CK}}}$ containing α is hyperarithmetical. The only properly Σ^1_1 equivalence class is the class consisting of the computable non well-orderings. As the complete relation R constructed above has infinitely many properly Σ^1_1 equivalence classes, it cannot be reduced to $E_{\omega_1^{\mathrm{CK}}}$. Thus $E_{\omega_1^{\mathrm{CK}}}$ is not complete. □

We would like to mention another natural example of an incomplete properly Σ^1_1 equivalence relation, namely, the relation of bi-embeddability on the class of linear orders studied in [**20**]. Recall the notion of Scott rank: it is a measure of model theoretic complexity of countable structures. For a computable structure, the Scott rank is at most $\omega_1^{\mathrm{CK}} + 1$ (see, for instance, [**3**] for a definition and an overview of results about the Scott rank of computable structures). In the class of computable linear orderings with the relation of bi-embeddability, the only equivalence class that contains structures of high (i.e., non-computable) Scott rank is the class of the dense linear order η. All other equivalence classes contain only structures of computable Scott rank (see [**20**] for details). If bi-embeddability on linear orderings were complete, it would necessarily have infinitely many equivalence classes with structures of high Scott rank. Therefore, bi-embeddability on linear orders cannot be complete.

3 Embedding Σ^1_1 sets into Σ^1_1 relations

For the reasons stated in Fact 1.10, we are interested in the structure of properly Σ^1_1 equivalence relations, i.e., relations that are Σ^1_1 but not Δ^1_1. In this section we prove the following theorem:

Theorem 3.1 *The structure of properly Σ^1_1 sets with the relation of m-reducibility is order-preservingly (and effectively) embedded into the structure of properly Σ^1_1 equivalence relations with the relation of FF-reducibility, i.e., one can assign to every properly Σ^1_1 set A a properly Σ^1_1 equivalence relation E_A such that, for all properly Σ^1_1 sets A, B,*

$$A \leq_m B \iff E_A \leq_{FF} E_B.$$

Before we give the proof of this theorem we will show the following:

Theorem 3.2 *The structure of properly Σ^1_1 sets with the relation of* 1*-reducibility is order-preservingly (and effectively) embedded into the structure of properly Σ^1_1 equivalence relations with the relation of FF-reducibility where the reducing function is* 1-1.

Proof. Let A be a properly Σ^1_1 set. Define the relation E_A in the following way:

$$\begin{aligned} xE_Ay \iff & x, y \in A \\ & \text{or } x = y. \end{aligned}$$

The relation E_A is properly Σ^1_1.

Lemma 3.3 *For all properly Σ^1_1 sets A, B,*

$$A \leq_1 B \iff E_A \leq_{FF} E_B,$$

where the FF-reducibility is witnessed by a computable 1-1 *function.*

Proof. The direction from right to left is obvious. To prove the direction from left to right suppose $A \leq_1 B$ via a computable 1-1 function f. Consider x, y such that xE_Ay. By definition of E_A, E_B and by properties of f,

$$xE_Ay \iff x, y \in A \text{ or } x = y \iff f(x), f(y) \in B \text{ or } f(x) = f(y) \iff f(x)E_Bf(y).$$

We use the fact that f is injective to prove the equivalence of the third and the second statement. □

The lemma proves the theorem. □

Remark 3.4 Relations of this kind for Σ^0_1 sets were considered in [**12**].

Proposition 3.5 *There exists an effective procedure which transforms a properly Σ^1_1 set A into a properly Σ^1_1 set A^* in such a way that*

$$A \leq_m B \Rightarrow A^* \leq_1 B^*;$$

$$A^* \leq_m B^* \Rightarrow A \leq_m B.$$

Proof. For every set A, define $A^* = A \times \omega = \{\langle x, i\rangle \mid x \in A, i \in \omega\}$. For every i, denote by A_i the set $\{\langle x, i\rangle \mid x \in A\}$. Then $A^* = \cup_i A_i$. Note that, by definition of A^*,

$$x \in A \iff \forall i \langle x, i\rangle \in A^* \iff \exists i \langle x, i\rangle \in A^*.$$

Suppose $A \leq_m B$ via a computable function f. We define a computable function h in the following way: for $x' = \langle x, i\rangle$ let $h(x') = \langle f(x), \langle x, i\rangle\rangle$, i.e., we send every $x' \in A_i$ to an element of $B_{x'}$. It guarantees that the function h is 1-1. Thus we only need to show that h witnesses the 1-reduction of A^* to B^*:

$$x' \in A^* \iff x \in A \iff f(x) \in B \iff \langle f(x), \langle x, i\rangle\rangle \in B^*.$$

Now suppose that $A^* \leq_m B^*$ via a computable function h. Define $f(x) = y \iff l(h(\langle x, 0\rangle)) = y$, i.e., $h(\langle x, 0\rangle) = \langle y, j\rangle$, for some $j \in \omega$. Then the function f m-reduces A to B:

$$x \in A \iff \langle x, 0\rangle \in A^* \iff h(\langle x, 0\rangle) = \langle y, j\rangle \in B^* \iff y \in B. \qquad \square$$

Proof of Theorem 3.1. The proof now follows directly from Proposition 3.5 and Theorem 3.2. □

Corollary 3.6 *For any $1 \leq n \leq \omega$, there exists an effective embedding of the structure of properly Σ^1_1 sets under m-reducibility into the structure of properly Σ^1_1 relations with exactly n properly Σ^1_1 equivalence classes under FF-reducibility.*

In Section 1.3 we introduced the notion of hm-reducibility on sets, which is a hyperarithmetical analogue of m-reducibility. We showed that the structure of hm-degrees of Σ^1_1 sets is complicated. Consider now a hyperarithmetical version of the 1-reducibility of subsets of ω:

Definition 3.7 Let A, B be subsets of ω. We say that A is *hyperarithmetically* 1*-reducible* to B, denoted by $A \leq_{h1} B$, iff there exists a hyperarithmetical 1-1 function f such that, for every $n \in \omega$,

$$n \in A \iff f(n) \in B.$$

Using this definition and ideas from above one can show the following:

Theorem 3.8 *The structure of properly Σ^1_1 sets with the relation of hm-reducibility is order-preservingly (and effectively) embedded into the structure of properly Σ^1_1 equivalence relations with the relation of h-reducibility, i.e., one can assign to every properly Σ^1_1 set A a properly Σ^1_1 equivalence relation E_A such that, for all properly Σ^1_1 sets A, B,*

$$A \leq_{hm} B \iff E_A \leq_h E_B.$$

Moreover, for every $n \leq \omega$, there is such an embedding into the structure of properly Σ^1_1 equivalence relations with exactly n properly Σ^1_1 equivalence classes.

Thus, the structure of h-degrees of Σ^1_1 equivalence relations even with just one properly Σ^1_1 equivalence class is at least as rich as the structure of Σ^1_1 sets under hm-reducibility.

4 Properly Σ^1_1 equivalence relations with only hyperarithmetical equivalence classes

In this section we show that a properly Σ^1_1 equivalence relation need not contain properly Σ^1_1 equivalence classes. Moreover, the example we present contains only equivalence classes of size 1 or 2.

Let A be a Σ^1_1 subset of ω which is not Δ^1_1. Define the corresponding equivalence relation F_A on $\omega \times 2$ in the following way:

$$\begin{aligned}(m_0, n_0) F_A (m_1, n_1) \iff\ & m_0 = m_1 \in A \\ & \text{or } (m_0, n_0) = (m_1, n_1).\end{aligned}$$

The relation F_A is Σ^1_1. The equivalence classes of F_A are of the form $\{(m, n) \mid 1 \leq n \leq 2\}$ if $m \in A$, and $\{(m, n)\}$ if $m \notin A$. In particular, every equivalence class has size 1 or 2. Again, similar relations constructed from Σ^0_1 sets were considered in [**12**].

Claim 4.1 The equivalence relation F_A is properly Σ^1_1.

Proof. If F_A were Δ^1_1, so would be the set A, as $A = \{m \mid (m,0)F_A(m,1)\}$, which is a contradiction. □

One can easily modify the example to get an equivalence relation with classes of size at most (and including) k, for $2 \leq k < \omega$.

Definition 4.2 Following [**12**], we call an equivalence relation *k-bounded* if all its equivalence classes have size at most k.

Theorem 4.3 *There exists a properly Σ^1_1 equivalence relation S^{k+1} with all its equivalence classes containing at most $k+1$ elements such that for no Σ^1_1 equivalence relation R with its equivalence classes containing at most k elements do we have $R^{k+1} \leq_h S$ (hence, for no such R do we have $R^{k+1} \leq_{FF} S$).*

Proof. As shown in [**12**], the analogous result is true for the case of c.e. relations. Simple transformation of this argument proves the theorem for Σ^1_1 equivalence relations. □

5 Equivalence relations with finitely many properly Σ^1_1 classes

One can modify the example from the proof of Proposition 2.6 to get, for every finite $k \geq 2$, a Σ^1_1 equivalence relation which has exactly k properly Σ^1_1 equivalence classes:

Proposition 5.1 *For every finite $k \geq 1$ there exists a Σ^1_1 equivalence relation on ω with infinitely many equivalence classes, such that exactly k of them are properly Σ^1_1.*

Proof. Let $A_1, \ldots, A_k$ be disjoint properly Σ^1_1 sets. Consider the relation $E_{A_1,\ldots,A_k}$:

$$xE_{A_1,\ldots,A_k}y \iff x = y \vee x,y \in A_1 \vee \ldots \vee x,y \in A_k.$$

Then $E_{A_1,\ldots,A_k}$ has the desired properties. □

We give another example of equivalence relations with exactly k properly Σ^1_1 classes, for $k \geq 1$. The reason is that in the next section we will use a generalization of this example.

Again consider ω as a set of codes for linear orders. We will define relations F_k, for $k \geq 1$, on pairs of linear orders. First of all, we define additional hyperarithmetical equivalence relations E_k (here we identify natural numbers k, k' with ordinals):

$$\begin{aligned} n_1 E_k n_2 \iff & \text{ either } L_{n_1} \cong L_{n_2} \cong k' < k-1 \\ & \text{ or both } n_1, n_2 \text{ are not codes for well-orders of type } k' < k-1. \end{aligned}$$

By definition, E_k is hyperarithmetical and has exactly k equivalence classes. We now define F_k as follows: for $(m_i, n_i) \in \omega^2, i = 1,2$,

$$\begin{aligned} (m_1,n_1)F_k(m_2,n_2) \iff & \text{ either } (L_{m_1}, L_{m_2} \text{ are not well-orders and } n_1 E_k n_2) \\ & \text{ or } (L_{m_1} \cong L_{m_2}). \end{aligned}$$

The idea is that we "cut" the properly Σ^1_1 class of $E_{\omega_1^{\mathrm{CK}}}$ (the relation defined in Proposition 2.6) into k properly Σ^1_1 pieces. The relations F_k, $k \geq 1$, have the necessary properties. Moreover,

Proposition 5.2 *For all $1 \leq k_1 < k_2 < \omega$, $F_{k_1} <_h F_{k_2}$.*

Proof. Let f be a hyperarithmetical function which witnesses $E_{k_1} <_h E_{k_2}$. Consider the function $g(m,n) = (m, f(n))$. It is hyperarithmetical and reduces F_{k_1} to F_{k_2}. The reduction is strict, as F_{k_1} has fewer properly Σ^1_1 equivalence classes than F_{k_2}. □

Remark 5.3 No F_k, for $k \geq 1$, is complete as no Σ^1_1 equivalence relation with only finitely many properly Σ^1_1 equivalence classes can be complete for the class of Σ^1_1 equivalence relations.

6 Equivalence relations with infinitely many properly Σ^1_1 classes

In this section we show that an infinite number of properly Σ^1_1 equivalence classes does not guarantee the h- or FF-completeness of a Σ^1_1 equivalence relation.

Indeed, it is easy to construct a non-complete Σ^1_1 equivalence relation with infinitely many properly Σ^1_1 equivalence classes. Take a computable sequence $(A_n)_{n\in\omega}$ of disjoint Σ^1_1 sets such that none of them is complete, and consider the relation R_∞ defined as follows:

$$xR_\infty y \iff x = y \vee \exists n(x, y \in A_n).$$

As the sequence $(A_i)_{i\in\omega}$ is computable, the relation R_∞ is Σ^1_1. Moreover, it is not complete as, for example, the relation R_B for a complete Σ^1_1 set B constructed as in Section 3 is not reducible to R_∞.

By Corollary 2.4, an h-complete (a FF-complete) Σ^1_1 equivalence relation must have infinitely many equivalence classes that are hm-complete (m-complete) as Σ^1_1 sets. Below we will show that this condition is not sufficient:

Theorem 6.1 *There exists a non-h-complete (non-FF-complete) Σ^1_1 equivalence relation with infinitely many classes that are hm-complete (m-complete) among Σ^1_1 sets.*

The proof of the theorem will follow from Proposition 6.3 below.

For every computable infinite ordinal α, we define equivalence relations E_α and F_α in the following way:

$$\begin{aligned} n_1 E_\alpha n_2 \iff & \text{ either } L_{n_1} \cong L_{n_2} \cong \alpha' < \alpha \\ & \text{ or [neither } n_1 \text{ nor } n_2 \text{ code well-orders of type } < \alpha]. \end{aligned}$$

In other words, for each $\alpha' < \alpha$, there is an equivalence class consisting of linear orders isomorphic to α'. All the linear orders that are not isomorphic to any $\alpha' < \alpha$ form a single equivalence class. By definition, if α is computable, then E_α is hyperarithmetical with infinitely many equivalence classes, provided α is infinite. Indeed, for a fixed $\alpha < \omega_1^{\mathrm{CK}}$ it is hyperarithmetical to check whether or not some $n \in \omega$ is a code for a well-order of type α or of type $< \alpha$. Then both the first and the second line of the definition give hyperarithmetical conditions. Hence, for infinite $\alpha < \omega_1^{\mathrm{CK}}$, all E_α are hyperarithmetical and h-equivalent to each other. Notice that there is some non-uniformity in the definition of E_α for finite (defined in the previous section) and infinite α.

Now define:

$$\begin{aligned} (m_1, n_1) F_\alpha (m_2, n_2) \iff & \text{ either } L_{m_1}, L_{m_2} \text{ are not well-orders and } n_1 E_\alpha n_2 \\ & \text{ or } L_{m_1} \cong L_{m_2}. \end{aligned}$$

Proposition 6.2 *For all computable infinite α_1, α_2, $F_{\alpha_1} \equiv_h F_{\alpha_2}$.*

Proof. Consider the function h that witnesses the h-equivalence of the corresponding $E_{\alpha_1}, E_{\alpha_2}$. The function h' which sends a pair (m,n) into the pair $(m,h(n))$ gives the equivalence of $F_{\alpha_1}, F_{\alpha_2}$. □

Recall the definition of the relation $E_{\omega_1^{\mathrm{CK}}}$ from Section 2:

$$\begin{aligned} mE_{\omega_1^{\mathrm{CK}}}n \iff &\text{ either } L_m, L_n \text{ are not well-orders (i.e., } m,n \notin WO) \\ &\text{ or } L_m \cong L_n. \end{aligned}$$

Finally, we define an equivalence relation $F_{\omega_1^{\mathrm{CK}}}$ as follows:

$$\begin{aligned} (m_1,n_1)F_{\omega_1^{\mathrm{CK}}}(m_2,n_2) \iff &\text{ either } L_{m_1}, L_{m_2} \text{ are not well-orders and } n_1E_{\omega_1^{\mathrm{CK}}}n_2 \\ &\text{ or } L_{m_1} \cong L_{m_2}. \end{aligned}$$

Note that all F_α, $\alpha < \omega_1^{\mathrm{CK}}$, and $F_{\omega_1^{\mathrm{CK}}}$ have infinitely many equivalence classes that are m-complete (thus, also hm-complete) among Σ^1_1 sets.

Proposition 6.3 *For every computable α,*

$$F_\alpha <_h F_{\omega_1^{\mathrm{CK}}}.$$

Proof. Obviously, $F_\alpha \leq_h F_{\omega_1^{\mathrm{CK}}}$: let f reduce E_α to $E_{\omega_1^{\mathrm{CK}}}$; then $g(m,n) = (m,f(n))$ reduces F_α to $F_{\omega_1^{\mathrm{CK}}}$. We only need to prove that $F_{\omega_1^{\mathrm{CK}}}$ is not reducible to F_α for any computable α. Suppose that for some computable α there were such a hyperarithmetical reduction h:

$$(m_1,n_1)F_{\omega_1^{\mathrm{CK}}}(m_2,n_2) \iff h((m_1,n_1))F_\alpha h((m_2,n_2)).$$

Consider $n_1, n_2 \in \omega$. For every $m \notin WO$ we have:

$$n_1E_{\omega_1^{\mathrm{CK}}}n_2 \iff (m,n_1)F_{\omega_1^{\mathrm{CK}}}(m,n_2) \iff h((m,n_1))F_\alpha h((m,n_2)) \iff$$

$$\iff L_{m_1} \cong L_{m_2} \cong \gamma, \text{ where } \gamma \text{ is an ordinal, or } [m_1, m_2 \notin WO \text{ and } l_1E_\alpha l_2],$$

where $h(m,n_i) = (m_i, l_i), i = 1,2$. Fix this notation for the rest of the proof.

If there exists an $m \notin WO$ such that for all n_1, n_2 the corresponding $m_1, m_2 \notin WO$, then the proposition is proved. Indeed, fix such an m. Then, for all $n_1, n_2 \in \omega$,

$$n_1E_{\omega_1^{\mathrm{CK}}}n_2 \iff (m,n_1)F_{\omega_1^{\mathrm{CK}}}(m,n_2) \iff (m_1,l_1)F_\alpha(m_2,l_2) \iff l_1E_\alpha l_2,$$

which gives a hyperarithmetical reduction of $E_{\omega_1^{\mathrm{CK}}}$ to E_α, a contradiction.

Suppose now that for every $m \notin WO$ there exist $n_1, n_2 \in \omega$ such that $L_{m_1} \cong L_{m_2} \cong \gamma$ for some $\gamma < \omega_1^{\mathrm{CK}}$. Define a Π^1_1 relation $R(m,n)$ as follows:

$$\begin{aligned} R(m,n) \iff &(m \in WO \wedge m = n) \\ &\text{or } (n \in WO \wedge L_n \cong L_{m_1} \cong L_{m_2} \text{ associated to some } h(m,n_1), h(m,n_2)). \end{aligned}$$

By Uniformization, R can be uniformized by a Π^1_1 function f. The function f is total, thus hyperarithmetical from ω to WO. By Bounding, the range of f is bounded by a computable ordinal γ_0.

Consider now all $m \in WO$ for which there exist n_1, n_2 such that $L_{m_1} \cong L_{m_2} \cong \gamma_0$. Then there is a computable bound α_0 on ordinals coded by such elements m.

Now we have

$$WO = \{m \mid m \neq \mathit{code}(\beta) \text{ for } \beta \leq \alpha_0 \text{ and } \exists n_1, n_2 L_{m_1} \cong L_{m_2} \cong \gamma \leq \gamma_0\},$$

which gives a hyperarithmetical definition of WO, a contradiction. □

Remark. The process above of constructing Σ^1_1 equivalence relations may be iterated further. In particular, the relation $F_{\omega_1^{\mathrm{CK}}}$ is not complete among Σ^1_1 equivalence relations.

7 More results

The following result from [**12**] shows the difference between the theory of Σ^0_1 equivalence relations and that of Σ^1_1 equivalence relations:

Theorem 7.1 *Let $A_1, \ldots, A_n$ be disjoint c.e. sets the complement of whose union is infinite. Then*

$$\mathrm{id}_\omega \leq_{FF} R_{A_1,\ldots,A_n} \iff A_1 \cup \ldots \cup A_n \textit{ is not simple.}$$

Here

$$xR_{A_1,\ldots,A_n}y \iff x = y \vee \exists i \leq n(x, y \in A_i).$$

In the case of h-reducibility and disjoint Σ^1_1 sets $A_1, \ldots, A_n$,

$$\mathrm{id}_\omega \leq_h R_{A_1,\ldots,A_n}$$

always holds. Indeed, the complement C of $\bigcup_{i\leq n} A_i$ is a Π^1_1 set, thus it contains a hyperarithmetical subset B. Then a 1-1 hyperarithmetical function from ω onto B witnesses the reduction.

The analogy with c.e. equivalence relations might be more complete if we considered Π^1_1 equivalence relations.

Using ideas from [**12**] one can show the following:

Theorem 7.2 *There exist properly Σ^1_1 equivalence relations that are m-complete (hm-complete) as Σ^1_1 sets but FF-incomparable (respectively, h-incomparable) as Σ^1_1 equivalence relations.*

Proof. Let A be an m-complete, hence, also hm-complete Σ^1_1 set. Let E_A be a Σ^1_1 equivalence relation built from A as in Section 3. Let F_A be a Σ^1_1 equivalence relation with all its equivalence classes finite built from A as in Section 4. Then E_A and F_A are neither FF-comparable nor h-comparable.

Suppose E_A is reducible to F_A via a computable (or hyperarithmetical) function f. Fix an arbitrary $x_0 \in A$ and let $y_0 = f(x_0)$. Then $A = \{x \mid f(x)F_A y_0\}$; therefore $A \leq [y_0]_{F_A}$, where $[y_0]_{F_A}$ is finite. Thus A is computable (hyperarithmetical), a contradiction.

Suppose now that F_A is reducible to E_A via g. Consider the set $B = \{g(x) \mid x \in \omega\}$. Then $B \cap A \neq \emptyset$, otherwise F_A would be reducible to id_ω, thus hyperarithmetical. Now let $C = \{x \mid g(x) \in A\}$; then C is an equivalence class of F_A. Pick an arbitrary $y \in A$ and define $h(x)$ in the following way:

$$h(x) = \begin{cases} y & \text{if } x \in C, \\ g(x) & \text{otherwise.} \end{cases}$$

All equivalence classes of F_A are finite, thus h is a computable (hyperarithmetical) function which reduces F_A to the equality on ω. □

8 Questions

If an equivalence relation E is reducible to an equivalence relation E' (under any of the two reducibilities considered here) then E is reducible to E' as sets (under the corresponding reducibility). On the other hand, if a Σ^1_1 equivalence relation is m-complete

(hm-complete) as a Σ^1_1 set, it does not guarantee that it is FF-complete (h-complete) as a Σ^1_1 equivalence relation. Indeed, let A be an m-(hm-)complete Σ^1_1 set. Let E_A be a Σ^1_1 equivalence relation built as in Sections 3 or 4. Then E_A is not complete among Σ^1_1 relations but it is obviously complete as a Σ^1_1 set. One can also build such equivalence relations with any number of properly Σ^1_1 equivalence classes.

As it follows from Theorem 7.2, two Σ^1_1 equivalence relations may be incomparable while both being m-complete among Σ^1_1 sets. However in the above example one of the relations had only finite classes while the other relation had an infinite class and all the other classes of size 1. Thus the following set of questions arises naturally:

Question 8.1 Let E, E' be Σ^1_1 equivalence relations with only finite (or hyperarithmetical) equivalence classes. Suppose E, E' are both complete as sets (under m- or hm-reducibility). As follows from Theorem 4.3, it may be the case that $E < E'$. Is it possible that E and E' are incomparable?

Question 8.2 The same for relations with a fixed number of properly Σ^1_1 (Σ^0_1) equivalence classes.

We studied properly Σ^1_1 equivalence relations according to the number of their properly Σ^1_1 equivalence classes. We saw examples of equivalence relations with only hyperarithmetical classes, with exactly n properly Σ^1_1 equivalence classes, for $n \in \omega$ and with infinitely many properly Σ^1_1 equivalence classes.

Question 8.3 Does there exist a properly Σ^1_1 equivalence relation on (a hyperarithmetical subset of) ω with infinitely many equivalence classes such that all its classes are properly Σ^1_1?

References

[1] C. Bernardi, A. Sorbi, *Classifying positive equivalence relations*, J. Symbolic Logic 48 (1983), 529–538.

[2] W. Calvert, D. Cenzer, V. Harizanov, A. Morozov, *Effective categoricity of equivalence structures*, Ann. Pure Appl. Logic 141 (2006), 61–78.

[3] W. Calvert, E. Fokina, S. Goncharov, J. Knight, O. Kudinov, A. Morozov, V. Puzarenko, *Index sets for classes of high rank structures*, J. Symbolic Logic 72 (2007), 4, 1418–1432.

[4] D. Cenzer, V. Harizanov, J. Remmel, Σ^0_1 *and* Π^0_1 *equivalence structures*, Proceedings of "Computability in Europe 2009", Heidelberg, Lecture Notes in Computer Science 5635, 99–108, 2009.

[5] S. D. Denisov, *Structure of the upper semilattice of recursively enumerable m-degrees and related questions*, Algebra and Logic 17 (1978), 6, 643–683.

[6] Yu. L. Ershov, Definability and computability, Siberian School of Algebra and Logic, Consultants Bureau, New York, 1996.

[7] E. Fokina, S. Friedman, *Equivalence relations on classes of computable structures*, Proceedings of "Computability in Europe 2009", Heidelberg, Lecture Notes in Computer Science 5635, 198–207, 2009.

[8] E. Fokina, S. Friedman, A. Törnquist, *The effective theory of Borel equivalence relations*, Ann. Pure Appl. Logic 161 (2010), 837–850.

[9] E. Fokina, S. Friedman, V. Harizanov, J. Knight, C. McCoy, A. Montalbán, *Isomorphism and bi-embeddability relations on computable structures*, to appear in J. Symbolic Logic.

[10] H. Friedman, L. Stanley, *A Borel reducibility theory for classes of countable structures*, J. Symbolic Logic 54 (1989), 894–914.

[11] S. Gao, Invariant Descriptive Set Theory, Pure and Applied Mathematics, CRC Press/Chapman & Hall, 2009.

[12] S. Gao, P. Gerdes, *Computably enumerable equivalence relations*, Studia Logica 67 (2001), 27–59.

[13] S. S. Goncharov, J. F. Knight, *Computable structure and non-structure theorems*, Algebra and Logic 41 (2002), 351–373 (English translation).

[14] V. Kanovei, Borel Equivalence Relations: Structure and Classification, University Lecture Series, 44, American Mathematical Society, 2008.

[15] A. Kechris, *New directions in descriptive set theory*, Bull. Symbolic Logic 5 (1999), 2, 161–174.

[16] B. Khoussainov, F. Stephan, Y. Yang, *Computable categoricity and the Ershov hierarchy*, Ann. Pure Appl. Logic 156 (2008), 86–95.

[17] A. Lachlan, *A note on positive equivalence relations*, Zeit. Mat. Logik Grundl. Math. 33 (1987), 43–46.

[18] A. Lachlan, *Recursively enumerable many-one degrees*, Algebra and Logic 11 (1972), 3, 326–358.

[19] A. Louveau, B. Velickovic, *A note on Borel equivalence relations*, Proc. Amer. Math. Soc. 120 (1994), 1, 255–259.

[20] A. Montalbán, *On the equimorphism types of linear orderings*, Bull. Symbolic Logic 13 (2007), 71–99.

[21] S. Yu. Podzorov, *The universal Lachlan semilattice without the greatest element*, Algebra and Logic 46 (2007), 3, 163–187.

[22] H. Rogers, Theory of Recursive Functions and Effective Computability, McGraw-Hill, 1967.

[23] G. Sacks, Higher Recursion Theory, Springer-Verlag, 1989.

[24] R. Soare, Recursively enumerable sets and degrees: A study of computable functions and computably generated sets, Perspectives in Mathematical Logic, Springer-Verlag, Berlin, 1987.

Transfinite machines, analysis and determinacy

Philip D. Welch[*]

[*] School of Mathematics, University of Bristol, UK
p.welch@bristol.ac.uk

Abstract. We survey recent connections between inductive operators, discrete transfinite machine models of computation, and determinacy. These are all at low levels of the arithmetic hierarchy, and concern results provable within second-order number theory.

Introduction

Transfinite machine models are interesting in their own right, and seeing how they may computes integers or reals can be an intricate question of programme design. However, dealing as they do essentially with the infinite, deeper insights, we claim, are obtained when relating their action to some well-known set-theoretic hierarchy, such as L, the Gödel universe of constructible sets; or else to a much studied concept from the early 70's, that of (monotone) inductive definitions. Infinite games enter in to this discussion, as the inductive operators then considered can be given a generalised quantifier expression in often in terms of infinite sequences of ordinary alternating $\forall\ \exists$ quantifiers which represent game moves.

We survey only some of these connections, concentrating for the most part on the Infinite Time Turing Machines (ITTM's) of Hamkins and Kidder ([**4**]) in the first section. Then we give a very brief survey of a few results relating games and operators. Our aim here is to concentrate on Σ^0_3 —the last pointclass left which seems to have been not well studied, or perhaps understood, in terms of its relation to L and describing the locations of strategies for games with payoff sets in this pointclass. The case of Σ^0_3 has now been given a classification by a characterisation of a level of the L_β hierarchy (Theorem 2.8 below), but somewhat unsatisfactorily in terms of non-wellfounded end extensions with a certain technical property. So there is some more work to be done here to get theorems like those of Solovay on Σ^0_2 below.

1 Infinite Time Turing Machines

We sketch a model which is minor variant of that proposed by Hamkins and Kidder in [**4**]. Allow a standard Turing machine to run through transfinite stages using one of the usual Turing programs $\langle P_e \mid e \in \mathbb{N}\rangle$. We simply have to specify some *limit rules* to declare how the machine behaves at limit stages λ of time. We take an alphabet of three symbols: $\{0, 1, \mathrm{B(lank)}\}$.

- Enumerate the cells of the tape $\langle C_k \mid k \in \mathbb{N}\rangle$ with contents at time τ as $C_k(\tau)$.
- Let the current instruction about to be performed at time τ be $I_{p(\tau)}$; and let the current cell being inspected be $C_{i(\tau)}$.

[*]The author thanks the John Templeton Foundation for its support through Project #13152, The Myriad Aspects of Infinity.

- Behaviour at successor stages $\alpha \to \alpha+1$: use the Turing program procedures just as normal.
- At limit times λ:
  (a) we specify cell values by
  $$C_i(\lambda) = \begin{cases} k & \text{iff } \exists\alpha < \lambda \forall\beta < \lambda(\alpha < \beta \to C_i(\beta) = k) \text{ for } k \in \{0,1,\mathrm{B}\} \\ B & \text{otherwise;} \end{cases}$$
  (b) we put the R/W to cell $C_{i(\lambda)}$, where
  $$i(\lambda) = \mathrm{Liminf}^*_{\alpha\mapsto\lambda}\{i(\beta) \mid \alpha < \beta < \lambda\};$$
  (c) we set $p(\lambda) = \mathrm{Liminf}_{\alpha\mapsto\lambda}\{p(\beta) \mid \alpha < \beta < \lambda\}$.

In the above, we define Liminf* to be the usual Liminf if the latter value is finite, and set it to be 0 if it is has become infinite.

This together with the value of the next instruction number p has the virtue of putting the R/W at the beginning of the outermost loop of any the outermost subroutine $I(\alpha)$ called unboundedly often in λ.

Hamkins and Lewis proved there is a *universal machine*, an S^m_n*-Theorem*, and a *Recursion Theorem* for ITTM's, and a wealth of results on the resulting ITTM-*degree theory* to which we refer the reader. Halting sets may be defined in the usual way:

$$H = \{(e,x) \mid e \in \mathbb{N}, x \in 2^{\mathbb{N}} \wedge P_e(x)\downarrow\};$$
$$H_0 = \{(e,0) \mid e \in \mathbb{N} \wedge P_e(0)\downarrow\}.$$

Natural occurring questions then are the following:

- What is H or H_0?
- How long do we have to wait to discover if $e \in H_0$ or not?
- What are the ITTM (semi)-decidable sets of integers? Or reals?

These questions can all be answered. However first we would like some type of an *"ITTM Normal Form Theorem"*:

Theorem 1.1 ([**9**, Corollary 36]) *There is a* universal predicate $\mathfrak{T}$ *which satisfies* $\forall e \forall x$*:*

$$P_e(x)\downarrow z \quad \leftrightarrow \quad \exists y \in 2^{\mathbb{N}}[\mathfrak{T}(e,x,y) \wedge \mathit{Last}(y) = z].$$

However for this to occur we need to know whether the ordinal length of any computation is capable of being output or written by a(nother) computation. (The *"Clockables = Writables" Problem* of [**4**].)

Theorem 1.2 (The λ,ζ,Σ-Theorem [**9**, Corollary 32]) *Let* Σ *be the least ordinal so that there exists* $\zeta < \Sigma$ *with the property that*

$$L_\zeta \prec_{\Sigma_2} L_\Sigma.$$

(i) *Then the universal ITTM machine first enters a loop at time* ζ*.*
Let λ *be the least ordinal satisfying*

$$L_\lambda \prec_{\Sigma_1} L_\zeta.$$

(ii) *Then*

$$\begin{aligned}\lambda &= \sup\{\alpha \mid \exists e\ P_e(0)\downarrow \textit{ in } \alpha \textit{ steps}\} \\ &= \sup\{\alpha \mid \exists e\ P_e(0)\downarrow y \in \mathit{WO} \wedge ||y|| = \alpha\}.\end{aligned}$$

The Normal Form Theorem follows as a corollary as does:

Corollary 1.3

(i) ([**9**, Theorem 41]) *The complete ITTM semi-decidable subset of* ω *is recursively isomorphic to the* $T^1_\lambda =_{df} \Sigma_1\text{-}Th(L_\lambda)$.

(ii) (Corollary 34 op. cit.) *The ITTM-decidable reals are precisely those of* L_λ.

2 Games and inductive definitions

The theory of inductive definitions and Gale–Stewart infinite games of perfect information over $\mathbb{N}$ is well studied. (We refer the reader to Moschovakis [**7**] for all notions of this section.)

We recall the definition of the *game* quantifier $\Game$:

Definition 2.1 A set $A \subseteq \mathbb{N}$ is $\Game\Gamma$ if there is $B \in \mathbb{N} \times \mathbb{R}$ so that

$$n \in A \iff \text{Player } I \text{ has a winning strategy in } G_{A_n}$$

where $A_n = \{x \in \mathbb{R} \mid B(n,x)\}$.

It is a result of Spector that monotone Π^1_1 inductive definitions do not lead out of the pointclass Π^1_1. We write (mon.-Π^1_1)-IND for the class of sets (1-1) reducible to fixed points of such monotone inductions starting with $\emptyset$. Part of Spector's analysis was that such inductions result in a fixed point at an ordinal at most the least non-recursive ordinal ω_1^{ck}: it is thus the 'closure ordinal' for monotone Π^1_1-inductive definitions.

Theorem 2.2 (Folklore) *A set* $A \subseteq \mathbb{N}$ *is* (mon.-Π^1_1)-IND *iff it is* $\Game\Sigma^0_1$ *iff it is* $\Sigma_1(L_{\omega_1^{ck}})$.

A closely related theorem:

Theorem 2.3 *For any* Σ^0_1 *game, if it is a win for Player* I*, then she has a* (mon.-Π^1_1)-IND *winning strategy. (And thus it is also* $\Sigma_1(L_{\omega_1^{ck}})$.*)*

We now consider over $\mathbb{N}$ the result of inductive definitions in the dual class: Σ^1_1-IND.

Theorem 2.4 (Solovay) *A set* $A \subseteq \mathbb{N}$ *is* Σ^1_1-IND *iff it is* $\Game\Sigma^0_2$ *iff it is* $\Sigma_1(L_{\sigma^1_1})$*, where* σ^1_1 *is the closure ordinal of* Σ^1_1*-inductive definitions.*

The corresponding closely related theorem to this is:

Theorem 2.5 (Solovay) *Any* Σ^0_2 *game, if it is a win for Player I then she has a* Σ^1_1-IND *winning strategy. (And thus also* $\Sigma_1(L_{\sigma^1_1})$.*)*

The reader by now naturally expects an answer to the question at the level of Σ^0_3? Recall that the results of Harvey Friedman on the non-provability of levels of determinacy in the Borel hierarchy in the system of analysis, Z_2, start at Σ^0_4-Det (Martin, refining Friedman [**2**]).

- $Z_2 \nvdash \Sigma^0_4$-Det.

Recently Montalbán and Shore have shown:

- $Z_2 \nvdash$ Bool(Σ^0_3)-Det (Montalbán–Shore, [**6**]).

However Det(Σ^0_3) has been known provable in analysis since Morton Davis's proof of this fact ([**1**]). Perhaps there is some link between such and the 'quasi-inductive' definitions that ITTM's constitute?

Question: Are strategies for Σ^0_3 sets ITTM-semi-decidable? Thus, are they $\Sigma_1(L_\Sigma)$?

Definition 2.6 Let "ITTM" abbreviate "$\forall X(H^X_0$ exists)"; in other words, "the complete ITTM-semi-decidable-in-X set exists".

Theorem 2.7 ([**11**]) *The theories*

$$\mathsf{\Pi}^1_3\text{-}\mathsf{CA}_0,\ \mathsf{\Delta}^1_3\text{-}\mathsf{CA}_0 + \Sigma^0_3\text{-}\mathsf{Det},\ \mathsf{\Delta}^1_3\text{-}\mathsf{CA}_0 + \mathsf{ITTM},\ \mathsf{\Delta}^1_3\text{-}\mathsf{CA}_0$$

are in strictly descending order of strength.

But where are the strategies? For sufficiently large β the following can happen: there is a non-wellfounded end extension $\mathcal{M}$ of L_β whose wellfounded part is precisely L_β but which contains an "infinite nesting" of Σ_2 elementary substructures. More precisely, there could be a sequence of $\mathcal{M}$-ordinals

$$\zeta(i) \leq \zeta(i+1) < \cdots < s(i+1) < s(i) \text{ for } i \in \omega$$

with both $(\zeta(i))_i$ and $(s(i))_i$ converging on β, and $L_{\zeta(i)} \prec_{\Sigma_2} L_{s(i)}$.

Theorem 2.8 *Let β be least so that L_β is the WFP($\mathcal{M}$) for some illfounded end-extension of L_β with this "infinite Σ_2 elementary extension nesting" configuration as above. Then β is also the least such that any Σ^0_3 game has a winning strategy definable over L_β.*

3 Hypermachines

Can we find 'machines' that will lift the Σ_2 "Liminf" of [HL] to a Σ_n-rule at limit stages? Yes, we can:

Theorem 3.1 ([**3**]) *For any $n \geq 2$ there is such a Σ_n limit rule, so that for a machine running under such a rule, the universal Σ_n-machine on integer input first enters a loop at the least $\zeta(n)$ such that $\exists\Sigma(n) > \zeta(n)$ with*

$$L_{\zeta(n)} \prec_{\Sigma_n} L_{\Sigma(n)}.$$

Such machines then compute, taken as a whole with n varying, all the reals of the least β-model of analysis $2^\omega \cap L_{\beta_0}$. It has to be said that with increasing n the machine intuition becomes more distanced, and considerations concerning stability of ordinals in L play an essential role in order for the universal Σ_n-hypermachine to satisfy the theorem.

Nevertheless, e.g. using Montalbán–Shore ([**6**]), strategies for n-Σ^0_3 games are computable by the $\Sigma(n+2)$-machines.

4 Open questions

We finish with some questions which we think merit further study.

Question 4.1 Give another description of the least β over which strategies for Σ^0_3 sets are definable.

Conjecture 4.2 Lubarsky's "Feedback-ITTM's" ([**5**]) are related to this.

Lubarsky has generalised the Hamkins–Kidder ITTM to one that allows calls to other ITTM's as sub-routines for enhanced input to the main program. In order not to have a nested, and so non-wellfounded, computation tree of machines calling each other he associates ordinals with each call, and insists they are descending in the usual ordering with each sub-call made. The computation times of such machines are then related to Mahlo limits of the Σ_2-extendible ordinal operation so-to-speak. He points out that one can consider lifting the restriction on ordinal labels and consider ill-founded computation trees. These may return no output of course and so are deemed to be non-terminating. Allowing this model somewhat more curious patters emerge: the "feedback-writable" reals turn out to be those that occur on any tape of any of these generalised computations, whether halting or not. This is something that does not occur on the usual ITTM models or the hypermachines. However it seems to us with the analysis of Theorem 2.8 above that the parallel between nested (and so ill-founded) calls of machines stuck inside a nest of Σ_2-extendible loops is strongly suggestive of a connection with the ordinal β of that theorem.

Another conjecture concerning this β relates it to the more standard theory of monotone inductive operators.

Conjecture 4.3 Does β equal the closure ordinal of mon.-$\Game\Sigma^0_3$-inductive operators?

Question 4.4 Develop an ordinal-theoretic analysis of the theory ITTM.

We note that if the Rathjen approach **[8]** to the ordinal analysis of Π^1_2-CA is taken to apply to that for Π^1_3-CA then the analysis he makes use of for chains of models $L_{\gamma_n} \prec_{\Sigma_1} L_{\gamma_{n+1}}$ for arbitrary $n < \omega$ must be lifted by one, to an analysis of chains of models of the form $L_{\zeta_n} \prec_{\Sigma_2} L_{\zeta_{n+1}}$. The ITTM's require a chain of just one link: $L_\zeta \prec_{\Sigma_2} L_\Sigma$, and so an ordinal analysis of the theory stating that the universe is closed under such chains is but a first, but necessary, step along this path.

Question 4.5 Find the β_n where strategies for n-Σ^0_3 games can be located.

Define semi-decidable sets of reals using the ITTM's (and Σ_n-hypermachines) in a standard way; this yields pointclasses $\boldsymbol{\Gamma_n}$ strictly within $\boldsymbol{\Delta^1_2}$.

Question 4.6 Quantify Det$(\boldsymbol{\Gamma_n})$.

Usually levels of determinacy (beyond that of Π^1_1) are calibrated by associating it with an embedding of a core model or a level of $L(\mathbb{R})$. A sample theorem of what is known:

Theorem 4.7 ([**10**]) ZFC + Det$(\boldsymbol{\Gamma_2}) \Rightarrow$ *There is an inner model with a proper class of strong cardinals.*

References

[1] M. Davis. Infinite games of perfect information. *Annals of Mathematical Studies*, 52:85–101, 1964.

[2] H. Friedman. Higher set theory and mathematical practice. *Annals of Mathematical Logic*, 2(3):325–327, 1970.

[3] S. D. Friedman and P. D. Welch. Hypermachines. *Journal of Symbolic Logic*, 76(2):620–636, 2011.

[4] J. D. Hamkins and A. Lewis. Infinite time Turing machines. *Journal of Symbolic Logic*, 65(2):567–604, 2000.

[5] R. Lubarsky. Well founded iterations of infinite time turing machines. In R.-D. Schindler (ed.), *Ways of Proof Theory.* Ontos, 2010.

[6] A. Montalbán and R. Shore. The limits of determinacy in second order number theory. *Proceedings of the London Mathematical Society*, to appear.

[7] Y. N. Moschovakis. *Elementary Induction on Abstract Structures*, volume 77 of *Studies in Logic series.* North-Holland, Amsterdam, 1974.

[8] M. Rathjen. An ordinal analysis of parameter-free Π^1_2 comprehension. *Archive for Mathematical Logic*, 44(3):263–362, 2005.

[9] P. D. Welch. Characteristics of discrete transfinite Turing machine models: halting times, stabilization times, and normal form theorems. *Theoretical Computer Science*, 410:426–442, 2009.

[10] P. D. Welch. Determinacy in strong cardinal models. *Journal of Symbolic Logic*, 76(2):719–728, 2011.

[11] P. D. Welch. Weak systems of determinacy and arithmetical quasi-inductive definitions. *Journal of Symbolic Logic*, 76(2):418–436, 2011.

Part IV

History and Philosophy of Set Theory

Foundational implications of the inner model hypothesis

Tatiana Arrigoni[†], **Sy-David Friedman**[‡]

[†] Fondazione Bruno Kessler, Trento, Italy
arrigoni@fbk.eu

[‡] Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
sdf@logic.univie.ac.at

Introduction

The goal of this paper is to bring the *Inner Model Hypothesis* (IMH), an axiomatic approach formulated by the second author in [**6**], into the current debate on the implications of independence results in set theory. We argue that the IMH provides an alternative to the two main contenders in this debate: the view that the universe of sets is *inherently undetermined*, its essential features being exhausted by the axioms of ZFC, and the opposing view that the next step toward the goal of making our knowledge of the universe of sets more determinate consists in the search for a suitable extension of the system "ZFC + large cardinal axioms". Both perspectives are objectionable in principle and the Inner Model Hypothesis confirms this in fact.

A brief overview of the current situation with regard to independence in set theory is given in Section 2. Section 3 illustrates the main views in the current debate on the implications of independence phenomena. Criticism against these views is presented in Section 4, while the implications of the Inner Model Hypothesis are discussed in Section 5.

Both authors wish to thank the John Templeton Foundation for its generous support of this work in the framework of the Infinity Project at the Centre de Recerca Matemàtica, Bellaterra, Spain.

1 A puzzling state of affairs

As a consequence of Gödel's construction of the *inner model* L and Cohen's introduction of *forcing* techniques in set theory, the existence of alternative universes satisfying the accepted axioms (i.e., the axioms of the system ZFC) has emerged as an inescapable fact. In addition to ZFC, the universe L of constructible sets satisfies the Generalized Continuum Hypothesis (GCH) —and therefore the Singular Cardinal Hypothesis (SCH)—, the assertion that there is a definable non-measurable set of reals, and the Singular Square Principle; it fails to satisfy the Suslin Hypothesis, the Whitehead Conjecture, the Borel Conjecture and the existence of a Borel bijection between any two non-Borel analytic sets.[1] On the other hand, many of these principles behave differently in forcing

To appear in *Annals of Pure and Applied Logic*.

[†]The first author is supported by *Provincia Autonoma di Trento*, Italy (Bando Post-Doc 2007).

[1] GCH is the assertion that for any cardinal number κ, $2^\kappa = \kappa^+$, while the SCH, implied by GCH, is the same assertion for κ a singular strong limit cardinal. For the other principles mentioned see [**13**].

extensions of L and, relative to the existence of large cardinals, they all behave differently in some model of ZFC.[2]

As a natural move in the attempt to decide statements independent from ZFC and thereby make our picture of the universe of sets more determinate, candidate axioms for extending ZFC have been proposed and investigated. In line with a suggestion of Gödel, a prominent role in this investigation has been played by *large cardinal axioms*.[3] With reference to such axioms Gödel says:

> It is not impossible that [...] some completeness theorem would hold which would say that every proposition expressible in set theory is decidable from the present axioms plus some true assertion about the largeness of the universe of all sets. ([**4**, pp. 150–153])

What came to be known as "Gödel's program for new axioms" did not however produce the desired results as far as independence is concerned. The statement of greatest interest which is independent from ZFC, Cantor's Continuum Hypothesis, is also independent from "ZFC + large cardinal axioms". But a relevant general fact emerged: The study of large cardinal axioms took the form of a strictly mathematical venture ("the theory is assumed and theorems are proved in the ordinary mathematical manner" [**5**, ix]), and its *mathematical success* was used as a source of evidence in set theory. *Success* is meant here as Gödel intended it, i.e., as consisting in axioms being "fruitful in consequences, exactly in 'verifiable' consequences, i.e., consequences demonstrable without the new axiom, whose proofs by means of the new axiom, however, are considerably simpler and easier to discover [...]",[4] as well as in axioms shedding light "upon a whole discipline, and furnishing [...] powerful methods for solving given problems" ([**8**, p. 183]).[5]

It is however worth noting that mathematical success can be reasonably ascribed to extensions of ZFC incompatible with "ZFC + large cardinal axioms". ZFC + $V = L$, for instance, is fruitful in consequences, furnishes powerful methods for solving problems and introduces the concept of *constructibility*, important throughout set theory.[6] Of course this theory is incompatible with "ZFC + there exists a measurable cardinal".[7]

How the mathematical success of large cardinal axioms is related to the program of making the picture of the set-theoretical universe more determinate —and, more generally, to the aim of producing *definitive* set-theoretical hypotheses— is discussed in the next two sections.

[2] Specifically, there are forcing extensions of L in which GCH is false, definable sets of reals are measurable and the Suslin Hypothesis, Whitehead Conjecture and the Borel Conjecture are true. Models of the negation of SCH, the negation of the Singular Square Principle and the existence of a Borel bijection between any two non-Borel analytic sets can be obtained assuming the existence of a hypermeasurable cardinal, a supercompact cardinal and a measurable cardinal, respectively.

[3] Large cardinal axioms assert the existence of cardinals κ with various strong properties, always implying that the family of sets of hereditary cardinality $< \kappa$ is a model of ZFC.

[4] "[...] and make it possible to condense into one proof many different proofs", [**8**, p. 183].

[5] On the success of large cardinal axioms, see [**11**] and [**1**].

[6] Inner and core models for large cardinals can be regarded as generalizations of the universe L of constructible sets; see [**14**].

[7] That if a *measurable cardinal* exists, then $V \neq L$ was proved by Scott in 1961; see [**13**] for details.

2 Reactions

Faced with the situation described in Section 2, set-theorists show diverse reactions. The existence of mutually incompatible, successful extensions of ZFC led some to the conclusion that the notion *set-theoretic universe* is inherently undetermined. This position is clearly expressed by Shelah in [**16**]:

> I do not feel "a universe of ZFC" is like "the Sun", it is rather like "a human being" or "a human being of some fixed nationality". [...] You may think "does CH, i.e., $2^{\aleph_0} = \aleph_1$ hold?" is like "Can a typical American be Catholic?" ([**16**, p. 211])

A different attitude is endorsed by those who, due to the success of large cardinal axioms, regard ZFC as "the twentieth century choice" for the axioms of set theory and consider "ZFC + large cardinal axioms" to be the contemporary theory of sets, "to be adopted by all, as part of a broadest point of view".[8] In fact these authors do not draw conclusions similar to Shelah's from the fact that large cardinals are preserved under forcing, and hence models of "ZFC + large cardinal axioms" exist in which mutually exclusive propositions are true. They put stress not on the failure of large cardinals to produce a determinate picture of the universe of sets but instead on the mathematical success of large cardinal axioms, and explicitly take this as providing evidence for the *correctness* (or *truth*) of these axioms, even regarding them as definitive hypotheses.[9] At the same time the hope is expressed that new correct (true) axioms will emerge that decide questions independent from the system "ZFC + large cardinals". As a result, the program of making the picture of the set-theoretical universe more determinate is placed in the restricted form: find suitable axiomatic extensions of "ZFC + large cardinals".

This forms part of Woodin's conclusions in [**19**], where an axiomatic proposal is advanced that is intended to play the same role with regard to third order number theory, in which the Continuum Hypothesis (CH) can be formulated, that is played by large cardinal axioms with regard to second order number theory.[10]

> So, is the Continuum Hypothesis solvable? Perhaps I am not completely confident that the "solution" I have sketched is the solution, but it is for me convincing evidence that there *is* a solution. [...] The universe of all sets is a large place. We have just barely begun to understand it. ([**19**, p. 690])

[8] See, respectively, [**19**] and [**17**], where the point is made that the "broadest point of view" proviso is meant to exclude from attention the temporary adoption of restrictive assumptions as a convenient device for avoiding irrelevant structure" (e.g., "$V = L$ is often temporarily assumed for such reasons by set-theorists who do not believe it [...]" [**17**, p. 422]).

[9] E.g., *Projective Determinacy* (PD), implied by the existence of infinitely many Woodin cardinals, is said in [**19**] to be "the *correct* axiom for the projective sets", yielding forcing invariant answers to questions independent of ZFC (e.g., the measurability of projective sets), which, when first formulated, were considered unsolvable; see [**19**, p. 570]. By *forcing invariance* is here meant that no sentence in the language of second order arithmetic, in which properties of projective sets are formulated, can be shown to be independent of the existence of large cardinals implying PD by the method of set-forcing. In fact, by a theorem of Woodin, if you suppose that every set belongs to an iterable inner model satisfying "there are ω Woodin cardinals", then, if M and N are set-generic extensions of V, you have $L(\mathbb{R})^M \equiv L(\mathbb{R})^N$; see [**19**].

[10] Second and third order number theory are presented in [**19**] as the theories of the structures $\langle H(\omega_1), \in\rangle$ and $\langle H(\omega_2), \in\rangle$. See [**19**] for details.

Both Shelah's and Woodin's positions are not immune to criticism. Objections to them are advanced in the next section.

3 Criticism

Let us start with positions like those expressed by Woodin regarding extensions of "ZFC + large cardinal axioms". According to them, successful, hence correct (true), set-theoretic axioms (large cardinal axioms) have been discovered that settle some notable questions independent from ZFC. This implies that the program for making the picture of the universe more determined cannot but consist in extending "ZFC + large cardinal axioms". We argue that the implication "success → correctness (or truth)" presupposed by this view is objectionable, and makes it ultimately untenable.

Observe first that by assuming the implication: "success → correctness (or truth)", one cannot do justice to the existence of mutually incompatible successful systems of set theory (like "ZFC + large cardinal axioms" and "ZFC + $V = L$"). For correctness (truth) is commonly intended as a matter of all or nothing, ruling out the possibility of equally correct (true) but mutually exclusive axiomatic systems. This would be the case, though, if evidence due to success were to imply correctness (truth) in set theory. On the other hand, assuming the implication "success → correctness (or truth)" and denying correctness or truth to e.g. "ZFC + $V = L$", one would *ipso facto* deny its mathematical success, which is undeniable.

The success of the axiom of constructibility ($V = L$) is often regarded as a counterexample to the view that success is all there is to correctness and truth in set theory.

> A favorite example against the pragmatic view that we accept an axiom because of its elegance (simplicity) and power (usefulness) is the constructibility hypothesis. It should be accepted according to the pragmatic view but is not generally accepted as true. ([**18**, p. 196])

Wang suggests what would be necessary and sufficient conditions for an axiomatic system to be accepted (as correct or true). Beyond being successful, the system should be explicitly suggested by the meaning of set.

> [$V = L$] is likely to be false according to the iterative concept of set. Basically it is felt that the pragmatic view leaves out the criterion of intuitive plausibility. ([**18**, p. 196])

Wang's argument, however, does not apply to most large cardinal axioms and, especially, to the ones discussed by Woodin. "Correct" ("true") principles like Projective Determinacy, and the large cardinal axioms implying it, lack any clear direct link to the iterative concept, which Wang calls upon as the meaning of set. In fact referring to these axioms, and explicitly describing them as "true", Woodin comments:

> There are natural questions about $H(\omega_1)$ which are not solvable from ZFC. However, there are axioms for $H(\omega_1)$ which resolve these questions [...] and which are clearly *true*. But the truth of these axioms became evident only *after* a great deal of work. ([**19**, p. 569])

Moreover, also the implication "success *and* intuitive plausibility (adherence to the iterative concept) → correctness (truth)" is objectionable. For it can be plausibly suggested that the iterative concept is a concept that arose alongside successful set-theoretic developments, and as such is a metaphorical reformulation of the insights delivered by the latter.[11] The same holds for methodological maxims that are often presented as inspired by the iterative concept, like e.g. "maximize", the view that the universe of sets should be high and wide, so "the more sets one proves to exist, the better". A mathematical concept could only be attached to the sentence "the universe is maximal" only after Scott's result that if a measurable cardinal exists then $V \neq L$ was obtained. Viewing the iterative concept and methodological principles like "maximization" in this way leads one to reject Wang's suggestion that "intuitive plausibility" (i.e., adherence to the iterative concept or "maximization") is sufficient, in conjunction with success, to produce truth or correctness in set theory. For, along with every system of set theory that turns out to be successful (according to Gödel's characterization of success), a distinguished concept of set and a system of preferred methodological maxims is likely to emerge.[12] Since competing successful systems of axioms exist in set theory, taking the conjunction "success *and* intuitive plausibility" to imply correctness (or truth), would still leave one with mutually exclusive, correct (true) systems of axioms. This contrasts with how the term *correct* (*true*) is meant to be used.

It is also worth noting that methodological maxims are very far from suggesting unique proposals for axiomatic extensions of ZFC. E.g., "maximization" may suggest the principle "there exists a $j\colon V \to V$", which is incompatible with the *Axiom of Choice*, also in line with maximality considerations.[13] The *Inner Model Hypothesis*, incompatible with large cardinal axioms, offers yet another example of the ambiguity of the concept of "maximization" (see the next section).

One might still object to our criticism by asserting that success comes in degrees in set theory, making it possible to draw a distinction between incompatible successful systems according to their degree of success, and suggest that it is only the most successful set-theoretic system that deserves to be regarded as correct or true. That mathematical success comes in degrees seems to be the case. According to Gödel's characterization of success, in fact, the term "successful" is to be applied to mathematical developments through which a link is established between formerly unrelated mathematical facts. A link may consist in one theory's enabling the interpretation of another in its own terms. Under these circumstances, the former would reveal itself to be "more successful" than the latter. In fact, as an implication of Scott's theorem, the universe L could be seen as a proper sub-universe of V and studied "from within" V under large cardinal axioms, thereby convincing some of the superior success of "ZFC + large cardinals" over "ZFC + $V = L$". Supposing "maximality" to be essentially a matter of maximizing interpretative power, Steel says the following with regard to ZFC + $V = L$:

> In this light we can see why most set-theorists reject $V = L$ as restrictive: adopting it restricts the interpretative power of the language of set theory. The language of set theory as used by the believer of $V = L$ can certainly be translated into the language of set theory as used by the believer in measurable cardinals, via the translation $\phi \mapsto \phi^L$. There is no translation in the other direction. While it is true that adopting $V = L$ enables one to settle new formal sentences, this is in fact a completely

[11] See [**1**] and [**2**].

[12] This view is presented and motivated in [**2**].

[13] This point is made in [**12**]. The principle "there exists a $j\colon V \to V$" (there is a nontrivial elementary embedding of the universe into itself) was proved to be contradictory with Choice by Kunen. See [**13**] for details.

> sterile move, because one settles ϕ by giving it the same interpretation as ϕ^L which can be settled in anyone's theory. ([**17**, p. 423])

Yet it remains that while one may accept that success comes in degrees, this is usually not the case as far as correctness and truth are concerned. Accordingly, correctness (truth) might well be supposed to be an attribute of the "most successful system of set theory", but this could not be done by arguing that correctness (truth) is an implication of success. The only possible way for one to coherently say that a successful axiomatic system for sets is correct (true) seems to be that of explicitly presenting one's position as a deliberate act, an act based on the decision to attach correctness (truth) to success "at the highest degree", as well as on a shared agreement as to what the most successful axiomatic system for sets currently is. However, at the moment, there is no agreement among set-theorists as to what the most successful theory of sets is.[14] Skeptical positions on the status of large cardinal axioms have been expressed (see, e.g., [**16**]). Arguments like Steel's to the effect that an interpretation of "ZFC + $V = L$" in terms of "ZFC + large cardinal axioms" is possible but not vice-versa, have been contested as well. Jensen, for instance, maintains that the relation between "ZFC + large cardinal axioms" and "ZFC + $V = L$" is one of mutual interpretability. For L itself can see the existence of "natural" models for large cardinal axioms if there are such cardinals in V. As a consequence of *Shoenfield's Absoluteness Lemma*, in fact, L and V have transitive countable models for the same large cardinal hypotheses.[15] "Hence we could just assume ourselves to be in a countable segment of L when we assume H".[16]

To sum up: the view that success furnishes evidence for correctness (truth), though not *per se* contradictory, does not help in defending the view that the program for making our picture of the universe more determinate must consist in finding suitable extensions of "ZFC + large cardinal axioms". At most it suggests that one should be cautious in taking as correct (true) what one regards as the most successful axiomatic system for sets, as there exist views about success that run contrary to one's own.

Let us add that, in fact, neither a simple identification of correctness (truth) with success, nor the view that correctness (truth) is conventionally attached to success "at the highest degree", seems to underlie positions like Woodin's. A Platonistic attitude appears to be at work. This is explicitly admitted by Foreman in [**5**]; with regard to consistency results involving large cardinal axioms, he observes:

> This type of unifying deep structure is taken as strong evidence that the axioms proposed reflect some underlying reality and so is often cited as a primary reason for accepting the existence of large cardinals. ([**5**, x])

Under these circumstances, correctness (truth) rests no longer on success. Success may well be regarded as a clue to it —if it is supposed that it is correctness (truth), meant as a matter of "reflecting some underlying reality", that ultimately implies success (or, better, success "at the highest degree"). Moreover, by regarding correctness (truth) as sufficent,

[14] Nor is there, one may guess, on the "conventional" view of correctness and truth introduced here.

[15] In fact, if the hypothesis H holds in V, then by reflection H should have a model that is a level V_κ of V (note this informal step in the argument) for some cardinal κ. By the *Löwenheim–Skolem* theorem, there is a countable elementary sub-model of V_κ, call it N, in which H holds. By *Mostowski's Collapsing Theorem* there is a transitive $\overline{N}$ that is a countable model of H. Let be $a \in \mathbf{R}$ be a code for $\overline{N}$. The formula asserting the existence of such an a is Σ^1_2. By *Shoenfield's Absoluteness Lemma*, it is true in L. That is, L sees the existence of a transitive countable set model for H.

[16] Quoted with permission from the handout of a talk given by Jensen in Krakow in 1999.

as opposed to necessary, to success, an explanation would be given, too, for the existence of mutually exclusive successful systems of set theory. For, under these circumstances, the existence of successful set theories that cannot be said to be correct (true) is no longer contradictory. However, one should still justify Platonism in order for this position to be sound. This is no easy task. Neither pursuing this justification nor criticizing it belongs to the aims of the present paper.

Having focused on the positions of Woodin, Steel and Foreman, let us now return to Shelah's views. Here one abdicates the search for new axioms that may yield solutions to questions independent of ZFC, solutions to which correctness or truth can be attached as the end-stage of a process through which a shared consensus is reached that certain mathematical developments, and the axioms that make them possible, are the most successful ones. This abdication may have positive consequences. It may work as a heuristic for exploiting the available resources (ZFC), to the effect that light is shed on still undiscovered implications of them, perhaps relevant with regard to independence phenomena. Shelah's *pcf theory*, developed entirely within ZFC, has a bearing on questions of cardinal arithmetic like the Generalized Continuum Hypothesis.[17] However, it might be felt that whereas positions like Shelah's are supported by the existence of incompatible successful set-theoretical developments, they also prescribe a halt to such developments by regarding ZFC as all there is to be said about sets. Shelah's conclusions also sound arbitrary. Why should the view that a universe of ZFC be not like "the Sun" but like "a human being of some fixed nationality" be a definitive one? Why not regard it as a description of a state of affairs that need not be permanent, merely reflecting the *actual* situation in set theory, where no development stands out as the most successful (and hence, one may add, the correct or true) one? As it seems premature to say that convincing evidence is available that the correct answer to the question "Is CH true/false?" is given by a suitable extension of "ZFC + large cardinal axioms", so seems it premature to rule this out and be content with the view that the notion *universe of all sets* is inherently undetermined.

As a case study supporting the above criticisms, we discuss the second author's *Inner Model Hypothesis* (IMH) in the following section. The IMH also provides a striking example of a phenomenon alluded to above, the ambiguity of the concept of "maximization".

4 The Inner Model Hypothesis

We begin with a restatement of our thesis. Objections can be raised against the view that the notion *universe of all sets* can only be made determinate by finding axiomatic extensions of "ZFC + large cardinal axioms" which successfully decide questions independent of the latter. In advancing this view it is assumed that mathematical success provides evidence for the correctness or truth of large cardinal axioms, which renders these axioms definitive set-theoretic principles that one can only "extend" but not contradict. In assuming that success implies correctness (truth), however, one is either tacitly committed to Platonism or faces the embarrassing situation that mutually exclusive and successful axiomatic systems for sets coexist. On the other hand, no a priori ground seems to exist for ruling out the possibility of making the notion *universe of all sets* more determinate than it is now through the introduction of new axiomatic proposals.

[17] See [**16**, p. 220]: "Cardinal arithmetic is loaded with consistency results because we ask the wrong questions. [...]. We should replace cardinality by cofinality, as explained below (pcf theory)".

By advancing the *Inner Model Hypothesis*, one *de facto* remains open to the possibility of making the universe of sets more determinate. At the same time, one does not impose the restriction of consistency with "ZFC + large cardinal axioms". The approach of the Inner Model Hypothesis is not to "determine" the universe by directly postulating what sets exist in it (which is done when e.g. large cardinals are assumed to exist in V), but to state from a metatheoretical perspective what properties the universe of sets is supposed to possess.

Let us discuss the hypothesis in more detail. How can metatheoretical properties be identified which one may wish the universe V of sets to have? The suggestion made in [**6**] is that one start from ZFC (or from a theory for sets and classes like Gödel–Bernays) and provisionally regard V as a model for it endowed with *countably many* sets (and classes). For a countable universe V many techniques are available for creating not only inner universes of V but also outer universes of V, i.e., universes V^* such that $V \subseteq V^*$, to which V can be compared. These techniques not only include (set and class) forcing, but also methods that arise from further generalizations of the forcing method (such as hyperclass forcing) or from infinitary model theory. Being able to compare V to a multitude of other universes enables one to better formulate properties that one wishes the intended universe V to obey. The *Inner Model Hypothesis* takes advantage of this method of comparison:

> *If a statement ϕ without parameters holds in an inner universe of some outer universe of V (i.e., in some universe compatible with V), then it already holds in some inner universe of V.*

Equivalently: statements that are *internally consistent* with respect to an outer universe of V are already *internally consistent* in V, where a statement is *internally consistent* if it holds in some inner universe. It follows that by enlarging V one gains nothing as far as internal consistency is concerned. So according to the Inner Model Hypothesis, V is *maximal* with respect to internal consistency.[18]

Although the IMH is formulated by supposing V to be countable, the Inner Model Hypothesis can also be formulated as a (weaker) hypothesis for an uncountable V. This is done by restricting the notion of outer universe to the set- and class-generic extensions of the given universe that preserve the Gödel–Bernays axioms, thereby reducing the hypothesis to a principle of ordinary class theory. Alternatively, one may regard the IMH as saying that although V itself is not countable, it should satisfy sentences that are true in countable universes which are maximal with respect to internal consistency. It is also worth noting that having the universe maximize internal consistency via the IMH *generalizes* a phenomenon known to hold for formulas (without parameters) proved to be consistent by set-forcing.[19]

One knows a lot about the consistency strength of the *Inner Model Hypothesis.* It is established by the following results.[20] 1) Assume that there is a Woodin cardinal and a larger inaccessible cardinal. Then there are universes which maximize internal consistency, so the Inner Model Hypothesis is consistent. 2) The Inner Model Hypothesis

[18] To put it in other terms, if $\mathfrak{L}$ = language of set/class theory and, for a universe W, $\Phi(W)$ = all sentences of $\mathfrak{L}$ which are true in some inner universe of W, then, under the Inner Model Hypothesis, if $V \subseteq W$ then $\Phi(V) = \Phi(W)$.

[19] See [**6**] for the details of this claim.

[20] See [**7**].

implies that there are inner models with measurable cardinals of arbitrarily large Mitchell order.

Note that by adopting the Inner Model Hypothesis, while not extending "ZFC + large cardinal axioms", one does appeal to large cardinals in two respects. First, large cardinal axioms are invoked for establishing its consistency strength. This acknowledges the major feature of the mathematical success of large cardinal axioms, their ability to prove consistency. The relevance of these axioms is seen here as metamathematical rather than as mathematical. Second, one asks whether the Inner Model Hypothesis has relevant implications with regard to large cardinals. This is in fact the case. Among the consequences of the Inner Model Hypothesis is that no *inaccessibles*, hence no large cardinals, exist in V and that the real numbers are not closed under the $\sharp$ operation. That is to say: not only is the Inner Model Hypothesis not an extension of the system "ZFC + large cardinals"; it is also incompatible with it! The *consistency* of large cardinal axioms is however preserved under the IMH (V sees inner models for them); it is only their *existence* that is contradicted.

This latter point also has important consequences for the methodological notion of "maximization". The IMH clearly asserts a *maximal* property of the universe of sets, namely that internal consistency has been maximized. But it is at the same time in conflict with the existence of large cardinals. This is despite the fact that large cardinal axioms have also been traditionally assumed to assert a form of maximality for the universe of sets. Let us return to Gödel:

> From an axiom in some sense opposite to $[V = L]$, the negation of Cantor's conjecture could perhaps be derived. I am thinking of an axiom which ... would state some maximum property of the system of all sets, whereas $[V = L]$ states a minimum property. Note that only a maximum property would seem to harmonize with the concept of set. ([**9**, pp. 262–263])

Note that there is no implication in this quote that "maximization" must be based on large cardinal axioms. And indeed, the IMH provides an alternative way of maximizing the universe of sets, thereby revealing the profound ambiguity of this concept.

What about questions which are independent from ZFC? Some of them are decided under the Inner Model Hypothesis, e.g., the Singular Cardinal Hypothesis and the existence of a projective non-measurable set of reals, which turn out to be true, and the existence of a Borel bijection between any two non-Borel analytic sets, which, instead, turns out to be false.[21] The Continuum Hypothesis remains undecided, though. For, suppose that V satisfies the Inner Model Hypothesis. One can create, by set forcing, a larger universe $V[G]$, in which CH is true (using a "Lévy collapse"). Since V is contained in $V[G]$, The Inner Model Hypothesis is also true in $V[G]$. So the hypothesis is consistent with CH. It cannot imply its negation. Similarly, one can create a larger universe $V[H]$

[21] Theorem 15 in [**6**] proves that that IMH implies the existence of a real R such that ZFC fails in $L_\alpha[R]$ for all ordinals α. This property implies that (a) for some real R, $\aleph_1 = \aleph_1^{L[R]}$, which in turn implies that (b) for some real R, $R^\sharp$ does not exist, which is equivalent to (c): for some real R, Jensen's covering property holds relative to $L[R]$ (i.e., every uncountable set of ordinals is a subset of a set in $L[R]$ of the same size). The truth of the *Singular Cardinal Hypothesis* and the *Singular Square Principle* and the falsity of the existence of a Borel-isomorphism of non-Borel analytic sets (via the results presented in [**10**]) follow from (c), while the existence of a projective non-measurable set of reals (via the results in [**15**]) follows from (a).

in which CH is false (by adding $\aleph_2$ Cohen reals), the Inner Model Hypothesis being true in $V[H]$. So the Inner Model Hypothesis cannot imply CH either. One needs a stronger version of the Inner Model Hypothesis to settle CH, i.e., the hypothesis for formulas with globally absolute parameters.[22] A consistency proof for the resulting *Strong Inner Model Hypothesis* (SIMH) is however still lacking.

Let us conclude with a bold question. Will the *Inner Model Hypothesis*, and its implications, be accepted as a definitive feature of the universe, making it more determinate than it is now? According to the views presented throughout this paper, the considerable mathematical success of the IMH is to play a decisive role in this respect, whether or not one deliberately decides to attach correctness (truth) to the most successful set-theoretic hypotheses. But the philosophical implications of the IMH are clear, as it presents an important challenge to two widely-shared views in contemporary set theory.

References

[1] Tatiana Arrigoni. *What is meant by V? Reflections on the Universe of all Sets.* Mentis Verlag, Paderborn, 2007.

[2] Tatiana Arrigoni. $V = L$ and intuitive plausibility in set theory. A case study. *Bulletin of Symbolic Logic*, 17(3):337–360, 2011.

[3] Paul Benacerraf and Hilary Putnam (eds.). *Philosophy of Mathematics. Selected Readings. Second Edition.* Cambridge University Press, 1983.

[4] S. Feferman, J. Dawson, S. Kleene, G. Moore, and J. Van Heijenoort (eds.). *Kurt Gödel. Collected Works, Volume II.* Oxford University Press, New York, 1990.

[5] Matthew Foreman and Akihiro Kanamori (eds.). *Handbook of Set Theory.* Springer, 2010.

[6] S. D. Friedman. Internal consistency and the inner model hypothesis. *Bulletin of Symbolic Logic*, 12(4):591–600, 2006.

[7] S. D. Friedman, P. Welch, and H. Woodin. On the consistency strength of the inner model hypothesis. *Journal of Symbolic Logic*, 73(2):391–400, 2008.

[8] Kurt Gödel. What is Cantor's continuum problem? *American Mathematical Monthly*, 54, 1947. Reprinted in [**4**], 176–187.

[9] Kurt Gödel. What is Cantor's continuum problem? In P. Benacerraf and H. Putnam (eds.), *Philosophy of Mathematics. Selected Readings*, pages 258–273. 1964. Revised and expanded version of [**8**]. Reprinted in [**3**], 470–485 and [**4**], 254–269. Quoted from [**4**].

[10] Leo A. Harrington. Analytic determinacy and $0^{\#}$. *Journal of Symbolic Logic*, 43(4):685–693, 1978.

[11] Kai Hauser. Was sind und was sollen neue Axiome. In G. Link (ed.), *One Hundred Years of Russell's Paradox*, pages 93–117. De Gruyter, Berlin, 2004.

[12] Kai Hauser. Is Choice self-evident? *American Philosophical Quarterly*, 42:237–261, 2005.

[13] Thomas Jech. *Set Theory. The Third Millenium Edition, Revised and Expanded.* Springer-Verlag, Berlin, Heidelberg, New York, 2003.

[14] Ronald Jensen. Inner models and large cardinals. *The Bulletin of Symbolic Logic*, 1:393–407, 1995.

[15] Saharon Shelah. Can you take Solovay's inaccessible away? *Israel Journal of Mathematics*, 48(1):1–47, 1984.

[16] Saharon Shelah. Logical dreams. *Bulletin of the American Mathematical Society*, 40(2):203–228, 2003.

[17] John Steel. Mathematics needs new axioms. *The Bulletin of Symbolic Logic*, 4:422–433, 2000.

[18] Hao Wang. *From Mathematics to Philosophy*, chapter VI. The concept of set, pages 181–223. Routledge and Kegan Paul, London, 1974.

[19] Hugh Woodin. The Continuum Hypothesis, i-ii. *Notices of the American Mathematical Society*, 48(7):567–576, 681–690, 2001.

[22] See [**6**].

Part V

Models and Sets

Amalgamation, absoluteness, and categoricity

John T. Baldwin[†]

[†] Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, USA
jbaldwin@uic.edu

Abstract. We describe the major result on categoricity in $L_{\omega_1,\omega}$, placing it in the context of more general work in abstract elementary classes. In particular, we illustrate the role of higher dimensional amalgamations and sketch the role of a weak extension of ZFC in the proof. We expound the translation of the problem to studying atomic models of first order theories. We provide a simple example of the failure of amalgamation for a complete sentence of $L_{\omega_1,\omega}$. We prove some basic results on the absoluteness of various concepts in the model theory of $L_{\omega_1,\omega}$ and publicize the problem of absoluteness of $\aleph_1$-categoricity in this context. Stemming from this analysis, we prove the following theorem: The class of countable models whose automorphism groups admit a complete left invariant metric is Π^1_1 but not Σ^1_1.

Introduction

The study of infinitary logic dates from the 1920's. Our focus here is primarily on the work of Shelah using stability theoretic methods in the field (beginning with [**30**]). In the first four sections we place this work in the much broader context of abstract elementary classes (AEC), but do not develop that subject here. The main result discussed, Shelah's categoricity transfer theorem for $L_{\omega_1,\omega}$, explicitly uses a weak form of the GCH. This raises questions about the absoluteness of fundamental notions in infinitary model theory. Sections 5–7 and the Appendix due to David Marker describe the complexity and thus the absoluteness of such basic notions as satisfiability, completeness, ω-stability, and excellence.[1] We state the question, framed in this incisive way by Laskowski, of the absoluteness of $\aleph_1$-categoricity. And from the model theoretic characterization of non-extendible models we derive the theorem stated in the abstract on the complexity of automorphism groups. Most of the results reported here in Sections 1–4 are due to Shelah; the many references to [**2**] are to provide access to a unified exposition. I do not know anywhere that the results in Section 5 have been published. The techniques are standard. Our main goal was to provide a reference for this material; but the distinction between the complexity of various notions for atomic classes as opposed to sentences of $L_{\omega_1,\omega}$ seems to be a new observation. The result in Section 6 is new but easy.

Appeared in the Proceedings of the 11th Asia Logic Conference, Toshiuasu Arai, Qi Feng, Bjungang Kim, Guoha Wu, Yue Yang (eds.), 2012, World Scientific Publishing Co. Pte. Ltd., Singapore, pp. 22–51.

[†]This article is a synthesis of the paper given in Singapore with later talks, including the Mittag-Leffler Institute in 2009 and CRM Barcelona in 2010. It reflects discussions with set theorists during my stay at Mittag-Leffler and discussion with the Infinity Project members in Barcelona. The author wishes to thank the John Templeton Foundation for its support through Project #13152, Myriad Aspects of Infinity, hosted during 2009–2011 at the Centre de Recerca Matemàtica, Bellaterra, Spain. Baldwin was partially supported by NSF-0500841.

[1]David Marker is partially supported by National Science Foundation grant DMS-0653484.

1 The universe is wide or deep

Shelah made the following rough conjecture: Let $\boldsymbol{K}$ be a *reasonable* class of models. Either for some λ there are many models of cardinality λ or there are models of arbitrarily large cardinality.

Our metaphor requires some explanation. 'The universe' should perhaps be 'each universe'; universe refers to all models in a specific class. Further we are taking 'or' in the inclusive sense. Certainly, there are classes (e.g., dense linear orders) which are both wide and deep. Perhaps, taking narrow, as meaning there are few models in each cardinality, the aphorism better reads. A narrow universe is deep. It turns out that this question depends very much on the choice of 'reasonable'. It also seems to be sensitive to the choice of axioms of set theory. In order to give a precise formulation of the conjecture we have to specify 'many' and the notion of a 'reasonable class'. In general 'many' should mean 2^λ; but in important cases that have been proved, it is slightly smaller.

As is often the case there are some simplifying assumptions in this area that have been internalized by specialists but obscure the issues for other logicians. We try to explain a few of these simplifications and sketch some of the major results.

Some historical background will help clarify the issues. Much model theoretic research in the 60's focused on general properties of first order and infinitary logic. A number of results seemed to depend heavily on extensions of ZFC. For example, both Keisler's proof that two structures are elementarily equivalent if and only if they have isomorphic ultrapowers and Chang's proof of two cardinal transfer required GCH. In general, even the existence of saturated models depends on the GCH. Shelah removed the set-theoretic hypothesis from Keisler's theorem. But various versions of two cardinal transfer were proven to require GCH and even large cardinal hypotheses; see [**7**].

The invention of stability theory radically recast the subject of model theory. E.g., for various classes in the stability hierarchy, it is straightforward to characterize in ZFC exactly in which cardinals there are saturated models. And for the best behaved theories the answers is: all cardinals. Further, for countable stable theories Shelah and Lachlan independently showed that two cardinal transfer between any pair of cardinalities is true in ZFC. Moreover, the fundamental notions of first order stability theory are absolute.

For first order logic, our guiding question is trivial.[2] If a theory has an infinite model then it has arbitrarily large models. The question is interesting for theories in logics which fail the upward Löwenheim–Skolem theorem. The notion of an Abstract Elementary Class (AEC) provides a general framework for analyzing such classes. But as we show in the next section the conjecture is trivially false in that case. It is not too difficult to find in ZFC examples (Example 2.1) of AEC that have no model above $\aleph_1$ but that are $\aleph_1$-categorical [**2, 34**]. And in $L_{\omega_1,\omega}(Q)$, it is consistent (via Martin's axiom) that there are $\aleph_1$-categorical sentences with no model of cardinality greater than $2^{\aleph_0}$. But those sentences have many models in $2^{\aleph_0}$. In this note we describe how for $L_{\omega_1,\omega}$ there are major advances on the target problem. They use extensions of ZFC but rather mild ones; the initials below refer to the 'Weak Continuum Hypothesis' and the 'Very Weak Continuum Hypothesis':

WGCH: Cardinal exponentiation is an increasing function.

[2] The main gap theorem, every first order theory either eventually has the maximal number of models or the number of models is bounded by a small function, has the same flavor. And in fact the argument for this result arose after Shelah's consideration of the infinitary problems.

VWGCH: Cardinal exponentiation is an increasing function below $\aleph_\omega$.

This leaves us with two more precise questions:

(1) Does the proof of the conjecture for $L_{\omega_1,\omega}$ (see Section 4) really need VWGCH?

(2) Is the conjecture 'eventually true' for AEC's? [3]

Much of core mathematics studies either properties of particular structures of size at most the continuum or makes assertions that are totally cardinal independent. E.g., if every element of a group has order two then the group is abelian. Model theory and even more clearly infinitary model theory allows the investigation of 'structural properties' that are cardinal dependent such as: existence of models, spectra of stability, and number of models and existence of decompositions. Often these properties can be tied to global conditions such as the existence of a 'good' notion of dependence.

2 Abstract elementary classes

We begin by discussing the notion of an abstract elementary class. The examples show that this is too broad a class to be 'reasonable' for our target problem. But some positive results can be proved in this general setting; this generality exposes more clearly what is needed for the argument by avoiding dependence on accidental syntactic features.

An abstract elementary class[4] $(\boldsymbol{K}, \prec_{\boldsymbol{K}})$ is a collection of structures for a fixed vocabulary τ that satisfy the following, where $A \prec_{\boldsymbol{K}} B$ means in particular that A is a substructure of B:

(1) If $A, B, C \in \boldsymbol{K}$, $A \prec_{\boldsymbol{K}} C$, $B \prec_{\boldsymbol{K}} C$ and $A \subseteq B$ then $A \prec_{\boldsymbol{K}} B$.

(2) Closure under direct limits of $\prec_{\boldsymbol{K}}$-embeddings.

(3) Downward Löwenheim–Skolem: If $A \subset B$ and $B \in \boldsymbol{K}$ then there is an A' with $A \subseteq A' \prec_{\boldsymbol{K}} B$ and $|A'| \leq |A| + \mathrm{LS}(\boldsymbol{K})$.

The invariant $\mathrm{LS}(\boldsymbol{K})$ is a crucial property of the class. The class of well-orderings satisfies the other axioms (under end extension) but is not an AEC.

Two easy examples are: First order and $L_{\omega_1,\omega}$-classes; $L(Q)$ classes have Löwenheim–Skolem number $\aleph_1$. For the second case one has to be careful about the definition of $\prec_{\boldsymbol{K}}$ —being an $L(Q)$-elementary submodel does not work (a union of a chain can make $(Qx)\phi(x)$ become true even if it is false along the chain).

The notion of AEC has been reinterpreted in terms of category theory by Kirby: "Abstract Elementary Categories" **[18]** and by Lieberman: "AECs as accessible categories" **[22]**. It is easy to see that just AEC is too weak a condition for the general conjecture.

Example 2.1 The class of well-orderings with order-type at most ω_1 with $\prec_{\boldsymbol{K}}$ as initial segment is an AEC with $\aleph_1$ countable models. It is $\aleph_1$-categorical and satisfies both amalgamation and joint embedding but is not ω-Galois stable **[19]**. And in fact there is no model of cardinality $\aleph_2$. So this universe is neither wide nor deep.

Let us clarify the specific meaning of the amalgamation property in this context. The arrows here denote morphisms in the abstract elementary class; various strengthening requiring certain maps to be inclusions are well-known.

[3] For much positive work in this direction, see **[34]**.

[4] Naturally we require that both $\boldsymbol{K}$ and $\prec_{\boldsymbol{K}}$ are closed under isomorphism.

Definition 2.2 The class $\boldsymbol{K}$ satisfies the *amalgamation property* if, for any situation with $A, M, N \in \boldsymbol{K}$,

$$\begin{tikzcd}
 & N \\
A \arrow[ur] \arrow[dr] & \\
 & M,
\end{tikzcd}$$

there exists an $N_1 \in \boldsymbol{K}$ such that

$$\begin{tikzcd}
 & N \arrow[dr] & \\
A \arrow[ur] \arrow[dr] & & N_1. \\
 & M \arrow[ur] &
\end{tikzcd}$$

Note that we have required the base structure A to be in $\boldsymbol{K}$; this is sometimes referred to as 'model amalgamation'. Requiring amalgamation over arbitrary substructures A is a much stronger condition, which fails for important natural examples such as Zilber's pseudo-exponential field [**40**]. There is much work in homogenous model theory where the stronger homogeneity condition is assumed.

The existence of amalgamations is an absolutely fundamental problem for AEC and for any study of infinitary logic. In first order logic it is easy to show that for complete theories amalgamation always holds over models with $\prec$ as elementary extension. And it holds over arbitrary subsets of models if T admits elimination of quantifiers. Here is a basic example of failure for a complete sentence of $L_{\omega_1,\omega}$.

Example 2.3 Let T be the first order theory in a language with binary relation symbols $\langle E_i : i < \omega\rangle$ that asserts that the E_i are infinitely many refining equivalence relations with binary splitting.

Using $L_{\omega_1,\omega}$, the equivalence relation E_∞ —the intersection of the given equivalence relations— is definable. Add two unary predicates (blue and red) and the following infinitary axioms:

(1) Each E_∞-class contains infinitely many elements.
(2) Every element of an E_∞-class is red or every element is blue.
(3) Blue and red divide the E_∞-classes into dense and codense subsets of the natural linear order of the paths.

Now it is easy to check that these axioms are $\aleph_0$-categorical but fail amalgamation (since a new path may be either red or blue).

We introduced the notion of abstract elementary class in this paper in order to state 'one completely general result' which can be found in [**34**, I.3.8] or [**2, 32**].

Theorem 2.4 (WGCH) *Suppose $\lambda \geq \mathrm{LS}(\boldsymbol{K})$ and $\boldsymbol{K}$ is λ-categorical. If amalgamation fails in λ, then there are 2^{λ^+} models in $\boldsymbol{K}$ of cardinality λ^+.*

As opposed to many other results in the study of abstract elementary classes which rely on an additional collection of model theoretic hypotheses, this result is about *all* AEC's. Moreover, variants of the proposition recur repeatedly in the proof of the main result being expounded. The argument uses weak diamond and is primarily combinatorial; it proceeds directly from the definition of an AEC. The result fails under $MA + \neg CH$. An example is presented in [**34, 38**] and a simpler one in [**2**]. It is an AEC (even given by a theory in $L(Q)$) which fails amalgamation in $\aleph_0$, but becomes $\aleph_1$-categorical in a forcing extension. But it remains open whether there are such examples in $L_{\omega_1,\omega}$. Easy examples ([4]) show that categoricity is a necessary condition for Theorem 2.4. This has a fundamental impact on the structure of the main proof. Because of this we must pass to complete sentences and gain categoricity in $\aleph_0$. One strategy in Shelah's approach through frames in [**34**] evades the categoricity difficulty by restricting to subclasses of the AEC, e.g., the λ-saturated models.

Amalgamation plays a fundamental role in the study of AEC's. One line of research pioneered by Shelah [**33**] and highly developed by Grossberg, VanDieren, and Lessmann in a series of papers (e.g., [**12**]) assumes arbitrarily large models, joint embedding, and amalgamation; under strong model theoretic assumptions the results are proved in ZFC. An account of this work with full references to the published papers appears in Part II of [**2**]. In this paper we focus on earlier work on $L_{\omega_1,\omega}$, which is a little more concrete as the logic is fixed. But it is more general in another way. Rather than assuming amalgamation, failure of amalgamation is shown to create width. Both amalgamation and the existence of large models are proved for narrow classes; this brings the set theoretic difficulties into view. The work of Hyttinen and Kesala on finitary AEC (e.g., [**15**]) continues the program of assuming arbitrarily large models and amalgamation. However, even stronger model theoretic assumptions lead to the development of a geometric stability theory. Several further directions of study in AEC are explored in [**34**]. The introduction to that book surveys the field and explains Shelah's viewpoint. The method of frames, expounded in [**34**], provides an approach to the problem of building larger models from categoricity in one or several successive uncountable cardinals; he attempts to avoid the traces of compactness that simplify the work starting at $\aleph_0$ and $\aleph_1$ in $L_{\omega_1,\omega}$. In other papers, Shelah (e.g., [**39**]) considers the general problem of eventual categoricity assuming large cardinal axioms.

3 From $L_{\omega_1,\omega}$ to first order

We begin by translating the problem from infinitary logic into the study of specific subclasses of models of first order theories. This removes the distraction of developing new notions of each syntactic idea (e.g., type) for each fragment of $L_{\omega_1,\omega}$. More subtly, for technical reasons we need to restrict to complete sentences in $L_{\omega_1,\omega}$. (This restriction to complete sentences is automatic in the first order case but its legitimacy is only proved in certain cases for infinitary logic.)

Definition 3.1 For Δ a fragment of $L_{\omega_1,\omega}$, a Δ-theory T is Δ-*complete* if, for every Δ-sentence ϕ, either $T \models \phi$ or $T \models \neg\phi$. We may write *complete* when $\Delta = L_{\omega_1,\omega}$.

Definition 3.2

(1) A model M of a first order theory is called *atomic* if each finite sequence from M realizes a principal type over the empty set —one generated by a single formula.

(2) An *atomic class* is an AEC, consisting of the atomic models of a complete first order theory T with elementary submodel as the notion of strong submodel. Thus, $\mathbb{M}$ is a large saturated model of T, usually not atomic. A set $A \subset \mathbb{M}$ is an *atomic set* if each finite sequence from A realizes a principal type over the empty set-generated by a single formula.

The study of categoricity (at least from $\aleph_1$ upwards), in $L_{\omega_1,\omega}$ can be translated to the study of atomic models of a first order theory. This is non-trivial. The argument begins with a fundamental result from the early 60's.

Theorem 3.3 (Chang and López-Escobar) *Let ϕ be a sentence in $L_{\omega_1,\omega}$ in a countable vocabulary τ. Then there is a countable vocabulary τ' extending τ, a first order τ'-theory T, and a countable collection of τ'-types Γ such that reduct is a* 1-1 *map from the models of T which omit Γ onto the models of ϕ.*

The proof is straightforward. E.g., for any formula ψ of the form $\bigwedge_{i<\omega} \phi_i$, add to the language a new predicate symbol $R_\psi(\mathbf{x})$. The theory T will contain the sentences for each subformula ψ of ϕ:

$$(\forall \mathbf{x})[R_\psi(\mathbf{x}) \rightarrow \phi_i(\mathbf{x})]$$

for $i < \omega$ and omit the type $p = \{\neg R_\psi(\mathbf{x})\} \cup \{\phi_i : i < \omega\}$. There are similar requirements for other steps in the inductive definition of θ.

Thus we have restricted to the models of a theory that omit a family Γ of types, but that may realize some non-principal types. Shelah observed that if T had only countably many types then a similar expansion of the vocabulary gives a T' such that the required interpretation is obtained by omitting *all* non-principal types. That is, the object of study is the atomic models of T'. This further reduction is technically important. In particular it implies ω-categoricity.

But why can we assume that the T associated with θ has only countably many types over the empty set? We need a few definitions to give an explanation.

Definition 3.4 Fix a sentence $\phi \in L_{\omega_1,\omega}$ and let Δ be a countable fragment of $L_{\omega_1,\omega}$ containing ϕ.

(1) A τ-structure M is Δ-*small* if M realizes only countably many Δ-types (over the empty set).

(2) An $L_{\omega_1,\omega}$-sentence ϕ is Δ-*small* if there is a countable set X of complete Δ-types over the empty set and each model realizes some subset of X.

Here 'small' means $\Delta = L_{\omega_1,\omega}$.

It is easy to see that if M is small then M satisfies a complete sentence. If ϕ is small then Scott's argument for countable models generalizes and there is a complete sentence ψ_ϕ such that $\phi \wedge \psi_\phi$ has a countable model. So ψ_ϕ implies ϕ. But ψ_ϕ is not in general unique. For example ϕ might be just the axioms for algebraically closed fields. Two choices for ψ_ϕ are the Scott sentence of the prime field and the Scott sentence for the model of transcendence degree $\aleph_0$. Only the second has an uncountable model.

We can make an appropriate choice of ψ_ϕ if ϕ is $\aleph_1$-categorical. There are two ingredients in the choice.

Theorem 3.5 (Shelah) *If ϕ has an uncountable model M that is Δ-small for every* countable Δ *and ϕ is $\aleph_1$-categorical then ϕ is implied by a complete sentence ψ with a model of cardinality $\aleph_1$.*

This result appears first in [**31**]. It is retold in [**2**]; in [**1**], we adapt the argument to give a model theoretic proof of a result of Makkai (obtained by admissible set theory) that a counterexample to Vaught's conjecture is not $\aleph_1$-categorical. The crux of Shelah's argument is an appeal to the non-definability of well-order in $L_{\omega_1,\omega}$.

The second step is to require that for each countable fragment Δ there are only countably many Δ-types over the empty set. If ϕ has arbitrarily large models this is easy by using Ehrenfeucht–Mostowski models. But if not, the only known argument is from few models in $\aleph_1$ and depends on a subtle argument of Keisler [**17**] (see also Appendix C of [**2**]).

Theorem 3.6 (Keisler) *If ϕ has $< 2^{\aleph_1}$ models of cardinality $\aleph_1$, then each model of ϕ is Δ-small for every* countable Δ.

Now Theorems 3.5 and 3.6 immediately yield:

Theorem 3.7 (Shelah) *If ϕ has $< 2^{\aleph_1}$ models of cardinality $\aleph_1$, then there is a complete sentence ψ such that ψ implies ϕ and ψ has an uncountable model. In particular, if ϕ is $\aleph_1$-categorical there is a Scott sentence for the model in $\aleph_1$, i.e., the model in $\aleph_1$ is small. So an atomic class $\boldsymbol{K}$ is associated with ϕ .*

It is easy to construct a sentence ϕ such that no completion has an uncountable model, i.e., no uncountable model is small. Let τ contain binary relations E_n for $n < \omega$. Let ϕ assert that the E_n are refining equivalence relations with binary splitting, and that there do not exist two distinct points that are E_n equivalent for all n. And add a countable set A of constants that realize a dense set of paths. Now every uncountable model realizes uncountably many distinct types over A.

We have the following question, which is open if $\kappa > \aleph_1$.

Question 3.8 If ϕ is κ-categorical must the model in κ be small?

Thus for technical work we will consider the class of atomic models of first order theories. Our notion of type will be the usual first order one —but we must define a restricted Stone space.

Definition 3.9 Let A be an atomic set; $S_{\rm at}(A)$ is the collection of $p \in S(A)$ such that if $\boldsymbol{a} \in \mathbb{M}$ realizes p then $A\boldsymbol{a}$ is atomic.

Here $\mathbb{M}$ is the monster model for the ambient theory T; in interesting cases it is not atomic. And the existence[5] of a monster model for the atomic class associated with a sentence categorical in some set of cardinals is a major project. (It follows from excellence; after Theorem 4.3, we see under VWGCH categoricity up to $\aleph_\omega$ is sufficient.)

Definition 3.10 $\boldsymbol{K}$ is λ-*stable* if for every *model* M in $\boldsymbol{K}$ (thus necessarily atomic) with cardinality λ, $|S_{\rm at}(M)| = \lambda$.

The insistence that M be a model is essential. The interesting examples of pseudo-exponential fields, covers of Abelian varieties and the basic examples of Marcus and Julia Knight all are ω-stable but have countable sets A with $|S_{\rm at}(A)| > \aleph_0$.

With somewhat more difficulty than the first order case, one obtains:

Theorem 3.11 *For a class $\boldsymbol{K}$ of atomic models, ω-stable implies stable in κ for all κ.*

A fundamental result in model theory is Morley's proof that an $\aleph_1$-categorical first order theory is ω-stable. This argument depends on the compactness theorem in a number of ways. The key idea is to construct an Ehrenfeucht–Mostowski model over a well-order of cardinality $\aleph_1$. Such a model realizes only countably many types over any countable submodel. But the existence of the model depends on a compactness argument in the

[5] The difficulties we discuss here concern obtaining amalgamation. For simplicity, think only of gaining a monster model in λ with $\lambda^{<\lambda} = \lambda$. Weakening that hypothesis is a different project (see [**2, 14**] or any first order stability book for comments on the cardinality question).

proof of the Ehrenfeucht–Mostowski theorem. Further, this only contradicts ω-stability because amalgamation allows the construction from a model M_0 in $\aleph_0$ that has uncountably many types over it an elementary extension M_1 of M_0 with power $\aleph_1$ that realizes all of them. And again amalgamation in the first order case is a consequence of compactness. In $L_{\omega_1,\omega}$, the work of Keisler and Shelah evades the use of compactness —but at the cost of set theoretic hypotheses.

Theorem 3.12 (Keisler–Shelah) *Let $\boldsymbol{K}$ be the atomic model of a countable first order theory. If $\boldsymbol{K}$ is $\aleph_1$-categorical and $2^{\aleph_0} < 2^{\aleph_1}$ then $\boldsymbol{K}$ is ω-stable.*

This proof uses WCH directly and weak diamond via the 'one completely general result'. That is, from amalgamation failure of ω-stability yields a model of cardinality $\aleph_1$ that realizes uncountably many types from $S_{\mathrm{at}}(M)$ for a countable model M. Naming the elements of M yields a theory which has uncountably many types over the empty set. Thus by Theorem 3.6 the new theory has $2^{\aleph_1}$ models in $\aleph_1$ and (since $2^{\aleph_0} < 2^{\aleph_1}$) so does the original theory. Is CH is necessary?

Example 3.13 There are examples [**2, 37**] of an AEC $\boldsymbol{K}$ and even one given by a sentence of $L_{\omega_1,\omega}(Q)$ such that MA + ¬ CH imply that $\boldsymbol{K}$ is $\aleph_1$-categorical, but $\boldsymbol{K}$

(a) is not ω-stable;

(b) does not satisfy amalgamation even for countable models.

Laskowski (unpublished) showed that the example proposed for $L_{\omega_1,\omega}$ by Shelah [**34, 38**] fails. The question of whether such an $L_{\omega_1,\omega}$-example exists is a specific strategy for answering the next question.

Question 3.14 Is categoricity in $\aleph_1$ of a sentence of $L_{\omega_1,\omega}$ absolute (with respect to suitable forcings)?

By suitable, I mean that, e.g., it is natural to demand cardinal preserving. This result has resisted a number of attempts although, as we lay out in Section 5, many other fundamental notions of the model theory of $L_{\omega_1,\omega}$ are absolute.

4 The conjecture for $L_{\omega_1,\omega}$

Using the notion of splitting, a nice theory of independence (Definition 5.6) can be defined for ω-stable atomic classes [**2, 31, 32**]. This allows the formulation of the crucial notion of excellence and the proof of a version of Morley's theorem. We will not discuss the details but outline some aspects of the argument. These results are non-trivial but the exposition of the entire situation in [**2**] occupies less than 100 pages.

The concept of an independent system of models is hard to grasp although it is playing an increasing role in many areas of model theory. Rather than repeating the notation-heavy definition (see [**2, 21, 32**] or various first order stability texts), I give a simple example. Let X be a set of n algebraically independent elements in an algebraically closed field. For each $Y \subsetneq X$, let M_Y be the algebraic closure of Y. The M_Y form an independent system of $2^n - 1$-models. This is exactly the concept needed in Zilber's theory of quasiminimal excellence. For Shelah's more general approach the notion is axiomatized using the independence notion from the previous paragraph. In the example, there is clearly a prime model over the union of the independent system. In various more complicated algebraic examples (e.g., [**6**]) the existence of such a prime model is non-trivial. Here we discuss how to find one from model theoretic hypotheses.

Definition 4.1 An atomic class $\boldsymbol{K}$ is *excellent* if $\boldsymbol{K}$ is ω-stable and any of the following equivalent conditions hold.

For any finite independent system of countable models with union C,

(1) $S_{\mathrm{at}}(C)$ is countable;
(2) there is a unique primary model over C;
(3) the isolated types are dense in $S_{\mathrm{at}}(C)$.

The key point is that this is a condition of 'n-dimensional amalgamation'. A primary model is a particulary strong way of choosing a prime model over C. Thus, condition (2) specifies the existence of a strong kind of amalgamation of n independent models. This definition emphasizes the contrast of the current situation with first order logic; condition (1) does *not* follow from ω-stability. See [**2**] for details of the notation.

Note that excellence is a condition on countable models. It has the following consequence for models in *all* cardinalities. The key to this extension is the proof that n-dimensional amalgamation in $\aleph_n$ implies $(n-1)$-dimensional amalgamation in $\aleph_{n+1}$. Thus amalgamation for all n in $\aleph_0$ implies amalgamation for all n below $\aleph_\omega$ and then for all cardinals by a short argument.

Theorem 4.2 (Shelah, ZFC) *If an atomic class $\boldsymbol{K}$ is excellent and has an uncountable model, then*

(1) *$\boldsymbol{K}$ has models of arbitrarily large cardinality;*
(2) *categoricity in one uncountable power implies categoricity*[6] *in all uncountable powers.*

This result is in ZFC but extensions of set theory are used to obtain excellence. Recall that by VWGCH we mean the assertion $2^{\aleph_n} < 2^{\aleph_{n+1}}$ for $n < \omega$. The following is an immediate corollary of Theorem 4.6.

Theorem 4.3 (Shelah, VWGCH) *An atomic class $\boldsymbol{K}$ that is categorical in $\aleph_n$ for each $n < \omega$ is excellent.*

We remarked after Definition 3.9 on the difficulty of constructing a monster model for an atomic class associated with a sentence categorical in some power. Of course such a monster model in appropriate cardinalities is immediate from the amalgamation property. But, even assuming categoricity up to $\aleph_\omega$, we need to use the VWGCH to get excellence, then derive amalgamation and finally a monster model.

The requirement of categoricity below $\aleph_\omega$ in Theorem 4.3 is essential. Indeed, Baldwin and Kolesnikov [**3**] (refining [**13**]) show:

Theorem 4.4 *For each $2 \leq k < \omega$ there is an $L_{\omega_1,\omega}$-sentence ϕ_k such that:*

(1) *ϕ_k has an atomic model in every cardinal;*
(2) *ϕ_k is categorical in μ if $\mu \leq \aleph_{k-2}$;*
(3) *ϕ_k is not categorical in any μ with $\mu > \aleph_{k-2}$;*
(4) *ϕ_k has the (disjoint) amalgamation property.*

Note that of course the ϕ_k are not excellent. There is one further refinement on the 'wide' versus 'deep' metaphor. How wide?

Definition 4.5 We say that

(1) $\boldsymbol{K}$ has *few* models in power λ if $I(\boldsymbol{K},\lambda) < 2^\lambda$;

[6] In contrast to some authors, we say $\boldsymbol{K}$ is categorical in κ if there is *exactly* one model in cardinality κ.

(2) $\boldsymbol{K}$ has *very few* models in power $\aleph_n$ if $I(\boldsymbol{K}, \aleph_n) \leq 2^{\aleph_{n-1}}$.

These are equivalent under GCH. And Shelah argues on the last couple of pages of [**32**] (see also [**36**]) that they are equivalent under $\neg 0^+$. But in general we have a theorem and a conjecture [**31, 32**], which differ only in the word 'very'.

Theorem 4.6 (Shelah) (For $n < \omega$, $2^{\aleph_n} < 2^{\aleph_{n+1}}$) *An atomic class $\boldsymbol{K}$ that has at least one uncountable model and that has* very *few models in $\aleph_n$ for each $n < \omega$ is excellent.*

Conjecture 4.7 (Shelah) (For $n < \omega$, $2^{\aleph_n} < 2^{\aleph_{n+1}}$) An atomic class $\boldsymbol{K}$ that has at least one uncountable model and that has *few* models in $\aleph_n$ for each $n < \omega$ is excellent.

The proof of Theorem 4.6 uses the technology of atomic classes very heavily. But the calculation of the categoricity spectrum in Theorem 4.2(2) can be lifted to arbitrary sentences of $L_{\omega_1,\omega}$ by a calculation [**31, 32**], reported as Theorem 25.19 of [**2**].

5 Absoluteness of properties of atomic classes

As remarked in the introduction, one of the significant attributes of first order stability theory is that the basic notions: stable, ω-stable, superstable, $\aleph_1$-categoricity can be seen absolute in very strong ways. We sketch proofs of similar results, except the open $\aleph_1$-categoricity, for $L_{\omega_1,\omega}$. This section and the appendix tie together some results which are folklore with the use of well-known methods which are systematically applied to discuss the case of $L_{\omega_1,\omega}$. We are indebted for discussions with Alf Dolich, Paul Larson, Chris Laskowski, and Dave Marker for clarifying the issues. Among the few places model theoretic absoluteness issues have recently been addressed in print is [**35**]. Earlier accounts include [**5, 28**].

For example a first order theory T is unstable just if there is a formula $\phi(\mathbf{x}, \mathbf{y})$ such that, for every n,

$$T \models (\exists \mathbf{x}_1, \dots \mathbf{x}_n \exists \mathbf{y}_1, \dots \mathbf{y}_n) \bigwedge_{i<j} \phi(\mathbf{x}_i, \mathbf{y}_j) \wedge \bigwedge_{i \geq j} \neg\phi(\mathbf{x}_i, \mathbf{y}_j).$$

This is an arithmetic statement and so is absolute by basic properties of absoluteness [**16, 20**]. In first order logic, ω-stability is Π^1_1; there is no consistent tree[7]

$$\{\phi_i^{\sigma(i)}(x_\sigma, \boldsymbol{a}_\sigma \restriction n) : \sigma \in 2^\omega, i < \omega\}.$$

With a heavier use of effective descriptive set theory, suggested by David Marker, the same applies for the atomic class case.

To demonstrate absoluteness of various concepts of infinitary logic we need the full strength of the Shoenfield absoluteness lemma. In this section, we work with *atomic classes* (Definition 3.2). We noted Shelah's observation Theorem 3.7 that each $\aleph_1$-categorical sentence of $L_{\omega_1,\omega}$ determines such a class. In this section we first show absoluteness for various properties of atomic classes. In the last theorem, we show that the properties for sentences of $L_{\omega_1,\omega}$ remain absolute although in some cases they are more complex. The Appendix (written by David Marker) makes a precise definition of a formula in $L_{\omega_1,\omega}$ as a subset of $\omega^{<\omega}$ so that we can apply descriptive set theoretic techniques. It gives an effective analysis of the transformation in Theorem 3.3. The Appendix fixes some notation for the rest of the paper and clarifies the complexity of a number of basic notions; e.g., that the collection of complete sentences in $L_{\omega_1,\omega}$ is complete Π^1_1.

[7] We use the convention that $\phi^{\sigma(i)}\phi(x)$ denotes $\phi(x)$ or $\neg\phi(x)$ depending on whether $\sigma(i)$ is 0 or 1.

Theorem 5.1 (Shoenfield Absoluteness Lemma) *If*

(1) *$V \subset V'$ are models of ZF with the same ordinals, and*
(2) *ϕ is a lightface Π^1_2 predicate of a set of natural numbers,*

then for any $A \subset N$, $V \models \phi(A)$ iff $V' \models \phi(A)$.

Note that this trivially gives the same absoluteness results for Σ^1_2-predicates.

Lemma 5.2 (Atomic models)

(1) *'T has an atomic model' is an arithmetic property of T.*
(2) *'M is an atomic model of T' is an arithmetic property of M and T.*
(3) *For any vocabulary τ, the class M of countable atomic τ-structures is Borel.*

Proof. The first condition is given by: for every formula $\phi(\mathbf{x})$ there is a formula $\psi(\mathbf{x})$, consistent with T, such that $\psi(\mathbf{x}) \to \phi(\mathbf{x})$ and, for every $\chi(\mathbf{x})$, either $\psi(\mathbf{x}) \to \chi(\mathbf{x})$ or $\psi(\mathbf{x}) \to \neg\chi(\mathbf{x})$. Let $\theta(M, T)$ be the arithmetic predicate of the reals M, T asserting that T is the theory of M. The third condition is a Δ^1_1-predicate of M given by: there exists (for all) T such that $\theta(M, T)$ and, for every $\boldsymbol{a} \in M$, there exists a T-atom $\psi(\mathbf{x})$ such that $M \models \psi(\boldsymbol{a})$. $\square_{5.2}$

Earlier versions of this paper had weaker characterizations; e.g., a Σ^1_2 characterization of ω-stability and Π^1_2 characterization of excellence. Marker pointed out the application of Harrison's theorem, Fact 5.4(ii), to improve the result to Π^1_1.

Definition 5.3 We say that $x \in \omega^\omega$ is *hyperarithmetic* if $x \in \Delta^1_1$, and x is *hyperarithmetic in y*, written $x \leq_{\mathrm{hyp}} y$, if $x \in \Delta^1_1(y)$.

Fact 5.4

(i) The predicate $\{(x, y) : x \leq_{\mathrm{hyp}} y\}$ is Π^1_1.
(ii) If $K \subset \omega^\omega$ is Σ^1_1, then for any y, K contains an element which is not hyperarithmetic in y if and only if K contains a perfect set.

The unrelativized version of statement (i) is [**29**, II.1.4.ii]; the relativized version is [**25**, 7.15]. Again, the unrelativized version of statement (ii) is [**29**, III.6.2]; in this case the relativization is routine. $\square_{5.4}$

In the next theorem, the atomic set A must be regarded as an element of ω^ω. There are at least two ways to think of this: 1) a pair (M, A) where is M is a countable atomic model of T and A is a subset (automatically atomic) of M, or 2) as a pair (A, Φ) where Φ is the diagram of A as an atomic subset of the monster model $\mathbb{M}$.

Lemma 5.5 (Marker) *Let $\boldsymbol{K}$ be an atomic class (Definition 3.2) with a countable complete first order theory T.*

(1) *Let A be a countable atomic set. The predicate of p and A, 'p is in $S_{\mathrm{at}}(A)$', is arithmetic.*
(2) *'$S_{\mathrm{at}}(A)$ is countable' is a Π^1_1-predicate of A.*

Proof. (1) Note first that '$q(\mathbf{x})$ is a principal type over $\emptyset$ in T' is an arithmetic property. Now p is in $S_{\mathrm{at}}(A)$ if and only if for all $\boldsymbol{a} \in A$, $p \restriction \boldsymbol{a}$ is a principal type. So this is also arithmetic.

(2) By (1), the set of p such that 'p is in $S_{\mathrm{at}}(A)$' is arithmetic (*a fortiori* Σ^1_1) in A, so by Fact 5.4(ii), each such p is hyperarithmetic in A. Since the Continuum Hypothesis

holds for Σ^1_1-sets, '$S_{\mathrm{at}}(A)$ is countable' is formalized by:

$$(\forall p)[p \in S_{\mathrm{at}}(A) \rightarrow (p \leq_{\mathrm{hyp}} A)],$$

which is Π^1_1. $\square_{5.5}$

In order to show the absoluteness of excellence we need some more detail on the notion of independence. We will use item (1) of Definition 4.1. The independent families of models [**2, 32**] in that definition are indexed by subsets of n with strictly less than n elements; we denote this partial order by $\mathcal{P}^-(n)$. We will show that independence of models is an arithmetic property.

Definition 5.6

(1) A complete type p over A *splits* over $B \subset A$ if there are $\mathbf{b}, \mathbf{c} \in A$ which realize the same type over B and a formula $\phi(\mathbf{x}, \mathbf{y})$ with $\phi(\mathbf{x}, \mathbf{b}) \in p$ and $\neg\phi(\mathbf{x}, \mathbf{c}) \in p$.

(2) Let ABC be atomic. We write $A \underset{C}{\perp} B$ and say A is *free* or *independent* from B over C if for any finite sequence $\boldsymbol{a}$ from A, $\mathrm{tp}(\boldsymbol{a}/B)$ does not split over some finite subset of C.

Lemma 5.7 *Let T be a complete countable first order theory. The properties that the class of atomic models of T is*

(1) *ω-stable;*

(2) *excellent;*

are each given by a Π^1_1-formula of set theory and so are absolute.

Proof. (1) The class of atomic models of T is ω-stable if and only if for every atomic model M, '$S_{\mathrm{at}}(M)$ is countable'. This property is Π^1_1 by Lemma 5.5.

(2) The class of atomic models of T is excellent if and only if for any finite set of countable atomic models $\{A_s : s \in \mathcal{P}^-(n)\}$ that form an independent system, with $A = \bigcup\{A_s : s \in \mathcal{P}^-(n)\}$, $S_{\mathrm{at}}(A)$ is countable. Here we have universal quantifiers over finite sequences of models (using a pairing function, this is quantifying over a single real). The stipulation that the diagram is independent requires repeated use of the statement $A \underset{C}{\perp} B$, where A, B, C are finite unions of the models in the independent system. This requires quantification over finite sequences from the A_s; thus, it is arithmetic. The assertion '$S_{\mathrm{at}}(A)$ is countable' is again π^1_1 by Lemma 5.5 and we finish. $\square_{5.7}$

Lemma 5.8 *The property that an atomic class $\boldsymbol{K}$ has arbitrarily large models is absolute. In fact it is Σ^1_1.*

Proof. Let $\boldsymbol{K}$ be the class of atomic models of a first order theory T in a vocabulary τ. $\boldsymbol{K}$ has arbitrarily large models if and only if there are $\hat{T}$, $\hat{\tau}$, M and C such that $\hat{T}$ is a Skolemization of T in a vocabulary $\hat{\tau}$ and M is a countable model of $\hat{T}$ such that $M \restriction \tau$ is atomic and M contains an infinite set C of $\hat{\tau}$-indiscernibles. This formula is Σ^1_1. $\square_{5.8}$

Finally, following Lessmann [**2, 21**], we prove that the absolute 'Baldwin–Lachlan' characterization of first order $\aleph_1$-categoricity has a natural translation to the $L_{\omega_1,\omega}$ situation; the resulting property of atomic classes is absolute and in ZFC it implies $\aleph_1$-categoricity. But we do not see how to derive it from $\aleph_1$-categoricity without using the Continuum Hypothesis. We need some definitions. To be a bit more specific we speak of Vaughtian triples instead of Vaughtian pairs.

Definition 5.9 The formula $\phi(x, \mathbf{c})$ with $\mathbf{c} \in M \in \boldsymbol{K}$ is *big* if for any $M' \supseteq A$ with $M' \in \boldsymbol{K}$ there exists an N' with $M' \prec_{\boldsymbol{K}} N'$ and with a realization of $\phi(x, \mathbf{c})$ in $N' - M'$.

This definition has no requirements on the cardinality of M, M', N' so it is saying that $\phi(\mathbf{x}, \mathbf{c})$ has as many solutions as the size of the largest models in $\boldsymbol{K}$. This condition is equivalent to one on countable models. A translation of Lemma 25.2 of [**2**] gives:

Lemma 5.10 *Let $A \subseteq M$ and $\phi(x, \mathbf{c})$ be over A. The following are equivalent:*

(1) *There is an N with $M \prec N$ and $c \in N - M$ satisfying $\phi(x, \mathbf{c})$;*
(2) *$\phi(x, \mathbf{c})$ is big.*

The significance of this remark is that it makes '$\phi(x, \mathbf{c})$ is big' a Σ^1_1 predicate.

Definition 5.11

(1) A triple (M, N, ϕ) where $M \prec N \in \boldsymbol{K}$ with $M \neq N$, ϕ defined over M, ϕ big, and $\phi(M) = \phi(N)$ is called a *Vaughtian triple*.
(2) We say $\boldsymbol{K}$ admits (κ, λ), witnessed by ϕ, if there is a model $N \in \boldsymbol{K}$ with $|N| = \kappa$ and $|\phi(N)| = \lambda$ and ϕ is big.

Now we have the partial characterization.

Lemma 5.12 *Let $\boldsymbol{K}$ be a class of atomic models. If $\boldsymbol{K}$ is ω-stable and has no Vaughtian triples then $\boldsymbol{K}$ is $\aleph_1$-categorical. The hypothesis of this statement is Π^1_1.*

Proof. The sufficiency of the condition is found by tracing results in [**2**]: ω-stability gives the existence of a quasiminimal formula ϕ. Note from the proof of Theorem 24.1 in [**2**] that ω-stability is sufficient to show that there are prime models over independent subsets of cardinality $\aleph_1$. (The point of excellence is that higher dimensional amalgamation is needed to extend this result to larger sets.) So if $|M| = \aleph_1$, there is an $N \prec_{\boldsymbol{K}} M$ which is prime over a basis for $\phi(M)$. As noted in [**2**, Chapter 2], this determines N up to isomorphism (again without use of excellence because we are in $\aleph_1$). So we are done unless $N \precnsim M$. But then Löwenheim–Skolem gives us a countable Vaughtian triple, contrary to the hypothesis. $\square_{5.12}$

Since the second condition below is true if $2^{\aleph_0} < 2^{\aleph_1}$ and we have shown that the conclusion of this condition is absolute, we have:

Corollary 5.13 *$\aleph_1$-categoricity is absolute for atomic classes if and only if in ZFC $\aleph_1$-categoricity implies countable amalgamation and ω-stabity.*

Consequence 5.14 *Let $\boldsymbol{K}$ be a class of atomic models. Then $\aleph_1$-categoricity of $\boldsymbol{K}$ is absolute between models of set theory that satisfy either of the following conditions:*

(1) *$\boldsymbol{K}$ has arbitrarily large members and $\boldsymbol{K}$ has amalgamation in $\aleph_0$, or*
(2) $2^{\aleph_0} < 2^{\aleph_1}$.

Proof. Each hypothesis implies the characterization in Lemma 5.12. $\square_{5.14}$

Note the hypothesis of condition (1) is absolute. It seems unlikely that $\aleph_1$-categoricity implies the existence of arbitrarily large models in $\boldsymbol{K}$; but no counterexample has yet been constructed. The use of the Continuum Hypothesis is central to current proofs that $\aleph_1$-categoricity implies amalgamation and ω-stability. For general AEC, Example 3.13 shows ZFC does not imply the assertion (A): $\aleph_1$-categoricity implies amalgamation in $\aleph_0$ and ω-stability. But [**9**] have shown (employing standard forcings) that for each AEC

$\boldsymbol{K}$ that fails amalgamation in $\aleph_0$, there is a model of set theory such that in that model $2^{\aleph_0} = 2^{\aleph_1}$, $\boldsymbol{K}$ continues to fail amalgamation in $\aleph_0$, and $\boldsymbol{K}$ has $2^{\aleph_1}$ models in $\aleph_1$. So assertion (A) does not imply CH.

Consequence 5.15 *Let $\boldsymbol{K}$ be a class of atomic models. Categoricity in all cardinals is absolute between models of set theory that satisfy the VWGCH.*

Proof. Under VWGCH, categoricity in all powers is equivalent to the Π^1_1-condition: excellence with no two cardinal models. $\square_{5.15}$

Theorem 5.16 *Each of the properties that a complete sentence of $L_{\omega_1,\omega}$ is ω-stable, excellent, or has no two-cardinal models is Σ^1_2.*

Proof. Let $Q(T)$ denote any of the conditions above as a property of the first order theory T in a vocabulary τ^*. Now write the following properties of the complete sentence ϕ in vocabulary τ:

(1) ϕ is a complete sentence.
(2) There exists a $\tau^* \supseteq \tau$ and τ^* theory T satisfying the following:
  (a) T is a complete theory that has an atomic model.
  (b) The reduct to τ of any atomic model of T satisfies ϕ.
  (c) There is a model M of ϕ and there exists an expansion of M to an atomic model of T.
  (d) $Q(T)$.

Proof. We know that condition (1) is Π^1_1. Condition (2) is an existential function quantifier followed by conditions which are at worst Π^1_1. $\square_{5.16}$

So, as far as we know the conditions on sentences of $L_{\omega_1,\omega}$ are more complicated than those for atomic classes and the application of Harrison's lemma[8] was needed to obtain absoluteness of these conditions for sentences of $L_{\omega_1,\omega}$.

6 Complexity

We prove the following claim. This result was developed in conversation with Martin Koerwien and Sy Friedman at the CRM Barcelona and benefitted from further discussion with Dave Marker.

Claim 6.1 *The class of countable models whose automorphism groups admit a complete left invariant metric is Π^1_1 but not Σ^1_1.*

Our proof is by propositional logic from known results of Gao [**10**] and Deissler [**8**].

Definition 6.2 A countable model is *minimal* (equivalently *non-extendible*) if it has no proper $L_{\omega_1,\omega}$-elementary submodel.

We showed in Lemma 5.2 that the class of atomic structures is Borel. The following claim is an easy back and forth.

Claim 6.3 *If M is atomic, τ-elementary submodel is the same as $L_{\omega_1,\omega}(\tau)$-elementary submodel.*

[8] Grossberg has pointed out that by suitably modifying the rank for ω-stable atomic classes the result could be given a direct model theoretic proof. This is slightly tricky because this rank will only be defined on some atomic sets.

Claim 6.3 shows that an atomic model is minimal iff it is minimal in first order logic. Note that the class of first order minimal models is obviously Π^1_1. Now if the class of minimal models were Borel, it would follow that the class of minimal atomic (equal first order minimal prime) models is also Borel. But Corollary 2.6 of Deissler [**8**] asserts for first order theories:

Lemma 6.4 (Deissler) *There is a countable relational vocabulary τ such that the class of minimal prime models for τ is not Σ^1_1.*

Gao [**10**] characterized non-extendible models in terms of metrics on their automorphism group.

Lemma 6.5 (Gao) *The following are equivalent:*

(1) $\mathrm{Aut}(M)$ *admits a compatible left-invariant complete metric.*
(2) *There is no $L_{\omega_1,\omega}$-elementary embedding from M into itself which is not onto.*

So we can transfer to the characterization of automorphism groups and prove Claim 6.1.

Gao pointed out to me that Malicki [**23**] recently proved a related result: the class of Polish groups with a complete left invariant metric is $\boldsymbol{\Pi}^1_1$ but not $\boldsymbol{\Sigma}^1_1$. We now analyze the connection between the two results and show that the properties studied are Borel equivalent. This observation was made jointly with Christian Rosendal.

Recall that S_∞ is is a Borel subspace of N^N. We denote by $\mathbb{SG}(S_\infty)$ the collection of closed subgroups of S_∞. It is contained in $\mathbb{F}$, the hyperspace of closed subsets of S_∞. $\mathbb{F}$ is a standard Borel space with the Effros–Borel structure generated by

$$\{F \in \mathbb{F} : F \cap U \neq \emptyset\}$$

for some open $U \subset S_\infty$. Proposition 1 of [**23**] implies that with this topology $\mathbb{SG}(S_\infty)$ is a standard Borel space.

Claim 6.6 *The map A taking M to* $\mathrm{Aut}(M)$ *mapping the standard Borel space of countable atomic models models into* $\mathbb{SG}(S_\infty)$ *is Borel.*

Proof. We have to show that for any basic open set $X \in \mathbb{SG}(S_\infty)$, $A^{-1}(X)$ is a Borel subset of $\mathcal{A}$. That is, for fixed open U, if X is the set of F with $F \cap U \neq \emptyset$, the inverse image of X is Borel in the space of atomic models.

Say U is all permutations mapping $\boldsymbol{a}$ to $\mathbf{b}$ where $\boldsymbol{a}, \mathbf{b} \in N^n$. Now there is $g \in \mathrm{Aut}(M)$ mapping $\boldsymbol{a}$ to $\mathbf{b}$ if and only if $\boldsymbol{a}$ and $\mathbf{b}$ realize the same type in M if and only if they satisfy the same formulas over the empty set, which is a Borel condition. $\square_{6.6}$

Corollary 6.7 *The class of Polish groups with a complete left invariant metric is Π^1_1 but not Σ^1_1.*

Conversely, we want to reduce the CLI groups to the class of minimal atomic models. The reduction is a map B from a group G acting on N to a structure M on N with $\mathrm{Aut}(M) = G$. This is easily done by mapping G to a structure with universe N which has a predicate for each orbit of G on N^n.

Deissler also uses a vocabulary with infinitely many n-ary predicates for each n so the vocabulary is in fact the same for both directions of reduction.

7 Conclusion

The spectrum problem for first order theories motivated many technical developments that eventually had significant algebraic consequences. A similar possibility for application of infinitary logic to algebraic problems is suggested by Zilber's program [**40, 41**]. But the basic development is far more difficult and less advanced. The notion of excellence provides one useful context. And others are being developed under the guise of abstract elementary classes and metric abstract elementary classes. But while first order stability theory is developed in ZFC, the current development of the model theory of $L_{\omega_1,\omega}$ uses a (rather weak) extension of set theory: the VWGCH. This raises both model theoretic and set theoretic questions. The proof of the 'one completely general result', Theorem 2.4, is a fundamentally combinatorial argument using no sophisticated model theoretic lemmas. The current proof uses $2^\lambda < 2^{\lambda^+}$. Can this hypothesis be removed?

Like first order logic such fundamental definitions of $L_{\omega_1,\omega}$ as satisfaction, ω-stability, and excellence are absolute. And in fact the complexity of their description can often be computed. But while $\aleph_1$-categoricity is seen (by a model theoretic argument) to be absolute in the first order case, this issue remains open for $L_{\omega_1,\omega}$.

We have also investigated the complexity of various properties of $L_{\omega_1,\omega}$-sentences and associated atomic classes. It is shown in Lemma 8.7 that the graph of the translation from a sentence to a finite diagram (T,Γ) is arithmetic. In Theorem 5.16, we avoided a precise calculation of the translation from a complete sentence to the atomic models of a first order theory. The tools of the appendix should allow a careful computation of this complexity. Note that while, for example, we showed that ω-stability was Π^1_1 as a property of an atomic class, we only showed it to be Σ^1_2 as a property of the $L_{\omega_1,\omega}$-sentence.

8 Appendix: Basic definability notions for $L_{\omega_1,\omega}$

by David Marker

Fix a vocabulary τ and let $\mathbb{X}_\tau$ be the Polish space of countable τ-structures with universe ω. Our first goal is to describe the collection of codes for $L_{\omega_1,\omega}(\tau)$-formulas. This is analogous to the construction of Borel codes in descriptive set theory.

Definition 8.1

(1) A *labeled tree* is a non-empty tree $T \subseteq \omega^{<\omega}$ with functions l and v with domain T such that for any $\sigma \in T$ one of the following holds:
- σ is a terminal node of T then $l(\sigma) = \psi$ where ψ is an atomic τ-formula and $v(\sigma)$ is the set of free variables in ψ;
- $l(\sigma) = \neg$, $\sigma\hat{\ }0$ is the unique successor of σ in T and $v(\sigma) = v(\sigma\hat{\ }0)$;
- $l(\sigma) = \exists v_i$, $\sigma\hat{\ }0$ is the unique successor of σ in T and $v(\sigma) = v(\sigma\hat{\ }0) \setminus \{i\}$;
- $l(\sigma) = \bigwedge$ and $v(\sigma) = \bigcup_{\sigma\hat{\ }i\in T} v(\sigma\hat{\ }i)$ is finite.

(2) A *formula* ϕ is a well founded labeled tree (T,l,v). A *sentence* is a formula where $v(\emptyset) = \emptyset$.

Proposition 8.2 *The set of labeled trees is arithmetic and the set of formulas is complete-*Π^1_1*, as is the set of sentences.*

Now it is easy to see:

Proposition 8.3 *There is $R(x,y) \in \Pi^1_1$ and $S(x,y) \in \Sigma^1_1$ such that if ϕ is a sentence and $M \in \mathbb{X}_\tau$, then*

$$M \models \phi \iff R(M,\phi) \iff S(M,\phi).$$

In particular, $\{(M,\phi) : \phi$ is a sentence and $M \models \phi\}$ is Π^1_1. However, for any fixed ϕ, $\mathrm{Mod}(\phi) = \{M \in \mathbb{X}_\tau : M \models \phi\}$ is Borel, indeed $\Delta^1_1(\phi)$.

Proof. We define a predicate 'f is a *truth definition* for the labeled tree (T,l,v) in M' as follows:

- The domain of f is the set of pairs (σ,μ) where $\sigma \in T$ and $\mu\colon v(\sigma) \to M$ is an assignment of the free variables at node σ and $f(\sigma,\mu) \in \{0,1\}$.
- If $l(\sigma) = \psi$ an atomic formula, then $f(\sigma,\mu) = 1$ if and only if ψ is true in M when we use μ to assign the free variables.
- If $l(\sigma) = \neg$, then $f(\sigma,\mu) = 1$ if and only if $f(\sigma\,\hat{}\,0,\mu) = 0$.
- If $l(\sigma) = \exists v_i$ there are two cases. If $v_i \in v(\sigma\,\hat{}\,0)$, then $f(\sigma,\mu) = 1$ if and only if there is $a \in M$ such that $f(\sigma\,\hat{}\,0,\mu^*) = 1$, where $\mu^* \supset \mu$ is the assignment where $\mu^*(v_i) = a$. Otherwise, $f(\sigma,\mu) = f(\sigma\,\hat{}\,0,\mu)$.
- If $l(\sigma) = \bigwedge$, then $f(\sigma,\mu) = 1$ if and only if $f(\sigma\,\hat{}\,i,\mu|v)(\sigma\,\hat{}\,i) = 1$ for all i such that $\sigma\,\hat{}\,i \in T$.

This predicate is arithmetic. If ϕ is a sentence, there is a unique truth definition f for ϕ in M. Let $R(x,y) \Leftrightarrow x \in \mathbb{X}_\tau$ and y is a labeled tree and $f(\emptyset,\emptyset) = 1$ for all truth definitions f for y in x, and let $S(x,y) \Leftrightarrow y$ is a labeled tree and there is a truth definition f for y in x such that $f(\emptyset,\emptyset) = 1$. $\square_{8.3}$

Notation 8.4 We write that a property of a set of reals is $\Pi^1_1 \wedge \Sigma^1_1$ if it is defined by the conjunction of a Π^1_1 and a Σ^1_1 formula.

Proposition 8.5 *$\{\phi : \phi$ is a satisfiable sentence$\}$ is $\Pi^1_1 \wedge \Sigma^1_1$, but neither $\mathbf{\Pi}^1_1$ nor $\mathbf{\Sigma}^1_1$.*

Proof. 'ϕ is a sentence' is Π^1_1; 'there is a model for ϕ' is equivalent to $\exists x\ S(x,\phi)$, which is Σ^1_1. The set of satisfiable sentences is not $\mathbf{\Sigma}^1_1$ since otherwise the set of underlying trees would be a $\mathbf{\Sigma}^1_1$-set of trees and there would be a countable bound (e.g., [**24**, Theorem 3.12]), on their heights.

We show that the set of satisfiable sentences is not $\mathbf{\Pi}^1_1$ by constructing a reduction of non-well ordered linear orders to satisfiable sentences.

Let $\tau = \{U, V, <, s, f, 0, c_n : n \in \omega\}$. For each linear order $\prec$ of ω we write down an $L_{\omega_1,\omega}$ sentence $\phi_\prec$ asserting:

- the universe is the disjoint union of U and V;
- $U = \{c_0, c_1, \dots\}$ all of which are distinct;
- $<$ is a linear order of U;
- $c_n < c_m$, if $n \prec m$;
- s is a successor function on V and $V = \{0, s(0), s(s(0)), \dots\}$;
- $f : V \to U$ and $f(s(n)) < f(n)$ for all n.

It $\prec$ is not a well order, and $n_0 \succ n_1 \succ \dots$ is an infinite descending chain, then by defining $f(n) = c_n$ we get a model of $\phi_\prec$. On the other hand if $\prec$ is a well order we can find no model of $\phi_\prec$.

Thus $\prec \mapsto \phi_\prec$ is a reduction of non-well-ordered linear orders to $\{\phi : \phi$ is satisfiable$\}$ which is impossible if satisfiability is $\mathbf{\Pi}^1_1$. $\square_{8.5}$

We now effectivize Chang's observation (Lemma 3.3) that for each sentence ϕ in $L_{\omega_1,\omega}$ we can find a first order theory T^* in a vocabulary τ^* and a countable set Γ of partial τ^*-types such that the models of ϕ are exactly the τ-reducts of models of T^* that omit all the types in Γ.

Definition 8.6 A *Chang-assignment* to a labeled tree (T,l,v) is a pair of functions S,γ with domain T such that $S(\sigma)$ is a set of sentences in the vocabulary $\tau_\sigma = \tau \cup \{R_\tau : \tau \supseteq \sigma\}$, where τ and σ are in T and R_τ is a relation symbol in $|v(\tau)|$-variables and $\gamma(\sigma)$ is a function with domain ω such that each $\gamma(\sigma)(n)$ is a partial τ_σ type.[9] We also require:

- if $l(\sigma) = \psi$ is atomic, $S(\sigma) = \{\forall \overline{v}(\sigma)(R_\sigma(\overline{v}) \leftrightarrow \psi\}$, and each $\gamma(\sigma)(i) = \{v_1 \neq v_1\}$;
- if $l(\sigma) = \neg$, then $S(\sigma) = S(\sigma\,\hat{}\,0) \cup \{\forall \overline{v}(\sigma) R_\sigma \leftrightarrow Neg R_{\sigma\,\hat{}\,0}\}$ and $\gamma(\sigma) = \gamma(\sigma\,\hat{}\,0)$;
- if $l(\sigma) = \exists v_i$, then $S(\sigma) = S(\sigma\,\hat{}\,0) \cup \{\forall \overline{v}(\sigma) R_\sigma \leftrightarrow \exists v_i R_{\sigma\,\hat{}\,0}\}$ and $\gamma(\sigma) = \gamma(\sigma\,\hat{}\,0)$;
- if $l(\sigma) = \bigwedge$; then $S(\sigma) = \bigcup_{\sigma\,\hat{}\,i \in T} S(\sigma\,\hat{}\,i) \cup \{\forall \overline{v}(\sigma)(R_\sigma \to R_{\sigma\,\hat{}\,i}) : \sigma\,\hat{}\,i \in T\}$.

Fix a pairing function $\mu \colon \omega \times \omega \to \omega$. Let

$$\gamma(\sigma)(0) = \{R_\sigma, \neg R_{\sigma\,\hat{}\,i} : \sigma\,\hat{}\,i \in T\}$$

and

$$\gamma(\sigma)(\mu(i,n)+1) = \begin{cases} \gamma(\sigma\,\hat{}\,i)(n) & \text{if } \sigma\,\hat{}\,i \in T \\ \{v_1 \neq v_1\} & \text{otherwise.} \end{cases}$$

In other words, $\gamma(\sigma)$ lists all the types listed by the successors of σ and the additional type $\{R_\sigma, \neg R_{\sigma\,\hat{}\,i} : \sigma\,\hat{}\,i \in T\}$.

It is now easy to see:

Lemma 8.7 *The predicate "(S,γ) is a Chang-assignment for the labeled tree (T,l,v)" is arithmetic. If ϕ is a sentence then there is a unique Chang-assignment for ϕ.*

To simplify notation we will call (T,Γ) the Chang-assignment where T is the theory $S(\emptyset)$ and Γ is the set of types $\gamma(\emptyset)(0), \gamma(\emptyset)(1), \ldots$.

The following remark is implicit in [**11**].

Lemma 8.8 *The property that a sentence ϕ of $L_{\omega_1,\omega}$ has arbitrarily large models is absolute. In fact it is $\Pi^1_1 \wedge \Sigma^1_1$, but neither $\boldsymbol{\Pi}^1_1$ nor $\boldsymbol{\Sigma}^1_1$.*

Proof. A τ-sentence ϕ has arbitrarily large models if and only if there is a Chang-assignment (T,Γ), $\tau^* \supseteq \tau$ and $T^* \supseteq T$ a Skolemized τ^*-theory such that there is a model of T^* omitting all types in Γ and containing an infinite set of τ^*-indiscernibles. This condition is Σ^1_1 once we restrict to the Π^1_1-set of sentences.

For any sentence ϕ let ϕ^* be the sentence which asserts we have two sorts, the first of which is a model of ϕ and the second is an infinite set with no structure. Then ϕ is satisfiable if and only if ϕ^* has arbitrarily large models. Thus $\phi \mapsto \phi^*$ is a reduction of satisfiable sentences to sentences with arbitrarily large models. By Proposition 8.5, the set of sentences with arbitrarily large models is neither $\boldsymbol{\Sigma}^1_1$ nor $\boldsymbol{\Pi}^1_1$. $\square_{8.8}$

Recall that an $L_{\omega_1,\omega}$-sentence is *complete* if and only if it is satisfiable and any two countable models are isomorphic. This is easily seen to be Π^1_2. Drawing on some results of Nadel, we show that in fact:

Theorem 8.9 $\{\phi : \phi$ *is a complete sentence*$\}$ *is complete*-Π^1_1.

[9] We allow relation symbols in 0 variables, but these could easily be eliminated.

The argument requires some preparation. We begin by recalling the usual Karp–Scott back-and-forth analysis.

Definition 8.10 If M and N are τ-structures, we inductively define $\sim_\alpha$ by: $(M, \boldsymbol{a}) \sim_0 (N, \mathbf{b})$ if $M \models \phi(\boldsymbol{a})$ if and only if $N \models \phi(\mathbf{b})$ for all atomic τ-formulas ϕ. For all ordinals α, $(M, \boldsymbol{a}) \sim_{\alpha+1} (N, \mathbf{b})$ if for all $c \in M$ there is $d \in N$ such that $(M, \boldsymbol{a}, c) \sim_\alpha (N, \mathbf{b}, d)$ and for all $d \in N$ there is $c \in M$ such that $(M, \boldsymbol{a}, c) \sim_\alpha (N, \mathbf{b}, d)$. For all limit ordinals β, $(M, \boldsymbol{a}) \sim_\beta (N, \mathbf{b})$ if and only if $(M, \boldsymbol{a}) \sim_\alpha (N, \mathbf{b})$ for all $\alpha < \beta$.

A classical fact is that $(M, \boldsymbol{a}) \sim_\alpha (N, \mathbf{b})$ if and only if $M \models \phi(\boldsymbol{a}) \Leftrightarrow N \models \phi(\mathbf{b})$ for all formulas ϕ of quantifier rank at most α.

We say that ϕ has *Scott rank* α if α is the least ordinal such that if $M, N \models \phi$ and $(M, \boldsymbol{a}) \sim_\alpha (N, \mathbf{b})$ then $(M, \boldsymbol{a}) \sim_\beta (N, \mathbf{b})$ for all ordinals β.

We need to analyze the complexity of $\sim_\alpha$.

Definition 8.11 Let WO* (the class of pseudo-well-orders) be the set of all linear orders R with domain ω such that:

(i) 0 is the R-least element;

(ii) if n is not R-maximal, then there is y such that xRy and there is no z such that xRz and zRx, we say y is the R-successor of x and write $y = s_R(x)$.

If $n \neq 0$ is not an R-successor, we say it is an R-limit.

Note that WO*, $s_R(n) = m$ and 'n is an R-limit' are arithmetic.

Definition 8.12 We say that z is an *R-analysis of M and N* if

(i) $z \subseteq \omega \times \bigcup_{n\in\omega}(\omega^n \times \omega^n)$;

(ii) $(0, \boldsymbol{a}, \mathbf{b}) \in z$ if and only if $M \models \phi(\boldsymbol{a}) \leftrightarrow N \models \phi(\mathbf{b})$ for all quantifier free ϕ;

(iii) if $(n, \boldsymbol{a}, \mathbf{b})$ and mRn, then $(m, \boldsymbol{a}, \mathbf{b})$;

(iv) $(s_R(n), \boldsymbol{a}, \mathbf{b}) \in z$ if and only if for all $c \in \omega$ there is $d \in \omega$ such that $(n, \boldsymbol{a}\hat{\ } c, \mathbf{b}\hat{\ } d) \in z$ and for all $d \in \omega$ there is $c \in \omega$ such that $(n, \boldsymbol{a}\hat{\ } c, \mathbf{b}\hat{\ } d) \in z$;

(v) if n is an R-limit, then $(n, \boldsymbol{a}, \mathbf{b}) \in z$ if and only if $(m, \boldsymbol{a}, \mathbf{b}) \in z$ for all mRn.

Note:

- '$\{(z, R, M, N)$: 'z is an R-analysis'$\}$ is arithmetic.
- Suppose R is a well-order of order type α. Let $\beta(n) < \alpha$ be the order type of $\{m : mRn\}$. If z is an R-analysis of M, N, then

$$(n, \boldsymbol{a}, \mathbf{b}) \in z \text{ if and only if } (M, \boldsymbol{a}) \sim_{\beta(n)} (N, \mathbf{b}).$$

  In particular, there is a unique R-analysis of M, N.

We need two results of Mark Nadel.

Theorem 8.13 (Nadel)

(a) *If ϕ is complete, then there is $M \models \phi$ with $M \leq_{\text{hyp}} \phi$.*

(b) *If ϕ is complete then the Scott rank of ϕ is at most* $\operatorname{qr}(\phi) + \omega$ *where* $\operatorname{qr}(\phi)$ *is the quantifier rank of ϕ.*

Here (a) is [**27**, Theorem 2], while (b) is [**26**, Theorem 5.1]. For completeness we sketch the proofs.

For (a), add new constants $c_1, c_2, \ldots$ to τ. Let F be a countable fragment such that $\phi \in F$ —we can choose F arithmetic in ϕ. Let $S = \{s : s$ a finite set of F-sentences using only finitely many c_i such that $\phi \models \exists \overline{v} \bigwedge_{\psi\in s} \psi(\overline{v})\}$. S is a consistency property. Since ϕ is complete,

$$\phi \models \theta \Leftrightarrow \forall M \; (M \models \phi \to M \models \theta) \Leftrightarrow \exists M(M \models \phi \wedge M \models \theta).$$

It follows that S is $\Delta^1_1(\phi, F)$ and hence $S \leq_{\text{hyp}} \phi$. Using the consistency property S, one can easily construct $M \models \phi$ with $M \leq_{\text{hyp}} \phi$.

Towards (b), let F be as above. Since ϕ is complete, there are only countably many F-types. By the Omitting Types Theorem for $L_{\omega_1,\omega}$, there is a model of ϕ where every element satisfies an F-complete formula. Since ϕ is complete, this is true in the unique countable model M.

The usual arguments show that we can do a back and forth in M with F-types. Thus if $\boldsymbol{a}, \mathbf{b}$ in M and $(M, \boldsymbol{a}) \equiv_F (M, \mathbf{b})$ then there is an automorphism of M mapping $\boldsymbol{a}$ to $\mathbf{b}$. If we pick α such that every ψ is F has quantifier rank below α and $(M, \boldsymbol{a}) \sim_\alpha (M, \mathbf{b})$, then $(M, \boldsymbol{a}) \sim_\beta (M, \mathbf{b})$ for all β. Thus the Scott rank of ϕ is at most α.

If F is the smallest fragment containing ϕ, every formula in F has Scott rank below $\text{qr}(\phi) + \omega$, so this is an upper bound on the Scott rank. $\square_{8.13}$

Proof of Theorem 8.9. First note that if α is a bound on the Scott rank of models of ϕ, then any two countable models M and N of ϕ are isomorphic if and only if we can do a back-and forth construction using $\sim_\alpha$. Thus by Nadel's Theorems, a sentence ϕ is complete if and only if

(i) $(\exists M) M \leq_{\text{hyp}} \phi \wedge M \models \phi$, and

(ii) $\exists \alpha$ recursive in ϕ such that for all $M, N \models \phi$ if $\boldsymbol{a} \in M, \mathbf{b} \in N$ and $(M, \boldsymbol{a}) \sim_\alpha (N, \mathbf{b})$, then for all $c \in M$ there is $d \in \overline{N}$ such that $(M, \boldsymbol{a}, c) \sim_\alpha (N, \mathbf{b}, d)$.

Here (i) is easily seen to be Π^1_1, using Fact 5.4, while (ii) is equivalent to $\forall M, N \models \phi$ $(\exists R, \exists z) z \leq_{\text{hyp}} \phi$, $R \in WO^*$ and z is an R-analysis of M and N and there is an n such that if $\boldsymbol{a}, c \in M, \mathbf{b} \in N$ with $(n, \boldsymbol{a}, \mathbf{b}) \in z$, then there is $d \in N$ such that $(n, \boldsymbol{a}, c, \mathbf{b}, d) \in z$. This is also Π^1_1, again using Fact 5.4.

Finally, to each linear order $\prec$ of ω we will assign an $L_{\omega_1,\omega}$ sentence $\phi_\prec$ such that $\prec$ is a well order if and only if $\phi_\prec$ is complete. This will show that $\{\phi : \phi \text{ is complete}\}$ is $\boldsymbol{\Pi}^1_1$-complete.

The vocabulary τ is $\{P_n : n \in \omega\}$ where P_n is a unary predicate.

- We say that every element is in some P_n.
- We say that each P_n is infinite and that if $n \prec m$, then $P_n \subset P_m$ and $P_m \setminus P_n$ is infinite.
- Moreover if $\forall m \prec n \exists k \; m \prec k \prec n$, then we also say that $P_n \setminus \bigcup_{m \prec n} P_m$ is infinite.

If $\prec$ is a well ordering, then $\phi_\prec$ is $\aleph_0$-categorical as for each n we just put $\aleph_0$ elements in each $P_n \setminus \bigcup_{m \prec n} P_m$. On the other hand if $n_0 \succ n_1 \succ \ldots$ is an infinite descending chain, let $X = \{m : m \prec n_i \text{ for all } i\}$. We can put any number of elements in

$$\bigcap_{i=0}^{\infty} P_{n_i} \setminus \bigcup_{m \in X} P_m,$$

so $\phi_\prec$ is not complete. $\square_{8.9}$

References

[1] J. T. Baldwin. The Vaught conjecture: Do uncountable models count? *Notre Dame Journal of Formal Logic*, 48(1):79–92, 2007.

[2] J. T. Baldwin. *Categoricity.* No. 51 in University Lecture Notes. American Mathematical Society, 2009. `www.math.uic.edu/\~\jbaldwin`.

[3] J. T. Baldwin and A. Kolesnikov. Categoricity, amalgamation, and tameness. *Israel Journal of Mathematics*, 170, 2009. `www.math.uic.edu/\~\jbaldwin`.

[4] J. T. Baldwin, A. Kolesnikov, and S. Shelah. The amalgamation spectrum. *Journal of Symbolic Logic*, 74:914–928, 2009.

[5] J. Barwise (ed.). *Admissible Sets and Structures.* Perspectives in Mathematical Logic. Springer-Verlag, 1975.

[6] M. Bays and B. I. Zilber. Covers of multiplicative groups of an algebraically closed field of arbitrary characteristic. arXiv math.AC/0401301, 2004.

[7] C. C. Chang and H. J. Keisler. *Model Theory.* North-Holland, 1973. 3rd edition, 1990.

[8] R. Deissler. Minimal models. *Journal of Symbolic Logic*, 42:254–260, 1977.

[9] S.-D. Friedman and M. Koerwien. On absoluteness of categoricity in abstract elementary classes. *Notre Dame Journal of Formal Logic*, 52(4):395–402, 2011.

[10] S. Gao. On automorphism groups of countable structures. *Journal of Symbolic Logic*, 63:891–896, 1996.

[11] R. Grossberg and S. Shelah. On the number of non isomorphic models of an infinitary theory which has the order property, part A. *Journal of Symbolic Logic*, 51:302–322, 1986.

[12] R. Grossberg and M. VanDieren. Categoricity from one successor cardinal in tame abstract elementary classes. *Journal of Mathematical Logic*, 6:181–201, 2006.

[13] B. Hart and S. Shelah. Categoricity over P for first order T or categoricity for $\phi \in \mathrm{l}_{\omega_1\omega}$ can stop at $\aleph_k$ while holding for $\aleph_0, \ldots, \aleph_{k-1}$. *Israel Journal of Mathematics*, 70:219–235, 1990.

[14] W. Hodges. *Model Theory.* Cambridge University Press, 1993.

[15] T. Hyttinen and M. Kesälä. Superstability in simple finitary AECs. *Fundamenta Mathematicae*, 195(3):221–268, 2007.

[16] T. Jech. *Multiple Forcing*, vol. 88 of *Cambridge Topics in Mathematics.* Cambridge University Press, 1987.

[17] H. J. Keisler. *Model Theory for Infinitary Logic.* North-Holland, 1971.

[18] J. Kirby. Abstract elementary categories. `http://arxiv.org/abs/1006.0894v1`, 2008.

[19] D. W. Kueker. Abstract elementary classes and infinitary logics. *Annals of Pure and Applied Logic*, 156:274–286, 2008.

[20] K. Kunen. *Set Theory, An Introduction to Independence Proofs.* North-Holland, 1980.

[21] O. Lessmann. An introduction to excellent classes. In Yi Zhang (ed.), *Logic and its Applications*, Contemporary Mathematics, 231–261. American Mathematical Society, 2005.

[22] M. Lieberman. Accessible categories vrs aecs. `www.math.lsa.umich.edu/~liebermm/vita.html`.

[23] M. Malicki. On Polish groups admitting a compatible complete left-invariant metric. *Journal of Symbolic Logic*, 76(2):437–447, 2011,

[24] R. Mansfield and G. Weitkamp. *Recursive Aspects of Descriptive Set Theory.* Oxford University Press, 1985.

[25] D. Marker. Descriptive set theory, 2002. `http://www.math.uic.edu/~marker/math512/dst.pdf`.

[26] M. Nadel. More Löwenheim–Skolem results for admissible sets. *Israel J. Math.*, 18:53–64, 1974.

[27] M. Nadel. Scott sentences and admissible sets. *Annals of Mathematical Logic*, 7:267–294, 1974.

[28] G. Sacks. *Saturated Model Theory.* Benjamin, Reading, Mass., 1972.

[29] G. Sacks. *Higher Recursion Theory.* Springer-Verlag, Berlin, Heidelberg, 1990.

[30] S. Shelah. Categoricity in $\aleph_1$ of sentences in $L_{\omega_1,\omega}(Q)$. *Israel Journal of Mathematics*, 20:127–148, 1975. Paper 48.

[31] S. Shelah. Classification theory for nonelementary classes, I. The number of uncountable models of $\psi \in L_{\omega_1\omega}$ part A. *Israel Journal of Mathematics*, 46:3:212–240, 1983. Paper 87a.

[32] S. Shelah. Classification theory for nonelementary classes, I. The number of uncountable models of $\psi \in L_{\omega_1\omega}$ part B. *Israel Journal of Mathematics*, 46:3:241–271, 1983. Paper 87b.

[33] S. Shelah. Categoricity for abstract classes with amalgamation. *Annals of Pure and Applied Logic*, 98:261–294, 1999. Paper 394. Consult Shelah for post-publication revisions.

[34] S. Shelah. *Classification Theory for Abstract Elementary Classes.* Studies in Logic. College Publications. www.collegepublications.co.uk, 2009. Binds together papers 88r, 600, 705, 734 with introduction E53.

[35] S. Shelah. Model theory without choice? Categoricity. *Journal of Symbolic Logic*, 74:361–401, 2009.

[36] S. Shelah. Non-structure in λ^{++} using instances of WGCH. Paper 838.

[37] S. Shelah. Abstract elementary classes near $\aleph_1$ sh88r. Revision of Classification of nonelementary classes II, Abstract elementary classes; on the Shelah archive.

[38] S. Shelah. Classification of nonelementary classes II, abstract elementary classes. In J. T. Baldwin (ed.), *Classification Theory (Chicago, IL, 1985)*, 419–497. Springer, Berlin, 1987. Paper 88: Proceedings of the USA-Israel Conference on Classification Theory, Chicago, December 1985; vol. 1292 of Lecture Notes in Mathematics.

[39] S. Shelah. Categoricity of theories in $L_{\kappa\omega}$ when κ is a measurable cardinal, part II. *Fundamenta Mathematicae*, 170:165–196, 2001.

[40] B. I. Zilber. Pseudo-exponentiation on algebraically closed fields of characteristic 0. *Annals of Pure and Applied Logic*, 132:67–95, 2004.

[41] B. I. Zilber. Covers of the multiplicative group of an algebraically closed field of characteristic zero. *Journal of the London Mathematical Society*, 74(1):41–58, 2006.

Beyond first order logic: from number of structures to structure of numbers, part I

John T. Baldwin[†], **Tapani Hyttinen**[‡], **Meeri Kesälä**[‡]

[†] Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, USA
jbaldwin@uic.edu

[‡] Department of Mathematics and Statistics, University of Helsinki, Finland
tapani.hyttinen@helsinki.fi, meeri.kesala@helsinki.fi

Abstract. The paper studies the history and recent developments in non-elementary model theory focusing in the framework of *abstract elementary classes.* We discuss the role of syntax and semantics and the motivation to generalize first order model theory to non-elementary frameworks and illuminate the study with concrete examples of classes of models.

This first part introduces the main conceps and philosophies and discusses two research questions, namely categoricity transfer and the stability classification.

Introduction

Model theory studies classes of structures. These classes are usually a collection of structures that satisfy an (often complete) set of sentences of first order logic. Such sentences are created by closing a family of basic relations under finite conjunction, negation and quantification over individuals. *Non-elementary logic* enlarges the collection of sentences by allowing longer conjunctions and some additional kinds of quantification. In this paper we first describe for the general mathematician the history, key questions, and motivations for the study of non-elementary logics and distinguish it from first order model theory. We give more detailed examples accessible to model theorists of all sorts. We conclude with questions about countable models which require only a basic background in logic.

For the last 50 years most research in model theory has focused on first order logic. Motivated both by intrinsic interest and the ability to better describe certain key mathematical structures (e.g., the complex numbers with exponentiation), there has recently been a revival of 'non-elementary model theory'. We develop contrasts between first order and non-elementary logic in a more detailed way than just noting 'failure of compactness'. We explain the sense in which we use the words syntax and semantics in Section 1. Many of the results and concepts in this paper will reflect a tension between these two viewpoints. In Part II, as we move from the study of classes that are defined syntactically to those that are defined semantically, we will be searching for a replacement for the fundamental notion of first order model theory, i.e., the notion of a complete theory. Section 1 also defines the basic notions of non-elementary model theory. Section 2 describes some of the research streams in more detail and illuminates some of the distinctions between elementary and non-elementary model theory. Subsection 2.1 describes the founding result

Accepted for publication in 2012 in the Bulletin of the Iranian Mathematical Society.

[†] Partially supported by grant NSF-0500841.

[‡] Partially supported by the Academy of Finland, grant number 1123110.

of modern first order model theory, Morley's categoricity theorem, and sketches Shelah's generalization of it to $L_{\omega_1\omega}$. In Part II we study several generalizations of the result to abstract elementary classes (AEC). The remainder of Section 2 studies the so-called stability classification and provides specific mathematical examples that illustrate some key model theoretic notions. We describe concrete examples explaining the concepts and problems in non-elementary model theory and a few showing connections with other parts of mathematics. Two of these illustrate the phrase 'to structure of numbers' in the title. Example 2.11, initiated by Zilber, uses infinitary methods to study complex exponentiation and covers of abelian varieties. The example in Subsection 2.3 of Part II studies models of Peano Arithmetic and the notion of elementary end-extension. This is the first study of models of Peano arithmetic as an AEC. Furthermore, Part II contains new results and explores the proper analogy to complete theory for AECs; it answers a question asked by David Kueker and includes Kossak's example of a class of models of PA interesting from the standpoint of AEC.

Neither of the standard approaches, $L_{\kappa\omega}$-definable class or AECs, has been successful in studying the countable models of an infinitary sentence. The first approach is too specific. It rapidly reduces to a complete infinitary sentence which has only one countable model. Results so far in studying general AECs give little information about countable models. We seek to find additional conditions on an AEC that lead to a fruitful study of the class of countable models. In particular we would like to find tools for dealing with one famous and one not so famous problem of model theory. The famous problem is Vaught's conjecture. Can a sentence of $L_{\omega_1\omega}$ have strictly between $\aleph_0$ and $2^{\aleph_0}$ countable models? The second problem is more specific. What if we add the condition that the class is $\aleph_1$-categorical? Can we provide sufficient conditions for having less than $2^{\aleph_0}$ countable models or for actually counting the number of countable models? In Part II we describe two sets of concepts for addressing this issue; unfortunately so far not very successfully. The first is the notion of a simple finitary AEC and the second is an attempt to define a notion of a 'complete AEC', which like a complete first order theory imposes enough uniformity to allow analysis of the models but without trivializing the problem to one model.

One thesis of this paper is that the importance of non-elementary model theory lies not only in widening the scope of applications of model theory but also in shedding light on the essence of the tools, concepts, methods and conventions developed and found useful in elementary model theory.

We thank Jouko Väänänen and Juliette Kennedy from the University of Helsinki for discussions that led to better understanding on the history of non-elementary model theory, the philosophical issues discussed in Section 1, and for helpful references.

1 Non-elementary model theory

In this section we study the history of non-elementary model theory during the second half of the twentieth century and compare that to the development of more 'mainstream' first order model theory. We identify two different trends in the development. In both the 'elementary' and non-elementary cases the focus of research has moved from 'syntactic' consideration towards 'semantic' ones —we will explain what we mean by this. We see some of the cyclic nature of science. Non-elementary classes bloom in the 60's and 70's; the bloom fades for some decades, overshadowed by the success and applications arising from the 'elementary' field. But around the turn of the 21st century, innovative examples and further internal developments lead to a rebirth.

We will focus on some 'motivating questions' that have driven both the elementary and non-elementary approaches, such as the categoricity transfer problem. While counting models seems a rather mundane problem, new innovations and machinery developed for the solution have led to the recognition of systems of invariants that are new to mathematics and in the first order case to significant mathematical advances in e.g. number theory [**13**]. It is hoped that the deeper developments of infinitary logic will have similar interactions with core mathematics. Boris Zilber's webpage contains many beginnings.

1.1 Syntax and semantics

The distinction between *syntax* and *semantics* has been present throughout the history of modern logic starting from the late 19th century: *completeness theorems* build a bridge between the two by asserting that a sentence is provable if and only if it is true in all models. By syntax we refer to the formalism of logic, objects of language as strings of symbols and deductions as manipulations of these strings according to certain rules. Semantics, however, has to do with interpretations, 'meaning' and 'sense' of the language. By the *semantics* for a language we mean a 'truth definition' for the sentences of the language, a description of the conditions when a structure is considered to be a model for that sentence. 'Semantic properties' have to do with properties of such models.

In fact these two notions can also be seen as methodologies or attitudes toward logic. The extreme (formalist) view of the syntactic method avoids reference to any 'actual' mathematical objects or meaning for the statements of the language, considering these to be 'metaphysical objects'. The semantic attitude is that logic arises from the tradition of mathematics. The method invokes a trace of Platonism, a search for the 'truth' of statements with less regard for formal language. The semantic method would endorse 'proof in metamathematics or set theory' while the syntactic method seeks a 'proof in some formal system'. Traditionally model theory is seen as the intersection of these two approaches. Chang and Keisler [**17**] write: *universal algebra + logic = model theory.* Juliette Kennedy [**33**] discusses ideas of 'formalism freeness', found in the work of Kurt Gödel. Motivated by issues of incompleteness and faithfulness and hence the 'failure' of first order logic to capture truth and reasoning, Gödel asked if there is some (absolute) concept of proof (or definability) 'by all means imaginable'. One interpretation of this absolute notion (almost certainly not Gödel's) is as the kind of semantic argument described above. We will spell out this contrast in many places below.

Model theory by definition works with the semantic aspect of logic, but the dialectics between the syntactic and semantic attitudes is central. This becomes even clearer when discussing questions arising from *non-elementary model theory.* Non-elementary model theory studies formal languages other than 'elementary' or first order logic; most of them extend first order. We began by declaring that model theory studies classes of models. Traditionally, each class is the collection of models that satisfy some (set of) sentence(s) in a particular logic. Abstract elementary classes provide new ways of determining classes: a class of structures in a fixed vocabulary is characterized by semantic properties. The notion of AEC does not designate the models of a collection of sentences in some formal language, although many examples arise from such syntactic descriptions. In first order logic, the most fruitful topic is classes of models of complete theories. A *theory* T is a set of sentences in a given language. We say that T is *complete* if for every sentence ϕ in the language, either T implies ϕ, or T implies $\neg\phi$. In Part II we seek an analogue to completeness for AEC.

Model-Theoretic Logics, edited by Barwise and Feferman [**8**], summarizes the early study of non-elementary model theory. In this book, 'abstract model theory' is a study comparing different logics with regard to such properties as interpolation, expansions, relativizations and projections, notions of compactness, Hanf and Löwenheim–Skolem numbers.

A vocabulary[1] L consists of constant symbols, relation symbols and function symbols, which have a prescribed number of arguments (arity). An L-structure consists of a universe, which is a set, and interpretations for the symbols in L. When L' is a subset of a vocabulary L, and M is an L-structure, we can talk about the reduct of M to L', written $M \upharpoonright L'$. Then M is the expansion of $M \upharpoonright L'$ to L. If M and N are two L-structures, we say that M is an *L-substructure* of N if the domain of M is contained in the domain of N and the interpretations of all the symbols in L in M agree with the restriction of N to M.

A *formal language* or logic in the vocabulary L is a collection of formulas that are built by certain rules from the symbols of the vocabulary and from some 'logical symbols'. In this paper we focus on countable vocabularies but do not needlessly restrict definitions to this case.

L-terms are formed recursively from variables and the constant and function symbols of the vocabulary by composing in the natural manner. With a given interpretation for the constants and assignment of values for the variables in a structure, each term designates an element in the structure.

An *atomic formula* is an expression $R(t_1, \dots, t_n)$ where R is an n-ary relation symbol (including equality) of the vocabulary and each t_i is a term.

Definition 1.1 (The language $L_{\lambda\kappa}$) Assume that L is a vocabulary. The language $L_{\lambda\kappa}$ consists of formulas $\phi(\overline{x})$, where the free variables of the formula are contained in the finite sequence $\overline{x}$ and where the formulas are built with the following operations:

- $L_{\lambda\kappa}$ contains all atomic formulas in the vocabulary L.
- If $\phi(\overline{x})$, $\psi(\overline{x})$ are in $L_{\lambda\kappa}$, then the negation $\neg\phi(\overline{x})$ and implication $(\phi(\overline{x}) \rightarrow \psi(\overline{x}))$ are in $L_{\lambda\kappa}$.
- If $\phi_i(\overline{x})$ is in $L_{\lambda\kappa}$ for every i in the index set I, and $|I| < \lambda$, the conjunction $\bigwedge_{i\in I} \phi_i(\overline{x})$ and disjunction $\bigvee_{i\in I} \phi_i(\overline{x})$ are in $L_{\lambda\kappa}$.
- If $\phi(y_i, \overline{x})$ is in $L_{\lambda\kappa}$ for each i in the well-ordered index set I, and $|I| < \kappa$, then the quantified formula $(Q_i y_i)_{i\in I}\phi(\overline{x})$ is in $L_{\lambda\kappa}$, where each quantifier Q_i is either $\forall$ ('for all y_i') or $\exists$ ('there exists y_i').

First order logic is the language $L_{\omega\omega}$, i.e., only finite operations are allowed. We define that $L_{\infty}\kappa$ is the union of all $L_{\lambda\kappa}$ for all cardinal numbers λ.

The languages $L_{\lambda\omega}$ allowing only finite strings of quantifiers are much better behaved. We will later introduce *abstract elementary classes* generalizing, among other things, classes of structures definable with a sentence in $L_{\lambda\omega}$. The definition of the *truth* of a

[1] Another convention specifies the vocabulary by a small Greek letter and the L with decorations describes the particular logic. What we call a vocabulary is sometimes called a *language*. We have written *language* or *logic* for the collections of sentences; more precisely, this might be called the language and the logic would include proof rules and even semantics.

formula in a structure is crucial. For a formula $\phi(\overline{x})$, with the sequence $\overline{x}$ containing all the free variables of ϕ, we define what it means that the formula $\phi(\overline{x})$ is true in an L-structure M with the variables $\overline{x}$ interpreted in a particular way as elements $\overline{a}$, written $M \models \phi(\overline{a})$. The definition is done by induction on the complexity of the formula, following the inductive definition of the formula in Definition 1.1.

Definition 1.2 (The language $L(Q)$) The language $L(Q)$ is formed as the first order logic $L_{\omega\omega}$, but we allow also formulas of the form $Qy\phi(y, \overline{x})$ with the following truth definition: $M \models Qy\phi(y, \overline{a})$ if there are uncountable many $b \in M$ such that $M \models \phi(b, \overline{a})$.

Definition 1.3 (Elementary substructure with respect to a fragment) A subset $\mathcal{F} \subseteq L$ is a *fragment* of some formal language L if it contains all atomic formulas and is closed under subformulas, substitution of variables with L-terms, finite conjunction and disjunction, negation and the quantifiers $\forall$ and $\exists$, applied finitely many times. For two L-structures M and N, we say that M is an *$\mathcal{F}$-elementary substructure* of N, written $M \preccurlyeq_{\mathcal{F}} N$, if M is an L-substructure of N and for all formulas $\phi(\overline{x})$ of $\mathcal{F}$ and sequences $\overline{a}$ of elements in M, $M \models \phi(\overline{a})$ if and only if $N \models \phi(\overline{a})$.

Definition 1.4 (Elementary class and PC-class) An *elementary class* $\mathbb{K}$ of L-structures is the class of all models of a given theory in first order logic. A *pseudo-elementary* (PC) class $\mathbb{K}$ is the class of reducts $M \restriction L$ of some elementary class in a larger vocabulary $L' \supseteq L$.

We say that a formal language (logic) L is *compact* if whenever a set of sentences is *inconsistent*, that is, has no model, then there is some finite subset which is already inconsistent. This is a crucial property that, along with the upwards Löwenheim–Skolem property, fails in most non-elementary logics.

The Löwenheim–Skolem number and the Hanf number are defined for a formal logic L (i.e., 'the Löwenheim–Skolem or Hanf number of L'). In the following definitions $\mathbb{K}$ is a class definable with a sentence of L, $\preccurlyeq_{\mathbb{K}}$ is given as the $\mathcal{F}$-elementary substructure relation in some given fragment $\mathcal{F}$ of L, usually the smallest fragment containing the sentence defining $\mathbb{K}$, and the collection $\mathcal{C}$ is the collection of all classes definable with a sentence L.

Definition 1.5 (Löwenheim–Skolem number) The Löwenheim–Skolem number LS($\mathbb{K}$) for a class of structures $\mathbb{K}$ and a relation $\preccurlyeq_{\mathbb{K}}$ between the structures is the smallest cardinal number λ with the following property: For any $M \in \mathbb{K}$ and a subset $A \subseteq M$ there is a structure $N \in \mathbb{K}$ containing A such that $N \preccurlyeq_{\mathbb{K}} M$ and $|N| \leq \max\{\lambda, |A|\}$.

Definition 1.6 (Hanf number) The Hanf number H for a collection $\mathcal{C}$ of classes of structures is the smallest cardinal number with the property: for any $\mathbb{K} \in \mathcal{C}$, if there is $M \in \mathbb{K}$ of size at least H, then $\mathbb{K}$ contains arbitrarily large structures.

Modern model theory began in the 1950's. Major achievements in the mid 60's and early 70's included Morley's categoricity transfer theorem in 1965 [**43**] and Shelah's development of stability theory [**49**]. These works give results on counting the number of isomorphism types of structures in a given cardinality and establishing invariants in order to classify the isomorphism types. Such invariants arise naturally in many concrete classes: the dimension of a vector space or the transcendence degree of an algebraically closed field are prototypical examples. A crucial innovation of model theory is to see how to describe structures by *families* of dimensions. The general theory of dimension appears in e.g. ([**45, 49**]); it is further developed and applied to valued fields in [**23**].

Non-elementary model theory thrived in the mid 60's and early 70's. Results such as Lindström theorem in 1969, Barwise's compactness theorem for *admissible fragments* of $L_{\omega_1\omega}$ published in 1969, Mostowski's work on generalized quantifiers in 1957 [**44**] and Keisler's beautiful axiomatization of $L(Q)$ in [**31**] gave the impression of a treasury of new formal languages with amenable properties, a possibility to extend the scope of definability and maybe get closer to the study of provability with 'all means imaginable'. However, the general study turned out to be very difficult. For example, the study of the languages $L_{\lambda\kappa}$ got entangled with the set-theoretical properties of the cardinals λ and κ. Since the real numbers are definable as the unique model of a sentence in $L_{2^\omega\omega}$, the continuum hypothesis would play a major role. But perhaps the study was focused too much on the syntax and trying to study the model theory of *languages*? Why not study the properties of classes of structures, defined semantically. One might replace compactness with, say, closure under unions of chains?

One can argue that a major achievement of non-elementary model theory has been to *isolate* properties that are crucial for classifying structures, properties that might not be visible to a mathematician working with only a specific application or even restricted to the first order case. Excellence (see below) is a crucial example. Some examples of applications of non-elementary model theory to 'general mathematics' are presented in the chapter 'Applications to Algebra' by Eklof in [**8**]. In many of these applications we can see that some class of structures is definable in $L_{\omega_1\omega}$ or in $L_{\infty\omega}$ and then use the model theory of these languages to, for example, count the number of certain kind of structures or classify them in some other way. Barwise writes in *Model-Theoretic Logics* [**8**]:

> Most important in the long run, it seems, is where logic contributes to mathematics by leading to the formation of concepts that allow the right questions to be asked and answered. A simple example of this sort stems from 'back-and-forth arguments' and leads to the concept of partially isomorphic structures, which plays such an important role in extended model theory. For example, there is a classical theorem by Erdős, Gillman and Henriksen; two real-closed fields of order type η_1 and cardinality $\aleph_1$ are isomorphic. However, this way of stating the theorem makes it vacuous unless the continuum hypothesis is true, since without this hypothesis there are no fields which satisfy both hypotheses. But if one looks at the proof, there is obviously something going on that is quite independent of the size of the continuum, something that needs a new concept to express. This concept has emerged in the study of logic, first in the work of Ehrenfeucht and Fraïssé in first-order logic, and then coming into its own with the study of infinitary logic. And so in his chapter (in [**8**]), Dickmann shows that the theorem can be reformulated using partial isomorphisms as: Any two real-closed fields of order type η_1, of any cardinality whatsoever, are strongly partially isomorphic. There are similar results on the theory of abelian torsion groups which place Ulm's theorem in its natural setting. (...) Extended model theory provides a framework within which to understand existing mathematics and push it forward with new concepts and tools.

One of the foundational discoveries of abstract model theory was Per Lindström's theorem that first order logic is the strongest logic which has both the compactness property and a countable Löwenheim–Skolem number. In order to study such concepts as 'the strongest logic', one has to define the notion of an 'abstract logic'. The book [**8**] presents the syntax as a crucial part: an abstract logic is a class of sentences with a satisfaction relation between the sentences and the structures, where this relation satisfies certain properties. However, Barwise comments on Lindström's formulation of his theorem [**38**]:

> To get around the difficulties of saying just what a logic is, they dealt entirely with classes of structures and closure conditions on these classes, thinking of the classes definable in some logic. That is, they avoided the problem of formulating a notion of a logic in terms of syntax, semantics, and satisfaction, and dealt purely with their semantic side.

Lindström defined a logic to be a non-empty set of objects called sentences, but the role of these is only to name a class of structures as 'structures modeling one sentence'. Then it is possible to define for example compactness as the property that if a countable intersection of such classes is empty, then already some finite intersection must be empty.

Saharon Shelah built on these insights and introduced abstract elementary classes in [**51**]. Semantic properties of a class of structures $\mathbb{K}$ and a relation $\preccurlyeq_{\mathbb{K}}$ are prescribed, which are sufficient to isolate interesting classes of structures. But more than just the class is described; the relation $\preccurlyeq$ between the structures in $\mathbb{K}$ provides additional information that, as examples in Subsection 2.2 illustrate, may be crucial.

Definition 1.7 For any vocabulary τ, a class of τ-structures $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ is *an abstract elementary class* (AEC) if

(1) Both $\mathbb{K}$ and the binary relation $\preccurlyeq_{\mathbb{K}}$ are closed under isomorphism.
(2) If $\mathcal{A} \preccurlyeq_{\mathbb{K}} \mathcal{B}$, then $\mathcal{A}$ is a substructure of $\mathcal{B}$.
(3) $\preccurlyeq_{\mathbb{K}}$ is a partial order on $\mathbb{K}$.
(4) If $\langle \mathcal{A}_i : i < \delta \rangle$ is an $\preccurlyeq_{\mathbb{K}}$-increasing chain:
  (a) $\bigcup_{i<\delta} \mathcal{A}_i \in \mathbb{K}$;
  (b) for each $j < \delta$, $\mathcal{A}_j \preccurlyeq_{\mathbb{K}} \bigcup_{i<\delta} \mathcal{A}_i$;
  (c) if each $\mathcal{A}_i \preccurlyeq_{\mathbb{K}} \mathcal{M} \in \mathbb{K}$, then $\bigcup_{i<\delta} \mathcal{A}_i \preccurlyeq_{\mathbb{K}} \mathcal{M}$.
(5) If $\mathcal{A}, \mathcal{B}, \mathcal{C} \in \mathbb{K}$, $\mathcal{A} \preccurlyeq_{\mathbb{K}} \mathcal{C}$, $\mathcal{B} \preccurlyeq_{\mathbb{K}} \mathcal{C}$ and $\mathcal{A} \subseteq \mathcal{B}$ then $\mathcal{A} \preccurlyeq_{\mathbb{K}} \mathcal{B}$.
(6) There is a Löwenheim–Skolem number $\mathrm{LS}(\mathbb{K})$ such that if $\mathcal{A} \in \mathbb{K}$ and $B \subset \mathcal{A}$ a subset, there is $\mathcal{A}' \in \mathbb{K}$ such that $B \subset \mathcal{A}' \preccurlyeq_{\mathbb{K}} \mathcal{A}$ and $|\mathcal{A}'| = |B| + \mathrm{LS}(\mathbb{K})$.

When $\mathcal{A} \preccurlyeq_{\mathbb{K}} \mathcal{B}$, we say that $\mathcal{B}$ is an $\mathbb{K}$-extension of $\mathcal{A}$ and $\mathcal{A}$ is an $\mathbb{K}$-submodel of $\mathcal{B}$. If $\mathcal{A}, \mathcal{B} \in \mathbb{K}$ and $f\colon \mathcal{A} \to \mathcal{B}$ is an embedding such that $f(\mathcal{A}) \preccurlyeq_{\mathbb{K}} \mathcal{B}$, we say that f is a $\mathbb{K}$-*embedding.* Category-theoretic versions of the axioms are studied by Kirby [**34**], Liebermann [**37**] and Beke and Rosický [**11**].

A basic example of an AEC is the class of models defined by some sentence $\phi \in L_{\infty\omega}$, where $\preccurlyeq_{\mathbb{K}}$ is taken as the elementary substructure relation in the smallest fragment of $L_{\infty\omega}$ containing ϕ. Then the Löwenheim–Skolem number is the size of the fragment. An even simpler example is that of an elementary class, where ϕ is a complete theory in first order logic.

A class defined with a sentence in $L_{\omega_1\omega}(Q)$ with the quantifier $Qx\phi(x)$ standing for 'there exists uncountably many x such that $\phi(x)$ holds' can be an AEC. The natural syntactic notion of elementary submodel is inadequate but substitutes are available. Arbitrary pseudo-elementary classes are often not AEC. For example, If $\mathbb{K}$ is the class of all structures A in a language L with a single unary predicate such that $|A| \leq 2^{|U(A)|}$ then $\mathbb{K}$ fails to be an AEC with respect to L-elementary submodels as it is not closed under unions of chains; see [**4**, Chapter 5 and 4.29].

In contemporary first order model theory, the most fundamental concept is the class of models of a complete theory in first order logic. This can be seen as a form of *focusing*; instead of studying different vocabularies, expansions and projections, one fixes one class: the class of differentially closed fields of fixed characteristic (see [**41**]) or the class of models of 'true' arithmetic. This focus on classes and of properties determining 'similar' classes has become a crucial tool in applications to algebra. The difference from the 'Lindström-style' study of classes of structures is significant: we do not study many classes of structures each corresponding to the 'models of one sentence', but focus on a fixed class, 'models of a theory'. *Abstract elementary classes*, which will be one of the main notions studied in this paper, takes the 'semantic view' to the extreme by eliminating the syntactic definition.

2 Several research lines in non-elementary logic

2.1 Categoricity transfer in $L_{\omega\omega}$ and $L_{\omega_1\omega}$

Definition 2.1 (Categoricity) Let κ be a cardinal. We say that a class of structures $\mathbb{K}$ is *κ-categorical* if there is exactly one model of size κ in $\mathbb{K}$, up to isomorphism. A theory T is *κ-categorical* if $\mathrm{Mod}(T)$, the class of models of T, is κ-categorical.

The transition to the focus on classes of models begins with Morley's theorem:

Theorem 2.2 (Morley's categoricity transfer theorem) *Assume that T is a complete theory in $L_{\omega\omega}$, where L is countable. If there exists an uncountable cardinal κ such that T is κ-categorical, then T is λ-categorical for all uncountable cardinals λ.*

Categoricity transfer will be our first example of a motivating question in the history of model theory. Its proof gave many new tools and concepts that are nowadays contained in every basic course in model theory. Furthermore, both the tools and the theorem itself have been generalized to different frameworks. A categoricity transfer theorem for elementary classes in an uncountable vocabulary was proved by Shelah in [**47**] (announced in 1970): if the language has cardinality κ and a theory is categorical in some uncountable cardinal greater than κ then it is categorical in all cardinalities greater than κ. This widening of scope led to many tools, such as weakly minimal sets and a greater focus on the properties of individual formulas, that proved fruitful for countable vocabularies. We will look more closely at some of the many extensions of categoricity results to non-elementary classes.

We consider a syntactical *type* in some logic $\mathcal{L}$ as a collection of $\mathcal{L}$-formulas in some finite sequence of variables $\overline{x}$ with *parameters* from a given subset A of a structure M such that an element $\overline{b}$ in an $\mathcal{L}$-elementary extension N of M realizes (simultaneously satisfies) p. If no such sequence exists in a model N, we say that the type is *omitted* in N. In elementary classes, the compactness theorem implies all *finitely consistent* such collections p of formulas are really *realized.* If there is a structure N and a finite sequence $\overline{b} \in N$ such that $M \preccurlyeq N$ and

$$p = \{\phi(\overline{x}, \overline{a}) : \overline{a} \in A \subseteq M, N \models \phi(\overline{b}, \overline{a})\},$$

then p is called a *complete type* over A for two reasons. Semantically, it gives a complete description of the relation of $\overline{b}$ and A. Syntactically, every formula $\phi(\overline{x}, \overline{a})$ over A or its negation is in p.

An essential concept in Morley's argument is a *saturated structure* M: M is *saturated* if all *consistent* types over parameter sets of size strictly less than $|M|$ are realized in M. Two saturated models of T of size κ are always isomorphic. Morley shows that if T is categorical in some uncountable power, saturated models exist in each infinite cardinality. Then he concludes that if T is *not* categorical in some uncountable power λ, there is a model of power λ which is *not* saturated or even $\aleph_1$-saturated; some type over a countable subset is omitted. But then he shows that if some model of uncountable power λ omits a type over a countable set, then in any other uncountable power κ some model omits the type. Hence, T cannot be categorical in κ either. This method, *saturation transfer*, generalizes to many other frameworks. While proving saturation transfer for elementary classes he introduced many new concepts such as a *totally transcendental theory* ($\aleph_0$*-stable* theory), *prime models over sets* and *Morley sequences.*

Keisler generalized many of the ideas from Morley's proof to the logic $L_{\omega_1\omega}$; see [**32**]. He studies a class of structures $(\mathbb{K}, \preccurlyeq_{\mathcal{F}})$, where $\mathbb{K}$ is definable with a sentence in $L_{\omega_1\omega}$ and $\mathcal{F}$ is some countable fragment of $L_{\omega_1\omega}$ containing the sentence. He uses a concept of *homogeneity*, which is closely related to saturation.

Definition 2.3 For L-structures M and N and a fragment $\mathcal{F}$ of $L_{\omega_1\omega}$, $A \subset M$ a subset and $f\colon A \to N$ a function, write $(M, A) \equiv_{\mathcal{F}} (N, f(A))$ if for every formula $\phi(\overline{x}) \in \mathcal{F}$ and every $\overline{a} \in A$,

$$M \models \phi(\overline{a}) \text{ if and only if } N \models \phi(f(\overline{a})).$$

A model is $(\kappa, \mathcal{F})$-homogeneous if for every set $A \subseteq M$ of cardinality strictly less than κ and every $f\colon A \to M$, if $(M, A) \equiv_{\mathcal{F}} (M, f(A))$, then for all $b \in M$ there exists $c \in M$ such that

$$(M, A \cup \{b\}) \equiv_{\mathcal{F}} (M, f(A) \cup \{c\}).$$

Keisler proved the following theorem ([**32**, Theorem 35]):

Theorem 2.4 (Keisler 1971, announced in 1969) *Let $\mathcal{F}$ be a countable fragment of $L_{\omega_1\omega}$, $T \subseteq \mathcal{F}$ a set of sentences and $\kappa, \lambda > \omega$. Assume that:*

(1) *T is κ-categorical.*

(2) *For every countable model M of T, there are models N of T of arbitrarily large power such that $M \preccurlyeq_{\mathcal{F}} N$.*

(3) *Every model M of power κ is $(\omega_1, \mathcal{F})$-homogeneous.*

Then T is λ-categorical. Moreover, every model of T of power λ is $(\lambda, \mathcal{F})$-homogeneous.

One stage in the transition from strictly syntactic to semantic means of defining classes is Shelah's version of Theorem 2.4. To understand it, we need the following fact, which stems from Chang, Scott and López-Escobar (see for example [**16**] from 1968); the current formulation is Theorem 6.1.8 in the book [**4**].

Theorem 2.5 (Chang, Scott and López-Escobar) *Let ϕ be a sentence in $L_{\omega_1\omega}$ in a countable vocabulary L. Then there is a countable vocabulary L' extending L, a first-order L' theory T and a countable collection Σ of L'-types such that reduct is a* 1-1 *map from the models of T which omit Σ onto the models of ϕ.*

A crucial point is that the infinitary aspects are translated to a first order context, at the cost of expanding the vocabulary. If ϕ is a complete sentence, the pair (T, Σ) can be chosen so that the associated class of models is the class of atomic models of T (every tuple realizes a principal type). Saharon Shelah generalized this idea to develop a more general context, *finite diagrams* [**46**]. A *finite diagram* D is a set of types over the empty set and the class of structures consists of the models which only realize types from D. Shelah defined a structure M to be (D, λ)-homogeneous if it realizes only types from D and is $(|M|, L_{\omega\omega})$-homogenous (in the sense of Definition 2.3). He (independently) generalized Theorem 2.4 to finite diagrams. His argument, like Keisler's, required the assumption of homogeneity. Thus, [**46**] is the founding paper of *homogeneous model theory*, which was further developed in for example [**15, 21, 29, 30**]. The compact case ('Kind II' in [**48**]) was transformed into the study of continuous logics and abstract metric spaces [**12**] and finally generalized to metric abstract elementary classes [**24**]. These last developments have deep connections with Banach space theory.

Baldwin and Lachlan in 1971 [**7**] give another method for first order categoricity transfer. They develop some *geometric* tools to study structures of a theory categorical in some uncountable cardinal: any model of such a theory is prime over a *strongly minimal* set and the isomorphism type is determined by a certain *dimension* of the strongly minimal set. This gives a new proof for the Morley theorem for elementary classes but also the *Baldwin–Lachlan Theorem*: if an elementary class is categorical in some uncountable cardinal, it has either just one or $\aleph_0$-many countable models. The geometric analysis of uncountably categorical elementary classes was developed even further by Zilber (see [**57**], earlier Russian version [**56**]), giving rise to *geometric stability theory.* We discuss the number of countable models of an $\aleph_1$-categorical *non-elementary* class in Part II.

A further semantic notion closely tied to categoricity is Shelah's 'excellence'. Excellence is a kind of generalized amalgamation (details in [**4**]). The rough idea is to posit a type of unique *prime models* over certain *independent diagrams* of models. 'Excellence' was discovered independently by Boris Zilber while studying the model theory of an algebraically closed field with *pseudo-exponentiation* (a homomorphism from $(F, +)$ to $(F^*, \cdot)$. He defines the notion of a quasiminimal excellent (qme) class by 'semantic conditions'; Kirby [**35**] proved they can be axiomatized in $L_{\omega_1\omega}(Q)$. Zilber showed that any qme class is categorical in all uncountable powers and finds such a class of pseudo-exponential fields. Natural algebraic characterizations of excellence have been found in context of algebraic groups by Zilber and Bays [**9, 10, 60**]. Excellence implies that the class of structures has models in all cardinalities, has the *amalgamation property* (see Part II), and admits full categoricity transfer. Zilber's notion of 'excellence' specializes Shelah's notion of excellence for sentences in $L_{\omega_1\omega}$, invented while proving the following general theorem for transferring categoricity for sentences in $L_{\omega_1\omega}$ [**50**].[2] The theorem uses a minor assumption on cardinal arithmetic.

Theorem 2.6 (Shelah 1983) *Assume that* $2^{\aleph_n} < 2^{\aleph_{n+1}}$ *for all* $n < \omega$. *Let* $\phi \in L_{\omega_1\omega}$ *be a sentence which has an uncountable model, but strictly less than the maximal number of models in each cardinality* $\aleph_n$ *for* $0 < n < \omega$. *Then the sentence is* excellent.

(ZFC) *Assume that a sentence* ϕ *in* $L_{\omega_1\omega}$ *is excellent and categorical in some uncountable cardinality. Then* ϕ *is categorical in every uncountable cardinality.*

The excellence property is defined only for complete sentences in $L_{\omega_1\omega}$, more precisely for the associated classes of *atomic models* (each model omits all non-isolated types) of a first order theory T in an extended vocabulary. Excellent classes have been further

[2] The important first order notion of the OTOP discussed in Subsection 2.2 was derived from the earlier concept of excellence for $L_{\omega_1\omega}$.

studied in [**20, 27, 36**]. Theorem 2.6, expounded in [**4**], extends easily to incomplete sentences:

Corollary 2.7 *Assume that $2^{\aleph_n} < 2^{\aleph_{n+1}}$ for all $n < \omega$. Let $\phi \in L_{\omega_1\omega}$ be a sentence which is categorical in $\aleph_n$ for each $n < \omega$. Then ϕ is categorical in every cardinality.*

Shelah and Hart [**22**], made more precise in [**6**], show the necessity of considering categoricity up to $\aleph_\omega$; there are examples of $L_{\omega_1\omega}$-sentences ϕ_n which are categorical in each $\aleph_k$ for $k \leq n$ but have the maximal number of models in $\aleph_{n+1}$. However, it is not known whether the assumption on cardinal arithmetic can be removed from the theorem.

In the discussion above we isolated properties such as *homogeneity* and *excellence*, which enable one to prove categoricity transfer theorems. More importantly, they support the required tools for classifying and analyzing structures with model-theoretic methods; both generated subfields: *homogeneous model theory* and *model theory for excellent classes.* These properties have applications to 'general mathematics': $L_{\infty\omega}$-free algebras [**42**] for homogeneous model theory or Zilber's pseudo-exponentiation and the work on covers of abelian varieties [**59**] for excellence. We argue that finding such *fundamental properties* for organizing mathematics is one of the crucial tasks of model theory.

The investigation of $L_{\omega_1\omega}$ surveyed in this section makes no assumption that the class studied has large models; the existence of large models is deduced from sufficient categoricity in small cardinals. Shelah pursues a quite different line in [**52**]. He abandons the syntactic hypothesis of definability in a specific logic. In attempting to prove eventual categoricity, he chooses smaller AEC's in successive cardinalities. Thus he attempts to construct a smaller class which is categorical in all powers. Crucially, this work does not assume the existence of arbitarily large models.

We discuss more on categoricity transfer in AECs in Part II. There we will concentrate on certain type of AECs, namely Jónsson classes, where some categoricity transfer results are known and some stability theory along with a natural notion of type can be constructed. These classes are generalizations of homogeneous and excellent classes and they have arbitrarily large models and for example the *amalgamation property* by assumption.

2.2 The stability classification: first order vs. non-elementary

One of the major themes of contemporary model theory is the notion of classification theory. Classification is used in two senses. On the one hand models in a particular class can be classified by some assignment of structural invariants. On the other hand, the classes of models[3] are split into different groups according to common properties, which may be semantic or syntactic; many examples are given below. Shelah (e.g. [**52**]) has stressed the importance of certain properties of theories, those which are dividing lines: both the property and its negation have strong consequences. In the following we discuss various important classes of theories and emphasize those properties which are dividing lines.

Saharon Shelah originated *stability theory* for elementary classes [**49**] and produced much of the early work. Now, however, the field embraces much of model theory and the

[3] The word *class* is vastly overloaded in this context. In first order logic, a complete theory is a natural unit. In studying infinitary logic, the natural unit often becomes an AEC (in the first order case this would be the class of models of the theory).

tools are pervasive in modern applications of model theory. Among the many texts are [**2, 14, 45**].

We can define *stability* in λ as the property that there are no more than λ many distinct complete types over any subset of size λ. However, stability has many equivalent definitions in elementary classes. A remarkable consequence of the analysis is that counting the number of types is related to the geometry of the structures in the class. For example, if the class of structures is stable in any cardinal at all, one can define a notion of *independence* between arbitrary subsets of any model, which is a useful tool to analyze the properties of the structures in the class. The importance of such a notion of independence is well established and such *independence calculus* has been generalized to some unstable elementary classes such as classes given by *simple* [**54**] or *NIP* theories [**1**]. Stability theory has evolved to such fields as *geometric stability theory* [**45**], which is the major source for applications of model theory to 'general mathematics'.

Stability theory divides classes into four basic categories. This division is called the *stability hierarchy*:

(1) $\aleph_0$-stable classes;
(2) superstable classes, that is, classes stable from some cardinal onwards;
(3) stable classes, that is, stable in at least one cardinal;
(4) unstable classes.

In elementary classes, $\aleph_0$-stable classes are stable in all cardinalities and hence we get a hierarchy of implications (1) $\Rightarrow$ (2) $\Rightarrow$ (3). Uncountably categorical theories are always $\aleph_0$-stable whereas non-superstable classes have the maximal number of models in each uncountable cardinal. An $\aleph_0$-stable or superstable class can also have the maximal number of models, e.g., if it has one of the properties DOP or OTOP, discussed in Examples 2.9 and 2.10.

Developing stability theory for non-elementary classes is important not only because it widens the scope of applications but also because it forces further analysis of the tools and concepts developed for elementary classes. Which of the tools are there only because first order logic 'happens' to be compact and which could be cultivated to extend to non-elementary classes? Especially, can we *distinguish* some core properties enabling the process? What are the problems met in, say, categoricity transfer or developing independence calculus? Why does the number of *types* realized in the structure seem to affect the geometric properties of structures and can we analyze the possible geometries arising from different frameworks? For example, Hrushovski [**25**] proved a famous theorem in geometric stability theory: under assumptions of a logical nature the geometry given by the notion of independence on the realizations of a regular type, must fall into one of three natural categories involving group actions. In the available non-elementary versions of the same theorem ([**26, 28**]), we cannot rule out a fourth possibility: existence of a so-called *non-classical group*, a non-abelian group admitting an ω-homogeneous pre-geometry. We can identify some quite peculiar properties of such groups. Even their existence is open.

The established notion of type for abstract elementary classes is a so-called *Galois type*, which we will define more carefully in Part II. Then κ-stability is defined with respect to these types: A class of structures is stable in a cardinal κ if no structure in the class realizes more than κ many Galois types over an $\preccurlyeq_{\mathbb{K}}$-elementary substructure of size $\leq \kappa$. For the remainder of this section the reader can think of the following descriptive notion on a Galois type: Let $\mathcal{A} \preccurlyeq_{\mathbb{K}} \mathcal{B}$ and a, b be elements in $\mathcal{B}$. We say that a and b

have the same Galois type over the structure $\mathcal{A}$ if there is $\mathcal{C}$ such that $\mathcal{B} \preccurlyeq_{\mathbb{K}} \mathcal{C}$ and an automorphism of $\mathcal{C}$ fixing $\mathcal{A}$ pointwise and mapping a to b.

We present here some examples of AECs where the choice of the relation $\preccurlyeq_{\mathbb{K}}$ affects the placement of the class in the stability hierarchy. How 'coincidental' is the division of elementary classes according to the stability hierarchy? The placement of a class of structures in the hierarchy has been shown to affect a huge number of properties that at first sight do not seem to have much to do with the number of types. Which of these connections are 'deep' or 'semantic', or especially, which extend to non-elementary frameworks? Can an appropriate hierarchy be found?

The moral of these examples is that properties of the 'same' class of structures might look different if definitions in logics with more expressive power are allowed or a different notion $\preccurlyeq_{\mathbb{K}}$ for an abstract elementary class is chosen.

Example 2.8 (Abelian groups) Let $\mathbb{K}$ be the class of all abelian groups. Then $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ is an $\aleph_0$-stable AEC with the notion $\preccurlyeq_{\mathbb{K}}$ as the substructure relation.

However, the same class of structures is strictly stable (stable but not superstable) if we take as $\preccurlyeq_{\mathbb{K}}$ the following notion: $M \preccurlyeq_{\mathbb{K}} N$ if and only if M is a subgroup and for each $a \in M$ and $n \in \mathbb{N} \setminus \{0\}$, n divides a in M if and only if n divides a in N.

The model theory of abelian groups is studied in Eklof–Fischer [**18**], where the latter notion of $\preccurlyeq_{\mathbb{K}}$ is in the focus of study. AECs induced by tilting and co-tilting modules are studied in Baldwin–Eklof–Trlifaj [**5**]. In [**53**], a more semantic notion of $\preccurlyeq_{\mathbb{K}}$ is provided and the classes of abelian groups are strictly stable except in one degenerate case.

A number of properties in first order classification theory induce 'bad behavior' for an elementary class of structures, signaled by the existence of the maximal number of models in a given cardinality. The most basic of these are instability and unsuperstability. Others include OTOP, 'the omitting types order property', and DOP, 'the dimensional order property', with a version ENI-DOP, which gives many countable models. Especially, these play a role in classifying countable complete first order theories; their negations NOTOP, NDOP and ENI-NDOP have 'good' implications from the viewpoint of classification theory; they aid in the assigning of invariants.

One equivalent definition for unstability is that there is a formula which in the models of a first order theory defines an infinite ordering. Then, by compactness, the elementary class must contain models interpreting various different orderings, which (nontrivially) forces the number of models to the maximum. Similarly the properties DOP and OTOP cause certain kind of orderings to appear in the structures; however, the orderings are not defined by a single first order formula. Just as in Example 2.8, the unsuperstability of the class of abelian groups is not visible to quantifier-free formulas, the only ones 'seen' by the substructure relation, OTOP and DOP are a form of instability not visible to first order formulas.

The following two examples illustrate the properties OTOP and DOP. In each case we 'define' an arbitrary graph (e.g., an ordering) on $P \times P$ by describing a column above each point of the plane. The two methods of description, by a type or a single formula, distinguish OTOP and DOP.

Example 2.9 (An example with OTOP) Let the vocabulary L consist of two predicates P and Q and ternary relations R_n for each $n < \omega$.

By ternary predicates $R_n(x, y, z)$ we define a decreasing chain of sets $R_n(a, b, z)$ of subsets of Q over each pair (a, b) in $P \times P$. The sets $R_0(a, b, z)$ are disjoint as the pairs

(a, b) vary. And there is exactly one element $c_n^{a,b}$ in $R_n(a, b, z)$ but not in $R_{n+1}(a, b, z)$. Thus the types $p_{ab}(x) = \{R_n(a, b, x),\ x \neq c_n^{a,b} : n < \omega\}$ can be independently omitted or realized.

The resulting elementary class is $\aleph_0$-stable but it has the maximal number of models in each infinite cardinality. Any directed graph (especially any ordering) can be coded by a structure the following way:

$$\text{there exists an edge from } x \text{ to } y \iff \exists z \bigwedge_{n<\omega} R_n(x, y, z).$$

We can study the same class $\mathbb{K}$ of structures but replace first order elementary substructure by $\preccurlyeq_{\mathbb{K}}$, elementary submodel in a fragment of $L_{\omega_1\omega}$ containing all first order formulas and the formula

$$\phi(x, y) = \exists z \bigwedge_{n<\omega} R_n(x, y, z).$$

The relation $\preccurlyeq_{\mathbb{K}}$ 'sees' the complexity caused by the formula, and the class $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ is unstable in the sense of the fragment. But this means it is also unstable as an abstract elementary class. Galois types always refine syntactic types if the submodel notion has a syntactic definition.

This example also has ENI-DOP and thus DOP. From ENI-DOP, we can define another notion $\preccurlyeq_{\mathbb{K}}$ for that class so that the new AEC is unstable but still has Löwenheim–Skolem number $\aleph_0$. Namely, let $M \preccurlyeq_{\mathbb{K}} N$ if M is an elementary substructure of N and whenever there are only finitely many z such that $M \models \bigwedge_{n<\omega} R_n(x, y, z)$, then the number of such elements z is not increased in N.

Example 2.10 (An example with DOP) Let the vocabulary L consist of predicates X_1, X_2 and P and two binary relation symbols π_1 and π_2. We define a theory in first order logic, with definable projections from P to each X_i and study the dimensions of pre-images of pairs in $X_1 \times X_2$. We require that

- the universe of a structure consists of three disjoint infinite predicates X_1, X_2 and P,
- the relations π_i determine surjective functions $\pi_i : P \to X_i$ for $i = 1, 2$, and
- for each $x \in X_1$ and $y \in X_2$ there are infinitely many $z \in P$ such that $\pi_1(z) = x$ and $\pi_2(z) = y$.

Again we get an $\aleph_0$-stable elementary class, which is $\aleph_0$-categorical but has the maximal number of models in each uncountable cardinality. Now we can code an ordering $(I, <)$ on the pairs $(x_i, y_i)_{i\in I}$ in an uncountable model so that $(x_i, y_i) < (x_j, y_j)$ if and only if the set $\{z \in P : \pi_1(z) = x_i \text{ and } \pi_2(z) = y_j\}$ is uncountable.

Furthermore, we get an unstable abstract elementary class for the same class of structures $\mathbb{K}$ as follows: strengthen $\preccurlyeq_{\mathbb{K}}$ so that $M \preccurlyeq_{\mathbb{K}} N$ implies that for all pairs (x, y) in the set $X_1 \times X_2$ of the structure M, if there are only countably many z in the set P of M such that $\pi_1(z) = x$ and $\pi_2(z) = y$, then no such z is added to the set P of the structure N. Since automorphisms must preserve the cardinalities of sets described on the right hand side of the above displayed equivalence, the class is unstable for Galois types. This notion $\preccurlyeq_{\mathbb{K}}$ does not have *finite character* (see Part II) and the new $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ has Löwenheim–Skolem number $\aleph_1$.

Similar phenomena appear in *differentially closed fields* of characteristic zero, whose elementary theory is $\aleph_0$-stable with ENI-DOP, and thus DOP. They have the maximal number of models in each infinite cardinality. See the survey articles by Marker **[39, 40]**.

The following examples exhibit the difference between a traditional first order approach and a non-elementary approach.

Example 2.11 (Exponential maps of abelian varietes) Martin Bays, Misha Gavrilovich, Anand Pillay, and Boris Zilber [**9, 10, 19**] study 'exponential maps' or 'universal group covers' $\pi\colon (\mathbb{C}^g,+) \to A(\mathbb{C})$, where $(\mathbb{C}^g,+)$ is the additive group of the complex numbers to power g and $A(\mathbb{C})$ is an abelian variety. The kernel Λ of π is a free abelian subgroup of $(\mathbb{C}^g,+)$. Two approaches appear in the work: the structures modeling the first order theory of such a map and the structures modeling the $L_{\omega_1\omega}$-theory. The $L_{\omega_1\omega}$-sentence describing the map is quasi-minimal excellent and so categorical in each uncountable cardinality. All the models of the sentence share the same Λ and are determined up to the transcendence degree of the field interpreted in $A(\mathbb{C})$. However, the first order theory is also 'classifiable', it is superstable with NDOP and NOTOP and is 'shallow', although not categorical. Each model of the first-order theory is described by choosing a lattice Λ and a transcendence degree for the field in $A(\mathbb{C})$.

In this case, the non-elementary framework was understood first; the elementary class gives a little more information. Both depend on rather deep algebraic number theory. This topic is an offshoot of trying to understand the model theory of the complex exponentiation $\exp\colon (\mathbb{C},+,\times) \to (\mathbb{C},+,\times)$, which has a very ill-behaved theory in first order logic; see [**3, 58**] for more discussion on the subject.

Example 2.12 (Valued fields) The recent book by Haskell, Hrushovski and Macpherson [**23**] greatly develops the first order model theory of *algebraically closed valued fields.* The elementary class is unstable and not even simple, and hence the structure theory has involved developing new extensions of the stability-theoretic machinery investigating the class of theories without the independence property.

A valued field consists of a field K together with a homomorphism from its multiplicative group to an ordered abelian group Γ, which satisfies the ultrametric inequality. The problems in the elementary theory of valued fields reduce to that of the value group Γ and the so called *residue field.*

However, we can study valued fields as an AEC fixing the value group as $(\mathbb{R},+,<)$ and taking all substructures as elementary substructures, requiring also that the value group stays fixed. This class is stable and contains those valued fields that are of most interest. The cases where $(\Gamma,+,<)$ is not embeddable to $(\mathbb{R},+,<)$ are often called *Krull valuations.* They are forced to be in the scope of study in the first order approach since first order logic cannot separate them from the usual ones. The non-elementary class fixing the value group can be seen as 'almost compact'; see the work of Itaï Ben Yaacov [**55**].

Acknowledgements

We thank the Mittag-Leffler Institute in Djursholm, Sweden, for providing ideal surroundings for mathematical discussions. The authors and Professor Kossak attended during the 2009 fall semester the scientific program *Mathematical Logic: Set Theory and Model Theory* at the institute. The authors wish to thank the John Templeton Foundation for its support through Project #13152, Myriad Aspects of Infinity, hosted during 2009–2011 at the Centre de Recerca Matemàtica, Bellaterra, Spain.

References

[1] H. Adler. An introduction to theories without the independence property. To appear in Archive for Mathematical Logic.

[2] J. T. Baldwin. *Fundamentals of Stability Theory.* Perspectives in Mathematical Logic. Springer-Verlag, 1988.

[3] J. T. Baldwin. The complex numbers and complex exponentiation: Why infinitary logic is necessary! 2006.

[4] J. T. Baldwin. *Categoricity*, vol. 50 of *University Lecture Series.* AMS, 2009.

[5] J. T. Baldwin, P. C. Eklof, and J. Trlifaj. ${}^{\perp}N$ as an abstract elementary class. *Annals of Pure and Applied Logic*, 149(1-3):25–39, 2007.

[6] J. T. Baldwin and A. Kolesnikov. Categoricity, amalgamation, and tameness. *Israel Journal of Mathematics*, 170:411–443, 2009.

[7] J. T. Baldwin and A. H. Lachlan. On strongly minimal sets. *The Journal of Symbolic Logic*, 36:79–96, 1971.

[8] J. Barwise and S. Feferman. *Model-Theoretic Logics.* Springer-Verlag, New York, 1985.

[9] M. Bays. Model theory of exponential maps of abelian varieties. Talk at the meeting Geometric Model Theory, 25th-28th March 2010, Oxford.

[10] M. Bays and B. Zilber. Covers of multiplicative groups of an algebraically closed field of arbitrary characteristic. Preprint: arXiv math.AC/0401301, 2004.

[11] T. Beke and J. Rosický. Abstract elementary classes and accessible categories. Preprint, 2010.

[12] I. Ben-Yaacov, A. J. Berenstein, C. W. Henson, and A. Usvyatsov. Model theory for metric structures. In Z. Chatzidakis, D. Macpherson, A. Pillay, and A. Wilkie, eds., *Model Theory with Applications to Algebra and Analysis vol.* 2, London Math Soc. Lecture Note Series 350. Cambridge University Press, Almaty, 2008.

[13] E. Bouscaren, ed. *Model Theory and Algebraic Geometry: An Introduction to E. Hrushovski's Proof of the Geometric Mordell–Lang Conjecture.* Springer-Verlag, 1999.

[14] S. Buechler. *Essential Stability Theory.* Springer-Verlag, 1991.

[15] S. Buechler and O. Lessmann. Simple homogeneous models. *Journal of the American Mathematical Society*, 16(1):91–121, 2003.

[16] C. C. Chang. Some remarks on the model theory of infinitary languages. In J. Barwise, ed., *The syntax and semantics of infinitary languages*, 36–64. Springer-Verlag, 1968.

[17] C. C. Chang and H. Keisler. *Model Theory.* North-Holland, 1973.

[18] P. C. Eklof and E. R. Fischer. The elementary theory of abelian groups. *Annals of Pure and Applied Logic*, 4:115–171, 1972.

[19] M. Gavrilovich. *Model theory of universal covering spaces of complex analytic varieties.* PhD thesis, Balliol College Oxford.

[20] R. Grossberg and B. Hart. The classification of excellent classes. *The Journal of Symbolic Logic*, 54(4):1359–1381, 1989.

[21] R. Grossberg and O. Lessmann. Shelah's stability spectrum and homogeneity spectrum in finite diagrams. *Archive of Mathematical Logic*, 41(1):1–31, 2002.

[22] B. Hart and S. Shelah. Categoricity over P for first order T or categoricity for $\phi \in \mathrm{L}_{\omega_1\omega}$ can stop at $\aleph_k$ while holding for $\aleph_0, \ldots, \aleph_{k-1}$. *Israel Journal of Mathematics*, 70:219–235, 1990.

[23] D. Haskell, E. Hrushovski, and D. Macpherson. *Stable domination and independence in algebraically closed valued fields.* Lecture Notes in Logic. Cambridge University Press, 2008.

[24] Å. Hirvonen and T. Hyttinen. Categoricity in homogeneous complete metric spaces. *Archive of Mathematical Logic*, 48(3-4):269–322, 2009.

[25] E. Hrushovski. Almost orthogonal regular types. *Annals of Pure and Applied Logic*, 45(2):139–155, 1989.

[26] T. Hyttinen and M. Kesälä. Interpreting Groups and Fields in Simple, Finitary AECs. To appear in the Journal of Symbolic Logic, 2011.

[27] T. Hyttinen and O. Lessmann. Simplicity and uncountable categoricity in excellent classes. *Annals of Pure and Applied Logic*, 139(1-3):110–137, 2006.

[28] T. Hyttinen, O. Lessmann, and S. Shelah. Interpreting groups and fields in some nonelementary classes. *Journal of Mathematical Logic*, 5(1):1–47, 2005.

[29] T. Hyttinen and S. Shelah. Strong splitting in stable homogeneous models. *Annals of Pure and Applied Logic*, 103:201–228, 2000.

[30] T. Hyttinen and S. Shelah. Main gap for locally saturated elementary submodels of a homogeneous structure. *The Journal of Symbolic Logic*, 66:1286–1302, 2001.

[31] H. J. Keisler. Logic with the quantifier "there exists uncountably many". *Annals of Mathematical Logic*, 1(1):1–93, 1970.

[32] H. J. Keisler. *Model Theory for Infinitary Logic.* Studies in Logic and the Foundations of Mathematics. North-Holland Publishing Company, 1971.

[33] J. Kennedy. Gödel and formalism freeness. Preprint, 2010.

[34] J. Kirby. Abstract elementary categories. Preprint, 2008.

[35] J. Kirby. On quasiminimal excellent classes. *The Journal of Symbolic Logic*, 75(2):551–564, 2010.

[36] O. Lessmann. An introduction to excellent classes. In A. Blass and Y. Zhang, eds., *Logic and its Applications*, vol. 380 of *Contemporary Mathematics*, 231–259. AMS, 2005.

[37] M. Lieberman. *Topological and category-theoretic aspects of abstract elementary classes.* PhD thesis, University of Michigan, 2009.

[38] P. Lindström. On extensions of elementary logic. *Theoria*, 35:1–11, 1969.

[39] D. Marker. Model theory of differential fields. In *Model Theory, Algebra, and Geometry*, 53–63. Cambridge University Press, Cambridge, 2000.

[40] D. Marker. The number of countable differentially closed fields. *Notre Dame Journal of Formal Logic*, 48(1):99–113, 2007.

[41] D. Marker, M. Messmer, and A. Pillay. *Model Theory of Fields.* Springer-Verlag, 1996.

[42] A. H. Mekler and S. Shelah. $L_{\infty\omega}$-free algebras. *Algebra Universalis*, 26:351–366, 1989.

[43] M. D. Morley. Categoricity in power. *Transactions of the American Mathematical Society*, 114:514–538, 1965.

[44] A. Mostowski. On a generalization of quantifiers. *Polska Akademia Nauk. Fundamenta Mathematicae*, 44:12–36, 1957.

[45] A. Pillay. *Geometric Stability Theory.* Oxford University Press, 1996.

[46] S. Shelah. Finite diagrams stable in power. *Annals of Mathematical Logic*, 2:293–300, 1970.

[47] S. Shelah. Categoricity of uncountable theories. In *Proceedings of the Tarski Symposium (Proc. Sympos. Pure Math., vol. XXV, University of California, Berkeley, California, 1971)*, 187–203, AMS, Providence, RI, 1974.

[48] S. Shelah. The lazy model-theoretician's guide to stability. *Logique et Analyse. Nouvelle Série*, 18(71-72):241–308, 1975.

[49] S. Shelah. *Classification Theory and the Number of Nonisomorphic Models.* North-Holland, 1978. Second Revised Edition, 1990.

[50] S. Shelah. Classification theory for for nonelementary classes, I. The number of uncountable models of $\psi \in L_{\omega_1,\omega}$. Parts A and B. *Israel Journal of Mathematics*, 46:212–273, 1983.

[51] S. Shelah. Classification of non elementary classes II, Abstract elementary classes. In J. T. Baldwin, ed. *Classification Theory, Proceedings, Chicago 1985*, 419–497. Springer-Verlag, Berlin, 1987.

[52] S. Shelah. *Classification Theory for Abstract Elementary Classes*, vol. 18 of *Studies in Logic, Mathematical Logic and Foundations.* College Publications, 2009.

[53] J. Trlifaj. Abstract elementary classes induced by tilting and cotilting modules have finite character. *Proceedings of the American Mathematical Society*, 137(3):1127–1133, 2009.

[54] F. O. Wagner. *Simple Theories.* Springer, 2000.

[55] I. B. Yaacov. Model theoretic properties of metric valued fields. arXiv:0907.4560, 2009.

[56] B. Zil'ber. Strongly minimal countably categorical theories III. *Siberian Mathematics Journal*, 25:559–571, 1984.

[57] B. Zilber. *Uncountably categorical theories*, vol. 117 of *Translations of Mathematical Monographs.* American Mathematical Society, Providence, RI, 1993.

[58] B. Zilber. Pseudo-exponentiation on algebraically closed fields of characteristic 0. *Annals of Pure and Applied Logic*, 132:67–95, 2004.

[59] B. Zilber. A categoricity theorem for quasi-minimal excellent classes. In *Logic and its Applications*, vol. 380 of *Contemporary Mathematics*, 297–306. Amer. Math. Soc., Providence, RI, 2005.

[60] B. Zilber. Covers of the multiplicative group of an algebraically closed field of characteristic zero. *Journal of the London Mathematical Society*, 74(1):41–58, 2006.

Beyond first order logic: from number of structures to structure of numbers, part II

John T. Baldwin[†], **Tapani Hyttinen**[‡], **Meeri Kesälä**[‡]

[†] Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, USA
jbaldwin@uic.edu

[‡] Department of Mathematics and Statistics, University of Helsinki, Finland
tapani.hyttinen@helsinki.fi, meeri.kesala@helsinki.fi

Abstract. The paper studies the history and recent developments in non-elementary model theory focusing in the framework of *abstract elementary classes.* We discuss the role of syntax and semantics and the motivation to generalize first order model theory to non-elementary frameworks and illuminate the study with concrete examples of classes of models.

This second part continues to study the question of categoricity transfer and counting the number of structures of certain cardinality. We discuss more thoroughly the role of countable models, search for a non-elementary counterpart for the concept of completeness and present two examples: One example answers a question asked by David Kueker and the other investigates models of Peano Arihmetic and the relation of an elementary end-extension in the terms of an abstract elementary class.

Introduction

In the article *Beyond first order logic: from number of structures to structure of numbers, part I*, we studied the basic concepts in non-elementary model theory, such as *syntax* and *semantics*, the languages $L_{\lambda\kappa}$ and the notion of a complete theory in *first order logic* (i.e., in the language $L_{\omega\omega}$), which determines an *elementary class* of structures. Classes of structures which cannot be axiomatized as the models of a first-order theory, but might have some other 'logical' unifying attribute, are called *non-elementary.*

We discussed the categoricity transfer problem and how this led to the development of a so-called stability classification. We emphasized how research questions in counting the number of models of the class in a given cardinality had led to better understanding of the structures of the class, enabled classification via invariants and found out to have applications beyond the original research field.

We mentioned two procedures for proving a categoricity transfer theorem: the *saturation transfer method* and the *dimension method.* Especially, we discussed *types* and how the question whether or how many times certain types are *realized* in a structure was essential. Here we describe how these methods have been applied for abstract elementary classes.

The study of complete sentences in $L_{\omega_1\omega}$ gives little information about countable models as each sentence is $\aleph_0$-categorical. Another approach to the study of countable models of infinitary sentences is via the study of simple finitary AEC, which are expounded in Subsection 1.1. However, while complete sentences in $L_{\omega_1\omega}$ is too strong a notion, some

Accepted for publication in 2012 in the Bulletin of the Iranian Mathematical Society.

[†]Partially supported by grant NSF-0500841.

[‡]Partially supported by the Academy of Finland, grant number 1123110.

strengthening of simple finitary AEC is needed to solve even such natural questions as, ‘When must an $\aleph_1$-categorical class have at most countably many countable models?’. In Section 2 we focus on *countable* models and study the concept of *completeness* for abstract elementary classes. Some interesting examples of models of Peano Arithmetic enliven the discussion.

1 Abstract elementary classes and Jónsson classes

We recall the definition of an abstract elementary class.

Definition 1.1 For any vocabulary τ, a class of τ-structures $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ is an *abstract elementary class* (AEC) if:

(1) Both $\mathbb{K}$ and the binary relation $\preccurlyeq_{\mathbb{K}}$ are closed under isomorphism.
(2) If $\mathcal{A} \preccurlyeq_{\mathbb{K}} \mathcal{B}$, then $\mathcal{A}$ is a substructure of $\mathcal{B}$.
(3) $\preccurlyeq_{\mathbb{K}}$ is a partial order on $\mathbb{K}$.
(4) If $\langle \mathcal{A}_i : i < \delta \rangle$ is an $\preccurlyeq_{\mathbb{K}}$-increasing chain:
  (a) $\bigcup_{i<\delta} \mathcal{A}_i \in \mathbb{K}$;
  (b) for each $j < \delta$, $\mathcal{A}_j \preccurlyeq_{\mathbb{K}} \bigcup_{i<\delta} \mathcal{A}_i$;
  (c) if each $\mathcal{A}_i \preccurlyeq_{\mathbb{K}} \mathcal{M} \in \mathbb{K}$, then $\bigcup_{i<\delta} \mathcal{A}_i \preccurlyeq_{\mathbb{K}} \mathcal{M}$.
(5) If $\mathcal{A}, \mathcal{B}, \mathcal{C} \in \mathbb{K}$, $\mathcal{A} \preccurlyeq_{\mathbb{K}} \mathcal{C}$, $\mathcal{B} \preccurlyeq_{\mathbb{K}} \mathcal{C}$ and $\mathcal{A} \subseteq \mathcal{B}$ then $\mathcal{A} \preccurlyeq_{\mathbb{K}} \mathcal{B}$.
(6) There is a Löwenheim–Skolem number $\mathrm{LS}(\mathbb{K})$ such that if $\mathfrak{A} \in \mathbb{K}$ and $B \subset \mathfrak{A}$ a subset, there is $\mathfrak{A}' \in \mathbb{K}$ such that $B \subset \mathfrak{A}' \preccurlyeq_{\mathbb{K}} \mathfrak{A}$ and $|\mathfrak{A}'| = |B| + \mathrm{LS}(\mathbb{K})$.

Abstract elementary classes arise from very different notions $\preccurlyeq_{\mathbb{K}}$, which do not necessarily have a background in some logic traditionally studied in model theory. If a class $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ is an AEC, many tools of model theory can be applied to study that class. The first essential observation is that an analog of the Chang–Scott–López-Escobar Theorem (see Theorem 2.5 in Part I) holds for any AEC. Here, purely semantic conditions on a class imply that it has a syntactic definition.

Theorem 1.2 (Shelah) *Assume that* $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ *is an abstract elementary class of* L*-structures, where* $|L| \leq \mathrm{LS}(\mathbb{K})$*. There is a vocabulary* $L' \supseteq L$ *with cardinality* $|\mathrm{LS}(\mathbb{K})|$*, a first order* L'*-theory* T *and a set* Σ *of at most* $2^{\mathrm{LS}(\mathbb{K})}$ *partial types such that* $\mathbb{K}$ *is the class of reducts of models of* T *omitting* Σ *and* $\preccurlyeq_{\mathbb{K}}$ *corresponds to the* L'*-substructure relation between the expansions of structures to* L'*.*

This theorem has interesting corollaries, since it enables us to use the tools available for *pseudo-elementary classes*: for example, we can count an upper bound for the Hanf number. To extend the notion of *Hanf number* (see Definition 1.6 in Part I) to AEC, take $\mathcal{C}$ in the definition as the collection of all abstract elementary classes for a fixed vocabulary and a fixed Löwenheim–Skolem number. (For a more general account of Hanf numbers, see [**2**, p. 32].) There is an interesting interplay between syntax and semantics: we can compute the Hanf number for AECs with a given $\mathrm{LS}(\mathbb{K})$, a semantically defined class. But the proof relies on the methods available only for an associated syntactically defined class of structures in an extended vocabulary.

The following properties of an AEC play a crucial role in advanced work:

Definition 1.3 (Amalgamation and joint embedding)

(1) We say that $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ has the *amalgamation property* (AP) if it satisfies the following: If $\mathcal{A}, \mathcal{B}, \mathcal{C} \in \mathbb{K}$, $\mathcal{A} \preccurlyeq_{\mathbb{K}} \mathcal{B}$, $\mathcal{A} \preccurlyeq_{\mathbb{K}} \mathcal{C}$ and $\mathfrak{B} \cap C = \mathfrak{A}$, there is $\mathcal{D} \in \mathbb{K}$ and a map $f\colon \mathcal{B} \cup \mathcal{C} \to \mathcal{D}$ such that $f \restriction \mathcal{B}$ and $f \restriction \mathcal{C}$ are $\mathbb{K}$-embeddings.

(2) We say that $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ has the *joint embedding property* (JEP) if for all $\mathfrak{A}, \mathfrak{B} \in \mathbb{K}$ there is $C \in \mathbb{K}$ and $\mathbb{K}$-embeddings $f\colon \mathfrak{A} \to C$ and $g\colon \mathfrak{B} \to C$.

The notion of AEC is naturally seen as a generalization of Jónsson's work in the 50's on universal and homogeneous-universal relational systems; we introduce new terminology for those AEC's close to his original notion.

Definition 1.4 (Jónsson class) An abstract elementary class is a *Jónsson class* if the class has arbitrarily large models and the joint embedding and amalgamation properties.

The models of a first order theory under elementary embedding form a Jónsson class in which complete first order type (over a model) coincides exactly with the Galois types described below and the usual notion of a monster model is the one we now explain.

A standard setting, stemming from Jónsson's [**10**] version of Fraïssé limits of classes of structures, builds a 'large enough' *monster model* $\mathfrak{M}$ (or universal domain) for an elementary class of structures via amalgamation and unions of chains. A monster model is *universal* and *homogeneous* in the sense that

- all 'small enough' structures can be elementarily embedded in $\mathfrak{M}$, and
- all *partial elementary maps* from $\mathfrak{M}$ to $\mathfrak{M}$ with 'small enough' domain extend to automorphisms of $\mathfrak{M}$.

Here 'small enough' refers to the possibility to find all structures 'of interest' inside the monster model; further cardinal calculation can be done to determine the actual size of the monster model.

The situation is more complicated for AEC. We consider here *Jónsson classes*, where we are able to construct a *monster model.* However, even then the outcome differs crucially from the monster in elementary classes, since we get only *model-homogeneity*, that is, the monster model for a Jónsson class is a model $\mathfrak{M}$ such that:

- For any 'small enough' model $M \in \mathbb{K}$ there is a $\mathbb{K}$-embedding $f\colon M \to \mathfrak{M}$.
- Any isomorphism $f\colon M \to N$ between 'small enough' $\mathbb{K}$-elementary substructures $M, N \preccurlyeq_{\mathbb{K}} \mathfrak{M}$ extends to an automorphism of $\mathfrak{M}$.

The first order case has homogeneity over sets; AEC's have homogeneity only over models.

The first problem in stability theory for abstract elementary classes is to define 'type', since now it cannot be just a collection of formulas. We note two definitions of *Galois type.*

Definition 1.5 (Galois type)

(1) For an arbitrary AEC $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ and models $M \preccurlyeq_{\mathbb{K}} N \in \mathbb{K}$, consider the following relation for triples $(\overline{a}, M, N)$, where $\overline{a}$ is a finite tuple in N:

$$(\overline{a}, M, N) \equiv (\overline{b}, M, N')$$

if there are a model $N'' \in \mathbb{K}$ and $\mathbb{K}$-embeddings $f\colon N \to N''$, $g\colon N' \to N''$ such that $f \restriction M = g \restriction M$ and $f(\overline{a}) = \overline{b}$. Take the transitive closure of this relation. The equivalence class of a tuple $\overline{a}$ in this relation, written $\mathrm{tp}^{\mathrm{g}}(\overline{a}, M, N)$, is called the *Galois type* of $\overline{a}$ in N over M.

(2) Assume that $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ is a Jónsson class and $\mathfrak{M}$ is a fixed monster model for the class. We say that the tuples $\overline{a}$ and $\overline{b}$ in $\mathfrak{M}$ have the same *Galois type* over a subset $A \subseteq \mathfrak{M}$,

$$\mathrm{tp}^{\mathrm{g}}(\overline{a}/A) = \mathrm{tp}^{\mathrm{g}}(\overline{b}/A),$$

if there is an automorphism f of $\mathfrak{M}$ fixing A pointwise such that $f(\overline{a}) = \overline{b}$.

Fruitful use of Definition 1.5(2) depends on the class having the amalgamation property over the 'parameter sets' A. Thus, even with amalgamation, there is a good notion of Galois types only over models and not over arbitrary subsets.

The monster model is λ-saturated for a 'big enough' λ. That is, all Galois types over $\preccurlyeq_{\mathbb{K}}$-elementary substructures M of size $\leq \lambda$, which are realized in some $\preccurlyeq_{\mathbb{K}}$-extension of M, are realized in $\mathfrak{M}$. When M is a $\mathbb{K}$-elementary substructure of the monster model $\mathfrak{M}$, the two notions of a Galois type $\mathrm{tp}^{\mathrm{g}}(\overline{a}, M, \mathfrak{M})$ agree. As in the first order case, the set of realization of a Galois type of $\overline{a}$ (over a model) is exactly *the orbits of the tuple* $\overline{a}$ under automorphisms of $\mathfrak{M}$ fixing the model M pointwise. That is,

$$\mathrm{tp}^{\mathrm{g}}(\overline{a}, M, \mathfrak{M}) = \mathrm{tp}^{\mathrm{g}}(\overline{b}, M, \mathfrak{M})$$

if and only if there is an automorphism f of $\mathfrak{M}$ fixing M pointwise such that $f(\overline{a}) = \overline{b}$. Furthermore, if $N \preccurlyeq_{\mathbb{K}} \mathfrak{M}$ is any $\mathbb{K}$-extension of M containing $\overline{a}$, $\mathrm{tp}^{\mathrm{g}}(\overline{a}, M, N)$ equals $\mathrm{tp}^{\mathrm{g}}(\overline{a}, M, \mathfrak{M}) \cap N$. Hence in Jónsson classes we fix a monster model $\mathfrak{M}$ and use a simpler notation for a Galois type, $\mathrm{tp}^{\mathrm{g}}(\overline{a}/M)$, which abbreviates $\mathrm{tp}^{\mathrm{g}}(\overline{a}, M, \mathfrak{M})$. Since we can also study automorphisms of $\mathfrak{M}$ fixing some *subset* A of $\mathfrak{M}$, also the notion of a Galois type over a set A becomes amenable. But over sets, the two forms are not equivalent.

The notion of Galois type lacks many properties that the compactness of first order logic guarantees for first order types. In the first order case, we can always realize a *union* of an increasing chain of types in the monster model and types have *finite character*: the types of $\overline{a}$ and $\overline{b}$ agree over a subset A if and only if they agree over every finite subset of A. Many such nice properties disappear for arbitrary Galois types. But we restrict to better-behaved Jónsson classes. Grossberg and VanDieren [**4**] isolated the concept of *tameness* that is crucial in the study of categoricity transfer for Jónsson classes.

Definition 1.6 (Tameness) We say that a Jónsson class $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ is (κ, λ)-tame for $\kappa \leq \lambda$ if the following are equivalent for a model M of size at most λ:

- $\mathrm{tp}^{\mathrm{g}}(\overline{a}/M) = \mathrm{tp}^{\mathrm{g}}(\overline{b}/M)$;
- $\mathrm{tp}^{\mathrm{g}}(\overline{a}/M') = \mathrm{tp}^{\mathrm{g}}(\overline{b}/M')$ for each $M \preccurlyeq_{\mathbb{K}} M$ with $|M'| \leq \kappa$.

Furthermore, we say that the class is *κ-tame* if it is (κ, λ)-tame for all cardinals λ and *tame* if it is $\mathrm{LS}(\mathbb{K})$-tame.

Giving up compactness also has benefits: 'non-standard structures' that realize unwanted types, which are forced by compactness, can now be avoided. For example, we might study real vector spaces in a two sorted language and demand that the reals be standard.

The first 'test question' for AECs was to ask if one can prove a categoricity transfer theorem. Shelah stated the following conjecture:

Conjecture 1.7 There exists a cardinal number κ (depending only on $\mathrm{LS}(\mathbb{K})$) such that if an AEC with a given number $\mathrm{LS}(\mathbb{K})$ is categorical in some cardinality $\lambda > \kappa$, then it is categorical in every cardinality $\lambda > \kappa$.

Shelah introduced the notion of a Jónsson class (not the name) in 1999 [**18**] and proved the following categoricity transfer result ([**2**, Part II]).

Theorem 1.8 (Shelah) *Let $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ be a Jónsson class. Then there is a calculable cardinal H_2, depending only on $\mathrm{LS}(\mathbb{K})$, such that if $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ is categorical in some cardinal $\lambda^+ > \mathrm{H}_2$, then $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ is categorical in all cardinals in the interval $[\mathrm{H}_2, \lambda^+]$.*

We remark that this almost settles the Categoricity Conjecture for Jónsson classes: for each such AEC with a fixed Löwenheim–Skolem number LS, let $\mu_{\mathbb{K}}$ be the sup (if it exists) of the successor cardinals in which $\mathbb{K}$ is categorical. Since there does not exist a proper class of such AECs, there is a supremum for such $\mu_{\mathbb{K}}$. Denote this number by $\Lambda(\mathrm{LS})$. Now if a Jónsson class with Löwenheim–Skolem number LS is categorical in some successor cardinal $\lambda > \mu = \sup(\Lambda(\mathrm{LS}), \mathrm{H}_2)$, it is categorical in all cardinals in $[\mathrm{H}_2, \lambda^+]$, and in arbitrarily large successor cardinals, and hence in all cardinals above H_2. Two problems remain in this area. Remove the restriction to successor cardinals in Theorem 1.8; this would avoid the completely non-effective appeal to $\Lambda(\mathrm{LS})$. Make a more precise calculation of the cardinal H_2 in the successor case ([**2**, Problem D.1.5]).

Shelah proves a downward categoricity transfer theorem and also shows categoricity for $\lambda^+ > H_2$ implies certain kind of 'tameness' for Galois types over models of size $\leq H_2$, which enables the transfer of categoricity up to all cardinals in the interval $[\mathrm{H}_2, \lambda^+]$. Grossberg and VanDieren separated out the upward categoricity transfer argument, and realized that tameness was the only additional condition needed to transfer categoricity arbitrarily high. The downward step uses the *saturation transfer* method, where saturation is with respect to *Galois types*; the upwards induction uses the *dimension method.*

Theorem 1.9 (Grossberg and VanDieren) *Assume that a χ-tame Jónsson class* $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ *is categorical in λ^+, where $\lambda > \mathrm{LS}(\mathbb{K})$ and $\lambda \geq \chi$. Then* $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ *is categorical in each cardinal $\geq \lambda^+$.*

Lessmann [**15**] extends the result to $\mathrm{LS}(\mathbb{K})^+$-categoricity in the case $\mathrm{LS}(\mathbb{K}) = \aleph_0$. The restriction to countable Löwenheim cardinal number reflects a significant combinatorial obstacle. In these two results the categoricity transfer is only from successor cardinals and the proof is essentially an induction on dimension. In Subsection 1.1 we discuss further use of the saturation transfer method for simple, finitary AECs by Hyttinen and Kesälä in [**11**].

1.1 Simple finitary AECs

Simple finitary AECs were defined particularly to study independence and stability theory in a framework without compactness. The idea was both to find a common extension for homogeneous model theory and the study of excellent sentences in $L_{\omega_1\omega}$ (see Part I) and also clarify the 'core' properties which support a successful dimension theory. The property *finite character* is essential for this analysis.

Definition 1.10 (Finite character) We say that $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ has *finite character* if for any two models $\mathfrak{A}, \mathfrak{B} \in \mathbb{K}$ such that $\mathfrak{A} \subseteq \mathfrak{B}$ the following are equivalent:

(1) $\mathfrak{A} \preccurlyeq_{\mathbb{K}} \mathfrak{B}$.

(2) For every finite sequence $\overline{a} \in \mathfrak{A}$ there is a $\mathbb{K}$-embedding $f \colon \mathfrak{A} \to \mathfrak{B}$ such that $f(\overline{a}) = \overline{a}$.

Definition 1.11 (Finitary AEC) An abstract elementary class is *finitary* if it is a Jónsson class with countable Löwenheim–Skolem number that has finite character.

Definition 1.11 slightly modifies Hyttinen–Kesälä [**6**]; in particular the formulation of finite character is from Kueker [**14**]. Elementary classes are finitary AECs. However, a class defined by an arbitrary sentence in $L_{\omega_1\omega}$, the relation $\preccurlyeq_{\mathbb{K}}$ being the one given by the corresponding fragment, may not have AP, JEP or even arbitrarily large models. A relation $\preccurlyeq_{\mathbb{K}}$ given by any fragment of $L_{\infty\omega}$ will have finite character. Most abstract elementary classes definable in $L_{\omega_1\omega}(Q)$ do not have finite character. An easy example of a class without finite character, due to Kueker [**14**], is a class of structures with a countable predicate P, where $M \preccurlyeq_{\mathbb{K}} N$ if and only if $M \subseteq N$ and $P(M) = P(N)$.

The notion of weak type is just Galois type with built-in finite character: two tuples $\overline{a}$ and $\overline{b}$ have the same *weak type* over a set A, written

$$\mathrm{tp}^{\mathrm{w}}(\overline{a}/A) = \mathrm{tp}^{\mathrm{w}}(\overline{b}/A),$$

if they have the same Galois type over each finite subset $A' \subseteq A$. Furthermore, we say that a model M is *weakly saturated* if it realizes all weak types over subsets of size $< M$.

Basic stability theory with a categoricity transfer result for simple finitary AEC's is carried out in the papers [**5, 6, 7**]. However, some parts of the theory hold also for arbitrary Jónsson classes; this is expounded in [**9**]. Kueker [**14**] has clarified when AEC admit syntactic definitions and particularly the connection of finite character to definability in $L_{\infty\omega}$ definablity of AEC's; unlike in Theorem 1.2, no extra vocabulary is needed for these results.

Theorem 1.12 (Kueker) *Assume that* $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ *is an abstract elementary class such that* $\mathrm{LS}(\mathbb{K}) = \kappa$. *Then:*

(1) *The class* $\mathbb{K}$ *is closed under* $L_{\infty\kappa^+}$*-elementary equivalence.*

(2) *If* $\mathrm{LS}(\mathbb{K}) = \aleph_0$ *and* $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ *contains at most* λ *models of cardinality* $\leq \lambda$ *for some cardinal* λ *such that* $\lambda^\omega = \lambda$, *then* $\mathbb{K}$ *is definable with a sentence in* $L_{\lambda^+\omega_1}$.

(3) *If* $\kappa = \aleph_0$ *and* $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ *has* finite character, *the class is closed under* $L_{\infty\omega}$*-elementary equivalence.*

(4) *Furthermore, if* $\kappa = \aleph_0$, $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ *has finite character and at most* λ *many models of size* $\leq \lambda$ *for some infinite* λ, $\mathbb{K}$ *is definable with sentence in* $L_{\lambda^+\omega}$.

The notion of an *indiscernible sequence* of tuples further illustrates the distinction between the syntactic and semantic viewpoint. Classically a sequence is indiscernible if each increasing n-tuple of elements realize the same (syntactic) type. In AEC, a sequence $(\overline{a}_i)_{i<\kappa}$ is *indiscernible* over a set A (or A*-indiscernible*) if the sequence can be extended to any 'small enough' length $\kappa' > \kappa$ so that any order-preserving partial permutation of the larger sequence extends to an automorphism of the monster model fixing the set A.

Note that two tuples lying on the same A-indiscernible sequence is a much stronger condition than two tuples having the same Galois type over A. However, 'lying on the same sequence' is not a transitive relation and hence not an equivalence relation; the notion of *Lascar strong type* is obtained by taking the transitive closure of this relation.

Using indiscernible sequences we can define a notion of *independence* based on *Lascar splitting.*[1] Furthermore, we say that the class is *simple* if this notion satisfies that each type is independent over its domain. Under further stability hypotheses (both $\aleph_0$-stability

[1] The notions are defined 'for weak types' since they are preserved under the equivalence of weak types.

Definition 1.13 (Independence) A type $\mathrm{tp}^{\mathrm{w}}(\overline{a}/A)$ *Lascar-splits* over a finite set $E \subseteq A$ if there is a strongly indiscernible sequence $(\overline{a}_i)_{i<\omega}$ such that $\overline{a}_0, \overline{a}_1$ are in the set A but

$$\mathrm{tp}^{\mathrm{w}}(\overline{a}_0/E \cup \overline{a}) \neq \mathrm{tp}^{\mathrm{w}}(\overline{a}_1/E \cup \overline{a}).$$

We write that a set B is *independent* of a set C over a set A, written $B \downarrow_A C$, if for any finite tuple $\overline{a} \in B$ there is a finite set $E \subseteq A$ such that for all sets D containing $A \cup C$ there is $\overline{b}$ realizing the type $\mathrm{tp}^{\mathrm{w}}(\overline{a}/A \cup C)$ such that $\mathrm{tp}^{\mathrm{w}}(\overline{b}/D)$ does *not* Lascar-split over E.

[**5, 6**] and superstability [**7, 9**] have been developed) we get an *independence calculus* for subsets of the monster model. Unlike in elementary stability theory, stability or even categoricity does not imply simplicity; it is a further assumption. However, we show that if *any* reasonable independence calculus exists for arbitrary sets and not just over models, the class must be simple and the notion of independence must agree with the one defined by Lascar splitting; see [**5**].

The saturation transfer method was further analyzed for simple, finitary AECs by Hyttinen and Kesälä in [**11**]. It was noted there that, even without tameness, *weak saturation* transfers between different uncountable cardinalities. Assuming simplicity, they developed much of the stability theoretic machinery for these classes and hence were able to remove the assumption in Theorems 1.8 and 1.9 that the categoricity cardinal is a successor.

Theorem 1.14 *Assume that* $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ *is a simple finitary AEC,* $\kappa > \omega$*, and each model of size* κ *is weakly saturated. Then:*

(1) *For any* $\lambda > \min\{(2^{\aleph_0})^+), \kappa\}$*, each model of size* λ *is weakly saturated.*

(2) *Furthermore, each uncountable* $\aleph_0$*-saturated model is weakly saturated.*

If in addition $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ *is* $\aleph_0$*-tame, all weakly saturated models with a common cardinality are isomorphic.*

What then is the role of finite character of $\preccurlyeq_{\mathbb{K}}$? If it happens that there are only countably many Galois types over any finite set (this holds for example if the class is $\aleph_0$-stable), the finite character property provides a 'finitary' sufficient condition for a substructure M of $\mathfrak{M}$ to be in $\mathbb{K}$: If all Galois types over finite subsets are realized in M, M is back-and forth-equivalent to an $\aleph_0$-saturated $\mathbb{K}$-elementary substructure N of $\mathfrak{M}$ with $|N| = |M|$; a chain argument and finite character give that $N \approx M$. Even without the condition on the number of Galois types, finite character enables many constructions involving building models from finite sequences. It implies, for example, that under simplicity and superstability, two tuples with the same *Lascar type* over a countable set can be mapped to each other by an automorphism fixing the set (i.e., they have the same Galois type over the set); see [**9**]. These Lascar types (also called *weak Lascar strong types*) are a major tool in geometric stability theory for finitary classes [**8**], since they have finite character.

2 Countable models and completeness

We recall that a theory T in the first order logic $L_{\omega\omega}$ is said to be *complete* if for any sentence $\phi \in L_{\omega\omega}$ either ϕ or its negation can be deduced from T.

A famous open conjecture for elementary classes was stated by Vaught in [**21**]:

Conjecture 2.1 (Vaught) The number of countable models of a countable and complete first order theory must be either countable or $2^{\aleph_0}$.

The conjecture can be resolved by the continuum hypothesis, which is independent of the axioms of set theory: If there is no cardinality between $\aleph_0$ and $2^{\aleph_0}$, the conjecture is trivially true. The problem is to determine the value in ZFC. Morley [**17**] proved the most significant result: not just for first order theories but for any sentence of $L_{\omega_1\omega}$ the number of countable models is either $\leq \aleph_1$ or $2^{\aleph_0}$. He used a combination of descriptive set theoretic and model theoretic techniques. There has been much progress using descriptive set theory. The study of this conjecture has also led to many new innovations in model theory: a positive solution for $\aleph_0$-stable theories was shown by Harrington, Makkai and Shelah in [**19**] and a more general positive solution for superstable theories of finite rank by Buechler in [**3**]. However, the full conjecture is still open. The article [**1**] provides connections with the methods of this paper.

An easier question for elementary classes is the number of countable models of a theory, which has only one model, up to isomorphism, in some *uncountable* power. Morley [**16**] showed that the number of countable models of an uncountably categorical elementary class must be countable. We consider as a useful 'motivating question':

Question 2.2 Must an AEC categorical in $\aleph_1$ or in some uncountable cardinal have only countably many countable models?

As asked, the answer is opposite to the first order case. For example, we can define a sentence ψ in $L_{\omega_1\omega}$ as a disjunct of two sentences, one totally categorical and one having uncountably many countable models but no uncountable models. This problem does not occur in the first order case because categoricity implies completeness. $L_{\omega_1\omega}$ poses two difficulties to this approach. First, deducing completeness from categoricity is problematic; there are several completions. Secondly, $L_{\omega_1\omega}$-completeness is too strong; it implies $\aleph_0$ categoricity and there are interesting $\aleph_1$-categorical sentences that are not $\aleph_0$-categorical. But sentences like ψ lack 'good' semantic properties such as joint embedding. We might ask a further question: Are there some semantic properties that allow the dimensional analysis of the Baldwin–Lachlan proof for an abstract elementary class? For example, does the question have a negative answer for, say, finitary AECs? (See Subsection 2.1.) What can we say on the number of countable models in different frameworks? Some results and conjectures were stated for *admissible* infinitary logics already by Kierstead in 1980 [**12**].

For a non-elementary class with a better toolbox for dimension-theoretic considerations it might be possible to say more on such questions. For example, *excellent* sentences of $L_{\omega_1\omega}$ have a well-behaved model theory; but such sentences are *complete*, so their countable model is unique up to isomorphism. An essential benefit of the approach of *finitary abstract elementary classes* is that the framework also enables the study of incomplete sentences of $L_{\omega_1\omega}$. The Vaught conjecture is false for finitary abstract elementary classes: Kueker [**14**] gives an example, well-orders of length $\leq \omega_1$, where $\preccurlyeq_{\mathbb{K}}$ is taken as end-extension. This example has exactly $\aleph_1$ many countable models. The example is categorical in $\aleph_1$, but is not a finitary AEC since it does not have arbitrarily large models. However, we can transform the example to a finitary AEC, by adding a sort with a totally categorical theory; but we lose categoricity.

Contrast the semantic and syntactic approach. If we require definability in some specific language, $L_{\omega\omega}$ or $L_{\omega_1\omega}$, the Vaught conjecture is a hard problem, but it has an 'easy' solution under the 'semantic' requirements we have suggested, such as, a finitary AEC. Is there a similar difference for Question 2.2, maybe in the opposite direction? Kueker had a special reason for asking Question 2.2 for *finitary* AECs. Recall that, by Theorem 1.12(4), if $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ is an AEC with finite character, $\mathrm{LS}(\mathbb{K}) = \aleph_0$, and $\mathbb{K}$ contains at most λ models of cardinality $\leq \lambda$, then it is definable in $L_{\lambda^+\omega}$. Hence if $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ is $\aleph_1$-categorical and has only countably many countable models, it is definable in $L_{\omega_1\omega}$. But under what circumstances can we gain this? Clearly if $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ is $\aleph_0$-categorical, this holds. Kueker asks the following, refining Question 2.2:

Question 2.3 (Kueker) Does categoricity in some uncountable cardinal imply that a finitary AEC $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ is definable with a sentence in $L_{\omega_1\omega}$?

Answering the following question positively would suffice:

Question 2.4 (Kueker) Does categoricity in some uncountable cardinal imply that a finitary AEC $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ has only countably many countable models?

Unfortunately, Example 2.5 gives a negative answer to Question 2.4, leaving the first question open.

Kueker's results illuminate the distinction between semantic and syntactic properties. Abstract elementary classes were defined with only semantic properties in mind; Kueker provides additional semantic conditions which imply definability in a specific syntax. Thus, the ability to choose a notion of $\preccurlyeq_{\mathbb{K}}$ for an AEC to make it finitary has definability consequences. The concept of *finite character* concerns the relation $\preccurlyeq_{\mathbb{K}}$ between the models in an AEC; Kueker's results conclude definability for the class $\mathbb{K}$ of structures. He does prove some, but remarkably weaker, definability results without assuming finite character.

2.1 An example answering Kueker's second question

The following example is a simple, finitary AEC, which is categorical in each uncountable power but has uncountably many countable models. Hence the example gives a negative answer to Kueker's second question.

Example 2.5 We define a language $L = \{Q, (P_n)_{n<\omega}, E, f)$, where Q and P_n are unary predicates, E is a ternary relation and f is a unary function. We consider the following axiomatization in $L_{\omega_1\omega}$:

(1) The predicates Q and $\langle P_n : n < \omega\rangle$ partition the universe.
(2) Q has at most one element.
(3) If $E(x, y, z)$ then $x \in Q$ and z, y are not in Q.
(4) If Q is empty, we have that for each $n < \omega$, $|P_{n+1}| \leq |P_n| + 1$.
(5) If P_0 is nonempty, then Q is nonempty.
(6) For all $x \in Q$, the relation $E(x, -, -)$ is an equivalence relation where each class intersects each P_n exactly once.
(7) $f(x) = x$ for all $x \in Q$ and $y \in P_n$ implies $f(y) \in P_{n+1}$.
(8) f is one-to-one.
(9) For $x \in Q$, $y \in P_n$ and $z \in P_{n+1}$, $E(x, y, z)$ if and only if $f(y) = z$.

Now we define the class $\mathbb{K}$ to be the L-structures satisfying the axioms above and the relation $\preccurlyeq_{\mathbb{K}}$ to be the substructure relation.

The example has two kinds of countable models. When there is no element in Q, the predicate P_n may have at most n elements, and either $|P_{n+1}| = |P_n|$ or P_{n+1} is one element larger. If any P_n has more than n elements, the predicate Q gets an element. When there is an element x in Q, all predicates P_n have equal cardinality, since the relation $E(x, -, -)$ gives a bijection between the predicates.

Thus we can characterize the countable models of $\mathbb{K}$: There are countably many models with nonempty Q: one where each P_n is countably infinite and one where each P_n has size k for $1 \leq k < \omega$. If Q is empty, the model is characterized by a function $f\colon \omega \to \{0, 1\}$ so that $f(n) = 1$ if and only if $|P_{n+1}| > |P_n|$. Hence there are $2^{\aleph_0}$ countable models.

This example is an AEC with $\mathrm{LS}(\mathbb{K}) = \aleph_0$. The key to establish closure under unions of chains is to note that if the union of a chain has a nonempty Q, some model in the chain must already have one. This example clearly has finite character, joint embedding and arbitrarily large models. Furthermore, the class is categorical in all uncountable cardinals.

We prove that the class has amalgamation. For this, let M, M' and M'' be in $\mathbb{K}$ such that M' and M'' extend M. We need to amalgamate M' and M'' over M. The case where $Q(M)$ is nonempty is easier and we leave it as an exercise. Hence we assume that $Q(M)$ is empty. By taking isomorphic copies if necessary we may assume that the intersection $P_n(M'') \cap P_m(M')$ is $P_n(M)$ for $n = m$ and empty otherwise. Furthermore, we extend both M' and M'' if necessary so that $Q(M')$ and $Q(M'')$ become nonempty and each $P_n(M')$ and $P_n(M'')$ become infinite. We amalgamate as follows: For two elements $x \in P_n(M')$ and $y \in P_n(M'')$, if there is $k < \omega$ such that $f^k(x) = f^k(y)$ in $P_{n+k}(M)$, then we identify x and y. Otherwise, we take a disjoint union.

We prove that the class is simple. For this, define the following notion of independence for A, B, C subsets of the monster model:

$$A \downarrow_C B \iff \begin{array}{l}\text{For any } a \in A, b \in B, \text{ if we have that } E(x,a,b),\\ \text{then there is } c \in C \text{ with } E(x,a,c).\end{array}$$

This notion satisfies invariance, monotonicity, finite character, local character, extension, transitivity, symmetry and uniqueness of free extensions. Furthermore, $\overline{a} \not\downarrow_C B$ if and only if for some $D \supseteq B$ and every $\overline{b} \models \mathrm{tp}^{\mathrm{w}}(\overline{a}/C \cup B)$, the type $\mathrm{tp}^{\mathrm{w}}(\overline{b}/D \cup C)$ Lascar-splits over C. Hence the notion is the same as the independence notion defined for general finitary AECs. This ends the proof.

We can divide this AEC into two disjoint subclasses, both of which are AECs with the same Löwenheim–Skolem number. The class of models where there is no element in Q has uncountably many countable models and is otherwise 'badly-behaved'; all models are countable and the amalgamation property fails. However, the class of models where Q is nonempty is an uncountably categorical finitary AEC with only countably many countable models. This resembles the example of the sentence in $L_{\omega_1\omega}$, mentioned in the beginning of this section, which was a disjunction of two sentences, a totally categorical one and one with uncountably many countable models and no uncountable ones. Is this 'incompleteness' the reason for categoricity not implying countably many countable models? Can we obtain the conjecture if we require the AEC to be somehow 'complete'? These concepts and questions are explored in the next section.

Jonathan Kirby recently suggested another example with similar properties. This example might feel more natural to some readers, since it consists of 'familiar' structures.

Example 2.6 Let $\mathbb{K}$ be the class of all fields of characteristic 0 which are either algebraically closed or (isomorphic to) subfields of the complex algebraic numbers $\mathbb{Q}^{\mathrm{alg}}$. Let $\preccurlyeq_{\mathbb{K}}$ be the substructure relation. Then $\mathbb{K}$ is categorical in all uncountable cardinalities and has $2^{\aleph_0}$ countable models which all embed in the uncountable models. Also $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ is a simple finitary AEC. Further, this class can be divided into smaller AEC's. For example, we can take all algebraically closed fields of characteristic 0, *except* those isomorphic to subfields of $\mathbb{Q}^{\mathrm{alg}}$ as one class and all fields isomorphic to a subfield of $\mathbb{Q}^{\mathrm{alg}}$ as the other.

2.2 Complete, irreducible and minimal AECs

We define several concepts to describe the 'completeness' or 'incompleteness' of an abstract elementary class. A nonempty collection $\mathbb{C}$ of structures of an AEC $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ is a *sub-AEC* of $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ if

- $\mathbb{C}$ is an abstract elementary class with $\preccurlyeq_{\mathbb{C}} = \preccurlyeq_{\mathbb{K}} \cap \mathbb{C}^2$;
- $\mathrm{LS}(\mathbb{K}) = \mathrm{LS}(\mathbb{C})$, that is, the Löwenheim–Skolem numbers are the same.

This allows both 'extreme cases' that $\mathbb{C}$ is $\mathbb{K}$ or that $\mathbb{C}$ consists of only one structure, up to isomorphism. The latter can happen if the only structure in $\mathbb{C}$ is of size $\mathrm{LS}(\mathbb{K})$ and is not isomorphic to a proper $\preccurlyeq_{\mathbb{K}}$-substructure of itself.

Definition 2.7 (Minimal AEC) We say that an AEC is *minimal* if it does not contain a proper sub-AEC.

Definition 2.8 (Irreducible AEC) We say that an AEC $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ is *irreducible* if there are no two proper sub-AECs $\mathbb{C}_1$ and $\mathbb{C}_2$ of $\mathbb{K}$ such that $\mathbb{C}_1 \cup \mathbb{C}_2 = \mathbb{K}$.

Definition 2.9 (Complete AEC) We say that an AEC $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ is *complete* if there are no two sub-AECs $\mathbb{C}_1$ and $\mathbb{C}_2$ of $\mathbb{K}$ such that $\mathbb{C}_1 \cup \mathbb{C}_2 = \mathbb{K}$ and $\mathbb{C}_1 \cap \mathbb{C}_2 = \emptyset$.

Example 2.5 is not complete, not irreducible and not minimal. The sub-AEC of Example 2.5, which contains the models where Q is nonempty, is also not complete: One abstract elementary class can be formed by taking all such models where each P_n is of equal size $\leq M$ for some finite M, and the rest of the models of the class form another AEC.

We make a few remarks that follow from the definitions.

Remark 2.10

(1) Minimality implies irreducibility, which implies completeness.
(2) Minimality implies the joint embedding property for models of size $\mathrm{LS}(\mathbb{K})$.
(3) Completeness and the amalgamation property imply joint embedding.
(4) If T is a complete first order theory, then the elementary class of models of T is not necessarily complete in the sense above.

Item 1 is obvious. Item 2 holds, since if there are a pair M_0, M_1 of models in $\mathbb{K}$ with size $\mathrm{LS}(\mathbb{K})$, which do not have a common extension, those structures of $\mathbb{K}$ which $\mathbb{K}$-embed M_0 form a proper sub-AEC. For item 3, note that if the class has the amalgamation property, the following classes are disjoint sub-AECs:

$$\{M \in \mathbb{K} : M \text{ can be jointly embedded with } M_0\}$$

and

$$\{M \in \mathbb{K} : M \text{ cannot be jointly embedded with } M_0\}.$$

Furthermore, the amalgamation property gives that joint embedding for models of size $\mathrm{LS}(\mathbb{K})$ implies joint embedding for all models. Note that an $\aleph_1$- but not $\aleph_0$-categorical countable first order theory is not complete as an AEC.

Example 2.5 has joint embedding and amalgamation but is not complete nor minimal, hence the implications of items 2 and 3 are not reversible. Is one or both of the implications of item 1 of Remark 2.10 reversible? If $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ is an $\aleph_0$-stable elementary class which is not $\aleph_0$-categorical, the class of $\aleph_0$-saturated models of T is a proper sub-AEC, so the class is not minimal. Example 2.18 below gives a class which is complete but not irreducible, minimal or $\aleph_0$-categorical. However, this example is not finitary: it does not have finite character or even arbitrarily large models.

To discuss the relationship between minimality and LS($\mathbb{K}$)-categoricity, it is important to specify the meaning of LS($\mathbb{K}$)-categoricity. We define an AEC to be LS($\mathbb{K}$)-categorical if it has only one model up to isomorphism, of size *at most* LS($\mathbb{K}$). We have forbidden smaller models because models of an AEC which are strictly smaller than the number LS($\mathbb{K}$) can cause quite irrational and one could say insignificant changes to the class. We could add, say, one finite model which is not embeddable in any member of the class; this would give non-minimality, since the one model constitutes an AEC. However, an AEC with one model of size LS($\mathbb{K}$) and no smaller models is automatically minimal: For any sub-AEC $\mathbb{K}'$ we can show by induction on the size of the models in $\mathbb{K}$, using the union and Löwenheim–Skolem axioms, that all models of $\mathbb{K}$ are actually contained in $\mathbb{K}'$.

Here are some further questions:

Question 2.11

(1) If an AEC is uncountably categorical and *complete*, can it have uncountably many countable models?
(2) Is there a minimal AEC which is not LS($\mathbb{K}$)-categorical?
(3) Is there an irreducible AEC which is not minimal?

2.3 An example of models of Peano Arithmetic: completeness does not imply irreducibility

In this section we present an example of a class of models of Peano Arithmetic suggested by Roman Kossak. The example shows that completeness does not imply irreducibility. The properties of the class are from Section 1.10 and Chapter 10 of the book *The Structure of Models of Peano Arithmetic* [**13**].

A model M of Peano Arithmetic (PA) is *recursively saturated* if for all finite tuples $\overline{b} \in M$ and recursive types $p(v, \overline{w})$, if $p(v, \overline{b})$ is finitely realizable then $p(v, \overline{b})$ is realized in M. Clearly an elementary union of recursively saturated models is recursively saturated. For M, a nonstandard model of PA, define $SSy(M)$, the *standard system* of M, as follows:

$$SSy(M) = \{X \subseteq \mathbb{N} : \exists Y \text{ definable in } M \text{ such that } X = Y \cap \mathbb{N}\}.$$

Lemma 2.12 ([**13**, Proposition 1.8.1]) *Let N, M be two recursively saturated models of Peano Arithmetic. Then $M \equiv_{\infty\omega} N$ if and only if $M \equiv N$ and $SSy(M) = SSy(N)$.*

It follows that any countable recursively saturated elementary end-extension of a recursively saturated M is isomorphic to M.

We say that $N \models PA$ is *ω_1-like* if it has cardinality $\aleph_1$ and every proper initial segment of N is countable. We say that $N \models PA$ is an *elementary cut* in M if M is an elementary end-extension of N.

Theorem 2.13 ([**13**, Corollary 10.3.3]) *Every countable recursively saturated model $M \models PA$ has $2^{\aleph_1}$ pairwise non-isomorphic recursively saturated ω_1-like elementary end-extensions.*

The following abstract elementary class $(\mathbb{K}, \preccurlyeq_{\mathbb{K}})$ has one countable model, $2^{\aleph_1}$ models of size $\aleph_1$ and no bigger models. We will use it to generate the counterexample.

Example 2.14 Let M be a countable recursively saturated model of Peano Arithmetic. Let $\mathbb{K}$ be the smallest class, closed under isomorphism, containing M and all ω_1-like recursively saturated elementary end-extensions of M. Let $\preccurlyeq_{\mathbb{K}}$ be elementary end-extension.

Lemma 2.15 *The AEC of Example* 2.14 *does not have finite character.*

Proof. Let M be a recursively saturated countable model of PA. Let M' be a recursively saturated elementary substructure of M (not necessarily a cut) and let $\overline{a}$ be a finite tuple in M'. We construct a $\preccurlyeq_{\mathbb{K}}$-map $f\colon M' \to M$ fixing $\overline{a}$. When M' is not a cut we contradict finite character. For this, we will find an elementary cut M'' of M and an isomorphism $f\colon M' \to M''$ such that $f(\overline{a}) = \overline{a}$. Since M and M' are recursively saturated, both $(M, \overline{a})$ and $(M', \overline{a})$ are recursively saturated. Furthermore, $(M, \overline{a})$ is elementarily equivalent to $(M', \overline{a})$. Now let M'' be an elementary cut in M such that $(M, \overline{a})$ is an elementary end-extension of $(M'', \overline{a})$ and $(M'', \overline{a})$ is recursively saturated. Then $(M', \overline{a}) \cong (M'', \overline{a})$. □

From now on, let M be a fixed countable recursively saturated model of PA.

Now we construct a complete but not irreducible AEC. Let $\prec_{\text{end}}$ denote elementary end-extension. We define

$$M(a) = \bigcap \{K \prec_{\text{end}} M : a \in K\},$$

$$M[a] = \bigcup \{K \prec_{\text{end}} M : a \notin K\},$$

where $M[a]$ can be empty. Then let $\operatorname{gap}(a)$ denote $M(a) \setminus M[a]$.

It is easy to see that an equivalent definition is the following: Let $\mathcal{F}$ be the set of definable functions $f\colon M \to M$ for which $x < y$ implies $x \leq f(x) \leq f(y)$. Let a be an element in M. Then $\operatorname{gap}(a)$ in M is the smallest subset C of M containing a such that whenever $b \in C$, $f \in \mathcal{F}$ and $b \leq x \leq f(b)$ or $x \leq b \leq f(x)$, then $x \in C$.

We say that $N \models PA$ is *short* if it is of the form $N(a)$ for some $a \in N$. Equivalently, N has a *last gap.* A short model $N(a)$ is not recursively saturated, since it omits the type

$$p(v, a) = \{t(a) < v : t \text{ a Skolem term}\}.$$

If N is not short, it is called *tall.* The following three properties are exercises in [**13**].

(1) The union of any ω-chain of end-extensions of short elementary cuts in M is tall.
(2) Any tall elementary cut in M is recursively saturated and thus isomorphic to M.
(3) If K is an elementary cut in M and is *not* recursively saturated, then $K = M(a)$ for some $a \in M$.

It also follows that the union of any ω-chain of elementary end-extensions of models isomorphic to short elementary cuts in M is isomorphic to M. For the following theorem, see [**20**].

Theorem 2.16 *Two short elementary cuts $M(a)$ and $M(b)$ are not isomorphic if and only if the sets of complete types realised in* $\operatorname{gap}(a)$ *and* $\operatorname{gap}(b)$ *respectively are* disjoint. *There are countably many pairwise non-isomorphic short elementary cuts in M.*

Lemma 2.17 *If $a \notin M(0)$, the model $M(a)$ is isomorphic to some proper initial segment $M(a')$ of $M(a)$, which is an elementary cut of $M(a)$.*

Proof. Define the recursive type

$$p(x, a) = \{\phi(x) \leftrightarrow \phi(a) : \phi(x) \in L\} \cup \{t(x) < a : t \text{ is a Skolem term}\}.$$

Any finite subset of $\operatorname{tp}(a/\emptyset)$ is realized in $M(0)$ since $M(0) \prec M$. Thence $p(x, a)$ is consistent as $M(0)$ is closed under the Skolem terms. Let $a' \in M$ realize $p(x, a)$. Then $\operatorname{tp}(a') = \operatorname{tp}(a)$ and $M(a') < a$. Hence $M(a)$ is isomorphic to $M(a')$ by Theorem 2.16. Furthermore, $M(a')$ is an elementary cut in $M(a)$. □

Lemma 2.17 implies that elementary $\prec_{\text{end}}$-chains can be formed from isomorphic copies of one $M(a)$, when $a \notin M_0$. Hence, each of the following classes $\mathbb{K}_\alpha$ is an abstract elementary class extending the $\aleph_0$-categorical class $\mathbb{K}$ from Example 2.14 and $\mathbb{K}_\alpha$ has α many countable models, where $\alpha \in \omega \cup \{\omega\}$.

Example 2.18 Let α be a finite number or ω. Choose $(M(a_i))_{i<\alpha}$ to be pairwise non-isomorphic short elementary cuts in M, where each a_i is non-standard. Let $\mathbb{K}_\alpha$ be the smallest class, closed under isomorphism, containing $\mathbb{K}$ and $M(a_i)$ for all $1 \leq i < \alpha$. Let $\preccurlyeq_{\mathbb{K}}$ be elementary end-extension.

The countable models of $\mathbb{K}_\alpha$ are exactly M and $M(a_i)$ for $1 \leq i < \alpha$. This class is closed under $\preccurlyeq_{\mathbb{K}}$-unions: if $\langle M_j,\, j < \beta\rangle$ is a $\preccurlyeq_{\mathbb{K}}$-chain of models in $\mathbb{K}_\alpha$, we have that for every countable limit ordinal β, $\bigcup_{j<\beta} M_\beta$ is tall and hence isomorphic to M, and if β is uncountable, the union is isomorphic to some ω_1-like recursively saturated model in $\mathbb{K}$. (Note that the union is also an end-extension of M.)

Any abstract elementary class containing a short elementary cut $M(a)$ for some $a \in M$ must contain M, as M is a union of models isomorphic to $M(a)$ elementarily end-extending each other. Hence any abstract elementary class containing $M(a)$ contains M.

It follows that $\mathbb{K}_\alpha$ is complete since it has no disjoint sub-AECs. Furthermore, the class $\mathbb{K}_\alpha$ is not irreducible for $\alpha > 2$, since we can divide it into two classes, one containing $M(a_i)$ but not $M(a_j)$ and one vice versa, for any $i \neq j < \alpha$.

However, Example 2.18 is neither a Jónsson class (all models have cardinality below the continuum) nor a finitary AEC. We ask:

Question 2.19 Is there a Jónsson class which is complete but not irreducible or minimal? Furthermore, is there such a finitary AEC?

Acknowledgements

We thank the Mittag-Leffler Institute in Djursholm, Sweden, for providing ideal surroundings for mathematical discussions. The authors and Professor Kossak attended during the 2009 fall semester the scientific program *Mathematical Logic: Set Theory and Model Theory* at the Institute. The authors wish to thank the John Templeton Foundation for its support through Project #13152, Myriad Aspects of Infinity, hosted during 2009–2011 at the Centre de Recerca Matemàtica, Bellaterra, Spain.

References

[1] J. T. Baldwin. The Vaught conjecture: do uncountable models count? *Notre Dame Journal of Formal Logic*, 48(1):79–92, 2007.

[2] J. T. Baldwin. *Categoricity*, vol. 50 of University Lecture Series. AMS, 2009.

[3] S. Buechler. Vaught's conjecture for superstable theories of finite rank. *Annals of Pure and Applied Logic*, 155(3):135–172, 2008.

[4] R. Grossberg and M. VanDieren. Galois-stability in tame abstract elementary classes. *Journal of Mathematical Logic*, 6(1):25–49, 2006.

[5] T. Hyttinen and M. Kesälä. Categoricity transfer in simple finitary abstract elementary classes. To appear in the Journal of Symbolic Logic.

[6] T. Hyttinen and M. Kesälä. Independence in finitary abstract elementary classes. *Annals of Pure and Applied Logic*, 143(1-3):103–138, 2006.

[7] T. Hyttinen and M. Kesälä. Superstability in simple finitary AEC. *Fundamenta Mathematicae*, 195(3):221–268, 2007.

[8] T. Hyttinen and M. Kesälä. Interpreting groups and fields in simple, finitary AECs. To appear in the Journal of Symbolic Logic.

[9] T. Hyttinen and M. Kesälä. Lascar types and Lascar automorphisms in Abstract Elementary Classes. *Notre Dame Journal of Formal Logic*, 52(1):39–54, 2011.

[10] B. Jónsson. Homogeneous universal relational systems. *Mathematica Scandinavica*, 8:137–142, 1960.

[11] M. Kesälä. *Finitary Abstract Elementary Classes.* PhD thesis, University of Helsinki, Department of Mathematics and Statistics, 2006.

[12] H. A. Kierstead. Countable models of ω_1-categorical theories in admissible languages. *Annals of Mathematical Logic*, 19(1-2):127–175, 1980.

[13] R. Kossak and J. Schmerl. *The Structure of Models of Peano Arithmetic.* Number 50 in Oxford Logic Guides. Oxford University Press, 2006.

[14] D. W. Kueker. Abstract elementary classes and infinitary logics. *Annals of Pure and Applied Logic*, 156(2-3):274–286, 2008.

[15] O. Lessmann. Upward categoricity from a successor cardinal for tame abstract classes with amalgamation. *The Journal of Symbolic Logic*, 70(2):639–660, 2005.

[16] M. Morley. Countable models of $\aleph_1$-categorical theories. *Israel Journal of Mathematics*, 5:65–72, 1970.

[17] M. Morley. The number of countable models. *The Journal of Symbolic Logic*, 35:14–18, 1970.

[18] S. Shelah. Categoricity of abstract classes with amalgamation. *Annals of Pure and Applied Logic*, 98:261–294, 1999. Shelah [Sh:394].

[19] S. Shelah, L. Harrington, and M. Makkai. A proof of Vaught's conjecture for ω-stable theories. *Israel Journal of Mathematics*, 49(1-3):259–280, 1984.

[20] C. Smoryński. Elementary extensions of recursively saturated models of arithmetic. *Notre Dame Journal of Formal Logic*, 22(3):193–203, 1981.

[21] R. L. Vaught. Denumerable models of complete theories. In *Infinitistic Methods (Proc. Sympos. Foundations of Math., Warsaw, 1959)*, 303–321. Pergamon, Oxford, 1961.

On Borel equivalence relations in generalized Baire space

Sy-David Friedman[†], **Tapani Hyttinen**[‡]

[†] Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
sdf@logic.univie.ac.at

[‡] Department of Mathematics and Statistics, University of Helsinki, Finland
tapani.hyttinen@helsinki.fi

Abstract. We construct two Borel equivalence relations on the generalized Baire space κ^κ, assuming $\kappa^{<\kappa} = \kappa > \omega$, with the property that neither of them is Borel reducible to the other. A small modification of the construction shows that the straightforward generalization of the Glimm–Effros dichotomy fails.

Introduction

By λ^κ we denote the set of all functions $\kappa \to \lambda$. We define a topology to $(\lambda^\kappa)^n$ by letting the sets

$$N_{(\eta_1,\dots,\eta_n)} = \{(f_1,\dots,f_n) \in (\lambda^\kappa)^n \mid \eta_i \subseteq f_i \text{ for all } 1 \le i \le n\}$$

be the basic open sets, where for some $\alpha < \kappa$ and for all $1 \le i \le n$, η_i is a function from α to λ. We write N_η for $N_{(\eta)}$. For $\kappa > \omega$, the spaces κ^κ are called generalized Baire spaces. The study of these spaces started already in [**5**] and since then many papers have been written on these; more on the history can be found from [**2**]. Most of the study of these spaces (for $\kappa > \omega$) is done under the assumption that $\kappa^{<\kappa} = \kappa$ and we make this assumption also.

By closing open sets under complementation and unions of size $\le \kappa$, we get the class of Borel sets. A function between these spaces is Borel if the inverse image of every open set is Borel. As in the case $\kappa = \omega$, a Borel function F is continuous on a co-meager set, i.e., there are open dense sets U_i, $i < \kappa$, such that $F \restriction (\bigcap_{i<\kappa} U_i)$ is continuous; see [**2**].

Let $X, Y \in \{\kappa^\kappa, 2^\kappa\}$ and let $E \subseteq X^2$ and $E' \subseteq Y^2$ be equivalence relations. We say that E is Borel reducible to E' and write $E \le_B E'$ if there is a Borel function $F\colon X \to Y$ such that, for all $f, g \in X$, fEg if and only if $F(f)E'F(g)$. We say that they are Borel bi-reducible if both $E \le_B E'$ and $E' \le_B E$ hold.

In [**2**] these Borel reductions were studied. We were mostly interested in equivalence relations like isomorphism among (codes of) models of some first-order theory, but also some general theory was developed. And we were annoyed when we found out that we could not find Borel equivalence relations which are incomparable with respect to Borel reducibility. Let us see why one cannot just take some example from the case $\kappa = \omega$ and carry out a straightforward generalization.

To appear in *Archive for Mathematical Logic.*

[†] Supported by the John Templeton Foundation under grant number 13152, and the Austrian Science Fund (FWF), grant number P20835-N13.

[‡] Partially supported by the John Templeton Foundation, grant number 13152, and the Academy of Finland, grant number 1123110.

Arguments like the one in [**4**] use machinery available only in the case $\kappa = \omega$. But with some basic results like the Borel incomparability of E_1 and E_0^ω this is not the case. And indeed one can generalize the definitions of these relations in a straightforward way and prove that E_0^κ is not Borel reducible to E_1; see the proof of Lemma 5. Also if one takes the classical proof of the other direction (see Theorem 8.2 in [**3**]), one notices that everything in the proof holds also in the case $\kappa > \omega$. However, this does not prove that the result is true for $\kappa > \omega$. In the proof two functions are constructed by induction on $i < \kappa$, and, if $\kappa > \omega$, one needs to go over limits and at least without major changes in the construction, this cannot be done (and one can prove this).

In this paper, we will modify the definitions of E_1 and E_0^κ and prove the following theorem:

Theorem 1 *Suppose $\kappa^{<\kappa} = \kappa > \omega$. Then there are Borel equivalence relations on κ^κ such that neither of them is Borel reducible to the other.*

The rest of this paper gives a proof for this theorem.

1 Proof

Before defining the equivalence relations, we want to point out that if $\lambda < \kappa$ and we define $\mathrm{id}_\lambda \subseteq (2^\kappa)^2$ to be the set of pairs (f, g) such that $|\{\alpha < \kappa : f(\alpha) \neq g(\alpha)\}| < \lambda$, then id_λ is Borel bi-reducible with the identity; see [**2**].

For $\alpha, \beta < \kappa$, by $\alpha - \beta$ we denote the (unique) ordinal γ such that $\alpha + \gamma = \beta$ or $\beta + \gamma = \alpha$.

Definition 1.1 We let E_* be the set of pairs (f, g) of functions from κ to κ such that, for some $\alpha < \kappa$, $f(\beta) - g(\beta) < \alpha$ for all $\beta < \kappa$.

Clearly E_* is a Borel equivalence relation on κ^κ.

Let $\gamma \leq \kappa$ and $\pi \colon \kappa \to \gamma \times \kappa$ be one to one and onto. We define a topology on $2^{\gamma\times\kappa}$ so that $f \mapsto g$, $g(\alpha) = f(\pi(\alpha))$, is a homeomorphism from $2^{\gamma\times\kappa}$ onto 2^κ. For $f \in 2^{\gamma\times\kappa}$ and $\alpha < \gamma$, by f_α we mean the function $f_\alpha(x) = f(\alpha, x)$.

Definition 1.2

(I) We let E_0' be the set of pairs $(f, g) \in (2^\kappa)^2$ such that $\{\alpha < \kappa : f(\alpha) \neq g(\alpha)\}$ is a finite union of intervals bounded in κ, i.e., a finite union of sets of the form $[\gamma, \delta)$, $\gamma < \delta < \kappa$.

(II) We let E^* be the set of pairs $(f, g) \in (2^{\kappa\times\kappa})^2$ such that $f_\alpha E_0' g_\alpha$ for all $\alpha < \kappa$.

Clearly both E_0' and E^* are Borel and it is easy to see that they are also equivalence relations. It is also easy to see that E^* is Borel bi-reducible with a Borel equivalence relation on κ^κ (also E_* is Borel bi-reducible with a Borel equivalence relation on 2^κ). So to prove Theorem 1, it is enough to prove Lemmas 1.3 and 1.4 below.

Lemma 1.3 *$E_* \not\leq_B E^*$.*

Proof. For a contradiction, suppose that $F \colon \kappa^\kappa \to 2^{\kappa\times\kappa}$ is a Borel reduction of E_* to E^*. As mentioned above, there are dense and open subsets U_i, $i < \kappa$, of κ^κ such that F is continuous on $U = \bigcap_{i<\kappa} U_i$.

By induction on $i < \kappa$ we construct ordinals $\alpha_i, \beta_i < \kappa$ and functions $f_i^0, f_i^1 \colon \alpha_i \to \kappa$ and $g_i^0, g_i^1 \colon \beta_i \times \beta_i \to 2$ so that

(i) for $i < j$, $\alpha_i < \alpha_j$, $\beta_i < \beta_j$, $f^0_i \subseteq f^0_j$, $f^1_i \subseteq f^1_j$, $g^0_i \subseteq g^0_j$ and $g^1_i \subseteq g^1_j$;
(ii) $N_{f^0_i} \cup N_{f^1_i} \subseteq U_j$ for all $j < i$;
(iii) $F(N_{f^0_i} \cap U) \subseteq N_{g^0_i}$ and $F(N_{f^1_i} \cap U) \subseteq N_{g^1_i}$;
(iv) for some $\gamma < \alpha_{i+1}$, $f^0_{i+1}(\gamma) - f^1_{i+1}(\gamma) \geq i$;
(v) for all $i < \kappa$ and $\gamma < \beta_i$, there is $\beta_i \leq \delta < \beta_{i+1}$ such that $g^0_{i+1}(\gamma, \delta) = g^1_{i+1}(\gamma, \delta)$.

Notice that if we can construct these so that (i)–(v) hold, then $f^0 = \cup_{i<\kappa} f^0_i$ and $f^1 = \cup_{i<\kappa} f^1_i$ belong to U, $F(f^0) = g^0 = \cup_{i<\kappa} g^0_i$, $F(f^1) = g^1 = \cup_{i<\kappa} g^1_i$, and f^0 and f^1 are not in the relation E_*. Also it is not hard to see (as shown below) that for all $i < \kappa$ there are $h^0 \supseteq g^0_i$ and $h^1 \supseteq g^1_i$ such that they are in the relation E^*.

For $i = 0$, we let $\alpha_i = \beta_i = 0$ (and $f^0_i = f^1_i = g^0_i = g^1_i = \emptyset$) and at limits we take unions. Clearly these are as required.

So suppose we have constructed these for i and we construct then for $i + 1$. For all $j < \kappa$ we construct first $h^0_j, h^1_j \in \kappa^{\gamma_j}$, $\gamma_j < \kappa$, as follows: $h^0_0 = f^0_i$ and $h^1_0 = f^1_i$ and at limits we take unions. For $j = 2k + 1$, we choose first $h^1_j \supseteq h^1_{2k}$ so that $N_{h^1_j} \subseteq U_k$ and h^1_j properly extends h^1_{2k}, and then we choose $h^0_j \supseteq h^0_{2k}$ so that $\mathrm{dom}(h^0_j) = \mathrm{dom}(h^1_j)$ and for all $\gamma \in \mathrm{dom}(h^0_j) - \mathrm{dom}(h^0_{2k})$, $h^0_j(\gamma) = h^1_j(\gamma) + i$. For $j = 2k + 2$ we do the reverse, i.e., $N_{h^0_j} \subseteq U_k$ and for all $\gamma \in \mathrm{dom}(h^1_j) - \mathrm{dom}(h^1_{2k+1})$, $h^1_j(\gamma) = h^0_j(\gamma) + i$. Then $h^0 = \cup_{j<\kappa} h^0_j$ and $h^1 = \cup_{j<\kappa} h^1_j$ belong to U and they are E_*-equivalent. Then $F(h^0)$ and $F(h^1)$ are E^*-equivalent and since at stage i the elements satisfy (iii), $F(h^0) \supseteq g^0_i$ and $F(h^1) \supseteq g^1_i$. And so by choosing β_{i+1} large enough and letting $g^0_{i+1} = F(h^0) \upharpoonright (\beta_{i+1} \times \beta_{i+1})$ and $g^1_{i+1} = F(h^1) \upharpoonright (\beta_{i+1} \times \beta_{i+1})$ the requirement (v) and relevant parts of (i) are satisfied. Since F is continuous on U and $h^0, h^1 \in U$, by choosing α_{i+1} large enough and letting $f^0_{i+1} = h^0 \upharpoonright \alpha_{i+1}$ and $f^1_{i+1} = h^1 \upharpoonright \alpha_{i+1}$, the rest of the requirements can be satisfied.

So now we have f^0 and f^1 and since they are not E_*-equivalent, $g^0 = \cup_{i<\kappa} g^0_i = F(f^0)$ and $g^1 = \cup_{i<\kappa} g^1_i = F(f^1)$ are not E^*-equivalent. Let $\alpha < \kappa$ witness this, i.e., $(g^0)_\alpha$ and $(g^1)_\alpha$ are not E'_0-equivalent. By (v) from the construction of g^0 and g^1, it is not possible that $(g^0)_\alpha(\gamma) \neq (g^1)_\alpha(\gamma)$ for all large enough γ. Thus there must exist an increasing sequence $(\gamma_i)_{i<\omega}$ of ordinals $< \kappa$ such that $(g^0)_\alpha(\gamma_i) = (g^1)_\alpha(\gamma_i)$ iff i is odd.

Now choose $i^* < \kappa$ so that $\beta_{i^*} > \alpha \cup \bigcup_{i<\omega} \gamma_i$. Then there are no h^0 and h^1 extending $g^0_{i^*}$ and $g^1_{i^*}$, respectively, so that h^0 and h^1 are E^*-equivalent. As pointed out above, this is a contradiction. □

Lemma 1.4 $E^* \not\leq_B E_*$.

Proof. Towards a contradiction, suppose that $F\colon 2^{\kappa\times\kappa} \to \kappa^\kappa$ is a Borel reduction of E^* to E_*. As above, there are open and dense subsets U_i, $i < \kappa$, of $2^{\kappa\times\kappa}$ such that F is continuous on $U = \cap_{i<\kappa} U_i$. By induction on $i < \kappa$, we construct ordinals $\alpha_i, \beta_i < \kappa$ and functions $f^0_i, f^1_i\colon \alpha_i \times \alpha_i \to 2$ and $g^0_i, g^1_i\colon \beta_i \to \kappa$ so that

(i) for $i < j$, $\alpha_i < \alpha_j$, $\beta_i < \beta_j$, $f^0_i \subseteq f^0_j$, $f^1_i \subseteq f^1_j$, $g^0_i \subseteq g^0_j$, $g^1_i \subseteq g^1_j$ and $\alpha_0 = \beta_0 = 0$;
(ii) $N_{f^0_i} \cup N_{f^1_i} \subseteq U_j$ for all $j < i$;
(iii) $F(N_{f^0_i} \cap U) \subseteq N_{g^0_i}$ and $F(N_{f^1_i} \cap U) \subseteq N_{g^1_i}$;
(iv) for some $\gamma < \alpha_{i+1}$, $g^0_{i+1}(\gamma) - g^1_{i+1}(\gamma) \geq i$;
(v) for all $\alpha_i \leq \alpha < \alpha_{i+1} \leq \alpha_j$, the following hold:
 (a) for all $\gamma < \alpha_{i+1}$, $(f^0_j)_\alpha(\gamma) \neq (f^1_j)_\alpha(\gamma)$;
 (b) for all $\alpha_{i+1} \leq \gamma < \alpha_j$, $(f^0_j)_\alpha(\gamma) = (f^1_j)_\alpha(\gamma)$.

If we can construct these so that (i)–(v) hold, we have a contradiction: By (v), $f^0 = \bigcup_{i<\kappa} f^0_i$ and $f^1 = \bigcup_{i<\kappa} f^1_i$ are E^*-equivalent. By (ii) and (iii), $F(f^0) = g^0 = \bigcup_{i<\kappa} g^0_i$ and $F(f^1) = g^0 = \bigcup_{i<\kappa} g^1_i$, and by (iv) these are not E_*-equivalent.

For $i = 0$, we let $\alpha_i = \beta_i = 0$ (and $f^0_i = f^1_i = g^0_i = g^1_i = \emptyset$) and at limits we take unions. Clearly these are as required.

So suppose that we have constructed these for $j \leq i$ and we construct them for $i+1$. First we want to find $h^0, h^1 \colon \kappa \times \kappa \to 2$ such that $f^0_i \subseteq h^0$, $f^1_i \subseteq h^1$, $h^0, h^1 \in U$ and for all $(\delta, \delta') \in (\kappa \times \kappa) - (\alpha_i \times \alpha_i)$, $h^0(\delta, \delta') = h^1(\delta, \delta')$ if and only if $\delta < \alpha_i$. For this we construct increasing sequences $(h^0_j)_{j<\kappa}$ and $(h^1_j)_{j<\kappa}$ of functions $h^0_j, h^1_j \colon \gamma_j \times \gamma_j \to 2$ as follows:

For $j = 0$, we let $\gamma_j = \alpha_i$, $h^0_j = f^0_i$ and $h^1_j = f^1_i$ and at limits we take unions. For $j = 2k+1$, choose the h^0_j, h^1_j as follows: We let $\gamma_j > \gamma_{2k}$ and $h^0_j \colon \gamma_j \times \gamma_j \to 2$ be such that $h^0_{2k} \subseteq h^0_j$ and $N_{h^0_j} \subseteq U_k$, and we let $h^1_j \colon \gamma_j \times \gamma_j \to 2$ be such that $h^1_{2k} \subseteq h^1_j$ and for all $(\delta, \delta') \in (\gamma_j \times \gamma_j) - (\gamma_{2k} \times \gamma_{2k})$, $h^1_j(\delta, \delta') = h^0_j(\delta, \delta')$ if and only if $\delta < \alpha_i$. For $j = 2k+2$ we do the reverse, i.e., we let $\gamma_j > \gamma_{2k+1}$ and $h^1_j \colon \gamma_j \times \gamma_j \to 2$ be such that $h^1_{2k+1} \subseteq h^1_j$ and $N_{h^1_j} \in U_k$, and we let $h^0_j \colon \gamma_j \times \gamma_j \to 2$ be such that $h^0_{2k+1} \subseteq h^1_j$ and for all $(\delta, \delta') \in (\gamma_j \times \gamma_j) - (\gamma_{2k+1} \times \gamma_{2k+1})$, $h^0_j(\delta, \delta') = h^1_j(\delta, \delta')$ if and only if $\delta < \alpha_i$. Then $h^0 = \bigcup_{j<\kappa} h^0_j$ and $h^1 = \bigcup_{j<\kappa} h^1_j$ are as wanted.

Since h^0 and h^1 are not E^*-equivalent and $h^0, h^1 \in U$, $F(h^0) \supseteq g^0_i$ and $F(h^1) \supseteq g^1_i$ are not E_*-equivalent. So by choosing $\beta_{i+1} > \beta_i$ large enough, $g^0_{i+1} = F(h^0) \restriction \beta_{i+1}$ and $g^1_{i+1} = F(h^1) \restriction \beta_{i+1}$ satisfy (iv) and the relevant parts of (i). Since F is continuous on U, by choosing $\alpha_{i+1} > \alpha_i$ large enough the rest of the requirements can be satisfied. □

2 Open questions

We finish this paper with several open questions. But before this, we make some definitions and observations.

Let id be the set of pairs $(f, g) \in (2^\kappa)^2$ such that $f = g$ and E_0 be the set of pairs $(f, g) \in (2^\kappa)^2$ such that, for some $\alpha < \kappa$, $f(\gamma) = g(\gamma)$ for all $\gamma > \alpha$. Then these are Borel equivalence relations and clearly id $\leq_B E_0$ and similarly id $\leq_B E'_0$ (Definition 1.2(I)). As pointed out in [**2**], $E_0 \not\leq_B$ id since if F is a reduction of E_0 to id and continuous on a co-meager set U, one can find $\alpha < \kappa$ and $\eta, \xi \colon \alpha \to 2$ and $\alpha' < \kappa$ and $\eta', \xi' \colon \alpha' \to 2$ so that $\eta' \neq \xi'$ and $F(N_\eta \cap U) \subseteq N_{\eta'}$ and $F(N_\xi \cap U) \subseteq N_{\xi'}$. But this is impossible because there are $f \in N_\eta \cap U$ and $g \in N_\xi \cap U$ which are E_0-equivalent. Similarly $E'_0 \not\leq_B$ id and by repeating this argument ω times, one can see the following lemma:

Lemma 2.1 $E_0 \not\leq E'_0$.

Proof. (Sketch) For a contradiction, suppose that $F \colon 2^\kappa \to 2^\kappa$ is a reduction, which is continuous on a co-meager set U. As in the proof of $E_0 \not\leq$ id, one can find increasing sequences $(\alpha_i)_{i<\omega}$ and $(\gamma_i)_{i<\omega}$ of ordinals and increasing sequences of functions $\eta_i, \xi_i \colon \alpha_i \to 2$ and $\eta'_i, \xi'_i \colon \gamma_i \to 2$ such that

(i) $F(N_{\eta_i} \cap U) \subseteq N_{\eta'_i}$ and $F(N_{\xi_i} \cap U) \subseteq N_{\xi'_i}$;

(ii) for all $i < \omega$ there are $\gamma_i \leq \beta < \beta' < \gamma_{i+1}$ such that $\eta'_{i+1}(\beta) = \xi'_{i+1}(\beta)$ and $\eta'_{i+1}(\beta') \neq \xi'_{i+1}(\beta')$.

But now we have a contradiction, since there are $f \in U \cap \bigcap_{i<\omega} N_{\eta_i}$ and $g \in U \cap \bigcap_{i<\omega} N_{\xi_i}$ which are E_0-equivalent. □

Open Question 2.2 Is E_0' Borel reducible to E_0?

Open Question 2.3 In the case $\kappa = \omega$, by the Glimm–Effros dichotomy (see e.g. [BK]), for all Borel equivalence relations E above id, either $E \leq_B$ id or $E_0 \leq_B E$. By what is above, E_0' witnesses that this is not true for uncountable κ. However, notice that, for $\kappa = \omega$, $E_0 = E_0'$, and one can ask: is Glimm–Effros true with E_0' in place of E_0 (for $\kappa > \omega$)?

Open Question 2.4 Let us look at the structure $(BE, \leq_B)$ where BE is the set of all Borel equivalence relations on 2^κ. By what is said above, $(BE, \leq_B)$ contains antichains of length at least 2 and above id, chains of length at least 4 (id $<_B E_0' <_B E^* <_B$ "$E^* \times E_*$"; essentially as in the proof of Lemma 1.4 one can show that E^* is not Borel reducible to E_0'). Can one find longer chains and antichains?

In Open Question 2.4 we mean equivalence relations that can be defined for all $\kappa = \kappa^{<\kappa} > \omega$. For large κ, the following gives a long chain: For all γ such that $\aleph_\gamma < \kappa$, let E_0^γ be the set of pairs $(f, g) \in (2^\kappa)^2$ such that there is an increasing and continuous sequence $(\alpha_i)_{i \leq \beta}$, $\beta < \aleph_{\gamma+1}$, such that $\alpha_0 = 0$, for all $\delta \geq \alpha_\beta$, $f(\delta) = g(\delta)$ and for all $i < \beta$, either for all $\alpha_i \leq \delta < \alpha_{i+1}$, $f(\delta) = g(\delta)$ or for all $\alpha_i \leq \delta < \alpha_{i+1}$, $f(\delta) \neq g(\delta)$. Then for $\alpha > 0$, $\aleph_\alpha < \kappa$, we define $E_0^{<\alpha}$ to be the set of all pairs $(f, g) \in (2^{\alpha \times \kappa})^2$ such that for all $\gamma < \alpha$, $f_\gamma E_0^\gamma g_\gamma$.

It is easy to see that these are Borel equivalence relations, for all $\gamma < \beta$ and $0 < \alpha < \beta$, $E_0^\gamma, E_0^{<\alpha} \leq_B E_0^{<\beta}$ and, as in the proof of Lemma 2.1, one can see that $E_0^\alpha \not\leq_B E_0^{<\alpha}$ and thus $E_0^{<\beta} \not\leq_B E_0^{<\alpha}$.

References

[1] H. Becker and A. Kechris, *Descriptive Set Theory of Polish Group Actions*, London Math. Soc. Lecture Note Series, vol. 232, Cambridge University Press, 1996.

[2] S. Friedman, T. Hyttinen and V. Kulikov, Generalized descriptive set theory and classification theory, Preprint no. 999, Centre de Recerca Matemàtica, Bellaterra, 2011, 1–99.

[3] G. Hjorth, *Classification and Orbit Equivalence Relations*, AMS Mathematical Surveys and Monographs, vol. 75, 2000.

[4] A. Louveau and B. Velickovic, A note on Borel equivalence relations, *Proc. Amer. Math. Soc.*, 120 (1994), 255–259.

[5] R. Vaught, Invariant sets in topology and logic, *Fund. Math.*, 82 (1974/75), 269–294.

Non-absoluteness of model existence in uncountable cardinals for $L_{\omega_1,\omega}$

Sy-David Friedman[†], **Tapani Hyttinen**[‡], **Martin Koerwien**[§]

[†] Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
sdf@logic.univie.ac.at

[‡] Department of Mathematics and Statistics, University of Helsinki, Finland
tapani.hyttinen@helsinki.fi

[§] Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
koerwien@math.uic.edu

Abstract. For sentences ϕ of $L_{\omega_1,\omega}$, we investigate the question of absoluteness of ϕ having models in uncountable cardinalities. We first observe that having a model in $\aleph_1$ is an absolute property, but having a model in $\aleph_2$ is not, as it may depend on the validity of the Continuum Hypothesis. We then consider the GCH context and provide sentences for any $\alpha \in \omega_1 \setminus \{0, 1, \omega\}$ for which the existence of a model in $\aleph_\alpha$ is non-absolute (relative to large cardinal hypotheses). Finally, we present a *complete* sentence for which model existence in $\aleph_3$ is non-absolute.

Introduction

Throughout, we assume that ϕ is an $L_{\omega_1,\omega}$ sentence which has infinite models. By the downward Löwenheim–Skolem theorem, ϕ must have a countable model, so the property "having a countable model" is an absolute property of such sentences in the sense that its validity does not depend on the properties of the set-theoretic universe we work in. More precisely, if $V \subseteq W$ are transitive models of ZFC with the same ordinals and $\phi \in V$, $V \models$ "ϕ is an $L_{\omega_1,\omega}$-sentence" (with a natural set-theoretic coding of such sentences), then $V \models$ "ϕ has a countable model" if and only if $W \models$ "ϕ has a countable model". The purpose of this paper is to investigate the question of how far we can replace "countable" by higher cardinalities.

A main tool for absoluteness considerations is Shoenfield's absoluteness theorem (Theorem 25.20 in [**9**]). It states that any property expressed by either a $\mathbf{\Sigma}^1_2$ or a $\mathbf{\Pi}^1_2$ formula is absolute between transitive models of ZFC with the same ordinals. As John Baldwin observed in [**1**], it follows from results of [**7**] that the property of ϕ having arbitrarily large models is absolute (it can be expressed in form of the existence of an infinite indiscernible sequence, which by Shoenfield is absolute). Since the Hanf number of the logic $L_{\omega_1,\omega}$ equals $\beth_{\omega_1}$, it follows that the existence of models in cardinalities above that number is absolute. Therefore the context we are interested in is where ϕ (absolutely) does not have a model of size $\beth_{\omega_1}$.

Submitted for publication in the Notre Dame Journal of Formal Logic.

[†] Supported by the John Templeton Foundation under grant #13152, The Myriad Aspects of Infinity, and by the Austrian Science Fund (FWF) through Project Number P 22430-N13.

[‡] Partially supported by the Academy of Finland under grant number 1123110.

[§] Supported by the John Templeton Foundation under grant #13152, The Myriad Aspects of Infinity.

1 The case $\aleph_1$

For *complete* sentences ϕ (meaning that any model of ϕ satisfies the same $L_{\omega_1,\omega}$ sentences), having a model in $\aleph_1$ is an absolute notion. We have the following characterization (which appears also in [**1**] as well as in [**5**]) of ϕ having a model of size $\aleph_1$ (which is a $\mathbf{\Sigma}^1_1$ property and therefore absolute by Shoenfield's absoluteness theorem):

(∗) There exist two countable models M, N of ϕ such that M is a proper elementary (in the fragment of ϕ) substructure of N.

To see that this is a characterization, note first that if ϕ has an uncountable model, then (∗) holds by Löwenheim–Skolem. For the converse, we use the completeness of ϕ which implies that any two countable models of ϕ are isomorphic (by Scott's isomorphism theorem, since ϕ is complete and thus characterizes its countable models up to isomorphism). Then, as $N \cong M$, we can find a proper countable $L_{\omega_1,\omega}$-elementary extension of N as well and continue this procedure ω_1 many times (taking unions at limit stages). The union of this elementary chain will then be a model of ϕ of size $\aleph_1$.

If the sentence is not complete, criterion (∗) does not obviously imply the existence of an uncountable model. By a theorem of Gregory's (see [**6**]), it can be seen that it actually does. We will however provide a different criterion ((∗∗) below) for which we have a relatively basic proof (essentially only using the omitting types theorem for $L_{\omega_1,\omega}$) that it is equivalent to ϕ having an uncountable model. We thus have that for any (even incomplete) $L_{\omega_1,\omega}$-sentence, model existence in $\aleph_1$ is absolute.[1]

In the following, we consider the sentence ϕ as a set-theoretic object using standard coding of formulas of $L_{\omega_1,\omega}$. Thus ϕ can be regarded as a hereditarily countable set.

The following property which (again by Shoenfield) is absolute, characterizes ϕ having a model of size $\aleph_1$:

(∗∗) There is a countable transitive model U of ZFC$^-$ (ZFC without the power set axiom) containing ϕ with $U \models$ "ω_1 exists, ϕ is hereditarily countable, and there is a model of ϕ with universe ω_1".

First, suppose ϕ has a model M of size $\aleph_1$, say one with universe ω_1. As both ϕ and M are elements of H_{ω_2} (the collection of sets hereditarily of size at most $\aleph_1$), we have $H_{\omega_2} \models \text{ZFC}^- +$ "there is a model of ϕ with universe ω_1". Now it suffices to take a countable (first order) elementary substructure $U \prec H_{\omega_2}$ containing ϕ, and U will have the properties of (∗∗).

Conversely, assuming that (∗∗) holds for some countable U, we can take an elementary extension U' of U where all (in the sense of U) hereditarily countable sets are unchanged and all (in U) uncountable ones become sets of size $\aleph_1$. This can be achieved using Corollary A of Theorem 36 in [**11**], noting that it holds for models of ZFC$^-$ (instead of full ZFC as the Corollary originally assumes), as the powerset axiom is not used for it. In particular this is true for the ω_1 of U' on which we know a model M of ϕ lives (note that $U' \models (M \models \phi)$ implies that $M \models \phi$ in the real universe; to see this, use that U' contains the fragment of ϕ and satisfaction for formulas in this fragment is absolute between U' and the real universe). So we get a model of ϕ of size $\aleph_1$.

There is another absolute criterion characterizing ϕ having an uncountable model, but it requires going beyond the logic $L_{\omega_1,\omega}$. Let us consider the extension $L_{\omega_1,\omega}(Q)$ of $L_{\omega_1,\omega}$

[1] This has also been observed recently by Paul Larson. His argument uses iterated generic ultrapowers. Rami Grossberg points out that he knew of this fact already in the 1980's but did not publish it, and that others like Shelah, Barwise and Keisler most likely knew of it even earlier.

obtained by adding an extra quantifier Q with the semantics "there exist uncountably many". As is shown in [**2**], $L_{\omega_1,\omega}(Q)$ admits a completeness theorem which actually has a very natural (absolute) deduction calculus. Now the statement

$(* * *)$ There is a proof of $\neg Qx(x = x)$ starting from ϕ.

characterizes ϕ having only countable models. Thus the negation of $(***)$ is an (absolute) property characterizing ϕ having an uncountable model. Note that this argument shows that model existence in $\aleph_1$ is absolute even for $L_{\omega_1,\omega}(Q)$ sentences.

2 Going beyond $\aleph_1$

It is not generally true that the existence of a model of size $\aleph_2$ is an absolute property.

A very simple way to see this is to take any sentence ϕ that has models exactly up to size continuum. We easily find even *complete* sentences with this property. Then, clearly, ϕ has a model of size $\aleph_2$ if and only if the continuum hypothesis fails.

More generally, such a sentence has a model of size $\aleph_\alpha$ if and only if $2^{\aleph_0} \geq \aleph_\alpha$. So, for any $\alpha > 1$, the existence of a model of size $\aleph_\alpha$ is non-absolute.

There are many examples of complete $L_{\omega_1,\omega}$-sentences in the literature having models exactly up to size continuum, but they are mostly more complicated than necessary for our purposes, because their authors have been interested in additional properties. Therefore we provide here a very simple such example which uses the idea of coding full binary trees. This same idea has been used in Malitz's examples showing that the Hanf number for *complete* $L_{\omega_1,\omega}$-sentences equals $\beth_{\omega_1}$ (see [**12**]).

Let the language L consist of countably many binary relation symbols E_n $(n < \omega)$, and let $\sigma \in L_{\omega_1,\omega}$ be the conjunction of

- all E_n are equivalence relations such that E_0 has two classes and each E_n-class is the union of exactly two E_{n+1}-classes;
- $\forall x, y((\bigwedge_{n<\omega} E_n(x, y)) \rightarrow x = y)$.

It is an easy back-and-forth argument to show that any two countable models of σ are isomorphic, so σ is complete. Every model represents a set of branches through a full binary tree, so there cannot be models greater than the continuum. On the other hand, the Cantor space 2^ω together with the relations "$E_n(x, y)$ if and only if x and y coincide on the $n + 1$ first components" is a model of σ of size continuum.

3 Going beyond $\aleph_1$ under the assumption of GCH

As we have seen, playing with the cardinal exponential function provides trivial examples for the non-absoluteness of the existence of models of cardinality greater than $\aleph_1$. A next natural question is if this is the only non-absoluteness phenomenon there is. That is, under the additional assumption of GCH, does the existence of models in cardinalities greater than $\aleph_1$ become an absolute notion? We will provide different incomplete sentences and later on even a complete one that show that the answer is negative.

3.1 A reminder about two-cardinal properties

As we will see later, there is an interesting connection between classical first-order two-cardinal properties and model existence for $L_{\omega_1,\omega}$-sentences. We recall the following definition:

Definition 3.1 Let T be a first-order theory in a signature containing a unary predicate P. Given two infinite cardinals $\kappa \geq \lambda$, we say that T *admits* (κ, λ) if there is a model of T of size κ such that $P^M = \{a \in M \mid M \models P(a)\}$ has cardinality λ.

As is already exposed in Chang–Keisler's classical textbook [**3**, Chapter 7.2], admitting certain pairs (κ, λ) is a non-absolute property for certain theories. There, examples are given where admitting (κ^+, κ) is equivalent to the existence of a special κ^+-Aronszajn tree or where admitting (κ^{++}, κ) is equivalent to the existence of a κ^+ Kurepa tree (or equivalently a κ^+ Kurepa family).

3.2 Some set theory

We now recall the two classical concepts of Kurepa families and special Aronszajn trees. The first-order examples in [**3**] showing non-absoluteness of the existence of certain two-cardinal models and our later exposed examples of $L_{\omega_1,\omega}$-sentences showing non-absoluteness of model existence in certain cardinalities code those objects in their models. The coding is such that the existence of a certain two-cardinal model or the existence of a model in a certain cardinality is equivalent to the existence of such an object (which is independent from $\mathrm{ZFC} + \mathrm{GCH}$ as we will see in the following).

Definition 3.2 Let κ be any infinite cardinal. A κ^+ *Kurepa family* is a family $\mathcal{F}$ of subsets of some set A with $|A| = \kappa^+$, such that $|\mathcal{F}| > \kappa^+$ and, for any subset $B \subset A$ with $|B| = \kappa$, $|\{X \cap B : X \in \mathcal{F}\}| \leq \kappa$.

Let KH_{κ^+} be the statement that there exists a κ^+ Kurepa family.

It is folklore that the existence of Kurepa families in different $\aleph_\alpha$ $(\alpha < \omega_1)$ is independent from one another. We will now describe the formal arguments for the cases we need (essentially the same arguments would work more generally for "switching on and off" independently of the existence of Kurepa families in different $\aleph_\alpha$). In the constructible universe, KH_{κ^+} is true for all cardinals κ (this follows from the fact that $\diamondsuit^+$ holds at successor cardinals in L; see [**10**]). On the other hand we have:

Theorem 3.3 *The consistency of "*ZFC + *there are uncountably many inaccessible cardinals" implies the consistency of "*ZFC + GCH + $\forall\alpha < \omega_1 \neg\mathrm{KH}_{\aleph_{\alpha+1}}$*".*

Proof. This is a slight generalisation of Silver's argument that if κ is inaccessible then after forcing with $\mathrm{Coll}(\omega_1, < \kappa)$, the forcing to convert κ into $\aleph_2$ with countable conditions, $\mathrm{KH}_{\aleph_1}$ fails (see [**9**, Theorem 27.9]).

Assume GCH; let κ_0 be $\aleph_1$ and define $(\kappa_\beta)_{0<\beta<\omega_1}$ inductively: set $\kappa_{\beta+1}$ the least inaccessible cardinal greater than κ_β and for $\beta < \omega_1$ a limit ordinal set $\kappa_\beta = \sup\{\kappa_\gamma \mid \gamma < \beta\}^+$. Let P be the fully supported product of the forcings $\mathrm{Coll}(\kappa_\beta, < \kappa_{\beta+1})$ for $\beta < \omega_1$. Then in the extension, κ_β equals $\aleph_{\beta+1}$, while the GCH still holds. We claim that $\mathrm{KH}_{\kappa_\beta}$ fails for each $\beta < \omega_1$.

Indeed, the forcing P can be factored as $P(< \beta) \times P(\geq \beta)$ where $P(< \beta)$ refers only to the collapses $\mathrm{Coll}(\kappa_\gamma, < \kappa_{\gamma+1})$ for $\gamma < \beta$ and $P(\geq \beta)$ refers only to the collapses $\mathrm{Coll}(\kappa_\gamma, < \kappa_{\gamma+1})$ for $\gamma \geq \beta$. Similarly, $V[G]$ factors as $V[G(< \beta)][G(\geq \beta)]$. In the model $V[G(< \beta)]$, $\kappa_{\beta+1}$ is still inaccessible, so we can apply Silver's argument to conclude that $\mathrm{KH}_{\kappa_\beta}$ fails in $V[G(< \beta)][G(\geq \beta)] = V[G]$, using the closure of the forcing $P(\geq \beta)$ under sequences of length less than κ_β. □

Definition 3.4 A *tree* is a partially ordered set $(T, <)$ such that, for any element $t \in T$, the set $\{x \mid x < t\}$ is well ordered by $<$. The *rank* $\mathrm{rk}(t)$ of t is the order type of $\{x \mid x < t\}$. For any ordinal α, let $T_\alpha = \{t \in T \mid \mathrm{rk}(t) = \alpha\}$.

For any cardinal κ, a κ^+*-tree* is a tree T such that $T_{\kappa^+} = \emptyset$ and, for all $\alpha < \kappa^+$, $0 < |T_\alpha| < \kappa^+$. A tree T is *normal* if

- $|T_0| = 1$;
- every element has at least two immediate successors;
- for any $t \in T$ and α with $\mathrm{rk}(t) < \alpha < \kappa^+$, there is some $t' > t$ with $\mathrm{rk}(t') = \alpha$.

A normal κ^+-tree T is a *special* κ^+*-Aronszajn tree* if there is some set A of size κ and a function $f\colon T \to A$ such that for all $t, t' \in T$, $t < t'$ implies $f(t) \neq f(t')$.

It is a consequence of GCH that special κ-Aronszajn trees exist for all successor cardinals κ that are not successors of singular cardinals (see [**13**]). Moreover, in the constructible universe, special Aronszajn trees exist even in successors of singular cardinals (this is a consequence of $\square_\kappa$; see [**10**]).

On the other hand, the consistency of "ZFC $+ \exists\kappa(\kappa$ supercompact)" implies the consistency of "ZFC $+$ GCH $+$ there are no special $\aleph_\alpha$-Aronszajn trees for all countable limit successors α": We start with a model of GCH with a supercompact cardinal κ and force with $\mathrm{Coll}(\omega_1, <\kappa)$. As is argued in [**4**], this forcing preserves a stationary reflection property sufficient to ensure that Weak Square fails at $\aleph_\lambda$ for λ a limit ordinal of countable cofinality. By a result of Jensen in [**10**], Weak Square at a cardinal κ is equivalent to the existence of a special Aronszajn tree on κ^+.

3.3 Connecting first-order two-cardinal properties with $L_{\omega_1,\omega}$-model existence

We will describe how a first-order theory T can be turned into an $L_{\omega_1,\omega}$-sentence σ in such a way that T admitting certain (κ, λ) is equivalent to the existence of a model of σ of size κ.

We start with the definition of an $L_{\omega_1,\omega}$-sentence σ_0^α characterizing $\aleph_\alpha$ (for $\alpha < \omega_1$), which means that it (absolutely) has a model of size $\aleph_\alpha$, but no bigger model. We wish to point out that the idea we use here of characterizing cardinals using κ-like orderings for various κ is not new. Also, there exist other ways of characterizing cardinals in the literature, most notably Hjorth's examples presented in [**8**] that are even *complete* sentences.

Let $L_0^\alpha = \{Q_\beta, a_n, <, F\}_{\beta \leq \alpha;\ n<\omega}$, where the Q_β are unary predicates, the a_n are constant symbols, $<$ is a binary and F a ternary relation symbol.

Let $\sigma_0^\alpha \in (L_0^\alpha)_{\omega_1,\omega}$ be the conjunction of the following sentences:

- The universe is the union of all Q_β.
- $Q_0 = \{a_n \mid n < \omega\}$ where all a_n designate distinct elements.
- For any $\beta < \alpha$, $Q_{\beta+1}$ is disjoint from any Q_γ for all $\gamma \leq \beta$.
- For any limit ordinal $\beta \leq \alpha$, $Q_\beta = \bigcup_{\gamma<\beta} Q_\gamma$.
- $<$ linearly orders $Q_{\beta+1}$ for every $\beta < \alpha$ and $x < y$ implies that for some $\beta < \alpha$, both x and y belong to $Q_{\beta+1}$.
- $F(a, b, c)$ implies that for some $\beta < \alpha$, $a \in Q_{\beta+1}$, $b < a$ and $c \in Q_\beta$.
- For every $\beta < \alpha$ and every $a \in Q_{\beta+1}$, $F(a, \cdot, \cdot)$ defines a total injective function from $\{x \mid x < a\}$ into Q_β.

Note that for β a limit ordinal or zero, Q_β is not ordered by $<$ and, if $\alpha = 0$, both $<$ and F are empty relations.

Clearly, if $M \models \sigma_0^\alpha$, then in M the ordering of $Q_{\beta+1}$ must be $|Q_\beta|$-like (i.e., any proper initial segment has cardinality at most $|Q_\beta|$). This implies that $|Q_{\beta+1}|$ is at most $|Q_\beta|^+$ and, since Q_0 is countable by definition, we see inductively that the cardinality of each Q_β is bounded by $\aleph_\beta$. Also, there clearly exist models such that $|Q_\beta| = \aleph_\beta$ for all $\beta \leq \alpha$.

Now suppose that we have a first-order theory T in a language containing a unary predicate P. For $\beta < \alpha < \omega_1$, we define the $L_{\omega_1,\omega}$-sentence $\sigma_T^{\alpha,\beta}$ as the conjunction of

- T;
- σ_0^α;
- $P = Q_\beta$.

Proposition 3.5 *Let $\beta < \omega_1$ and $0 < n < \omega$. T admits $(\aleph_{\beta+n}, \aleph_\beta)$ if and only if $\sigma_T^{\beta+n,\beta}$ has a model of cardinality $\aleph_{\beta+n}$.*

Proof. If $M \models \sigma_T^{\beta+n,\beta}$ has cardinality $\aleph_{\beta+n}$, we must have $|Q_\beta| = \aleph_\beta$ in that model (here we use that n is finite!). Now the reduct of M to the language of T is a model of size $\aleph_{\beta+n}$ where P has size $\aleph_\beta$.

Conversely, given a model of T of size $\aleph_{\beta+n}$ where P has size $\aleph_\beta$, it is easy to expand this model to be a model of $\sigma_T^{\beta+n,\beta}$. □

Note that this proposition becomes false if n is allowed to be infinite.

3.4 Examples of incomplete sentences: successor cardinals

We quote Chang–Keisler's results 7.2.11 and 7.2.13 from **[3]** (adapting the notation slightly):

- There is a sentence ϕ_1 in a finite language L such that, for all infinite cardinals λ, ϕ_1 admits (λ^+, λ) if and only if there exists a special λ^+-Aronszajn tree.
- There is a sentence ϕ_2 in a suitable language such that, for all infinite cardinals λ, ϕ_2 admits (λ^{++}, λ) if and only if a λ^+ Kurepa family exists.

From the preceding section we get thus infinitary sentences $\sigma_{\phi_1}^{\alpha+1,\alpha}$ and $\sigma_{\phi_2}^{\alpha+2,\alpha}$ such that

- $\sigma_{\phi_1}^{\alpha+1,\alpha}$ has a model of cardinality $\aleph_{\alpha+1}$ if and only if a special $\aleph_{\alpha+1}$-Aronszajn tree exists;
- $\sigma_{\phi_2}^{\alpha+2,\alpha}$ has a model of cardinality $\aleph_{\alpha+2}$ if and only if an $\aleph_{\alpha+1}$ Kurepa family exists.

Now, recalling the set-theoretic facts from Section 3.2, we get the following results:

Theorem 3.6 *Let $\alpha < \omega_1$ be a limit ordinal. Assuming ZFC, GCH and the existence of a supercompact cardinal, model-existence in $\aleph_{\alpha+1}$ is non-absolute for $L_{\omega_1,\omega}$-sentences.*

Theorem 3.7 *Let $\alpha < \omega_1$. Assuming ZFC, GCH and the existence of uncountably many inaccessible cardinals, model-existence in $\aleph_{\alpha+2}$ is non-absolute for $L_{\omega_1,\omega}$-sentences.*

At this point, we have covered all cases of successor cardinals $\aleph_\alpha$ for $1 < \alpha < \aleph_{\omega_1}$.

3.5 Examples of incomplete sentences: limit cardinals

We would also like to find examples of (incomplete) sentences where model existence in $\aleph_\alpha$ is non-absolute modulo $\mathrm{ZFC} + \mathrm{GCH}$ for countable limit ordinals α. With a slight variation of our examples involving special Aronszajn trees, we can deal with limits that are greater than ω.

Since the construction is rather straightforward, we will only give an informal description of it.

The sentence ϕ_1 used to prove Theorem 3.6 which is given explicitly in [**3**] involves essentially a binary relation T coding a tree and a unary predicate U and has the property that whenever $M \models \phi_1$ and $|M| = |U^M|^+$, then T has a subtree which is a special $|M|$-Aronszajn tree.

Now, fixing some $\alpha < \omega_1$ greater than ω, we start with the sentence σ_0^α (see Section 3.3) and for all $\beta < \alpha$, we add the sentence ϕ_1 relativised to $\bigcup_{\gamma \leq \beta+1} Q_\gamma$ (i.e., the set $\bigcup_{\gamma \leq \beta+1} Q_\gamma$ with the induced structure in the language of ϕ_1 is a model of ϕ_1) with Q_β taking the role of U. That is, we are coding special Aronszajn trees at *every* level $Q_{\beta+1}$ where $|Q_{\beta+1}| = |Q_\beta|^+$.

The result is a sentence σ_1^α for which (assuming consistency of supercompact cardinals) the existence of a model of size $\aleph_\alpha$ is non-absolute modulo $\mathrm{ZFC} + \mathrm{GCH}$. The reason is that if no special $\aleph_{\omega+1}$-Aronszajn tree exists, the maximum cardinality of a model of σ_1^α is $\aleph_\omega$ since whenever for some $\gamma < \alpha$, $|Q_{\gamma+1}| = |Q_\gamma|^+ = \aleph_{\omega+1}$, a special $\aleph_{\omega+1}$-Aronszajn tree will be coded in the model. Note that in any case, σ_1^α will have models of size $\aleph_\omega$ since GCH implies the existence of special α_n-Aronszajn trees for all finite $n > 0$. Therefore these examples do not show non-absoluteness of model existence in $\aleph_\omega$.

4 A complete sentence

Both the first-order examples from [**3**] and our $L_{\omega_1,\omega}$-examples from the preceding section are highly incomplete (i.e., many first-order or $L_{\omega_1,\omega}$-statements are undecided) and it seems a very non-trivial task to turn them into complete theories while conserving the properties that matter to us.

We will now introduce a method of completing incomplete $L_{\omega_1,\omega}$-sentences that has the benefits of providing fairly explicit axiomatizations as well as some means of constructing models of the resulting complete sentence with certain properties. This method will then be applied to an incomplete sentence coding $\aleph_2$ Kurepa trees (similar to the examples from the preceding section).

Definition 4.1 Let $\sigma \in L_{\omega_1,\omega}$.

- A *σ-chain* is a family $(M_\alpha)_{\alpha<\lambda}$ of models of σ such that, whenever $\alpha < \beta < \lambda$, we have $M_\alpha \subset M_\beta$.
- σ is *preserved under chains* if, for any σ-chain $(M_\alpha)_{\alpha<\lambda}$, $M = \bigcup_{\alpha<\lambda} M_\alpha$ is a model of σ.

As in the classical first-order case, it is still true that any Π_2-sentence is preserved under chains, i.e., any sentence of the form $\forall \overline{x}\, \exists \overline{y}\, \psi(\overline{x}, \overline{y})$ where ψ is quantifier-free (but possibly infinitary). We have to be a little careful with the definition of Π_2 as for example infinite disjunctions of universal formulas might not be preserved under chains. A simple example is given by the sentence

$$\sigma = \bigvee_{S \subset \omega \text{ finite}} \forall x \Big(U(x) \leftrightarrow \bigvee_{i \in S} x = a_i \Big)$$

in the language of countably many constants a_i and a unary predicate U. This sentence expresses that U is finite.

Definition 4.2 Let $\sigma \in L_{\omega_1,\omega}$.

- Set $S_{\mathrm{qf}}(\sigma) = \{\mathrm{tp}_{\mathrm{qf}}(\overline{a}) \mid \exists M \models \sigma \ (\overline{a} \in M)\}$ (where $\mathrm{tp}_{\mathrm{qf}}(\overline{a})$ is the quantifier-free type of $\overline{a}$).
- σ is *qf-small* if $S_{\mathrm{qf}}(\sigma)$ is countable.

Note that, by the downward Löwenheim–Skolem theorem, we can define $S_{\mathrm{qf}}(\sigma)$ by referring only to countable models of σ.

Definition 4.3 Suppose σ is qf-small.

- For any pair $p(\overline{x}), q(\overline{xy}) \in S_{\mathrm{qf}}(\sigma)$, define $\sigma_{p,q} = \forall \overline{x}(p(\overline{x}) \rightarrow \exists \overline{y}\ q(\overline{xy}))$.
- Set $\sigma^* = \sigma \wedge \bigwedge_{p,q \in S_{\mathrm{qf}}(\sigma);\, p \subset q} \sigma_{p,q}$.

If σ is preserved under chains, then σ^* is as well. However, there are consistent σ for which σ^* is inconsistent. An example would be the sentence $\sigma = \forall a, b, c, d(R(a, b) \wedge R(c, d) \rightarrow a = c \wedge b = d)$, which expresses that exactly two points are R-related.

Proposition 4.4 *For any σ, if σ^* is consistent, then it is complete.*

Proof. We show $\aleph_0$-categoricity. Let $M, N \models \sigma^*$ be countable and suppose f is a finite partial isomorphism mapping a tuple $\overline{a} \in M$ to a tuple $\overline{b} \in N$. Now let $c \in M$ be any point and set $p = \mathrm{tp}_{\mathrm{qf}}(\overline{a})$ $(= \mathrm{tp}_{\mathrm{qf}}(\overline{b}))$ and $q = \mathrm{tp}_{\mathrm{qf}}(\overline{a}c)$. Since $N \models \sigma_{p,q}$, we find a $d \in N$ with $\overline{b}d \models q$, so we can extend f by mapping c to d. Now after enumerating both M and N we can construct a total isomorphism as the union of finite partial isomorphisms by adding every point of M to the domain and every point of N to the range eventually. □

Definition 4.5 A sentence $\sigma \in L_{\omega_1,\omega}$ has the *extension property for countable models* (EPC) if, for any countable $M \models \sigma$ and $p(\overline{x}) \subset q(\overline{xy})$ in $S_{\mathrm{qf}}(\sigma)$, whenever some $\overline{a} \in M$ realizes p, there is a countable $N \models \sigma$ with $M \subset N$ containing some $\overline{b}$ with $\overline{a}\overline{b} \models q$.

Theorem 4.6 *Suppose that $\sigma \in L_{\omega_1,\omega}$ is preserved under chains, is qf-small and has the EPC. Then:*

(1) *σ^* is consistent.*

(2) *Any countable model of σ has an extension that is a model of σ^*.*

(3) *σ^* is the only completion of σ with property* (2) *that is preserved under chains.*

Proof. Let $M \models \sigma$ be countable. Enumerate all possible pairs $(\overline{a}, q)$ where $\overline{a} \in M$ and $\mathrm{tp}_{\mathrm{qf}}(\overline{a}) \subset q \in S_{\mathrm{qf}}(\sigma)$ as $((\overline{a}_n, q_n))_{n<\omega}$. Construct a $\subset$-chain $(M_n)_{n<\omega}$ of models of σ such that in M_n we add a tuple $\overline{b}_n$ with the property that $\overline{a}_n\overline{b}_n \models q_n$. Let $M^1 = \bigcup_{n<\omega} M_n$. Do the same procedure for M^1 in place of M to get some M^2. Repeat this ω many more times and set $N = \bigcup_{k<\omega} M^k$. Since σ is preserved under chains we still have $N \models \sigma$, and we just added all necessary witnesses in the chains to satisfy all $\sigma_{p,q}$ as well, so we have constructed a model of σ^* that contains the model M we started with.

The uniqueness of σ^* follows from the fact that if some τ has the same properties, including being preserved under chains, we can form a $\subset$-chain $(M_n)_{n<\omega}$ with $M_{2n} \models \sigma^*$ and $M_{2n+1} \models \tau$ for all n. Then by preservation under chains, the union must be a model of both σ^* and τ and we conclude by completeness of both sentences. □

Now we turn to the definition of an incomplete sentence coding $\aleph_2$ Kurepa families, which we will then complete by the described technique.

Our language will be $\mathcal{L} = \{S, L, U, V, E_n, <, R, F, G, H\}_{n<\omega}$, where S and L are unary predicates, all E_n as well as $U, V, <, R$ are binary relations and F, G and H are ternary relations.

Before we give the formal definition of our sentence, we describe informally what a model of it looks like:

- $(L, <)$ is a linear order.
- The elements of S code subsets of L via the relation R such that any two of them coincide on an initial segment of L with a maximum element and are disjoint above that initial segment.
- F defines a binary function mapping two elements of S to the point of L where they become disjoint.
- For every $a \in L$, U and V, define sets $U_a = \{x \mid U(a,x)\}$, $V_a = \{x \mid V(a,x)\}$ and all those sets are pairwise disjoint.
- The E_n are such that every set U_a and V_a with the restrictions of the E_n satisfies the theory of binary splitting equivalence relations, given in Section 2. In particular, all these sets have size at most $2^{\aleph_0} = \aleph_1$.
- G codes bijections between every initial segment $\{x \mid x < a\}$ and the set U_a. This makes $(L, <)$ $\aleph_2$-like.
- H codes intersections of sets coded by elements of S with initial segments $\{x \mid x < a\}$ as elements of V_a. Consequently, on each initial segment, there are at most $\aleph_1$ many possibilities for the sets coded by elements of S.

Let σ be the conjunction of the following statements:

(A1) Both $U(x,y)$ or $V(x,y)$ imply $x \in L$. Writing $U_x = \{y \mid U(x,y)\}$ and $V_x = \{y \mid V(x,y)\}$, the sets L, S, U_x, V_x (for all $x \in L$) are pairwise disjoint and their union is everything.

(A2) All E_n define equivalence relations on every set U_x and V_x where on every U_x or V_x, E_0 has exactly two classes and every E_n-class is the union of exactly two E_{n+1}-classes. In addition, $\bigwedge_{n<\omega} xE_ny$ implies $x = y$.

(A3) $<$ is a linear ordering of L. For $x \in L$, we write $L_{<x} = \{y \in L \mid y < x\}$ and $L_{\leq x} = L_{<x} \cup \{x\}$.

(A4) $F(s,t,x)$ implies $s, t \in S$ and $x \in L$. F defines a symmetric function from $S \times S$ to L.

(A5) $R \subset S \times L$. For $s \in S$, we write $R_s = \{x \in L \mid R(s,x)\}$. For any two distinct $s, t \in S$, R_s and R_t are identical on $L_{\leq F(s,t)}$ and disjoint on $L \setminus L_{\leq F(s,t)}$.

(A6) $G(x,y,z)$ implies $x \in L$, $y < x$ and $z \in U_x$. For every $x \in L$, $G(x,\cdot,\cdot)$ defines a bijective function $G_x \colon L_{<x} \to U_x$ by $G_x(y) = z$ if and only if $G(x,y,z)$.

(A7) $H(x,y,z)$ implies $x \in L$, $y \in S$ and $z \in V_x$. For every $x \in L$, $H(x,\cdot,\cdot)$ defines a surjective function $H_x \colon S \to V_x$ by $H_x(y) = z$ if and only if $H(x,y,z)$. H_x has the property that $H_x(s) = H_x(t)$ if and only if $F(s,t) \geq x$.

It is easy to construct a model of σ, but σ is not a complete sentence. We verify that it satisfies the hypotheses of Theorem 4.6. The axioms are all at most Π_2-statements, so we have preservation under chains. Also, since the equivalence relations E_n are refining and $\mathcal{L} \setminus \{E_n\}_{n<\omega}$ is finite, $S_{\mathrm{qf}}(\sigma)$ is countable.

Towards showing EPC, let $M \models \sigma$ be countable, $\overline{a} = (a_1, \ldots, a_n) \in M$ and let $p(\overline{x}), q(\overline{x}, y) \in S_{\mathrm{qf}}(\sigma)$ with $\overline{a} \models p$ and $p \subset q$ (note that it suffices to consider a single variable y instead of an arbitrary tuple). We want to find some countable $N \supset M$ and $b \in N$ such that $\overline{a}b \models q$. There are several cases:

- Suppose $S(y) \in q(\overline{x}, y)$. We will add a new point y to S and define a set R_y respecting the requirements of q and the axioms of σ. The requirements can be $R(y, x_i)$, $\neg R(y, x_i)$ as well as $F(y, x_j) = x_i$, $F(y, x_j) \neq x_i$ and $H_{x_i}(y) = x_j$, $H_{x_i}(y) \neq x_j$, $H_{x_i}(y) E_n x_j$, $\neg H_{x_i}(y) E_n x_j$ for components x_i, x_j in $\overline{x}$ and $n < \omega$ (G does not matter here since it does not involve elements from S; also note that conditions like $F(y, x_j) > x_i$ translate to $F(y, x_j) = x_k \wedge x_k > x_i$ since F is not a function but a relation symbol).

  Consider the set of all elements $z \in L$ occurring in $\overline{x}$ such that one of the following holds:

  (i) $q \vdash F(y, x_i) = z$ for some x_i in $\overline{x}$, or
  (ii) $q \vdash R(y, z)$ and there is some $s \in S$ with $M \models R(s, z)$, or
  (iii) $q \vdash H_z(y) = x_i$ and $M \models H_z(s) = x_i$ for some x_i in $\overline{x}$ and $s \in S$.

  Let $A = \{a \in L \mid q \vdash R(y, a)\}$. We now have two cases:

  - There is no such element. In this case, we choose any $c \in L$ that is smaller than any element of $\overline{x}$, as well as an arbitrary element $s \in S$. We set $R_y = A \cup (R_s \cap L_{\leq c})$ and naturally $F(y, s) = c$ (note that R_y and every R_t ($t \in S$) are disjoint above c since (ii) fails).
  - There is such an element and let z be the maximal such. We set $R_y = A \cup (R_{x_i} \cap L_{\leq z})$ if z satisfies (i) and $R_y = A \cup (R_s \cap L_{\leq z})$ in the cases (ii) and (iii) (choose any such s arbitrarily). If we are in the case (ii) or (iii) and q implies $F(y, s) \neq z$, we also add a new element w to L which is greater than z and smaller than any element of $\overline{x}$ that is larger than z, and we declare $R(s, w)$, $R(y, w)$, $F(s, y) = w$.

  In either of the two cases, we will have to turn M with the additional y (and possibly w) into a model of σ. We have to set the F- and H-relations which can be done straightforwardly (respecting possible requirements from q for H; we may have to add new points to sets V_a for $a > z$). In case we added the point w, we also have to add new sets U_w, V_w as well as a new point to each U_a for $a > w$, and extend G accordingly.
- Now suppose $L(y) \in q(\overline{x}, y)$. Add a new element z to L for y in an arbitrary cut that complies with the conditions $x_i < y$ or $x_i > y$ contained in q. Add $R(x_i, z)$ whenever demanded by q and for any other $s \in S$ add $R(s, z)$ if and only if $R(t, z)$ and $F(s, t) > z$ for some element $t \in S$. Finally, we have to add new sets U_z and V_z as well as a new point a to each U_w with $w > z$ and declare $G(w, z, a)$. We may have to add a new point to sets V_w for $w > z$ too.
- Should $U_{x_i}(y)$ or $V_{x_i}(y)$ belong to q, it is easy to see that there must already be some $b \in M$ with $\overline{a}b \models q$.

Now we apply Theorem 4.6 to σ. Immediately we see that σ^* implies:

- The ordering on L is dense without endpoints.
- Every set R_s is dense (and thus unbounded) and co-dense in L.
- $s \neq t$ implies $R_s \neq R_t$ ("R is extensional").

But we know more about the properties of σ^*. The countable model of σ^* is extendible, so there is an uncountable model. In addition, we have seen in the verification of EPC that we have a lot of freedom in adding new elements to countable models of σ, and thus to models of σ^*, so that we can conclude the existence of models of σ^* with:

- $(L, <)$ isomorphic to a proper initial segment of $\eta_1 \cdot \omega_2$, where η_1 is the saturated dense linear order without endpoints of size $\aleph_1$ (we assume GCH).
- All $(U_x, E_n)_{n<\omega}$ and $(V_x, E_n)_{n<\omega}$ isomorphic to $(2^\omega, F_n)$ where we define $\xi F_n \rho$ if and only if $\xi(k) = \rho(k)$ for all $k \leq n$.

Now we consider the class $\mathbb{P}$ of all such models with the additional properties:

- $(L, <)$ is an initial segment of $(\eta_1 \cdot \omega_2, <)$.
- S is a subset of ω_3 of size $\aleph_1$ (so all models in $\mathbb{P}$ will have size $\aleph_1$).
- The sets U_x and V_x $(x \in L)$ equal $2^\omega \times \{(x, 0)\}$ and $2^\omega \times \{(x, 1)\}$ respectively and the E_n defined on them are the natural ones (compare with F_n above).

We order the elements of $\mathbb{P}$ by the superstructure relation $\supset$. Since σ^* is preserved under unions, the poset $(\mathbb{P}, \supset)$ is ω_2-closed (meaning that every sequence of length less than ω_2 of elements of $\mathbb{P}$ has a lower $\supset$-bound; clearly the union of the chain of models will do).

Now we show that $(\mathbb{P}, \supset)$ has the ω_3-cc. Take any $X \subset \mathbb{P}$ of size $\aleph_3$. We shall find two elements of X which have a common extension. By the pigeonhole-principle and the delta-system-lemma, we may assume that

- the collection of the underlying sets of the models in X form a delta-system;
- the L-part of all models in X is identical;
- the $\mathcal{L}$-structure of all models in X is identical of the root of the delta-system;
- the collection of sets R_s $(s \in S)$ is identical for all elements of X.

Two models $M, N \in X$ may only differ on their S-part. We would like to make the union $M \cup N$ into a model of σ. The problem is that if the models are not already identical, there will be $x \in S^M$, $y \in S^N$ outside the root such that $R_x = R_y$, so $F(x, y)$ cannot be defined in such a way that axiom (A5) holds. The solution is to end-extend L in order to make R_x and R_y disjoint on a final segment.

Suppose that in $\eta_1 \cdot \omega_2$, L is an initial segment contained in $\{x \mid x < a\}$. Enumerate the elements of $S^M \setminus S^N$ as $(s_\alpha)_{\alpha<\mu}$ (for some $\mu \leq \aleph_1$). Now inductively do the following: given $\alpha < \omega_1$ there is a unique $t \in S^N \setminus S^M$ such that $R_{s_\alpha} = R_t$. Set $R(s_\alpha, a)$, $R(t, a)$, $F(s_\alpha, t) = a$ and $R(s_\alpha, a_\alpha)$ (but *not* $R(t, a_\alpha)$), where $a_\alpha \in \eta_1 \cdot \omega_2$ is greater than a and any already chosen a_β $(\beta < \alpha)$. Now we have to add sets U_{a_α} and V_{a_α} and extend G and H to get a model M' of σ containing both M and N. Note that we do not have to add any point to the U_x, V_x for $x \in L^M$, which is fortunate since that would be impossible.

Having obtained a model M' of σ containing both M and N as submodels, our final task in proving ω_3-cc is to extend M' to an element M'' of $\mathbb{P}$. In particular, we want M'' to have the following properties:

- M'' must be a model of σ^*.
- $L^{M''}$ must be an initial segment of $\eta_1 \cdot \omega_2$.
- The sets $U_x^{M''}$ and $V_x^{M''}$ must be equal to $2^\omega \times \{(x, 0)\}$ and $2^\omega \times \{(x, 1)\}$ respectively.

We will contruct a continuous chain $(M_\alpha)_{\alpha<\omega_1}$ of models of σ starting from $M_0 = M'$ such that $M'' = \bigcup_{\alpha<\omega_1} M_\alpha$ satisfies our requirements. We have several sets of "tasks" (each enumerated in order type ω_1) that we want to perform along that chain:

- Let W be an enumerated set of the tasks "add the element w to the L-part of the so far constructed model" for any $w \in \eta_1 \cdot \omega_2$ that is smaller than some a_α (we constructed the elements a_α above). Thus, after performing all tasks in W, the L-part of M'' will be an initial segment of $\eta_1 \cdot \omega_2$ (the smallest one containing all a_α).

- Having reached stage α of the chain, enumerate all pairs $(\overline{a}, q)$ with $\overline{a} \in M_\alpha$ and $\mathrm{tp}_{\mathrm{qf}}(\overline{a}) \subset q \in S_{\mathrm{tp}_{\mathrm{qf}}}(\sigma)$ as $T_\alpha = ((\overline{a}_\beta, q_\beta))_{\beta<\omega_1}$ (compare with the proof of Theorem 4.6). Designate the set of tasks "add a tuple $\bar{b}$ such that $\overline{a}_\beta \bar{b} \models q_\beta$" by T_α.
- Having reached stage α of the chain, let X_α and Y_α respectively be enumerated sets of the tasks "add the element $(\sigma, (x, 0))$ to U_x" and "add the element $(\sigma, (x, 1))$ to V_x" for all $\sigma \in 2^\omega$ and $x \in L^{M_\alpha}$.

At each stage α of the chain, add elements to the model M_α such that the least task in W as well as in all $T_\beta, X_\beta, Y_\beta$ $(\beta \le \alpha)$ is performed. Then remove those tasks from the sets $W, T_\beta, X_\beta, Y_\beta$. By possibly adding additional elements, we can do this while obtaining a model $M_{\alpha+1}$ of σ (the arguments and techniques are the same as in the proof that σ has EPC).

Note that at each stage we only have to add countably many elements in each U_x, V_x and the L-part of M_α, so we do not encounter the problem of saturating at a countable stage of the chain-construction the U_x, V_x for $x > a$ (a as defined above) or any part of L above a. This would be a serious problem as for example adding a new element w to the order requires adding a *new* element to the U_x with $x > w$ (because of the properties of G). We are thus able to carry out the construction through all countable ordinals and obtain M'' as the union of the chain with the required properties. This concludes the proof of ω_3-cc.

Let G be a $\mathbb{P}$-generic filter over V. Then $\bigcup G$ will be a model of σ^* of size $\aleph_3^V$. But since the forcing is ω_2-closed and has ω_3-cc, all cardinals are preserved and in particular $\aleph_3^{V[G]} = \aleph_3^V$. That is, we get a model of σ^* of size $\aleph_3$ in a generic extension. On the other hand, any such model codes an $\aleph_2$ Kurepa family which means that it is consistent with ZFC + GCH (assuming the existence of an inaccessible cardinal and noting that the forcing preserves GCH[2]) that σ^* has no model of size $\aleph_3$.

5 Final observations

The question of absoluteness of model-existence (under ZFC + GCH) in $\aleph_\omega$ remains open. On the other hand, the technique of finding complete examples described in Section 4 should be applicable more widely to obtain complete examples of non-absoluteness of model existence (under ZFC + GCH) in cardinals greater than $\aleph_3$. Interestingly, however, this method seems to be problematic for finding examples for model existence in $\aleph_2$, at least with the approach of trying to code Kurepa families. The reason is that it seems difficult to code an ω_1-like ordering without making many elements definable over others

[2] This is a standard argument: For any (infinite) κ, each subset of κ added by the forcing is of the form $\{\alpha < \kappa \mid G \cap A_\alpha \text{ is nonempty}\}$, where $\vec{A} = (A_\alpha)_{\alpha<\kappa}$ is in the ground model and each A_α is an antichain in the forcing. This is because we can take a name σ for the given set, let B_α be a maximal antichain consisting of conditions which decide "$\alpha \in \sigma$" and take A_α to consist of the elements of B_α which force "$\alpha \in \sigma$".

As GCH holds in the ground model and the forcing has ω_3-cc, the fact that the forcing has size ω_3 implies that there are only $((\omega_3)^{\omega_2})^\kappa = (\omega_3)^\kappa$ many (in the sense of the ground model) such sequences $\vec{A}$. For $\kappa \ge \omega_2$, this is $2^\kappa = \kappa^+$. As the forcing does not add subsets of ω_1, the GCH will also hold at ω and ω_1.

(or even getting infinite definable closures over finite tuples), which destroys any chance to have EPC.

As a last remark, our use of the concept of Kurepa families has the slight flaw that in order to find set-theoretic universes which do not contain such families, we have to assume the existence of inaccessible cardinals. For the special Aronszajn technique, we even have to assume the consistency of supercompact cardinals. It would be nice to find $L_{\omega_1,\omega}$ sentences for which under GCH the existence of models of certain cardinalities is not absolute, without assuming the existence of large cardinals.

Acknowledgements

We would like to express our gratitude to the John Templeton Foundation for supporting the Infinity Project (hosted by the CRM in Barcelona) in which the presented work has been accomplished. We also greatly thank John Baldwin, Fred Drueck, Rami Grossberg and Andrés Villaveces for helpful discussions at the CRM, especially on the idea of coding Kurepa families into models to show non-absoluteness of model existence. Rami Grossberg in particular suggested generalizing a theory coding an $\aleph_1$ Kurepa family to higher cardinalities to achieve non-absoluteness of model existence in cardinalities above $\aleph_2$. He also was the first to recognize a problem in dealing with $\aleph_\omega$ by this construction.

References

[1] Baldwin, J. T., *Amalgamation, absoluteness and categoricity*, Proceedings of Southeast Asia Logic Conference (June 2009), to appear. Available at www.math.uic.edu/jbaldwin.

[2] Barwise, K. J., *The role of the omitting types theorem in infinitary logic*, Archiv für Mathematische Logik und Grundlagenforschung, vol. 21, 1981, pp. 55–68.

[3] Chang, C. C. and Keisler, H. J., *Model Theory*, North-Holland Publishing Company, Amsterdam, 1990.

[4] Cummings, J., Foreman, M. and Magidor, M., *Squares, scales and stationary reflection*, Journal of Mathematical Logic, vol. 1, 2001, pp. 35–98.

[5] Gao, S., *On automorphism groups of countable structures*, The Journal of Symbolic Logic, vol. 63(3), 1998, pp. 891–896.

[6] Gregory, J., *Elementary extensions and uncountable models for infinitary finite quantifier language fragments. Preliminary report*, Notices of the American Mathematical Society, vol. 17, 1970, pp. 967–968.

[7] Grossberg, R., Shelah, S., *On the number of non isomorphic models of an infinitary theory which has the order property, part A*, Journal of Symbolic Logic, vol. 51, 1986, pp. 302–322.

[8] Hjorth, G., *Knight's model, its automorphism group, and characterizing the uncountable cardinals*, Journal of Mathematical Logic, vol. 2(1), 2002, pp. 113–144.

[9] Jech, T., *Set Theory*, Springer Monographs in Mathematics, 2002.

[10] Jensen, R. B., *The fine structure of the constructible hierarchy. With a section by Jack Silver*, Annals of Mathematical Logic, vol. 4, 1972, pp. 229–308.

[11] Keisler, H. J., *Model Theory for Infinitary Logic*, North-Holland Publishing Company, Amsterdam, 1971.

[12] Malitz, J., *The Hanf number for complete $L_{\omega_1,\omega}$-sentences*, in: The Syntax and Semantics of Infinitary Languages (J. Barwise, ed.), Lecture Notes in Mathematics, vol. 72, Springer, 1968, pp. 166–181.

[13] Specker, E., *Sur un problème de Sikorski*, Colloquium Mathematicum, vol. 2, 1949, pp. 9–12.

Generalized descriptive set theory and classification theory

Sy-David Friedman[†], **Tapani Hyttinen**[‡], **Vadim Kulikov**[‡]

[†] Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
sdf@logic.univie.ac.at

[‡] Department of Mathematics and Statistics, University of Helsinki, Finland
tapani.hyttinen@helsinki.fi, vadim.kulikov@helsinki.fi

Abstract. Descriptive set theory is mainly concerned with studying subsets of the space of all countable binary sequences. In this paper we study the generalization where countable is replaced by uncountable. We explore properties of generalized Baire and Cantor spaces, equivalence relations and their Borel reducibility. The study shows that the descriptive set theory looks very different in this generalized setting compared to the classical, countable case. We also draw the connection between the stability theoretic complexity of first-order theories and the descriptive set theoretic complexity of their isomorphism relations. Our results suggest that Borel reducibility on uncountable structures is a model theoretically natural way to compare the complexity of isomorphism relations.

History and motivation

There is a long tradition in studying connections between Borel structure of Polish spaces (descriptive set theory) and model theory. The connection arises from the fact that any class of countable structures can be coded into a subset of the space 2^ω provided all structures in the class have domain ω. A survey on this topic is given in [**9**]. Suppose X and Y are subsets of 2^ω and let E_1 and E_2 be equivalence relations on X and Y respectively. If $f\colon X \to Y$ is a map such that $E_1(x,y) \Leftrightarrow E_2(f(x),f(y))$, we say that f is a *reduction of* E_1 *to* E_2. If there exists a Borel or continuous reduction, we say that E_1 is Borel or continuously *reducible* to E_2, denoted $E_1 \leqslant_B E_2$ or $E_1 \leqslant_c E_2$. The mathematical meaning of this is that *f classifies E_1-equivalence in terms of E_2-equivalence.*

The benefit of various reducibility and irreducibility theorems is roughly the following. A reducibility result, say $E_1 \leqslant_B E_2$, tells us that E_1 is at most as complicated as E_2; once you understand E_2, you understand E_1 (modulo the reduction). An irreducibility result, $E_1 \nleqslant_B E_2$ tells that there is no hope in trying to classify E_1 in terms of E_2, at least in a "Borel way". From the model theoretic point of view, the isomorphism relation, and the elementary equivalence relation (in some language) on some class of structures are the equivalence relations of main interest. But model theory in general does not restrict itself to countable structures. Most of stability theory and Shelah's classification theory characterizes first-order theories in terms of their uncountable models. This leads to the generalization adopted in this paper. We consider the space 2^κ for an uncountable cardinal κ with the idea that models of size κ are coded into elements of that space.

This approach to connect such uncountable descriptive set theory with model theory began in the early 1990's. One of the pioneering papers was authored by Mekler and

The authors wish to thank the John Templeton Foundation for its generous support of this research through the CRM Infinity Project (Grant ID 13152).

Väänänen [**24**]. A survey of the research done in the 1990's can be found in [**35**], and a discussion of the motivational background for this work in [**36**]. A more recent account is given in [**37**, Chapter 9.6].

Let us explain how our approach differs from the earlier ones and why it is useful. For a first-order complete countable theory in a countable vocabulary T and a cardinal $\kappa \geqslant \omega$, define

$$S_T^\kappa = \{\eta \in 2^\kappa \mid \mathcal{A}_\eta \models T\} \quad \text{and} \quad \cong_T^\kappa = \{(\eta,\xi) \in (S_T^\kappa)^2 \mid \mathcal{A}_\eta \cong \mathcal{A}_\xi\},$$

where $\eta \mapsto \mathcal{A}_\eta$ is some fixed coding of (all) structures of size κ. We can now define the partial order on the set of all theories as above by

$$T \leqslant^\kappa T' \iff \cong_T^\kappa \leqslant_B \cong_{T'}^\kappa .$$

As pointed out above, $T \leqslant^\kappa T'$ says that $\cong_T^\kappa$ is at most as difficult to classify as $\cong_{T'}^\kappa$. But does this tell us whether T is a simpler theory than T'? Rough answer: *If $\kappa = \omega$, then no, but if $\kappa > \omega$, then yes.*

To illustrate this, let $T = \mathrm{Th}(\mathbb{Q}, \leqslant)$ be the theory of the order of the rational numbers (DLO) and let T' be the theory of a vector space over the field of rational numbers. Without loss of generality we may assume that they are models of the same vocabulary. It is easy to argue that the model class defined by T' is strictly simpler than that of T. (For instance, there are many questions about T, unlike T', that cannot be answered in ZFC; say existence of a saturated model.) On the other hand $\cong_T^\omega \leqslant_B \cong_{T'}^\omega$ and $\cong_{T'}^\omega \not\leqslant_B \cong_T^\omega$ because there is only one countable model of T and there are infinitely many countable models of T'. But for $\kappa > \omega$ we have $\cong_T^\kappa \not\leqslant_B \cong_{T'}^\kappa$ and $\cong_{T'}^\kappa \leqslant_B \cong_T^\kappa$, since there are 2^κ equivalence classes of $\cong_T^\kappa$ and only one equivalence class of $\cong_{T'}^\kappa$.

Another example, introduced in Martin Koerwien's Ph.D. thesis and his article [**19**], shows that there exists an ω-stable theory without DOP and without OTOP with depth 2 for which $\cong_T^\omega$ is not Borel, while we show here that, for $\kappa^{<\kappa} = \kappa > 2^\omega$, $\cong_T^\kappa$ is Borel for all classifiable shallow theories (*shallow* is the opposite of *deep*). The converse holds for all κ with $\kappa^{<\kappa} = \kappa > \omega$: if $\cong_T^\kappa$ is Borel, then T is classifiable and shallow; see Theorems 4.1, 4.6 and 4.7 starting from page 520.

Our results suggest that the order $\leqslant^\kappa$ for $\kappa > \omega$ corresponds naturally to the classification of theories in stability theory: the more complex a theory is from the viewpoint of stability theory, the higher it seems to sit in the ordering $\leqslant^\kappa$ and vice versa. Since dealing with uncountable cardinals often implies the need for various cardinality or set theoretic assumptions beyond ZFC, the results are not always as simple as in the case $\kappa = \omega$, but they tell us a lot. For example, our results easily imply the following (modulo some mild cardinality assumptions on κ):

- If T is deep and T' is shallow, then $\cong_T \not\leqslant_B \cong_{T'}$.
- If T is unstable and T' is classifiable, then $\cong_T \not\leqslant_B \cong_{T'}$.

1 Introduction

1.1 Notations and conventions

1.1.1 Set theory

We use standard set theoretical notation:

- $A \subset B$ means that A is a subset of B or is equal to B.
- $A \subsetneq B$ means proper subset.

- Union, intersection and set theoretical difference are denoted respectively by $A \cup B$, $A \cap B$ and $A \setminus B$. For larger unions and intersections, $\bigcup_{i \in I} A_i$ etc.
- Symmetric difference: $A \triangle B = (A \setminus B) \cup (B \setminus A)$.
- $\mathcal{P}(A)$ is the power set of A and $[A]^{<\kappa}$ is the set of subsets of A of size $< \kappa$.

Usually the Greek letters κ, λ and μ will stand for cardinals, and α, β and γ for ordinals, but this is not strict. Also η, ξ, ν are usually elements of κ^κ or 2^κ, and p, q, r are elements of $\kappa^{<\kappa}$ or $2^{<\kappa}$. We denote by $\operatorname{cf}(\alpha)$ the cofinality of α (the least ordinal β for which there exists an increasing unbounded function $f\colon \beta \to \alpha$).

By S^κ_λ we mean $\{\alpha < \kappa \mid \operatorname{cf}(\alpha) = \lambda\}$. A λ-*cub set* is a subset of a limit ordinal (usually of cofinality $> \lambda$) which is unbounded and contains suprema of all bounded increasing sequences of length λ. A set is *cub* if it is λ-cub for all λ. A set is *stationary* if it intersects all cub sets and λ-*stationary* if it intersects all λ-cub sets. Note that $C \subset \kappa$ is λ-cub if and only if $C \cap S^\kappa_\lambda$ is λ-cub and $S \subset \kappa$ is λ-stationary if and only if $S \cap S^\kappa_\lambda$ is (just) stationary.

If $(\mathbb{P}, \leqslant)$ is a forcing notion, we write $p \leqslant q$ if p and q are in $\mathbb{P}$ and q forces more than p. Usually $\mathbb{P}$ is a set of functions equipped with inclusion and $p \leqslant q \Leftrightarrow p \subset q$. In that case, $\varnothing$ is the weakest condition and we write $\mathbb{P} \Vdash \varphi$ to mean $\varnothing \Vdash_{\mathbb{P}} \varphi$. By *Cohen forcing* or *standard Cohen forcing* we mean the partial order $2^{<\kappa}$ of partial functions from κ to $\{0,1\}$ ordered by inclusion, where κ depends on the context.

1.1.2 Functions

We denote by $f(x)$ the value of x under the mapping f and by $f[A]$ or just fA the image of the set A under f. Similarly $f^{-1}[A]$ or just $f^{-1}A$ indicates the inverse image of A. Domain and range are denoted respectively by $\operatorname{dom} f$ and $\operatorname{ran} f$.

If it is clear from the context that f has an inverse, then f^{-1} denotes that inverse. For a map $f\colon X \to Y$, *injective* means the same as *one-to-one*, and *surjective* the same as *onto*.

Suppose that $f\colon X \to Y^\alpha$ is a function with range consisting of sequences of elements of Y of length α. The projection pr_β is a function $Y^\alpha \to Y$ defined by $\operatorname{pr}_\beta((y_i)_{i<\alpha}) = y_\beta$. For the coordinate functions of f we use the notation $f_\beta = \operatorname{pr}_\beta \circ f$ for all $\beta < \alpha$.

By the *support* of a function f we mean the subset of $\operatorname{dom} f$ in which f takes non-zero values, whatever "zero" means depending on the context (hopefully never unclear). The support of f is denoted by $\operatorname{sprt} f$.

1.1.3 Model theory

In the section *Coding models*, on page 478, we fix a countable vocabulary and assume that all theories are theories in this vocabulary. Moreover we assume that they are first-order, complete and countable. By $\operatorname{tp}(\overline{a}/A)$ we denote the complete type of $\overline{a} = (a_1, \dots, a_{\operatorname{len} \overline{a}})$ over A, where $\operatorname{len} \overline{a}$ is the length of the sequence $\overline{a}$.

We think of models as tuples $\mathcal{A} = \langle \operatorname{dom} \mathcal{A}, P_n^{\mathcal{A}} \rangle_{n<\omega}$ where the P_n are relation symbols in the vocabulary and the $P_n^{\mathcal{A}}$ are their interpretations. If a relation R has arity n (a property of the vocabulary), then for its interpretation it holds that $R^{\mathcal{A}} \subset (\operatorname{dom} \mathcal{A})^n$. In the section *Coding models* we adopt more conventions concerning this.

In sections *The Silver dichotomy for isomorphism relations* (page 492) and *Complexity of isomorphism relations* (page 520) we will use the following stability theoretical notions: stable, superstable, DOP, OTOP, shallow and $\kappa(T)$. Classifiable means superstable with no DOP nor OTOP, and the least cardinal in which T is stable is denoted by $\lambda(T)$.

1.1.4 Reductions

Let $E_1 \subset X^2$ and $E_2 \subset Y^2$ be equivalence relations on X and Y respectively. A function $f\colon X \to Y$ is *a reduction* of E_1 to E_2 if for all $x, y \in X$ we have that $xE_1y \Leftrightarrow f(x)E_2f(y)$. Suppose in addition that X and Y are topological spaces. Then we say that E_1 is *continuously reducible to* E_2 if there exists a continuous reduction from E_1 to E_2, and we say that E_1 is *Borel reducible to* E_2 if there is a Borel reduction. For the definition of Borel adopted in this paper, see Definition 1.17. We denote the fact that E_1 is continuously reducible to E_2 by $E_1 \leqslant_c E_2$ and respectively Borel reducibility by $E_1 \leqslant_B E_2$.

We say that relations E_2 and E_1 are (Borel) *bireducible* to each other if $E_2 \leqslant_B E_1$ and $E_1 \leqslant_B E_2$.

1.2 Ground work

1.2.1 Trees and topologies

Throughout the paper κ is assumed to be an uncountable regular cardinal which satisfies

$$\kappa^{<\kappa} = \kappa \tag{1.1}$$

(For justification of this, see below.) We look at the space κ^κ (the generalized Baire space), i.e., the functions from κ to κ and the space formed by the initial segments $\kappa^{<\kappa}$. It is useful to think of $\kappa^{<\kappa}$ as a tree ordered by inclusion and of κ^κ as a topological space of the branches of $\kappa^{<\kappa}$; the topology is defined below. Occasionally we work in 2^κ (the generalized Cantor space) and $2^{<\kappa}$ instead of κ^κ and $\kappa^{<\kappa}$.

Definition 1.1 A *tree* t is a partial order with a root in which the sets $\{x \in t \mid x < y\}$ are well ordered for each $y \in t$. A *branch* in a tree is a maximal linear suborder.

A tree is called *a $\kappa\lambda$-tree* if there are no branches of length λ or higher and no element has $\geqslant \kappa$ immediate successors. If t and t' are trees, we write $t \leqslant t'$ to mean that there exists an order preserving map $f\colon t \to t'$ such that $a <_t b \Rightarrow f(a) <_{t'} f(b)$.

Convention Unless otherwise said, by a tree $t \subset (\kappa^{<\kappa})^n$ we mean a tree with domain being a downward closed subset of

$$(\kappa^{<\kappa})^n \cap \{(p_0, \dots, p_{n-1}) \mid \operatorname{dom} p_0 = \dots = \operatorname{dom} p_{n-1}\}$$

ordered as follows: $(p_0, \dots, p_{n-1}) < (q_0, \dots, q_{n-1})$ if $p_i \subset q_i$ for all $i \in \{0, \dots, n-1\}$. It is always a $\kappa^+, \kappa + 1$-tree.

Example 1.2 Let $\alpha < \kappa^+$ be an ordinal and let t_α be the tree of descending sequences in α ordered by end extension. The root is the empty sequence. It is a $\kappa^+\omega$-tree. Such t_α can be embedded into $\kappa^{<\omega}$, but note that not all subtrees of $\kappa^{<\omega}$ are $\kappa^+\omega$-trees (there are also $\kappa^+, \omega + 1$-trees).

In fact the trees $\kappa^{<\beta}$, $\beta \leqslant \kappa$ and t_α are universal in the following sense:

Fact 1.3 ($\kappa^{<\kappa} = \kappa$) Assume that t is a $\kappa^+, \beta + 1$-tree, $\beta \leqslant \kappa$ and t' is $\kappa^+\omega$-tree. Then

(1) there is an embedding $f\colon t \to \kappa^{<\beta}$

(2) and a strictly order preserving map $f\colon t' \to t_\alpha$ for some $\alpha < \kappa^+$ (in fact there is also such an embedding f).

Define the topology on κ^κ as follows. For each $p \in \kappa^{<\kappa}$ define the basic open set

$$N_p = \{\eta \in \kappa^\kappa \mid \eta \restriction \operatorname{dom}(p) = p\}.$$

Open sets are precisely the empty set and the sets of the form $\bigcup X$, where X is a collection of basic open sets. Similarly for 2^κ.

There are many justifications for the assumption (1.1) which will be most apparent after seeing the proofs of our theorems. The crucial points can be summarized as follows: if (1.1) does not hold, then

- the space κ^κ does not have a dense subset of size κ;
- there are open subsets of κ^κ that are not κ-unions of basic open sets which makes controlling Borel sets difficult (see Definition 1.17 on page 480);
- Vaught's generalization of the López-Escobar theorem (Theorem 2.2, page 483) fails —see Remark 2.3 on page 485;
- the model theoretic machinery we are using often needs this cardinality assumption (see e.g. Theorem 2.8, page 487, and proof of Theorem 4.9, page 525).

Initially the motivation to assume (1.1) was simplicity. Many statements concerning the space $\kappa^{<\kappa}$ are independent of ZFC and using (1.1) we wanted to make the scope of such statements neater. In the statements of (important) theorems we mention the assumption explicitly.

Because the intersection of less than κ basic open sets is either empty or a basic open set, we get the following.

Fact 1.4 ($\kappa^{<\kappa}=\kappa$) The following hold for a topological space $P\in\{2^\kappa,\kappa^\kappa\}$:

(1) The intersection of less than κ basic open sets is either empty or a basic open set.
(2) The intersection of less than κ open sets is open.
(3) Basic open sets are closed.
(4) $|\{A\subset P : A \text{ is basic open}\}|=\kappa$.
(5) $|\{A\subset P : A \text{ is open}\}|=2^\kappa$.

In the space $\kappa^\kappa\times\kappa^\kappa=(\kappa^\kappa)^2$ we define the ordinary product topology.

Definition 1.5 A set $Z\subset\kappa^\kappa$ is Σ^1_1 if it is a projection of a closed set $C\subset(\kappa^\kappa)^2$. A set is Π^1_1 if it is the complement of a Σ^1_1-set. A set is Δ^1_1 if it is both Σ^1_1 and Π^1_1.

As in standard descriptive set theory ($\kappa=\omega$), we have the following:

Theorem 1.6 *For $n<\omega$ the spaces $(\kappa^\kappa)^n$ and κ^κ are homeomorphic.* □

Remark This standard theorem can be found for example in Jech's book [**16**]. Applying this theorem we can extend the concepts of Definition 1.5 to subsets of $(\kappa^\kappa)^n$. For instance a subset A of $(\kappa^\kappa)^n$ is Σ^1_1 if, for a homeomorphism $h\colon(\kappa^\kappa)^n\to\kappa^\kappa$, $h[A]$ is Σ^1_1 according to Definition 1.5.

1.2.2 Ehrenfeucht–Fraïssé games

We will need Ehrenfeucht–Fraïssé games in various connections. It serves also as a way of coding isomorphisms.

Definition 1.7 (Ehrenfeucht–Fraïssé games) Let t be a tree, κ a cardinal and $\mathcal{A}$ and $\mathfrak{B}$ structures with domains A and B respectively. Note that t might be an ordinal. The game $\mathrm{EF}^\kappa_t(\mathcal{A},\mathfrak{B})$ is played by players **I** and **II** as follows. Player **I** chooses subsets of $A\cup B$ and climbs up the tree t, and player **II** chooses partial functions $A\to B$ as follows. Suppose that a sequence

$$(X_i,p_i,f_i)_{i<\gamma}$$

has been played (if $\gamma = 0$, then the sequence is empty). Player **I** picks a set $X_\gamma \subset A \cup B$ of cardinality strictly less than κ such that $X_\delta \subset X_\gamma$ for all ordinals $\delta < \gamma$. Then player **I** picks a $p_\gamma \in t$ which is $<_t$-above all p_δ where $\delta < \gamma$. Then player **II** chooses a partial function $f_\gamma \colon A \to B$ such that $X_\gamma \cap A \subset \operatorname{dom} f_\gamma$, $X_\gamma \cap B \subset \operatorname{ran} f_\gamma$, $|\operatorname{dom} f_\gamma| < \kappa$ and $f_\delta \subset f_\gamma$ for all ordinals $\delta < \gamma$. The game ends when player **I** cannot go up the tree anymore, i.e., $(p_i)_{i<\gamma}$ is a branch. Player **II** wins if

$$f = \bigcup_{i<\gamma} f_i$$

is a partial isomorphism. Otherwise player **I** wins.

A *strategy* of player **II** in $\mathrm{EF}_t^\kappa(\mathcal{A}, \mathfrak{B})$ is a function

$$\sigma \colon ([A \cup B]^{<\kappa} \times t)^{<\mathrm{ht}(t)} \longrightarrow \bigcup_{I \in [A]^{<\kappa}} B^I,$$

where $[R]^{<\kappa}$ is the set of subsets of R of size $< \kappa$ and $\mathrm{ht}(t)$ is the *height* of the tree, i.e.,

$$\mathrm{ht}(t) = \sup\{\alpha \mid \alpha \text{ is an ordinal and there is an order preserving embedding } \alpha \to t\}.$$

A strategy of **I** is similarly a function

$$\tau \colon \left(\bigcup_{I \in [A]^{<\kappa}} B^I \right)^{<\mathrm{ht}(t)} \longrightarrow [A \cup B]^{<\kappa} \times t.$$

We say that a strategy τ of player **I** *beats* strategy σ of player **II** if the play $\tau * \sigma$ is a win for **I**. The play $\tau * \sigma$ is just the play where **I** uses τ and **II** uses σ. Similarly σ beats τ if $\tau * \sigma$ is a win for **II**. We say that a strategy is a *winning strategy* if it beats all opponent strategies. The notation $X \uparrow \mathrm{EF}_t^\kappa(\mathcal{A}, \mathfrak{B})$ means that player X has a winning strategy in $\mathrm{EF}_t^\kappa(\mathcal{A}, \mathfrak{B})$.

Remark By our convention, $\operatorname{dom} \mathcal{A} = \operatorname{dom} \mathfrak{B} = \kappa$, so while player **I** picks a subset of $\operatorname{dom} \mathcal{A} \cup \operatorname{dom} \mathfrak{B}$ he actually just picks a subset of κ, but as a small analysis shows, this does not alter the game.

Consider the game $\mathrm{EF}_t^\kappa(\mathcal{A}, \mathfrak{B})$, where $|\mathcal{A}| = |\mathfrak{B}| = \kappa$, $|t| \leqslant \kappa$ and $\mathrm{ht}(t) \leqslant \kappa$. The set of strategies can be identified with κ^κ, for example as follows. The moves of player **I** are members of $[A \cup B]^{<\kappa} \times t$ and the moves of player **II** are members of $\bigcup_{I \in [A]^{<\kappa}} B^I$. By our convention, $\operatorname{dom} \mathcal{A} = \operatorname{dom} \mathfrak{B} = A = B = \kappa$, so these become $V = [\kappa]^{<\kappa} \times t$ and $U = \bigcup_{I \in [\kappa]^{<\kappa}} \kappa^I$. By our cardinality assumption $\kappa^{<\kappa} = \kappa$, these sets are of cardinality κ.

Let

$$f \colon U \to \kappa, \quad g \colon U^{<\kappa} \to \kappa, \quad h \colon V \to \kappa, \quad k \colon V^{<\kappa} \to \kappa$$

be bijections. Let us assume that $\tau \colon U^{<\kappa} \to V$ is a strategy of player **I** (there cannot be more than κ moves in the game because we assumed $\mathrm{ht}(t) \leqslant \kappa$). Let $\nu_\tau \colon \kappa \to \kappa$ be defined by

$$\nu_\tau = h \circ \tau \circ g^{-1}$$

and, if $\sigma \colon V^{<\kappa} \to U$ is a strategy of player **II**, let ν_σ be defined by

$$\nu_\sigma = f \circ \sigma \circ k^{-1}.$$

We say that ν_τ *codes* τ.

Theorem 1.8 ($\kappa^{<\kappa} = \kappa$) *Let $\lambda \leqslant \kappa$ be a cardinal. The set*

$$C = \{(\nu, \eta, \xi) \in (\kappa^\kappa)^3 \mid \nu \text{ codes a winning strategy of } \mathbf{II} \text{ in } \mathrm{EF}^\kappa_\lambda(\mathcal{A}_\eta, \mathcal{A}_\xi)\} \subset (\kappa^\kappa)^3$$

is closed. If $\lambda < \kappa$, then also the corresponding set for player **I**

$$D = \{(\nu, \eta, \xi) \in (\kappa^\kappa)^3 \mid \nu \text{ codes a winning strategy of } \mathbf{I} \text{ in } \mathrm{EF}^\kappa_\lambda(\mathcal{A}_\eta, \mathcal{A}_\xi)\} \subset (\kappa^\kappa)^3$$

is closed.

Compare with Theorem 1.15.

Proof. Assuming $(\nu_0, \eta_0, \xi_0) \notin C$, we will show that there is an open neighborhood U of (ν_0, η_0, ξ_0) such that $U \subset (\kappa^\kappa)^3 \setminus C$. Denote the strategy that ν_0 codes by σ_0. By the assumption there is a strategy τ of **I** which beats σ_0. Consider the game in which **I** uses τ and **II** uses σ_0. Denote the γ-th move in this game by (X_γ, h_γ) where $X_\gamma \subset A_{\eta_0} \cup A_{\xi_0}$ and $h_\gamma \colon A_{\eta_0} \to A_{\xi_0}$ are the moves of the players. Since player **I** wins this game, there is $\alpha < \lambda$ for which h_α is not a partial isomorphism between $\mathcal{A}_{\eta_0}$ and $\mathcal{A}_{\xi_0}$. Let

$$\varepsilon = \sup(X_\alpha \cup \operatorname{dom} h_\alpha \cup \operatorname{ran} h_\alpha)$$

(recall $\operatorname{dom} \mathcal{A}_\eta = A_\eta = \kappa$ for any η by convention). Let π be the coding function defined in Definition 1.14 on page 478. Let

$$\beta_1 = \pi[\varepsilon^{<\omega}] + 1.$$

The idea is that $\eta_0 \restriction \beta_1$ and $\xi_0 \restriction \beta_1$ decide the models $\mathcal{A}_{\eta_0}$ and $\mathcal{A}_{\xi_0}$ as far as the game has been played. Clearly $\beta_1 < \kappa$.

Up to this point, player **II** has applied her strategy σ_0 precisely to the sequences of the moves made by her opponent, namely to $S = \{(X_\gamma)_{\gamma<\beta} \mid \beta < \alpha\} \subset \operatorname{dom} \sigma_0$. We can translate this set to represent a subset of the domain of ν_0: $S' = k[S]$, where k is as defined before the statement of the present theorem. Let $\beta_2 = (\sup S') + 1$ and let

$$\beta = \max\{\beta_1, \beta_2\}.$$

Thus $\eta_0 \restriction \beta$, $\xi_0 \restriction \beta$ and $\nu_0 \restriction \beta$ decide the moves $(h_\gamma)_{\gamma<\alpha}$ and the winner.

Now

$$\begin{aligned} U &= \{(\nu, \eta, \xi) \mid \nu \restriction \beta = \nu_0 \restriction \beta \wedge \eta \restriction \beta = \eta_0 \restriction \beta \wedge \xi \restriction \beta = \xi_0 \restriction \beta\} \\ &= N_{\nu_0 \restriction \beta} \times N_{\eta_0 \restriction \beta} \times N_{\xi_0 \restriction \beta} \end{aligned}$$

is the desired neighborhood. Indeed, if $(\nu, \eta, \xi) \in U$ and ν codes a strategy σ, then τ beats σ on the structures $\mathcal{A}_\eta$, $\mathcal{A}_\xi$, since the first α moves are exactly as in the corresponding game of the triple (ν_0, η_0, ξ_0).

Let us now turn to D. The proof is similar. Assume that $(\nu_0, \eta_0, \xi_0) \notin D$ and ν_0 codes strategy τ_0 of player **I**. Then there is a strategy of **II** which beats τ_0. Let $\beta < \kappa$ be, as before, an ordinal such that all moves have occurred before β and the relations of the substructures generated by the moves are decided by $\eta_0 \restriction \beta$, $\xi_0 \restriction \beta$ as well as the strategy τ_0. Unlike for player **I**, the win of **II** is determined always only in the end of the game, so β can be $\geqslant \lambda$. This is why we made the assumption $\lambda < \kappa$, by which we can always have $\beta < \kappa$ and so

$$\begin{aligned} U &= \{(\nu, \eta, \xi) \mid \nu \restriction \beta = \nu_0 \restriction \beta \wedge \eta \restriction \beta = \eta_0 \restriction \beta \wedge \xi \restriction \beta = \xi_0 \restriction \beta\} \\ &= N_{\nu_0 \restriction \beta} \times N_{\eta_0 \restriction \beta} \times N_{\xi_0 \restriction \beta} \end{aligned}$$

is an open neighborhood of (ν_0, η_0, ξ_0) in the complement of D. □

Let us list some theorems concerning Ehrenfeucht–Fraïssé games which we will use in the proofs.

Definition 1.9 Let T be a theory and $\mathcal{A}$ a model of T of size κ. The $L_{\infty\kappa}$-*Scott height* of $\mathcal{A}$ is

$$\sup\{\alpha \mid \exists \mathfrak{B} \models T(\mathcal{A} \not\cong \mathfrak{B} \wedge \mathbf{II} \uparrow \mathrm{EF}^{\kappa}_{t_\alpha}(\mathcal{A}, \mathfrak{B}))\},$$

if the supremum exists, and ∞ otherwise, where t_α is as in Example 1.2 and the subsequent fact.

Remark Sometimes the Scott height is defined in terms of quantifier ranks, but this gives an equivalent definition by Theorem 1.11 below.

Definition 1.10 The *quantifier rank* $R(\varphi)$ of a formula $\varphi \in L_{\infty\infty}$ is an ordinal defined by induction on the length of φ as follows. If φ is quantifier free, then $R(\varphi) = 0$. If $\varphi = \exists \bar{x}\psi(\bar{x})$, then $R(\varphi) = R(\psi(\bar{x})) + 1$. If $\varphi = \neg\psi$, then $R(\varphi) = R(\psi)$. If $\varphi = \bigwedge_{\alpha<\lambda} \psi_\alpha$, then $R(\varphi) = \sup\{R(\psi_\alpha \mid \alpha < \lambda)\}$.

Theorem 1.11 *Models $\mathcal{A}$ and $\mathfrak{B}$ satisfy the same $L_{\infty\kappa}$-sentences of quantifier rank $< \alpha$ if and only if $\mathbf{II} \uparrow \mathrm{EF}^{\kappa}_{t_\alpha}(\mathcal{A}, \mathfrak{B})$.* □

The following theorem is a well known generalization of a theorem of Karp [**17**]:

Theorem 1.12 *Models $\mathcal{A}$ and $\mathfrak{B}$ are $L_{\infty\kappa}$-equivalent if and only if $\mathbf{II} \uparrow \mathrm{EF}^{\kappa}_{\omega}(\mathcal{A}, \mathfrak{B})$.* □

Remark 1.13 Models $\mathcal{A}$ and $\mathfrak{B}$ of size κ are $L_{\kappa^+\kappa}$-equivalent if and only if they are $L_{\infty\kappa}$-equivalent. For an extensive and detailed survey on this and related topics, see [**37**].

1.2.3 Coding models

There are various degrees of generality to which the content of this text is applicable. Many of the results generalize to vocabularies with infinitary relations or to uncountable vocabularies, but not all. We find it reasonable though to fix the used vocabulary to make the presentation clearer.

Models can be coded to models with just one binary predicate. Function symbols often make situations unnecessarily complicated from the point of view of this paper.

Thus our approach is, without great loss of generality, to fix our attention to models with finitary relation symbols of all finite arities.

Let us fix L to be the countable relational vocabulary consisting of the relations P_n, $n < \omega$, $L = \{P_n \mid n < \omega\}$, where each P_n is an n-ary relation: the interpretation of P_n is a set consisting of n-tuples. We can assume without loss of generality that the domain of each L-structure of size κ is κ, i.e., $\operatorname{dom} \mathcal{A} = \kappa$. If we restrict our attention to these models, then the set of all L-models has the same cardinality as κ^κ.

We will next present the way we code the structures and the isomorphisms between them into the elements of κ^κ (or equivalently —as will be seen— to 2^κ).

Definition 1.14 Let π be a bijection $\pi\colon \kappa^{<\omega} \to \kappa$. If $\eta \in \kappa^\kappa$, define the structure $\mathcal{A}_\eta$ to have $\operatorname{dom}(\mathcal{A}_\eta) = \kappa$ and if $(a_1, \dots a_n) \in \operatorname{dom}(\mathcal{A}_\eta)^n$, then

$$(a_1, \dots, a_n) \in P_n^{\mathcal{A}_\eta} \iff \eta(\pi(a_1, \dots, a_n)) > 0.$$

In that way the rule $\eta \mapsto \mathcal{A}_\eta$ defines a surjective (onto) function from κ^κ to the set of all L-structures with domain κ. We say that η *codes* $\mathcal{A}_\eta$.

Remark Define the equivalence relation on κ^κ by $\eta \sim \xi \Leftrightarrow \operatorname{sprt}\eta = \operatorname{sprt}\xi$, where sprt means support; see the section *Functions* on page 473. Now we have $\eta \sim \xi \Leftrightarrow \mathcal{A}_\eta = \mathcal{A}_\xi$, i.e., the identity map $\kappa \to \kappa$ is an isomorphism between $\mathcal{A}_\eta$ and $\mathcal{A}_\xi$ when $\eta \sim \xi$ and vice versa. On the other hand $\kappa^\kappa/\sim\, \cong 2^\kappa$, so the coding can be seen also as a bijection between models and the space 2^κ.

The distinction will make little difference, but it is convenient to work with both spaces depending on context. To illustrate the insignificance of the choice between κ^κ and 2^κ, note that $\sim$ is a closed equivalence relation and the identity on 2^κ is bireducible with $\sim$ on κ^κ (see page 474).

1.2.4 Coding partial isomorphisms

Let $\xi, \eta \in \kappa^\kappa$ and let p be a bijection $\kappa \to \kappa \times \kappa$. Let $\nu \in \kappa^\alpha$, $\alpha \leqslant \kappa$. The idea is that, for $\beta < \alpha$, $p_1(\nu(\beta))$ is the image of β under a partial isomorphism and $p_2(\nu(\beta))$ is the inverse image of β. That is, for a $\nu \in \kappa^\alpha$, define a relation $F_\nu \subset \kappa \times \kappa$:

$$(\beta,\gamma) \in F_\nu \iff \big(\beta < \alpha \wedge p_1(\nu(\beta)) = \gamma\big) \vee \big(\gamma < \alpha \wedge p_2(\nu(\gamma)) = \beta\big).$$

If ν happens to be such that F_ν is a partial isomorphism $\mathcal{A}_\xi \to \mathcal{A}_\eta$, then we say that *$\nu$ codes a partial isomorphism between $\mathcal{A}_\xi$ and $\mathcal{A}_\eta$*, this isomorphism being determined by F_ν. If $\alpha = \kappa$ and ν codes a partial isomorphism, then F_ν is an isomorphism and we say that *ν codes an isomorphism.*

Theorem 1.15 *The set*

$$C = \{(\nu,\eta,\xi) \in (\kappa^\kappa)^3 \mid \nu \textit{ codes an isomorphism between } \mathcal{A}_\eta \textit{ and } \mathcal{A}_\xi\}$$

is a closed set.

Proof. Suppose that $(\nu,\eta,\xi) \notin C$, i.e., ν does not code an isomorphism $\mathcal{A}_\eta \cong \mathcal{A}_\xi$. Then (at least) one of the following holds:

(1) F_ν is not a function,
(2) F_ν is not one-to-one,
(3) F_ν does not preserve relations of $\mathcal{A}_\eta$, $\mathcal{A}_\xi$.

(Note that F_ν is always onto if it is a function and $\operatorname{dom}\nu = \kappa$.) If (1), (2) or (3) holds for ν, then respectively (1), (2) or (3) holds for any triple (ν',η',ξ') where $\nu' \in N_{\nu\restriction\gamma}$, $\eta' \in N_{\eta\restriction\gamma}$ and $\xi' \in N_{\xi\restriction\gamma}$, so it is sufficient to check that (1), (2) or (3) holds for $\nu\restriction\gamma$ for some $\gamma < \kappa$.

Let us check the above in the case that (3) holds. The other cases are left to the reader. Suppose that (3) holds. Then there is $(a_0,\dots,a_{n-1}) \in (\operatorname{dom}\mathcal{A}_\eta)^n = \kappa^n$ such that $(a_0,\dots,a_{n-1}) \in P_n$ and $(a_0,\dots,a_{n-1}) \in P_n^{\mathcal{A}_\eta}$ and $(F_\nu(a_0),\dots,F_\nu(a_{n-1})) \notin P_n^{\mathcal{A}_\xi}$. If β is greater than

$$\max(\{\pi(a_0,\dots,a_{n-1}),\pi(F_\nu(a_0),\dots,F_\nu(a_{n-1}))\} \cup \{a_0,\dots a_{n-1},F_\nu(a_0),\dots,F_\nu(a_{n-1})\}),$$

then it is easy to verify that any $(\eta',\xi',\nu') \in N_{\eta\restriction\beta} \times N_{\xi\restriction\beta} \times N_{\nu\restriction\beta}$ satisfies (3) as well. □

Corollary 1.16 *The set* $\{(\eta,\xi) \in (\kappa^\kappa)^2 \mid \mathcal{A}_\eta \cong \mathcal{A}_\xi\}$ *is* Σ_1^1.

Proof. It is the projection of the set C of Theorem 1.15. □

1.3 Generalized Borel sets

Definition 1.17 We have already discussed Δ_1^1-sets which generalize Borel subsets of Polish space in one way. Let us see how else can we generalize usual Borel sets to our setting.

- ([**5, 24**]) The collection of λ-*Borel* subsets of κ^κ is the smallest set which contains the basic open sets of κ^κ and is closed under complementation and under taking intersections of size λ. Since we consider only κ-Borel sets, we write Borel = κ-Borel.
- The collection $\Delta_1^1 = \Sigma_1^1 \cap \Pi_1^1$.
- ([**5, 24**]) The collection of *Borel** subsets of κ^κ. A set A is Borel* if there exists a $\kappa^+\kappa$-tree t in which each increasing sequence of limit order type has a unique supremum and a function

$$h\colon \{\text{branches of } t\} \longrightarrow \{\text{basic open sets of } \kappa^\kappa\}$$

such that $\eta \in A \Leftrightarrow$ player **II** has a winning strategy in the game $G(t,h,\eta)$. The game $G(t,h,\eta)$ is defined as follows. At the first round, player **I** picks a minimal element of the tree; on successive rounds he picks an immediate successor of the last move played by player **II**, and, if there is no last move, he chooses an immediate successor of the supremum of all previous moves. Player **II** always picks an immediate successor of player **I**'s choice. The game ends when the players cannot go up the tree anymore, i.e., have chosen a branch b. Player **II** wins if $\eta \in h(b)$; otherwise **I** wins.

A *dual* of a Borel* set B is the set

$$B^d = \{\xi \mid \mathbf{I} \uparrow G(t,h,\xi)\}$$

where t and h satisfy the equation $B = \{\xi \mid \mathbf{II} \uparrow G(t,h,\xi)\}$. The dual is not unique.

Remark Suppose that t is a $\kappa^+\kappa$ tree and $h\colon \{\text{branches of } t\} \to \text{Borel}^*$ is a labeling function taking values in Borel* sets instead of basic open sets. Then $\{\eta \mid \mathbf{II} \uparrow G(t,h,\eta)\}$ is a Borel* set.

Thus if we change the basic open sets to Borel* sets in the definition of Borel*, we get Borel*.

Remark 1.18 Blackwell [**2**] defined Borel* sets in the case $\kappa = \omega$ and showed that in fact Borel = Borel*. When κ is uncountable it is not the case. But it is easily seen that if t is a $\kappa^+\omega$-tree, then the Borel* set coded by t (with some labeling h) is a Borel set, and vice versa: each Borel set is a Borel* set coded by a $\kappa^+\omega$-tree. We will use this characterization of Borel.

It was first explicitly proved in [**24**] that these are indeed generalizations:

Theorem 1.19 ([**24**], $\kappa^{<\kappa} = \kappa$) Borel $\subset \Delta_1^1 \subset$ Borel* $\subset \Sigma_1^1$.

Proof. (Sketch) If A is Borel*, then it is Σ_1^1; intuitively, because $\eta \in A$ if and only if *there exists* a winning strategy of player **II** in $G(t,h,\eta)$ where (t,h) is a tree that codes A (here one needs the assumption $\kappa^{<\kappa} = \kappa$ to be able to code the strategies into the elements of κ^κ). By Remark 1.18 above, if A is Borel, then there is also such a tree. Since Borel $\subset$ Borel* by Remark 1.18 and Borel is closed under taking complements, Borel sets are Δ_1^1.

The fact that Δ_1^1-sets are Borel* is a more complicated issue; it follows from a separation theorem proved in [**24**]. The separation theorem says that any two disjoint Σ_1^1-sets can be separated by Borel* sets. It is proved in [**24**] for $\kappa = \omega_1$, but the proof generalizes to any κ (with $\kappa^{<\kappa} = \kappa$). □

Additionally we have the following results:

Theorem 1.20

(1) Borel $\subsetneq \Delta^1_1$.
(2) $\Delta^1_1 \subsetneq \Sigma^1_1$.
(3) *If* $V = L$*, then* Borel* $= \Sigma^1_1$.
(4) $\Delta^1_1 \subsetneq$ Borel* *holds if* $V = L$*, and also in every* $\mathbb{P}$*-generic extension starting from a ground model with* $\kappa^{<\kappa} = \kappa$*, where*

$$\mathbb{P} = \{p \mid p \textit{ is a function, } |p| < \kappa, \operatorname{dom} p \subset \kappa \times \kappa^+, \operatorname{ran} p \subset \{0,1\}\}.$$

Proof. (Sketch)

(1) The following universal Borel set is not Borel itself, but is Δ^1_1:

$$B = \{(\eta, \xi) \in 2^\kappa \times 2^\kappa \mid \eta \text{ is in the set coded by } (t_\xi, h_\xi)\},$$

where $\xi \mapsto (t_\xi, h_\xi)$ is a continuous coding of ($\kappa^+\omega$-tree, labeling)-pairs in such a way that for all $\kappa^+\omega$-trees $t \subset \kappa^{<\omega}$ and labelings h there is ξ with $(t_\xi, h_\xi) = (t, h)$. It is not Borel since if it were, then the diagonal's complement

$$D = \{\eta \mid (\eta, \eta) \notin B\}$$

would be a Borel set, yet it is not, since it cannot be coded by any (t_ξ, h_ξ). On the other hand, its complement $C = (2^\kappa)^2 \setminus B$ is Σ^1_1, because $(\eta, \xi) \in C$ if and only if *there exists* a winning strategy of player **I** in the Borel-game $G(t_\xi, h_\xi, \eta)$ and the latter can be coded to a Borel set. It is left to the reader to verify that when $\kappa > \omega$, then the set

$$F = \{(\eta, \xi, \nu) \mid \nu \text{ codes a winning strategy for } \mathbf{I} \text{ in } G(t_\xi, h_\xi, \eta)\}$$

is closed.

The existence of an isomorphism relation which is Δ^1_1 but not Borel follows from Theorems 4.7 and 4.8.

(2) Similarly as above (and similarly as in the case $\kappa = \omega$), take a universal Σ^1_1-set $A \subset 2^\kappa \times 2^\kappa$ with the property that if $B \subset 2^\kappa$ is any Σ^1_1-set, then there is $\eta \in 2^\kappa$ such that $B \times \{\eta\} \subset A$. This set can be constructed as in the case $\kappa = \omega$; see **[16]**. The diagonal $\{\eta \mid (\eta, \eta) \in A\}$ is Σ^1_1 but not Π^1_1.

(3) Suppose $V = L$ and $A \subset 2^\kappa$ is Σ^1_1. There exists a formula $\varphi(x, \xi)$ with parameter $\xi \in 2^\kappa$ which is Σ_1 in the Lévy hierarchy (see **[16]**) and for all $\eta \in 2^\kappa$ we have

$$\eta \in A \iff L \models \varphi(\eta, \xi).$$

Now we have that $\eta \in A$ if and only if the set

$$\big\{\alpha < \kappa \mid \exists\beta\big(\eta\restriction\alpha, \xi\restriction\alpha \in L_\beta,\ L_\beta \models \big(\text{ZF}^- \wedge (\alpha \text{ is a cardinal}) \wedge \varphi(\eta\restriction\alpha, \xi\restriction\alpha)\big)\big)\big\}$$

contains an ω-cub set.

But the ω-cub filter is Borel* so A is also Borel*.

(4) The first part follows from clauses (2) and (3) of this theorem and the second part from clauses (3.20), (3.20) and (3.20) of Theorem 3.20 on page 502; see especially the proof of (7). □

Open Problem Is it consistent that Borel* is a proper subclass of Σ^1_1, or even equals Δ^1_1? Is it consistent that all the inclusions are proper at the same time: $\Delta^1_1 \subsetneq$ Borel* $\subsetneq \Sigma^1_1$?

Theorem 1.21 *For a set $S \subset \kappa^\kappa$, the following are equivalent:*

(1) *S is Σ_1^1.*
(2) *S is a projection of a Borel set.*
(3) *S is a projection of a Σ_1^1-set.*
(4) *S is a continuous image of a closed set.*

Proof. Let us go in order.

(1) ⇒ (2)**:** Closed sets are Borel.
(2) ⇒ (3)**:** The same proof as in the standard case $\kappa = \omega$ gives that Borel sets are Σ_1^1 (see for instance [**16**]).
(3) ⇒ (4)**:** Let $A \subset \kappa^\kappa \times \kappa^\kappa$ be a Σ_1^1-set which is the projection of A, $S = \mathrm{pr}_0 A$. Then let $C \subset \kappa^\kappa \times \kappa^\kappa \times \kappa^\kappa$ be a closed set such that $\mathrm{pr}_1 C = A$. Here $\mathrm{pr}_0 \colon \kappa^\kappa \times \kappa^\kappa \to \kappa^\kappa$ and $\mathrm{pr}_1 \colon \kappa^\kappa \times \kappa^\kappa \times \kappa^\kappa \to \kappa^\kappa \times \kappa^\kappa$ are the obvious projections. Let $f \colon \kappa^\kappa \times \kappa^\kappa \times \kappa^\kappa \to \kappa^\kappa$ be a homeomorphism. Then S is the image of the closed set $f[C]$ under the continuous map $\mathrm{pr}_0 \circ \mathrm{pr}_1 \circ f^{-1}$.
(4) ⇒ (1)**:** The image of a closed set under a continuous map f is the projection of the graph of f restricted to that closed set. It is a basic topological fact that a graph of a continuous partial function with closed domain is closed (provided the range is Hausdorff). □

Theorem 1.22 ([**24**]) *Borel* sets are closed under unions and intersections of size κ.* □

Definition 1.23 A Borel* set B is *determined* if there exists a tree t and a labeling function h such that the corresponding game $G(t, h, \eta)$ is determined for all $\eta \in \kappa^\kappa$ and

$$B = \{\eta \mid \mathbf{II} \text{ has a winning strategy in } G(t, h, \eta)\}.$$

Theorem 1.24 ([**24**]) *Δ_1^1-sets are exactly the determined Borel* sets.* □

2 Borel sets, Δ_1^1-sets and infinitary logic

2.1 The language $L_{\kappa^+\kappa}$ and Borel sets

The interest in the class of Borel sets is explained by the fact that the Borel sets are relatively simple yet at the same time this class includes many interesting definable sets. Below we prove Vaught's theorem (Theorem 2.2), which equates "invariant" Borel sets with those definable in the infinitary language $L_{\kappa^+\kappa}$. Recall that two models $\mathcal{A}$ and $\mathfrak{B}$ of size κ are $L_{\kappa^+\kappa}$-equivalent if and only if they are $L_{\infty\kappa}$-equivalent. Vaught proved his theorem for the case $\kappa = \omega_1$ assuming CH in [**38**], but the proof works for arbitrary κ assuming $\kappa^{<\kappa} = \kappa$.

Definition 2.1 Denote by S_κ the set of all permutations of κ. If $u \in \kappa^{<\kappa}$, denote

$$\overline{u} = \{p \in S_\kappa \mid p^{-1} \restriction \mathrm{dom}\, u = u\}.$$

Note that $\overline{\varnothing} = S_\kappa$ and, if $u \in \kappa^\alpha$ is not injective, then $\overline{u} = \varnothing$.

A permutation $p \colon \kappa \to \kappa$ acts on 2^κ by

$$p\eta = \xi \iff p \colon \mathcal{A}_\eta \to \mathcal{A}_\xi \text{ is an isomorphism.}$$

The map $\eta \mapsto p\eta$ is well defined for every p and it is easy to check that it defines an action of the permutation group S_κ on the space 2^κ. We say that a set $A \subset 2^\kappa$ is *closed under permutations* if it is a union of orbits of this action.

Theorem 2.2 ([**38**], $\kappa^{<\kappa} = \kappa$) *A set $B \subset \kappa^\kappa$ is Borel and closed under permutations if and only if there is a sentence φ in $L_{\kappa^+\kappa}$ such that $B = \{\eta \mid \mathcal{A}_\eta \models \varphi\}$.*

Proof. Let φ be a sentence in $L_{\kappa^+\kappa}$. Then $\{\eta \in 2^\kappa \mid \mathcal{A}_\eta \models \varphi\}$ is closed under permutations, because, if $\eta = p\xi$, then $\mathcal{A}_\eta \cong \mathcal{A}_\xi$ and $\mathcal{A}_\eta \models \varphi \iff \mathcal{A}_\xi \models \varphi$ for every sentence φ. If φ is a formula with parameters $(a_i)_{i<\alpha} \in \kappa^\alpha$, one easily verifies by induction on the complexity of φ that the set

$$\{\eta \in 2^\kappa \mid \mathcal{A}_\eta \models \varphi((a_i)_{i<\alpha})\}$$

is Borel. This of course implies that for every sentence φ the set $\{\eta \mid \mathcal{A}_\eta \models \varphi\}$ is Borel.

The converse is less trivial. Note that the set of permutations $S_\kappa \subset \kappa^\kappa$ is Borel, since

$$S_\kappa = \bigcap_{\beta<\kappa} \bigcup_{\alpha<\kappa} \underbrace{\{\eta \mid \eta(\alpha) = \beta\}}_{\text{open}} \cap \bigcap_{\alpha<\beta<\kappa} \underbrace{\{\eta \mid \eta(\alpha) \neq \eta(\beta)\}}_{\text{open}}. \tag{2.1}$$

For a set $A \subset \kappa^\kappa$ and $u \in \kappa^{<\kappa}$, define

$$A^{*u} = \big\{\eta \in 2^\kappa \mid \{p \in \overline{u} \mid p\eta \in A\} \text{ is co-meager in } \overline{u}\big\}.$$

From now on in this section we will write "$\{p \in \overline{u} \mid p\eta \in A\}$ is co-meager", when we really mean "co-meager in $\overline{u}$".

Let us show that the set

$$Z = \{A \subset 2^\kappa \mid A \text{ is Borel and } A^{*u} \text{ is } L_{\kappa^+\kappa}\text{-definable for all } u \in \kappa^{<\kappa}\}$$

contains all the basic open sets, is closed under intersections of size κ and under complementation in the three steps (a), (b) and (c) below. This implies that Z is the collection of all Borel sets. We will additionally keep track of the fact that the formula which defines A^{*u} depends only on A and $\operatorname{dom} u$, i.e., for each $\beta < \kappa$ and Borel set A there exists $\varphi = \varphi^A_\beta$ such that for all $u \in \kappa^\beta$ we have $A^{*u} = \{\eta \mid \mathcal{A}_\eta \models \varphi((u_i)_{i<\beta})\}$. Setting $u = \varnothing$, we have the intended result, because $A^{*\varnothing} = A$ for all A which are closed under permutations and φ is a sentence (with no parameters).

If A is fixed we denote $\varphi^A_\beta = \varphi_\beta$.

(a) Assume $q \in 2^{<\kappa}$ and let N_q be the corresponding basic open set. Let us show that $N_q \in Z$. Let $u \in \kappa^\beta$ be arbitrary. We have to find $\varphi^{N_q}_\beta$. Let θ be a quantifier free formula with α parameters such that

$$N_q = \{\eta \in 2^\kappa \mid \mathcal{A}_\eta \models \theta((\gamma)_{\gamma<\alpha})\}.$$

Here $(\gamma)_{\gamma<\alpha}$ denotes both an initial segment of κ as well as an α-tuple of the structure. Suppose $\alpha \leqslant \beta$. We have $p \in \overline{u} \Rightarrow u \subset p^{-1}$, so

$$\begin{aligned}
\eta \in N_q^{*u} &\iff \{p \in \overline{u} \mid p\eta \in N_q\} \text{ is co-meager} \\
&\iff \{p \in \overline{u} \mid \mathcal{A}_{p\eta} \models \theta((\gamma)_{\gamma<\alpha})\} \text{ is co-meager} \\
&\iff \{p \in \overline{u} \mid \mathcal{A}_\eta \models \theta((p^{-1}(\gamma))_{\gamma<\alpha})\} \text{ is co-meager} \\
&\iff \{p \in \overline{u} \mid \underbrace{\mathcal{A}_\eta \models \theta((u_\gamma)_{\gamma<\alpha})}_{\text{independent of } p}\} \text{ is co-meager} \\
&\iff \mathcal{A}_\eta \models \theta((u_\gamma)_{\gamma<\alpha}).
\end{aligned}$$

Then $\varphi_\beta = \theta$.

Assume then that $\alpha > \beta$. By the above, we still have

$$\eta \in N_q^{*u} \iff E = \big\{p \in \overline{u} \mid \mathcal{A}_\eta \models \theta\big((p^{-1}(\gamma))_{\gamma<\alpha}\big)\big\} \text{ is co-meager.}$$

Assume that $w = (w_\gamma)_{\gamma<\alpha} \in \kappa^\alpha$ is an arbitrary sequence with no repetition and such that $u \subset w$. Since $\overline{w}$ is an open subset of $\overline{u}$ and E is co-meager, there is $p \in \overline{w} \cap E$. Because $p \in E$, we have $\mathcal{A}_\eta \models \theta\big((p^{-1}(\gamma))_{\gamma<\alpha}\big)$. On the other hand, $p \in \overline{w}$, so we have $w \subset p^{-1}$, i.e., $w_\gamma = w(\gamma) = p^{-1}(\gamma)$ for $\gamma < \alpha$. Hence

$$\mathcal{A}_\eta \models \theta((w_\gamma)_{\gamma<\alpha}). \tag{2.2}$$

On the other hand, if for every injective $w \in \kappa^\alpha$, $w \supset u$, we have (2.2), then in fact $E = \overline{u}$ and is trivially co-meager. Therefore we have an equivalence:

$$\eta \in N_q^{*u} \iff (\forall w \supset u)(w \in \kappa^\alpha \wedge w \text{ injective} \Rightarrow \mathcal{A}_\eta \models \theta((w_\gamma)_{\gamma<\alpha})).$$

But the latter can be expressed in the language $L_{\kappa^+\kappa}$ by the formula $\varphi_\beta((w_i)_{i<\beta})$:

$$\bigwedge_{i<j<\beta}(w_i \neq w_j) \wedge \Big(\underset{\beta\leqslant i<\alpha}{\forall} w_i\Big)\Big(\bigwedge_{i<j<\alpha}(w_i \neq w_j) \to \theta((w_i)_{i<\alpha})\Big).$$

Here θ was defined to be a formula defining N_q with parameters. It is clear thus that θ is independent of u. Furthermore the formulas constructed above from θ depend only on $\beta = \operatorname{dom} u$ and on θ. Hence the formulas defining N_q^{*u} and N_q^{*v} for $\operatorname{dom} u = \operatorname{dom} v$ are the same modulo parameters.

(b) For each $i < \kappa$ let $A_i \in Z$. We want to show that $\bigcap_{i<\kappa} A_i \in Z$. Assume that $u \in \kappa^{<\kappa}$ is arbitrary. It suffices to show that

$$\bigcap_{i<\kappa}(A_i^{*u}) = \Big(\bigcap_{i<\kappa} A_i\Big)^{*u},$$

because then $\varphi_\beta^{\bigcap_i A_i}$ is just the κ-conjunction of the formulas $\varphi_\beta^{A_i}$, which exist by the induction hypothesis. Clearly the resulting formula depends again only on $\operatorname{dom} u$ if the previous did. Note that a κ-intersection of co-meager sets is co-meager. Now

$$\begin{aligned}\eta \in \bigcap_{i<\kappa}(A_i^{*u}) &\iff (\forall i < \kappa)(\{p \in \overline{u} \mid p\eta \in A_i\} \text{ is co-meager})\\ &\iff (\forall i < \kappa)(\forall i < \kappa)(\{p \in \overline{u} \mid p\eta \in A_i\} \text{ is co-meager})\\ &\iff \bigcap_{i<\kappa}\{p \in \overline{u} \mid p\eta \in A_i\} \text{ is co-meager}\\ &\iff \{p \in \overline{u} \mid p\eta \in \bigcap_{i<\kappa} A_i\} \text{ is co-meager}\\ &\iff \eta \in \Big(\bigcap_{i<\kappa} A_i\Big)^{*u}.\end{aligned}$$

(c) Assume that $A \in Z$, i.e., that A^{*u} is definable for any u. Let $\varphi_{\operatorname{dom} u}$ be the formula which defines A^{*u}. Let now $u \in \kappa^{<\kappa}$ be arbitrary and let us show that $(A^c)^{*u}$ is definable. We will show that

$$(A^c)^{*u} = \bigcap_{v\supset u}(A^{*v})^c,$$

i.e., for all η,

$$\eta \in (A^c)^{*u} \iff \forall v \supset u(\eta \notin A^{*v}). \tag{2.3}$$

Granted this, one can write the formula "$\forall v \supset u \neg\varphi_{\operatorname{dom} u}((v_i)_{i<\operatorname{dom} v})$", which is not of course the real $\varphi_\beta^{A^c}$ which we will write in the end of the proof.

To prove (2.3), we have to show first that, for all $\eta\in\kappa^\kappa$, the set

$$B=\{p\in\overline{u}\mid p\eta\in A\}$$

has the property of Baire (P.B.); see Section 3.3.

The set of all permutations $S_\kappa\subset\kappa^\kappa$ is Borel by (2.1) on page 483. The set $\overline{u}$ is an intersection of S_κ with an open set. Again the set $\{p\in\overline{u}\mid p\eta\in A\}$ is the intersection of $\overline{u}$ and the inverse image of A under the continuous map $(p\mapsto p\eta)$, so it is Borel and hence has the property of Baire.

We can now turn into proving the equivalence (2.3). First "$\Leftarrow$":

$$\begin{aligned}\eta\notin(A^c)^{*u}&\Rightarrow B=\{p\in\overline{u}\mid p\eta\in A\}\text{ is not meager in }\overline{u}\\&\Rightarrow\text{By P.B. of }B\text{ there is a non-empty open }U\text{ such that }U\setminus B\text{ is meager}\\&\Rightarrow\text{There is non-empty }\overline{v}\subset\overline{u}\text{ such that }\overline{v}\setminus B\text{ is meager}\\&\Rightarrow\text{There exists }\overline{v}\subset\overline{u}\text{ such that }\{p\in\overline{v}\mid p\eta\in A\}=\overline{v}\cap B\text{ is co-meager}\\&\Rightarrow\exists v\supset u(\eta\in A^{*v}).\end{aligned}$$

And then the other direction "$\Rightarrow$":

$$\begin{aligned}\eta\in(A^c)^{*u}&\Rightarrow\{p\in\overline{u}\mid p\eta\in A\}\text{ is meager}\\&\Rightarrow\text{For all }\overline{v}\subset\overline{u}\text{ the set }\{p\in\overline{v}\mid p\eta\in A\}\text{ is meager}\\&\Rightarrow\forall\overline{v}\subset\overline{u}(\eta\notin A^{*v}).\end{aligned}$$

Let us now write the formula $\psi=\varphi_\beta^{A^c}$ such that

$$\forall\overline{v}\subset\overline{u}(\eta\notin A^{*v})\iff\mathcal{A}_\eta\models\psi((u_i)_{i<\beta}),$$

where $\beta=\operatorname{dom}u$: let $\psi((u_i)_{i<\beta})$ be

$$\bigwedge_{\beta\leqslant\gamma<\kappa}\ \forall_{i<\gamma}x_i\left(\Big[\bigwedge_{j<\beta}(x_j=u_j)\wedge\bigwedge_{i<j<\gamma}(x_i\neq x_j)\Big]\to\neg\varphi_\gamma((x_i)_{i<\gamma})\right).$$

One can easily see that this is equivalent to $\forall v\supset u\big(\neg\varphi_{\operatorname{dom}v}((v_i)_{i<\operatorname{dom}v})\big)$ and that ψ depends only on $\operatorname{dom}u$ modulo parameters. □

Remark 2.3 If $\kappa^{<\kappa}>\kappa$, then the direction from right to left of the above theorem does not in general hold. Let $\langle\kappa,\lessdot,A\rangle$ be a model with domain κ, $A\subset\kappa$ and $\lessdot$ a well ordering of κ of order type κ. Shelah and Väänänen have shown [**32**, Corollary 17] that if $\kappa=\lambda^+$, $\kappa^{<\kappa}>\kappa$, $\lambda^{<\lambda}=\lambda$, and a forcing axiom holds (and $\omega_1^L=\omega_1$ if $\lambda=\omega$), then there is a sentence of $L_{\kappa\kappa}$ defining the set

$$\mathrm{STAT}=\{\langle\kappa,\lessdot,A\rangle\mid A\text{ is stationary}\}.$$

If now STAT is Borel, then so would be the set CUB defined in Section 3.3, but by Theorem 3.20, page 502, this set cannot be Borel since Borel sets have the property of Baire by Theorem 3.16 on page 501.

Open Problem Does the direction left to right of Theorem 2.2 hold without the assumption $\kappa^{<\kappa}=\kappa$?

2.2 The language $M_{\kappa^+\kappa}$ and Δ_1^1-sets

In this section we will present a theorem similar to Theorem 2.2. It is also a generalization of the known result which follows from [**24**] and [**35**]:

Theorem 2.4 ([**24, 35**]) *Let $\mathcal{A}$ be a model of size ω_1. Then the isomorphism type $I = \{\eta \mid \mathcal{A}_\eta \cong \mathcal{A}\}$ is Δ_1^1 if and only if there is a sentence φ in $M_{\kappa^+\kappa}$ such that $I = \{\eta \mid \mathcal{A}_\eta \models \varphi\}$ and $2^\kappa \setminus I = \{\eta \mid \mathcal{A}_\eta \models \sim\varphi\}$, where $\sim\theta$ is the dual of θ.*

The idea of the proof of the following theorem is due to Sam Coskey and Philipp Schlicht:

Theorem 2.5 ($\kappa^{<\kappa} = \kappa$) *A set $D \subset 2^\kappa$ is Δ_1^1 and closed under permutations if and only if there is a sentence φ in $M_{\kappa^+\kappa}$ such that $D = \{\eta \mid \mathcal{A}_\eta \models \varphi\}$ and $\kappa^\kappa \setminus D = \{\eta \mid \mathcal{A}_\eta \models \sim\varphi\}$, where $\sim\theta$ is the dual of θ.*

We have to define these concepts before the proof.

Definition 2.6 (Karttunen [**18**]) Let λ and κ be cardinals. The language $M_{\lambda\kappa}$ is defined to be the set of pairs $(t, \ll)$ of a tree t and a labeling function $\ll$. The tree t is a $\lambda\kappa$-tree where the limits of increasing sequences of t exist and are unique. The labeling $\ll$ is a function satisfying the following conditions:

(1) $\ll : t \to a \cup \bar{a} \cup \{\bigwedge, \bigvee\} \cup \{\exists x_i \mid i < \kappa\} \cup \{\forall x_i \mid i < \kappa\}$ where a is the set of atomic formulas and $\bar{a}$ is the set of negated atomic formulas.
(2) If $x \in t$ has no successors, then $\ll(t) \in a \cup \bar{a}$.
(3) If $x \in t$ has exactly one immediate successor then $\ll(t)$ is either $\exists x_i$ or $\forall x_i$ for some $i < \kappa$.
(4) Otherwise $\ll(t) \in \{\bigvee, \bigwedge\}$.
(5) If $x < y$, $\ll(x) \in \{\exists x_i, \forall x_i\}$ and $\ll(y) \in \{\exists x_j, \forall x_j\}$, then $i \neq j$.

Definition 2.7 Truth for $M_{\lambda\kappa}$ is defined in terms of a semantic game. Let $(t, \ll)$ be the pair which corresponds to a particular sentence φ and let $\mathcal{A}$ be a model. The semantic game $S(\varphi, \mathcal{A}) = S(t, \ll, \mathcal{A})$ for $M_{\lambda\kappa}$ is played by players **I** and **II** as follows. At the first move the players are at the root and later in the game at some other element of t. Let us suppose that they are at the element $x \in t$. If $\ll(x) = \bigvee$, then player **II** chooses a successor of x and the players move to that chosen element. If $\ll(x) = \bigwedge$, then player **I** chooses a successor of x and the players move to that chosen element. If $\ll(x) = \forall x_i$ then player **I** picks an element $a_i \in \mathcal{A}$ and if $\ll(x) = \exists x_i$ then player **II** picks an element a_i and they move to the immediate successor of x. If they come to a limit, they move to the unique supremum. If x is a maximal element of t, then they plug the elements a_i in place of the corresponding free variables in the atomic formula $\ll(x)$. Player **II** wins if this atomic formula is true in $\mathcal{A}$ with these interpretations. Otherwise player **I** wins.

We define $\mathcal{A} \models \varphi$ if and only if **II** has a winning strategy in the semantic game.

Given a sentence φ, the *dual* sentence $\sim\varphi$ is defined by modifying the labeling function as follows. The atomic formulas are replaced by their negations, the symbols $\bigvee$ and $\bigwedge$ switch places and the quantifiers $\forall$ and $\exists$ switch places. A sentence $\varphi \in M_{\lambda\kappa}$ is *determined* if for all models $\mathcal{A}$ either $\mathcal{A} \models \varphi$ or $\mathcal{A} \models \sim\varphi$.

Now the statement of Theorem 2.5 makes sense. Theorem 2.5 concerns a sentence φ whose dual defines the complement of the set defined by φ among the models of size κ, so it is determined in that model class. Before the proof let us recall a separation theorem for $M_{\kappa^+\kappa}$, Theorem 3.9 from [**34**]:

Theorem 2.8 *Assume $\kappa^{<\kappa} = \lambda$ and let $\exists R\varphi$ and $\exists S\psi$ be two Σ^1_1 sentences where φ and ψ are in $M_{\kappa^+\kappa}$ and $\exists R$ and $\exists S$ are second order quantifiers. If $\exists R\varphi \wedge \exists S\psi$ does not have a model, then there is a sentence $\theta \in M_{\lambda^+\lambda}$ such that, for all models $\mathcal{A}$,*

$$\mathcal{A} \models \exists R\varphi \Rightarrow \mathcal{A} \models \theta \text{ and } \mathcal{A} \models \exists S\psi \Rightarrow \mathcal{A} \models \sim\theta. \qquad \square$$

Definition 2.9 For a tree t, let σt be the tree of downward closed linear subsets of t ordered by inclusion.

Proof of Theorem 2.5. Let us first show that if φ is an arbitrary sentence of $M_{\kappa^+\kappa}$, then $D_\varphi = \{\eta \mid \mathcal{A}_\eta \models \varphi\}$ is Σ^1_1. The proof has the same idea as the proof of Theorem 1.19 that Borel* $\subset \Sigma^1_1$. Note that this implies that if $\sim\varphi$ defines the complement of D_φ in 2^κ, then D_φ is Δ^1_1.

A strategy in the semantic game $S(\varphi, \mathcal{A}_\eta) = S(t, \ll, \mathcal{A}_\eta)$ is a function

$$\upsilon\colon \sigma t \times (\operatorname{dom} \mathcal{A}_\eta)^{<\kappa} \longrightarrow t \cup (t \times \operatorname{dom} \mathcal{A}_\eta).$$

This is because the previous moves always form an initial segment of a branch of the tree together with the sequence of constants picked by the players from $\operatorname{dom} \mathcal{A}_\eta$ at the quantifier moves, and a move consists either of going to some node of the tree or going to a node of the tree together with choosing an element from $\operatorname{dom} \mathcal{A}_\eta$. By the convention that $\operatorname{dom} \mathcal{A}_\eta = \kappa$, a strategy becomes a function

$$\upsilon\colon \sigma t \times \kappa^{<\kappa} \longrightarrow t \cup (t \times \kappa).$$

Because t is a $\kappa^+\kappa$-tree, there are fewer than κ moves in a play (there are no branches of length κ and the players go up the tree on each move). Let

$$f\colon \sigma t \times \kappa^{<\kappa} \longrightarrow \kappa$$

be any bijection and let

$$g\colon t \cup (t \times \kappa) \longrightarrow \kappa$$

be another bijection. Let F be the bijection

$$F\colon (t \cup (t \times \kappa))^{\sigma t \times \kappa^{<\kappa}} \longrightarrow \kappa^\kappa$$

defined by $F(\upsilon) = g \circ \upsilon \circ f^{-1}$. Let

$$C = \{(\eta, \xi) \mid F^{-1}(\xi) \text{ is a winning strategy of } \mathbf{II} \text{ in } S(t, \ll, \mathcal{A}_\eta)\}.$$

Clearly D_φ is the projection of C. Let us show that C is closed. Consider an element (η, ξ) in the complement of C. We shall show that there is an open neighborhood of (η, ξ) outside C. Denote $\upsilon = F^{-1}(\xi)$. Since υ is not a winning strategy, there is a strategy τ of $\mathbf{I}$ that beats υ. There are $\alpha + 1 < \kappa$ moves in the play $\tau * \upsilon$ (by definition all branches have successor order type). Assume that $b = (x_i)_{i \leqslant \alpha}$ is the chosen branch of the tree and $(c_i)_{i<\alpha}$ the constants picked by the players. Let $\beta < \kappa$ be an ordinal with the properties $\{f((x_i)_{i<\gamma}, (c_i)_{i<\gamma}) \mid \gamma \leqslant \alpha + 1\} \subset \beta$ and

$$\eta' \in N_{\eta\restriction\beta} \to \mathcal{A}_{\eta'} \not\models \ll(x_\alpha)((c_i)_{i<\alpha}). \tag{2.4}$$

Such a β exists, since $|\{f((x_i)_{i<\gamma}, (c_i)_{i<\gamma}) \mid \gamma \leqslant \alpha + 1\}| < \kappa$ and $\ll(x_\alpha)$ is a (possibly negated) atomic formula which is not true in $\mathcal{A}_\eta$, because $\mathbf{II}$ lost the game $\tau * \upsilon$ and because already a fragment of size $< \kappa$ of $\mathcal{A}_\eta$ decides this. Now if $(\eta', \xi') \in N_{\eta\restriction\beta} \times N_{\xi\restriction\beta}$ and $\upsilon' = F^{-1}(\xi')$, then $\upsilon * \tau$ is the same play as $\tau * \upsilon'$. So $\mathcal{A}_{\eta'} \not\models \ll(x_\alpha)((c_i)_{i<\alpha})$ by (2.4), (η', ξ') is not in C, and $N_{\eta\restriction\beta} \times N_{\xi\restriction\beta}$ is the intended open neighborhood of (η, ξ) outside C. This completes the "if"-part of the proof.

Now for a given $A \in \Delta^1_1$ which is closed under permutations we want to find a sentence $\varphi \in M_{\kappa^+\kappa}$ such that $A = \{\eta \mid \mathcal{A}_\eta \models \varphi\}$ and $2^\kappa \setminus A = \{\eta \mid \mathcal{A}_\eta \models \sim\varphi\}$. By our assumption $\kappa^{<\kappa} = \kappa$ and Theorems 1.24 and 2.8, it is enough to show that for a given Borel* set B which is closed under permutations, there is a sentence $\exists R\psi$ which is Σ^1_1 over $M_{\kappa^+\kappa}$ (as in the formulation of Theorem 2.8), such that $B = \{\eta \mid \mathcal{A}_\eta \models \exists R\psi\}$.

The sentence "R is a well ordering of the universe of order type κ", is definable by the formula $\theta = \theta(R)$ of $L_{\kappa^+\kappa} \subset M_{\kappa^+\kappa}$:

$$\begin{aligned}
&\text{“}R\text{ is a linear ordering on the universe”}\\
&\wedge \quad \Big(\mathop{\forall}_{i<\omega} x_i\Big)\Big(\bigvee_{i<\omega} \neg R(x_{i+1}, x_i)\Big)\\
&\wedge \quad \forall x \bigvee_{\alpha<\kappa} \mathop{\exists}_{i<\alpha} y_i \left[\big(\forall y (R(y,x) \to \bigvee_{i<\alpha} y_i = y)\big)\right]. \qquad (2.5)
\end{aligned}$$

(We assume $\kappa > \omega$, so the infinite quantification is allowed. The second row says that there are no descending sequences of length ω and the third row says that the initial segments are of size less than κ. This ensures that $\theta(R)$ says that R is a well ordering of order type κ.)

Let t and h be the tree and the labeling function corresponding to B. Define the tree $t^\star$ as follows.

(1) Assume that b is a branch of t with $h(b) = N_{\xi\restriction\alpha}$ for some $\xi \in \kappa^\kappa$ and $\alpha < \kappa$. Then attach a sequence of order type α^* on top of b where

$$\alpha^* = \bigcup_{s \in \pi^{-1}[\alpha]} \operatorname{ran} s,$$

where π is the bijection $\kappa^{<\omega} \to \kappa$ used in the coding; see Definition 1.14 on page 478.

(2) Do this to each branch of t and add a root r to the resulting tree.

After doing this, the resulting tree is $t^\star$. Clearly it is a $\kappa^+\kappa$-tree, because t is. Next, define the labeling function $\ll$. If $x \in t$ then either $\ll(x) = \bigwedge$ or $\ll(x) = \bigvee$ depending on whether it is player **I**'s move or player **II**'s move: formally let $n < \omega$ be such that $\mathrm{OTP}(\{y \in t^\star \mid y \leqslant x\}) = \alpha + n$ where α is a limit ordinal or 0; then if n is odd, put $\ll(x) = \bigwedge$ and otherwise $\ll(x) = \bigvee$. If $x = r$ is the root, then $\ll(x) = \bigwedge$. Otherwise, if x is not maximal, define

$$\beta = \mathrm{OTP}\{y \in t^\star \setminus (t \cup \{r\}) \mid y \leqslant x\}$$

and set $\ll(x) = \exists x_\beta$.

Next we will define the labeling of the maximal nodes of $t^\star$. By definition these should be atomic formulas or negated atomic formulas, but it is clear that they can be replaced without loss of generality by any formula of $M_{\kappa^+\kappa}$; this fact will make the proof simpler. Assume that x is maximal in $t^\star$. Then $\ll(x)$ will depend only on $h(b)$ where b is the unique branch of t leading to x. Let us define $\ll(x)$ to be the formula of the form $\theta \wedge \Theta_b((x_i)_{i<\alpha^*})$, where θ is defined above and Θ_b is defined below. The idea is that

$$\mathcal{A}_\eta \models \Theta_b((a_\gamma)_{\gamma<\alpha^*})\} \iff \eta \in h(b) \text{ and } \forall\gamma < \alpha^*(a_\gamma = \gamma).$$

Let us define such a Θ_b. Suppose that ξ and α are such that $h(b) = N_{\xi\restriction\alpha}$. Define for $s \in \pi^{-1}[\alpha]$ the formula A^s_b as follows:

$$A^s_b = \begin{cases} P_{\operatorname{dom} s}, & \text{if } \mathcal{A}_\xi \models P_{\operatorname{dom} s}((s(i))_{i\in\operatorname{dom} s}),\\ \neg P_{\operatorname{dom} s}, & \text{if } \mathcal{A}_\xi \not\models P_{\operatorname{dom} s}((s(i))_{i\in\operatorname{dom} s}). \end{cases}$$

Then define

$$\begin{aligned} \psi_0((x_i)_{i<\alpha^*}) &= \bigwedge_{i<\alpha^*} \Big[\forall y(R(y,x_i) \Leftrightarrow \bigvee_{j<i}(y = x_j))\Big];\\ \psi_1((x_i)_{i<\alpha^*}) &= \bigwedge_{s\in\pi^{-1}[\alpha]} A^s_b((x_{s(i)})_{i\in\operatorname{dom} s});\\ \Theta_b &= \psi_0 \wedge \psi_1. \end{aligned}$$

The disjunction over the empty set is considered false.

Claim 1 Suppose that, for all η, R is the standard order relation on κ. Then

$$(\mathcal{A}_\eta, R) \models \Theta_b((a_\gamma)_{\gamma<\alpha^*}) \iff \eta \in h(b) \wedge \forall\gamma < \alpha^*(\alpha_\gamma = \gamma).$$

Proof of Claim 1. Suppose $\mathcal{A}_\eta \models \Theta((a_\gamma)_{\gamma<\alpha^*})$. Then by $\mathcal{A}_\eta \models \psi_0((a_\gamma)_{\gamma<\alpha^*})$ we have that $(a_\gamma)_{\gamma<\alpha^*}$ is an initial segment of $\operatorname{dom}\mathcal{A}_\eta$ with respect to R. But $(\operatorname{dom}\mathcal{A}_\eta, R) = (\kappa, <)$, so $\forall\gamma < \alpha^*(\alpha_\gamma = \gamma)$. Assume that $\beta < \alpha$ and $\eta(\beta) = 1$ and denote $s = \pi^{-1}(\beta)$. Then $\mathcal{A}_\eta \models P_{\operatorname{dom} s}((s(i))_{i\in\operatorname{dom} s})$. Since Θ is true in $\mathcal{A}_\eta$ as well, we must have $A^s_b = P_{\operatorname{dom} s}$, which by definition means that $\mathcal{A}_\xi \models P_{\operatorname{dom} s}((s(i))_{i\in\operatorname{dom} s})$ and hence $\xi(\beta) = \xi(\pi(s)) = 1$. In the same way one shows that if $\eta(\beta) = 0$, then $\xi(\beta) = 0$ for all $\beta < \alpha$. Hence $\eta\restriction\alpha = \xi\restriction\alpha$.

Assume then that $a_\gamma = \gamma$ for all $\gamma < \alpha^*$ and that $\eta \in N_{\xi\restriction\alpha}$. Then $\mathcal{A}_\eta$ trivially satisfies ψ_0. Suppose that $s \in \pi^{-1}[\alpha]$ is such that $\mathcal{A}_\xi \models P_{\operatorname{dom} s}((s(i))_{i\in\operatorname{dom} s})$. Then $\xi(\pi(s)) = 1$ and since $\pi(s) < \alpha$, also $\eta(\pi(s)) = 1$, so $\mathcal{A}_\eta \models P_{\operatorname{dom} s}((s(i))_{i\in\operatorname{dom} s})$. Similarly one shows that if

$$\mathcal{A}_\xi \not\models P_{\operatorname{dom} s}((s(i))_{i\in\operatorname{dom} s}),$$

then $\mathcal{A}_\eta \not\models P_{\operatorname{dom} s}((s(i))_{i\in\operatorname{dom} s})$. This shows that $\mathcal{A}_\eta \models A^s_b((s(i))_{i\in\operatorname{dom} s})$ for all s. Hence $\mathcal{A}_\eta$ satisfies ψ_1, so we have $\mathcal{A}_\eta \models \Theta$. $\square_{\text{Claim 1}}$

Claim 2 t, h, $t^\star$ and $\ll$ are such that, for all $\eta \in \kappa^\kappa$,

$$\mathbf{II} \uparrow G(t,h,\eta) \iff \exists R \subset (\operatorname{dom}\mathcal{A}_\eta)^2 \;\; \mathbf{II} \uparrow S(t^\star, \ll, \mathcal{A}_\eta).$$

Proof of Claim 2. Suppose σ is a winning strategy of **II** in $G(t,h,\eta)$. Let R be the well ordering of $\operatorname{dom}\mathcal{A}_\eta$ such that $(\operatorname{dom}\mathcal{A}_\eta, R) = (\kappa, <)$. Consider the game $S(t^\star, \ll, \mathcal{A}_\eta)$. On the first move the players are at the root and player **I** chooses where to go next. They go to a minimal element of t. From here on **II** uses σ as long as they are in t. Let us see what happens if they got to a maximal element of t, i.e., they picked a branch b from t. Since σ is a winning strategy of **II** in $G(t,h,\eta)$, we have $\eta \in h(b)$ and $h(b) = N_{\xi\restriction\alpha}$ for some ξ and α. For the next α moves, the players climb up the tower defined in item (1) of the definition of $t^\star$. All labels are of the form $\exists x_\beta$, so player **II** has to pick constants from $\mathcal{A}_\eta$. She picks them as follows: for the variable x_β she picks $\beta \in \kappa = \operatorname{dom}\mathcal{A}_\eta$. She wins now if $\mathcal{A}_\eta \models \Theta((\beta)_{\beta<\alpha^*})$ and $\mathcal{A}_\eta \models \theta$. But $\eta \in h(b)$, so by Claim 1 the former holds and the latter holds because we chose R to be a well ordering of order type κ.

Let us assume that there is no winning strategy of **II** in $G(t,h,\eta)$. Let R be an arbitrary relation on $\operatorname{dom}\mathcal{A}_\eta$. Here we shall finally use the fact that B is closed under permutations. Suppose R is not a well ordering of the universe of order type κ. Then after the players reached the final node of $t^\star$, player **I** chooses to go to θ and player **II** loses. So we can assume that R is a well ordering of the universe of order type κ. Let $p\colon \kappa\to\kappa$ be a bijection such that $p(\alpha)$ is the α-th element of κ with respect to R. Now p is a permutation and $\{\eta \mid \mathcal{A}_{p\eta}\in B\} = B$ since B is closed under permutations. So by our assumption that $\eta\notin B$ (i.e., $\mathbf{II} \not\uparrow G(t,h,\eta)$), we also have $p\eta\notin B$, i.e., player **II** has no winning strategy in $G(t,h,p\eta)$ either.

Suppose σ is any strategy of **II** in $S(t^\star,\ll,\mathcal{A}_\eta)$. Player **I** imagines that σ is a strategy in $G(t,h,p\eta)$ and picks a strategy τ that beats it. In the game $S(t^\star,\ll,\mathcal{A}_\eta)$, as long as the players are still in t, player **I** uses τ that would beat σ if they were playing $G(t,h,p\eta)$ instead of $S(t^\star,\ll,\eta)$. Suppose they picked a branch b of t. Now $p\eta\notin h(b)$. If **II** wants to satisfy ψ_0 of the definition of Θ_b, she is forced to pick the constants $(a_i)_{i<\alpha^*}$ such that a_i is the i-th element of $\operatorname{dom}\mathcal{A}_\eta$ with respect to R. Suppose that $\mathcal{A}_\eta \models \psi_1((a_i)_{i<\alpha^*})$ (recall $\Theta_b=\psi_0\wedge\psi_1$). But then $\mathcal{A}_{p\eta}\models\psi_1((\gamma)_{\gamma<\alpha^*})$ and also $\mathcal{A}_{p\eta}\models\psi_0((\gamma)_{\gamma<\alpha^*})$, so by Claim 1 we should have $p\eta\in h(b)$, which is a contradiction. $\square_{\text{Claim 2}}$ $\square$

3 Generalizing classical descriptive set theory

3.1 Simple generalizations

3.1.1 The identity relation

Denote by id the equivalence relation $\{(\eta,\xi)\in(2^\kappa)^2 \mid \eta=\xi\}$. If we want to emphasize the set on which the identity relation lies, we denote it by id_X if the set is X. With respect to our choice of topology, the natural generalization of the equivalence relation

$$E_0=\{(\eta,\xi)\in 2^\omega\times 2^\omega \mid \exists n<\omega\forall m>n(\eta(m)=\xi(m))\}$$

is equivalence modulo sets of size $<\kappa$:

$$E_0^{<\kappa}=\{(\eta,\xi)\in 2^\kappa\times 2^\kappa \mid \exists\alpha<\kappa\forall\beta>\alpha(\eta(\beta)=\xi(\beta))\},$$

although the equivalences modulo sets of size $<\lambda$ for $\lambda<\kappa$ can also be studied:

$$E_0^{<\lambda}=\{(\eta,\xi)\in 2^\kappa\times 2^\kappa \mid \exists A\subset\kappa[|A|<\lambda\wedge\forall\beta\notin A(\eta(\beta)=\xi(\beta))]\},$$

but for $\lambda<\kappa$ these turn out to be bireducible with id (see below). Similarly one can define $E_0^{<\lambda}$ on κ^κ instead of 2^κ.

It makes no difference whether we define these relations on 2^κ or κ^κ since they become bireducible to each other:

Theorem 3.1 *Let $\lambda\leqslant\kappa$ be a cardinal and let $E_0^{<\lambda}(P)$ denote the equivalence relation $E_0^{<\lambda}$ on $P\in\{2^\kappa,\kappa^\kappa\}$ (notation defined above). Then*

$$E_0^{<\lambda}(2^\kappa)\leqslant_c E_0^{<\lambda}(\kappa^\kappa) \text{ and } E_0^{<\lambda}(\kappa^\kappa)\leqslant_c E_0^{<\lambda}(2^\kappa).$$

Note that, when $\lambda=1$, we have $E_0^{<1}(P)=\mathrm{id}_P$.

Proof. In this proof we think of functions $\eta,\xi\in\kappa^\kappa$ as graphs $\eta=\{(\alpha,\eta(\alpha))\mid\alpha<\kappa\}$. Fix a bijection $h\colon\kappa\to\kappa\times\kappa$. Let $f\colon 2^\kappa\to\kappa^\kappa$ be the inclusion, $f(\eta)(\alpha)=\eta(\alpha)$. Then f is easily seen to be a continuous reduction $E_0^{<\lambda}(2^\kappa)\leqslant_c E_0^{<\lambda}(\kappa^\kappa)$. Define $g\colon\kappa^\kappa\to 2^\kappa$ as follows. For $\eta\in\kappa^\kappa$, let $g(\eta)(\alpha)=1$ if $h(\alpha)\in\eta$ and $g(\eta)(\alpha)=0$ otherwise. Let us show that g is a continuous reduction $E_0^{<\lambda}(\kappa^\kappa) \leqslant_c E_0^{<\lambda}(2^\kappa)$. Suppose $\eta, \xi \in \kappa$ are $E_0^{<\lambda}(\kappa^\kappa)$-equivalent. Then clearly $|\eta \triangle \xi| < \lambda$. On the other hand,

$$I = \{\alpha \mid g(\eta)(\alpha) \neq g(\xi)(\alpha)\} = \{\alpha \mid h(\alpha) \in \eta \triangle \xi\}$$

and, because h is a bijection, we have that $|I| < \lambda$.

Suppose η and ξ are not $E_0^{<\lambda}(\kappa^\kappa)$-equivalent. Then $|\eta \triangle \xi| \geqslant \lambda$ and the argument above shows that also $|I| \geqslant \lambda$, so $g(\eta)(\alpha)$ is not $E_0^{<\lambda}(2^\kappa)$-equivalent to $g(\xi)(\alpha)$.

The function g is easily seen to be continuous. □

We will need the following lemma, which is a straightforward generalization of the case $\kappa = \omega$:

Lemma 3.2 *Borel functions are continuous on a co-meager set.*

Proof. For each $\eta \in \kappa^{<\kappa}$ let V_η be an open subset of κ^κ such that $V_\eta \triangle f^{-1}N_\eta$ is meager. Let

$$D = \kappa^\kappa \setminus \bigcup_{\eta \in \kappa^{<\kappa}} V_\eta \triangle f^{-1}N_\eta.$$

Then D is as intended. Clearly it is co-meager, since we took away only a κ-union of meager sets. Let $\xi \in \kappa^{<\kappa}$ be arbitrary. The set $D \cap f^{-1}N_\xi$ is open in D since $D \cap f^{-1}N_\xi = D \cap V_\xi$ and so $f \restriction D$ is continuous. □

Theorem 3.3 ($\kappa^{<\kappa} = \kappa$) *$E_0^{<\lambda}$ is an equivalence relation on 2^κ for all $\lambda \leqslant \kappa$ and*

(1) *$E_0^{<\lambda}$ is Borel;*
(2) *$E_0^{<\kappa} \not\leqslant_B \mathrm{id}$;*
(3) *If $\lambda \leqslant \kappa$, then $\mathrm{id} \leqslant_c E_0^{<\lambda}$;*
(4) *If $\lambda < \kappa$, then $E_0^{<\lambda} \leqslant_c \mathrm{id}$.*

Proof. $E_0^{<\lambda}$ is clearly reflexive and symmetric. Suppose $\eta E_0^{<\lambda} \xi$ and $\xi E_0^{<\lambda} \zeta$. Denote $\eta = \eta^{-1}\{1\}$ and similarly for ξ, ζ. Then $|\eta \triangle \xi| < \lambda$ and $|\xi \triangle \zeta| < \lambda$; but $\eta \triangle \zeta \subset (\eta \triangle \xi) \cup (\xi \triangle \zeta)$. Thus $E_0^{<\lambda}$ is indeed an equivalence relation.

(1) $E_0^{<\lambda} = \bigcup_{A \in [\kappa]^{<\lambda}} \bigcap_{\alpha \notin A} \underbrace{\{(\eta, \xi) \mid \eta(\alpha) = \xi(\alpha)\}}_{\text{open}}$.

(2) Assume that there is a Borel reduction $f\colon 2^\kappa \to 2^\kappa$ witnessing $E_0 \leqslant_B \mathrm{id}$. By Lemma 3.2 there are dense open sets $(D_i)_{i<\kappa}$ such that $f \restriction \bigcap_{i<\kappa} D_i$ is continuous. If $p, q \in 2^\alpha$ for some α and $\xi \in N_p$, let us denote $\xi^{(p/q)} = q^\frown(\xi \restriction (\kappa \setminus \alpha))$, and if $A \subset N_p$, denote

$$A^{(p/q)} = \{\eta^{(p/q)} \mid \eta \in A\}.$$

Let C be the collection of sets, each of which is of the form

$$\bigcup_{q \in 2^\alpha} [D_i \cap N_p]^{(p/q)}$$

for some $\alpha < \kappa$ and some $p \in 2^\alpha$. It is easy to see that each such set is dense and open, so C is a collection of dense open sets. By the assumption $\kappa^{<\kappa} = \kappa$, C has size κ. Also C contains the sets D_i for all $i < \kappa$, (taking $\alpha = 0$). Denote $D = \bigcap_{i<\kappa} D_i$. Let $\eta \in \bigcap C$, $\xi = f(\eta)$ and $\xi' \neq \xi$, $\xi' \in \operatorname{ran}(f \restriction D)$. Now ξ and ξ' have disjoint open neighborhoods V and V' respectively. Let α and $p, q \in 2^\alpha$ be such that $\eta \in N_p$ and such that $D \cap N_p \subset f^{-1}[V]$ and $D \cap N_q \subset f^{-1}[V']$. These p and q exist by the continuity of f on D. Since $\eta \in \bigcap C$ and $\eta \in N_p$, we have

$$\eta \in [D_i \cap N_q]^{(q/p)}$$

for all $i < \kappa$, which is equivalent to

$$\eta^{(p/q)} \in [D_i \cap N_q]$$

for all $i < \kappa$, i.e., $\eta^{(p/q)}$ is in $D \cap N_q$. On the other hand (since $D_i \in C$ for all $i < \kappa$ and because $\eta \in N_p$), we have $\eta \in D \cap N_p$. This implies that $f(\eta) \in V$ and $f(\eta^{(p/q)}) \in V'$, which is a contradiction, because V and V' are disjoint and $(\eta, \eta^{p/q}) \in E_0$.

(3) Let $(A_i)_{i<\kappa}$ be a partition of κ into pieces of size κ: if $i \neq j$ then $A_i \cap A_j = \varnothing$, $\bigcup_{i<\kappa} A_i = \kappa$ and $|A_i| = \kappa$. Obtain such a collection for instance by taking a bijection $h\colon \kappa \to \kappa \times \kappa$ and defining $A_i = h^{-1}[\kappa \times \{i\}]$. Let $f\colon 2^\kappa \to 2^\kappa$ be defined by $f(\eta)(\alpha) = \eta(i) \Leftrightarrow \alpha \in A_i$. Now if $\eta = \xi$, then clearly $f(\eta) = f(\xi)$ and so $f(\eta) E_0^{<\lambda} f(\xi)$. If $\eta \neq \xi$, then there exists i such that $\eta(i) \neq \xi(i)$ and we have that

$$A_i \subset \{\alpha \mid f(\eta)(\alpha) \neq f(\xi)(\alpha)\}$$

and A_i is of size $\kappa \geqslant \lambda$.

(4) Let $P = \kappa^{<\kappa} \setminus \kappa^{<\lambda}$. Let $f\colon P \to \kappa$ be a bijection. It induces a bijection $g\colon 2^P \to 2^\kappa$. Let us construct a map $h\colon 2^\kappa \to 2^P$ such that $g \circ h$ is a reduction $E_0^{<\lambda} \to \mathrm{id}_{2^\kappa}$. Let us denote by $E^{<\lambda}(\alpha)$ the equivalence relation on 2^α such that two subsets X, Y of α are $E^{<\lambda}(\alpha)$-equivalent if and only if $|X \triangle Y| < \lambda$.

For each α in $\lambda < \alpha < \kappa$ let h_α be any reduction of $E^{<\lambda}(\alpha)$ to id_{2^α}. This exists because both equivalence relations have 2^α many classes. Now reduce $E_0^{<\lambda}$ to $\mathrm{id}_{\kappa^{<\kappa}}$ by $f(A) = (h_\alpha(A \cap \alpha) \mid \lambda \leqslant \alpha < \kappa)$. If A, B are $E_0^{<\lambda}$-equivalent, then $f(A) = f(B)$. Otherwise $f_\alpha(A \cap \alpha)$ differs from $f_\alpha(B \cap \alpha)$ for large enough $\alpha < \kappa$ because λ is less than κ and κ is regular. Continuity of h is easy to check. □

3.2 On the Silver dichotomy

To begin with, let us define the Silver dichotomy and the perfect set property:

Definition 3.4 Let $C \in \{\text{Borel}, \Delta_1^1, \text{Borel}^*, \Sigma_1^1, \Pi_1^1\}$. By *the Silver dichotomy*, or more specifically, *κ-SD for C* we mean the statement that there are no equivalence relations E in the class C such that $E \subset 2^\kappa \times 2^\kappa$ and E has more than κ equivalence classes such that $\mathrm{id} \not\leqslant_B E$, $\mathrm{id} = \mathrm{id}_{2^\kappa}$.

Similarly, the *perfect set property*, or *κ-PSP for C*, means that each member A of C has either size $\leqslant \kappa$ or there is a Borel injection $2^\kappa \to A$. Using Lemma 3.2, it is not hard to see that this definition is equivalent to the game definition given in [**24**].

3.2.1 The Silver dichotomy for isomorphism relations

Although the Silver dichotomy for Borel sets is not provable from ZFC for $\kappa > \omega$ (see Theorem 3.12 on page 499), it holds when the equivalence relation is an isomorphism relation, if $\kappa > \omega$ is an inaccessible cardinal:

Theorem 3.5 *Assume that κ is inaccessible. If the number of equivalence classes of $\cong_T$ is greater than κ, then* $\mathrm{id} \leqslant_c \cong_T$.

Proof. Suppose that there are more than κ equivalence classes of $\cong_T$. We will show that then $\mathrm{id}_{2^\kappa} \leqslant_c \cong_T$. If T is not classifiable, then, as was done in [**29**], we can construct a tree $t(S)$ for each $S \subset S^\kappa_\omega$ and Ehrenfeucht–Mostowski-type models $M(t(S))$ over these trees such that if $S \bigtriangleup S'$ is stationary then $M(t(S)) \not\cong M(t(S'))$. Now it is easy to construct a reduction $f\colon \mathrm{id}_{2^\kappa} \leqslant_c E_{S^\kappa_\omega}$ (see notation defined in Section 1.1), so then $\eta \mapsto M(t(f(\eta)))$ is a reduction $\mathrm{id} \leqslant_c \cong_T$.

Assume now that T is classifiable. By $\lambda(T)$ we denote the least cardinal in which T is stable. By [**30**, Theorem XIII.4.8] (this is also mentioned in [**8**, Theorem 2.5]), assuming that $\cong_T$ has more than κ equivalence classes, it has depth at least 2 and so there are: a $\lambda(T)^+$-saturated model $\mathfrak{B} \models T$, $|\mathfrak{B}| = \lambda(T)$, and a $\lambda(T)^+$-saturated elementary submodel $\mathcal{A} \preccurlyeq \mathfrak{B}$ and $a \notin \mathfrak{B}$ such that $\mathrm{tp}(a/\mathfrak{B})$ is orthogonal to $\mathcal{A}$. Let $f\colon \kappa \to \kappa$ be strictly increasing and such that, for all $\alpha < \kappa$, $f(\alpha) = \mu^+$ for some μ with the properties $\lambda(T) < \mu < \kappa$, $\mathrm{cf}(\mu) = \mu$ and $\mu^{2^\omega} = \mu$. For each $\eta \in 2^\kappa$ with $\eta^{-1}\{1\}$ unbounded we will construct a model $\mathcal{A}_\eta$. As above, it will be enough to show that $\mathcal{A}_\eta \not\cong \mathcal{A}_\xi$ whenever $\eta^{-1}\{1\} \bigtriangleup \xi^{-1}\{1\}$ is λ-stationary where $\lambda = \lambda(T)^+$. Fix $\eta \in 2^\kappa$ and let $\lambda = \lambda(T)^+$.

For each $\alpha \in \eta^{-1}\{1\}$ choose $\mathfrak{B}_\alpha \supset \mathcal{A}$ such that

(1) $\exists \pi_\alpha\colon \mathfrak{B} \cong \mathfrak{B}_\alpha$, $\pi_\alpha \restriction \mathcal{A} = \mathrm{id}_{\mathcal{A}}$;
(2) $\mathfrak{B}_\alpha \downarrow_{\mathcal{A}} \bigcup\{\mathfrak{B}_\beta \mid \beta \in \eta^{-1}\{1\}, \beta \neq \alpha\}$.

Note that (2) implies that, if $\alpha \neq \beta$, then $\mathfrak{B}_\alpha \cap \mathfrak{B}_\beta = \mathcal{A}$. For each $\alpha \in \eta^{-1}\{1\}$ and $i < f(\alpha)$, choose tuples a^α_i with the properties

(3) $\mathrm{tp}(a^\alpha_i/\mathfrak{B}_\alpha) = \pi_\alpha(\mathrm{tp}(a/\mathfrak{B}))$;
(4) $a^\alpha_i \downarrow_{\mathfrak{B}_\alpha} \bigcup\{a^\alpha_j \mid j < f(\alpha), j \neq i\}$.

Let $\mathcal{A}_\eta$ be F^s_λ-primary over

$$S_\eta = \bigcup\{B_\alpha \mid a < \eta^{-1}\{1\}\} \cup \bigcup\{a^\alpha_i \mid \alpha < \eta^{-1}\{1\}, i < f(\alpha)\}.$$

It remains to show that, if $S^\kappa_\lambda \cap \eta^{-1}\{1\} \bigtriangleup \xi^{-1}\{1\}$ is stationary, then $\mathcal{A}_\eta \not\cong \mathcal{A}_\xi$. Without loss of generality we may assume that $S^\kappa_\lambda \cap \eta^{-1}\{1\} \setminus \xi^{-1}\{1\}$ is stationary. Let us make a counter assumption, namely that there is an isomorphism $F\colon \mathcal{A}_\eta \to \mathcal{A}_\xi$.

Without loss of generality there exist singletons b^η_i and sets B^η_i, $i < \kappa$ of size $< \lambda$ such that $\mathcal{A}_\eta = S_\eta \cup \bigcup_{i<\kappa} b^\eta_i$ and $(S_\eta, (b^\eta_i, B^\eta_i)_{i<\kappa})$ is an F^s_λ-construction.

Let us find an ordinal $\alpha < \kappa$ and sets $C \subset \mathcal{A}_\eta$ and $D \subset \mathcal{A}_\xi$ with the properties listed below:

(a) $\alpha \in \eta^{-1}\{1\} \setminus \xi^{-1}\{1\}$;
(b) $D = F[C]$;
(c) $\forall \beta \in (\alpha + 1) \cap \eta^{-1}\{1\}(\mathfrak{B}_\beta \subset C)$ and $\forall \beta \in (\alpha + 1) \cap \xi^{-1}\{1\}(\mathfrak{B}_\beta \subset D)$;
(d) for all $i < f(\alpha)$, $\forall \beta \in \alpha \cap \eta^{-1}\{1\}(a^\beta_i \in C)$ and $\forall \beta \in \alpha \cap \xi^{-1}\{1\}(a^\beta_i \in D)$;
(e) $|C| = |D| < f(\alpha)$;
(f) for all β, if $\mathfrak{B}_\beta \cap C \setminus \mathcal{A} \neq \varnothing$, then $\mathfrak{B}_\beta \subset C$, and if $\mathfrak{B}_\beta \cap D \setminus \mathcal{A} \neq \varnothing$, then $\mathfrak{B}_\beta \subset D$;
(g) C and D are λ-saturated;
(h) if $b^\eta_i \in C$, then $B^\eta_i \subset [S_\eta \cup \bigcup\{b^\eta_i \mid j < i\}] \cap C$, and if $b^\xi_i \in D$, then $B^\xi_i \subset [S_\xi \cup \bigcup\{b^\xi_i \mid j < i\}] \cap D$.

This is possible, because $\eta^{-1}\{1\} \setminus \xi^{-1}\{1\}$ is stationary and we can close under the properties (b)–(h).

Now $\mathcal{A}_\eta$ is F^s_λ-primary over $C \cup S_\eta$ and $\mathcal{A}_\xi$ is F^s_λ-primary over $D \cup S_\eta$, and thus $\mathcal{A}_\eta$ is F^s_λ-atomic over $C \cup S_\eta$ and $\mathcal{A}_\xi$ is F^s_λ-atomic over $D \cup S_\xi$. Let

$$I_\alpha = \{a^\alpha_i \mid i < f(\alpha)\}.$$

Now $|I_\alpha \setminus C| = f(\alpha)$, because $|C| < f(\alpha)$, and so $I_\alpha \setminus C \neq \varnothing$. Let $c \in I_\alpha \setminus C$ and let $A \subset S_\xi \setminus D$ and $B \subset D$ be such that $\mathrm{tp}(F(c)/A \cup B) \vdash \mathrm{tp}(F(c)/D \cup S_\xi)$ and $|A \cup B| < \lambda$. Since $\alpha \notin \xi^{-1}\{1\}$, we can find (just take disjoint copies) a sequence $(A_i)_{i<f(\alpha)^+}$ such that $A_i \subset I_\alpha \cap \mathcal{A}_\xi$, $\mathrm{tp}(A_i/D) = \mathrm{tp}(A/D)$ and $A_i \downarrow_D \bigcup\{A_j \mid j \neq i, j < f(\alpha)^+\}$.

Now we can find $(d_i)_{i<f(\alpha)^+}$ such that

$$\mathrm{tp}(d_i{}^\frown A_i{}^\frown B_i/\varnothing) = \mathrm{tp}(F(c)^\frown A^\frown B/\varnothing).$$

Then it is a Morley sequence over D and, for all $i < f(\alpha)^+$,

$$\mathrm{tp}(d_i/D) = \mathrm{tp}(F(c)/D),$$

which implies that

$$\mathrm{tp}(F^{-1}(d_i)/C) = \mathrm{tp}(c/C)$$

for some i, since for some i we have $c = a^\alpha_i$. Since by (c), $\mathfrak{B}_\alpha \subset C$, the above implies that

$$\mathrm{tp}(F^{-1}(d_i)/\mathfrak{B}_\alpha) = \mathrm{tp}(a^\alpha_i/\mathfrak{B}_\alpha),$$

which by the definition of a^α_i, item 3, implies

$$\mathrm{tp}(F^{-1}(d_i)/\mathfrak{B}_\alpha) = \pi_\alpha(\mathrm{tp}(a/\mathfrak{B})).$$

Thus the sequence $(F^{-1}(d_i))_{i<f(\alpha)^+}$ witnesses that the dimension of $\pi_\alpha(\mathrm{tp}(a/\mathfrak{B}))$ in $\mathcal{A}_\eta$ is greater than $f(\alpha)$. Denote that sequence by J. Since $\pi_\alpha(\mathrm{tp}(a/\mathfrak{B}))$ is orthogonal to $\mathcal{A}$, we can find $J' \subset J$ such that $|J'| = f(\alpha)^+$ and J' is a Morley sequence over S_η. Since $f(\alpha)^+ > \lambda$, this contradicts Theorem 4.9(2) of Chapter IV of [**30**]. □

Open Problem Under what conditions on κ does the conclusion of Theorem 3.5 hold?

3.2.2 Theories bireducible with id

Theorem 3.6 *Assume $\kappa^{<\kappa} = \kappa = \aleph_\alpha > \omega$, κ is not weakly inaccessible and $\lambda = |\alpha + \omega|$. Then the following are equivalent:*

(1) *There is $\gamma < \omega_1$ such that $\beth_\gamma(\lambda) \geqslant \kappa$.*
(2) *There is a complete countable T such that* $\mathrm{id} \leqslant_B \cong_T$ *and* $\cong_T \leqslant_B \mathrm{id}$.

Proof. (2)⇒(1): Suppose that (1) is not true. Notice that then $\kappa > 2^\omega$. Then every shallow classifiable theory has $< \kappa$ many models of power κ (see [**8**], item 6 of the theorem which is on the first page of the article) and thus $\mathrm{id} \not\leqslant_B \cong_T$. On the other hand if T is not classifiable and shallow, $\cong_T$ is not Borel by Theorem 4.7 and thus it is not Borel reducible to id by Fact 5.1.

(1)⇒(2): Since $\mathrm{cf}(\kappa) > \omega$, (1) implies that there is $\alpha = \beta + 1 < \omega_1$ such that $\beth_\alpha(\lambda) = \kappa$. But then there is an L^*-theory T^* which has exactly κ many models in cardinality κ (up to isomorphism, use [**8**], Theorem 6.1 items 2 and 8). But then it has exactly κ many models of cardinality $\leqslant \kappa$; let $\mathcal{A}_i$, $i < \kappa$ list these. Such a theory must be classifiable and shallow. Let L be the vocabulary we get from L^* by adding one binary relation symbol E. Let $\mathcal{A}$ be an L-structure in which E is an equivalence relation with infinitely many equivalence classes such that for every equivalence class a/E, $(\mathcal{A} \restriction a/E) \restriction L^*$ is a model of T^*. Let $T = \mathrm{Th}(\mathcal{A})$.

We show first that identity on $\{\eta \in 2^\kappa |\ \eta(0) = 1\}$ reduces to $\cong_T$. For all $\eta \in 2^\kappa$, let $\mathfrak{B}_\eta$ be a model of T of power κ such that if $\eta(i) = 0$, then the number of equivalence classes isomorphic to $\mathfrak{B}_i$ is countable and otherwise the number is κ. Clearly we can code $\mathfrak{B}_\eta$ as $\xi_\eta \in 2^\kappa$ so that $\eta \mapsto \xi_\eta$ is the required Borel reduction.

We show then that $\cong_T$ Borel reduces to identity on

$$X = \{\eta \colon \kappa \to (\kappa + 1)\}.$$

Since T^* is classifiable and shallow, for all $\delta, i < \kappa$ the set

$$\{\eta \in X |\ (\mathcal{A}_\eta \restriction \delta/E) \restriction L^* \cong \mathcal{A}_i\}$$

is Borel. But then for all cardinals $\theta \leqslant \kappa$ and $i < \kappa$, the set

$$\{\eta \in X \mid |(|\{\delta/E \mid\ \delta < \kappa,\ (\mathcal{A}_\eta \restriction \delta/E) \restriction L^* \cong \mathcal{A}_i\}) = \theta\}$$

is Borel. Then $\eta \mapsto \xi_\eta$ is the required reduction when

$$\xi_\eta(i) = |\{\delta/E \mid\ \delta < \kappa,\ (\mathcal{A}_\eta \restriction \delta/E) \restriction L^* \cong \mathcal{A}_i\}|. \qquad \square$$

In the above it was assumed that κ is not inaccessible. If κ is inaccessible, then (2) of the above theorem always holds:

Theorem 3.7 *Suppose κ is inaccessible and $\kappa^{<\kappa} = \kappa$. Then there is a theory T such that $\cong_T$ is bireducible with* id_{2^κ}.

Proof. Let $\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow$ be the model with domain $M = \mathrm{dom}\,\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow = \omega \cup (\omega \times \omega)$ and a binary relation R which is interpreted as

$$R^{\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow} = \{(a, (b, c)) \in M^2 \mid a \in \omega, (b, c) \in \omega \times \omega, a = b\}.$$

Then our intended theory is the complete first-order theory of $T = \mathrm{Th}(\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow)$.

Let $\hat{C} = \{\aleph_\beta \mid \beta \leqslant \kappa\}$ and $C = \omega \cup \hat{C}$. Let $\mathcal{A}$ be a model of T of size κ and let $f_\mathcal{A} \colon \hat{C} \to C$ be a function such that

$$f_\mathcal{A}(\aleph_\beta) = |(|\{x \in A \mid |(|\{(a, b) \in A \mid R(x, (a, b))\}\}) = \aleph_\beta\}), \tag{3.1}$$

i.e., $f_\mathcal{A}(\aleph_\beta)$ equals the number of elements which are R-related to exactly $\aleph_\beta$ elements. Clearly $\mathcal{A} \cong \mathfrak{B}$ is equivalent to $f_\mathcal{A} = f_\mathfrak{B}$.

Let $g_0 \colon \hat{\mu} \to \hat{C}$ and $g_1 \colon \mu \to C$ be bijections. Let us define the function F by

$$F(\xi) = g_1^{-1} \circ f_{\mathcal{A}_\xi} \circ g_0.$$

Now F is a reduction $\cong_T \leqslant \mathrm{id}_{\kappa^\kappa}$. By Theorem 3.1, page 490, $\mathrm{id}_{\kappa^\kappa}$ is continuously bireducible with id_{2^κ}. Let us show that F is Borel. In order to do it, we will use the easy direction (right to left) of Theorem 2.2 on page 483. Because every basic open set in κ^κ is an intersection of the sets of the form

$$U_{\gamma\delta} = \{\eta \in \kappa^\kappa \mid \eta(\gamma) = \delta\},$$

it is enough to show that $F^{-1}[U_{\gamma\delta}]$ is Borel for any $\gamma, \delta \in \kappa$.

Note that $\eta \in F^{-1}[U_{\gamma\delta}]$ is equivalent to

($\star$) *there exist exactly $g_1(\delta)$ elements in $F^{-1}(\eta)$ which are R-related to exactly $g_0(\gamma)$ elements.*

We can express ($\star$) in $L_{\kappa^+\kappa}$. First, let us define the formula φ_λ for $\lambda < \kappa$ which says that the variable x is R-related to exactly λ elements:

$$\varphi_\lambda(x) : \exists_{i<\lambda} y_i \Big[\Big(\bigwedge_{j_0<j_1<\lambda} \neg y_{j_0} = y_{j_1}\Big) \wedge \bigwedge_{i<\lambda} R(x, y_i) \wedge \forall z \Big(R(x,z) \to \bigvee_{i<\lambda} z = y_i\Big)\Big].$$

Then one can write the formula which says that there are exactly $\nu < \kappa$ such x_k that satisfy φ_λ:

$$\psi_{\lambda\nu} : \exists_{k<\nu} x_k \Big[\Big(\bigwedge_{i<j<\nu} \neg x_i = x_j\Big) \wedge \bigwedge_{k<\nu} \varphi_\lambda(x_k) \wedge \forall z \Big(\varphi_\lambda(z) \to \bigvee_{k<\nu} (z = x_k)\Big)\Big].$$

For the cases $\gamma = \kappa$, $\delta = \kappa$, define

$$\varphi_\kappa(x_k) : \bigwedge_{\beta<\kappa} \forall_{i<\beta} y_i \Big[\exists y_\beta \Big[\Big(\bigwedge_{i<\beta} (y_\beta \neq y_i)\Big) \wedge R(x_k, y_\beta)\Big]\Big]$$

and

$$\psi_{\kappa\lambda} : \bigwedge_{\beta<\kappa} \forall_{k<\beta} x_k \Big[\exists x_\beta \Big[\Big(\bigwedge_{k<\beta} (x_\beta \neq x_k)\Big) \wedge \varphi_\lambda(x_\beta)\Big]\Big].$$

Note that the last formulas say “for all $\beta < \kappa$ there exist more than β...”, but it is equivalent to “there exist exactly κ...” in our class of models, because the models are all of size κ.

Thus $\psi_{g_0(\gamma),g_1(\delta)}$ is defined for all $\gamma \leqslant \kappa$ and $\delta \leqslant \kappa$. By the direction right to left of Theorem 2.2, this implies that the sets $F^{-1}U_{\gamma\delta}$ are Borel. This proves $\cong_T \leqslant_B \mathrm{id}_{2^\kappa}$.

Since κ is inaccessible, the other direction follows from Theorem 3.5, page 492. On the other hand one easily constructs such a reduction from scratch. Let us do it for the sake of completeness.

Let us show that id $\leqslant_c \cong_T$. Let us modify the setting a little; let $C_{<\kappa} = \{\lambda < \kappa \mid \lambda \text{ is a cardinal}\}$ and $C^\omega_{<\kappa} = C_{<\kappa} \setminus \omega$, and let

$$h_0 \colon \kappa \longrightarrow C^\omega_{<\kappa}$$

and

$$h_1 \colon \kappa \longrightarrow C_{<\kappa}$$

be increasing bijections. Suppose $\eta \in \kappa^\kappa$ and define $f_\eta \colon C^\omega_{<\kappa} \to C_{<\kappa}$ by

$$f_\eta(\lambda) = [(h_1 \circ \eta \circ h_0^{-1})(\lambda)]^+$$

(recall that κ is inaccessible). Let us now build the model $\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta$:

$$\operatorname{dom} \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta = \bigcup_{\lambda \in C^\omega_{<\kappa}} \{(\lambda, f_\eta(\lambda))\} \times [f_\eta(\lambda) \cup f_\eta(\lambda) \times \lambda]$$

(that is, formally $\operatorname{dom} \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta$ consists of pairs and triples the first projection being a pair of the form $(\lambda, f_\eta(\lambda)))$ and, for all $x, y \in \operatorname{dom} \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta$,

$$R(x,y) \iff \exists\lambda\exists\alpha\exists\beta\big(x = ((\lambda, f_\eta(\lambda)), \alpha) \wedge y = ((\lambda, f_\eta(\lambda)), \alpha, \beta)\big).$$

Denote the mapping $\eta \mapsto \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta$ by G, i.e., $G(\eta) = \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta$. Clearly $\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta \models T$. Let us show that

$$\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta \cong \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\xi \iff \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta = \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\xi \iff \eta = \xi.$$

The implications from right to left are evident. Suppose $h\colon \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta \to \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\xi$ is an isomorphism. Since it preserves relations, the restrictions send bijectively the λ-levels to some other λ'-levels:

$$h\restriction\{(\lambda, f_\eta(\lambda))\} \times [\{\alpha\} \cup \{\beta\} \times \lambda] \to \{(\lambda', f_\eta(\lambda'))\} \times [\{\alpha'\} \cup \{\beta'\} \times \lambda']$$

is a bijection which implies $\lambda = \lambda'$. Further, by bijectivity, the map $\alpha \mapsto \alpha'$ induced by these restrictions is also bijective (by preservation of relations, pairs are sent to pairs), so this map $\alpha \mapsto \alpha'$ is a bijection between $f_\eta(\lambda)$ and $f_\xi(\lambda)$, thus they are the same cardinal for all λ, i.e., $f_\eta = f_\xi$.

For a model of the form $\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta$ and $\alpha < \kappa$, let

$$\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_{\eta\restriction\alpha} = \bigcup_{\lambda \in C^\omega_{<\kappa},\ \lambda < h_0(\alpha)} \{(\lambda, f_\eta(\lambda))\} \times [f_\eta(\lambda) \cup f_\eta(\lambda) \times \lambda]$$

equipped with the relation $R^{\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_{\eta\restriction\alpha}} = R^{\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow} \cap (\operatorname{dom} \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_{\eta\restriction\alpha})^2$.

Let us fix a well ordering of $\operatorname{dom}\mathcal{A}$ for each model $\mathcal{A} \in \operatorname{ran} G$ as follows. If $x, y \in \operatorname{dom}\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta$, then

$$\begin{aligned} x \lessdot y \iff\ & \operatorname{pr}_1(x) < \operatorname{pr}_1(y) \\ & \text{or } \operatorname{pr}_1(x) = \operatorname{pr}_1(y) \wedge \operatorname{pr}_2(x) < \operatorname{pr}_2(y) \\ & \text{or } \operatorname{pr}_1(x) = \operatorname{pr}_1(y) \wedge \operatorname{pr}_2(x) = \operatorname{pr}_2(y) \wedge \operatorname{pr}_3(x) < \operatorname{pr}_3(y). \end{aligned}$$

Note that in the last case it might happen that there is no third projection of x. In that case define $\operatorname{pr}_3(x)$ to be -1. (If $\operatorname{pr}_3(y)$ were also undefined, then we had $x = y$.) The initial segments with respect to $\lessdot$ are of size less than κ, because $f_\eta(\lambda)$ and λ are elements of $C_{<\kappa}$ and $\lessdot$ is clearly a well ordering. Moreover, since we added the $+$ in the definition of $f_\eta(\lambda)$, we have that $\forall\lambda\forall\eta(f_\eta(\lambda) > 0)$, so we get the following:

($\star\star$) Suppose x is the γ-th element of the model with respect to $\lessdot$. Then $\operatorname{pr}_1(x) \leqslant \gamma$. Hence, for any η,

$$\begin{aligned} \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta \cap \{x \in \operatorname{dom}\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta \mid \operatorname{OTP}_\lessdot(x) < \gamma\} \\ \subset \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_{\eta\restriction(\gamma+1)}. \end{aligned}$$

Note also that if $\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_{\eta\restriction\alpha} = \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_{\xi\restriction\alpha}$, then the identity map $\operatorname{id}\colon \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_{\eta\restriction\alpha} = \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_{\xi\restriction\alpha}$ preserves $\lessdot$.

Recall the coding $\eta \mapsto \mathcal{A}_\eta$ of Definition 1.14. In the definition it is assumed that $\operatorname{dom}\mathcal{A} = \kappa$, but instead of that we can use the well-ordering $\lessdot$. More precisely, for a given model $\mathcal{A}$, let $c(\mathcal{A})$ denote some η such that there is an isomorphism $f\colon \mathcal{A}_\eta \cong \mathcal{A}$ which preserves the ordering of the domain: $f(\alpha)$ is the α-th element of $\operatorname{dom}\mathcal{A}$ with respect to $\lessdot$. In our present case, $c\colon \operatorname{ran} G \to \kappa^\kappa$.

Let us show that the map $F = c \circ G\colon \eta \mapsto c(\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta)$ is continuous and therefore is the intended bijection. For that purpose let us equip $\operatorname{ran} G$ with a topology τ. We will then show that G is continuous with respect to that topology and then show that also c is continuous.

Let τ be the topology on $\operatorname{ran} G$ generated by

$$U_p = \{\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta \mid p \subset \eta\}$$

for $p \in \kappa^{<\kappa}$. In fact τ is the topology co-induced by G, so it trivially makes G continuous:

$$G^{-1}U_p = N_p.$$

Let us show that

$$(\star\star\star)\qquad U_p = \{\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow \in \operatorname{ran} G \mid \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_p \subset \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow\}.$$

Suppose $\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta \in U_p$ for some η. This is equivalent to assuming that there is ξ with $p \subset \xi$ such that $\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta = \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\xi$. This is in turn equivalent with $p \subset \eta$, since necessarily $\eta = \xi$. So $\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta \in U_p$ implies that

$$\begin{aligned}\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_p &= \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_{\eta\restriction\operatorname{dom} p}\\ &= \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta \cap \bigcup_{\lambda\in C^\omega_{<\kappa},\ \lambda<h_0(\operatorname{dom} p)} \{\lambda\}\times[f_\eta(\lambda)\cup f_\eta(\lambda)\times\lambda]\\ &\subset \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta.\end{aligned}$$

Assume that $\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow \in \operatorname{ran} G$, $\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_p \subset \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow$ and that η is such that $\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow = \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta$. Let us assume that ξ is such that $p \subset \xi$ and let us show that $\xi\restriction\operatorname{dom} p \subset \eta$. Let $\lambda < h_0(\operatorname{dom} p)$. Then because $f_\xi(\lambda) > 0$, we have

$$(\lambda, f_\xi(\lambda), 0) \in \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_p.$$

By the assumption $\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_p \subset \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta$, this implies $(\lambda, f_\xi(\lambda), 0) \in \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta$. By definition, this can only happen if $f_\eta(\lambda) = f_\xi(\lambda)$. Thus for all $\lambda < h_0(\operatorname{dom} p)$, we have $f_\eta(\lambda) = f_\xi(\lambda)$. Recall that h_1 and h_0 are increasing bijections, so

$$\begin{aligned}&[\forall\lambda < h_0(\operatorname{dom} p)](f_\eta(\lambda) = f_\xi(\lambda))\\ &\quad\iff [\forall\lambda < h_0(\operatorname{dom} p)]((h_1\circ\eta\circ h_0^{-1})(\lambda) = (h_1\circ\xi\circ h_0^{-1})(\lambda))\\ &\quad\iff [\forall\alpha < \operatorname{dom} p]((h_1\circ\eta)(\alpha) = (h_1\circ\xi)(\alpha))\\ &\quad\iff [\forall\alpha < \operatorname{dom} p](\eta(\alpha) = \xi(\alpha))\\ &\quad\iff [\forall\alpha < \operatorname{dom} p](\eta(\alpha) = p(\alpha)),\end{aligned}$$

and this implies that $p \subset \eta$.

Consider now the coding $c\colon \operatorname{ran} G \to \kappa^\kappa$. Let $N_{\xi\restriction\alpha}$ be a basic open set of κ^κ. Let $\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow$ be a model in $c^{-1}N_{\xi\restriction\alpha}$. Let us show that there is an open τ-neighborhood of $\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow$ inside $c^{-1}N_{\xi\restriction\alpha}$. We know that $\xi\restriction\alpha$ decides a segment of $\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow$ that is below the γ-th element with respect to $\lessdot$, for some γ. Denote that segment by $S \subset \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow$. Let η be such that $\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow = \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta$. From $(\star\star)$ we have

$$\begin{aligned}S &\subset \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta \cap \{x \in \operatorname{dom}\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_\eta \mid \mathrm{OTP}_\lessdot(x) < \gamma\}\\ &\subset \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_{\eta\restriction(\gamma+1)}.\end{aligned}$$

Let us show that $U_{\eta\restriction(\gamma+1)}$ is an open neighborhood of $\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow$ inside $c^{-1}[N_{\xi\restriction\alpha}]$. Suppose $\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow \in U_{\eta\restriction(\gamma+1)}$ and $c(\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow) = \zeta$. Then by $(\star\star\star)$ we have $\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_{\eta\restriction(\gamma+1)} \subset \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow$. Let $S' \subset \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow$ be the subset of $\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow$ decided by $\zeta\restriction\alpha$. Thus

$$\{\mathrm{OTP}_\lessdot(x) \mid x \in S'\} = \{\mathrm{OTP}_\lessdot(x) \mid x \in S\},$$

but, by the note after $(\star\star)$, we have $S = S'$, and, since $S \subset \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_{\eta\restriction(\gamma+1)}$ and $\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_{\eta\restriction(\gamma+1)} = \mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow_{\zeta\restriction(\gamma+1)}$, the codings must coincide and we have $\zeta\restriction\alpha = \xi\restriction\alpha$, i.e., $c(\mathcal{M}\rfloor\dashv\updownarrow\rfloor\dashv\updownarrow) \in N_{\xi\restriction\alpha}$. □

3.2.3 Failures of Silver's dichotomy

There are well-known dichotomy theorems for Borel equivalence relations on 2^ω. Two of them are:

Theorem 3.8 (Silver [**33**]) *Let $E \subset 2^\omega \times 2^\omega$ be a Π^1_1 equivalence relation. If E has uncountably many equivalence classes, then* $\mathrm{id}_{2^\omega} \leqslant_B E$. □

Theorem 3.9 (Generalized Glimm–Effros dichotomy [**7**]) *Let $E \subset 2^\omega \times 2^\omega$ be a Borel equivalence relation. Then either $E \leqslant_B \mathrm{id}_{2^\omega}$ or else $E_0 \leqslant_c E$.* □

As in the case $\kappa = \omega$ we have the following also for uncountable κ (see Definition 3.4, page 492):

Theorem 3.10 *If κ-SD for Π^1_1 holds, then the κ-PSP holds for Σ^1_1-sets. More generally, if $C \in \{\mathrm{Borel}, \Delta^1_1, \mathrm{Borel}^*, \Sigma^1_1, \Pi^1_1\}$, then κ-SD for C implies κ-PSP for C', where elements in C' are all the complements of those in C.*

Proof. Let us prove this for $C = \Pi^1_1$, as the other cases are similar. Suppose we have a Σ^1_1-set A. Let

$$E = \{(\eta, \xi) \mid \eta = \xi \text{ or } ((\eta \notin A) \wedge (\xi \notin A))\}.$$

Now $E = \mathrm{id} \cup (2^\kappa \setminus A)^2$. Since A is Σ^1_1, $(2^\kappa \setminus A)^2$ is Π^1_1 and because id is Borel, also E is Π^1_1. Obviously $|A|$ is the number of equivalence classes of E provided A is infinite. Then suppose $|A| > \kappa$. Then there are more than κ equivalence classes of E, so by κ-SD for Π^1_1, there is a reduction $f\colon \mathrm{id} \leqslant E$. This reduction in fact witnesses the PSP of A. □

The idea of using Kurepa trees for this purpose arose already in the paper [**24**] by Mekler and Väänänen.

Definition 3.11 If $t \subset 2^{<\kappa}$ is a tree, a *path* through t is a branch of length κ. A *κ-Kurepa tree* is a tree $K \subset 2^{<\kappa}$ which satisfies the following:

(a) K has more than κ paths;
(b) K is downward closed;
(c) for all $\alpha < \kappa$, the levels are small: $|\{p \in K \mid \mathrm{dom}\, p = \alpha\}| \leqslant |\alpha + \omega|$.

Theorem 3.12 *Assume one of the following:*

(1) *κ is regular but not strongly inaccessible and there exists a κ-Kurepa tree $K \subset 2^{<\kappa}$.*
(2) *κ is regular (might be strongly inaccessible), $2^\kappa > \kappa^+$ and there exists a tree $K \subset 2^{<\kappa}$ with more than κ but less than 2^κ branches.*

Then the Silver dichotomy for κ does not hold. In fact there is an equivalence relation $E \subset 2^\kappa \times 2^\kappa$ which is the union of a closed and an open set, has more than κ equivalence classes but $\mathrm{id}_{2^\kappa} \not\leqslant_B E$.

Proof. Let us break the proof according to the assumptions (1) and (2). So first let us consider the case where κ is not strongly inaccessible and there is a κ-Kurepa tree.

(1): Let us carry out the proof in the case $\kappa = \omega_1$. It should be obvious then how to generalize it to any κ not strongly inaccessible. So let $K \subset 2^{<\omega_1}$ be an ω_1-Kurepa tree. Let P be the collection of all paths of K. For $b \in P$, denote $b = \{b_\alpha \mid \alpha < \omega_1\}$ where b_α is an element of K with domain α.

Let

$$C = \{\eta \in 2^{\omega_1} \mid \eta = \bigcup_{\alpha < \omega_1} b_\alpha, b \in P\}.$$

Clearly C is closed. Let $E = \{(\eta,\xi) \mid (\eta \notin C \wedge \xi \notin C) \vee (\eta \in C \wedge \eta = \xi)\}$. In words, E is the equivalence relation whose equivalence classes are the complement of C and the singletons formed by the elements of C. E is the union of the open set $\{(\eta,\xi) \mid \eta \notin C \wedge \xi \notin C\}$ and the closed set $\{(\eta,\xi) \mid \eta \in C \wedge \eta = \xi\} = \{(\eta,\eta) \mid \eta \in C\}$. The number of equivalence classes equals the number of paths of K, so there are more than ω_1 of them by the definition of Kurepa tree.

Let us show that $\mathrm{id}_{2^{\omega_1}}$ is not embeddable to E. Suppose that $f\colon 2^{\omega_1} \to 2^{\omega_1}$ is a Borel reduction. We will show that then K must have a level of size $\geqslant \omega_1$ which contradicts the definition of Kurepa tree. By Lemma 3.2, page 491, there is a co-meager set D on which $f \restriction D$ is continuous. There is at most one $\eta \in 2^{\omega_1}$ whose image $f(\eta)$ is outside C, so without loss of generality $f[D] \subset C$. Let p be an arbitrary element of K such that $f^{-1}[N_p] \neq \varnothing$. By continuity there is a $q \in 2^{<\omega_1}$ with $f[N_q \cap D] \subset N_p$. Since D is co-meager, there are η and ξ such that $\eta \neq \xi$, $q \subset \eta$ and $q \subset \xi$. Let $\alpha_1 < \omega_1$ and p_0 and p_1 be extensions of p with the properties $p_0 \subset f(\eta)$, $p_1 \subset f(\xi)$, $\alpha_1 = \operatorname{dom} p_0 = \operatorname{dom} p_1$, $f^{-1}[N_{p_0}] \neq \varnothing \neq f^{-1}[N_{p_1}]$ and $N_{p_0} \cap N_{p_1} = \varnothing$. Note that p_0 and p_1 are in K. Then, again by continuity, there are q_0 and q_1 such that $f[N_{q_0} \cap D] \subset N_{p_0}$ and $f[N_{q_1} \cap D] \subset N_{p_1}$. Continue in the same manner to obtain α_n and $p_s \in K$ for each $n < \omega$ and $s \in 2^{<\omega}$ so that $s \subset s' \Leftrightarrow p_s \subset p_{s'}$ and $\alpha_n = \operatorname{dom} p_s \Leftrightarrow n = \operatorname{dom} s$. Let $\alpha = \sup_{n<\omega} \alpha_n$. Now clearly the α's level of K contains continuum many elements: by (b) in the definition of Kurepa tree it contains all the elements of the form $\bigcup_{n<\omega} p_{\eta \restriction n}$ for $\eta \in 2^\omega$ and $2^\omega \geqslant \omega_1$.

If κ is any regular not strongly inaccessible cardinal, then the proof is the same, but instead of ω steps one has to do λ steps where λ is the least cardinal with $2^\lambda \geqslant \kappa$.

(2): The argument is even simpler. Define the equivalence relation E exactly as above. Now E is again closed and has as many equivalence classes as is the number of paths in K. Thus the number of equivalence classes is $> \kappa$ but id cannot be reduced to E since there are less than 2^κ equivalence classes. □

Remark 3.13 Some related results:

1. In L, the PSP fails for closed sets for all uncountable regular κ. This is because "weak Kurepa trees" exist (see the proof sketch of (3) below for the definition of "weak Kurepa tree").
2. (P. Schlicht) In Silver's model where an inaccessible κ is made into ω_2 by Lévy collapsing each ordinal below to ω_1 with countable conditions, every Σ^1_1 subset X of 2^{ω_1} obeys the PSP.
3. Supercompactness does not imply the PSP for closed sets.

Sketch of a proof of item (3). Suppose κ is supercompact and by a reverse Easton iteration add to each inaccessible α a "weak Kurepa tree", i.e., a tree T_α with α^+ branches whose β-th level has size β for stationary many $\beta < \alpha$. The forcing at stage α is α-closed and the set of branches through T_κ is a closed set with no perfect subset. If $j\colon V \to M$ witnesses λ-supercompactness ($\lambda > \kappa$) and G is the generic then we can find G^* which is $j(P)$-generic over M containing $j[G]$: Up to λ we copy G, between λ and $j(\kappa)$ we build G^* using λ^+ closure of the forcing and of the model M, and at $j(\kappa)$ we form a master condition out of $j[G(\kappa)]$ and build a generic below it, again using λ^+ closure. □

Corollary 3.14 *The consistency of the Silver dichotomy for Borel sets on ω_1 with CH implies the consistency of a strongly inaccessible cardinal. In fact, if there is no equivalence relation witnessing the failure of the Silver dichotomy for ω_1, then ω_2 is inaccessible in L.*

Proof. By a result of Silver, if there are no ω_1-Kurepa trees, then ω_2 is inaccessible in L; see Exercise 27.5 in Part III of [**16**]. □

Open Problem Is the Silver dichotomy for uncountable κ consistent?

3.3 Regularity properties and definability of the CUB filter

In the standard descriptive theory ($\kappa = \omega$), the notions of Borel, Δ^1_1 and Borel* coincide and one of the most important observations in the theory is that such sets have the property of Baire and that the Σ^1_1-sets obey the perfect set property. In the case $\kappa > \omega$ the situation is more complicated as the following shows. It was already pointed out in the previous section that Borel $\subsetneq \Delta^1_1$. In this section we focus on the cub filter

$$\mathrm{CUB} = \{\eta \in 2^\kappa \mid \eta^{-1}\{1\} \text{ contains a cub}\}.$$

The set CUB is easily seen to be Σ^1_1: the set

$$\{(\eta,\xi) \mid (\eta^{-1}\{1\} \subset \xi^{-1}\{1\}) \wedge (\eta^{-1}\{1\} \text{ is cub})\}$$

is Borel. CUB (restricted to cofinality ω; see Definition 3.19 below) will serve (consistently) as a counterexample to $\Delta^1_1 = $ Borel*, but we will show that it is also consistent that CUB is Δ^1_1. The latter implies that it is consistent that Δ^1_1-sets do not have the property of Baire and we will also show that in a forcing extension of L, Δ^1_1-sets all have the property of Baire.

Definition 3.15 A *nowhere dense set* is a subset of a set whose complement is dense and open. Let $X \subset \kappa^\kappa$. A subset $M \subset X$ is *κ-meager in X* if $M \cap X$ is the union of no more than κ nowhere dense sets,

$$M = \bigcup_{i<\kappa} N_i.$$

We usually drop the prefix "κ-".

Clearly κ-meager sets form a κ-complete ideal. A *co-meager* set is a set whose complement is meager.

A subset $A \subset X$ *has the property of Baire*, or shorter *P.B.*, if there exists an open $U \subset X$ such that the symmetric difference $U \vartriangle A$ is meager.

Halko showed in [**5**] that

Theorem 3.16 ([**5**]) *Borel sets have the property of Baire.* □

(The same proof as when $\kappa = \omega$ works.) This is independent of the assumption $\kappa^{<\kappa} = \kappa$. Borel* sets do not in general have the property of Baire.

Definition 3.17 ([**11, 23, 24**]) A $\kappa^+\kappa$-tree t is a *$\kappa\lambda$-canary tree* if for all stationary $S \subset S^\kappa_\lambda$ it holds that, if $\mathbb{P}$ does not add subsets of κ of size less than κ and $\mathbb{P}$ kills the stationarity of S, then $\mathbb{P}$ adds a κ-branch to t.

Remark Hyttinen and Rautila [**11**] use the term *κ-canary tree* for our *$\kappa^+\kappa$-canary tree.*

It was shown by Mekler and Shelah [**23**] and Hyttinen and Rautila [**11**] that it is consistent with ZFC + GCH that there is a $\kappa^+\kappa$-canary tree *and* it is consistent with ZFC + GCH that there are no $\kappa^+\kappa$-canary trees. The same proof as in [**11, 23**] gives the following:

Theorem 3.18 *Assume GCH and assume $\lambda < \kappa$ are regular cardinals. Let $\mathbb{P}$ be the forcing which adds κ^+ Cohen subsets of κ. Then in the forcing extension there are no $\kappa\lambda$-canary trees.* □

Definition 3.19 Suppose $X \subset \kappa$ is stationary. For each such X define the set

$$\mathrm{CUB}(X) = \{\eta \in 2^\kappa \mid X \setminus \eta^{-1}\{1\} \text{ is non-stationary}\},$$

so $\mathrm{CUB}(X)$ is "cub in X".

Theorem 3.20 *In the following κ satisfies $\kappa^{<\kappa} = \kappa > \omega$.*

(1) $\mathrm{CUB}(S^\kappa_\omega)$ *is Borel*.*

(2) *For all regular $\lambda < \kappa$,* $\mathrm{CUB}(S^\kappa_\lambda)$ *is not Δ^1_1 in the forcing extension after adding κ^+ Cohen subsets of κ.*

(3) *If $V = L$, then for every stationary $S \subset \kappa$, the set* $\mathrm{CUB}(S)$ *is not Δ^1_1.*

(4) *Assume GCH and that κ is not a successor of a singular cardinal. For any stationary set $Z \subset \kappa$ there exists a forcing notion $\mathbb{P}$ which has the κ^+-c.c., does not add bounded subsets of κ and preserves GCH and stationary subsets of $\kappa \setminus Z$ such that* $\mathrm{CUB}(\kappa \setminus Z)$ *is Δ^1_1 in the forcing extension.*

(5) *Let the assumptions for κ be as in* (3.20)*. For all regular $\lambda < \kappa$,* $\mathrm{CUB}(S^\kappa_\lambda)$ *is Δ^1_1 in a forcing extension as in* (3.20)*.*

(6) $\mathrm{CUB}(X)$ *does not have the property of Baire for stationary $X \subset \kappa$. (Proved by Halko and Shelah in* **[6]** *for $X = \kappa$.)*

(7) *It is consistent that all Δ^1_1-sets have the property of Baire. (Independently known to P. Lücke and P. Schlicht.)*

Proof of Theorem 3.20.

Proof of item (1). Let $t = [\kappa]^{<\omega}$ (increasing functions ordered by end extension) and, for all branches $b \subset t$,

$$h(b) = \left\{ \xi \in 2^\kappa \mid \xi\Big(\sup_{n<\omega} b(n)\Big) \neq 0 \right\}.$$

Now if $\kappa \setminus \xi^{-1}\{0\}$ contains an ω-cub set C, then player **II** has a winning strategy in $G(t, h, \xi)$: for her n-th move she picks an element $x \in t$ with domain $2n+2$ such that $x(2n+1)$ is in C. Suppose the players picked a branch b in this way. Then the condition $\xi(b(2n+1)) \neq 0$ holds for all $n < \omega$ and because C is cub outside $\xi^{-1}\{0\}$, we have $\xi(\sup_{n<\omega} b(n)) \neq 0$.

Suppose, on the contrary, that $S = \xi^{-1}\{0\}$ is stationary. Let σ be any strategy of player **II**. Let C_σ be the set of ordinals closed under this strategy. It is a cub set, so there is an $\alpha \in C_\sigma \cap S$. Player **I** can now easily play towards this ordinal to force $\alpha = \sup_{n<\omega} b(n)$ and so $\xi(\sup_{n<\omega} b(b)) = 0$, so σ cannot be a winning strategy. $\square_{\text{item (1)}}$

Proof of item (2). It is not hard to see that $\mathrm{CUB}^\kappa_\lambda$ is Δ^1_1 if and only if there exists a $\kappa\lambda$-canary tree. This fact is proved in detail in **[24]** in the case $\kappa = \omega_1$, $\lambda = \omega$ and the proof generalizes easily to any regular uncountable κ along with the assumption $\kappa^{<\kappa} = \kappa$. So the statement follows from Theorem 3.18. $\square_{\text{item (2)}}$

Proof of item (7). Suppose that φ is Σ_1 and for simplicity assume that φ has no parameters. Then for $x \subset \kappa$ we have:

Claim $\varphi(x)$ holds if and only if the set A of those α for which there exists $\beta > \alpha$ with

$$L_\beta \models \big(\mathrm{ZF}^- \wedge (\omega < \alpha \text{ is regular}) \wedge ((S \cap \alpha) \text{ is stationary }) \wedge \varphi(x \cap \alpha)\big)$$

contains $C \cap S$ for some cub set C.

Proof of the claim. "$\Rightarrow$". If $\varphi(x)$ holds then choose a continuous chain $(M_i \mid i < \kappa)$ of elementary submodels of some large ZF^- model L_θ so that x and S belong to M_0 and the intersection of each M_i with κ is an ordinal α_i less than κ. Let C be the set of α_i's, cub in κ. Then any α in $C \cap S$ belongs to A by condensation.

"$\Leftarrow$". If $\varphi(x)$ fails then let C be any cub in κ and let D be the cub of $\alpha < \kappa$ such that $H(\alpha)$ is the Skolem Hull in some large L_θ of α together with $\{\kappa, S, C\}$ contains no ordinals in the interval $[\alpha, \kappa)$. Let α be the least element of $S \cap \lim(D)$. Then α does not belong to A: If L_β satisfies $\varphi(x \cap \alpha)$ then β must be greater than $\overline{\beta}$ where $\overline{H(\alpha)} = L_{\overline{\beta}}$ is the transitive collapse of $H(\alpha)$, because $\varphi(x \cap \alpha)$ fails in $\overline{H(\alpha)}$. But as $\lim(D) \cap \alpha$ is an element of $L_{\overline{\beta}+2}$ and is disjoint from S, it follows that either α is singular in L_β or $S \cap \alpha$ is not stationary in $L_{\overline{\beta}+2}$ and hence not in L_β. Of course α does belong to C so we have shown that A does not contain $S \cap C$ for an arbitrary cub C in κ. $\square_{\text{Claim}}$

It follows from the above that any Σ_1 subset of 2^κ is Δ_1 over $(L^+_\kappa, \mathrm{CUB}(S))$ and therefore, if $\mathrm{CUB}(S)$ were Δ_1, then every Σ_1 subset of 2^κ would be Δ_1, which is a contradiction. $\square_{\text{item (7)}}$

Proof of item (4). If $X \subset 2^\kappa$ is Δ^1_1, then $\{\eta \in X \mid \eta^{-1}\{1\} \subset \kappa \setminus Z\}$ is Δ^1_1, so it is sufficient to show that we can force a set $E \subset Z$ which has the claimed property. So we force a set $E \subset Z$ such that E is stationary but $E \cap \alpha$ is non-stationary in α for all $\alpha < \kappa$ and $\kappa \setminus E$ is fat. A set is *fat* if its intersection with any cub set contains closed increasing sequences of all order types $< \kappa$.

This can be easily forced with

$$\mathbb{R} = \{p\colon \alpha \to 2 \mid \alpha < \kappa,\ p^{-1}\{1\} \cap \beta \subset Z \text{ is non-stationary in } \beta \text{ for all } \beta \leqslant \alpha\},$$

ordered by end-extension. For any $\mathbb{R}$-generic G, the set $E = (\cup G)^{-1}\{1\}$ satisfies the requirements. Also $\mathbb{R}$ does not add bounded subsets of κ and has the κ^+-c.c. and does not kill stationary sets.

Without loss of generality, assume that such E exists in V and that $0 \in E$.

Next let $\mathbb{P}_0 = \{p\colon \alpha \to 2^{<\alpha} \mid \alpha < \kappa,\ p(\beta) \in 2^\beta,\ p(\beta)^{-1}\{1\} \subset E\}$. This forcing adds a $\diamondsuit_E$-sequence $\langle A_\alpha \mid \alpha \in E\rangle$ (if G is generic, set $A_\alpha = (\cup G)(\alpha)^{-1}\{1\}$) such that for all $B \subset E$ there is a stationary $S \subset E$ such that $A_\alpha = B \cap \alpha$ for all $\alpha \in S$. This forcing $\mathbb{P}_0$ is $< \kappa$-closed and clearly has the κ^+-c.c., so it is easily seen that it does not add bounded subsets of κ and does not kill stationary sets.

Let $\psi(G, \eta, S)$ be a formula with parameters $G \in (2^{<\kappa})^\kappa$ and $\eta \in 2^\kappa$ and a free variable $S \subset \kappa$ which says:

$$\forall \alpha < \kappa (\alpha \in S \iff G(\alpha)^{-1}\{1\} = \eta^{-1}\{1\} \cap \alpha).$$

If $\langle G(\alpha)^{-1}\{1\}\rangle_{\alpha<\kappa}$ happens to be a $\diamondsuit_E$-sequence, then S satisfying ψ is always stationary. Thus if G_0 is $\mathbb{P}_0$-generic over V and $\eta \in 2^E$, then $(\psi(G_0, \eta, S) \to (S \text{ is stationary}))^{V[G_0]}$.

For each $\eta \in 2^E$, let $\dot{S}_\eta$ be a nice $\mathbb{P}_0$-name for the set S such that $V[G_0] \models \psi(G_0, \eta, S)$ where G_0 is $\mathbb{P}_0$-generic over V. By the definitions, $\mathbb{P}_0 \Vdash$ "$\dot{S}_\eta \subset \check{E}$ is stationary" and if $\eta \neq \eta'$, then $\mathbb{P}_0 \Vdash$ "$\dot{S}_\eta \cap \dot{S}_{\eta'}$ is bounded".

Let us enumerate $E = \{\beta_i \mid i < \kappa\}$ such that $i < j \Rightarrow \beta_i < \beta_j$ and for $\eta \in 2^E$ and $\gamma \in \kappa$ define $\eta + \gamma$ to be the $\xi \in 2^E$ such that $\xi(\beta_i) = 1$ for all $i < \gamma$ and $\xi(\beta_{\gamma+j}) = \eta(\beta_j)$ for $j \geqslant 0$. Let

$$F_0 = \{\eta \in 2^E \mid \eta(0) = 0\}^V. \tag{3.2}$$

Now for all $\eta, \eta' \in F_0$ and $\alpha, \alpha' \in \kappa$, $\eta + \alpha = \eta' + \alpha'$ implies $\eta = \eta'$ and $\alpha = \alpha'$. Let us now define the formula $\varphi(G, \eta, X)$ with parameters $G \in (2^{<\kappa})^\kappa$, $\eta \in 2^\kappa$ and a free variable $X \subset \kappa \setminus E$ which says:

$$\begin{aligned}(\eta(0) = 0) \wedge \forall \alpha < \kappa \big[&(\alpha \in X \rightarrow \exists S(\psi(G, \eta + 2\alpha, S) \wedge S \text{ is non-stationary})) \\ &\wedge (\alpha \notin X \rightarrow \exists S(\psi(G, \eta{+}2\alpha{+}1, S) \wedge S \text{ is non-stationary}))\big].\end{aligned}$$

Now, we will construct an iterated forcing $\mathbb{P}_{\kappa^+}$, starting with $\mathbb{P}_0$, which kills the stationarity of $\dot{S}_\eta$ for suitable $\eta \in 2^E$, such that if G is $\mathbb{P}_{\kappa^+}$-generic, then for all $S \subset \kappa \setminus E$, S is stationary if and only if

$$\exists \eta \in 2^E(\varphi(G_0, \eta, S))$$

where $G_0 = G \restriction \{0\}$. In this model, for each $\eta \in F_0$, there will be a unique X such that $\varphi(G_0, \eta, X)$, so let us denote this X by X_η. It is easy to check that the mapping $\eta \mapsto X_\eta$ defined by φ is Σ^1_1 so in the result, also $\mathcal{S} = \{S \subset \kappa \setminus E \mid S \text{ is stationary}\}$ is Σ^1_1. Since cub and non-stationarity are also Σ^1_1, we get that $\mathcal{S}$ is Δ^1_1, as needed.

Let us show how to construct the iterated forcing. For $S \subset \kappa$, we denote by $T(S)$ the partial order of all closed increasing sequences contained in the complement of S. Clearly $T(S)$ is a forcing that kills the stationarity of S. If the complement of S is fat and S is non-reflecting, then $T(S)$ has all the nice properties we need, as the following claims show. Let $f \colon \kappa^+ \setminus \{0\} \rightarrow \kappa^+ \times \kappa^+$ be a bijection such that $f_1(\gamma) \leqslant \gamma$.

$\mathbb{P}_0$ is already defined and it has the κ^+-c.c. and it is $< \kappa$-closed. Suppose that $\mathbb{P}_i$ has been defined for $i < \alpha$ and σ_i has been defined for $i < \cup\alpha$ such that σ_i is a (nice) $\mathbb{P}_i$-name for a κ^+-c.c. partial order. Also suppose that for all $i < \cup\alpha$, $\{(\dot{S}_{ij}, \delta_{ij}) \mid j < \kappa^+\}$ is the list of all pairs $(\dot{S}, \delta)$ such that $\dot{S}$ is a nice $\mathbb{P}_i$-name for a subset of $\check{\kappa} \setminus \check{E}$ and $\delta < \kappa$, and suppose that

$$g_\alpha \colon \{\dot{S}_{f(i)} \mid i < \alpha\} \longrightarrow F_0 \tag{3.3}$$

is an injective function, where F_0 is defined at (3.2).

If α is a limit, let $\mathbb{P}_\alpha$ consist of those $p \colon \alpha \rightarrow \bigcup_{i<\alpha} \operatorname{dom} \sigma_i$ with $|\operatorname{sprt}(p)| < \kappa$ (support; see page 473) such that for all $\gamma < \alpha$, $p \restriction \gamma \in \mathbb{P}_\gamma$ and let $g_\alpha = \bigcup_{i<\alpha} g_i$. Suppose α is a successor, $\alpha = \gamma + 1$. Let $\{(\dot{S}_{\gamma j}, \delta_{\gamma j}) \mid j < \kappa\}$ be the list of pairs as defined above. Let $(\dot{S}, \delta) = (\dot{S}_{f(\gamma)}, \delta_{f(\gamma)})$ where f is the bijection defined above. If there exists $i < \gamma$ such that $\dot{S}_{f(i)} = \dot{S}_{f(\gamma)}$ (i.e., $\dot{S}_i$ has been already under focus), then let $g_\alpha = g_\gamma$. Otherwise let

$$g_\alpha = g_\gamma \cup \{(\dot{S}_{f(\gamma)}, \eta)\},$$

where η is some element in $F_0 \setminus \operatorname{ran} g_\gamma$. Doing this, we want to make sure that in the end $\operatorname{ran} g_{\kappa^+} = F_0$. We omit the technical details needed to ensure that.

Denote $\eta = g(\dot{S}_{f(\gamma)})$. Let σ_γ be a $\mathbb{P}_\gamma$-name such that for all $\mathbb{P}_\gamma$-generic G_γ it holds that

$$\mathbb{P}_\gamma \Vdash \begin{cases} \sigma_\gamma = T(\dot{S}_{\eta+2\delta}), & \text{if } V[G_\gamma] \models [(\delta_{f(\gamma)} \in \dot{S}_{f(\gamma)}) \wedge (\dot{S}_{f(\gamma)} \text{ is stationary})] \\ \sigma_\gamma = T(\dot{S}_{\eta+2\delta+1}), & \text{if } V[G_\gamma] \models [(\delta_{f(\gamma)} \notin \dot{S}_{f(\gamma)}) \wedge (\dot{S}_{f(\gamma)} \text{ is stationary})] \\ \sigma_\gamma = \{\check{\varnothing}\}, & \text{otherwise.} \end{cases}$$

Now let $\mathbb{P}_\alpha$ be the collection of sequences $p = \langle \rho_i \rangle_{i \leqslant \gamma}$ such that $p \restriction \gamma = \langle \rho_i \rangle_{i<\gamma} \in \mathbb{P}_\gamma$, $\rho_\gamma \in \operatorname{dom} \sigma_\gamma$ and $p \restriction \gamma \Vdash_{\mathbb{P}_\gamma} \rho_\gamma \in \sigma_\gamma$ with the ordering defined in the usual way.

Let G be $\mathbb{P}_{\kappa^+}$-generic. Let us now show that the extension $V[G]$ satisfies what we want, namely that $S \subset \kappa \setminus E$ is stationary if and only if there exists $\eta \in 2^E$ such that $S = X_\eta$ (Claims 3 and 4 below).

Claim 1 For $\alpha \leqslant \kappa^+$ the forcing $\mathbb{P}_\alpha$ does not add bounded subsets of κ and the suborder

$$\mathbb{Q}_\alpha = \{p \mid p \in \mathbb{P}_\alpha, p = \langle \check{\rho}_i \rangle_{i<\alpha} \text{ where } \rho_i \in V \text{ for } i < \alpha\}$$

is dense in $\mathbb{P}_\alpha$.

Proof of Claim 1. Let us show this by induction on $\alpha \leqslant \kappa^+$. For $\mathbb{P}_0$ this is already proved and the limit case is left to the reader. Suppose this is proved for all $\gamma < \alpha < \kappa^+$ and $\alpha = \beta + 1$. Then suppose $p \in \mathbb{P}_\alpha$, $p = \langle \rho_i \rangle_{i<\alpha}$. Now $p \restriction \beta \Vdash \rho_\beta \in \sigma_\beta$. Since by the induction hypothesis $\mathbb{P}_\beta$ does not add bounded subsets of κ and $\mathbb{Q}_\beta$ is dense in $\mathbb{P}_\beta$, there exists a condition $r \in \mathbb{Q}_\beta$, $r > p \restriction \beta$ and a standard name $\check{q}$ such that $r \Vdash \check{q} = \rho_\beta$. Now $r^\frown(\check{q})$ is in $\mathbb{Q}_\alpha$, so it is dense in $\mathbb{P}_\alpha$. To show that $\mathbb{P}_\alpha$ does not add bounded sets, it is enough to show that $\mathbb{Q}_\alpha$ does not. Let us think of $\mathbb{Q}_\alpha$ as a suborder of the product $\prod_{i<\alpha} 2^{<\kappa}$. Assume that τ is a $\mathbb{Q}_\alpha$-name and $p \in \mathbb{Q}_\alpha$ forces that $|\tau| = \check{\lambda} < \check{\kappa}$ for some cardinal λ. Then let $\langle M_\delta \rangle_{\delta<\kappa}$ be a sequence of elementary submodels of $H(\kappa^+)$ such that, for all δ, β,

(a) $|M_\delta| < \kappa$;
(b) $\delta < \beta \Rightarrow M_\delta \preceq M_\beta$;
(c) $M_\delta \cap \kappa \subset M_\delta$;
(d) if β is a limit ordinal, then $M_\beta = \bigcup_{\alpha<\beta} M_\alpha$;
(e) if $\kappa = \lambda^+$, then $M_\delta^{<\lambda} \subset M_\delta$ and if κ is inaccessible, then $M_\delta^{|M_\delta|} \subset M_{\delta+1}$;
(f) $M_\alpha \in M_{\alpha+1}$;
(g) $\{p, \kappa, \mathbb{Q}_\alpha, \tau, \check{E}\} \subset M_0$.

This (especially (e)) is possible since κ is not a successor of a singular cardinal and GCH holds. Now the set $C = \{M_\delta \cap \kappa \mid \delta < \kappa\}$ is cub, so because $\kappa \setminus E$ is fat, there is a closed sequence s of length $\lambda+1$ in $C \setminus E$. Let $(\delta_i)_{i \leqslant \lambda}$ be the sequence such that $s = \langle M_{\delta_i} \cap \kappa \rangle_{i\leqslant\lambda}$. For $q \in \mathbb{Q}_\alpha$, let

$$m(q) = \inf_{\gamma \in \operatorname{sprt} q} \operatorname{ran} q(\gamma). \tag{3.4}$$

Let $p_0 = p$ and for all $i < \gamma$ let $p_{i+1} \in M_{\delta_{i+1}} \setminus M_{\delta_i}$ be such that $p_i < p_{i+1}$, p_{i+1} decides $i+1$ first values of τ (think of τ as a name for a function $\lambda \to \kappa$ and that p_i decides the first i values of that function) and $m(p_{i+1}) \geqslant M_{\delta_i} \cap \kappa$. This p_{i+1} can be found because clearly $p_i \in M_{\delta_{i+1}}$ and $M_{\delta_{i+1}}$ is an elementary submodel. If i is a limit, $i < \lambda$, then let p_i be an upper bound of $\{p_j \mid j < i\}$ which can be found in $M_{\delta_{i+1}}$ by the assumptions (f), (e) and (b), and because $M_{\delta_i} \cap \kappa \notin E$. Finally let p_λ be an upper bound of $\langle p_i \rangle_{i<\lambda}$ which exists because for all $\alpha \in \bigcup_{i<\lambda} \operatorname{sprt} p_i$ $\sup_{i<\lambda} \operatorname{ran} p_i(\alpha) = M_{\delta_\lambda} \cap \kappa$ is not in E and the forcing is closed under such sequences. So p_λ decides the whole τ. This completes the proof of the claim. $\square_{\text{Claim 1}}$

So, for simplicity, instead of $\mathbb{P}_{\kappa^+}$ let us work with $\mathbb{Q}_{\kappa^+}$.

Claim 2 Let G be $\mathbb{P}_{\kappa^+}$-generic over V. Suppose $S \subset \kappa$, $S \in V[G]$ and $\dot{S}$ is a nice name for a subset of κ such that $\dot{S}_G = S$. Then let γ be the smallest ordinal with $S \in V[G_\gamma]$. If $(S \subset \kappa \setminus E \text{ is stationary})^{V[G_\gamma]}$, then S is stationary in $V[G]$. If $\dot{S} = \dot{S}_\eta$ for some $\eta \in V$ and $V[G_\gamma] \models \sigma_\gamma \neq T((\dot{S}_\eta)_{G_\gamma \restriction \{0\}})$ for all $\gamma < \kappa^+$, then S is stationary in $V[G]$.

Proof of Claim 2. Recall that σ_γ is as in the construction of $\mathbb{P}_{\kappa^+}$. Suppose first that $S \subset \kappa \setminus E$ is a stationary set in $V[G_\gamma]$ for some $\gamma < \kappa^+$. Let us show that S is stationary in $V[G]$. Note that $V[G] = V[G_\gamma][G^\gamma]$ where $G^\gamma = G \restriction \{\alpha \mid \alpha \geqslant \gamma\}$. Let us show this in the case $\gamma = 0$ and $S \in V$, the other cases being similar. Let $\dot C$ be a name and p a condition which forces that $\dot C$ is cub. Let us show that then $p \Vdash \check S \cap \dot C \neq \check\varnothing$. For $q \in \mathbb{Q}_{\kappa^+}$ let $m(q)$ be defined as in (3.4) above.

Like in the proof of Claim 1, construct a continuous increasing sequence $\langle M_\alpha \rangle_{\alpha<\kappa}$ of elementary submodels of $H(\kappa^{++})$ such that $\{p, \kappa, \mathbb{P}_{\kappa^+}, \check S, \dot C\} \subset M_0$ and $M_\alpha \cap \kappa$ is an ordinal. Since $\{M_\alpha \cap \kappa \mid \alpha < \kappa, M_\alpha \cap \kappa = \alpha\}$ is cub, there exists $\alpha \in S$ such that $M_\alpha \cap \kappa = \alpha$ and because E does not reflect to α there exists a cub sequence

$$c \subset \{M_\beta \cap \kappa \mid \beta < \alpha, M_\beta \cap \kappa = \beta\} \setminus E,$$

$c = \langle c_i \rangle_{i<\mathrm{cf}(\alpha)}$. Now, similarly as in the proof of Claim 1, we can choose an increasing $\langle p_i \rangle_{i\leqslant\mathrm{cf}(\alpha)}$ such that $p_0 = p$, $p_i \in \mathbb{Q}_{\kappa^+}$ for all i, $p_{i+1} \Vdash \check\beta \in \dot C$ for some $c_i \leqslant \beta \leqslant c_{i+1}$, $p_{i+1} \in M_{c_{i+1}} \setminus M_{c_i}$ and $m(p_{i+1}) \geqslant c_i$. If i is a limit, let p_i be again an upper bound of $\{p_j \mid j < i\}$ in M_{c_i}. Since the limits are not in E, the upper bounds exist. Finally $p_{\mathrm{cf}(\alpha)} \Vdash \alpha \in \dot C$, which implies $p_{\mathrm{cf}(\alpha)} \Vdash \check S \cap \dot C \neq \varnothing$, because α was chosen from S.

Assume then that $\dot S = \dot S_\eta$ for some $\eta \in V$ such that

$$V[G_\gamma] \models \sigma_\gamma \neq T((\dot S_\eta)_{G_\gamma\restriction\{0\}})$$

for all $\gamma < \kappa^+$. To prove that $(\dot S_\eta)_G$ is stationary in $V[G]$, we carry the same argument as the above, a little modified. Let us work in $V[G_0]$ and let p_0 force that

$$\forall\gamma < \kappa^+(\sigma_\gamma \neq T(S_\eta)).$$

(This p_0 exists for example because there is at most one γ such that $\sigma_\gamma = T(S_\eta)$.) Build the sequences c, $\langle M_{c_i}\rangle_{i<\mathrm{cf}(\alpha)}$ and $\langle p_i \rangle_{i<\mathrm{cf}(\alpha)}$ in the same fashion as above, except that assume additionally that the functions g_{κ^+} and f, defined along with $\mathbb{P}_{\kappa^+}$, are in M_{c_0}.

At the successor steps one has to choose p_{i+1} such that for each $\gamma \in \operatorname{sprt} p_i$, p_{i+1} decides σ_γ. This is possible, since there are only three choices for σ_γ, namely $\{\varnothing\}$, $T(S_{\xi+2\alpha+1})$ or $T(S_{\xi+2\alpha})$ where ξ and α are justified by the functions g_{κ^+} and f. For all $\gamma \in \operatorname{sprt} p_i$ let us denote by ξ_γ the function such that $p_{i+1} \restriction \gamma \Vdash \sigma_\gamma = T(S_{\xi_\gamma})$. Clearly $\eta \neq \xi_\gamma$ for all $\gamma \in \operatorname{sprt} p_i$. Further demand that $m(p_{i+1}) > \sup(S_\eta \cap S_{\xi_\gamma})$ for all $\gamma \in \operatorname{sprt} p_i$. It is possible to find such p_{i+1} from M_{i+1} because M_{i+1} is an elementary submodel and such can be found in $H(\kappa^{++})$ since $\xi_\gamma \neq \eta$ and, by the definitions given, $S_\eta \cap S_{\xi_\gamma}$ is bounded. $\square_{\text{Claim 2}}$

Claim 3 In $V[G]$ the following holds: if $S \subset \kappa \setminus E$ is stationary, then there exists $\eta \in 2^E$ with $\eta(0) = 0$ such that $S = X_\eta$.

Proof of Claim 3. Recall the function g_{κ^+} from the construction of $\mathbb{P}_{\kappa^+}$ (defined at (3.3) and the paragraph below that). Let $\eta = g_{\kappa^+}(\dot S)$ where $\dot S$ is a nice name $\dot S \in V$ such that $\dot S_G = S$. If $\alpha \in S$, then there is the smallest γ such that $\dot S = S_{f(\gamma)}$ and $\alpha = \delta_{f(\gamma)}$ (where f is as in the definition of $\mathbb{P}_{\kappa^+}$). This stage γ is the only stage where it is possible that $V[G_\gamma] \models \sigma_\gamma = T(S_{\eta+2\alpha+1})$, but since $V[G_\gamma] \models \check\alpha \in \dot S$, by the definition of $\mathbb{P}_{\kappa^+}$ it is not the case, so the stationarity of $S_{\eta+2\alpha+1}$ has not been killed by Claim 2. On the other hand the stationarity of $S_{\eta+2\alpha}$ is killed at this level γ of the construction, so $\alpha \in X_\eta$ by the definitions of φ and X_η. Similarly if $\alpha \notin S$, we conclude that $\alpha \notin X_\eta$. $\square_{\text{Claim 3}}$

Claim 4 In $V[G]$ the following holds: if $S \subset \kappa \setminus E$ is not stationary, then for all $\eta \in 2^E$ with $\eta(0) = 0$ we have $S \neq X_\eta$.

Proof of Claim 4. It is sufficient to show that X_η is stationary for all $\eta \in 2^E$ with $\eta(0) = 0$. Suppose first that $\eta \in F_0 \subset V$. Then since g_{κ^+} is a surjection onto F_0 —see (3.3)— there exists a name $\dot{S}$ such that $S = \dot{S}_G$ is stationary, $S \subset \kappa \setminus E$ and $g_{\kappa^+}(S) = \eta$. Now the same argument as in the proof of Claim 3 implies that $X_\eta = S$, so X_η is stationary by Claim 2.

If $\eta \notin F_0$, then by the definition of $\eta \mapsto X_\eta$ it is sufficient to show that the $\diamondsuit$-sequence added by $\mathbb{P}_0$ guesses in $V[G]$ every new set on a stationary set.

Suppose that τ and $\dot{C}$ are nice $\mathbb{P}_{\kappa^+}$-names for subsets of $\check{\kappa}$ and let p be a condition forcing that $\dot{C}$ is cub. We want to find γ and $q > p$ such that

$$q \Vdash ((\cup\dot{G}_0)(\check{\gamma})^{-1}\{1\} = \tau \cap \check{\gamma}) \wedge (\check{\gamma} \in \dot{C}),$$

where $\dot{G}_0 = \dot{G} \restriction \{0\}$ is the name for the $\mathbb{P}_0$-generic. To do this, let $p_0 \geqslant p$ be such that $p_0 \Vdash \tau \notin \mathcal{P}(\check{\kappa})^V$.

Similarly as in the proofs above define a suitable sequence $\langle M_i \rangle_{i<\lambda}$ of elementary submodels, of length $\lambda < \kappa$, where λ is a cofinality of a point in E, such that $\sup_{i<\lambda}(M_i \cap \kappa) = \alpha \in E$ and $M_i \cap \kappa \notin E$ for all $i < \lambda$. Assume also that $p_0 \in M_0$. Suppose $p_i \in M_i$ is defined. Let $p_{i+1} > p_i$ be an element of $M_{i+1} \setminus M_i$ satisfying the following:

(1) p_{i+1} decides σ_β for all $\beta \in \operatorname{sprt} p_i$;

(2) for all $\beta \in \operatorname{sprt} p_i$ there is $\beta' \in M_{i+1}$ such that $p_{i+1} \Vdash \beta' \in \tau \bigtriangleup \xi_\beta$; where ξ_β is defined as in the proof of Claim 2 and p_{i+1} decides what it is;

(3) p_{i+1} decides τ up to $M_i \cap \kappa$;

(4) $p_{i+1} \Vdash \delta \in \dot{C}$ for some $\delta \in M_{i+1} \setminus M_i$;

(5) $m(p_{i+1}) > M_i \cap \kappa$, where $m(p)$ is defined at (3.4).

Item (1) is possible for the same reason as in the proof of Claim 2, and (2) is possible since $p_i \Vdash \forall \eta \in \mathcal{P}(\check{\kappa})^V (\tau \neq S_{\check{\eta}})$.

Since $M_i \cap \kappa \notin E$ for $i < \lambda$, this ensures that the sequence $p_0 \leqslant p_1 \leqslant \dots$ closes under limits $< \lambda$. Let $p_\lambda = \bigcup_{i<\lambda} p_i$ and let us define $q \supset p_\lambda$ as follows: $\operatorname{sprt} q = \operatorname{sprt} p_\lambda$, for $\delta \in \operatorname{sprt} p_\lambda \setminus \{0\}$ let $\operatorname{dom} q = \alpha + 1$, $p_\lambda(\delta) \subset q(\delta)$, $q(\alpha) = 1$ and $q(0)(\alpha) = \tau \cap \gamma$ (τ means here what have been decided by $\{p_i \mid i < \lambda\}$). Now q is a condition in the forcing notion.

Now certainly, if $q \in G$, then in the extension $\tau_G \cap \alpha = (\cup G_0)(\alpha)^{-1}\{1\}$ and $\alpha \in C$, so we finish. $\square_{\text{Claim 4}}$ $\square_{\text{item (4)}}$

Proof of item (3). If $\kappa = \lambda^+$, this follows from the result of Mekler and Shelah [**23**] and Hyttinen and Rautila [**11**] that the existence of a $\kappa\lambda$-canary tree is consistent. For arbitrary $\lambda < \kappa$ the result follows from item (4) of this theorem proved above (take $Z = \kappa \setminus S^\kappa_\lambda$). $\square_{\text{item (3)}}$

Proof of item (5). For $X = \kappa$ this was proved by Halko and Shelah in [**6**, Theorem 4.2]. For X any stationary subset of κ the proof is similar. It is sufficient to show that $2^\kappa \setminus \operatorname{CUB}(X)$ is not meager in any open set. Suppose U is an open set and $(D_\alpha)_{\alpha<\kappa}$ is a set of dense open sets and let us show that

$$(2^\kappa \setminus \operatorname{CUB}(X)) \cap U \cap \bigcap_{\alpha<\kappa} D_\alpha \neq \varnothing.$$

Let $p \in 2^{<\kappa}$ be such that $N_p \subset U$. Let $p_0 \geqslant p$ be such that $p_0 \in D_0$. Suppose p_β are defined for $\beta < \alpha + 1$. Let $p_{\alpha+1}$ be such that $p_{\alpha+1} \geqslant p_\alpha$, $p_{\alpha+1} \in D_{\alpha+1}$. Suppose p_β is defined for $\beta < \alpha$ and α is a limit ordinal. Let p_α be any element of $2^{<\kappa}$ such that

$p_\alpha > \bigcup_{\beta<\alpha} p_\beta$, $p_\alpha(\sup_{\beta<\alpha} \operatorname{dom} p_\beta) = 0$ and $p_\alpha \in D_\alpha$. Let $\eta = \bigcup_{\alpha<\kappa} p_\alpha$. The complement of $\eta^{-1}\{1\}$ contains a cub, so $X \setminus \eta^{-1}\{1\}$ is stationary whence $\eta \notin \mathrm{CUB}(X)$ and so $\eta \in 2^\kappa \setminus \mathrm{CUB}(X)$. Also clearly $\eta \in U \cap \bigcap_{\alpha<\kappa} D_\alpha$. $\square_{\text{item (5)}}$

Proof of item (6). Our proof is different from that given by Lücke and Schlicht. Suppose $\kappa^{<\kappa} = \kappa > \omega$. We will show that in a generic extension of V all Δ^1_1-sets have the property of Baire. Let

$$\mathbb{P} = \{p \mid p \text{ is a function, } |p| < \kappa, \operatorname{dom} p \subset \kappa \times \kappa^+, \operatorname{ran} p \subset \{0,1\}\}$$

with the ordering $p < q \Leftrightarrow p \subset q$ and let G be $\mathbb{P}$-generic over V. Suppose that $X \subset 2^\kappa$ is a Δ^1_1-set in $V[G]$. It is sufficient to show that for every $r \in 2^{<\kappa}$ there is $q \supset r$ such that either $N_q \setminus X$ or $N_q \cap X$ is co-meager. So let $r \in 2^{<\kappa}$ be arbitrary.

Now suppose that $\langle p_i \rangle_{i<\kappa}$ and $\langle q_i \rangle_{i<\kappa}$ are sequences in $V[G]$ such that $p_i, q_i \in (2^{<\kappa})^2$ for all $i < \kappa$ and X is the projection of

$$C_0 = (2^\kappa)^2 \setminus \bigcup_{i<\kappa} N_{p_i}$$

and $2^\kappa \setminus X$ is the projection of

$$C_1 = (2^\kappa)^2 \setminus \bigcup_{i<\kappa} N_{q_i}.$$

(By N_{p_i} we mean $N_{p_i^1} \times N_{p_i^2}$ where $p_i = (p_i^1, p_i^2)$.) Since these sequences have size κ, there is $\alpha_1 < \kappa^+$ such that they are already in $V[G_{\alpha_1}]$, where $G_{\alpha_1} = \{p \in G \mid \operatorname{dom} p \subset \kappa \times \alpha_1\}$. More generally, for $E \subset \mathbb{P}$ and $A \subset \kappa^+$, we will denote $E_A = \{p \in E \mid \operatorname{dom} p \subset \kappa \times A\}$ and if $p \in \mathbb{P}$, similarly $p_A = p \restriction (\kappa \times A)$.

Let $\alpha_2 \geqslant \alpha_1$ be such that $r \in G_{\{\alpha_2\}}$ (identifying $\kappa \times \{\alpha_2\}$ with κ). This is possible since G is generic. Let $x = G_{\{\alpha_2\}}$. In $V[G]$, $x \in X$ or $x \in 2^\kappa \setminus X$, so there are $\alpha_3 > \alpha_2$, $p \in G_{\alpha_3}$, $p_{\{\alpha_2\}} \supset r$ and a name τ such that p forces that $(x,\tau) \notin N_{p_i}$ for all $i < \kappa$ or $(x,\tau) \notin N_{q_i}$ for all $i < \kappa$. Without loss of generality assume that p forces $(x,\tau) \notin N_{p_i}$ for all $i < \kappa$. Also assume that τ is a $\mathbb{P}_{\alpha_3}$-name and that $\alpha_3 = \alpha_2 + 2$.

By working in $V[G_{\alpha_2}]$ we may assume that $\alpha_2 = 0$. For all $q \in \mathbb{P}_{\{1\}}$, $p_{\{1\}} \subseteq q$ and $i < \kappa$, let $D_{i,q}$ be the set of all $s \in \mathbb{P}_{\{0\}}$ such that $p_{\{0\}} \subseteq s$, $\operatorname{dom}(s) \geqslant \operatorname{dom}(p_i^1)$ and there is $q' \in \mathbb{P}_{\{1\}}$ such that $q \subseteq q'$ and $s \cup q'$ decides $\tau \restriction \operatorname{dom}(p_i^2)$. Clearly each $D_{i,q}$ is dense above $p_{\{0\}}$ in $\mathbb{P}_{\{0\}}$ and so it suffices to show that if $y \in 2^\kappa$ is such that for all $i < \kappa$ and q as above there is $\alpha < \kappa$ such that $y \restriction \alpha \in D_{i,q}$, then $y \in X$. So let y be such. Then we can find $z \in 2^\kappa$ such that for all $i < \kappa$ and q as above there are $\alpha, \beta < \kappa$ such that $\alpha \geqslant \operatorname{dom}(p_i^1)$ and $y \restriction \alpha \cup z \restriction \beta$ decides $t = \tau \restriction \operatorname{dom}(p_i^2)$. By the choice of p, $(y \restriction \operatorname{dom}(p_i^1), t) \neq p_i$. Letting τ^* be the function decided by y and z, $(y, \tau^*) \in C_0$ and so $y \in X$. $\square_{\text{item (6)}}$ $\square_{\text{Theorem 3.20}}$

Remark $(\mathrm{cf}(\kappa) = \kappa > \omega)$ There are some more results and strengthenings of the results in Theorem 3.20:

(1) (Independently known by S. Coskey and P. Schlicht) If $V = L$ then there is a Δ^1_1 well-order of $\mathcal{P}(\kappa)$ and this implies that there is a Δ^1_1-set without the Baire property.

(2) Suppose that $\omega < \kappa < \lambda$, κ regular and λ inaccessible. Then after turning λ into κ^+ by collapsing each ordinal less than λ to κ using conditions of size $< \kappa$, the Baire property holds for Δ^1_1 subsets of κ^κ.

Corollary 3.21 *For a regular $\lambda < \kappa$, let* NS_λ *denote the equivalence relation on* 2^κ *such that* $\eta \mathrm{NS}_\lambda \xi$ *if and only if* $\eta^{-1}\{1\} \triangle \xi^{-1}\{1\}$ *is not* λ*-stationary. Then* NS_λ *is not Borel and it is not* Δ^1_1 *in* L *or in the forcing extensions after adding* κ^+ *Cohen subsets of* κ*.*

Proof. Define a map $f\colon 2^\kappa \to (2^\kappa)^2$ by $\eta \mapsto (\varnothing, \kappa \setminus \eta)$. Suppose for a contradiction that NS_λ is Borel. Then

$$\mathrm{NS}_\varnothing = \mathrm{NS}_\lambda \cap \underbrace{\{(\varnothing, \eta) \mid \eta \in 2^\kappa\}}_{\text{closed}}$$

is Borel, and further $f^{-1}[\mathrm{NS}_\varnothing]$ is Borel by continuity of f. But $f^{-1}[\mathrm{NS}_\varnothing]$ equals CUB which is not Borel by Theorem 3.20 (5) and Theorem 3.16. Similarly, using items (2) and (7) of Theorem 3.20, one can show that NS_λ is not Δ^1_1 under the stated assumptions. □

3.4 Equivalence modulo the non-stationary ideal

In this section we investigate the relations defined as follows:

Definition 3.22 For $X \subset \kappa$, we denote by E_X the relation

$$E_X = \{(\eta, \xi) \in 2^\kappa \times 2^\kappa \mid (\eta^{-1}\{1\} \triangle \xi^{-1}\{1\}) \cap X \text{ is not stationary}\}.$$

The set X consists usually of ordinals of fixed cofinality, i.e., $X \subset S^\kappa_\mu$ for some μ. These relations are easily seen to be Σ^1_1. If $X \subset S^\kappa_\omega$, then it is in fact Borel*. To see this use the same argument as in the proof of Theorem 3.20 (1) that the $\mathrm{CUB}^\kappa_\omega$-set is Borel*.

3.4.1 An antichain

Theorem 3.23 *Assume GCH,* $\kappa^{<\kappa} = \kappa$ *is uncountable, and* $\mu < \kappa$ *is a regular cardinal such that if* $\kappa = \lambda^+$ *then* $\mu \leqslant \mathrm{cf}(\lambda)$*. Then, in a cofinality and GCH preserving forcing extension, there are stationary sets* $K(A) \subset S^\kappa_\mu$ *for each* $A \subset \kappa$ *such that* $E_{K(A)} \not\leqslant_B E_{K(B)}$ *if and only if* $A \not\subset B$*.*

Remark This was improved for Borel equivalence relations in [**20**].

Proof. In this proof we identify functions $\eta \in 2^{\leqslant\kappa}$ with the sets $\eta^{-1}\{1\}$: for example we write $\eta \cap \xi$ to mean $\eta^{-1}\{1\} \cap \xi^{-1}\{1\}$.

The embedding will look as follows. Let $(S_i)_{i<\kappa}$ be pairwise disjoint stationary subsets of

$$\lim S^\kappa_\mu = \{\alpha \in S^\kappa_\mu \mid \alpha \text{ is a limit of ordinals in } S^\kappa_\mu\}.$$

Let

$$K(A) = E_{\bigcup_{\alpha\in A} S_\alpha}. \tag{3.5}$$

If $X_1 \subset X_2 \subset \kappa$, then $E_{X_1} \leqslant_B E_{X_2}$, because $f(\eta) = \eta \cap X_1$ is a reduction. This guarantees that

$$A_1 \subset A_2 \Rightarrow K(A_1) \leqslant_B K(A_2).$$

Now suppose that for all $\alpha < \kappa$ we have killed (by forcing) all reductions from $K(\alpha) = E_{S_\alpha}$ to $K(\kappa \setminus \alpha) = E_{\bigcup_{\beta\neq\alpha} S_\beta}$ for all $\alpha < \kappa$. Then if $K(A_1) \leqslant_B K(A_2)$ it follows that $A_1 \subset A_2$: Otherwise choose $\alpha \in A_1 \setminus A_2$ and we have:

$$K(\alpha) \leqslant_B K(A_1) \leqslant_B K(A_2) \leqslant_B K(\kappa \setminus \alpha),$$

contradiction. So we have:

$$A_1 \subset A_2 \iff K(A_1) \leqslant_B K(A_2).$$

It is easy to obtain an antichain of length κ in $\mathcal{P}(\kappa)$ and so the result follows.

Suppose that $f\colon E_X \leqslant_B E_Y$ is a Borel reduction. Then $g\colon 2^\kappa \to 2^\kappa$ defined by $g(\eta) = f(\eta) \triangle f(0)$ is a Borel function with the following property:

$$\eta \cap X \text{ is stationary } \iff g(\eta) \cap Y \text{ is stationary.}$$

The function g is Borel, so by Lemma 3.2, page 491, there are dense open sets D_i for $i < \kappa$ such that $g \restriction D$ is continuous where $D = \bigcap_{i<\kappa} D_i$. Note that D_i are open so for each i we can write $D_i = \bigcup_{j<\kappa} N_{p(i,j)}$, where $(p(i,j))_{j<\kappa}$ is a suitable collection of elements of $2^{<\kappa}$.

Next define $Q_g\colon 2^{<\kappa} \times 2^{<\kappa} \to \{0,1\}$ by $Q_g(p,q) = 1 \Leftrightarrow N_p \cap D \subset g^{-1}[N_q]$ and $R_g\colon \kappa \times \kappa \to 2^{<\kappa}$ by $R_g(i,j) = p(i,j)$ where $p(i,j)$ are as above.

For any $Q\colon 2^{<\kappa} \times 2^{<\kappa} \to \{0,1\}$ define $Q^*\colon 2^\kappa \to 2^\kappa$ by

$$Q^*(\eta) = \begin{cases} \xi, & \text{such that } \forall \alpha < \kappa\, \exists \beta < \kappa\, Q(\eta \restriction \beta, \xi \restriction \alpha) = 1 \text{ if such exists,} \\ 0, & \text{otherwise.} \end{cases}$$

And for any $R\colon \kappa \times \kappa \to 2^{<\kappa}$ define

$$R^* = \bigcap_{i<\kappa} \bigcup_{j<\kappa} N_{R(i,j)}.$$

Now clearly $R^*_g = D$ and $Q^*_g \restriction D = g \restriction D$, i.e., (Q,D) *codes* $g \restriction D$ in this sense. Thus we have shown that if there is a reduction $E_X \leqslant_B E_Y$, then there is a pair (Q,R) which satisfies the following conditions:

(1) $Q\colon (2^{<\kappa})^2 \to \{0,1\}$ is a function.
(2) $Q(\varnothing, \varnothing) = 1$.
(3) If $Q(p,q) = 1$ and $p' > p$, then $Q(p',q) = 1$.
(4) If $Q(p,q) = 1$ and $q' < q$, then $Q(p,q') = 1$.
(5) Suppose $Q(p,q) = 1$ and $\alpha > \operatorname{dom} q$. There exist $q' > q$ and $p' > p$ such that $\operatorname{dom} q' = \alpha$ and $Q(p',q') = 1$.
(6) If $Q(p,q) = Q(p,q') = 1$, then $q \leqslant q'$ or $q' < q$.
(7) $R\colon \kappa \times \kappa \to 2^{<\kappa}$ is a function.
(8) For each $i \in \kappa$ the set $\bigcup_{j<\kappa} N_{R(i,j)}$ is dense.
(9) For all $\eta \in R^*$, $\eta \cap X$ is stationary if and only if $Q^*(\eta \cap X) \cap Y$ is stationary.

Let us call a pair (Q,R) which satisfies (1)–(9) *a code for a reduction (from E_X to E_Y)*. Note that it is not the same as the Borel code for the graph of a reduction function as a set. Thus we have shown that, if $E_X \leqslant_B E_Y$, then there exists a code for a reduction from E_X to E_Y. We will now prove the following lemma, which is stated in a general enough form so we can use it also in the next section:

Lemma 3.24 (GCH) *Suppose μ_1 and μ_2 are regular cardinals less than κ such that if $\kappa = \lambda^+$ then $\mu_2 \leqslant \operatorname{cf}(\lambda)$, and suppose X is a stationary subset of $S^\kappa_{\mu_1}$, Y is a subset of $S^\kappa_{\mu_2}$, $X \cap Y = \varnothing$ (relevant if $\mu_1 = \mu_2$) and if $\mu_1 < \mu_2$ then $\alpha \cap X$ is not stationary in α for all $\alpha \in Y$. Suppose that (Q,R) is an arbitrary pair. Denote by φ the statement "(Q,R) is not a code for a reduction from E_X to E_Y". Then there is a κ^+-c.c. $<\kappa$-closed forcing $\mathbb{R}$ such that $\mathbb{R} \Vdash \varphi$.*

Remark Clearly if $\mu_1 = \mu_2 = \omega$ then the condition $\mu_2 \leqslant \operatorname{cf}(\lambda)$ is of course true. We need this assumption in order to have $\nu^{<\mu_2} < \kappa$ for all $\nu < \kappa$.

Proof of Lemma 3.24. We will show that one of the following holds:

(1) φ already holds, i.e., $\{\varnothing\} \Vdash \varphi$;

(2) $\mathbb{P} = 2^{<\kappa} = \{p\colon \alpha \to 2 \mid \alpha < \kappa\} \Vdash \varphi$;
(3) $\mathbb{R} \Vdash \varphi$;
where

$$\mathbb{R} = \{(p,q) \mid p,q \in 2^{\alpha},\ \alpha < \kappa,\ X \cap p \cap q = \varnothing,\ q \text{ is } \mu_1\text{-closed}\}.$$

Above "q is μ_1-closed" means "$q^{-1}\{1\}$ is μ_1-closed" etc., and we will use this abbreviation below. Assuming that (1) and (2) do not hold, we will show that (3) holds.

Since (2) does not hold, there is a $p \in \mathbb{P}$ which forces $\neg\varphi$ and so $\mathbb{P}_p = \{q \in \mathbb{P} \mid q > p\} \Vdash \neg\varphi$. But $\mathbb{P}_p \cong \mathbb{P}$, so in fact $\mathbb{P} \Vdash \neg\varphi$, because φ has only standard names as parameters (names for elements in V, such as Q, R, X and Y). Let G be any $\mathbb{P}$-generic and let us denote the set $G^{-1}\{1\}$ also by G. Let us show that $G \cap X$ is stationary. Suppose that $\dot{C}$ is a name and $r \in \mathbb{P}$ is a condition which forces that $\dot{C}$ is cub. For an arbitrary q_0, let us find a $q > q_0$ which forces $\dot{C} \cap \dot{G} \cap \check{X} \neq \varnothing$. Make a counter assumption: no such $q > q_0$ exists. Let $q_1 > q_0$ and $\alpha_1 > \operatorname{dom} q_0$ be such that $q_1 \Vdash \check{\alpha}_1 \in \dot{C}$, $\operatorname{dom} q_1 > \alpha_1$ is a successor and $q_1(\max \operatorname{dom} q_1) = 1$. Then by induction on $i < \kappa$ let q_{i+1} and $\alpha_{i+1} > \operatorname{dom} q_i$ be such that $q_{i+1} \Vdash \check{\alpha}_{i+1} \in \dot{C}$, $\operatorname{dom} q_{i+1} > \alpha_{i+1}$ is a successor and $q_{i+1}(\max \operatorname{dom} q_{i+1}) = 1$. If j is a limit ordinal, let $q_j = \bigcup_{i<j} q_i \cup \{(\sup_{i<j} \operatorname{dom} q_i, 1)\}$ and $\alpha_j = \sup_{i<j} \alpha_i$. We claim that for some $i < \kappa$, the condition q_i is as needed, i.e.,

$$q_i \Vdash \dot{G} \cap \check{X} \cap \dot{C} \neq \varnothing.$$

Clearly, for limit ordinals j, we have $\alpha_j = \max \operatorname{dom} q_j$ and $q_j(\alpha_j) = 1$ and $\{\alpha_j \mid j \text{ limit}\}$ is cub. Since X is stationary, there exists a limit j_0 such that $\alpha_{j_0} \in X$. Because q_0 forces that $\dot{C}$ is cub, $q_j > q_i > q_0$ for all $i < j$, $q_i \Vdash \check{\alpha}_i \in \dot{C}$ and $\alpha_j = \sup_{i<j} \alpha_i$, we have $q_j \Vdash \alpha_j \in \dot{C} \cap \check{X}$. On the other hand $q_j(\alpha_j) = 1$, so $q_j \Vdash \alpha_j \in G$ and we finish.

So now we have in $V[G]$ that $G \cap X$ is stationary, $G \in R^*$ (since R^* is co-meager) and Q is a code for a reduction, so Q^* has the property (9) and $Q^*(G \cap X) \cap Y$ is stationary. Denote $Z = Q^*(G \cap X) \cap Y$. We will now construct a forcing $\mathbb{Q}$ in $V[G]$ such that

$$V[G] \models (\mathbb{Q} \Vdash \text{“}G \cap X \text{ is not stationary, but } Z \text{ is stationary”}).$$

Then $V[G] \models (\mathbb{Q} \Vdash \varphi)$ and hence $\mathbb{P} * \mathbb{Q} \Vdash \varphi$. On the other hand $\mathbb{Q}$ will be chosen such that $\mathbb{P} * \mathbb{Q}$ and $\mathbb{R}$ give the same generic extensions. So let

$$\mathbb{Q} = \{q\colon \alpha \to 2 \mid X \cap G \cap q = \varnothing,\ q \text{ is } \mu_1\text{-closed}\}. \tag{3.6}$$

Clearly $\mathbb{Q}$ kills the stationarity of $G \cap X$. Let us show that it preserves the stationarity of Z. For that purpose it is sufficient to show that for any nice $\mathbb{Q}$-name $\dot{C}$ for a subset of κ and any $p \in \mathbb{Q}$, if $p \Vdash$ " $\dot{C}$ is μ_2-cub", then $p \Vdash (\dot{C} \cap \check{Z} \neq \check{\varnothing})$.

So suppose $\dot{C}$ is a nice name for a subset of κ and $p \in \mathbb{Q}$ is such that

$$p \Vdash \text{“}\dot{C} \text{ is cub”}.$$

Let $\lambda > \kappa$ be a sufficiently large regular cardinal and let N be an elementary submodel of $\langle H(\lambda), p, \dot{C}, \mathbb{Q}, \kappa\rangle$ which has the following properties:

- $|N| = \mu_2$;
- $N^{<\mu_2} \subset N$;
- $\alpha = \sup(N \cap \kappa) \in Z$ (this is possible because Z is stationary).

Here we use the hypothesis that μ_2 is at most $\operatorname{cf}(\lambda)$ when $\kappa = \lambda^+$. Now by the assumption of the theorem, $\alpha \setminus X$ contains a μ_1-closed unbounded sequence of length μ_2, $\langle \alpha_i \rangle_{i<\mu_2}$. Let $\langle D_i \rangle_{i<\mu_2}$ list all the dense subsets of $\mathbb{Q}^N$ in N. Let $q_0 \geqslant p$, $q_0 \in \mathbb{Q}^N$ be arbitrary and suppose $q_i \in \mathbb{Q}^N$ is defined for all $i < \gamma$. If $\gamma = \beta + 1$, then define q_γ to be an extension of q_β such that $q_\gamma \in D_\beta$ and $\operatorname{dom} q_\gamma = \alpha_i$ for some $\alpha_i > \operatorname{dom} q_\beta$. To do that, for instance, choose $\alpha_i > \operatorname{dom} q_\beta$ and define $q' \supset q_\beta$ by $\operatorname{dom} q' = \alpha_i$, $q(\delta) = 0$ for all $\delta \in \operatorname{dom} q' \setminus \operatorname{dom} q_\beta$ and then q' to q_β in D_β. If γ is a limit ordinal with $\operatorname{cf}(\gamma) \neq \mu_1$, then let $q_\gamma = \bigcup_{i<\gamma} q_i$. If $\operatorname{cf}(\gamma) = \mu_1$, let

$$q_\gamma = \Big(\bigcup_{i<\gamma} q_i\Big)^\frown \Big\langle \sup_{i<\gamma} \operatorname{dom} q_i, 1\Big\rangle.$$

Since N is closed under taking sequences of length less than μ_2, $q_\gamma \in N$. Since we required elements of $\mathbb{Q}$ to be μ_1-closed but not γ-closed if $\operatorname{cf}(\gamma) \neq \mu_1$, $q_\gamma \in \mathbb{Q}$ when $\operatorname{cf}(\gamma) \neq \mu_1$. When $\operatorname{cf}(\gamma) = \mu_1$, the limit $\sup_{i<\gamma} \operatorname{dom} q_i$ coincides with a limit of a subsequence of $\langle \alpha_i \rangle_{i<\mu_2}$ of length μ_1, i.e., the limit is α_β for some β since this sequence is μ_1-closed. So by definition $\sup_{i<\gamma} \operatorname{dom} q_i \notin X$ and again $q_\gamma \in \mathbb{Q}$.

Then $q = \bigcup_{\gamma<\mu} q_\gamma$ is a $\mathbb{Q}^N$-generic over N. Since $X \cap Y = \varnothing$, also $(X \cap G) \cap Z = \varnothing$ and $\alpha \notin X \cap G$. Hence $q^\frown(\alpha, 1)$ is in $\mathbb{Q}$. We claim that $q \Vdash (\dot{C} \cap \check{Z} \neq \varnothing)$.

Because $p \Vdash$ "$\dot{C}$ is unbounded", also $N \models (p \Vdash$ "$\dot{C}$ is unbounded") by elementarity. Assuming that λ is chosen large enough, we may conclude that for all $\mathbb{Q}^N$-generic g over N, $N[g] \models$ "$\dot{C}_g$ is unbounded", thus in particular $N[g] \models$ "$\dot{C}_g$ is unbounded in κ". Let G_1 be $\mathbb{Q}$-generic over $V[G]$ with $q \in G_1$. Then $\dot{C}_{G_1} \supset \dot{C}_q$ which is unbounded in α by the above, since $\sup(\kappa \cap N) = \alpha$. Because $\dot{C}_{G_1}$ is μ_2-cub, α is in $\dot{C}_{G_1}$.

Thus $\mathbb{P} * \mathbb{Q} \Vdash \varphi$. It follows straightforwardly from the definition of iterated forcing that $\mathbb{R}$ is isomorphic to a dense suborder of $\mathbb{P} * \dot{\mathbb{Q}}$ where $\dot{Q}$ is a $\mathbb{P}$-name for a partial order such that $\dot{\mathbb{Q}}_G$ equals $\mathbb{Q}$ as defined in (3.6) for any $\mathbb{P}$-generic G.

Now it remains to show that $\mathbb{R}$ has the κ^+-c.c. and is $< \kappa$-closed. Since $\mathbb{R}$ is a suborder of $\mathbb{P} \times \mathbb{P}$, which has size κ, it trivially has the κ^+-c.c. Suppose $(p_i, q_i)_{i<\gamma}$ is an increasing sequence, $\gamma < \kappa$. Then the pair

$$(p, q) = \Big\langle \Big(\bigcup_{i<\gamma} p_i\Big)^\frown \langle \alpha, 0\rangle, \Big(\bigcup_{i<\gamma} q_i\Big)^\frown \langle \alpha, 1\rangle \Big\rangle$$

is an upper bound. $\square_{\text{Lemma 3.24}}$

Note that the forcing used in the previous proof is equivalent to κ-Cohen forcing.

Corollary 3.25 (GCH) *Let $K \colon A \mapsto E_{\bigcup_{\alpha\in A} S_\alpha}$ be as in the beginning of the proof. For each pair (Q, R) and each α there is a $<\kappa$-closed, κ^+-c.c. forcing $\mathbb{R}(Q, R, \alpha)$ such that*

$$\mathbb{R}(Q, R, \alpha) \Vdash \text{“}(Q, R) \textit{ is not a code for a reduction from } K(\{\alpha\}) \textit{ to } K(\kappa \setminus \{\alpha\})\text{”}.$$

Proof. By the above lemma one of the choices $\mathbb{R} = \{\varnothing\}$, $\mathbb{R} = 2^{<\kappa}$ or

$$\mathbb{R} = \{(p, q) \mid p, q \in 2^\beta,\ \beta < \kappa,\ S_\alpha \cap p \cap q = \varnothing,\ q \text{ is } \mu\text{-closed}\}$$

suffices. □

Start with a model satisfying GCH. Let $h \colon \kappa^+ \to \kappa^+ \times \kappa \times \kappa^+$ be a bijection such that $h_3(\alpha) < \alpha$ for $\alpha > 0$ and $h_3(0) = 0$. Let $\mathbb{P}_0 = \{\varnothing\}$. For each $\alpha < \kappa$, let $\{\sigma_{\beta\alpha 0} \mid \beta < \kappa^+\}$ be the list of all $\mathbb{P}_0$-names for codes for a reduction from $K(\{\alpha\})$ to $K(\kappa \setminus \{\alpha\})$. Suppose $\mathbb{P}_i$ and $\{\sigma_{\beta\alpha i} \mid \beta < \kappa^+\}$ are defined for all $i < \gamma$ and $\alpha < \kappa$, where $\gamma < \kappa^+$ is a successor $\gamma = \beta + 1$, $\mathbb{P}_i$ is $< \kappa$-closed and has the κ^+-c.c.

Consider $\sigma_{h(\beta)}$. By the above corollary, the following holds:

$$\mathbb{P}_\beta \Vdash \big[\exists \mathbb{R} \in \mathcal{P}(2^{<\kappa} \times 2^{<\kappa})(\mathbb{R} \text{ is } < \kappa\text{-closed, } \kappa^+\text{-c.c. p.o. and}$$
$$\mathbb{R} \Vdash \text{“ } \sigma_{h(\beta)} \text{ is not a code for a reduction”})\big].$$

So there is a $\mathbb{P}_\beta$-name ρ_β such that $\mathbb{P}_\beta$ forces that ρ_β is as $\mathbb{R}$ above. Define

$$\mathbb{P}_\gamma = \{(p_i)_{i<\gamma} \mid ((p_i)_{i<\beta} \in \mathbb{P}_\beta) \wedge ((p_i)_{i<\beta} \Vdash p_\beta \in \rho_\beta)\}.$$

And if $p = (p_i)_{i<\gamma} \in \mathbb{P}_\gamma$ and $p' = (p'_i)_{i<\gamma} \in \mathbb{P}_\gamma$, then

$$p \leqslant_{\mathbb{P}_\gamma} p' \iff [(p_i)_{i<\beta} \leqslant_{\mathbb{P}_\beta} (p'_i)_{i<\beta}] \wedge [(p'_i)_{i<\beta} \Vdash (p_\beta \leqslant_{\rho_\beta} p'_\beta)].$$

If γ is a limit, $\gamma \leqslant \kappa^+$, let

$$\mathbb{P}_\gamma = \{(p_i)_{i<\gamma} \mid \forall \beta(\beta < \gamma \to (p_i)_{i<\beta} \in \mathbb{P}_\beta) \wedge (|\operatorname{sprt}(p_i)_{i<\gamma}| < \kappa)\},$$

where sprt means support; see page 473. For every α, let $\{\sigma_{\beta\alpha\gamma} \mid \beta < \kappa^+\}$ list all $\mathbb{P}_\beta$-names for codes for a reduction. It is easily seen that $\mathbb{P}_\gamma$ is $< \kappa$-closed and has the κ^+-c.c. for all $\gamma \leqslant \kappa^+$

We claim that $\mathbb{P}_{\kappa^+}$ forces that, for all α, $K(\{\alpha\}) \nleqslant_B K(\kappa \setminus \{\alpha\})$, which suffices by the discussion in the beginning of the proof; see (3.5) for the notation.

Let G be $\mathbb{P}_{\kappa^+}$-generic and let $G_\gamma =$ “ $G \cap \mathbb{P}_\gamma$” for every $\gamma < \kappa$. Then G_γ is $\mathbb{P}_\gamma$-generic.

Suppose that, in $V[G]$, $f \colon 2^\kappa \to 2^\kappa$ is a reduction $K(\{\alpha\}) \leqslant_B K(\kappa \setminus \{\alpha\})$ and (Q, R) is the corresponding code for a reduction. By [**21**, Theorem VIII.5.14], there is a $\delta < \kappa^+$ such that $(Q, R) \in V[G_\delta]$. Let δ_0 be the smallest such δ.

Now there exists $\sigma_{\gamma\alpha\delta_0}$, a $\mathbb{P}_{\delta_0}$-name for (Q, R). By the definition of h, there exists a $\delta > \delta_0$ with $h(\delta) = (\gamma, \alpha, \delta_0)$. Thus

$$\mathbb{P}_{\delta+1} \Vdash \text{“}\sigma_{\gamma\alpha\delta_0} \text{ is not a code for a reduction”},$$

i.e., $V[G_{\delta+1}] \models (Q, R)$ is not a code for a reduction. Now one of the items (1)–(9) fails for (Q, R) in $V[G_{\delta+1}]$. We want to show that then one of them fails in $V[G]$. The conditions (1)–(8) are absolute, so if one of them fails in $V[G_{\delta+1}]$, then we are done. Suppose (1)–(8) hold but (9) fails. Then there is an $\eta \in R^*$ such that $Q^*(\eta \cap S_{\{\alpha\}}) \cap S_{\kappa\setminus\alpha}$ is stationary but $\eta \cap S_{\{\alpha\}}$ is not or vice versa. In $V[G_{\delta+1}]$ define

$$\mathbb{P}^{\delta+1} = \{(p_i)_{i<\kappa^+} \in \mathbb{P}_{\kappa^+} \mid (p_i)_{i<\delta+1} \in G_{\delta+1}\}.$$

Then $\mathbb{P}^{\delta+1}$ is $< \kappa$-closed. Thus it does not kill stationarity of any set. Hence, if $G^{\delta+1}$ is $\mathbb{P}_{\delta+1}$-generic over $V[G_{\delta+1}]$, then in $V[G_{\delta+1}][G^{\delta+1}]$, (Q, R) is not a code for a reduction. Now it remains to show that $V[G] = V[G_{\delta+1}][G^{\delta+1}]$ for some $G^{\delta+1}$. In fact putting $G^{\delta+1} = G$ we get $\mathbb{P}^{\delta+1}$-generic over $V[G_{\delta+1}]$ and of course $V[G_{\delta+1}][G] = V[G]$ (since $G_{\delta+1} \subset G$). $\square_{\text{Theorem 3.23}}$

Remark The forcing constructed in the proof of Theorem 3.23 above, combined with the forcing in the proof of item (4) of Theorem 3.20, page 502, gives that for $\kappa^{<\kappa} = \kappa > \omega_1$ not successor of a singular cardinal, we have in a forcing extension that $\langle \mathcal{P}(\kappa), \subset \rangle$ embeds into $\langle \mathrm{E}^{\Delta^1_1}, \leqslant_B \rangle$, i.e., the partial order of Δ^1_1-equivalence relations under Borel reducibility.

3.4.2 Reducibility between different cofinalities

Recall the notation defined in Section 1.1. In this section we will prove the following two theorems:

Theorem 3.26 *Suppose that κ is a weakly compact cardinal and that $V = L$. Then:*

(A) $E_{S^\kappa_\lambda} \leqslant_c E_{\mathrm{reg}(\kappa)}$ *for any regular $\lambda < \kappa$, where* $\mathrm{reg}(\kappa) = \{\lambda < \kappa \mid \lambda$ *is regular*$\}$.

(B) *In a forcing extension $E_{S^{\omega_2}_\omega} \leqslant_c E_{S^{\omega_2}_{\omega_1}}$. Similarly for λ, λ^+ and λ^{++} instead of ω, ω_1 and ω_2 for any regular $\lambda < \kappa$.*

Theorem 3.27 *For a cardinal κ which is a successor of a regular cardinal or κ inaccessible, there is a cofinality-preserving forcing extension in which, for all regular $\lambda < \kappa$, the relations $E_{S^\kappa_\lambda}$ are $\leqslant_B$-incomparable with each other.*

Let us begin by proving the latter.

Proof of Theorem 3.27. Let us show that there is a forcing extension of L in which $E_{S^{\omega_2}_{\omega_1}}$ and $E_{S^{\omega_2}_\omega}$ are incomparable. The general case is similar.

We shall use Lemma 3.24 with $\mu_1 = \omega$ and $\mu_2 = \omega_1$ and vice versa, and then a similar iteration as in the end of the proof of Theorem 3.23. First we force, like in the proof of Theorem 3.20 (4), a stationary set $S \subset S^{\omega_2}_\omega$ such that, for all $\alpha \in S^{\omega_2}_{\omega_1}$, $\alpha \cap S$ is non-stationary in α. Also for all $\alpha \in S^{\omega_2}_\omega$, $\alpha \cap S^{\omega_2}_{\omega_1}$ is non-stationary.

By Lemma 3.24, for each code for a reduction from E_S to $E_{S^{\omega_2}_{\omega_1}}$ there is a $< \omega_2$-closed ω_3-c.c. forcing which kills it. Similarly for each code for a reduction from $E_{S^{\omega_2}_{\omega_1}}$ to $E_{S^{\omega_2}_\omega}$. Making an ω_3-long iteration, similarly as in the end of the proof of Theorem 3.23, we can kill all codes for reductions from E_S to $E_{S^{\omega_2}_{\omega_1}}$ and from $E_{S^{\omega_2}_{\omega_1}}$ to $E_{S^{\omega_2}_\omega}$. Thus, in the extension there are no reductions from $E_{S^{\omega_2}_{\omega_1}}$ to $E_{S^{\omega_2}_\omega}$ and no reductions from $E_{S^{\omega_2}_\omega}$ to $E_{S^{\omega_2}_{\omega_1}}$. (Suppose there is one of a latter kind, $f\colon 2^{\omega_2} \to 2^{\omega_2}$. Then $g(\eta) = f(\eta \cap S)$ is a reduction from E_S to $E_{S^{\omega_2}_{\omega_1}}$.) □ Theorem 3.27

Definition 3.28 Let X, Y be subsets of κ and suppose that Y consists of ordinals of uncountable cofinality. We say that X *$\diamondsuit$-reflects to Y* if there exists a sequence $\langle D_\alpha \rangle_{\alpha \in Y}$ such that

(1) $D_\alpha \subset \alpha$ is stationary in α;

(2) if $Z \subset X$ is stationary, then $\{\alpha \in Y \mid D_\alpha = Z \cap \alpha\}$ is stationary.

Theorem 3.29 *If X $\diamondsuit$-reflects to Y, then $E_X \leqslant_c E_Y$.*

Proof. Let $\langle D_\alpha \rangle_{\alpha \in Y}$ be the sequence of Definition 3.28. For a set $A \subset \kappa$ define

$$f(A) = \{\alpha \in Y \mid A \cap X \cap D_\alpha \text{ is stationary in } \alpha\}. \qquad (i)$$

We claim that f is a continuous reduction. Clearly f is continuous. Assume that $(A \triangle B) \cap X$ is non-stationary. Then there is a cub set $C \subset \kappa \setminus [(A \triangle B) \cap X]$. Now

$$A \cap X \cap C = B \cap X \cap C. \qquad (ii)$$

The set $C' = \{\alpha < \kappa \mid C \cap \alpha$ is unbounded in $\alpha\}$ is also cub and if $\alpha \in Y \cap C'$, we have that $D_\alpha \cap C$ is stationary in α (*iii*). Therefore, for $\alpha \in Y \cap C'$, we have the following equivalences:

$$\begin{aligned}\alpha \in f(A) &\iff A \cap X \cap D_\alpha \text{ is stationary}\\ &\overset{(iii)}{\iff} A \cap X \cap C \cap D_\alpha \text{ is stationary}\\ &\overset{(ii)}{\iff} B \cap X \cap C \cap D_\alpha \text{ is stationary}\\ &\overset{(iii)}{\iff} B \cap X \cap D_\alpha \text{ is stationary}\\ &\overset{(i)}{\iff} \alpha \in f(B).\end{aligned}$$

Thus $(f(A) \triangle f(B)) \cap Y \subset \kappa \setminus C'$ and it is non-stationary.

Suppose $A \triangle B$ is stationary. Then either $A \setminus B$ or $B \setminus A$ is stationary. Without loss of generality suppose the former. Then

$$S = \{\alpha \in Y \mid (A \setminus B) \cap X \cap \alpha = D_\alpha\}$$

is stationary by the definition of the sequence $\langle D_\alpha \rangle_{\alpha \in Y}$. Thus for $\alpha \in S$ we have that $A \cap X \cap D_\alpha = A \cap X \cap (A \setminus B) \cap X \cap \alpha = (A \setminus B) \cap X \cap \alpha$ is stationary in α and $B \cap X \cap D_\alpha = B \cap X \cap (A \setminus B) \cap X \cap \alpha = \varnothing$ is not stationary in α. Therefore $(f(A) \triangle f(B)) \cap Y$ is stationary (as it contains S). □

Fact (Π^1_1-reflection) Assume that κ is weakly compact. If R is any binary predicate on V_κ and $\forall A\varphi$ is some Π^1_1-sentence where φ is a first-order sentence in the language of set theory together with predicates $\{R, A\}$ such that $(V_\kappa, R) \models \forall A\varphi$, then there exists stationary many $\alpha < \kappa$ such that $(V_\alpha, R \cap V_\alpha) \models \forall A\varphi$.

We say that X *strongly reflects to* Y if for all stationary $Z \subset X$ there exist stationary many $\alpha \in Y$ with $X \cap \alpha$ stationary in α.

Theorem 3.30 *Suppose $V = L$, κ is weakly compact and that $X \subset \kappa$ and $Y \subset \operatorname{reg}\kappa$. If X strongly reflects to Y, then X $\diamondsuit$-reflects to Y.*

Proof. Define D_α by induction on $\alpha \in Y$. For the purpose of the proof, define also C_α for each α as follows. Suppose (D_β, C_β) is defined for all $\beta < \alpha$. Let (D, C) be the L-least[1] pair such that

(1) C is cub subset of α;
(2) D is a stationary subset of $X \cap \alpha$;
(3) for all $\beta \in Y \cap C$, $D \cap \beta \neq D_\beta$.

If there is no such pair then set $D = C = \varnothing$. Then let $D_\alpha = D$ and $C_\alpha = C$. We claim that the sequence $\langle D_\alpha \rangle_{\alpha \in Y}$ is as needed. To show this, let us make a counter assumption: there is a stationary subset Z of X and a cub subset C of κ such that

$$C \cap Y \subset \{\alpha \in Y \mid D_\alpha \neq Z \cap \alpha\}. \tag{3.7}$$

Let (Z, C) be the L-least such pair. Let $\lambda > \kappa$ be regular and let M be an elementary submodel of L_λ such that

(1) $|M| < \kappa$;
(2) $\alpha = M \cap \kappa \in Y \cap C$;
(3) $Z \cap \alpha$ is stationary in α;
(4) $\{Z, C, X, Y, \kappa\} \subset M$.

[1] The least in the canonical definable ordering on L; see **[21]**.

Here (2) and (3) are possible by the definition of strong reflection. Let $\overline{M}$ be the Mostowski collapse of M and let $G\colon M \to \overline{M}$ be the Mostowski isomorphism. Then $\overline{M} = L_\gamma$ for some $\gamma > \alpha$. Since $\kappa \cap M = \alpha$, we have

$$(3.8)\quad G(Z) = Z \cap \alpha,\ G(C) = C \cap \alpha,\ G(X) = X \cap \alpha,\ G(Y) = Y \cap \alpha, \text{ and } G(\kappa) = \alpha.$$

Note that, by the definability of the canonical ordering of L, the sequence $\langle D_\beta \rangle_{\beta<\kappa}$ is definable. Let $\varphi(x, y, \alpha)$ be the formula which says

"(x, y) is the L-least pair such that x is contained in $X \cap \alpha$, x is stationary in α, y is cub in α and $x \cap \beta \neq D_\beta$ for all $\beta \in y \cap Y \cap \alpha$".

By the assumption,

$$L \models \varphi(Z, C, \kappa), \text{ so } M \models \varphi(Z, C, \kappa) \text{ and } L_\gamma \models \varphi(G(Z), G(C), G(\kappa)).$$

Let us show that this implies $L \models \varphi(G(Z), G(C), G(\kappa))$, i.e., $L \models \varphi(Z \cap \alpha, C \cap \alpha, \alpha)$. This will be a contradiction because then $D_\alpha = Z \cap \alpha$, which contradicts the assumptions (2) and (3.7) above.

By the relative absoluteness of being the L-least, the relativised formula with parameters $\varphi^{L_\gamma}(G(Z), G(C), G(\kappa))$ says

"$(G(Z), G(C))$ is the L-least pair such that $G(Z)$ is contained in $G(X)$, $G(Z)$ is $(\text{stationary})^{L_\gamma}$ in $G(\kappa)$, $G(C)$ is cub in $G(\kappa)$ and $G(Z) \cap \beta \neq D_\beta^{L_\gamma}$ for all $\beta \in G(C) \cap G(Y) \cap G(\kappa)$".

Written out, this is equivalent to

"$(Z \cap \alpha, C \cap \alpha)$ is the L-least pair such that $Z \cap \alpha$ is contained in $X \cap \alpha$, $Z \cap \alpha$ is $(\text{stationary})^{L_\gamma}$ in α, $C \cap \alpha$ is cub in α and $Z \cap \beta \neq D_\beta^{L_\gamma}$ for all $\beta \in C \cap Y \cap \alpha$".

Note that this is true in L. Since $Z \cap \alpha$ is stationary in α also in L by (3), it remains to show by induction on $\beta \in \alpha \cap Y$ that $Z \cap \alpha$ $D_\beta^{L_\gamma} = D_\beta^L$ and $C_\beta^{L_\gamma} = C_\beta^L$ and we are done. Suppose we have proved this for $\delta \in \beta \cap Y$ and $\beta \in \alpha \cap Y$. Then $(D_\beta^{L_\gamma}, C_\beta^{L_\gamma})$ is

(a) $(\text{the least } L\text{-pair})^{L_\gamma}$ such that
(b) $(C_\beta \text{ is a cub subset of } \beta)^{L_\gamma}$,
(c) $(D_\beta \text{ is a stationary subset of } \beta)^{L_\gamma}$,
(d) and for all $\delta \in Y \cap \beta$, $(D_\beta \cap \delta \neq D_\delta)^{L_\gamma}$,
(e) or there is no such pair and $D_\beta = \varnothing$.

The L-order is absolute as explained above, so (a) is equivalent to $(\text{the least } L\text{-pair})^L$. Being a cub subset of α is also absolute for L_γ so (b) is equivalent to $(C_\beta \text{ is a cub subset of } \alpha)^L$. All subsets of β in L are elements of $L_{|\beta|^+}$ (see [**21**]), and since α is regular and $\beta < \alpha \leqslant \gamma$, we have $\mathcal{P}(\beta) \subset L_\gamma$. Thus

$$(D_\beta \text{ is stationary subset of } \beta)^{L_\gamma} \iff (D_\beta \text{ is stationary subset of } \beta)^L.$$

Finally, the statement of (d), $(D_\beta \cap \delta \neq D_\delta)^{L_\gamma}$, is equivalent to $D_\beta \cap \delta \neq D_\delta^{L_\gamma}$ as it is defining D_β, but by the induction hypothesis $D_\delta^{L_\gamma} = D_\delta^L$, so we are done. For (e), the fact that

$$\mathcal{P}(\beta) \subset L_{|\beta|^+} \subset L_\alpha \subset L_\gamma$$

as above implies that if there is no such pair in L_γ, then there is no such pair in L. □

Proof of Theorem 3.26. In the case (A) we will show that S^κ_λ strongly reflects to $\mathrm{reg}(\kappa)$ in L, which suffices by Theorems 3.29 and 3.30. For (B) we will assume that κ is a weakly compact cardinal in L and then collapse it to ω_2 to get a $\diamondsuit$-sequence which witnesses that $S^{\omega_2}_\omega$ $\diamondsuit$-reflects to $S^{\omega_2}_{\omega_1}$ which is sufficient by Theorem 3.29. In the following we assume that $V = L$ and κ is weakly compact.

(A): Let us use Π^1_1-reflection. Let $X \subset S^\kappa_\lambda$. We want to show that the set

$$\{\lambda \in \mathrm{reg}(\kappa) \mid X \cap \lambda \text{ is stationary in } \lambda\}$$

is stationary. Let $C \subset \kappa$ be cub. The sentence

$$\text{“}(X \text{ is stationary in } \kappa) \wedge (C \text{ is cub in } \kappa) \wedge (\kappa \text{ is regular})\text{”}$$

is a Π^1_1-property of (V_κ, X, C). By Π^1_1-reflection we get $\delta < \kappa$ such that $(V_\delta, X \cap \delta, C \cap \delta)$ satisfies it. But then δ is regular, $X \cap \delta$ is stationary and δ belongs to C.

(B): Let κ be weakly compact and let us Lévy-collapse κ to ω_2 with the following forcing:

$$\mathbb{P} = \{f\colon \mathrm{reg}\,\kappa \to \kappa^{<\omega_1} \mid \mathrm{ran}(f(\mu)) \subset \mu,\ |\{\mu \mid f(\mu) \neq \varnothing\}| \leqslant \omega\}.$$

Order $\mathbb{P}$ by $f < g$ if and only if $f(\mu) \subset g(\mu)$ for all $\mu \in \mathrm{reg}(\kappa)$. For all μ, put $\mathbb{P}_\mu = \{f \in \mathbb{P} \mid \mathrm{sprt}\, f \subset \mu\}$ and $\mathbb{P}^\mu = \{f \in \mathbb{P} \mid \mathrm{sprt}\, f \subset \kappa \setminus \mu\}$, where sprt means support.

Claim 1 For all regular μ, $\omega < \mu \leqslant \kappa$, $\mathbb{P}_\mu$ satisfies the following:

(a) If $\mu > \omega_1$, then $\mathbb{P}_\mu$ has the μ-c.c.
(b) $\mathbb{P}_\mu$ and $\mathbb{P}^\mu$ are $< \omega_1$-closed.
(c) $\mathbb{P} = \mathbb{P}_\kappa \Vdash \omega_2 = \check{\kappa}$.
(d) If $\mu < \kappa$, then $\mathbb{P} \Vdash \mathrm{cf}(\check{\mu}) = \omega_1$.
(e) If $p \in \mathbb{P}$, σ a name and $p \Vdash$ “σ is cub in ω_2”, then there is a cub $E \subset \kappa$ such that $p \Vdash \check{E} \subset \sigma$.

Proof. Standard (see for instance [**16**]). □

We want to show that in the generic extension $S^{\omega_2}_\omega$ $\diamondsuit$-reflects to $S^{\omega_2}_{\omega_1}$. It is sufficient to show that $S^{\omega_2}_\omega$ $\diamondsuit$-reflects to some stationary $Y \subset S^{\omega_2}_{\omega_1}$ by letting $D_\alpha = \alpha$ for $\alpha \notin Y$. In our case $Y = \{\mu \in V[G] \mid (\mu \in \mathrm{reg}(\kappa))^V\}$. By (d) of Claim 1, $Y \subset S^{\omega_2}_{\omega_1}$, $(\mathrm{reg}(\kappa))^V$ is stationary in V (for instance by Π^1_1-reflection) and by (e) it remains stationary in $V[G]$.

It is easy to see that $\mathbb{P} \cong \mathbb{P}_\mu \times \mathbb{P}^\mu$. Let G be a $\mathbb{P}$-generic over (the ground model) V. Define

$$G_\mu = G \cap \mathbb{P}_\mu$$

and

$$G^\mu = G \cap \mathbb{P}^\mu.$$

Then G_μ is $\mathbb{P}_\mu$-generic over V. Also G^μ is $\mathbb{P}^\mu$-generic over $V[G_\mu]$ and $V[G] = V[G_\mu][G^\mu]$.

Let

$$E = \{p \in \mathbb{P} \mid (p > q) \wedge (p_\mu \Vdash p^\mu \in \dot{D})\}.$$

Then E is dense above q: if $p > q$ is an arbitrary element of $\mathbb{P}$, then $q \Vdash \exists p' > \check{p}^\mu (p' \in \dot{D})$. Thus there exists $q' > q$ with $q' > p_\mu$, $q' \in \mathbb{P}_\mu$ and $p' > p$, $p' \in \mathbb{P}^\mu$ such that $q' \Vdash p' \in \dot{D}$ and so $(q' \restriction \mu) \cup (p' \restriction (\kappa \setminus \mu))$ is above p and in E. So there is a $p \in G \cap E$. But then $p_\mu \in G_\mu$ and $p^\mu \in G^\mu$ and $p_\mu \Vdash p^\mu \in \dot{D}$, so $G^\mu \cap D \neq \varnothing$. Since D was arbitrary, this shows that G^μ is $\mathbb{P}^\mu$-generic over $V[G_\mu]$. Clearly $V[G]$ contains both G_μ and G^μ. On the other hand, $G = G_\mu \cup G^\mu$, so $G \in V[G_\mu][G^\mu]$. By the minimality of forcing extensions, we get $V[G] = V[G_\mu][G^\mu]$.

For each $\mu \in \operatorname{reg}(\kappa) \setminus \{\omega, \omega_1\}$, let

$$k_\mu \colon \mu^+ \longrightarrow \{\sigma \mid \sigma \text{ is a nice } \mathbb{P}_\mu \text{ name for a subset of } \mu\}$$

be a bijection. A nice $\mathbb{P}_\mu$ name for a subset of $\check{\mu}$ is of the form

$$\bigcup \{\{\check{\alpha}\} \times A_\alpha \mid \alpha \in B\},$$

where $B \subset \check{\mu}$ and, for each $\alpha \in B$, A_α is an antichain in $\mathbb{P}_\mu$. By (a) there are no antichains of length μ in $\mathbb{P}_\mu$ and $|\mathbb{P}_\mu| = \mu$, so there are at most $\mu^{<\mu} = \mu$ antichains and there are μ^+ subsets $B \subset \mu$, so there indeed exists such a bijection k_μ (these cardinality facts hold because $V = L$ and μ is regular). Note that if σ is a nice $\mathbb{P}_\mu$-name for a subset of $\check{\mu}$, then $\sigma \subset V_\mu$.

Let us define

$$D_\mu = \begin{cases} \Big[k_\mu\Big([(\cup G)(\mu^+)](0)\Big)\Big]_G & \text{if it is stationary,} \\ \mu & \text{otherwise.} \end{cases}$$

Now D_μ is defined for all $\mu \in Y$; recall that $Y = \{\mu \in V[G] \mid (\mu \in \operatorname{reg}\kappa)^V\}$. We claim that $\langle D_\mu \rangle_{\mu \in Y}$ is the needed $\diamondsuit$-sequence. Suppose it is not. Then there is a stationary set $S \subset S^{\omega_2}_\omega$ and a cub $C \subset \omega_2$ such that for all $\alpha \in C \cap Y$, $D_\alpha \neq S \cap \alpha$. By (e) there is a cub set $C_0 \subset C$ such that $C_0 \in V$. Let $\dot{S}$ be a nice name for S and p' such that p' forces that $\dot{S}$ is stationary. Let us show that

$$H = \{q \geqslant p' \mid q \Vdash D_\mu = \dot{S} \cap \check{\mu} \text{ for some } \mu \in C_0\}$$

is dense above p', which is obviously a contradiction. For that purpose, let $p > p'$ be arbitrary and let us show that there is $q > p$ in H. Let us now use Π^1_1-reflection. First let us redefine $\mathbb{P}$. Let $\mathbb{P}^* = \{q \mid \exists r \in \mathbb{P}(r \restriction \operatorname{sprt} r = q)\}$. Clearly $\mathbb{P}^* \cong \mathbb{P}$ but the advantage is that $\mathbb{P}^* \subset V_\kappa$ and $\mathbb{P}^*_\mu = \mathbb{P}^* \cap V_\mu$ where $\mathbb{P}^*_\mu$ is defined as $\mathbb{P}_\mu$. One easily verifies that all the above things (concerning $\mathbb{P}_\mu$, $\mathbb{P}^\mu$, etc.) translate between $\mathbb{P}$ and $\mathbb{P}^*$. From now on denote $\mathbb{P}^*$ by $\mathbb{P}$. Let

$$R = (\mathbb{P} \times \{0\}) \cup (\dot{S} \times \{1\}) \cup (C_0 \times \{2\}) \cup (\{p\} \times \{3\}).$$

Then $(V_\kappa, R) \models \forall A \varphi$, where φ says: "(if A is closed unbounded and $r > p$ arbitrary, then there exist $q > r$ and α such that $\alpha \in A$ and $q \Vdash_{\mathbb{P}} \check{\alpha} \in \dot{S}$)". So basically $\forall A \varphi$ says "$p \Vdash$ ($\dot{S}$ is stationary)". It follows from (e) that it is enough to quantify over cub sets in V. Let us explain why such a formula can be written for (V_κ, R). The sets (classes from the viewpoint of V_κ) $\mathbb{P}$, $\dot{S}$ and C_0 are coded into R, so we can use them as parameters. That $r > p$ and $q > r$ and A is closed and unbounded is expressible in first-order as well as $\alpha \in A$. How do we express $q \Vdash_{\mathbb{P}} \check{\alpha} \in \dot{S}$? The definition of $\check{\alpha}$ is recursive in α:

$$\check{\alpha} = \{(\check{\beta}, 1_{\mathbb{P}}) \mid \beta < \alpha\}$$

and is absolute for V_κ. Then $q \Vdash_{\mathbb{P}} \check{\alpha} \in \dot{S}$ is equivalent to saying that for each $q' > q$ there exists $q'' > q'$ with $(\check{\alpha}, q'') \in \dot{S}$ and this is expressible in first-order (as we have taken R as a parameter).

By Π^1_1-reflection there is $\mu \in C_0$ such that $p \in \mathbb{P}_\mu$ and $(V_\mu, R) \models \forall A \varphi$. Note that we may require that μ is regular, i.e., $(\check{\mu}_G \in Y)^{V[G]}$ and such that $\alpha \in S \cap \mu$ implies $(\check{\alpha}, \check{p}) \in \dot{S}$ for some $p \in \mathbb{P}_\mu$. Let $\dot{S}_\mu = \dot{S} \cap V_\mu$.

Thus $p \Vdash_{\mathbb{P}_\mu}$ “$\dot{S}_\mu$ is stationary”. Define q as follows: $\operatorname{dom} q = \operatorname{dom} p \cup \{\mu^+\}$, $q \restriction \mu = p \restriction \mu$ and $q(\mu^+) = f$, $\operatorname{dom} f = \{0\}$ and $f(0) = k_\mu^{-1}(\dot{S}_\mu)$. Then $q \Vdash_{\mathbb{P}} \dot{S}_\mu = D_\mu$ provided that $q \Vdash_{\mathbb{P}}$ “$\dot{S}_\mu$ is stationary”. The latter holds since $\mathbb{P}^\mu$ is $< \omega_1$-closed, and does not kill stationarity of $(\dot{S}_\mu)_{G_\mu}$ so $(\dot{S}_\mu)_{G_\mu}$ is stationary in $V[G]$ and, by the assumption on μ, $(\dot{S}_\mu)_{G_\mu} = (\dot{S}_\mu)_G$. Finally, it remains to show that in $V[G]$, $(\dot{S}_\mu)_G = S \cap \mu$. But this again follows from the definition of μ.

Instead of collapsing κ to ω_2, we could do the same for λ^{++} for any regular $\lambda < \kappa$ and obtain a model in which $E_{S^{\lambda^{++}}_{\lambda}} \leqslant_c E_{S^{\lambda^{++}}_{\lambda^+}}$. □

Open Problem Is it consistent that $S^{\omega_2}_{\omega_1}$ Borel reduces to $S^{\omega_2}_{\omega}$?

3.4.3 E_0 and $E_{S^\kappa_\lambda}$

In Subsection 3.4.2 above, Theorem 3.27, we showed that the equivalence relations of the form $E_{S^\kappa_\lambda}$ can form an antichain with respect to $\leqslant_B$. We will show that under mild set theoretical assumptions, all of them are strictly above

$$E_0 = \{(\eta, \xi) \mid \eta^{-1}\{1\} \triangle \xi^{-1}\{1\} \text{ is bounded}\}.$$

Theorem 3.31 *Let κ be regular and $S \subset \kappa$ stationary and suppose that $\diamondsuit_\kappa(S)$ holds (i.e., $\diamondsuit_\kappa$ holds on the stationary set S). Then E_0 is Borel reducible to E_S.*

Proof. The proof uses similar ideas than the proof of Theorem 3.29. Suppose that $\diamondsuit_\kappa(S)$ holds and let $\langle D_\alpha \rangle_{\alpha \in S}$ be the $\diamondsuit_\kappa(S)$-sequence. Define the reduction $f \colon 2^\kappa \to 2^\kappa$ by

$$f(X) = \{\alpha \in S \mid D_\alpha \text{ and } X \cap \alpha \text{ agree on a final segment of } \alpha\}.$$

If X, Y are E_0-equivalent, then $f(X)$, $f(Y)$ are E_S-equivalent, because they are in fact even E_0-equivalent as is easy to check. If X, Y are not E_0-equivalent, then there is a club C of α where X, Y differ cofinally in α; it follows that $f(X)$, $f(Y)$ differ on a stationary subset of S, namely the elements α of $C \cap S$ where D_α equals $X \cap \alpha$. □

Corollary 3.32 *Suppose $\kappa = \lambda^+ = 2^\lambda$. Then E_0 is Borel reducible to E_S where $S \subset \kappa \setminus S^\kappa_{\mathrm{cf}(\lambda)}$ is stationary.*

Proof. Gregory proved in [**4**] that if $2^\mu = \mu^+ = \kappa$, μ is regular and $\lambda < \mu$, then $\diamondsuit_\kappa(S^\kappa_\lambda)$ holds. Shelah extended this result in [**31**] and proved that if $\kappa = \lambda^+ = 2^\lambda$ and $S \subset \kappa \setminus S^\kappa_{\mathrm{cf}(\lambda)}$, then $\diamondsuit_\kappa(S)$ holds. Now apply Theorem 3.31. □

Corollary 3.33 (GCH) *Let us assume that κ is a successor cardinal. Then in a cofinality and GCH preserving forcing extension there is an embedding*

$$f \colon \langle \mathcal{P}(\kappa), \subset \rangle \longrightarrow \langle \mathrm{E}^{\Sigma^1_1}, \leqslant_B \rangle,$$

where $\mathrm{E}^{\Sigma^1_1}$ is the set of Σ^1_1-equivalence relations (see Theorem 3.23*) such that, for all $A \in \mathcal{P}(\kappa)$, E_0 is strictly below $f(A)$. If κ is not the successor of an ω-cofinal cardinal, we may replace Σ^1_1 above by Borel*.*

Proof. Suppose first that κ is not the successor of an ω-cofinal cardinal. By Theorem 3.23 there is a GCH and cofinality-preserving forcing extension such that there is an embedding

$$f \colon \langle \mathcal{P}(\kappa), \subset \rangle \longrightarrow \langle \mathrm{E}^{\mathrm{Borel}^*}, \leqslant_B \rangle.$$

From the proof of Theorem 3.23 one sees that $f(A)$ is of the form E_S where $S \subset S^\kappa_\omega$. Now E_0 is reducible to such relations by Corollary 3.32, as GCH continues to hold in the extension.

So it suffices to show that $E_S \not\leqslant_B E_0$ for stationary $S \subset S^\kappa_\omega$. By the same argument as in Corollary 3.21 on page 509, E_S is not Borel and by Theorem 3.3 on page 491, E_0 is Borel, so by Fact 5.1 on page 527, $E_{S^\kappa_\lambda}$ is not reducible to E_0.

Suppose κ is the successor of an ω-cofinal ordinal and $\kappa > \omega_1$. Then, in the proof of Theorem 3.23 replace μ by ω_1 and get the same result as above but for relations of the form E_S where $S \subset S^\kappa_{\omega_1}$.

The remaining case is $\kappa = \omega_1$. Let $\{S_\alpha \mid \alpha < \omega_1\}$ be a set of pairwise disjoint stationary subsets of ω_1. Let $\mathbb{P}$ be the forcing given by the proof of Theorem 3.23 such that in the $\mathbb{P}$-generic extension the function $f \colon \langle \mathcal{P}(\omega_1), \subset\rangle \to \langle \mathrm{E}^{\mathrm{Borel}^*}, \leqslant_B\rangle$ given by $f(A) = E_{\bigcup_{\alpha\in A} S_\alpha}$ is an embedding. This forcing preserves stationary sets, so as in the proof of clause (4) of Theorem 3.20, we can first force a $\diamondsuit$-sequence which guesses each subset of $\bigcup_{\alpha<\omega_1} S_\alpha$ on a set S such that $S \cap S_\alpha$ is stationary for all α. Then, by Corollary 3.32, E_0 is reducible to $E_{\bigcup_{\alpha\in A} S_\alpha}$ for all $A \subset \kappa$. □

Remark The embeddings from [**20**] are in contrast *strictly below* E_0.

4 Complexity of isomorphism relations

Let T be a countable complete theory. Let us turn to the question discussed in Section 5: "How is the set theoretic complexity of $\cong_T$ related to the stability theoretic properties of T?". The following theorems give some answers. As pointed out in Section 5, the assumption that κ is uncountable is crucial in the following theorems. For instance, the theory of dense linear orderings without end points is unstable, but $\cong_T$ is an open set in case $\kappa = \omega$, while we show below that for unstable theories T the set $\cong_T$ cannot be even Δ^1_1 when $\kappa > \omega$. Another example introduced by Martin Koerwien in his Ph.D. thesis and in [**19**] shows that there are classifiable shallow theories whose isomorphism is not Borel when $\kappa = \omega$, although we prove below that the isomorphism of such theories is always Borel, when $\kappa^{<\kappa} = \kappa > 2^\omega$. This justifies in particular the motivation for studying the space κ^κ for model theoretic purposes: the set theoretic complexity of $\cong_T$ positively correlates with the model theoretic complexity of T.

The following stability theoretical notions will be used: stable, superstable, DOP, OTOP, shallow, $\lambda(T)$ and $\kappa(T)$. Classifiable means superstable with no DOP nor OTOP and $\lambda(T)$ is the least cardinal in which T is stable.

Recall that by $\cong^\kappa_T$ we denote the isomorphism relation of models of T whose size is κ. The main theme in this section is exposed in the following two theorems:

Theorem 4.1 ($\kappa^{<\kappa} = \kappa$) *Assume that κ is a successor and let T be a complete countable theory. If $\cong^\kappa_T$ is Borel, then T is classifiable and shallow. If additionally $\kappa > 2^\omega$, then the converse holds: if T is classifiable and shallow, then $\cong^\kappa_T$ is Borel.*

Theorem 4.2 ($\kappa^{<\kappa} = \kappa$) *Assume that, for all $\lambda < \kappa$, $\lambda^\omega < \kappa$ and $\kappa > \omega_1$. Then in L and in the forcing extension after adding κ^+ Cohen subsets of κ we have: for any theory T, T is classifiable if and only if $\cong_T$ is Δ^1_1.*

The two theorems above are proved in many sub-theorems below. Our results are stronger than those given by 4.1 and 4.2 (for instance the cardinality assumption $\kappa > \omega_1$ is needed only in the case where T is superstable with DOP and the stable unsuperstable case is the only one for which Theorem 4.2 cannot be proved in ZFC). Theorem 4.1 follows from Theorems 4.6 and 4.7. Theorem 4.2 follows from Theorems 4.8, 4.9, 4.10 and items (2) and (7) of Theorem 3.20.

4.1 Preliminary results

The following Theorems 4.3 and 4.5 (page 523) will serve as bridges between the set theoretic complexity and the model theoretic complexity of an isomorphism relation.

Theorem 4.3 ($\kappa^{<\kappa} = \kappa$) *For a theory T, the set $\cong_T$ is Borel if and only if the following holds: there exists a $\kappa^+\omega$-tree t such that for all models $\mathcal{A}$ and $\mathfrak{B}$ of T, we have that $\mathcal{A} \cong \mathfrak{B} \Leftrightarrow \mathbf{II} \uparrow \mathrm{EF}^\kappa_t(\mathcal{A}, \mathfrak{B})$.*

Proof. Recall that we assume $\operatorname{dom} \mathcal{A} = \kappa$ for all models in the discourse. First suppose that there exists a $\kappa^+\omega$-tree t such that for all models $\mathcal{A}$ and $\mathfrak{B}$ of T, we have that $\mathcal{A} \cong \mathfrak{B} \Leftrightarrow \mathbf{II} \uparrow \mathrm{EF}^\kappa_t(\mathcal{A}, \mathfrak{B})$. Let us show that there exists a $\kappa^+\omega$-tree u which constitutes a Borel code for $\cong_T$ (see Remark 1.18 on page 480).

Let u be the tree of sequences of the form

$$\langle (p_0, A_0), f_0, (p_1, A_1), f_1, \ldots, (p_n, A_n), f_n \rangle$$

such that, for all $i \leqslant n$,

(1) (p_i, A_i) is a move of player $\mathbf{I}$ in EF^κ_t, i.e., $p_i \in t$ and $A_i \subset \kappa$ with $|A_i| < \kappa$;
(2) f_i is a move of player $\mathbf{II}$ in EF^κ_t, i.e., it is a partial function $\kappa \to \kappa$ with $|\operatorname{dom} f_i|, |\operatorname{ran} f_i| < \kappa$ and $A_i \subset \operatorname{dom} f_i \cap \operatorname{ran} f_i$;
(3) $\langle (p_0, A_0), f_0, (p_1, A_1), f_1, \ldots, (p_n, A_n), f_n \rangle$ is a valid position of the game, i.e., $(p_i)_{i \leqslant n}$ is an initial segment of a branch in t and $A_i \subset A_j$ and $f_i \subset f_j$ whenever $i < j \leqslant n$.

Order u by end extension. The tree u is a $\kappa^+\omega$-tree (because t is and by (3)).

Let us now define the function

$$h \colon \{\text{branches of } u\} \longrightarrow \{\text{basic open sets of } (\kappa^\kappa)^2\}.$$

Let $b \subset u$ be a branch,

$$b = \{\varnothing, \langle (p_0, A_0) \rangle, \langle (p_0, A_0), f_0 \rangle, \ldots, \langle (p_0, A_0), f_0, \ldots, (p_k, A_k), f_k \rangle\}.$$

It corresponds to a unique EF-game between some two structures with domains κ. In this game the players have chosen some set $A_k = \bigcup_{i \leqslant k} A_i \subset \kappa$ and some partial function $f_k = \bigcup_{i \leqslant k} f_i \colon \kappa \to \kappa$. Let $h(b)$ be the set of all pairs $(\eta, \xi) \in (\kappa^\kappa)^2$ such that $f_\kappa \colon \mathcal{A}_\eta \restriction A_\kappa \cong \mathcal{A}_\xi \restriction A_\kappa$ is a partial isomorphism. This is clearly an open set:

$$(\eta, \xi) \in h(b) \Rightarrow N_{\eta \restriction ((\sup A_\kappa)+1)} \times N_{\xi \restriction ((\sup A_\kappa)+1)} \subset h(b).$$

Finally we claim that $\mathcal{A}_\eta \cong \mathcal{A}_\xi \Leftrightarrow \mathbf{II} \uparrow G(u, h, (\eta, \xi))$. Here G is the game as in Definition 1.17 of Borel* sets, page 480 but played on the product $\kappa^\kappa \times \kappa^\kappa$. Assume $\mathcal{A}_\eta \cong \mathcal{A}_\xi$. Then $\mathbf{II} \uparrow \mathrm{EF}^\kappa_t(\mathcal{A}_\eta, \mathcal{A}_\xi)$. Let v denote the winning strategy. In the game $G(u, h, (\eta, \xi))$, let us define a winning strategy for player $\mathbf{II}$ as follows. By definition, at a particular move, say n, $\mathbf{I}$ chooses a sequence

$$\langle (p_0, A_0), f_0, \ldots, (p_n, A_n) \rangle.$$

Next $\mathbf{II}$ extends it according to v to

$$\langle (p_0, A_0), f_0, \ldots, (p_n, A_n), f_n \rangle,$$

where $f_n = \upsilon((p_0, A_0), \dots, (p_n, A_n))$. Since υ was a winning strategy, it is clear that $f_\kappa = \bigcup_{i<\kappa} f_i$ is going to be a isomorphism between $\mathcal{A}_\eta \restriction A_\kappa$ and $\mathcal{A}_\xi \restriction A_\kappa$, so $(\eta, \xi) \in h(b)$.

Assume that $\mathcal{A}_\eta \ncong \mathcal{A}_\xi$. Then by the assumption there is no winning strategy of **II**, so player **I** can play in such a way that $f_\kappa = \bigcup_{i \leqslant \kappa} f_i$ is not an isomorphism between $\mathcal{A}_\eta \restriction \cup A_i$ and $\mathcal{A}_\xi \restriction \cup A_i$, so (η, ξ) is not in $h(b)$. This completes the proof of the direction "$\Leftarrow$".

Let us prove "$\Rightarrow$". Suppose $\cong_T$ is Borel and let us show that there is a tree as in the statement of the theorem. We want to use Theorem 2.2 and formalize the statement "$\cong_T$ is definable in $L_{\kappa^+\kappa}$" by considering the space consisting of pairs of models.

Denote the vocabulary of $\mathcal{A}$ and $\mathfrak{B}$ as usual by L. Let P be a unary relation symbol not in L. We will now discuss two distinct vocabularies L and $L \cup \{P\}$ at the same time, so we have to introduce two distinct codings. Fix an $\eta \in 2^\kappa$. Let $\mathcal{A}_\eta$ denote the L-structure as defined in Definition 1.14 of our usual coding. Let $\rho \colon \kappa \cup \kappa^{<\omega} \to \kappa$ be a bijection and define $\mathcal{A}^\eta$ to be the model with $\operatorname{dom} \mathcal{A}^\eta = \kappa$ and if $a \in \operatorname{dom} \mathcal{A}^\eta$, then $\mathcal{A}^\eta \models P(a) \Leftrightarrow \eta(\rho(a)) = 1$ such that if $(a_1, \dots, a_n) \in (\operatorname{dom} \mathcal{A}^\eta)^n$, then $\mathcal{A}^\eta \models P_n(a_1, \dots, a_n) \Leftrightarrow \eta(\rho(a_1, \dots, a_n)) = 1$. Note that we are making a distinction here between κ and $\kappa^{\{0\}}$.

Claim 1 The set $W = \{\eta \in 2^\kappa \mid \kappa = |P^{\mathcal{A}^\eta}| = |\kappa \setminus P^{\mathcal{A}^\eta}|\}$ is Borel.

Proof of Claim 1. Let us show that the complement is Borel. By symmetry it is sufficient to show that

$$B = \{\eta \mid \kappa > |P^{\mathcal{A}^\eta}|\}$$

is Borel. Let $I \subset \kappa$ be a subset of size $< \kappa$. For $\beta \notin I$ define $U(I, \beta)$ to be the set

$$U(I, \beta) = \{\eta \mid \eta(\rho(\beta)) = 0\}.$$

Clearly $U(I, \beta)$ is open for all I, β. Now

$$B = \bigcup_{I \in [\kappa]^{<\kappa}} \bigcap_{\beta \notin I} U(I, \beta).$$

By the assumption $\kappa^{<\kappa} = \kappa$, this is Borel (in fact a union of closed sets). $\square_{\text{Claim 1}}$

Define a mapping $h \colon W \to (2^\kappa)^2$ as follows. Suppose $\xi \in W$. Let

$$r_1 \colon \kappa \longrightarrow P^{\mathcal{A}^\xi}$$

and

$$r_2 \colon \kappa \longrightarrow \kappa \setminus P^{\mathcal{A}^\xi}$$

be the order preserving bijections (note $P^{\mathcal{A}^\eta} \subset \kappa = \operatorname{dom} \mathcal{A}^\eta$).

Let η_1 be such that r_1 is an isomorphism

$$\mathcal{A}_{\eta_1} \longrightarrow (\mathcal{A}^\xi \cap P^{\mathcal{A}^\xi}) \restriction L$$

and η_2 such that r_2 is an isomorphism

$$\mathcal{A}_{\eta_2} \longrightarrow (\mathcal{A}^\xi \setminus P^{\mathcal{A}^\xi}) \restriction L.$$

Clearly η_1 and η_2 are unique, so we can define $h(\xi) = (\eta_1, \eta_2)$.

Claim 2 h is continuous.

Proof of Claim 2. Let $U = N_p \times N_q$ be a basic open set of $(2^\kappa)^2$, $p, q \in 2^{<\kappa}$ and let $\xi \in h^{-1}[U]$. Let $P^{\mathcal{A}^\xi} = \{\beta_i \mid i < \kappa\}$ be an enumeration such that $\beta_i < \beta_j \Leftrightarrow i < j$ and similarly $\kappa \setminus P^{\mathcal{A}^\xi} = \{\gamma_i \mid i < \kappa\}$. Let $\alpha = \max\{\beta_{\operatorname{dom} p}, \gamma_{\operatorname{dom} q}\} + 1$. Then $N_{\xi\restriction\alpha} \subset h^{-1}[U]$. Thus arbitrary ξ in $h^{-1}[U]$ have an open neighborhood in $h^{-1}[U]$, so it is open. $\square_{\text{Claim 2}}$

Recall our assumption that $E = \{(\eta, \xi) \in 2^\kappa \mid \mathcal{A}_\eta \cong \mathcal{A}_\xi\}$ is Borel. Since h is continuous and in particular Borel, this implies that

$$E' = \{\eta \mid \mathcal{A}_{h_1(\eta)} \cong \mathcal{A}_{h_2(\eta)}\} = h^{-1}E$$

is Borel in W. Because W is itself Borel, E' is Borel in 2^κ. Additionally, E' is closed under permutations: if $\mathcal{A}^\eta$ is isomorphic to $\mathcal{A}^\xi$, then $\mathcal{A}^\eta \cap P^{\mathcal{A}^\eta}$ is isomorphic to $\mathcal{A}^\xi \cap P^{\mathcal{A}^\xi}$ and $\mathcal{A}^\eta \setminus P^{\mathcal{A}^\eta}$ is isomorphic to $\mathcal{A}^\xi \setminus P^{\mathcal{A}^\xi}$, so if $\mathcal{A}^\eta \in E'$, then also $\mathcal{A}^\xi \in E'$ (and note that since $\eta \in W$, also $\xi \in W$). By Theorem 2.2 (page 483), there is a sentence θ of $L_{\kappa^+\kappa}$ over $L \cup \{P\}$ that defines E'. Thus by Theorem 1.11 (page 478) and Remark 1.13 (page 478) there is a $\kappa^+\omega$-tree t such that

$$\text{if } \eta \in E' \text{ and } \xi \notin E', \text{ then } \mathbf{II} \not\uparrow \mathrm{EF}_t^\kappa(\mathcal{A}^\eta, \mathcal{A}^\xi). \tag{4.1}$$

We claim that t is as needed, i.e., for all models $\mathcal{A}, \mathfrak{B}$ of T

$$\mathcal{A} \cong \mathfrak{B} \iff \mathbf{II} \uparrow \mathrm{EF}_t^\kappa(\mathcal{A}, \mathfrak{B}).$$

Suppose not. Then there are models $\mathcal{A} \not\cong \mathfrak{B}$ such that $\mathbf{II} \uparrow \mathrm{EF}_t^\kappa(\mathcal{A}, \mathfrak{B})$. Let η and ξ be such that $\mathcal{A}_{h_1(\eta)} = \mathcal{A}_{h_2(\eta)} = \mathcal{A}_{h_1(\xi)} = \mathcal{A}$ and $\mathcal{A}_{h_2(\xi)} = \mathfrak{B}$. Clearly $\eta \in E'$, but $\xi \notin E'$, so by (4.1) there is no winning strategy of $\mathbf{II}$ in $\mathrm{EF}_t^\kappa(\mathcal{A}^\eta, \mathcal{A}^\xi)$ which is clearly a contradiction, because $\mathbf{II}$ can apply her winning strategies in $\mathrm{EF}_t^\kappa(\mathcal{A}, \mathfrak{B})$ and $\mathrm{EF}_t^\kappa(\mathcal{A}, \mathcal{A})$ to win in $\mathrm{EF}_t^\kappa(\mathcal{A}^\eta, \mathcal{A}^\xi)$. $\square_{\text{Theorem 4.3}}$

We will use the following lemma from [**24**]:

Lemma 4.4 *If $t \subset (\kappa^{<\kappa})^2$ is a tree and $\xi \in \kappa^\kappa$, denote*

$$t(\xi) = \{p \in \kappa^{<\kappa} \mid (p, \xi \restriction \operatorname{dom} p) \in t\}.$$

Similarly, if $t \in (\kappa^{<\kappa})^3$, then

$$t(\eta, \xi) = \{p \in \kappa^{<\kappa} \mid (p, \eta \restriction \operatorname{dom} p, \xi \restriction \operatorname{dom} p) \in t\}.$$

Assume that Z is Σ^1_1. Then Z is Δ^1_1 if and only if for every tree $t \subset (\kappa^{<\kappa})^2$ such that

$$t(\xi) \textit{ has a } \kappa\textit{-branch} \iff \xi \in Z$$

there exists a $\kappa^+\kappa$-tree t' such that $\xi \in Z \Leftrightarrow t(\xi) \not\leqslant t'$. (Recall that $t \leqslant t'$ when there exists a strictly order preserving map $t \to t'$.)

Theorem 4.5 *Let T be a theory and assume that for every $\kappa^+\kappa$-tree t there exist $(\eta, \xi) \in (2^\kappa)^2$ such that $\mathcal{A}_\eta, \mathcal{A}_\xi \models T$, $\mathcal{A}_\eta \not\cong \mathcal{A}_\xi$ but $\mathbf{II} \uparrow \mathrm{EF}_t^\kappa(\mathcal{A}_\eta, \mathcal{A}_\xi)$. Then $\cong_T$ is not Δ^1_1.*

Proof. Let us abbreviate some statements:

A(t): $t \subset (\kappa^{<\kappa})^3$ is a tree and for all $(\eta, \xi) \in (\kappa^\kappa)^2$,

$$(\eta, \xi) \in \cong_T \iff t(\eta, \xi) \text{ contains a } \kappa\text{-branch}.$$

B(t): $t \subset (\kappa^{<\kappa})^3$ is a $\kappa^+\kappa$-tree and for all $(\eta, \xi) \in \kappa^\kappa$,

$$(\eta, \xi) \in \cong_T \iff t(\eta, \xi) \not\leqslant t'.$$

Now Lemma 4.4 implies that if $\cong_T$ is Δ^1_1, then $\forall t[A(t) \to \exists t' B(t,t')]$. We will show that $\exists t[A(t) \wedge \forall t' \neg B(t,t')]$, which by Lemma 4.4 suffices to prove the theorem. Let us define t. In the following, ν_α, η_α and ξ_α stand respectively for $\nu \restriction \alpha$, $\eta \restriction \alpha$ and $\xi \restriction \alpha$.

$$t = \{(\nu_\alpha, \eta_\alpha, \xi_\alpha) \mid \alpha < \kappa \text{ and } \nu \text{ codes an isomorphism between } \mathcal{A}_\eta \text{ and } \mathcal{A}_\xi\}.$$

Using Theorem 1.15 it is easy to see that t satisfies $A(t)$. Assume now that t' is an arbitrary $\kappa^+\kappa$-tree. We will show that $B(t,t')$ does not hold. For that purpose let $u = \omega \times t'$ be the tree defined by the set $\{(n,s) \mid n \in \omega, s \in t'\}$ and the ordering

$$(n_0, s_0) <_u (n_1, s_1) \iff \big(s_0 <_{t'} s_1 \vee (s_0 = s_1 \wedge n_0 <_\omega n_1)\big). \tag{4.2}$$

This tree u is still a $\kappa^+\kappa$-tree, so by the assumption of the theorem there is a pair (ξ_1, ξ_2) such that $\mathcal{A}_{\xi_1}$ and $\mathcal{A}_{\xi_2}$ are non-isomorphic, but $\mathbf{II} \uparrow \mathrm{EF}^\kappa_u(\mathcal{A}_{\xi_1}, \mathcal{A}_{\xi_2})$.

It is now sufficient to show that $t(\xi_1, \xi_2) \not\leqslant t'$.

Claim 1 There is no order preserving function $\sigma t' \to t'$, where $\sigma t'$ is as in Definition 2.9.

Proof of Claim 1. Assume that $g\colon \sigma t' \to t'$ is order preserving. Define $x_0 = g(\varnothing)$ and

$$x_\alpha = g\big(\{y \in t' \mid \exists \beta < \alpha (y \leqslant x_\beta)\}\big) \text{ for } 0 < \alpha < \kappa.$$

Then $(x_\alpha)_{\alpha<\kappa}$ contradicts the assumption that t' is a $\kappa^+\kappa$-tree. $\square_{\text{Claim 1}}$

Claim 2 There is an order preserving function

$$\sigma t' \longrightarrow t(\xi_1, \xi_2).$$

Proof of Claim 2. The idea is that players **I** and **II** play an EF-game for each branch of the tree t' and **II** uses her winning strategy in $\mathrm{EF}^\kappa_u(\mathcal{A}_{\xi_1}, \mathcal{A}_{\xi_2})$ to embed that branch into the tree of partial isomorphisms. A problem is that the winning strategy gives arbitrary partial isomorphisms while we are interested in those which are coded by functions defined on page 479. Now the tree u of (3) above becomes useful.

Let σ be a winning strategy of player **II** in $\mathrm{EF}^\kappa_u(\mathcal{A}_{\xi_1}, \mathcal{A}_{\xi_2})$. We define $g\colon \sigma t' \to t(\xi_1, \xi_2)$ recursively. Recall the function π from Definition 1.14 and define

$$C = \{\alpha \mid \pi[\alpha^{<\omega}] = \alpha\}.$$

Clearly C is cub. If $s \subset t'$ is an element of $\sigma t'$, then we assume that g is defined for all $s' <_{\sigma t'} s$ and that EF^κ_u is played up to $(0, \sup s) \in u$. If s does not contain its supremum, then put $g(s) = \bigcup_{s'<s} g(s')$. Otherwise let them continue playing the game for ω more moves; at the n-th of these moves player **I** picks $(n, \sup s)$ from u and a $\beta < \kappa$ where β is an element of C above

$$\max\{\operatorname{ran} f_{n-1}, \operatorname{dom} f_{n-1}\},$$

where f_{n-1} is the previous move by **II**. (If $n = 0$, it does not matter what **I** does.) In that way the function $f = \bigcup_{n<\omega} f_n$ is a partial isomorphism such that $\operatorname{dom} f = \operatorname{ran} f = \alpha$ for some ordinal α. It is straightforward to check that such an f is coded by some $\nu_\alpha\colon \alpha \to \kappa$. It is an isomorphism between $\mathcal{A}_{\xi_1} \cap \alpha$ and $\mathcal{A}_{\xi_2} \cap \alpha$ and since α is in C, there are ξ_1' and ξ_2' such that $\xi_1 \restriction \alpha \subset \xi_1'$, $\xi_2 \restriction \alpha \subset \xi_2'$ and there is an isomorphism $\mathcal{A}_{\xi_1'} \cong \mathcal{A}_{\xi_2'}$ coded by some ν such that $\nu_\alpha = \nu \restriction \alpha$. Thus $\nu_\alpha \in t(\xi_1, \xi_2)$ is suitable for setting $g(s) = \nu_\alpha$. $\square_{\text{Claim 2}}$

$\square_{\text{Theorem 4.5}}$

4.2 Classifiable

Throughout this section κ is a regular cardinal satisfying $\kappa^{<\kappa} = \kappa > \omega$.

Theorem 4.6 ($\kappa > 2^\omega$) *If the theory T is classifiable and shallow, then $\cong_T$ is Borel.*

Proof. If T is classifiable and shallow, then from [**30**, Theorem XIII.1.5 and Claim XIII.1.3] it follows that the models of T are characterized by a fragment of $L_{\kappa^+\kappa}$ which consists of formulas of bounded quantifier rank (the bound depends on depth of T). By the standard argument this implies that the game EF_t^κ characterized models of T of size κ up to isomorphism, where t is some $\kappa^+\omega$-tree (in fact a tree of descending sequences of an ordinal $\alpha < \kappa^+$). Hence by Theorem 4.3 the isomorphism relation of T is Borel. □

Theorem 4.7 *If the theory T is classifiable but not shallow, then $\cong_T$ is not Borel. If κ is not weakly inaccessible and T is not classifiable, then $\cong_T$ is not Borel.*

Proof. If T is classifiable but not shallow, then, by [**30**, XIII.1.8], the $L_{\infty\kappa}$-Scott heights of models of T of size κ are not bounded by any ordinal $< \kappa^+$ (see Definition 1.9 on page 478). Because any $\kappa^+\omega$-tree can be embedded into $t_\alpha = \{\text{decreasing sequences of } \alpha\}$ for some α (see Fact 1.3 on page 474), this implies that for any $\kappa^+\omega$-tree t there exists a pair of models $\mathcal{A}, \mathfrak{B}$ such that $\mathcal{A} \not\cong \mathfrak{B}$ but $\mathbf{II} \uparrow \mathrm{EF}_t^\kappa(\mathcal{A}, \mathfrak{B})$. Theorem 4.3 now implies that the isomorphism relation is not Borel.

If T is not classifiable, κ is not weakly inaccessible; then, by [**29**, Theorem 0.2 (Main Conclusion)], there are non-isomorphic models of T of size κ which are $L_{\infty\kappa}$-equivalent, so the same argument as above, using Theorem 4.3, gives that $\cong_T$ is not Borel. □

Theorem 4.8 *If the theory T is classifiable, then $\cong_T$ is Δ_1^1.*

Proof. Shelah's theorem [**30**, Theorem XIII.1.1] says that if a theory T is classifiable, then any two models that are $L_{\infty\kappa}$-equivalent are isomorphic. But $L_{\infty\kappa}$ equivalence is equivalent to $\mathrm{EF}_\omega^\kappa$-equivalence (see Theorem 1.12 on page 478). So in order to prove the theorem it is sufficient to show that if for any two models $\mathcal{A}$, $\mathfrak{B}$ of the theory T it holds that $\mathbf{II} \uparrow \mathrm{EF}_\omega^\kappa(\mathcal{A}, \mathfrak{B}) \Leftrightarrow \mathcal{A} \cong \mathfrak{B}$, then the isomorphism relation is Δ_1^1. The game $\mathrm{EF}_\omega^\kappa$ is a closed game of length ω and so determined. Hence we have $\mathbf{I} \uparrow \mathrm{EF}_\omega^\kappa(\mathcal{A}, \mathfrak{B}) \Leftrightarrow \mathcal{A} \not\cong \mathfrak{B}$. By Theorem 1.8, the set

$$\{(\nu, \eta, \xi) \in (\kappa^\kappa)^3 \mid \nu \text{ codes a winning strategy for } \mathbf{I} \uparrow \mathrm{EF}_\omega^\kappa(\mathcal{A}_\eta, \mathcal{A}_\xi))\}$$

is closed and thus $\{(\eta, \xi) \mid \mathcal{A}_\eta \not\cong \mathcal{A}_\xi\}$ is Σ_1^1, which further implies that $\cong_T$ is Δ_1^1 by Corollary 1.16. □

4.3 Unclassifiable

4.3.1 The unstable, DOP and OTOP cases

As before, κ is a regular cardinal satisfying $\kappa^{<\kappa} = \kappa > \omega$.

Theorem 4.9

(1) *If T is unstable, then $\cong_T$ is not Δ_1^1.*
(2) *If T is stable with OTOP, then $\cong_T$ is not Δ_1^1.*
(3) *If T is superstable with DOP and $\kappa > \omega_1$, then $\cong_T$ is not Δ_1^1.*
(4) *If T is stable with DOP and $\lambda = \mathrm{cf}(\lambda) = \lambda(T) + \lambda^{<\kappa(T)} \geqslant \omega_1$, $\kappa > \lambda^+$, and, for all $\xi < \kappa$, $\xi^\lambda < \kappa$, then $\cong_T$ is not Δ_1^1. (Note that $\kappa(T) \in \{\omega, \omega_1\}$.)*

Proof. For a model $\mathcal{A}$ of size κ of a theory T let us denote by $E(\mathcal{A})$ the following property: for every $\kappa^+\kappa$-tree t there is a model $\mathfrak{B}$ of T of cardinality κ such that $\mathbf{II} \uparrow \mathrm{EF}_t^\kappa(\mathcal{A}, \mathfrak{B})$ and $\mathcal{A} \ncong \mathfrak{B}$.

For (3) we need a result by Hyttinen and Tuuri [**15**, Theorem 6.2]:

Fact (Superstable with DOP) Let T be a superstable theory with DOP, $\kappa^{<\kappa} = \kappa > \omega_1$. Then there exists a model $\mathcal{A}$ of T of cardinality κ with the property $E(\mathcal{A})$.

For (4) we need a result by Hyttinen and Shelah from [**14**]:

Fact (Stable with DOP) Let T be a stable theory with DOP and $\lambda = \mathrm{cf}(\lambda) = \lambda(T) + \lambda^{<\kappa(T)} \geqslant \omega_1$, $\kappa^{<\kappa} = \kappa > \lambda^+$, and, for all $\xi < \kappa$, $\xi^\lambda < \kappa$. Then there is a model $\mathcal{A}$ of T of power κ with the property $E(\mathcal{A})$.

For (1), a result by Hyttinen and Tuuri [**15**, Theorem 4.9]:

Fact (Unstable) Let T be an unstable theory. Then there exists a model $\mathcal{A}$ of T of cardinality κ with the property $E(\mathcal{A})$.

And for (2) another result by Hyttinen and Tuuri [**15**, Theorem 6.6]:

Fact (Stable with OTOP) Suppose T is a stable theory with OTOP. Then there exists a model $\mathcal{A}$ of T of cardinality κ with the property $E(\mathcal{A})$.

Now (1), (2) and (4) follow immediately from Theorem 4.5. □

4.3.2 Stable unsuperstable

We assume $\kappa^{<\kappa} = \kappa > \omega$ in all theorems below.

Theorem 4.10 *Assume that for all $\lambda < \kappa$, $\lambda^\omega < \kappa$.*

(1) *If T is stable unsuperstable, then $\cong_T$ is not Borel.*
(2) *If κ is as above and T is stable unsuperstable, then $\cong_T$ is not Δ_1^1 in the forcing extension after adding κ^+ Cohen subsets of κ, or if $V = L$.*

Proof. By Theorem 5.13 on page 542, the relation $E_{S_\omega^\kappa}$ can be reduced to $\cong_T$. The theorem follows now from Corollary 3.21 on page 509. □

On the other hand, stable unsuperstable theories may behave nicely to some extent:

Lemma 4.11 *Assume that T is a theory and t is a $\kappa^+\kappa$-tree such that if $\mathcal{A}$ and $\mathfrak{B}$ are models of T, then $\mathcal{A} \cong \mathfrak{B} \Leftrightarrow \mathbf{II} \uparrow \mathrm{EF}_t^\kappa(\mathcal{A}, \mathfrak{B})$. Then $\cong$ of T is Borel*.*

Proof. Similar to the proof of Theorem 4.3. □

Theorem 4.12 *Assume $\kappa \in I[\kappa]$ and $\kappa = \lambda^+$ ("$\kappa \in I[\kappa]$" is known as the Approachability Property and follows from $\lambda^{<\lambda} = \lambda$). Then there exists an unsuperstable theory T whose isomorphism relation is Borel*.*

Proof. In [**12**] and [**13**] Hyttinen and Shelah show the following (Theorem 1.1 of [**13**], but the proof is essentially in [**12**]):

> Suppose $T = ((\omega^\omega, E_i)_{i<\omega})$, where $\eta E_i \xi$ if and only if, for all $j \leqslant i$, $\eta(j) = \xi(j)$. If $\kappa \in I[\kappa]$, $\kappa = \lambda^+$, and $\mathcal{A}$ and $\mathfrak{B}$ are models of T of cardinality κ, then $\mathcal{A} \cong \mathfrak{B} \Leftrightarrow \mathbf{II} \uparrow \mathrm{EF}_{\lambda\cdot\omega+2}^\kappa(\mathcal{A}, \mathfrak{B})$, where $+$ and $\cdot$ denote the ordinal sum and product, i.e., $\lambda \cdot \omega + 2$ is just an ordinal.

So taking the tree t to be $\lambda \cdot \omega + 2$ the claim follows from Lemma 4.11. □

Open Problem If $\kappa = 2^\omega$, is the isomorphism relation of all classifiable and shallow theories Borel on structures of size κ?

Open Problem We proved that, if $\kappa > 2^\omega$, the isomorphism relation of a theory T is Borel if and only if T is classifiable and shallow. Is there a connection between the depth of a shallow theory and the Borel degree of its isomorphism relation? Is one monotone in the other?

Open Problem Can it be proved in ZFC that if T is stable unsuperstable then $\cong_T$ is not Δ^1_1?

5 Reductions

Recall that in Section 4 we obtained a provable characterization of theories which are both classifiable and shallow in terms of the definability of their isomorphism relations. Without the shallowness condition we obtained only a consistency result. In this section we improve this to a provable characterization by analyzing isomorphism relations in terms of Borel reducibility.

Recall the definition of a reduction (section *Reductions*, page 474), and recall that if $X \subset \kappa$ is a stationary subset, we denote by E_X the equivalence relation defined by

$$\forall \eta, \xi \in 2^\kappa (\eta E_X \xi \Leftrightarrow (\eta^{-1}\{1\} \triangle \xi^{-1}\{1\}) \cap X \text{ is non-stationary}),$$

and by S^κ_λ we mean the ordinals of cofinality λ that are less than κ.

The equivalence relations E_X are Σ^1_1 ($\eta E_X \xi$ if and only if *there exists* a cub subset of $\kappa \setminus (X \cap (\eta \triangle \xi))$).

Simple conclusions can readily be made from the following observation that roughly speaking, the set theoretic complexity of a relation does not decrease under reductions:

Fact 5.1 If E_1 is a Borel (or Δ^1_1) equivalence relation and E_0 is an equivalence relation with $E_0 \leqslant_B E_1$, then E_0 is Borel (respectively Δ^1_1 if E_1 is Δ^1_1). □

The main theorem of this section is:

Theorem 5.2 *Suppose $\kappa = \lambda^+ = 2^\lambda > 2^\omega$ where $\lambda^{<\lambda} = \lambda$. Let T be a first-order theory. Then T is classifiable if and only if, for all regular $\mu < \kappa$, $E_{S^\kappa_\mu} \not\leqslant_B \cong^\kappa_T$.*

5.1 Classifiable theories

The following follows from [**30**, Theorem XIII.1.1] (see also the proof of Theorem 4.8 above):

Theorem 5.3 ([**30**]) *If a first-order theory T is classifiable and $\mathcal{A}$ and $\mathfrak{B}$ are non-isomorphic models of T of size κ, then $\mathbf{I} \uparrow \mathrm{EF}^\kappa_\omega(\mathcal{A}, \mathfrak{B})$.* □

Theorem 5.4 ($\kappa^{<\kappa} = \kappa$) *If a first-order theory T is classifiable, then, for all $\lambda < \kappa$,*

$$E_{S^\kappa_\lambda} \not\leqslant_B \cong^\kappa_T .$$

Proof. Let $\mathrm{NS} \in \{E_{S^\kappa_\lambda} \mid \lambda \in \mathrm{reg}(\kappa)\}$. Suppose $r \colon 2^\kappa \to 2^\kappa$ is a Borel function such that

$$\forall \eta, \xi \in 2^\kappa (\mathcal{A}_{r(\eta)} \models T \wedge \mathcal{A}_{r(\xi)} \models T \wedge (\eta \,\mathrm{NS}\, \xi \Leftrightarrow \mathcal{A}_{r(\eta)} \cong \mathcal{A}_{r(\xi)})). \tag{5.1}$$

By Lemma 3.2, page 491, let D be an intersection of κ-many dense open sets such that $R = r \restriction D$ is continuous. D can be coded into a function $v \colon \kappa \times \kappa \to \kappa^{<\kappa}$ such that

$D = \bigcap_{i<\kappa} \bigcup_{j<\kappa} N_{v(i,j)}$. Since R is continuous, it can also be coded into a single function $u\colon \kappa^{<\kappa} \times \kappa^{<\kappa} \to \{0,1\}$ such that

$$R(\eta) = \xi \iff (\forall \alpha < \kappa)(\exists \beta < \kappa)[u(\eta\restriction\beta, \xi\restriction\alpha) = 1].$$

(For example, define $u(p,q) = 1$ if $D \cap N_p \subset R^{-1}[N_q]$.) Let

$$\varphi(\eta, \xi, u, v) = (\forall \alpha < \kappa)(\exists \beta < \kappa)[u(\eta\restriction\beta, \xi\restriction\alpha) = 1] \wedge (\forall i < \kappa)(\exists j < \kappa)[\eta \in N_{v(i,j)}].$$

It is a formula of set theory with parameters u and v. It is easily seen that φ is absolute for transitive elementary submodels M of $H(\kappa^+)$ containing κ, u and v with $(\kappa^{<\kappa})^M = \kappa^{<\kappa}$.

Let $\mathbb{P} = 2^{<\kappa}$ be the Cohen forcing. Suppose $M \preccurlyeq H(\kappa^+)$ is a model as above, i.e., transitive, $\kappa, u, v \in M$ and $(\kappa^{<\kappa})^M = \kappa^{<\kappa}$. Note that then $\mathbb{P} \cup \{\mathbb{P}\} \subset M$. Then, if G is $\mathbb{P}$-generic over M, then $\cup G \in D$ and there is ξ such that $\varphi(\cup G, \xi, u, v)$. By the definition of φ and u, an initial segment of ξ can be read from an initial segment of $\cup G$. That is why there is a nice $\mathbb{P}$-name τ for a function (see [**21**]) such that

$$\varphi(\cup G, \tau_G, u, v)$$

whenever G is $\mathbb{P}$-generic over M.

Now since the game $\mathrm{EF}^\kappa_\omega$ is determined on all structures, (at least) one of the following holds:

(1) there is p such that $p \Vdash \mathbf{II} \uparrow \mathrm{EF}^\kappa_\omega(\mathcal{A}_\tau, \mathcal{A}_{r(\overline{0})})$;
(2) there is p such that $p \Vdash \mathbf{I} \uparrow \mathrm{EF}^\kappa_\omega(\mathcal{A}_\tau, \mathcal{A}_{r(\overline{0})})$,

where $\overline{0}$ is the constant function with value 0. Let us show that both of them lead to a contradiction.

Assume (1). Fix a nice $\mathbb{P}$-name σ such that

$$p \Vdash \text{“}\sigma \text{ is a winning strategy of } \mathbf{II} \text{ in } \mathrm{EF}^\kappa_\omega(\mathcal{A}_\tau, \mathcal{A}_{r(\overline{0})})\text{”}.$$

A strategy is a subset of $([\kappa]^{<\kappa})^{<\omega} \times \kappa^{<\kappa}$ (see Definition 1.7 on page 475), and the forcing does not add elements to that set, so the nice name can be chosen such that all names in $\operatorname{dom}\sigma$ are standard names for elements that are in $([\kappa]^{<\kappa})^{<\omega} \times \kappa^{<\kappa} \in H(\kappa^+)$.

Let M be an elementary submodel of $H(\kappa^+)$ of size κ such that

$$\{u, v, \sigma, r(\overline{0}), \tau, \mathbb{P}\} \cup (\kappa + 1) \cup M^{<\kappa} \subset M.$$

Listing all dense subsets of $\mathbb{P}$ in M, it is easy to find a $\mathbb{P}$-generic G over M which contains p and such that $(\cup G)^{-1}\{1\}$ contains a cub. Now in V, $\cup G \not{\mathrm{NS}}\, \overline{0}$. Since $\varphi(\cup G, \tau_G, u, v)$ holds, we have, by (5.1),

$$\mathcal{A}_{\tau_G} \not\cong \mathcal{A}_{r(\overline{0})}. \tag{5.2}$$

Let us show that σ_G is a winning strategy of player $\mathbf{II}$ in $\mathrm{EF}^\kappa_\omega(\mathcal{A}_{\tau_G}, \mathcal{A}_{r(\overline{0})})$ (in V) which by Theorem 5.3 above is a contradiction with (1).

Let μ be any strategy of player $\mathbf{I}$ in $\mathrm{EF}^\kappa_\omega(\mathcal{A}_{\tau_G}, \mathcal{A}_{r(\overline{0})})$ and let us show that σ_G beats it. Consider the play $\sigma_G * \mu$ and assume for a contradiction that it is a win for $\mathbf{I}$. This play is well defined, since the moves made by μ are in the domain of σ_G by the note after the definition of σ, and because $([\kappa]^{<\kappa})^{<\omega} \times \kappa^{<\kappa} \subset M$.

The play consists of ω moves and is a countable sequence in the set $([\kappa]^{<\kappa}) \times \kappa^{<\kappa}$. Since $\mathbb{P}$ is $< \kappa$ closed, there is $q_0 \in \mathbb{P}$ which decides $\sigma_G * \mu$ (i.e., $\sigma_{G_0} * \mu = \sigma_{G_1} * \mu$ whenever $q_0 \in G_0 \cap G_1$). Assume that G' is a $\mathbb{P}$-generic over V with $q_0 \in G'$. Then

$$(\sigma_{G'} * \mu)^{V[G']} = (\sigma_G * \mu)^{V[G']} = (\sigma_G * \mu)^V$$

(again, because $\mathbb{P}$ does not add elements of $\kappa^{<\kappa}$) and so

$$(\sigma_{G'} * \mu \text{ is a win for } \mathbf{I})^{V[G']}.$$

But $q_0 \Vdash$ "$\sigma * \mu$ is a win for $\mathbf{II}$", because q_0 extends p and by the choice of σ.

The case (2) is similar, just instead of choosing $\cup G$ such that $(\cup G)^{-1}\{1\}$ contains a cub, choose G such that $(\cup G)^{-1}\{0\}$ contains a cub. Then we should have $\mathcal{A}_{\tau_G} \cong \mathcal{A}_{r(\bar{0})}$ which contradicts (2) by the same absoluteness argument as above. □

5.2 Unstable and superstable theories

In this section we use Shelah's ideas on how to prove non-structure theorems using Ehrenfeucht–Mostowski models [**29**]. We use the definition of Ehrenfeucht–Mostowski models from [**15**, Definition 4.2.].

Definition 5.5 In the following discussion of linear orderings we use the following concepts:

- *Coinitiality* or *reverse cofinality* of a linear order η, denoted $\mathrm{cf}^*(\eta)$, is the smallest ordinal α such that there is a map $f\colon \alpha \to \eta$ which is strictly decreasing and $\operatorname{ran} f$ has no (strict) lower bound in η.
- If $\eta = \langle \eta, < \rangle$ is a linear ordering, by η^* we denote its mirror image: $\eta^* = \langle \eta, <^* \rangle$ where $x <^* y \Leftrightarrow y < x$.
- Suppose λ is a cardinal. We say that an ordering η is λ-*dense* if for all subsets A and B of η with the properties $\forall a \in A \, \forall b \in B \, (a < b)$ and $|A| < \lambda$ and $|B| < \lambda$ there is $x \in \eta$ such that $a < x < b$ for all $a \in A$, $b \in B$. *Dense* means ω-dense.

Theorem 5.6 *Suppose that $\kappa = \lambda^+ = 2^\lambda$ such that $\lambda^{<\lambda} = \lambda$. If T is unstable or superstable with OTOP, then $E_{S^\kappa_\lambda} \leqslant_c \cong_T$. If additionally $\lambda \geqslant 2^\omega$, then $E_{S^\kappa_\lambda} \leqslant_c \cong_T$ holds also for superstable T with DOP.*

Proof. We will carry out the proof for the case where T is unstable and shall make remarks on how certain steps of the proof should be modified in order for this to work for superstable theories with DOP or OTOP. First, for each $S \subset S^\kappa_\lambda$, let us construct the linear orders $\Phi(S)$ which will serve a fundamental role in the construction. The following claim is a special case of Lemma 7.17 in [**10**]:

Claim 1 For each cardinal μ of uncountable cofinality there exists a linear ordering $\eta = \eta_\mu$ which satisfies:

(1) $\eta \cong \eta + \eta$;
(2) for all $\alpha \leqslant \mu$, $\eta \cong \eta \cdot \alpha + \eta$;
(3) $\eta \cong \eta \cdot \mu + \eta \cdot \omega_1^*$;
(4) η is dense;
(5) $|\eta| = \mu$;
(6) $\mathrm{cf}^*(\eta) = \omega$.

Proof of Claim 1. Essentially the same as in [**10**]. $\square_{\text{Claim 1}}$

For a set $S \subset S^\kappa_\lambda$, define the linear order $\Phi(S)$ as follows:

$$\Phi(S) = \sum_{i<\kappa} \tau(i, S),$$

where $\tau(i,S)=\eta_\lambda$ if $i\notin S$ and $\tau(i,S)=\eta_\lambda\cdot\omega_1^*$, if $i\in S$. Note that $\Phi(S)$ is dense. For $\alpha<\beta<\kappa$, define

$$\Phi(S,\alpha,\beta)=\sum_{\alpha\leqslant i<\beta}\tau(i,S).$$

(These definitions are also as in [**10**] although the idea dates back to J. Conway's Ph.D. thesis from the 1960's; they are first referred to in [**25**].) From now on, denote $\eta=\eta_\lambda$.

Claim 2 If $\alpha\notin S$, then for all $\beta\geqslant\alpha$ we have $\Phi(S,\alpha,\beta+1)\cong\eta$, and if $\alpha\in S$, then for all $\beta\geqslant\alpha$ we have $\Phi(S,\alpha,\beta+1)\cong\eta\cdot\omega_1^*$.

Proof of Claim 2. Let us begin by showing the first part, i.e., assume that $\alpha\notin S$. This is also like in [**10**]. We prove the statement by induction on $\mathrm{OTP}(\beta\setminus\alpha)$. If $\beta=\alpha$, then $\Phi(S,\alpha,\alpha+1)=\eta$ by the definition of Φ. If $\beta=\gamma+1$ is a successor, then $\beta\notin S$, because S contains only limit ordinals, so $\tau(\beta,S)=\eta$ and

$$\Phi(S,\alpha,\beta+1)=\Phi(S,\alpha,\gamma+1+1)=\Phi(S,\alpha,\gamma+1)+\eta,$$

which, by the induction hypothesis and by (1), is isomorphic to η. If $\beta\notin S$ is a limit ordinal, then choose a continuous cofinal sequence $s\colon \mathrm{cf}(\beta)\to\beta$ such that $s(\gamma)\notin S$ for all $\gamma<\mathrm{cf}(\beta)$. This is possible since S contains only ordinals of cofinality λ. By the induction hypothesis $\Phi(S,\alpha,s(0)+1)\cong\eta$,

$$\Phi(S,s(\gamma)+1,s(\gamma+1)+1)\cong\eta$$

for all successor ordinals $\gamma<\mathrm{cf}(\beta)$,

$$\Phi(S,s(\gamma),s(\gamma+1)+1)\cong\eta$$

for all limit ordinals $\gamma<\mathrm{cf}(\beta)$, and so now

$$\Phi(S,\alpha,\beta+1)\cong\eta\cdot\mathrm{cf}(\beta)+\eta,$$

which is isomorphic to η by (2). If $\beta\in S$, then $\mathrm{cf}(\beta)=\lambda$ and we can again choose a cofinal sequence $s\colon\lambda\to\beta$ such that $s(\alpha)$ is not in S for all $\alpha<\lambda$. By the induction hypothesis, as above,

$$\Phi(S,\alpha,\beta+1)\cong\eta\cdot\lambda+\tau(\beta,S),$$

and, since $\beta\in S$, we have $\tau(\beta,S)=\eta\cdot\omega_1^*$, so we have

$$\Phi(S,\alpha,\beta+1)\cong\eta\cdot\lambda+\eta\cdot\omega_1^*,$$

which, by (3), is isomorphic to η.

Suppose $\alpha\in S$. Then $\alpha+1\notin S$, so by the previous part we have

$$\Phi(S,\alpha,\beta+1)\cong\tau(\alpha,S)+\Phi(S,\alpha+1,\beta+1)=\eta\cdot\omega_1^*+\eta=\eta\cdot\omega_1^*.\qquad\square_{\text{Claim 2}}$$

This gives us a way to show that the isomorphism type of $\Phi(S)$ depends only on the $E_{S^\kappa_\lambda}$-equivalence class of S:

Claim 3 If $S,S'\subset S^\kappa_\lambda$ and $S\bigtriangleup S'$ is non-stationary, then $\Phi(S)\cong\Phi(S')$.

Proof of Claim 3. Let C be a cub set outside $S\bigtriangleup S'$. Enumerate it as $C=\{\alpha_i\mid i<\kappa\}$ where $(\alpha_i)_{i<\kappa}$ is an increasing and continuous sequence. Now $\Phi(S)=\bigcup_{i<\kappa}\Phi(S,\alpha_i,\alpha_{i+1})$ and $\Phi(S')=\bigcup_{i<\kappa}\Phi(S',\alpha_i,\alpha_{i+1})$. Note that by the definitions these are disjoint unions, so it is enough to show that for all $i<\kappa$ the orders $\Phi(S,\alpha_i,\alpha_{i+1})$ and $\Phi(S',\alpha_i,\alpha_{i+1})$ are isomorphic. But, for all $i<\kappa$, $\alpha_i\in S\Leftrightarrow\alpha_i\in S'$, so by Claim 2 either

$$\Phi(S,\alpha_i,\alpha_{i+1})\cong\eta\cong\Phi(S',\alpha_i,\alpha_{i+1})$$

(if $\alpha_i \notin S$) or

$$\Phi(S, \alpha_i, \alpha_{i+1}) \cong \eta \cdot \omega_1^* \cong \Phi(S', \alpha_i, \alpha_{i+1})$$

(if $\alpha_i \in S$). $\square_{\text{Claim 3}}$

Definition 5.7 K^λ_{tr} is the set of L-models $\mathcal{A}$ where $L = \{<, \lessdot, (P_\alpha)_{\alpha \leqslant \lambda}, h\}$, with the following properties:

- $\operatorname{dom} \mathcal{A} \subset I^{\leqslant \lambda}$ for some linear order I;
- $\forall x, y \in A (x < y \Leftrightarrow x \subset y)$;
- $\forall x \in A (P_\alpha(x) \Leftrightarrow \operatorname{len}(x) = \alpha)$;
- $\forall x, y \in A[x \lessdot y \Leftrightarrow \exists z \in A((x, y \in \operatorname{Succ}(z)) \wedge (I \models x < y))]$;
- $h(x, y)$ is the maximal common initial segment of x and y.

For each S, define the tree $T(S) \in K^\lambda_{\text{tr}}$ by

$$\begin{aligned} T(S) = \Phi(S)^{<\lambda} \cup \{\eta \colon \lambda \to \Phi(S) \mid \eta \text{ increasing and} \\ \operatorname{cf}^*(\Phi(S) \setminus \{x \mid (\exists y \in \operatorname{ran} \eta)(x < y)\}) = \omega_1\}. \end{aligned}$$

The relations $<$, $\lessdot$, P_n and h are interpreted in the natural way.

Clearly an isomorphism between $\Phi(S)$ and $\Phi(S')$ induces an isomorphism between $T(S)$ and $T(S')$; thus $T(S) \cong T(S')$ if $S \mathbin{\triangle} S'$ is non-stationary.

Claim 4 Suppose T is unstable in the vocabulary v. Let T_1 be T with Skolem functions in the Skolemized vocabulary $v_1 \supset v$. Then there is a function

$$\mathcal{P}(S^\kappa_\lambda) \longrightarrow \{\mathcal{A}^1 \mid \mathcal{A}^1 \models T_1, |\mathcal{A}^1| = \kappa\},$$

$S \mapsto \mathcal{A}^1(S)$, which has the following properties:

(a) There is a mapping $T(S) \to (\operatorname{dom} \mathcal{A}^1(S))^n$ for some $n < \omega$, $\eta \mapsto a_\eta$, such that $\mathcal{A}^1(S)$ is the Skolem hull of $\{a_\eta \mid \eta \in T(S)\}$, i.e., $\{a_\eta \mid \eta \in T(S)\}$ is the skeleton of $\mathcal{A}^1(S)$. Denote the skeleton of $\mathcal{A}$ by $\operatorname{Sk}(\mathcal{A})$.

(b) $\mathcal{A}(S) = \mathcal{A}^1(S) \restriction v$ is a model of T.

(c) $\operatorname{Sk}(\mathcal{A}^1(S))$ is indiscernible in $\mathcal{A}^1(S)$, i.e., if $\overline{\eta}, \overline{\xi} \in T(S)$ and $\operatorname{tp}_{\text{q.f.}}(\overline{\eta}/\varnothing) = \operatorname{tp}_{\text{q.f.}}(\overline{\xi}/\varnothing)$, where $\operatorname{tp}_{\text{q.f.}}$ denotes the quantifier free type, then $\operatorname{tp}(a_{\overline{\eta}}/\varnothing) = \operatorname{tp}(a_{\overline{\xi}}/\varnothing)$, where $a_{\overline{\eta}} = (a_{\eta_1}, \ldots, a_{\eta_{\operatorname{len} \overline{\eta}}})$. This assignment of types in $\mathcal{A}^1(S)$ to q.f.-types in $T(S)$ is independent of S.

(d) There is a formula $\varphi \in L_{\omega\omega}(v)$ such that, for all $\eta, \nu \in T(S)$ and $\alpha < \lambda$, if $T(S) \models P_\lambda(\eta) \wedge P_\alpha(\nu)$, then $T(S) \models \eta > \nu$ if and only if $\mathcal{A}(S) \models \varphi(a_\eta, a_\nu)$.

Proof of Claim 4. The following is known:

(F1) Suppose that T is a complete unstable theory. Then, for each linear order η, T has an Ehrenfeucht–Mostowski model $\mathcal{A}$ of vocabulary v_1, where $|v_1| = |T| + \omega$ and order is definable by a first-order formula, such that the template (assignment of types) is independent of η.[2]

It is not hard to see that for every tree $t \in K^\omega_{\text{tr}}$ we can define a linear order $L(t)$ satisfying the following conditions:

(1) $\operatorname{dom}(L(t)) = (\operatorname{dom} t \times \{0\}) \cup (\operatorname{dom} t \times \{1\})$;

(2) for all $a \in t$, $(a, 0) <_{L(t)} (a, 1)$;

[2] This is from [**26**]; there is a sketch of the proof also in [**15**, Theorem 4.7].

(3) if $a, b \in t$, then $a <_t b \Leftrightarrow [(a,0) <_{L(t)} (b,0)] \wedge [(b,1) <_{L(t)} (a,1)]$;
(4) if $a, b \in t$, then

$$(a \nleq b) \wedge (b \nleq a) \iff [(b,1) <_{L(t)} (a,0)] \vee [(a,1) <_{L(t)} (b,0)].$$

Now for every $S \subset \kappa$, by (F1), there is an Ehrenfeucht–Mostowski model $\mathcal{A}^1(S)$ for the linear order $L(T(S))$ where order is definable by the formula ψ which is in $L_{\infty\omega}$. Suppose $\overline{\eta} = (\eta_0, \dots, \eta_n)$ and $\overline{\xi} = (\xi_0, \dots, \xi_n)$ are sequences in $T(S)$ that have the same quantifier free type. Then the sequences

$$\langle (\eta_0,0), (\eta_0,1), (\eta_1,0), (\eta_1,1), \dots, (\eta_n,0), (\eta_n,1) \rangle$$

and

$$\langle (\xi_0,0), (\xi_0,1), (\xi_1,0), (\xi_1,1), \dots, (\xi_n,0), (\xi_n,1) \rangle$$

have the same quantifier free type in $L(T(S))$. Now let the canonical skeleton of $\mathcal{A}^1(S)$ given by (F1) be $\{a_x \mid x \in L(T(S))\}$. Define the $T(S)$-skeleton of $\mathcal{A}^1(S)$ to be the set

$$\{a_{(\eta,0)}{}^\frown a_{(\eta,1)} \mid \eta \in T(S)\}.$$

Let us denote $b_\eta = a_{(\eta,0)}{}^\frown a_{(\eta,1)}$. This guarantees that (a), (b) and (c) are satisfied.

For (d), suppose that the order $L(T(S))$ is definable in $\mathcal{A}(S)$ by the formula $\psi(\overline{u}, \overline{c})$, i.e., $\mathcal{A}(S) \models \psi(a_x, a_y) \Leftrightarrow x < y$ for $x, y \in L(T(S))$. Let $\varphi(x_0, x_1, y_0, y_1)$ be the formula

$$\psi(x_0, y_0) \wedge \psi(y_1, x_1).$$

Suppose $\eta, \nu \in T(S)$ are such that $T(S) \models P_\lambda(\eta) \wedge P_\alpha(\nu)$. Then

$$\varphi((a_\nu, 0), (a_\nu, 1), (a_\eta, 0), (a_\eta, 1))$$

holds in $\mathcal{A}(S)$ if and only if $\nu <_{T(S)} \eta$. $\square_{\text{Claim 4}}$

Claim 5 Suppose that $S \mapsto \mathcal{A}(S)$ is a function as described in Claim 4 with identical notation. Suppose further that $S, S' \subset S^\kappa_\lambda$. Then $S \triangle S'$ is non-stationary if and only if $\mathcal{A}(S) \cong \mathcal{A}(S')$.

Proof of Claim 5. Suppose $S \triangle S'$ is non-stationary. Then, by Claim 3, $T(S) \cong T(S')$, which implies $L(T(S)) \cong L(T(S'))$ (defined in the proof of Claim 4), which in turn implies $\mathcal{A}(S) \cong \mathcal{A}(S')$.

Let us now show that if $S \triangle S'$ is stationary, then $\mathcal{A}(S) \ncong \mathcal{A}(S')$. Let us make a counter assumption, namely that there is an isomorphism

$$f \colon \mathcal{A}(S) \cong \mathcal{A}(S')$$

and that $S \triangle S'$ is stationary, and let us deduce a contradiction. Without loss of generality we may assume that $S \setminus S'$ is stationary. Denote

$$X_0 = S \setminus S'.$$

For all $\alpha < \kappa$ define $T^\alpha(S)$ and $T^\alpha(S')$ by

$$T^\alpha(S) = \{\eta \in T(S) \mid \operatorname{ran} \eta \subset \Phi(S, 0, \beta+1) \text{ for some } \beta < \alpha\}$$

and

$$T^\alpha(S') = \{\eta \in T(S) \mid \operatorname{ran} \eta \subset \Phi(S', 0, \beta+1) \text{ for some } \beta < \alpha\}.$$

Then we have:

(i) if $\alpha < \beta$, then $T^\alpha(S) \subset T^\beta(S)$;
(ii) if γ is a limit ordinal, then $T^\gamma(S) = \bigcup_{\alpha<\gamma} T^\alpha(S)$.

The same holds of course for S'. Note that, if $\alpha \in S \setminus S'$, then there is $\eta \in T^\alpha(S)$ cofinal in $\Phi(S, 0, \alpha)$ but there is no such $\eta \in T^\alpha(S')$ by definition of Φ: a cofinal function η is added only if $\mathrm{cf}^*(\Phi(S', \alpha, \kappa)) = \omega_1$, which is not the case if $\alpha \notin S'$. This is the key to achieving the contradiction.

But the clauses (i), (ii) are not sufficient to carry out the following argument, because we would like to have $|T^\alpha(S)| < \kappa$. That is why we want to define a different kind of filtration for $T(S)$, $T(S')$.

For all $\alpha \in X_0$, fix a function

$$\eta^\alpha_\lambda \in T(S) \tag{5.3}$$

such that $\operatorname{dom} \eta^\alpha_\lambda = \lambda$ and, for all $\beta < \lambda$, $\eta^\alpha_\lambda \restriction \beta \in T^\alpha(S)$ and $\eta^\alpha_\lambda \notin T^\alpha(S)$.

For arbitrary $A \subset T(S) \cup T(S')$, let $\mathrm{cl}_{\mathrm{Sk}}(A)$ be the set $X \subset \mathcal{A}(S) \cup \mathcal{A}(S')$ such that $X \cap \mathcal{A}(S)$ is the Skolem closure of $\{a_\eta \mid \eta \in A \cap T(S)\}$ and $X \cap \mathcal{A}(S')$ the Skolem closure of $\{a_\eta \mid \eta \in A \cap T(S')\}$. The following is easily verified:

There exists a λ-cub set C and a set $K^\alpha \subset T^\alpha(S) \cup T^\alpha(S')$ for each $\alpha \in C$ such that

(i') If $\alpha < \beta$, then $K^\alpha \subset K^\beta$;
(ii') If γ is a limit ordinal in C, then $K^\gamma = \bigcup_{\alpha \in C \cap \gamma} K^\alpha$;
(iii) for all $\beta < \alpha$, $\eta^\beta_\lambda \in K^\alpha$ (see (1) above);
(iv) $|K^\alpha| = \lambda$;
(v) $\mathrm{cl}_{\mathrm{Sk}}(K^\alpha)$ is closed under $f \cup f^{-1}$;
(vi) $\{\eta \in T^\alpha(S) \cup T^\alpha(S') \mid \operatorname{dom} \eta < \lambda\} \subset K^\alpha$;
(vii) K^α is downward closed.

Denote $K^\kappa = \bigcup_{\alpha < \kappa} K^\alpha$. Clearly K^κ is closed under $f \cup f^{-1}$ and so f is an isomorphism between $\mathcal{A}(S) \cap \mathrm{cl}_{\mathrm{Sk}}(K^\kappa)$ and $\mathcal{A}(S') \cap \mathrm{cl}_{\mathrm{Sk}}(K^\kappa)$. We will derive a contradiction from this, i.e., we will actually show that $\mathcal{A}(S) \cap \mathrm{cl}_{\mathrm{Sk}}(K^\kappa)$ and $\mathcal{A}(S') \cap \mathrm{cl}_{\mathrm{Sk}}(K^\kappa)$ cannot be isomorphic by f. Clauses (iii), (v), (vi) and (vii) guarantee that all elements we are going to deal with will be in K^κ.

Let $X_1 = X_0 \cap C$. For $\alpha \in X_1$, let us use the following abbreviations:

- By $\mathcal{A}_\alpha(S)$ we denote the Skolem closure of $\{a_\eta \mid \eta \in K^\alpha \cap T(S)\}$.
- By $\mathcal{A}_\alpha(S')$ we denote the Skolem closure of $\{a_\eta \mid \eta \in K^\alpha \cap T(S')\}$.
- $K^\alpha(S) = K^\alpha \cap T(S)$.
- $K^\alpha(S') = K^\alpha \cap T(S')$.

In the following we will often deal with finite sequences. When defining such a sequence we will use a bar, but afterwards we will not use the bar in the notation (e.g., let $a = \overline{a}$ be a finite sequence...).

Suppose $\alpha \in X_1$. Choose

$$\xi^\alpha_\lambda = \overline{\xi}^\alpha_\lambda \in T(S') \tag{5.4}$$

to be such that for some (finite sequence of) terms $\pi = \overline{\pi}$ we have

$$\begin{aligned} f(a_{\eta^\alpha_\lambda}) &= \pi(a_{\xi^\alpha_\lambda}) \\ &= \bigl\langle \pi_1\bigl(a_{\xi^\alpha_\lambda(1)}, \dots, a_{\xi^\alpha_\lambda(\mathrm{len}(\overline{\xi}^\alpha_\lambda))}\bigr), \dots, \pi_{\mathrm{len}\,\overline{\pi}}\bigl(a_{\xi^\alpha_\lambda(1)}, \dots, a_{\xi^\alpha_\lambda(\mathrm{len}(\overline{\xi}^\alpha_\lambda))}\bigr)\bigr\rangle. \end{aligned}$$

Note that ξ^α_λ is in K^κ by the definition of K^α's.

Let us denote by η^α_β the element $\eta^\alpha_\lambda \restriction \beta$, and let

$$\xi^\alpha_* = \{\nu \in T(S') \mid \exists \xi \in \xi^\alpha_\lambda (\nu < \xi)\}. \tag{5.5}$$

Also note that $\xi_*^\alpha \subset K^\beta$ for some β. Next define the function $g\colon X_1 \to \kappa$ as follows. Suppose $\alpha \in X_1$. Let $g(\alpha)$ be the smallest ordinal β such that $\xi_*^\alpha \cap K^\alpha(S') \subset K^\beta(S')$. We claim that $g(\alpha) < \alpha$. Clearly $g(\alpha) \leqslant \alpha$, so suppose that $g(\alpha) = \alpha$. Since ξ_λ^α is finite, there must be a $\xi_\lambda^\alpha(i) \in \xi_\lambda^\alpha$ such that for all $\beta < \alpha$ there exists γ such that $\xi_\lambda^\alpha(i) \restriction \gamma \in K^\alpha(S') \setminus K^\beta(S')$, i.e., $\xi_\lambda^\alpha(i)$ is cofinal in $\Phi(S', 0, \alpha)$, which it cannot be, because $\alpha \notin S'$.

Now by Fodor's lemma there exists a stationary set

$$X_2 \subset X_1$$

and γ_0 such that $g[X_2] = \{\gamma_0\}$.

Since there are only $< \kappa$ many finite sequences in $K_{\gamma_0}(S')$, there is a stationary set

$$X_3 \subset X_2$$

and a finite sequence $\xi = \bar\xi \in K^{\gamma_0}(S')$ such that, for all $\alpha \in X_3$, we have $\xi_*^\alpha \cap K^{\gamma_0}(S') = \xi_*$, where ξ_* is the set

$$\xi_* = \{\nu \in T(S') \mid \nu \leqslant \zeta \text{ for some } \zeta \in \bar\xi\} \subset K^{\gamma_0}(S').$$

Let us fix a (finite sequence of) term(s) $\pi = \bar\pi$ such that the set

$$X_4 = \{\alpha \in X_3 \mid f(a_{\eta_\lambda^\alpha}) = \pi(a_{\xi_\lambda^\alpha})\}$$

is stationary; see (5.3). Here $f(\bar a)$ means $\langle f(a_1), \dots, f(a_{\operatorname{len} \bar a})\rangle$ and $\bar\pi(\bar b)$ means

$$\langle \pi_1(b_1, \dots, b_{\operatorname{len} \bar a}), \dots, \pi_{\operatorname{len} \pi}(b_1, \dots, b_{\operatorname{len} \bar a})\rangle.$$

We can find π because there are only countably many such finite sequences of terms.

We claim that in $T(S')$ there are at most λ many quantifier free types over ξ_*. All types from now on are quantifier free. Let us show that there are at most λ many 1-types; the general case is left to the reader. To see this, note that a type p over ξ_* is described by the triple

$$(\nu_p, \beta_p, m_p) \tag{5.6}$$

defined as follows: if η satisfies p, then ν_p is the maximal element of ξ_* that is an initial segment of η, while β_p is the level of η and m_p tells how many elements of $\xi_* \cap P_{\operatorname{dom} \nu_p + 1}$ are there $\lessdot$-below $\eta(\operatorname{dom} \nu_p)$ (recall the vocabulary from Definition 5.7, page 531).

Since $\nu_p \in \xi_*$ and ξ_* is of size λ, $\beta_p \in (\lambda + 1) \cup \{\infty\}$ and $m_p < \omega$, there can be at most λ such triples.

Recall the notations (5.3), (5.4) and (5.5) above. We can pick ordinals $\alpha < \alpha'$, $\alpha, \alpha' \in X_4$, a term τ and an ordinal $\beta < \lambda$ such that

$$\eta_\beta^{\alpha'} \neq \eta_\beta^\alpha,$$

$$f(a_{\eta_\beta^\alpha}) = \tau(a_{\xi_\beta^\alpha}) \text{ and } f(a_{\eta_\beta^{\alpha'}}) = \tau(a_{\xi_\beta^{\alpha'}}) \text{ for some } \xi_\beta^\alpha, \xi_\beta^{\alpha'},$$

$$\operatorname{tp}(\xi_\lambda^\alpha/\xi_*) = \operatorname{tp}(\xi_\lambda^{\alpha'}/\xi_*),$$

and

$$\operatorname{tp}(\xi_\beta^\alpha/\xi_*) = \operatorname{tp}(\xi_\beta^{\alpha'}/\xi_*). \tag{5.7}$$

We claim that then in fact

$$\operatorname{tp}(\xi_\beta^\alpha/(\xi_* \cup \{\xi_\lambda^{\alpha'}\})) = \operatorname{tp}(\xi_\beta^{\alpha'}/(\xi_* \cup \{\xi_\lambda^{\alpha'}\})).$$

Let us show this. Denote

$$p = \operatorname{tp}(\xi_\beta^\alpha/(\xi_* \cup \{\xi_\lambda^{\alpha'}\}))$$

and

$$p' = \operatorname{tp}(\xi_\beta^{\alpha'}/(\xi_* \cup \{\xi_\lambda^{\alpha'}\})).$$

By the assumption (5.7), however, $p \restriction \xi_* = p' \restriction \xi_*$, so, because it is a tree, it suffices to show that $p \restriction \{\xi_\lambda^{\alpha'}\} = p' \restriction \{\xi_\lambda^{\alpha'}\}$. Since α and α' are in X_3 and X_2, we have $\xi_*^{\alpha'} \cap K^{\alpha'}(S') = \xi_*^{\alpha} \cap K^{\alpha}(S') = \xi_* \subset K^{\gamma_0}(S')$. On the other hand, $f \restriction \mathcal{A}_{\alpha'}(S)$ is an isomorphism between $\mathcal{A}_{\alpha'}(S)$ and $\mathcal{A}_{\alpha'}(S')$, because α and α' are in X_1, so $\xi_\beta^\alpha, \xi_\beta^{\alpha'} \in K^{\alpha'}(S')$. Thus ξ_β^α and $\xi_\beta^{\alpha'}$ are either both in ξ_*, whence they are the same, or not, in which case they are both not below $\xi_\lambda^{\alpha'}$. From (5.7) it follows that ξ_β^α and $\xi_\beta^{\alpha'}$ are on the same level and if $\xi_\lambda^{\alpha'}$ is also on the same level, then the above also implies that they are both $\lessdot$-below $\xi_\lambda^{\alpha'}$. From (5.7) and the above we also have that $h(\xi_\beta^\alpha, \xi^{\alpha'}) = h(\xi_\beta^{\alpha'}, \xi^{\alpha'})$; see Definition 5.7.

Now we have: ξ_λ^α and π are such that $f(a_{\eta_\lambda^\alpha}) = \pi(a_{\xi_\lambda^\alpha})$, and ξ_β^α and τ are such that $f(a_{\eta_\beta^\alpha}) = \tau(a_{\xi_\beta^\alpha})$. Similarly for α'. The formula φ is defined in Claim 4.

We know that

$$\mathcal{A}(S) \models \varphi(a_{\eta_\lambda^{\alpha'}}, a_{\eta_\beta^{\alpha'}})$$

and, because f is an isomorphism, this implies

$$\mathcal{A}(S') \models \varphi(f(a_{\eta_\lambda^{\alpha'}}), f(a_{\eta_\beta^{\alpha'}})),$$

which is equivalent to

$$\mathcal{A}(S') \models \varphi(\pi(a_{\xi_\lambda^{\alpha'}}), \tau(a_{\xi_\beta^{\alpha'}}))$$

because α, α' are in X_4. Since $T(S')$ is indiscernible in $\mathcal{A}(S')$ and $\xi_\beta^{\alpha'}$ and ξ_β^α have the same type over over $(\xi_* \cup \{\xi_\lambda^{\alpha'}\})$, we have

$$\mathcal{A}(S') \models \varphi(\pi(a_{\xi_\lambda^{\alpha'}}), \tau(a_{\xi_\beta^{\alpha'}})) \iff \varphi(\pi(a_{\xi_\lambda^{\alpha'}}), \tau(a_{\xi_\beta^{\alpha}})) \tag{5.8}$$

and so we get

$$\mathcal{A}(S') \models \varphi(\pi(a_{\xi_\lambda^{\alpha'}}), \tau(a_{\xi_\beta^{\alpha}})),$$

which is equivalent to

$$\mathcal{A}(S') \models \varphi(f(a_{\eta_\lambda^{\alpha'}}), f(a_{\eta_\beta^{\alpha}})),$$

and this is in turn equivalent to

$$\mathcal{A}(S) \models \varphi(a_{\eta_\lambda^{\alpha'}}, a_{\eta_\beta^{\alpha}}).$$

However, the latter cannot be true, because the definition of β, α and α' implies that $\eta_\beta^{\alpha'} \neq \eta_\beta^\alpha$. $\square_{\text{Claim 5}}$

Thus, the above Claims 1–5 justify the embedding of $E_{S^\kappa_\lambda}$ into the isomorphism relation on the set of structures that are models for T for unstable T. This embedding combined with a suitable coding of models gives a continuous map.

DOP and OTOP cases. The above proof was based on the fact (F1) that for unstable theories there are Ehrenfeucht–Mostowski models for any linear order such that the order is definable by a first-order formula φ and is indiscernible relative to $L_{\omega\omega}$ (see (c) on page 531); it is used in (5.8) above. For the OTOP case, we use instead the fact (F2):

(F2) Suppose that T is a theory with OTOP in a countable vocabulary v. Then for each dense linear order η we can find a model $\mathcal{A}$ of a countable vocabulary $v_1 \supset v$ such that $\mathcal{A}$ is an Ehrenfeucht–Mostowski model of T for η where order is definable by an $L_{\omega_1\omega}$-formula.[3]

Since the order $\Phi(S)$ is dense, it is easy to argue that if $T(S)$ is indiscernible relative to $L_{\omega\omega}$, then it is indiscernible relative to $L_{\infty\omega}$ (define this as in (c) on page 531 changing tp to $\mathrm{tp}_{L_{\infty\omega}}$). Other parts of the proof remain unchanged, because although the formula φ is not first-order anymore, it is still in $L_{\infty\omega}$.

In the DOP case we have the following fact:

(F3) Let T be a countable superstable theory with DOP of vocabulary v. Then there exists a vocabulary $v_1 \supset v$, $|v_1| = \omega_1$, such that for every linear order η there exists a v_1-model $\mathcal{A}$ which is an Ehrenfeucht–Mostowski model of T for η where order is definable by an $L_{\omega_1\omega_1}$-formula.[4]

Now the problem is that φ is in $L_{\infty\omega_1}$. By (c) of Claim 4, $T(S)$ is indiscernible in $\mathcal{A}(S)$ relative to $L_{\omega\omega}$ and by the above relative to $L_{\infty\omega}$. If we could require $\Phi(S)$ to be ω_1-dense, we would similarly get indiscernible relative to $L_{\infty\omega_1}$. Let us show how to modify the proof in order to do that. Recall that, in the DOP case, we assume $\lambda \geqslant 2^\omega$.

In Claim 1, page 529, we have to replace clauses (3), (4) and (6) by (3'), (4') and (6'):

(3') $\eta \cong \eta \cdot \mu + \eta \cdot \omega^*$;

(4') η is ω_1-dense;

(6') $\mathrm{cf}^*(\eta) = \omega_1$.

The proof that such an η exists is exactly as the proof of [**10**, Lemma 7.17] except that, instead of putting $\mu = (\omega_1)^V$, put $\mu = \omega$, build θ-many functions with domains being countable initial segments of ω_1 instead of finite initial segments of ω and instead of $\mathbb{Q}$ (the countable dense linear order) use an ω_1-saturated dense linear order —this order has size 2^ω and that is why the assumption $\lambda \geqslant 2^\omega$ is needed.

In the definition of $\Phi(S)$ (right after Claim 1), replace ω_1^* by ω^* and η by the new η satisfying (3'), (4') and (6') above. Note that $\Phi(S)$ becomes now ω_1-dense. In Claim 2 one has to replace ω_1^* by ω^*. The proof remains similar. In the proof of Claim 3 (page 530) one has to adjust the use of Claim 2. Then, in the definition of $T(S)$ replace ω_1 by ω.

Claim 4 for superstable T with DOP now follows with (c) and (d) modified: instead of indiscernible relative to $L_{\omega\omega}$, demand $L_{\infty\omega_1}$ and instead of $\varphi \in L_{\omega\omega}$ we have $\varphi \in L_{\infty\omega_1}$. The proof is unchanged except that the language is replaced by $L_{\infty\omega_1}$ everywhere and fact (F1) is replaced by (F3) above.

Everything else in the proof, in particular the proof of Claim 5, remains unchanged modulo some obvious things that are evident from the above explanation. $\square_{\text{Theorem 5.6}}$

5.3 Stable unsuperstable theories

In this section we provide a tree construction (Lemma 5.12) which is similar to Shelah's construction in [**29**], which he used to obtain (via Ehrenfeucht–Mostowski models) many pairwise non-isomorphic models. Then using a prime-model construction (proof of Theorem 5.13, page 542) we will obtain the needed result.

[3] Contained in the proof of [**27**, Theorem 2.5]; see also [**15**, Theorem 6.6].

[4] This is essentially from [**28**, Fact 2.5B]; a proof can be found also in [**15**, Theorem 6.1].

Definition 5.8 Let I be a tree of size κ. Suppose $(I_\alpha)_{\alpha<\kappa}$ is a collection of subsets of I such that the following hold:

- For each $\alpha < \kappa$, I_α is a downward closed subset of I.
- $\bigcup_{\alpha<\kappa} I_\alpha = I$.
- If $\alpha < \beta < \kappa$, then $I_\alpha \subset I_\beta$.
- If γ is a limit ordinal, then $I_\gamma = \bigcup_{\alpha<\gamma} I_\alpha$.
- For each $\alpha < \kappa$ the cardinality of I_α is less than κ.

Such a sequence $(I_\alpha)_{\alpha<\kappa}$ is called *κ-filtration* or just *filtration* of I.

Definition 5.9 Recall K^λ_{tr} from Definition 5.7. Let $K^\lambda_{\mathrm{tr}*} = \{A \restriction L^* \mid A \in K^\lambda_{\mathrm{tr}}\}$, where L^* is the vocabulary $\{<\}$.

Definition 5.10 Suppose $t \in K^\omega_{\mathrm{tr}*}$ is a tree of size κ (i.e., $t \subset \kappa^{\leqslant\omega}$), and let $\mathcal{I} = (I_\alpha)_{\alpha<\kappa}$ be a filtration of t. Define

$$S_{\mathcal{I}}(t) = \big\{\alpha < \kappa \mid (\exists \eta \in t)\big[(\operatorname{dom}\eta = \omega) \wedge \forall n < \omega(\eta\restriction n \in I_\alpha) \wedge (\eta \notin I_\alpha)\big]\big\}.$$

By $S \sim_{\mathrm{NS}} S'$ we mean that $S \bigtriangleup S'$ is not ω-stationary.

Lemma 5.11 *Suppose that t_0 and t_1 are isomorphic trees, and that $\mathcal{I} = (I_\alpha)_{\alpha<\kappa}$ and $\mathcal{J} = (J_\alpha)_{\alpha<\kappa}$ are κ-filtrations of t_0 and t_1 respectively. Then $S_{\mathcal{I}}(t_0) \sim_{\mathrm{NS}} S_{\mathcal{J}}(t_1)$.*

Proof. Let $f\colon t_0 \to t_1$ be an isomorphism. Then $f\mathcal{I} = (f[I_\alpha])_{\alpha<\kappa}$ is a filtration of t_1 and

$$\alpha \in S_{\mathcal{I}}(t_0) \iff \alpha \in S_{f\mathcal{I}}(t_1). \tag{5.9}$$

Define the set $C = \{\alpha \mid f[I_\alpha] = J_\alpha\}$. Let us show that it is cub. Let $\alpha \in \kappa$. Define $\alpha_0 = \alpha$ and by induction pick $(\alpha_n)_{n<\omega}$ such that $f[I_{\alpha_n}] \subset J_{\alpha_{n+1}}$ for odd n and $J_{\alpha_n} \subset f[I_{\alpha_{n+1}}]$ for even n. This is possible by the definition of a κ-filtration. Then $\alpha_\omega = \bigcup_{n<\omega} \alpha_n \in C$. Clearly C is closed and $C \subset \kappa \setminus S_{f\mathcal{I}}(t_1) \bigtriangleup S_{\mathcal{J}}(t_1)$, so now, by (5.9),

$$S_{\mathcal{I}}(t_0) = S_{f\mathcal{I}}(t_1) \sim_{\mathrm{NS}} S_{\mathcal{J}}(t_1). \qquad \square$$

Lemma 5.12 *Suppose that for $\lambda < \kappa$, $\lambda^\omega < \kappa$ and $\kappa^{<\kappa} = \kappa$. Then there exists a function $J\colon \mathcal{P}(\kappa) \to K^\omega_{\mathrm{tr}*}$ such that:*

- *For all $S \subset \kappa$, $|J(S)| = \kappa$.*
- *If $S \subset \kappa$ and $\mathcal{I}$ is a κ filtration of $J(S)$, then $S_{\mathcal{I}}(J(S)) \sim_{\mathrm{NS}} S$.*
- *If $S_0 \sim_{\mathrm{NS}} S_1$, then $J(S_0) \cong J(S_1)$.*

Proof. Let $S \subset S^\kappa_\omega$ and let us define a preliminary tree $I(S)$ as follows. For each $\alpha \in S$, let C_α be the set of all strictly increasing cofinal functions $\eta\colon \omega \to \alpha$. Let $I(S) = \underline{[\kappa]}^{<\omega} \cup \bigcup_{\alpha\in S} C_\alpha$ where $\underline{[\kappa]}^{<\omega}$ is the set of strictly increasing functions from finite ordinals to κ.

For ordinals $\alpha < \beta \leqslant \kappa$ and $i < \omega$, we adopt the notation

- $[\alpha,\beta] = \{\gamma \mid \alpha \leqslant \gamma \leqslant \beta\}$;
- $[\alpha,\beta) = \{\gamma \mid \alpha \leqslant \gamma < \beta\}$;
- $\tilde{f}(\alpha,\beta,i) = \bigcup_{i\leqslant j\leqslant\omega}\{\eta\colon [i,j) \to [\alpha,\beta) \mid \eta \text{ strictly increasing}\}$.

For each $\alpha,\beta < \kappa$ let us define the sets $P^{\alpha,\beta}_\gamma$, for $\gamma < \kappa$ as follows. If $\alpha = \beta = \gamma = 0$, then $P^{0,0}_0 = I(S)$. Otherwise let $\{P^{\alpha,\beta}_\gamma \mid \gamma < \kappa\}$ enumerate all downward closed subsets of $\tilde{f}(\alpha,\beta,i)$ for all i, i.e.,

$$\{P^{\alpha,\beta}_\gamma \mid \gamma < \kappa\} = \bigcup_{i<\omega} \mathcal{P}(\tilde{f}(\alpha,\beta,i)) \cap \{A \mid A \text{ is closed under inital segments}\}.$$

Define $\tilde{n}(P^{\alpha,\beta}_{\gamma})$ to be the natural number i such that $P^{\alpha,\beta}_{\gamma} \subset \tilde{f}(\alpha,\beta,i)$. The enumeration is possible, because by our assumption $\kappa^{<\kappa} = \kappa$ we have

$$\begin{aligned} \Big|\bigcup_{i<\omega} \mathcal{P}(\tilde{f}(\alpha,\beta,i))\Big| &\leqslant \omega \times |\mathcal{P}(\tilde{f}(0,\beta,0))| \\ &\leqslant \omega \times |\mathcal{P}(\beta^{\omega})| \\ &= \omega \times 2^{\beta^{\omega}} \\ &\leqslant \omega \times \kappa \\ &= \kappa. \end{aligned}$$

Let $S \subset \kappa$ be a set and define $J(S)$ to be the set of all $\eta\colon s \to \omega \times \kappa^4$ such that $s \leqslant \omega$ and the following conditions are met for all $i, j < s$:

(1) η is strictly increasing with respect to the lexicographical order on $\omega \times \kappa^4$;
(2) $\eta_1(i) \leqslant \eta_1(i+1) \leqslant \eta_1(i) + 1$;
(3) $\eta_1(i) = 0 \to \eta_2(i) = \eta_3(i) = \eta_4(i) = 0$;
(4) $\eta_1(i) < \eta_1(i+1) \to \eta_2(i+1) \geqslant \eta_3(i) + \eta_4(i)$;
(5) $\eta_1(i) = \eta_1(i+1) \to (\forall k \in \{2,3,4\})(\eta_k(i) = \eta_k(i+1))$;
(6) if for some $k < \omega$, $[i,j) = \eta_1^{-1}\{k\}$, then $\eta_5 \restriction [i,j) \in P^{\eta_2(i),\eta_3(i)}_{\eta_4(i)}$;
(7) if $s = \omega$, then either

$$(\exists m < \omega)(\forall k < \omega)(k > m \to \eta_1(k) = \eta_1(k+1))$$

or $\sup \operatorname{ran} \eta_5 \in S$;
(8) order $J(S)$ by inclusion.

Note that it follows from the definition of $P^{\alpha,\beta}_{\gamma}$ and conditions (6) and (4) that, for all $i < j < \operatorname{dom}\eta$, $\eta \in J(S)$,

(9) $i < j \to \eta_5(i) < \eta_5(j)$.

For each $\alpha < \kappa$, let

$$J^{\alpha}(S) = \{\eta \in J(S) \mid \operatorname{ran}\eta \subset \omega \times (\beta+1)^4 \text{ for some } \beta < \alpha\}.$$

Then $(J^{\alpha}(S))_{\alpha<\kappa}$ is a κ-filtration of $J(S)$ (see Claim 2 below). For the first item of the lemma, clearly $|J(S)| = \kappa$.

Let us observe that if $\eta \in J(S)$ and $\operatorname{ran}\eta_1 = \omega$, then

$$\sup \operatorname{ran} \eta_4 \leqslant \sup \operatorname{ran} \eta_2 = \sup \operatorname{ran} \eta_3 = \sup \operatorname{ran} \eta_5 \tag{5.10}$$

and if, in addition to that, $\eta \restriction k \in J^{\alpha}(S)$ for all k and $\eta \notin J^{\alpha}(S)$ or if $\operatorname{ran}\eta_1 = \{0\}$, then

$$\sup \operatorname{ran} \eta_5 = \alpha. \tag{5.11}$$

To see (5.10), suppose that $\operatorname{ran}\eta_1 = \omega$. By (9), $(\eta_5(i))_{i<\omega}$ is an increasing sequence. By (6), $\sup \operatorname{ran} \eta_3 \geqslant \sup \operatorname{ran} \eta_5 \geqslant \sup \operatorname{ran} \eta_2$. By (4), $\sup \operatorname{ran} \eta_2 \geqslant \sup \operatorname{ran} \eta_3$, and again by (4), $\sup \operatorname{ran} \eta_2 \geqslant \sup \operatorname{ran} \eta_4$. The inequality $\sup \operatorname{ran} \eta_5 \leqslant \alpha$ is an immediate consequence of the definition of $J^{\alpha}(S)$, so (5.11) follows now from the assumption that $\eta \notin J^{\alpha}(S)$.

Claim 1 Suppose that $\xi \in J^{\alpha}(S)$ and $\eta \in J(S)$. Then, if $\operatorname{dom}\xi < \omega$, $\xi \subsetneq \eta$ and $(\forall k \in \operatorname{dom}\eta \setminus \operatorname{dom}\xi)\big(\eta_1(k) = \xi_1(\max \operatorname{dom}\xi) \wedge \eta_1(k) > 0\big)$, it follows that $\eta \in J^{\alpha}(S)$.

Proof of Claim 1. Suppose that $\xi, \eta \in J^{\alpha}(S)$ are as in the assumption, and let us define $\beta_2 = \xi_2(\max \operatorname{dom}\xi)$, $\beta_3 = \xi_2(\max \operatorname{dom}\xi)$, and $\beta_4 = \xi_4(\max \operatorname{dom}\xi)$. Because $\xi \in J^{\alpha}(S)$, there is β such that $\beta_2, \beta_3, \beta_4 < \beta + 1$ and $\beta < \alpha$. Now by (5) $\eta_2(k) = \beta_2$, $\eta_3(k) = \beta_3$ and $\eta_4(k) = \beta_4$, for all $k \in \operatorname{dom}\eta \setminus \operatorname{dom}\xi$. Then by (6) for all $k \in \operatorname{dom}\eta \setminus \operatorname{dom}\xi$ we have that $\beta_2 < \eta_5(k) < \beta_3 < \beta + 1$. Since $\xi \in J^\alpha(S)$, also $\beta_4 < \beta + 1$, so $\eta \in J^\alpha(S)$. $\square_{\text{Claim 1}}$

Claim 2 $|J(S)| = \kappa$, $(J^\alpha(S))_{\alpha<\kappa}$ is a κ-filtration of $J(S)$, and if $S \subset \kappa$ and $\mathcal{I}$ is a κ-filtration of $J(S)$ then $S_{\mathcal{I}}(J(S)) \sim_{\text{NS}} S$.

Proof of Claim 2. For all $\alpha \leqslant \kappa$, $J^\alpha(S) \subset (\omega \times \alpha^4)^{\leqslant\omega}$, so by the cardinality assumption of the lemma, the cardinality of $J^\alpha(S)$ is $< \kappa$ if $\alpha < \kappa$ ($J^\kappa(S) = J(S)$). Clearly $\alpha < \beta$ implies $J^\alpha(S) \subset J^\beta(S)$. Continuity is verified by

$$\begin{aligned}\bigcup_{\alpha<\gamma} J^\alpha(S) &= \{\eta \in J(S) \mid \exists\alpha < \gamma, \exists\beta < \alpha(\operatorname{ran}\eta \subset \omega \times (\beta+1)^4)\}\\ &= \{\eta \in J(S) \mid \exists\beta < \cup\gamma(\operatorname{ran}\eta \subset \omega \times (\beta+1)^4)\},\end{aligned}$$

which equals $J^\gamma(S)$ if γ is a limit ordinal. By Lemma 5.11 it is enough to show that $S_{\mathcal{I}}(J(S)) \sim_{\text{NS}} S$ for $\mathcal{I} = (J^\alpha(S))_{\alpha<\kappa}$, and we will show that if $\mathcal{I} = (J^\alpha(S))_{\alpha<\kappa}$, then in fact $S_{\mathcal{I}}(J(S)) = S$.

Suppose $\alpha \in S_{\mathcal{I}}(J(S))$. Then there is $\eta \in J(S)$, $\operatorname{dom}\eta = \omega$, such that $\eta \restriction k \in J^\alpha(S)$ for all $k < \omega$ but $\eta \notin J^\alpha(S)$. Thus there is no $\beta < \alpha$ such that $\operatorname{ran}\eta \subset \omega \times (\beta+1)^4$ but on the other hand for all $k < \omega$ there is β such that $\operatorname{ran}\eta \restriction k \subset \omega \times (\beta+1)^4$. By (5) and (6) this implies that either $\operatorname{ran}\eta_1 = \omega$ or $\operatorname{ran}\eta_1 = \{0\}$. By (5.11) on page 538 it now follows that $\sup\operatorname{ran}\eta_5 = \alpha$ and, by (7), $\alpha \in S$.

Suppose then $\alpha \in S$. Let us show that $\alpha \in S_{\mathcal{I}}(J(S))$. Fix a function $\eta_\alpha \colon \omega \to \kappa$ with $\sup\operatorname{ran}\eta_\alpha = \alpha$. Then $\eta_\alpha \in I(S)$ and the function η such that $\eta(n) = (0,0,0,0,\eta_\alpha(n))$ is as required. (Recall that $P_0^{0,0} = I(S)$ in the definition of $J(S)$.) $\square_{\text{Claim 2}}$

Claim 3 Suppose that $S \sim_{\text{NS}} S'$. Then $J(S) \cong J(S')$.

Proof of Claim 3. Let $C \subset \kappa \setminus (S \triangle S')$ be the cub set which exists by the assumption. By induction on $i < \kappa$ we will define α_i and F_{α_i} such that:

(a) If $i < j < \kappa$, then $\alpha_i < \alpha_j$ and $F_{\alpha_i} \subset F_{\alpha_j}$.
(b) If i is a successor, then α_i is a successor and if i is limit, then $\alpha_i \in C$.
(c) If γ is a limit ordinal, then $\alpha_\gamma = \sup_{i<\gamma}\alpha_i$.
(d) F_{α_i} is a partial isomorphism $J(S) \to J(S')$.
(e) Suppose that $i = \gamma + n$, where γ is a limit ordinal or 0 and $n < \omega$ is even. Then $\operatorname{dom}F_{\alpha_i} = J^{\alpha_i}(S)$ (e1). If also $n > 0$ and $(\eta_k)_{k<\omega}$ is an increasing sequence in $J^{\alpha_i}(S)$ such that $\eta = \bigcup_{k<\omega}\eta_k \notin J(S)$, then $\bigcup_{k<\omega}F_{\alpha_i}(\eta_k) \notin J(S')$ (e2).
(f) If $i = \gamma + n$, where γ is a limit ordinal or 0 and $n < \omega$ is odd, then $\operatorname{ran}F_{\alpha_i} = J^{\alpha_i}(S')$ (f1). Further, if $(\eta_k)_{k<\omega}$ is an increasing sequence in $J^{\alpha_i}(S')$ such that $\eta = \bigcup_{k<\omega}\eta_k \notin J(S')$, then $\bigcup_{k<\omega}F_{\alpha_i}^{-1}(\eta_k) \notin J(S)$ (f3).
(g) If $\operatorname{dom}\xi < \omega$, $\xi \in \operatorname{dom}F_{\alpha_i}$, $\eta \restriction \operatorname{dom}\xi = \xi$ and $(\forall k \geqslant \operatorname{dom}\xi)\big(\eta_1(k) = \xi_1(\max\operatorname{dom}\xi) \wedge \eta_1(k) > 0\big)$, then $\eta \in \operatorname{dom}F_{\alpha_i}$. Similarly for $\operatorname{ran}F_{\alpha_i}$.
(h) If $\xi \in \operatorname{dom}F_{\alpha_i}$ and $k < \operatorname{dom}\xi$, then $\xi \restriction k \in \operatorname{dom}F_{\alpha_i}$.
(i) For all $\eta \in \operatorname{dom}F_{\alpha_i}$, $\operatorname{dom}\eta = \operatorname{dom}(F_{\alpha_i}(\eta))$.

First step. The first step and the successor steps are similar, but the first step is easier. Thus we give it separately in order to simplify the readability. Let us start with $i = 0$.

Let $\alpha_0 = \beta + 1$, for arbitrary $\beta \in C$. Let us denote by

$$\tilde{o}(\alpha)$$

the ordinal that is order isomorphic to $(\omega \times \alpha^4, <_{\text{lex}})$. Let γ be such that there is an isomorphism $h\colon P_\gamma^{0,\tilde{o}(\alpha_0)} \cong J^{\alpha_0}(S)$ and such that $\tilde{n}(P_\gamma^{0,\alpha_0}) = 0$, which exists by (1). Suppose that $\eta \in J^{\alpha_0}(S)$. Note that because P_γ^{0,α_0} and $J^{\alpha_0}(S)$ are closed under initial segments and by the definitions of $\tilde{n}$ and $P_\gamma^{\alpha,\beta}$, we have $\operatorname{dom} h^{-1}(\eta) = \operatorname{dom} \eta$. Define $\xi = F_{\alpha_0}(\eta)$ such that $\operatorname{dom} \xi = \operatorname{dom} \eta$ and, for all $k < \operatorname{dom} \xi$,

- $\xi_1(k) = 1$;
- $\xi_2(k) = 0$;
- $\xi_3(k) = \tilde{o}(\alpha_0)$;
- $\xi_4(k) = \gamma$;
- $\xi_5(k) = h^{-1}(\eta)(k)$.

Let us check that $\xi \in J(S')$. Conditions (1)–(5) and (7) are satisfied because ξ_k is constant for all $k \in \{1, 2, 3, 4\}$, $\xi_1(i) \neq 0$ for all i and ξ_5 is increasing. For (6), if $\xi_1^{-1}\{k\}$ is empty, the condition is verified since each $P_\gamma^{\alpha,\beta}$ is closed under initial segments and contains the empty function. If it is non-empty, then $k = 1$ and in that case $\xi_1^{-1}\{k\} = [0, \omega)$ and by the argument above $(\operatorname{dom} h^{-1}(\eta) = \operatorname{dom} \eta = \operatorname{dom} \xi)$ we have $\xi_5 = h^{-1}(\eta) \in P_\gamma^{0,\tilde{o}(\alpha_0)} = P_{\xi_4(0)}^{\xi_2(0),\xi_3(0)}$, so the condition is satisfied.

Let us check whether all the conditions (a)–(i) are met. In (a), (b), (c), (e2) and (f) there is nothing to check; (d) holds because h is an isomorphism; (e1) and (i) are immediate from the definition. Both $J^{\alpha_0}(S)$ and $P_\gamma^{0,\tilde{o}(\alpha_0)}$ are closed under initial segments, so (h) follows, because $\operatorname{dom} F_{\alpha_0} = J^{\alpha_0}(S)$ and $\operatorname{ran} F_{\alpha_0} = \{1\} \times \{0\} \times \{\tilde{o}(\alpha_0)\} \times \{\gamma\} \times P_\gamma^{0,\alpha_0}$. Claim 1 implies (g) for $\operatorname{dom} F_{\alpha_0}$. Suppose $\xi \in \operatorname{ran} F_{\alpha_0}$ and $\eta \in J(S')$ are as in the assumption of (g). Then $\eta_1(i) = \xi_1(i) = 1$ for all $i < \operatorname{dom} \eta$. By (5) it follows that $\eta_2(i) = \xi_2(i) = 0$, $\eta_3(i) = \xi_3(i) = \tilde{o}(\alpha_0)$ and $\eta_4(i) = \xi_4(i) = \gamma$ for all $i < \operatorname{dom} \eta$, so by (6) $\eta_5 \in P_\gamma^{0,\tilde{o}(\alpha_0)}$ and, since h is an isomorphism, $\eta \in \operatorname{ran} F_{\alpha_0}$.

Odd successor step. We want to handle the odd case but not the even case first, because the most important case is the successor of a limit ordinal; see $(\iota\iota\iota)$ below. Except for that, the even case is similar to the odd case.

Suppose that $j < \kappa$ is a successor ordinal. Then there exist β_j and n_j such that $j = \beta_j + n_j$ and β is a limit ordinal or 0. Suppose that n_j is odd and that α_l and F_{α_l} are defined for all $l < j$ such that the conditions (a)–(i) and (1)–(9) hold for $l < j$.

Let $\alpha_j = \beta + 1$ where β is such that $\beta \in C$, $\operatorname{ran} F_{\alpha_{j-1}} \subset J^\beta(S')$, $\beta > \alpha_{j-1}$. For convenience define $\xi(-1) = (0, 0, 0, 0, 0)$ for all $\xi \in J(S) \cup J(S')$. Suppose $\eta \in \operatorname{ran} F_{\alpha_{j-1}}$ has finite domain $\operatorname{dom} \eta = m < \omega$ and denote $\xi = F_{\alpha_{j-1}}^{-1}(\eta)$. Fix γ_η to be such that $\tilde{n}(P_{\gamma_\eta}^{\alpha,\beta}) = m$ and such that there is an isomorphism $h_\eta\colon P_{\gamma_\eta}^{\alpha,\beta} \to W$, where

$$W = \{\zeta \mid \operatorname{dom} \zeta = [m, s),\ m < s \leqslant \omega,\ \eta^\frown\langle m, \zeta(m)\rangle \notin \operatorname{ran} F_{\alpha_{j-1}},\ \eta^\frown\zeta \in J^{\alpha_j}(S')\},$$

$\alpha = \xi_3(m-1) + \xi_4(m-1)$ and $\beta = \alpha + \tilde{o}(\alpha_j)$ (defined in the beginning of the first step).

We will define F_{α_j} so that its range is $J^{\alpha_j}(S')$ and instead of F_{α_j} we will define its inverse. So let $\eta \in J^{\alpha_j}(S')$. We have three cases:

(ι) $\eta \in \operatorname{ran} F_{\alpha_{j-1}}$;

$(\iota\iota)$ $\exists m < \operatorname{dom} \eta (\eta \restriction m \in \operatorname{ran} F_{\alpha_{j-1}} \wedge \eta \restriction (m+1) \notin F_{\alpha_{j-1}})$;

$(\iota\iota\iota)$ $\forall m < \operatorname{dom} \eta (\eta \restriction (m+1) \in \operatorname{ran} F_{\alpha_{j-1}} \wedge \eta \notin \operatorname{ran} F_{\alpha_{j-1}})$.

Let us define $\xi = F^{-1}_{\alpha_j}(\eta)$ such that $\operatorname{dom}\xi = \operatorname{dom}\eta$. If ($\iota$) holds, define $\xi(n) = F^{-1}_{\alpha_{j-1}}(\eta)(n)$ for all $n < \operatorname{dom}\eta$. Clearly $\xi \in J(S)$ by the induction hypothesis. Suppose that ($\iota\iota$) holds and let m witness this. For all $n < \operatorname{dom}\xi$, let:

- If $n < m$, then $\xi(n) = F^{-1}_{\alpha_{j-1}}(\eta \restriction m)(n)$.
- Suppose $n \geqslant m$. Let
  - $\xi_1(n) = \xi_1(m-1) + 1$;
  - $\xi_2(n) = \xi_3(m-1) + \xi_4(m-1)$;
  - $\xi_3(n) = \xi_2(m) + \tilde{o}(\alpha_j)$;
  - $\xi_4(n) = \gamma_{\eta\restriction m}$;
  - $\xi_5(n) = h^{-1}_{\eta\restriction m}(\eta)(n)$.

Next we should check that $\xi \in J(S)$; let us check items (1) and (6) —the rest are left to the reader.

(1) By the induction hypothesis $\xi \restriction m$ is increasing. Next, $\xi_1(m) = \xi_1(m-1)+1$, so $\xi(m-1) <_{\text{lex}} \xi(m)$. If $m \leqslant n_1 < n_2$, then $\xi_k(n_1) = \xi_k(n_2)$ for all $k \in \{1,2,3,4\}$ and ξ_5 is increasing.

(6) Suppose that $[i,j) = \xi_1^{-1}\{k\}$. Since $\xi_1 \restriction [m,\omega)$ is constant, either $j < m$, when we are done by the induction hypothesis, or $i = m$ and $j = \omega$. In that case one verifies that $\eta \restriction [m,\omega) \in W = \operatorname{ran} h_{\eta\restriction m}$ and then, imitating the corresponding argument in the first step, that

$$\xi_5 \restriction [m,\omega) = h^{-1}_{\eta\restriction m}(\eta \restriction [m,\omega))$$

and hence $\operatorname{dom} h_{\eta\restriction m} = P^{\xi_2(m),\xi_3(m)}_{\xi_4(m)}$.

Suppose finally that ($\iota\iota\iota$) holds. Then $\operatorname{dom}\eta$ must be ω since otherwise the condition ($\iota\iota\iota$) is simply contradictory (because $\eta \restriction (\operatorname{dom}\eta - 1 + 1) = \eta$, except for the case $\operatorname{dom}\eta = 0$, but then condition ($\iota$) holds and we are done). By (g), we have $\operatorname{ran}\eta_1 = \omega$, because otherwise we had $\eta \in \operatorname{ran} F_{\alpha_{j-1}}$. Let $F^{-1}_{\alpha_j}(\eta) = \xi = \bigcup_{n<\omega} F^{-1}_{\alpha_{j-1}}(\eta \restriction n)$.

Let us check that it is in $J(S)$. Conditions (1)–(6) are satisfied by ξ, because they are satisfied by all its initial segments. Let us check (7).

First of all ξ cannot be in $J^{\alpha_{j-1}}(S)$, since otherwise, by (d) and (i),

$$F_{\alpha_{j-1}}(\xi) = \bigcup_{n<\omega} F_{\alpha_{j-1}}(\xi \restriction n) = \bigcup_{n<\omega} \eta \restriction n = \eta$$

would be again in $\operatorname{ran} F_{\alpha_{j-1}}$. If $j-1$ is a successor ordinal, then we are done: by (b) α_{j-1} is a successor and we assumed $\eta \in J(S')$, so by (e2) we have $\xi \in J(S)$. Thus we can assume that $j-1$ is a limit ordinal. Then by (b), α_{j-1} is a limit ordinal in C and by (a), (e) and (f), $\operatorname{ran} F_{\alpha_{j-1}} = J^{\alpha_{j-1}}(S')$ and $\operatorname{dom} F_{\alpha_{j-1}} = J^{\alpha_{j-1}}(S)$. This implies that $\operatorname{ran}\eta \not\subset \omega \times \beta^4$ for any $\beta < \alpha_{j-1}$ and by (5.11) on page 538 we must have $\sup\operatorname{ran}\eta_5 = \alpha_{j-1}$ which gives $\alpha_{j-1} \in S'$ by (7). Since $\alpha_{j-1} \in C \subset \kappa \setminus S \triangle S'$, we have $\alpha_{j-1} \in S$. Again by (5.11) and that $\operatorname{dom} F_{\alpha_{j-1}} = J^{\alpha_{j-1}}(S)$ by (e1), we have $\sup\operatorname{ran}\xi_5 = \alpha_{j-1}$, thus ξ satisfies the condition (7).

Let us check whether all the conditions (a)–(i) are met: (a), (b), (c) are common to the cases (ι), ($\iota\iota$) and ($\iota\iota\iota$) in the definition of $F^{-1}_{\alpha_j}$ and are easy to verify. Let us sketch a proof for (d); the rest is left to the reader.

(d) Let $\eta_1, \eta_2 \in \operatorname{ran} F_{\alpha_j}$ and let us show that

$$\eta_1 \subsetneq \eta_2 \iff F^{-1}_{\alpha_j}(\eta_1) \subsetneq F^{-1}_{\alpha_j}(\eta_2).$$

The case where both η_1 and η_2 satisfy $(\iota\iota)$ is the interesting one (as it implies all the others). So suppose $\eta_1, \eta_2 \in (\iota\iota)$. Then there exist m_1 and m_2 as described in the statement of $(\iota\iota)$. Let us show that $m_1 = m_2$. We have $\eta_1 \restriction (m_1 + 1) = \eta_2 \restriction (m_1 + 1)$ and $\eta_1 \restriction (m_1 + 1) \notin \operatorname{ran} F_{\alpha_{j-1}}$, so $m_2 \leqslant m_1$. If $m_2 \leqslant m_1$, then $m_2 < \operatorname{dom} \eta_1$, since $m_1 < \operatorname{dom} \eta_1$. Thus if $m_2 \leqslant m_1$, then $\eta_1 \restriction (m_2 + 1) = \eta_2 \restriction (m_2 + 1) \notin \operatorname{ran} F_{\alpha_{j-1}}$, which implies $m_2 = m_1$. According to the definition of $F^{-1}_{\alpha_j}(\eta_i)(k)$ for $k < \operatorname{dom} \eta_1$, $F^{-1}_{\alpha_j}(\eta_i)(k)$ depends only on m_i and $\eta \restriction m_i$ for $i \in \{1, 2\}$. Since $m_1 = m_2$ and $\eta_1 \restriction m_1 = \eta_2 \restriction m_2$, we have $F^{-1}_{\alpha_j}(\eta_1)(k) = F^{-1}_{\alpha_j}(\eta_2)(k)$ for all $k < \operatorname{dom} \eta_1$.

Let us now assume that $\eta_1 \not\subset \eta_2$. Then take the smallest $n \in \operatorname{dom} \eta_1 \cap \operatorname{dom} \eta_2$ such that $\eta_1(n) \neq \eta_2(n)$. It is now easy to show that $F^{-1}_{\alpha_j}(\eta_1)(n) \neq F^{-1}_{\alpha_j}(\eta_2)(n)$ by the construction.

Even successor step. Namely the one where $j = \beta + n$ and n is even. But this case goes exactly as the above completed step, except that we start with $\operatorname{dom} F_{\alpha_j} = J^{\alpha_j}(S)$ where α_j is big enough successor of an element of C such that $J^{\alpha_j}(S)$ contains $\operatorname{ran} F_{\alpha_{j-1}}$ and define $\xi = F_{\alpha_j}(\eta)$. Instead of (e) we use (f) as the induction hypothesis. This step is easier since one does not need to care about the successors of limit ordinals.

Limit step. Assume that j is a limit ordinal. Then let $\alpha_j = \bigcup_{i<j} \alpha_i$ and $F_{\alpha_j} = \bigcup_{i<j} F_{\alpha_i}$. Since α_i are successors of ordinals in C, $\alpha_j \in C$, so (b) is satisfied. Since each F_{α_i} is an isomorphism, also their union is, so (d) is satisfied. Because conditions (e), (f) and (i) hold for $i < j$, the conditions (e) and (i) hold for j. Condition (f) is satisfied because the premise is not true. Conditions (a) and (c) are clearly satisfied. Also (g) and (h) are satisfied by Claim 1 since now $\operatorname{dom} F_{\alpha_j} = J^{\alpha_j}(S)$ and $\operatorname{ran} F_{\alpha_j} = J^{\alpha_j}(S')$ (this is because (a), (e) and (f) hold for $i < j$).

Finally, $F = \bigcup_{i<\kappa} F_{\alpha_i}$ is an isomorphism between $J(S)$ and $J(S')$. $\square_{\text{Claim 3}}$

$\square_{\text{Lemma 5.12}}$

Theorem 5.13 *Suppose that κ is such that $\kappa^{<\kappa} = \kappa$ and for all $\lambda < \kappa$, $\lambda^{\omega} < \kappa$, and that T is a stable unsuperstable theory. Then $E_{S^\kappa_\omega} \leqslant_c \cong_T$.*

Proof. For $\eta \in 2^\kappa$ let $J_\eta = J(\eta^{-1}\{1\})$ where the function J is as in Lemma 5.12 above. For notational convenience, we assume that J_η is a downward closed subtree of $\kappa^{\leqslant\omega}$. Since T is stable unsuperstable, for all η and $t \in J_\eta$, there are finite sequences $a_t = a^\eta_t$ in the monster model such that:

(1) If $\operatorname{dom}(t) = \omega$ and $n < \omega$ then

$$a_t \underset{\bigcup_{m<n} a_t \restriction m}{\not\downarrow} a_{t\restriction n}.$$

(2) For all downward closed subtrees $X, Y \subset J_\eta$,

$$\bigcup_{t\in X} a_t \underset{\bigcup_{t\in X\cap Y} a_t}{\downarrow} \bigcup_{t\in Y} a_t.$$

(3) For all downward closed subtrees $X \subset J_\eta$ and $Y \subset J_{\eta'}$ the following holds: If $f\colon X \to Y$ is an isomorphism, then there is an automorphism F of the monster model such that, for all $t \in X$, $F(a^\eta_t) = a^{\eta'}_{f(t)}$.

Then we can find an F^f_ω-construction

$$\Big(\bigcup_{t\in J_\eta} a_t, (b_i, B_i)_{i<\kappa}\Big)$$

(here $(t(b/C), D) \in F^f_\omega$ if $D \subset C$ is finite and $b \downarrow_D C$; see [**30**]) such that

($\star$) for all $\alpha < \kappa$, c and finite $B \subset \bigcup_{t\in J_\eta} a_t \cup \bigcup_{i<\alpha} b_i$ there is $\alpha < \beta < \kappa$ such that $B_\beta = B$ and

$$\mathrm{stp}(b_\beta/B) = \mathrm{stp}(c/B).$$

Then

$$M_\eta = \bigcup_{t\in J_\eta} a_t \cup \bigcup_{i<\kappa} b_i \models T.$$

Without loss of generality we may assume that the trees J_η and the F^f_ω-constructions for M_η are chosen coherently enough such that one can find a code ξ_η for (the isomorphism type of) M_η so that $\eta \mapsto \xi_\eta$ is continuous. Thus we are left to show $\eta E_{S^\kappa_\omega} \eta' \Leftrightarrow M_\eta \cong M_{\eta'}$.

"$\Rightarrow$" Assume $J_\eta \cong J_{\eta'}$. By (3) it is enough to show that F^f_ω-constructions of length κ satisfying ($\star$) are unique up to isomorphism over $\bigcup_{t\in J_\eta} a_t$. But ($\star$) guarantees that the proof of the uniqueness of F-primary models from [**30**] works here.

"$\Leftarrow$" Suppose that $F\colon M_\eta \to M_{\eta'}$ is an isomorphism and, for a contradiction, suppose that $(\eta, \eta') \notin E_{S^\kappa_\omega}$. Let $(J^\alpha_\eta)_{\alpha<\kappa}$ be a filtration of J_η and $(J^\alpha_{\eta'})_{\alpha<\kappa}$ be a filtration of $J_{\eta'}$ (see Definition 5.8 above). For $\alpha < \kappa$, let

$$M^\alpha_\eta = \bigcup_{t\in J^\alpha_\eta} a_t \cup \bigcup_{i<\alpha} b_i$$

and, similarly for η',

$$M^\alpha_{\eta'} = \bigcup_{t\in J^\alpha_{\eta'}} a_t \cup \bigcup_{i<\alpha} b_i.$$

Let C be the cub set of those $\alpha < \kappa$ such that $F \restriction M^\alpha_\eta$ is onto $M^\alpha_{\eta'}$ and, for all $i < \alpha$, $B_i \subset M^\alpha_\eta$ and $B'_i \subset M^\alpha_{\eta'}$, where $\big(\bigcup_{t\in J_{\eta'}}, (b'_i, B'_i)_{i<b}\big)$ is in the construction of $M_{\eta'}$. Then we can find $\alpha \in \lim C$ such that in J_η there is t^* satisfying (a)–(c) below, but in $J_{\eta'}$ there is no such t^*:

(a) $\mathrm{dom}(t^*) = \omega$;

(b) $t^* \notin J^\alpha_\eta$;

(c) for all $\beta < \alpha$ there is $n < \omega$ such that $t^* \restriction n \in J^\alpha_\eta \setminus J^\beta_\eta$.

Note that

($\star\star$) if $\alpha \in C$ and $c \in M^\alpha_\eta$, there is a finite $D \subset \bigcup_{t\in J^\alpha_\eta} a_t$ with $(t(c, \bigcup_{t\in J_\eta} a_t), D) \in F^f_\omega$.

Let $c = F(a_{t^*})$. By the construction we can find finite $D \subset M^\alpha_{\eta'}$ and $X \subset J_{\eta'}$ such that

$$\Big(t(c, M^\alpha_{\eta'} \cup \bigcup_{t\in J_{\eta'}} a^{\eta'}_t), D \cup \bigcup_{t\in X} a^{\eta'}_t\Big) \in F^f_\omega.$$

But then there is $\beta \in C$, $\beta < \alpha$, such that $D \subset M^\beta_{\eta'}$ and if $u \leqslant t$ for some $t \in X$, then $u \in J^\beta_{\eta'}$ (since in $J_{\eta'}$ there is no element like t^* is in J_η). But then using ($\star\star$) and (2), it is easy to see that

$$c \underset{M^{\beta}_{\eta'}}{\downarrow} M^{\alpha}_{\eta'}.$$

On the other hand, using (1), (2), ($\star\star$) and the choice of t^* one can see that $a_{t^*} \underset{M^{\beta}_{\eta}}{\not\downarrow} M^{a}_{\eta}$, which is a contradiction. □

Open Problem If $\kappa = \lambda^+$, λ regular and uncountable, does equality modulo λ-non-stationary ideal, $E_{S^{\kappa}_{\lambda}}$, Borel reduce to T for all stable unsuperstable T?

6 Further research

In this section we merely list all the questions that have been prompted in the article:

Open Problem Is it consistent that Borel* is a proper subclass of Σ^1_1, or even equals Δ^1_1? Is it consistent that all the inclusions are proper at the same time: $\Delta^1_1 \subsetneq$ Borel* $\subsetneq \Sigma^1_1$?

Open Problem Does the direction left to right of Theorem 2.2 hold without the assumption $\kappa^{<\kappa} = \kappa$?

Open Problem Under what conditions on κ does the conclusion of Theorem 3.5 hold?

Open Problem Is the Silver dichotomy for uncountable κ consistent?

Open Problem Is it consistent that $S^{\omega_2}_{\omega_1}$ Borel reduces to $S^{\omega_2}_{\omega}$?

Open Problem We proved that the isomorphism relation of a theory T is Borel if and only if T is classifiable and shallow. Is there a connection between the depth of a shallow theory and the Borel degree of its isomorphism relation? Is one monotone in the other?

Open Problem Can it be proved in ZFC that if T is stable unsuperstable then $\cong_T$ is not Δ^1_1?

Open Problem If $\kappa = \lambda^+$, λ regular and uncountable, does equality modulo λ-non-stationary ideal, $E_{S^{\kappa}_{\lambda}}$, Borel reduce to T for all stable unsuperstable T?

Open Problem Let T_{dlo} be the theory of dense linear orderings without end points and T_{gr} the theory of random graphs. Does the isomorphism relation of T_{gr} Borel reduce to T_{dlo}, i.e., $\cong_{T_{\mathrm{gr}}} \leqslant_B \cong_{T_{\mathrm{dlo}}}$?

References

[1] S. Adams, A. S. Kechris, Linear algebraic groups and countable Borel equivalence relations, *J. Amer. Math. Soc.* 13 (2000), 909–943.

[2] D. Blackwell, Borel sets via games, *Ann. Probab.* 9 (1981), no. 2, 321–322.

[3] H. Enderton, *Elements of Set Theory*, Academic Press, 1977.

[4] J. Gregory, Higher Souslin trees and the generalized continuum hypothesis, *J. Symbolic Logic* 41 (1976), no. 3, 663–671.

[5] A. Halko, Negligible subsets of the generalized Baire space $\omega_1^{\omega_1}$, Ann. Acad. Sci. Fenn. Ser. Diss. Math. 107, Suomalainen Tiedeakatemia, 1996.

[6] A. Halko, S. Shelah, On strong measure zero subsets of ${}^{\kappa}2$, *Fund. Math.* 170 (2001), 219–229.

[7] L. Harrington, A. S. Kechris, A. Louveau, A Glimm–Effros dichotomy theorem for Borel equivalence relations, *J. Amer. Math. Soc.* 3 (1990), 903–928.

[8] B. Hart, E. Hrushovski, M. C. Laskowski, The uncountable spectra of countable theories, *Ann. of Math. (2)* 152 (2000), no. 1, 207–257.

[9] G. Hjorth, Group actions and countable models, A survey article presented at the ASL European Meeting in Utrecht, 1999.

[10] T. Huuskonen, T. Hyttinen, M. Rautila, On potential isomorphism and non-structure, *Arch. Math. Logic* 43 (2004), 85–120.

[11] T. Hyttinen, M. Rautila, The canary tree revisited, *J. Symbolic Logic* 66 (2001), no. 4, 1677–1694.

[12] T. Hyttinen, S. Shelah, Constructing strongly equivalent nonisomorphic models for unsuperstable theories, Part A, *J. Symbolic Logic* 59 (1994), no. 3, 984–996, Association for Symbolic Logic.

[13] T. Hyttinen, S. Shelah, Constructing strongly equivalent nonisomorphic models for unsuperstable theories, Part B, *J. Symbolic Logic* 60 (1995), no. 4, 1260–1272, Association for Symbolic Logic.

[14] T. Hyttinen, S. Shelah, Constructing strongly equivalent nonisomorphic models for unsuperstable theories, Part C, *J. Symbolic Logic* 64 (1999), no. 2, 634–642, Association for Symbolic Logic.

[15] T. Hyttinen, H. Tuuri, Constructing strongly equivalent nonisomorphic models, *Ann. Pure Appl. Logic* 52 (1991), 203–248.

[16] T. Jech, *Set Theory*, Springer-Verlag, Berlin Heidelberg New York, 2003.

[17] C. Karp, Finite-quantifier equivalence, in: *Theory of Models*, Proc. 1963 Internat. Sympos. Berkeley, 407–412, North-Holland, Amsterdam, 1965.

[18] M. Karttunen, Model theory for infinitely deep languages, Ann. Acad. Sci. Fenn. Ser. Diss. Math. 64, Suomalainen Tiedeakatemia, 1987.

[19] M. Koerwien, A complicated ω-stable depth 2 theory, *J. Symbolic Logic* 76 (2011), no. 1, 47–65.

[20] V. Kulikov, Borel reductions on the generalized Cantor space, submitted 2011.

[21] K. Kunen, *Set Theory – An Introduction to Independence Proofs*, Studies in Logic and Foundations of Mathematics 102, North-Holland, Amsterdam, 1980.

[22] A. Louveau, B. Velickovic, A note on Borel equivalence relations, *Proc. Amer. Math. Soc.* 120 (1994), no. 1, 255–259.

[23] A. Mekler, S. Shelah, The canary tree, *Canadian J. Math.* 36 (1993), 209–215.

[24] A. Mekler, J. Väänänen, Trees and Π^1_1-subsets of ${}^{\omega_1}\omega_1$, *J. Symbolic Logic* 58 (1993), no. 3, 1052–1070.

[25] M. Nadel, J. Stavi, $L_{\infty\lambda}$-equivalence, isomorphism and potential isomorphism, *Trans. Amer. Math. Soc.* 236 (1978), 51–74.

[26] S. Shelah, The number of non-isomorphic models of an unstable first-order theory, *Israel J. Math.* 9 (1971), 473–487.

[27] S. Shelah, A combinatorial problem; stability and order for models and theories in infinitary languages, *Pacific J. Math.* 41 (1972), 247–261.

[28] S. Shelah, The spectrum problem I: $\aleph_\varepsilon$-saturated models, the main gap, *Israel J. Math.* 43 (1982), 324–356.

[29] S. Shelah, Existence of many $L_{\infty,\lambda}$-equivalent, nonisomorphic models of T of power λ, *Ann. Pure Appl. Logic* 34 (1987), 291–310.

[30] S. Shelah, *Classification Theory*, revised ed., North-Holland, Amsterdam, 2000.

[31] S. Shelah, Diamonds, *Proc. Amer. Math. Soc.* 138 (2010), 2151–2161.

[32] S. Shelah, J. Väänänen, Stationary sets and infinitary logic, *J. Symbolic Logic* 65 (2000), 1311–1320.

[33] J. H. Silver, Counting the number of equivalence classes of Borel and coanalytic equivalence relations, *Ann. Math. Logic* 18 (1980), 1–28.

[34] H. Tuuri, Relative separation theorems for $L_{\kappa^+\kappa}$, *Notre Dame J. Formal Logic* 33 (1992), no. 3, 383–401.

[35] J. Väänänen, Games and trees in infinitary logic: A survey, in: M. Krynicki, M. Mostowski and L. Szczerba, eds., *Quantifiers*, Kluwer Academic Publishers, 1995, 105–138.

[36] J. Väänänen, How complicated can structures be? *Nieuw Archief voor Wiskunde*, June 2008, 117–121.

[37] J. Väänänen, *Models and Games*, Cambridge Studies in Advanced Mathematics 132, Cambridge University Press, 2011.

[38] R. Vaught, Invariant sets in topology and logic, *Fund. Math.* 82 (1974/75), 269–294.

Potential isomorphism of elementary substructures of a strictly stable homogeneous model

Sy-David Friedman[†], **Tapani Hyttinen**[‡], **Agatha C. Walczak-Typke**[‡]

[†] Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
sdf@logic.univie.ac.at

[‡] Department of Mathematics and Statistics, University of Helsinki, Finland
tapani.hyttinen@helsinki.fi, agatha.walczak-typke@helsinki.fi

Abstract. The results herein form part of a larger project to characterize the classification properties of the class of submodels of a homogeneous stable diagram in terms of the solvability (in the sense of [**1**]) of the potential isomorphism problem for this class of submodels.

We restrict ourselves to locally saturated submodels of the monster model $\mathfrak{M}$ of some power π. We assume that in Gödel's constructible universe $\mathbb{L}$, π is a regular cardinal at least the successor of the first cardinal in which $\mathfrak{M}$ is stable.

We show that the collection of pairs of submodels in $\mathbb{L}$ as above which are potentially isomorphic with respect to certain cardinal-preserving extensions of $\mathbb{L}$ is equiconstructible with $0^{\#}$. As $0^{\#}$ is highly "transcendental" over $\mathbb{L}$, this provides a very strong statement to the effect that potential isomorphism for this class of models not only fails to be set-theoretically absolute, but is of high (indeed of the highest possible) complexity.

The proof uses a novel method that does away with the need for a linear order on the skeleton.

Introduction

The results we give here are part of a larger project to prove strong non-structure results for non-elementary classes. The original impetus comes from work to generalize the results of [**2**] to the Homogeneous Model Theory context. The main theorem of that earlier work was:

Theorem ([**2**]) *Assume $0^{\#}$ exists, and let T be a constructible first-order theory which is countable in Gödel's constructible universe $\mathbb{L}$. Then the following are equivalent:*

(1) *The collection*

$$\{\langle \mathscr{A}, \mathscr{B}\rangle \in \mathbb{L} : \mathscr{A} \models T,\ \mathscr{B} \models T,\ \mathscr{A} \text{ and } \mathscr{B} \text{ have universe } (\aleph_2)^{\mathbb{L}} \text{ and are isomorphic in an extension of } \mathbb{L} \text{ with the same cardinals and reals as } \mathbb{L}\}$$

is constructible.

(2) *The theory T is superstable with NOTOP and NDOP.*

This work has been published: Potential isomorphism, $0^{\#}$ and classifiability of elementary substructures of a homogeneous model, *Journal of Symbolic Logic* 76(3):987–1004, 2011.

[†]The first author wishes to thank the Fonds zur Förderung der wissenschaftlichen Forschung (FWF) for its support through grant P19898-N18 and the John Templeton Foundation for its support through the CRM Infinity Project, Grant #13152.

[‡]The second and third authors gratefully acknowledge the hospitality of the Institut Mittag-Leffler, Djursholm, Sweden during Fall 2009. The work of the second and third authors is partially supported by the Academy of Finland, grant number 1123110.

This result was proved using strong non-structure theorems, following the cases found in the Main Gap Theorem [**15**].

We chose the Homogeneous Model Theory context to extend this result because of its well developed Main Gap Theorem [**10**]. Much of the difficulty lies in finding strong non-structure theorems in the Homogeneous Model Theory (HMT) context. While one can prove strong non-structure theorems in non-elementary contexts (e.g., Abstract Elementary Classes, or as in [**4**]) having the order property in exactly the same way as was done for unstable first order theories, strong non-structure theorems have not been proved for almost any other non-elementary classes. This is because the only first-order strong non-structure theorem that can be generalized in a straightforward manner is the one stemming from the order property.

In this paper, we prove a strong non-structure theorem for the strictly stable (stable but not superstable) case in HMT. In the first-order context, non-structure theorems for the strictly stable case are proved by first finding tree indiscernibles, and then using them as skeleta in Ehrenfeucht–Mostowski model constructions. In the HMT context, a major problem arises in simply generalizing the approach used in the first-order context: without large cardinals one cannot find tree-indiscernibles. Thus if one wants to carry out the constructions in $\mathbb{L}$, as we do in this paper, currently known methods do not allow for the use of Ehrenfeucht–Mostowski model constructions. Alternatively, if one were willing to assume large cardinals, then the ideas of [**1**] would have to be generalized from $\mathbb{L}$ to a larger core model, raising significant new set-theoretic challenges.

We hope that our exposition will be accessible both to model theorists and to set theorists. Those seeking definitions of set-theoretic concepts should consult, for example, [**12**]. On the other side, [**7**] or a similar introduction to methods in classification theory may help with the model-theoretic concepts. However, full comprehension of this paper requires knowledge of [**8**].

1 Preliminaries

1.1 Notation

Gödel's constructible universe will be denoted as $\mathbb{L}$. To differentiate, similarity types (languages) will be denoted with the calligraphic $\mathcal{L}$.

1.2 Set theory

1.2.1 Relative constructibility

This paper is concerned with examining the solvability (in the sense of [**1**]) of certain problems in the classification of structures that are not first-order axiomatizable. Our intuition is that if the collection of constructible objects that satisfy a particular condition is constructible (i.e., in $\mathbb{L}$), then we say that the condition's problem is *solvable*. On the other hand, if the collection is not in $\mathbb{L}$, then we say that the condition's problem is *unsolvable*.

We will demonstrate the unsolvability of a problem by *reducing* to it sets that are known to be non-constructible —indeed, to sets that are equiconstructible with $0^{\#}$.

First, some notation:

Definition 1.1 We have the following notion of *reduction*: Suppose that $\langle X_0, X_1\rangle$, $\langle Y_0, Y_1\rangle$ are pairs of disjoint subsets of the constructible universe $\mathbb{L}$; that is, they are disjoint collections of constructible sets. Note that the pairs $\langle X_0, X_1\rangle$ and $\langle Y_0, Y_1\rangle$ need not be constructible themselves. We write

$$\langle X_0, X_1\rangle \xrightarrow{\mathbb{L}} \langle Y_0, Y_1\rangle$$

if there exists a constructible function $g \in L$ such that

$$x \in X_0 \Rightarrow g(x) \in Y_0 \text{ and } x \in X_1 \Rightarrow g(x) \in Y_1.$$

We write X_0 instead of $\langle X_0, X_1\rangle$ in the case that X_0 is the complement of X_1 within some constructible set. We employ the analogous convention for Y_0 and Y_1.

The idea behind this notion of reduction is that if $\langle X_0, X_1\rangle$ is non-constructible, $X_0 \cup X_1$ is constructible, and $\langle X_0, X_1\rangle \xrightarrow{\mathbb{L}} \langle Y_0, Y_1\rangle$, then $\langle Y_0, Y_1\rangle$ is also non-constructible.

Definition 1.2

(1) A *cardinal preserving extension* ("*Cap-extension*") of $\mathbb{L}$ is a transitive model satisfying the Axiom of Choice containing all the ordinals, and which is contained in a set-generic extension of V and has the same cardinals as $\mathbb{L}$.

(2) A *cardinal- and real-preserving extension* ("*CaRp-extension*") of $\mathbb{L}$ is a transitive model satisfying the Axiom of Choice containing all the ordinals, and which is contained in a set-generic extension of V and has the same cardinals and real numbers as $\mathbb{L}$.

(3) For ν a cardinal, a *cardinal- and $\mathscr{P}(\nu)$-preserving extension* ("*Ca$\mathscr{P}(\nu)$-extension*") is defined analogously.

We also remind the reader of the following highly non-constructible object:

Definition 1.3 If there exists a non-trivial elementary embedding of the constructible universe $\mathbb{L}$ into itself, then there is a closed unbounded proper class of ordinals that are indiscernible for the structure $(\mathbb{L}, \in)$. Then, we can define $0^\#$ (*"zero-sharp"*) to be the real number that codes in the canonical way the Gödel numbers of the formulas that are true about the indiscernibles in $\mathbb{L}$.

The existence of $0^\#$ is independent of the axioms of set theory, ZFC. If ZFC is consistent, then so is ZFC with the assumption that $0^\#$ does not exist. It is commonly assumed that ZFC is consistent with the assumption that $0^\#$ does exist.

We assume throughout that $0^\#$ exists.

The real number $0^\#$ is a highly non-constructible object. Our intuition will be to show that a class of models is non-constructible by reducing $0^\#$ to it, in the sense above. In particular, we will use the following theorem. We denote by S^ν_ω the stationary set consisting of ordinals in ν of cofinality ω.

Theorem 1.4 ([**1**]) *Denote by $\mathcal{S}(\kappa)$ (resp. $\mathcal{S}_r(\kappa)$) the collection of sets $S \in \mathbb{L}$ such that $S \subseteq (S^\nu_\omega)^{\mathbb{L}}$ is stationary in $\mathbb{L}$ and in a cardinal- (and real-) preserving extension, $\nu \setminus S$ contains a club.*

Then, if κ is an uncountable regular cardinal in $\mathbb{L}$ and $(\kappa^+ = \nu)^{\mathbb{L}}$, then

$$0^\# \xrightarrow{\mathbb{L}} \mathcal{S}(\kappa)$$

and

$$0^\# \xrightarrow{\mathbb{L}} \mathcal{S}_r(\kappa).$$

1.3 Homogeneous model theory

1.3.1 Introduction and motivation for homogeneous model theory

Homogeneous Model Theory (HMT), introduced in [**13**] as "finite diagrams stable in power", is an approach to the model-theoretic classification of classes of non-elementary structures (i.e., structures not axiomatizable using a first-order theory). The motivation behind the development of this approach, as explained in [**3, 10**], was the aim to classify the class of models of an $\mathcal{L}_{\gamma^+\omega}$ sentence ψ, with $\preccurlyeq_{\mathcal{L}_{\gamma^+\omega}}$ as the substructure relation. We wish this class of models to be "well behaved" and so add the requirement that the class satisfies the amalgamation property. It was proved in [**13**] that it is equivalent to consider the class of elementary submodels of a homogeneous monster model $\mathfrak{M}$.

Thus, in practice the contrast to elementary (first-order) model theory where one assumes that all considerations take place within a large *saturated* monster model, is that we take away the assumption that the monster is saturated, and instead only insist that it be *homogeneous*. However, in the HMT context, a major difficulty arises because the compactness theorem fails. In return for this concession, we do gain a widening of the possible structures under consideration as opposed to elementary model theory. For example, the class of existentially closed models of an inductive theory can be studied within the framework of homogeneous model theory. In fact, for some γ big enough the class of submodels of a homogeneous model can be axiomatized in some theory $T^* \subset \mathcal{L}_{\gamma^+\omega}$. (Specifically, where $\gamma \geqslant |D(\mathrm{Th}(\mathfrak{M})) \setminus D|$, where D is the finite diagram. For more specifics, see [**3, 13**].)

1.3.2 Types and homogeneous monsters

We assume we work within a very large homogeneous model which can serve as a monster model. We will then be interested in the class of elementary submodels of this monster.

We work with $\mathfrak{M}$-consistent types:

Definition 1.5 ([**10**]) Let $A \subseteq \mathfrak{M}$, and let p be a (first-order) type over A. We say that p is *$\mathfrak{M}$-consistent* if it is realized in $\mathfrak{M}$.

We write $\mathrm{tp}_{\mathfrak{M}}(a/A)$ to indicate the $\mathfrak{M}$-consistent type of a over A. Similarly, we take $\mathrm{S}^m_{\mathfrak{M}}(A) = \{\mathrm{tp}_{\mathfrak{M}}(a/A) : a \in \mathfrak{M}, \text{length}(a) = m\}$, and $\mathrm{S}_{\mathfrak{M}}(A) = \bigcup_{m<\omega} \mathrm{S}^m_{\mathfrak{M}}(A)$.

Definition 1.6 A homogeneous monster model $\mathfrak{M}$ is said to be *stable in* λ if for every $B \subset \mathrm{dom}(\mathfrak{M})$ of cardinality at most λ, and for every $n < \omega$, we have $|\,\mathrm{S}^n_{\mathfrak{M}}(B)| \leqslant \lambda$.

The monster model $\mathfrak{M}$ is *stable* if it is stable in some λ.

The monster model $\mathfrak{M}$ is *unstable* if it is not stable.

We denote by $\lambda(\mathfrak{M})$ the least λ in which $\mathfrak{M}$ is stable, if it exists [**9**]. Denote by $\lambda_r(\mathfrak{M})$ the first regular cardinal $\geqslant \lambda(\mathfrak{M})$.

1.3.3 Indiscernibles and strong splitting independence

A standard notion from model theory follows. We include this definition to make the terminology clear, as the set-theoretic usage is sometimes at odds with accepted usage among model theorists.

Definition 1.7 An (indexed) set of tuples $\{\overline{a}_i : i < \alpha\}$ is called an *n-indiscernible sequence over A*, for $n < \omega$, if

$$\mathrm{tp}(\overline{a}_0, \ldots, \overline{a}_{n-1}/A) = \mathrm{tp}(\overline{a}_{i_0}, \ldots, \overline{a}_{i_{n-1}}/A),$$

for every $i_0 < \cdots < i_{n-1} < \alpha$. The set of tuples $\{\overline{a}_i : i < \alpha\}$ is an *indiscernible sequence over* A if it is an n-indiscernible sequence over A for every $n < \omega$. It is said to be an *indiscernible set* if the ordering induced by the indices does not matter.

Definition 1.8 ([**15**, III, p. 85, Definition 1.2]) A type $p \in \mathrm{S}^n(A)$ *splits strongly* over $B \subseteq A$ if there exists an indiscernible sequence $\{\overline{a}_i : i < \omega\}$ over B and a formula ϕ such that $\phi(\overline{x}, \overline{a}_0), \neg\phi(\overline{x}, \overline{a}_1) \in p$.

The following definitions are very similar to the definitions of independence and $\kappa(T)$ in the first-order context. However, here we use strong splitting instead of forking in the definitions. In the first order context, the definitions using forking and the definition as stated here are equivalent. In the HMT context, forking is ill-defined, so we take the strong splitting definition. Consequently, we lose some nice properties, among them transitivity of the independence relation.

Definition 1.9 ([**9**, p. 2]) We define $\kappa(\mathfrak{M})$ to be the least infinite cardinal such that there are no a, b_i, and c_i, $i < \kappa(\mathfrak{M})$, such that

(i) for all $i < \kappa(\mathfrak{M})$, there is an infinite indiscernible set I_i over $\bigcup_{j<i}(b_j \cup c_j)$ such that $b_i, c_i \in I_i$;
(ii) for all $i < \kappa(\mathfrak{M})$, there is $\phi_i(x, y)$ such that $\models \phi_i(a, b_i) \wedge \neg\phi_i(a, c_i)$.

Note that $\kappa(\mathfrak{M}) \leqslant \lambda(\mathfrak{M})$ by Corollary 1.3 of [**9**].

Definition 1.10 ([**9**, p. 17, remarks before Lemma 5.1]) We say that a monster model is *superstable* if $\kappa(\mathfrak{M}) = \aleph_0$. We will call a monster model *strictly stable* if it is stable, but not superstable.

Now we can define the notion of independence that we use in the HMT context.

Definition 1.11 ([**9**, Definition 3.1(i)]) We write $a \mathop{\smile\!\!\!\!\!\mid}_A B$ if there is $C \subseteq A$, $|C| < \kappa(\mathfrak{M})$, such that for all $D \supseteq A \cup B$ there is b which satisfies $\mathrm{tp}_{\mathfrak{M}}(b/A \cup B) = \mathrm{tp}_{\mathfrak{M}}(a/A \cup B)$ and $\mathrm{tp}_{\mathfrak{M}}(b/D)$ does not split strongly over C. We write $C \mathop{\smile\!\!\!\!\!\mid}_A B$ if for all $a \in C$, $a \mathop{\smile\!\!\!\!\!\mid}_A B$.

1.3.4 Primary model constructions

Most of the following definitions are given only in very general terms that allow one to apply the notions to a very wide range of contexts. We give here these definitions specifically in the way we need them in our context.

Definition 1.12 For the following, ν is a cardinal.

- We say that $\mathrm{tp}_{\mathfrak{M}}(a/A)$ is $\mathbf{F}^{\mathfrak{M}}_{\nu}$*-isolated* over B if there is $B \subseteq A$, $|B| < \nu$, such that for all b, $\mathrm{tp}_{\mathfrak{M}}(b/B) = \mathrm{tp}_{\mathfrak{M}}(a/B)$ implies $\mathrm{tp}_{\mathfrak{M}}(b/A) = \mathrm{tp}_{\mathfrak{M}}(a/A)$ ([**9**, Definition 5.2]).
- We say that an (elementary sub-)model $\mathscr{A}$ (of $\mathfrak{M}$) is $\mathbf{F}^{\mathfrak{M}}_{\nu}$*-saturated* if for all $A \subseteq \mathscr{A}$, $|A| < \nu$, and a, there is $b \in \mathscr{A}$ such that $\mathrm{tp}_{\mathfrak{M}}(b/A) = \mathrm{tp}_{\mathfrak{M}}(a/A)$ ([**9**, Definition 1.8(i)]).
- An $\mathbf{F}^{\mathfrak{M}}_{\nu}$*-construction* is a triple
$$\mathscr{A} = \langle A, \{\overline{a_i} : i < \alpha\}, \langle B_i : i < \alpha\rangle\rangle,$$
such that $\mathrm{tp}_{\mathfrak{M}}(\overline{a_i}/\bigcup\{\overline{a_j} : j < i\} \cup A)$ is $\mathbf{F}^{\mathfrak{M}}_{\nu}$-isolated over B_i ([**15**, IV, p. 155, Definition 1.2(1)]).

- We say that C_0 is $\mathbf{F}^{\mathfrak{m}}_{\nu}$*-constructible over* A_0 if there is some $\mathbf{F}^{\mathfrak{m}}_{\nu}$-construction
$$\mathscr{A} = \langle A_0, \{\overline{a_i} : i < \alpha\}, \langle B_i : i < \alpha\rangle\rangle$$
such that
$$C_0 = \bigcup\{\overline{a_i} : i < \alpha\} \cup A_0$$
([**15**, IV, p. 156, Definition 1.3]).
- If C is $\mathbf{F}^{\mathfrak{m}}_{\nu}$-constructible over A and C is $\mathbf{F}^{\mathfrak{m}}_{\nu}$-saturated then we say that C is $\mathbf{F}^{\mathfrak{m}}_{\nu}$*-primary over* A ([**15**, IV, p. 156, Definition 1.4(1)]).
- We say that C is $\mathbf{F}^{\mathfrak{m}}_{\nu}$*-primitive* over A if $A \subseteq C$, and for every $\mathbf{F}^{\mathfrak{m}}_{\nu}$-saturated C' such that $A \subseteq C'$, there is an elementary mapping f from C into C', where $f \restriction_A$ is the identity ([**15**, IV, p. 156, Definition 1.4(2)]).
- We say that C is $\mathbf{F}^{\mathfrak{m}}_{\nu}$*-prime* over A if it is $\mathbf{F}^{\mathfrak{m}}_{\nu}$-primitive over A and $\mathbf{F}^{\mathfrak{m}}_{\nu}$-saturated.
- We say A is $\mathbf{F}^{\mathfrak{m}}_{\nu}$*-atomic* over B if $B \subseteq A$ and for every $\overline{a} \in A$, $\mathrm{tp}_{\mathfrak{m}}(\overline{a}/B)$ is $\mathbf{F}^{\mathfrak{m}}_{\nu}$-isolated ([**15**, IV, p. 157, Definition 1.5]).

Fact 1.13 On the surface, the isolation notion $\mathbf{F}^{\mathfrak{m}}_{\nu}$ above is quite similar to the isolation notion $\mathbf{F}^{p}_{\nu}$ of [**15**, IV, p. 168, Definition 2.6], an isolation notion that does not satisfy certain axioms key in constructions.

However, as was noted in the last paragraph of the introduction to [**10**], under the assumption that $\mathfrak{m}$ is stable, one can easily show that the isolation notion $\mathbf{F}^{\mathfrak{m}}_{\nu}$, for $\nu \geqslant \lambda_r(\mathfrak{m})$ has properties very similar to the much better-behaved notion $\mathbf{F}^{s}_{\nu}$, a definition of which can be found in [**15**, IV, Definitions 2.1.1.ii and 2.1.2].

In our considerations, we will use (mostly) without comment properties of the isolation $\mathbf{F}^{\mathfrak{m}}_{\nu}$, $\nu \geqslant \lambda_r(\mathfrak{m})$, which are proved in [**9**].

Definition 1.14 ([**10**, Definition 0.1]) A model $\mathscr{A}$ is said to be *locally* $\mathbf{F}^{\mathfrak{m}}_{\nu}$*-saturated* if for all finite sets $A \subset \mathscr{A}$ there is an $\mathbf{F}^{\mathfrak{m}}_{\nu}$-saturated model $\mathscr{B}$ such that $A \subset \mathscr{B} \subset \mathscr{A}$.

2 The strictly stable case

Theorem 2.1 *Assume* $0^{\#}$ *exists. Suppose* $\mathcal{L} \in \mathbb{L}$ *is a signature such that* $(|\mathcal{L}| \leqslant \omega)^{\mathbb{L}}$. *Let* $\mathfrak{m} \in \mathbb{L}$ *be a strictly stable (stable, but not superstable) homogeneous monster model in similarity type* $\mathcal{L}$ *such that* $(|\mathfrak{m}| = \mu)^{\mathbb{L}}$, *for* μ *sufficiently large.*

Let π *be such that* $\pi = \mathrm{cf}(\pi) > \lambda_r(\mathfrak{m})$.

Let ${}^{Ca\mathscr{P}(\lambda_r)}PIP^{\mathfrak{m}}_{\pi}$ *be the collection of pairs* $(\mathscr{A}, \mathscr{B}) \in \mathbb{L}$ *of locally* $\mathbf{F}^{\mathfrak{m}}_{\lambda_r(\mathfrak{m})}$*-saturated elementary substructures of* $\mathfrak{m}$ *with universe* π *such that there is a cardinal- and* $\mathscr{P}(\lambda_r(\mathfrak{m}))$*-preserving extension of* $\mathbb{L}$ *in which* $\mathscr{A} \cong \mathscr{B}$. *(Here "PIP" stands for "potentially isomorphic pairs".)*

Then, ${}^{Ca\mathscr{P}(\lambda_r)}PIP^{\mathfrak{m}}_{\pi}$ *is equiconstructible with* $0^{\#}$.

We will show that, for each stationary set $S \subseteq S^{\pi}_{\omega}$, one can find two models $\mathscr{A}, \mathscr{B} \in \mathbb{L}$ of size π such that in any Cap-extension of $\mathbb{L}$, $\mathscr{A} \cong \mathscr{B}$ iff $\pi \setminus S$ contains a club set. We do this by constructing two trees of small height J_0, J_1, differing from one another only in that one codes S while the other does not. We will then perform a primary model constructions along these trees. We show then that these models are not isomorphic in the ground model, but become isomorphic in a suitable extension only if S is no longer stationary in that extension.

2.1 Defining the trees and other orderings

We define two trees I_0 and I_1, which will be used to define two trees J_0 and J_1. From J_0 and J_1 we will construct models $\mathscr{A}_{J_0}$ and $\mathscr{A}_{J_1}$, respectively, which are potentially isomorphic but not isomorphic. The trees I_i, J_i, $i = 0, 1$ all belong to a certain general family of trees K^ω_{tr}, defined below. Note that the trees we define here are precisely the trees that were used for the Ehrenfeucht–Mostowski constructions in the first order strictly stable case as analyzed in [**2**] and papers cited there.

Unlike the first-order context, without large cardinal assumptions non-structure results for strictly stable theories have only been shown for weakly $\mathbf{F}^{\mathfrak{m}}_{\lambda_r(\mathfrak{m})}$-saturated models, and not in general [**6, 8, 10**]. Ehrenfeucht–Mostowski constructions yield models that are insufficiently saturated to be able to use the existing non-structure results. We will thus instead use the technique of primary model constructions, which yield more saturated models. In addition, we cannot use Ehrenfeucht–Mostowski constructions in this case because we would need to find tree indiscernibles in the model, and to do so we would need large cardinals that are not available to us in $\mathbb{L}$. Because we need this different technique, we need to further define $K_i = \mathscr{P}^{<\omega}(J_i)$, the set of all finite subsets of J_i, $i = 0, 1$, as well as an ordering on the K_i. We will then carry out primary model constructions using sets indexed by the K_i.

We define first a general family of trees:

Definition 2.2 Let θ be a linear order, and let ${}^{\leqslant\omega}\theta$ be the set of all suborders of θ of length at most ω. We let $K^\omega_{\mathrm{tr}}(\theta)$ be the class of models that are isomorphic to a model of the form

$$\mathcal{I} = (I, \lessdot, \mathrm{DOM}_\alpha, <_{\mathrm{lex}}, \mathrm{MaxInSg})_{\alpha\leqslant\omega},$$

where

(1) $I \subseteq {}^{\leqslant\omega}\theta$ and is closed under initial segments;
(2) $\lessdot$ is the initial segment relation;
(3) $\mathrm{DOM}_\alpha = \{\eta \in I : \mathrm{dom}\,\eta = \alpha\}$;
(4) $<_{\mathrm{lex}}$ denotes the lexicographic ordering on I;
(5) $\mathrm{MaxInSg}(\zeta, \eta)$ is the maximal common initial segment of ζ and η.

Trees in the class $K^\omega_{\mathrm{tr}}(\theta)$ are called *ordered trees* in the literature. We define

$$K^\omega_{\mathrm{tr}} = \bigcup\{K^\omega_{\mathrm{tr}}(\theta) : \theta \text{ is a linear order}\}.$$

2.1.1 The first generation of trees

We fix some notation:

- Let $(\lambda = \lambda_r(\mathfrak{m}))^{\mathbb{L}}$. Because we have assumed that $\mathfrak{m}$ is strictly stable, $\lambda \geqslant \aleph_1$.
- Let $\pi \geqslant \lambda^+ \geqslant \aleph_2$ be an uncountable regular cardinal such that $\pi^\omega = \pi$.
- Let $S \subseteq (S^\pi_\omega)^L$ be a stationary set in $\mathbb{L}$.
- Let $\overline{S} = \langle \eta_\alpha : \alpha \in S \rangle$, where each η_α is an increasing cofinal sequence in α of order type ω (i.e., a π-club guessing sequence[1]). We are guaranteed the existence of this club guessing sequence because $\pi \geqslant \aleph_2$.

We next define our first pair of trees.

[1] For a definition, see p. 442 of [**12**].

Definition 2.3

- Let
$$I_0 = I(\pi, \overline{S})$$
be an ordered tree in $K^{\omega}_{\mathrm{tr}}(\pi)$, with cardinality $|I_0| = \pi$, having universe
$${}^{<\omega}\pi \cup \{\eta_\alpha : \eta_\alpha \in \overline{S}\} \subset {}^{\leqslant\omega}\pi,$$
where the relations are as always on ordered trees.
- Let
$$I_1 = I(\pi, \langle\,\rangle) = {}^{<\omega}\pi.$$
The tree I_1 is also in $K^{\omega}_{\mathrm{tr}}(\pi)$, and $|I_1| = \pi$.

2.1.2 The second generation of trees

Now we define the domains of our next generation of trees. This next generation is needed so that we have non-isomorphic $L_{\infty\pi}$-equivalent trees in $K^{\omega}_{\mathrm{tr}}(\pi)$ with certain further useful properties (see [**11**, Definition 8.19 and Lemma 8.20] and [**5**, Lemma 7.29]). The non-isomorphism of the pair of trees I_0 and I_1 is easy to detect. We therefore need a new pair of trees where this non-isomorphism is more "obscured". This construction is originally due to Shelah [**14**].

Let

- $\mathrm{LEX}({}^{<\omega}\pi)$ be a linear order with universe ${}^{<\omega}\pi$, ordered lexicographically.
- $\mathrm{OT}_\pi({}^{<\omega}\pi)$ be a linear (well) order with universe ${}^{<\omega}\pi$, ordered with order type π.
- $\theta = \mathrm{OT}_\pi({}^{<\omega}\pi) \cdot \mathrm{LEX}({}^{<\omega}\pi)$ be the product of the linear orders $\mathrm{OT}_\pi({}^{<\omega}\pi)$ and $\mathrm{LEX}({}^{<\omega}\pi)$ whose universe is $\mathrm{OT}_\pi({}^{<\omega}\pi) \times \mathrm{LEX}({}^{<\omega}\pi)$.

Let
$$\overline{I_0} = \langle I_0 \cap {}^{\leqslant\omega}\alpha : \alpha < \pi\rangle,$$
$$\overline{I_1} = \langle I_1 \cap {}^{\leqslant\omega}\alpha : \alpha < \pi\rangle = \langle {}^{<\omega}\pi \cap {}^{\leqslant\omega}\alpha : \alpha < \pi\rangle = \langle {}^{<\omega}\alpha : \alpha < \pi\rangle$$
be π-filtrations of I_0 and I_1, respectively.

The filtrations are used in Definition 2.5 to ensure that the trees we build are not "too similar".

Lemma 2.4 ([**5**, Lemma 7.24] or [**11**, Lemma 8.17]) *Let π be a cardinal. Suppose $\mathrm{LEX}({}^{<\omega}\pi)$ is as above. Then there is $E \subseteq \mathrm{LEX}({}^{<\omega}\pi)$ of cardinality π such that for any $a, b \in E$ there is an automorphism $g_{a,b}$ of $\mathrm{LEX}({}^{<\omega}\pi)$ which maps a to b.*

Let $E \subseteq \mathrm{LEX}({}^{<\omega}\pi)$ be as given by Lemma 2.4. Fix $c \in E$. Let g be a bijection $g\colon \{R \mid R \in \mathrm{rng}(\overline{I_0}) \cup \mathrm{rng}(\overline{I_1})\} \longrightarrow E \setminus \{c\}$.

Definition 2.5 Let $J_0 = J(c, g, \overline{I_0}, \overline{I_1})$ have a universe consisting of functions $\eta \in {}^{\leqslant\omega}\theta$, such that one of the following holds:

(1) $\eta \in {}^{<\omega}\theta$ (in other terms, $\eta \in \mathrm{DOM}_n$ for some $n \in \omega$; i.e., η is of finite length);

(2) there is $s \in I_0$ such that $\mathrm{dom}(s) = \omega$ and, for all $n < \omega$,
$$\eta(n) = \langle s \restriction_{(n+1)}, c\rangle;$$

(3) there are $m < \omega$, $R \in \mathrm{rng}(\overline{I_0}) \cup \mathrm{rng}(\overline{I_1})$, and $s \in R$ with $\mathrm{dom}(s) = \omega$ such that, for all finite $n \geqslant m$, $\eta(n) = \langle s \restriction_{(n+1)}, g(R)\rangle$.

Let $J_1 = J(c, g, \overline{I_1}, \overline{I_0})$ be defined analogously. Note that J_1 differs from J_0 only in that J_1 does not have any members satisfying condition (2) of the definition.

The trees J_0 and J_1 are isomorphic to ordered trees in $K^{\omega}_{\mathrm{tr}}(\theta)$, so we assume that $J_0, J_1 \in K^{\omega}_{\mathrm{tr}}(\theta)$.

Lemma 8.20 of [**11**] establishes that J_0 and J_1 are $\mathcal{L}_{\infty\pi}$-equivalent.

2.1.3 The third generation: a quasi-order

At this point in the construction, we can lose the $<_{\mathrm{lex}}$ ordering on J_i, since we do not need it for the primary model construction that follows. Indeed, we could have used a different construction in the second generation that did not feature $<_{\mathrm{lex}}$. However, we chose to take advantage of the existing construction from [**14**] to save some effort.

Let $K_i = \mathscr{P}^{<\omega}(J_i)$ be the set of all finite subsets of J_i, $i = 0, 1$, respectively.

We define relations as in [**8**]. Let $u, v \in K_i$. We define the "minimum" set of initials $\mathrm{MinSetIn}(u, v)$ to be the largest set X such that:

(1) $X \subseteq \{\mathrm{MaxInSg}(\zeta, \eta) : \zeta \in u, \eta \in v\}$;

(2) if $\eta_i, \eta_j \in X$ and η_i is an initial segment of η_j, then $\eta_i = \eta_j$.

Note that

$$\mathrm{MinSetIn}(u, u) = \{\zeta \in u : \neg\exists \eta \in u\,(\zeta \text{ is a proper initial segment of } \eta)\}.$$

The elements of K_i are ordered by $<^K$: $u <^K v$ iff for every $\zeta \in u$ there is $\eta \in v$ such that ζ is an initial segment of η. In other terms,

$$u \leq^K v \text{ iff } \mathrm{MinSetIn}(u, v) = \mathrm{MinSetIn}(u, u).$$

Note that $(K_i, <^K)$ cannot have infinite descending chains.

Definition 2.6 We call $s \in K_i$ *semi-good* if s is an antichain with regard to the $\lessdot$ relation in J_i.

Denote by $\overline{s}$ the downwards closure of s. We say that $r \in K_i$ is *good* if it is downwards closed and $r \subset \overline{s}$, where s is semi-good. We denote by $G(K_i)$ the collection of good elements of K_i.

2.2 Building the models: putting fat on the trees

We will base a primary model construction based on the trees J_i using the quasi-order K_i.

2.2.1 Cardinal assumptions

Recall that we assume in this section that we work within $\mathfrak{m}$, a strictly stable homogeneous monster model of cardinality $||\mathfrak{m}|| = \mu$. We let $\lambda(\mathfrak{m})$ be the first cardinal in which $\mathfrak{m}$ is stable, and we let $\lambda = \lambda_r(\mathfrak{m})$ be the first regular cardinal $\geqslant \lambda(\mathfrak{m})$. By our assumption that $\mathfrak{m}$ is strictly stable, $\kappa(\mathfrak{m}) \neq \omega$ (see 2.7 below). Thus, $\aleph_1 \leqslant \kappa(\mathfrak{m}) \leqslant \lambda(\mathfrak{m}) \leqslant \lambda$. Further, let π be a regular cardinal such that $\pi^{\omega} = \pi$ and $\lambda < \pi < \mu$. Thus $\pi \geqslant \aleph_2$. This π is the size of the models that we will be building, and is the cardinal upon which our trees have been built.

We proceed with the construction similarly to [**8**].

2.2.2 An initial ω-sequence of models

We restate the following lemma, which provides the seed for our construction:

Lemma 2.7 ([**9**, Lemma 5.1]) *The following are equivalent:*

(1) $\mathfrak{m}$ *is not superstable.*

(2) $\kappa(\mathfrak{M}) \neq \omega$.

(3) *There is an increasing sequence* $\mathscr{A}_n$, $n < \omega$ *of* $\mathbf{F}^{\mathfrak{M}}_{\lambda(\mathfrak{M})}$*-saturated models and an element* a *such that, for all* $n < \omega$, $a \mathop{\not\smile\hspace{-0.9em}\vert}_{\mathscr{A}_n} \mathscr{A}_{n+1}$.

Fact 2.8 The sequence $\mathscr{A}_n$, $n < \omega$ in Lemma 2.7 can be chosen to consist of models of size λ.

Proof. Let $\mathscr{A}_n$, $n < \omega$ be the sequence of models given by Lemma 2.7. It is easy to find such models that are quite large.

Each $\mathscr{A}_n$ is $\mathbf{F}^{\mathfrak{M}}_{\lambda}$-saturated, and hence strongly $\mathbf{F}^{\mathfrak{M}}_{\kappa(\mathfrak{M})}$-saturated by [**9**, Lemma 1.9(iv)]. Thus, by the monotonicity given by Lemmas 1.2(vi) and 1.13, and the proof of Lemma 3.2(iii) of that same paper, there exists an increasing sequence $B_n \subset \mathscr{A}_n$ of sets of size $< \kappa(\mathfrak{M})$ such that

$$a \mathop{\smile\hspace{-0.9em}\vert}_{B_n} \mathscr{A}_n.$$

We also have that $a \mathop{\not\smile\hspace{-0.9em}\vert}_{B_i} \mathscr{A}_{i+1}$. By the finite character of independence in our setting ([**9**, Corollary 3.5(i)]), there exist finite $b_{n+1} \in \mathscr{A}_{n+1}$ that witness $a \mathop{\not\smile\hspace{-0.9em}\vert}_{B_n} \mathscr{A}_{n+1}$ such that

$$a \mathop{\not\smile\hspace{-0.9em}\vert}_{B_n} b_{n+1}.$$

Choose $\mathbf{F}^{\mathfrak{M}}_{\lambda}$-saturated models $\mathscr{C}_n$ of size λ so that $B_n \subset \mathscr{C}_n \subset \mathscr{A}_n$ and $b_{n+1} \in \mathscr{C}_{n+1}$. We can do this by [**9**, Theorem 3.14].

We claim that $(\mathscr{C}_n)_{n<\omega}$ satisfy the requirements of Lemma 2.7. Assume the contrary, that $a \mathop{\smile\hspace{-0.9em}\vert}_{\mathscr{C}_n} \mathscr{C}_{n+1}$. Since $a \mathop{\smile\hspace{-0.9em}\vert}_{B_n} \mathscr{A}_n$, $a \mathop{\smile\hspace{-0.9em}\vert}_{B_n} \mathscr{C}_n$ by monotonicity. By transitivity and monotonicity, $a \mathop{\smile\hspace{-0.9em}\vert}_{B_n} \mathscr{C}_{n+1}$. Finally, monotonicity gives us

$$a \mathop{\smile\hspace{-0.9em}\vert}_{B_n} b_{n+1},$$

and hence a contradiction. $\square_{2.8}$

Construction element Thus, fix $(\mathscr{A}_j)_{j \leqslant \omega}$, a sequence of $\mathbf{F}^{\mathfrak{M}}_{\lambda(\mathfrak{M})}$-saturated models of size λ, and an element a with the properties as in Lemma 2.7.

Construction element Let $\mathscr{A}_\omega$ be a $\mathbf{F}^{\mathfrak{M}}_{\lambda_r(\mathfrak{M})}$-primary model over

$$a \cup \bigcup_{i<\omega} \mathscr{A}_i,$$

the existence of which is guaranteed by Theorem 5.3 of [**9**] (a proof is in [**13**]).

2.2.3 The construction

Construction element For all $\eta \in \pi^{\leqslant \omega}$, using analogous reasoning to that found in Section 1 of [**6**] (discussion of which begins after Theorem 1.15 and continues through the proof of Lemma 1.17 of that paper), we define models $\mathscr{A}_\eta$ such that

- for all $\eta \in {}^{\leqslant\omega}\pi$, there is an automorphism f_η of $\mathfrak{M}$ such that
$$f_\eta(\mathscr{A}_{\text{length}(\eta)}) = \mathscr{A}_\eta;$$
- if η is an initial segment of ζ, then
$$f_\zeta \restriction_{\mathscr{A}_{\text{length}(\eta)}} = f_\eta \restriction_{\mathscr{A}_{\text{length}(\eta)}};$$

- if $\eta \in {}^{<\omega}\pi$, $\alpha \in \pi$, and X is the set of those $\eta \in {}^{\leqslant\omega}\pi$ such that $\eta \smile (\alpha)$ is an initial segment of ζ, then
$$\bigcup_{\zeta\in X} \mathscr{A}_\zeta \underset{\mathscr{A}_\eta}{\perp} \bigcup_{\zeta\in({}^{\leqslant\omega}\pi\setminus X)} \mathscr{A}_\zeta;$$
- for all $\eta \in {}^{\omega}\pi$, we let $a_\eta = f_\eta(a)$.

We recall a definition which will allow us to carry out the construction in an orderly and controlled manner.

Definition 2.9 ([**8**, Definition 3]) Assume that $J \subseteq {}^{\leqslant\omega}\pi$ is closed under initial segments and $K = \mathscr{P}^{<\omega}(J)$. Let $\Sigma = \{A_u : u \in K\}$ be an indexed family of subsets of $\mathfrak{m}$ of cardinality $< \mu$. We say that Σ is *strongly independent* if

(1) for all $u, v \in K$, $u \leq^K v \to A_u \subseteq A_v$;
(2) if $u, u_i \in K$, $i < n \in \omega$, and $B \subseteq \bigcup_{i<n} A_{u_i}$ has cardinality $< \pi$, then there is an automorphism $f = f^{\Sigma,B}_{(u,u_0,\dots,u_{n-1})}$ of $\mathfrak{m}$ such that $f \restriction_{(B\cap A_u)} = \mathrm{id}_{B\cap A_u}$ and $f(B \cap A_{u_i}) \subseteq A_{\mathrm{MinSetIn}(u,u_i)}$.

Construction element Define
$$A^i_u = \bigcup_{\eta\in u} \mathscr{A}_\eta,$$
for $u \in K_i$.

We can apply [**8**, Lemma 6] to find that $\{A^i_u : u \in K_i\}$ is strongly independent.

Construction element We apply [**8**, Lemma 4] to $\{A^i_u : u \in K_i\}$, and so find models $\mathscr{A}^i_u \preccurlyeq \mathfrak{m}$, $u \in K_i$ which satisfy the following properties:

(1) For all $u, v \in K_i$, $u \leq^K v$ implies $\mathscr{A}^i_u \subseteq \mathscr{A}^i_v$.
(2) For all $u \in K_i$, $\mathscr{A}^i_u$ is $\mathbf{F}^{\mathfrak{m}}_{\lambda_r(\mathfrak{m})}$-primary over A^i_u. This implies that $\bigcup_{u\in K_i} \mathscr{A}^i_u$ is a model.
(3) If $v \leq^K u$, then $\mathscr{A}^i_u$ is $\mathbf{F}^{\mathfrak{m}}_{\lambda_r(\mathfrak{m})}$-atomic over $\bigcup_{u\in K_i} A^i_u$ and $\mathbf{F}^{\mathfrak{m}}_{\lambda_r(\mathfrak{m})}$-primary over $\mathscr{A}^i_v \cup A^i_u$.
(4) Note further that if $J' \subseteq J_i$ is closed under initial segments, and $u \in P^{<\omega}(J')$, then the union $\bigcup_{v\in P^{<\omega}(J')} \mathscr{A}_v$ is $\mathbf{F}^{\mathfrak{m}}_{\lambda_r(\mathfrak{m})}$-constructible over $\mathscr{A}_u \cup \bigcup_{v\in P^{<\omega}(J')} A_v$.

These models $\mathscr{A}^i_u$ arise via a $\mathbf{F}^{\mathfrak{m}}_{\lambda_r(\mathfrak{m})}$-construction, with points a_γ, and sets B_γ, $\gamma < \alpha$ chosen appropriately. See proof of [**8**, Lemma 4] for full details.

In addition, note that by the proof of [**8**, Lemma 4 (Claim)], the families of models $\{\mathscr{A}^i_u : u \in K_i\}$, where $i = 0$ or $i = 1$ are strongly independent.

Construction element Denote by
$$\mathscr{A}^{J_i} = \bigcup_{u\in K_i} \mathscr{A}^i_u$$
the resulting constructed models given by [**8**, Lemma 4].

2.3 Non-isomorphism when symmetric difference of S-invariants is stationary

We next prove some general facts concerning models built as above on arbitrary trees $J, J' \subseteq \pi^{\leqslant\omega}$. We will later apply these results to J_0 and J_1.

Definition 2.10 Denote by I_{NS^π} the ideal of non-stationary sets on π.

For $J \subseteq \pi^{\leqslant\omega}$, let $J^\alpha = J \cap \alpha^{\leqslant\omega}$.

For $K = \mathscr{P}^{\leqslant\omega}(J)$, let $K^\alpha = \mathscr{P}^{\leqslant\omega}(J^\alpha)$.

Define the *S-invariant* of J to be

$$S(\overline{J}) = \left\{\delta : \exists\eta \in J^\delta \left(\eta \notin \bigcup_{\alpha<\delta} J^\alpha\right)\right\} \text{ modulo } I_{NS^\pi}.$$

Lemma 2.11 *Let $\mathscr{A}^J$ and $\mathscr{A}^{J'}$ be models constructed as above for trees $J, J' \subseteq \pi^{\leqslant\omega}$. Assume $S(J) \bigtriangleup S(J') = (S(J) \setminus S(J')) \cup (S(J') \setminus S(J))$ is stationary. Then $\mathscr{A}^J \not\cong \mathscr{A}^{J'}$.*

Proof. We follow [**8**, Lemma 8]. Assume for a contradiction that $f\colon \mathscr{A}^J \to \mathscr{A}^{J'}$ is an isomorphism. Let $\overline{J} = (J^\alpha)_{\alpha<\pi}$, $\overline{J'} = (J'^\alpha)_{\alpha<\pi}$. Let $K = \mathscr{P}^{\leqslant\omega}(J)$, $K' = \mathscr{P}^{\leqslant\omega}(J')$, and let $K^\alpha = \mathscr{P}^{\leqslant\omega}(J^\alpha)$, $K'^\alpha = \mathscr{P}^{\leqslant\omega}(J'^\alpha)$.

Let $\mathscr{A}_J^\alpha = \bigcup_{s\in G(K^\alpha)} \mathscr{A}_s$, where $G(K^\alpha)$ is the collection of good elements of K^α, as defined in Definition 2.6.

We can find α and α_i, $i < \omega$ such that:

- $\eta = (\alpha_i)_{i<\omega}$ is strictly increasing for all $i < \omega$;
- $\alpha = \bigcup_{i<\omega} \alpha_i \in S(J) \bigtriangleup S(J')$;
- the restrictions
$$f \restriction_{\mathscr{A}_J^\alpha}\colon \mathscr{A}_J^\alpha \xrightarrow{\cong} \mathscr{A}_{J'}^\alpha$$
and
$$f \restriction_{\mathscr{A}_J^{\alpha_i}}\colon \mathscr{A}_J^{\alpha_i} \xrightarrow{\cong} \mathscr{A}_{J'}^{\alpha_i},\ \forall i < \omega$$
are isomorphisms.

Without loss of generality, we can assume that $\alpha \in S(J) \setminus S(J')$ and thus that $\eta \in J \setminus J'$.

Claim 1 $a_\eta \not\downarrow_{\mathscr{A}_J^{\alpha_i}} \mathscr{A}_J^{\alpha_{i+1}}$.

Recall from the construction that

$$a_\eta \underset{\mathscr{A}_{\eta\restriction_i}}{\not\downarrow} \mathscr{A}_{\eta\restriction_{i+1}}.$$

Since $\mathscr{A}_{\eta\restriction_i} \subset \mathscr{A}_J^{\alpha_i}$ and $\mathscr{A}_{\eta\restriction_{i+1}} \subset \mathscr{A}_J^{\alpha_{i+1}}$, and $\mathscr{A}_{\eta\restriction_{i+1}} \not\subset \mathscr{A}_J^{\alpha_i}$, by monotonicity ([**9**, Lemma 3.2(i)]), we have

$$a_\eta \underset{\mathscr{A}_{\eta\restriction_i}}{\not\downarrow} \mathscr{A}_{\eta\restriction_{i+1}} \Rightarrow a_\eta \underset{\mathscr{A}_{\eta\restriction_i}}{\not\downarrow} \mathscr{A}_J^{\alpha_{i+1}}.$$

Claim 1* Thus, to prove Claim 1, it is enough to show that

$$a_\eta \underset{\mathscr{A}_J^{\alpha_i}}{\not\downarrow} \mathscr{A}_{\eta\restriction_{i+1}}.$$

Assume for a contradiction that $a_\eta \downarrow_{\mathscr{A}_J^{\alpha_i}} \mathscr{A}_{\eta\restriction_{i+1}}$.

Claim 2 $a_\eta \downarrow_{\mathscr{A}_J^{\alpha_i}} \mathscr{A}_{\eta\restriction_{i+1}} \Rightarrow \mathscr{A}_J^{\alpha_i} \not\downarrow_{\mathscr{A}_{\eta\restriction_i}} \mathscr{A}_{\eta\restriction_{i+1}}$.

By assumption, $a_\eta \downarrow_{\mathscr{A}_J^{\alpha_i}} \mathscr{A}_{\eta\restriction_{i+1}}$. This implies that $\mathscr{A}_{\eta\restriction_{i+1}} \downarrow_{\mathscr{A}_J^{\alpha_i}} a_\eta$. We get this symmetry by using monotonicity to find that $a_\eta \downarrow_{\mathscr{A}_J^{\alpha_i}} \overline{b}$ for any finite $\overline{b} \in \mathscr{A}_{\eta\restriction_{i+1}}$. Then, since $\mathscr{A}_J^{\alpha_i}$ is $\mathbf{F}^{\mathfrak{m}}_{\lambda(\mathfrak{m})}$-saturated by construction ([**8**]), and hence strongly $\mathbf{F}^{\mathfrak{m}}_{\kappa(\mathfrak{m})}$-saturated, by [**9**, Lemma 3.6], $\bar{b} \perp_{\mathscr{A}_J^{\alpha_i}} a_\eta$. Since this is true for all $\bar{b} \in \mathscr{A}_{\eta\restriction_{i+1}}$, we get

$$\mathscr{A}_{\eta\restriction_{i+1}} \underset{\mathscr{A}_J^{\alpha_i}}{\perp} a_\eta.$$

Now, assume for a contradiction that $\mathscr{A}_J^{\alpha_i} \perp_{\mathscr{A}_{\eta\restriction_i}} \mathscr{A}_{\eta\restriction_{i+1}}$. By a similar symmetry argument, $\mathscr{A}_{\eta\restriction_{i+1}} \perp_{\mathscr{A}_{\eta\restriction_i}} \mathscr{A}_J^{\alpha_i}$. Thus, we have

$$\mathscr{A}_{\eta\restriction_{i+1}} \underset{\mathscr{A}_J^{\alpha_i}}{\perp} a_\eta \text{ and } \mathscr{A}_{\eta\restriction_{i+1}} \underset{\mathscr{A}_{\eta\restriction_i}}{\perp} \mathscr{A}_J^{\alpha_i}.$$

In addition, by [**8**, Lemma 3.2(iii)], we have $a_\eta \perp_{\mathscr{A}_J^{\alpha_i}} \mathscr{A}_J^{\alpha_i}$. We can thus apply [**8**, Lemma 3.8(iii)] to find that

$$\mathscr{A}_{\eta\restriction_{i+1}} \underset{\mathscr{A}_{\eta\restriction_i}}{\perp} a_\eta \cup \mathscr{A}_J^{\alpha_i}.$$

By monotonicity and symmetry, we get $a_\eta \perp_{\mathscr{A}_{\eta\restriction_i}} \mathscr{A}_{\eta\restriction_{i+1}}$, a contradiction. ☑$_{\text{Claim 2}}$

Thus, with our assumptions so far, we have $\mathscr{A}_{\eta\restriction_{i+1}} \not\perp_{\mathscr{A}_{\eta\restriction_i}} \mathscr{A}_J^{\alpha_i}$. We now show that this dependence causes a contradiction.

Since $\mathscr{A}_{\eta\restriction_i}$ is sufficiently saturated, by [**9**, Corollary 3.5(i)], there is $c \in \mathscr{A}_J^{\alpha_i}$ such that

$$\mathscr{A}_{\eta\restriction_{i+1}} \underset{\mathscr{A}_{\eta\restriction_i}}{\not\perp} c.$$

Since $\mathscr{A}_J^{\alpha_i} = \bigcup_{s \in G(K^{\alpha_i})} \mathscr{A}_s$, there is a good $s \in K^{\alpha_i}$ such that $c \in \mathscr{A}_s$.

Now, let $r = \{\eta \restriction_j : j \leqslant i+1\}$. Then, r is good and $r \cap J^{\alpha_i} = \{\eta \restriction_j : j \leqslant i\}$. Without loss of generality, we can assume that $\eta \restriction_i \in s$, since $\mathscr{A}_s$ cannot get smaller with this assumption.

However, by strong independence (see [**8**]), $\mathscr{A}_r \perp_{\mathscr{A}_{r\cap s}} \mathscr{A}_s$, which by definition, written otherwise

$$\mathscr{A}_{\eta\restriction_{i+1}} \underset{\mathscr{A}_{\eta\restriction_i}}{\perp} \mathscr{A}_s.$$

This gives a contradiction since $c \in \mathscr{A}_s$. ☑$_{\text{Claim 1}}$

Thus, there is $u \in K'$ such that for all $i < \omega$, $\mathscr{A}_u \not\perp_{\mathscr{A}_{J'}^{\alpha_i}} \mathscr{A}_{J'}^{\alpha_{i+1}}$. However, since $\alpha \notin S(J')$, this contradicts [**8**, Lemma 7(ii)]. Since the notation we use here is rather different from that in [**8**], note that we can find a model with the properties of what is in [**8**] defined by $\mathcal{A}_v$ in $\mathscr{A}_{J'}^{\alpha_i}$. Recall that $\mathscr{A}_{J'}^{\alpha_{i+1}}$ can be written as a union of appropriate models as in the notation found in [**8**]. $\square_{2.11}$

Corollary 2.12 *Let $\mathscr{A}^J$ and $\mathscr{A}^{J'}$ be models constructed as above for trees $J, J' \subseteq \pi^{\leqslant\omega}$. Assume that $S(J) = S \subset S^\pi_\omega$ and $S(J') = \emptyset$, thus $S(J) \bigtriangleup S(J') = S$ is stationary. Then $\mathscr{A}^J \not\cong \mathscr{A}^{J'}$ in any cardinal- and $\mathscr{P}(\lambda_r(\mathfrak{m}))$-preserving extension of the universe where the symmetric difference $S(J) \bigtriangleup S(J')$ remains stationary.*

Notice that the proof of Lemma 2.11 is in ZFC. In particular, the notion of independence is absolute for models where no small (of size $< \lambda_r(\mathfrak{m})$) subsets are added. Thus, two models $\mathscr{A}_J$ and $\mathscr{A}_{J'}$ which are non-isomorphic in the ground model remain non-isomorphic in any cardinal- and $\mathscr{P}(\lambda_r(\mathfrak{m}))$-preserving extension of the universe where the symmetric difference $S(J) \bigtriangleup S(J')$ remains stationary.

It is easy to see that $S(J_0) = S$ and $S(J_1) = \emptyset$. Thus, we can apply the previous lemma to find that $\mathscr{A}_{J_0} \ncong \mathscr{A}_{J_1}$ in $\mathbb{L}$.

2.4 Isomorphism of the models when S is killed

Theorem 2.13 *Assume that $J_0 \cong J_1$ in some extension of the set-theoretic universe which preserves cardinals and $\mathscr{P}(\lambda_r(\mathfrak{M}))$. Then, in that extension, $\mathscr{A}_{J_0} \cong \mathscr{A}_{J_1}$.*

Proof. Assume that $F\colon J_0 \to J_1$ is an isomorphism. We aim to find an isomorphism between $\mathscr{A}_{J_0}$ and $\mathscr{A}_{J_1}$.

We proceed by induction on good elements of K_0 along the ordering $\leq^K$ by building elementary maps G_u, $u \in K_0$. We ensure in this induction that if $u_i \leq^K u_j$ and $u_j \nleq^K u_i$ then G_{u_i} is constructed before G_{u_j}.

Base case: isomorphism for the first level of the tree G_0. For all $u \in K_0 = \mathscr{P}^{<\omega}(J_0)$, let $F(u) = \{F(\eta) : \eta \in u\}$. For $\eta \in J_0$, let $G_0 \restriction_{\mathscr{A}_\eta} = f_{F(\eta)} \circ f_\eta^{-1}$, where the f_η are as defined in Section 2.2.3. Thus,

$$G_0 : \bigcup_{\eta \in J_0} \mathscr{A}_\eta \longrightarrow \bigcup_{\eta \in J_1} \mathscr{A}_\eta.$$

Claim 3 The function G_0, which maps one strongly independent family to the other, is elementary.

We prove the claim by induction on good elements $s \in K_0$ along the ordering $\leq$. Denote by $G_0^\eta = f_{F(\eta)} \circ f_\eta^{-1}$, and by $G_0^s = \bigcup_{\xi \in s} G_0^\xi$, for $s \in G(J_0)$.

By construction, G_0^η, $\eta \in J_0$ is elementary.

Now, assume that G_0^s has been shown to be elementary. We wish to show that $G_0^{s'}$ for $s' \geq^K s$ is also elementary. Our ordering of $G(J_0)$ implies that it is enough to consider $s' = s \cup \{\eta\}$ for some $\eta \in J_0$. We thus have two cases: $\eta \in \pi^{<\omega}$ or $\eta \in \pi^\omega$. The arguments for both are similar.

If $\eta \in \pi^{<\omega}$, denote by $\eta^- = \eta \restriction_{(\text{length}(\eta)-1)}$. If $\eta \in \pi^\omega$ is an infinite branch, then we can then find $i < \omega$ such that $\forall \xi \in s$, $\xi \not> \eta \restriction_i$. Denote by $\eta^- = \eta \restriction_{i-1)}$.

Since we are working in a homogeneous monster model $\mathfrak{M}$, we can assume without loss of generality that $G_0^s \restriction_{A_s} = \mathrm{id}_{A_s}$.

In addition, we know from the construction that

$$\mathrm{tp}_{\mathfrak{M}}(\mathscr{A}_\eta / \mathscr{A}_{\eta^-}) = \mathrm{tp}_{\mathfrak{M}}(G_0^\eta(\mathscr{A}_\eta)/\mathscr{A}_{\eta^-}),$$

because G_0^η is elementary and $G_0^\eta \restriction_{\mathscr{A}_\eta} = \mathrm{id}$. We thus want to show that

$$\mathrm{tp}_{\mathfrak{M}}(\mathscr{A}_\eta / A_s) = \mathrm{tp}_{\mathfrak{M}}(G_0^\eta(\mathscr{A}_\eta)/A_s).$$

Since $\mathscr{A}_\eta$ is $\mathbf{F}^{\mathfrak{M}}_{\lambda(\mathfrak{M})}$-saturated, these types are stationary. Therefore, by definition of stationarity ([**9**, Definition 3.3]) it is enough to show that

$$\mathscr{A}_\eta \underset{\mathscr{A}_{\eta^-}}{\downarrow} A_s \text{ and } G_0^\eta(\mathscr{A}_\eta) \underset{\mathscr{A}_{\eta^-}}{\downarrow} A_s.$$

However, note that $\eta^- \in s = f(s)$, thus we have the independence by construction, and so the embedding is elementary. ☑Claim 3

Before we continue with the next step of the induction, we give some notation and reminders. Denote $\mathscr{A}_{\{\eta\}} = \mathscr{A}_\eta$. Recall that, by the construction, $A_u = \bigcup_{\eta \in u} \mathscr{A}_\eta$ for $u \in K_i$, and $\mathscr{A}_u$ is $\mathbf{F}^{\mathfrak{M}}_{\lambda_r(\mathfrak{M})}$-prime over A_u.

Ultimately, we aim to build an isomorphism $G\colon \mathscr{A}^{J_0} \to \mathscr{A}^{J_1}$ such that $G \restriction_{\mathscr{A}_\eta} = \mathrm{id}_{\mathscr{A}_\eta}$ for all $\eta \in J_i$. Since $\mathscr{A}^{J_i} = \bigcup_{u\in K_i} \mathscr{A}^i_u$, it is enough to construct $G_u\colon \mathscr{A}^0_u \to \mathscr{A}^1_u$ such that if $t \leq^K u$, then $G_t \subseteq G_u$. If we can show that $\bigcup_{t\leq^K u} G_t$ is elementary, then using homogeneity of $\mathfrak{M}$, we can find the desired isomorphism G_u. The full isomorphism will then be $G = \bigcup_{u\in G(K_i)} G_u$.

Inductive step. Assume we have shown that, for all $t \lneq u$, G_t are isomorphisms. We build an isomorphism $G_u\colon \mathscr{A}^0_u \to \mathscr{A}^1_u$.

Claim 4 The function $\bigcup_{t<^K u} G_t$ is elementary.

We assume for a contradiction that the inductive step fails at some point. Let u be the $\leq^K$-smallest such that $\bigcup_{t\leq^K u} G_t = G^*$ is not elementary.

This failure of elementariness is witnessed by some set of finitely many points $a_0, \dots, a_m \in \bigcup_{t\leq^K u} \mathscr{A}_t$. Then, in particular, $G^* \restriction_{\{a_0,\dots,a_m\}}$ is not elementary.

Subclaim 1 The points $a_0, \dots, a_m$ can be replaced with tuples $\overline{a}_i$, $i = 0, \dots, n$ which appear all at once at some step in the construction, that is, $\overline{a}_i \in \mathscr{A}_{t_i}$ and $\overline{a}_i \cap \bigcup_{t<^K t_i} \mathscr{A}_i = \emptyset$.

Consider $a_0, \dots, a_m$. For all $i \leqslant m$, there is $t^i <^K u$ such that $a_i \in \mathscr{A}_{t^i} \setminus \bigcup_{t<^K t^i} \mathscr{A}_t$. Let $\{t_0, \dots, t_n\}$ be an enumeration of the t^i so that $t_i \neq t_j$ if $i \neq j$ (i.e., we get rid of repetitions). In addition, we can assume without loss of generality that t_n is maximal in $\{t_0, \dots, t_n\}$ with respect to the ordering $\leq^K$.

Define $\overline{a}_i = \{a_j : t^j = t_i\}$. Then $\overline{a}_i$ is the desired tuple such that $\overline{a}_i \in \mathscr{A}_{t_i}$ and $\overline{a}_i \cap \bigcup_{t<t_i} \mathscr{A}_i = \emptyset$. ☑Subclaim 1

To save ink, we will denote the tuples $\overline{a}_i$ as a_i, and now consider the finite set of tuples $\{a_0, \dots, a_n\}$.

We wish to refine this choice of witnesses $\{a_0, \dots, a_n\}$ to minimize the t_n and the number n. To this end, we devise an ordering on $\mathscr{P}^{\leqslant\omega}(K_i)$:

Definition 2.14 For $t_i, u_i \in K_i$, we say that $\{t_i : i \leqslant n\} \Subset \{u_i : i \leqslant m\}$ iff for all $i \leqslant n$ there is $j \in m$ such that $t_i \leq u_j$ and there is u_j such that $u_j \not\leq t_i$ for every $i \leqslant n$.

We can minimize the choice of witnesses $\{a_0, \dots, a_n\}$ easily if there are only finitely many candidates which may be smaller than our initial choice. We will assume otherwise, and, using Ramsey's Theorem, come to a contradiction. Thus, assume for a contradiction that there are infinitely many choices of witnesses $\{a_0, \dots, a_n\} = \{a^0_0, \dots, a^0_{n_0}\}$, $\{a^1_0, \dots, a^1_{n_1}\}, \dots, \{a^j_0, \dots, a^j_{n_j}\}, \dots$ from $\mathscr{P}^{<\omega}(K)$ for which the associated $\{t_0, \dots, t_n\} = \{t^0_0, \dots, t^0_{n_0}\}$, $\{t^1_0, \dots, t^1_{n_1}\}$, $\dots$, $\{t^j_0, \dots, t^j_{n_j}\}$, $\dots$, are $\Subset$ than our original choice. These are quasi-ordered by $\Subset$.

Subclaim 2 The collection

$$\{t^0_0, \dots, t^0_{n_0}\}, \{t^1_0, \dots, t^1_{n_1}\}, \dots, \{t^j_0, \dots, t^j_{n_j}\}, \dots$$

is a quasi-ordering with no $\Subset$-infinite descending sequences.

For notational simplicity, we will write $X_j = \{t^j_i : i < n_j\}$, and consider them with the ordering $\Subset$.

Assume for a contradiction that there is an infinite descending chain. We assume, without loss of generality, that this chain is enumerated so that $X_{j+1} \Subset X_j$.

Let $u_j \in X_j$ be such that $u_j \not\leq^K t_k^{j+1}$ for every $k < n_{j+1}$ (by definition of $\Subset$, there is at least one such $u_j \in X_j$ for every j).

Thus, for all $j < i < \omega$, $u_j \not\leq^K u_i$. This is because if $i = j+1$, then this is simply the definition of u_j, and otherwise we can find $k < n_{j+1}$ such that $u_i \leq^K t_k^{j+1}$. So, if $u_j \leq^K u_i$, then $u_j \leq^K t_k^{j+1}$, a contradiction with the definition of u_i.

Since the u_j are finite antichains in J_i, it is easy to see that $\bigcup\{u_j : j < \omega\}$ does not contain infinite decreasing $\leq^J$-chains. By the same argument, there are also no infinite increasing $\leq^J$-sequences.

By Ramsey's Theorem, there must thus be an infinite $\leq^J$-antichain. Thus, we can find t_i^0, $i < n_0$, and an infinite set $X \subseteq \omega$ such that $\{u_j : j \in X\}$ is an $\leq^K$-antichain, and $u_j \leq^K t_i^0$ for all $j \in X$.

Let T be the tree composed of $\eta \in J$, such that $\eta < \xi$ for some $\xi \in t_i^0 \subset J$. We show that since such a tree has no maximal branches, the existence of an infinite $\leq^K$-antichain is not possible.

Note that for all $j < i$ and k, there is n such that $t_k^i \leq^K t_n^j$.

Without loss of generality, we can assume that $u_j = \{u_i^j : i < m\}$. To ensure this, we may need to make X smaller so that $|u_j| \leqslant n_0$, for all $i \in X$.

By applying the Ramsey Theorem m times, we can assume that one of the following holds for all $i < m$:

(1) for all $j < k$, $u_i^k <^K u_i^k$;
(2) for all $j < k$, $u_i^j \perp^K u_i^k$;
(3) for all $j < k$, $u_i^j \geq^K u_i^k$.

Clearly case 1 is not possible. Furthermore, it is not possible for case 3 to hold for all $i < m$. Thus, let $i < m$ be such that 2 holds. Then $\{u_i^j : j \in X\}$ is an infinite $\leq^J$-antichain in T, a contradiction. ☑$_{\text{Subclaim 2}}$

Assume now that our choice of $\{a_0, \ldots, a_n\}$ and $\{t_0, \ldots, t_n\}$ is minimal in $<^K$.

There is $C \subset \bigcup_{t<t_n} \mathscr{A}_t$, $|C| = \lambda_r(\mathfrak{M})$ such that

$$\operatorname{tp}(a_n/C) \models \operatorname{tp}(a_n/\bigcup_{t<t_n} \mathscr{A}_t).$$

Let $B = C \cup \{a_0, \ldots, a_{n-1}\}$.

On the one hand, let $H = f^B_{(t_n,t_0,\ldots,t_{n-1})}$ be as in Definition 2.9. That is, H is an automorphism of $\mathfrak{M}$ such that $H \restriction_{(B\cap\mathscr{A}_{t_n})} = \mathrm{id}_{B\cap\mathscr{A}_{t_n}}$ and, for $i < n$,

$$H(B \cap \mathscr{A}_{t_i}) \subseteq \mathscr{A}_{\mathrm{MinSetIn}(t_n,t_i)}.$$

Then, $H(a_i) \in \mathscr{A}_{\mathrm{MinSetIn}(t_n,t_i)}$. Since $\mathrm{MinSetIn}(t_n,t_i) < t_n$, $H(a_i) \in \bigcup_{t'<t_n} \mathscr{A}_{t'}$. Since $H \restriction_C = \mathrm{id}$, we have

$$\operatorname{tp}(a_0, \ldots, a_{n-1}/C) = \operatorname{tp}(H(a_0), \ldots, H(a_{n-1})/C)$$

and

$$\operatorname{tp}(a_n/C) \models \operatorname{tp}(a_n/C \cup \{H(a_0), \ldots, H(a_{n-1})\}),$$

so

$$\operatorname{tp}(a_n/C) \models \operatorname{tp}(a_n/C \cup \{a_0, \ldots, a_{n-1}\}).$$

On the other hand, consider G^*. Since $\{a_0, \ldots, a_n\}$ is a minimal witness that G^* is not elementary, the function $G^* \restriction_{C\cup\{a_0,\ldots,a_{n-1}\}}$ must be elementary.

Let G^+ be an automorphism of $\mathfrak{M}$ such that $G^+ \circ G^* \restriction_{C\cup\{a_0,\dots,a_{n-1}\}} = \mathrm{id}$.
Since $G^* \restriction_{\mathscr{A}_{t_n}} = G_{t_n}$ and $C \subseteq \mathscr{A}_{t_n}$, $G^* \restriction_{C\cup a_n}$ is elementary. Thus

$$\mathrm{tp}(G^+(G^*(a_n))/C) = \mathrm{tp}(a_n/C).$$

However,

$$\mathrm{tp}(G^+(G^*(a_n)), a_0, \dots, a_{n-1}/C) = \mathrm{tp}(G^*(a_n), G^*(a_0), \dots, G^*(a_{n-1})/G^*(C)),$$

thus

$$\mathrm{tp}(G^*(a_n), G^*(a_0), \dots, G^*(a_{n-1})/\emptyset) \models \mathrm{tp}(a_n, a_0, \dots, a_{n-1}/\emptyset).$$

This means that

$$\mathrm{tp}(a_n/C) \not\models \mathrm{tp}(a_n/C \cup a_0, \dots, a_{n-1}),$$

a contradiction. ☑$_{\text{Claim 4}}$

$\square_{2.13}$

Corollary 2.15 *Let $\mathscr{A}^{J_0}$ and $\mathscr{A}^{J_1}$ be models constructed as above for trees J_0 and J_1. Assume that in a cardinal-preserving extension of the universe, $S(J_0)$ is not stationary. Then $\mathscr{A}^{J_0} \cong \mathscr{A}^{J_1}$.*

Proof. Lemmas 7.15 and 7.31 of [**5**] demonstrate this in the extension, $J_0 \cong J_1$. We can then apply the previous Theorem 2.13. $\square_{2.15}$

2.5 Constructibility with respect to $0^\#$

We now have all the necessary ingredients to prove Theorem 2.1.

Proof. The result is a direct result of Theorem 1.4 and Corollaries 2.12 and 2.15. $\square_{2.1}$

References

[1] Sy-David Friedman. Cardinal-preserving extensions. *Journal of Symbolic Logic*, 68(4):1163–1170, 2003.

[2] Sy-David Friedman, Tapani Hyttinen, and Mika Rautila. Classification theory and $0^\#$. *Journal of Symbolic Logic*, 68(2):580–588, 2003.

[3] Rami Grossberg and Olivier Lessmann. Shelah's stability spectrum and homogeneity spectrum in finite diagrams. *Archive for Mathematical Logic*, 41(1):1–31, 2002.

[4] Rami Grossberg and Saharon Shelah. On the number of nonisomorphic models of an infinitary theory which has the infinitary order property. Part A. *Journal of Symbolic Logic*, 51(2):302–322, 1986.

[5] Taneli Huuskonen, Tapani Hyttinen, and Mika Rautila. On potential isomorphism and non-structure. *Archive for Mathematical Logic*, 43(1):85–120, 2004.

[6] Tapani Hyttinen. On nonstructure of elementary submodels of an unsuperstable homogeneous structure. *Mathematical Logic Quarterly*, 43(1):134–142, 1997.

[7] Tapani Hyttinen. *A Short Introduction to Classification Theory*, volume 2 of *Graduate Texts in Mathematics.* Department of Mathematics, University of Helsinki, Helsinki, 1997.

[8] Tapani Hyttinen and Saharon Shelah. On the number of elementary submodels of an unsuperstable homogeneous structure. *Mathematical Logic Quarterly*, 44:354–358, 1998.

[9] Tapani Hyttinen and Saharon Shelah. Strong splitting in stable homogeneous models. *Annals of Pure and Applied Logic*, 103(1–3):201–228, 2000.

[10] Tapani Hyttinen and Saharon Shelah. Main gap for locally saturated elementary submodels of a homogeneous structure. *Journal of Symbolic Logic*, 66(3):1286–1302, 2001.

[11] Tapani Hyttinen and Heikki Tuuri. Constructing strongly equivalent nonisomorphic models for unstable theories. *Annals of Pure and Applied Logic*, 52(3):203–248, 1991.

[12] Thomas Jech. *Set Theory.* The Third Millennium Edition, Revised and Expanded. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2003.

[13] Saharon Shelah. Finite diagrams stable in power. *Annals of Mathematical Logic*, 2:69–118, 1970.

[14] Saharon Shelah. Existence of many $L_{\infty,\lambda}$-equivalent, nonisomorphic models of T of power λ. *Annals of Pure and Applied Logic*, 34:291–310, 1987.

[15] Saharon Shelah. *Classification theory and the number of non-isomorphic models*, volume 92 of *Studies in Logic and the Foundations of Mathematics*. North-Holland, Amsterdam, 2nd revised edition, 1990.

On absoluteness of categoricity in AEC's

Sy-David Friedman[†], **Martin Koerwien**[‡]

[†] Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
sdf@logic.univie.ac.at

[‡] Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
koerwien@math.uic.edu

Abstract. Shelah has shown in [**4**] that $\aleph_1$-categoricity for Abstract Elementary Classes (AEC's) is not absolute in the following sense: There is an example K of an AEC (which is actually axiomatizable in the logic $L(Q)$) such that if $2^{\aleph_0} < 2^{\aleph_1}$ (the weak CH holds) then K has the maximum possible number of models of size $\aleph_1$, whereas if Martin's Axiom at $\aleph_1$ (denoted by $\mathrm{MA}_{\aleph_1}$) holds then K is $\aleph_1$-categorical. In this note we extract the properties from Shelah's example which make both parts work resulting in our definitions of condition A and condition B, and then we show that for any AEC satisfying these two conditions, neither of these implications can be reversed.

1 The model theoretic context

In Shelah's paper [**4**], the notion of *Abstract Elementary Classes* (AEC) was introduced, the idea being to write down basic properties of the first order elementary substructure relation.

Definition 1.1 Let K be a class of models of a given similarity type and let $\prec$ be a partial ordering on K refining the ordinary substructure relation. The pair $\mathcal{K} = (K, \prec)$ is an *AEC* if

(1) both K and $\prec$ are closed under isomorphism;
(2) $A \prec C$, $B \prec C$ and $A \subset B$ imply $A \prec B$;
(3) for any continuous $\prec$-chain $(A_\alpha)_{\alpha<\lambda}$,
  (a) $A = \bigcup_{\alpha<\lambda} A_\alpha \in K$;
  (b) for all $\alpha < \lambda$, $A_\alpha \prec A$;
  (c) if $A_\alpha \prec B$ for some B and all $\alpha < \lambda$, then $A \prec B$;
(4) there is a cardinal $\mathrm{LS}(K)$ such that for all $A \in K$ and any *subset* $A_0 \subset A$, there is $B \prec A$ containing A_0 with $|B| \leq |A_0| + \mathrm{LS}(K)$.

Many non-elementary classes can be made an AEC with appropriate relations $\prec$, such as classes axiomatized using an additional quantifier Q saying "there are uncountably many" (we will see an example of this later), or classes axiomatized by $L_{\omega_1,\omega}$-sentences (first

Appeared in Notre Dame Journal of Formal Logic, vol. 52(4), 2011, pp. 395–402.

[†]Partially supported by the Austrian Science Fund (FWF) under Project Number P 19375-N18 as well as by the John Templeton Foundation under Grant #13152, *The Myriad Aspects of Infinity.*

[‡]Partially supported by ESF short visit INFTY grant number 2986, as well as by the John Templeton Foundation under Grant #13152, *The Myriad Aspects of Infinity.* This author expresses his gratitude to the Mittag-Leffler Institute for hosting him during a very productive stay in which a major part of the present article has been developed. Special thanks also to John Baldwin and Tapani Hyttinen for helpful discussions.

order with infinite countable conjunctions and disjunctions) with $\prec$ being elementary substructure with respect to some countable fragment of $L_{\omega_1,\omega}$.

It becomes an interesting question to what extent results of first order model theory such as Morley's categoricity theorem extend to arbitrary AEC, or perhaps to AEC with some special properties. Some work in this direction is exposed in Baldwin's book [**2**], which has a particular emphasis on $L_{\omega_1,\omega}$.

2 Model theoretic properties: condition A and B

We now introduce two properties AEC can have. First, we have to fix some notation.

Notation 2.1 Let $(M_\alpha)_{\alpha<\beta}$ and $(N_\alpha)_{\alpha<\beta}$ be continous, strictly increasing (with respect to inclusion) sequences of structures.

- We write $(M_\alpha)_{\alpha<\beta} \cong (N_\alpha)_{\alpha<\beta}$ if there is a function $f\colon \bigcup_{\alpha<\beta} M_\alpha \to \bigcup_{\alpha<\beta} N_\alpha$ such that, for all $\alpha < \beta$, $f \restriction M_\alpha$ is an isomorphism between M_α and N_α. We call such an f a *filtration automorphism* if $M_\alpha = N_\alpha$ for all $\alpha < \beta$.
- Define rank: $\bigcup_{\alpha<\beta} M_\alpha \to \beta$ by $\mathrm{rank}(a) = \min\{\alpha \mid a \in M_\alpha\}$. Note that, by continuity of the chain, the range of rank is precisely the set of countable successor ordinals together with zero.
- For any finite tuple $\overline{a}$ in $\bigcup_{\alpha<\beta} \mathrm{M}_\alpha$ and $\alpha < \beta$, let $\overline{a}_\alpha$ be the subtuple of $\overline{a}$ of elements of rank α.
- Considering a tuple $\overline{a} = (a_0, a_1, \ldots, a_{n-1})$ as a function whose domain is $n = \{0, 1, \ldots, n-1\}$ (via $\overline{a}(i) = a_i$), let $s_{\overline{a}} = \mathrm{rank} \circ \overline{a}$ (i.e., $s_{\overline{a}}(i) = \mathrm{rank}(a_i)$ for all $i < n$).
- Let $\mathrm{tp}_{\mathrm{qf}}(\overline{a})$ denote the quantifier free type of $\overline{a}$ (over the empty set).

Definition 2.2 Let $(\mathbb{K}, \prec)$ be an AEC in a relational signature with Löwenheim–Skolem number $\aleph_0$. We say that

(1) $(\mathbb{K}, \prec)$ satisfies *condition A* if it is $\aleph_0$-categorical and fails amalgamation for countable models (i.e., there is a triple of countable models $M_0 \prec M_1, M_2$ such that there are no countable M_3 and embeddings $f_i\colon M_i \to M_3$ $(i = 1, 2)$ with $f[M_i] \prec M_3$ and $f_1 \restriction M_0 = f_2 \restriction M_0$).

(2) $(\mathbb{K}, \prec)$ satisfies *condition B* if there is an increasing and continous $\prec$-chain $(M_\alpha)_{\alpha<\omega_1}$ of countable models such that:

(i) (Decomposition) Any $N \in \mathbb{K}$ of size $\aleph_1$ can be written as $N = \bigcup_{\alpha<\omega_1} N_\alpha$ with $(N_\alpha)_{\alpha<\beta} \cong (M_\alpha)_{\alpha<\beta}$ for all $\beta < \omega_1$.

(ii) (Triviality) For any $N = \bigcup_{\alpha<\omega_1} N_\alpha$ as in (i), and any finite tuples $\overline{a}, \overline{b}, \overline{c}$ in N with $\max(s_{\overline{c}}) < \min(s_{\overline{a}})$, if $s_{\overline{b}} = s_{\overline{a}}$ and, for all α, $\mathrm{tp}_{\mathrm{qf}}(\overline{b}_\alpha\overline{c}) = \mathrm{tp}_{\mathrm{qf}}(\overline{a}_\alpha\overline{c})$, then $\mathrm{tp}_{\mathrm{qf}}(\overline{b}\overline{c}) = \mathrm{tp}_{\mathrm{qf}}(\overline{a}\overline{c})$.

(iii) (Homogeneity) Suppose $N = \bigcup_{\alpha<\omega_1} N_\alpha$ is as in (i) and $\overline{a}, \overline{b}$ are finite tuples in N such that there is an isomorphism $f\colon \overline{a} \to \overline{b}$ with $x \in N_\alpha$ if and only if $f(x) \in N_\alpha$ for all $x \in \mathrm{dom}(f)$ and $\alpha < \omega_1$. Then for any $\beta > \max(s_{\overline{a}}), \max(s_{\overline{b}})$, there is a filtration automorphism of $(N_\alpha)_{\alpha<\beta}$ extending f.

3 How set theory affects the number of models

Theorem 3.1 *If $2^{\aleph_0} < 2^{\aleph_1}$ and condition A holds, then $\mathbb{K}$ has $2^{\aleph_1}$ many non-isomorphic models of size $\aleph_1$.*

Proof. This result and its proof are exposed in [**2**, Theorem 17.11]. □

The proof of the following result is an abstract version of the proof given for Shelah's specific $L(Q)$-example (Theorem 6.6 in [**4**]). A simpler version can also be found in [**2**].

Theorem 3.2 *Martin's Axiom at $\aleph_1$ and condition B imply that $\mathbb{K}$ is $\aleph_1$-categorical.*

Proof. Let $N^i = \bigcup_{\alpha<\omega_1} N^i_\alpha$ (for $i < 2$) with $(N^i_\alpha)_{\alpha<\beta} \cong (M_\alpha)_{\alpha<\beta}$ for all $\beta < \omega_1$ (by decomposition). Let $\mathcal{F}$ be the set of finite partial isomorphisms f from N^0 to N^1 with $x \in N^0_\alpha$ if and only if $f(x) \in N^1_\alpha$ for all $x \in \operatorname{dom}(f)$ and $\alpha < \omega_1$. We show that the partial order $(\mathcal{F}, \supset)$ has the ccc:

Let $\{f_i \mid i < \omega_1\} \subset \mathcal{F}$. We attempt to find two distinct f_i whose union is an element of $\mathcal{F}$. By simple applications of the delta system lemma and the pigeonhole principle, we can assume the following:

- There is some $n < \omega$ such that, for all $i < \omega_1$, $|\operatorname{dom}(f_i)| = |\operatorname{ran}(f_i)| = n$.
- The sets $\{\operatorname{dom}(f_i) \mid i < \omega_1\}$ and $\{\operatorname{ran}(f_i) \mid i < \omega_1\}$ are delta systems with roots r and r' respectively, and for any $i < \omega_1$, $\max(s_r) < \min(s_{\operatorname{dom}(f_i)\setminus r})$ and $\max(s_{r'}) < \min(s_{\operatorname{ran}(f_i)\setminus r'})$.
- For all $i < j < \omega_1$, $f_i \restriction r = f_j \restriction r$ and $\operatorname{ran}(f_i \restriction r) = r'$.
- (Filtration disjointness) For all $i < j < \omega_1$, $\operatorname{ran}(s_{\operatorname{dom}(f_i)\setminus r})$ is disjoint from $\operatorname{ran}(s_{\operatorname{dom}(f_j)\setminus r})$ (and thus, since the f_i preserve the filtrations, the same holds for the ranges).

Now we claim that actually the union of any two f_i is an element of $\mathcal{F}$. Take $i < j < \omega_1$ and set $g = f_i \cup f_j$. Let $\overline{a} = \operatorname{dom}(f_i) \setminus r$, $\overline{b} = \operatorname{dom}(f_j) \setminus r$. For any relation symbol R in our signature, we want to show that $N^0 \models R(\overline{a}, \overline{b}, r)$ holds if and only if

$$N^1 \models R(g(\overline{a}), g(\overline{b}), r')$$

(not all elements of the tuples may actually occur in R). Let $\gamma < \omega_1$ be greater than $\max(s_{\overline{a}})$ and $\max(s_{\overline{b}})$ and (by decomposition) pick any h witnessing $(N^0_\alpha)_{\alpha<\gamma} \cong (N^1_\alpha)_{\alpha<\gamma}$. By homogeneity, we can assume that $h \restriction r = f_i \restriction r (= f_j \restriction r)$. Because $f_i, f_j \in \mathcal{F}$, $\operatorname{tp}_{\mathrm{qf}}(g(\overline{a}), r') = \operatorname{tp}_{\mathrm{qf}}(\overline{a}, r) = \operatorname{tp}_{\mathrm{qf}}(h(\overline{a}), r')$ and $\operatorname{tp}_{\mathrm{qf}}(g(\overline{b}), r') = \operatorname{tp}_{\mathrm{qf}}(\overline{b}, r) = \operatorname{tp}_{\mathrm{qf}}(h(\overline{b}), r')$ and thus by triviality (using filtration disjointness),

$$\operatorname{tp}_{\mathrm{qf}}(h(\overline{a}), h(\overline{b}), r') = \operatorname{tp}_{\mathrm{qf}}(g(\overline{a}), g(\overline{b}), r'). \tag{3.1}$$

This means that $N^0 \models R(\overline{a}, \overline{b}, r)$ if and only if $N^1 \models R(h(\overline{a}), h(\overline{b}), r')$ (h is an isomorphism), if and only if, by (3.1), $N^1 \models R(g(\overline{a}), g(\overline{b}), r')$. This finishes the proof of ccc.

Now we prove that $D_a = \{f \in \mathcal{F} \mid a \in \operatorname{dom}(f)\}$ and $R_b = \{f \in \mathcal{F} \mid b \in \operatorname{ran}(f)\}$ (for $a \in N^0$, $b \in N^1$) are dense in $(\mathcal{F}, \supset)$. Take any $g \in \mathcal{F}$, $a \in N^0$ and, using decomposition, an h witnessing $(N^0_\alpha)_{\alpha<\beta} \cong (N^1_\alpha)_{\alpha<\beta}$ for some β greater than $\max(s_{\operatorname{dom}(g)})$ and $\max(s_a)$. By homogeneity, there is a filtration automorphism k of $(N^1_\alpha)_{\alpha<\beta}$ mapping $h[\operatorname{dom}(g)]$ to $\operatorname{ran}(g)$ such that on $\operatorname{dom}(g)$ we have $k \circ h = g$. Now, $g' = k \circ h \restriction (\operatorname{dom}(g) \cup \{a\})$ is an extension of g with $g' \in D_a$. The same argument also works for R_b.

Finally, we apply Martin's Axiom to the partial order $(\mathcal{F}, \supset)$ in order to obtain a $\{D_a \mid a \in N^0\} \cup \{R_b \mid b \in N^1\}$-generic filter G. Then $\bigcup G$ is a total isomorphism between N^0 and N^1. Since the N^i were arbitrary models in $\mathbb{K}$ of size $\aleph_1$, $\aleph_1$-categoricity of $\mathbb{K}$ follows. □

The example given in the proof of the following theorem is due to Shelah and can be found in [**4**].

Theorem 3.3 *There is an AEC satisfying both condition A and condition B.*

Proof. Let ψ be the $L_{\omega_1,\omega}(Q)$-sentence in the signature $L = \{P, Q, R, E\}$ (P, Q unary predicates, R, E binary relations) stating:

(1) P, Q partition the universe and P is infinite, countable.

(2) E is an equivalence relation on Q with infinitely many classes, each countably infinite.

(3) $R \subset P \times Q$ has the following properties:

(3a) For any finite disjoint $F, G \subset Q$, there is some $a \in P$ such that, for all $b \in F \cup G$, $R(a, b)$ if and only if $b \in F$.

(3b) For any finite disjoint $F, G \subset P$, there is some $b \in Q$ *in each E-class* such that, for all $a \in F \cup G$, $R(a, b)$ if and only if $a \in F$.

It is easy to see that $\mathbb{K} = \mathrm{mod}(\psi)$ together with the substructure relation $\prec$ defined by

$$M \prec N \text{ if and only if } M \subset N,\ P^M = P^N \text{ and}$$
$$\text{no element of } N \setminus M \text{ is } E\text{-equivalent to an element of } M$$

is an AEC with $\mathrm{LS}(K) = \aleph_0$. Note that, by (3a), in any model of ψ, the collection of all sets $A_q = \{p \in P \mid R(p, q)\}$ $(q \in Q)$ is an *independent family* in the sense that any intersection of finitely many distinct sets or their complements is non-empty.

Amalgamation fails for countable models: take for M_0 any countable model and let M_1, M_2 be extensions where we add one E-class B_1, B_2 respectively to M_0 such that there are $b_1 \in B_1$ and $b_2 \in B_2$ with $R(a, b_1)$ if and only if $\neg R(a, b_2)$ for all $a \in P$. Such extensions exist by the facts that countable independent families are not maximal (even with the additional requirement of (3b)), and that an independent family stays independent if we replace some set with its complement.

Clearly, M_1 and M_2 do not amalgamate over M_0 because the amalgam would fail property (3a).

Now let M_0 be any countable model of ψ and define M_α for $\alpha < \omega_1$ by induction: at limits take unions and let $M_{\alpha+1}$ be such that $M_{\alpha+1} \setminus M_\alpha$ consists of exactly one E-class. We first show that the sequence $(M_\alpha)_{\alpha<\omega_1}$ witnesses decomposition.

Let N be any model of ψ of size $\aleph_1$, let $N_0 \prec N$ be countable and define inductively a continuous $\prec$-chain in N of models N_α such that $N_{\alpha+1} \setminus N_\alpha$ consists of exactly one E-class and such that $N = \bigcup_{\alpha<\omega_1} N_\alpha$. Let $\beta < \omega_1$ and f be a *finite partial* isomorphism $f\colon (N_\alpha)_{\alpha<\beta} \to (M_\alpha)_{\alpha<\beta}$. We want to extend f to a (still filtration preserving) partial isomorphism with domain $\mathrm{dom}(f) \cup \{a\}$ for any given $a \in N_\beta$. If $P(a)$, this is possible by (3b); if $Q(a)$, we use (3a).

This "filtration preserving extension property" for finite partial isomorphisms shows not only decomposition, but also $\aleph_0$-categoricity (since the models are countable; thereby also finishing the proof of condition A) and homogeneity (for this, apply the argument with $N_\alpha = M_\alpha$).

It remains to show triviality. Let $\overline{a}, \overline{c}$ be in M_β for some $\beta < \omega_1$ with $\max(s_{\overline{c}}) < \min(s_{\overline{a}})$ and let $\overline{b}$ be such that $s_{\overline{b}} = s_{\overline{a}}$ and $\mathrm{tp}_{\mathrm{qf}}(\overline{b}_\alpha \overline{c}) = \mathrm{tp}_{\mathrm{qf}}(\overline{a}_\alpha \overline{c})$ for all α. Since M_0 must contain all of P, $\max(s_{\overline{c}}) < \min(s_{\overline{a}})$ implies that all components of $\overline{a}$ lie in Q and then $s_{\overline{b}} = s_{\overline{a}}$ implies that $\overline{bc}$ and $\overline{ac}$ satisfy the same quantifier-free type with respect to formulas only involving E (here we use the fact that the M_α have been chosen to add exactly *one* E-class each time). But also with respect to the relation R, $\overline{bc}$ and $\overline{ac}$ have the same quantifier-free type because of $\mathrm{tp}_{\mathrm{qf}}(\overline{b}_\alpha\overline{c}) = \mathrm{tp}_{\mathrm{qf}}(\overline{a}_\alpha\overline{c})$, so we can conclude $\overline{bc} \models \mathrm{tp}_{\mathrm{qf}}(\overline{ac})$ as required. □

Shelah provides a second example of an AEC in [**4**] which is a modification of the presented $L(Q)$-example, axiomatizable in $L_{\omega_1,\omega}$. The basic idea is to make P countable by making it the countable union of finite definable sets. However, as Chris Laskowski proves in an unpublished note, this AEC has the maximum number of models in $\aleph_1$ under ZFC. In our terminology, that AEC satisfies condition A as well as decomposition and homogeneity, but it fails triviality. It remains an important open question if categoricity (in $\aleph_1$) is absolute for $L_{\omega_1,\omega}$-sentences.

4 Martin's axiom and WCH are sufficient but not necessary

Our main theorem is the following:

Theorem 4.1 *Let K be an AEC with* $\mathrm{LS}(K) = \aleph_0$.

(a) *Suppose condition A holds. If $2^{\aleph_0} < 2^{\aleph_1}$, then K has $2^{\aleph_1}$ models of size $\aleph_1$. However it is consistent that $2^{\aleph_0} = 2^{\aleph_1}$ and the same conclusion holds.*

(b) *Suppose condition B holds. Assuming Martin's Axiom at $\aleph_1$, K is $\aleph_1$-categorical. However it is consistent that $MA_{\aleph_1}$ fails and the same conclusion holds.*

The first statements in (a) and (b) are the contents of Theorems 3.1 and 3.2. We now turn to proofs of the second statements.

A model of ZFC *where $2^{\aleph_0} = 2^{\aleph_1}$ yet K has $2^{\aleph_1}$ models of size $\aleph_1$.* There are models M of ZFC in which $2^{\aleph_0} = \aleph_2$ and $2^{\aleph_1} = \aleph_3$. (In fact, Easton [**3**] showed that any reasonable behaviour of the generalised continuum function $\kappa \mapsto 2^\kappa$ for regular κ is possible.) Now over this model M apply $\aleph_2$-Cohen forcing P. This is the forcing whose conditions are of the form $p\colon |p| \to 2$, $|p| < \omega_2$, ordered by extension. This forcing is $\aleph_2$-closed, i.e., any descending ω_1-sequence of conditions has a lower bound. As a consequence, if G is P-generic over M, any subset of ω_1 in $M[G]$ already belongs to M. It follows that M and $M[G]$ have the same structures with universe ω_1 and the same isomorphisms between such structures; by the first statement of Theorem 4.1(a), K has $\aleph_3^M$ many models of size $\aleph_1$ in M. As $\aleph_2$ is the same in M and $M[G]$, it follows that K has at least $\aleph_2^{M[G]}$ many models in $M[G]$ and $2^{\aleph_0}$ is $\aleph_2$ in $M[G]$.

But $2^{\aleph_1}$ equals $\aleph_2$ in $M[G]$: Each subset of ω_1 in $M[G]$ can be described in $M[G]$ by an ω_1-sequence of subsets of P that belongs to M (a "canonical name" for it), and there are $\aleph_3^M$ many such sequences. If $g\colon \omega_2 \to 2$ is the union of the conditions in G, then every subset of ω_1 in M occurs as $\{i < \omega_1 \mid g(\alpha + i) = 1\}$ for some $\alpha < \omega_2$, and therefore $\aleph_3^M = |\mathcal{P}^M(\omega_1)| \leq \aleph_2$ in $M[G]$ (where $\mathcal{P}^M$ denotes the powerset operation of M).

So $M[G]$ is a model of ZFC in which $2^{\aleph_0} = 2^{\aleph_1} = \aleph_2$ and K has the maximum number of models of size $\aleph_1$, as claimed.

We now turn to the second statement of Theorem 4.1(b).

A model of ZFC *in which $MA_{\aleph_1}$ fails, yet K is $\aleph_1$-categorical.* We use iterated forcing with countable support to construct the desired model of ZFC. We first review the argument that $\mathrm{MA}_{\aleph_1}$ yields $\aleph_1$-categoricity. Given two models $\mathcal{A}$, $\mathcal{B}$ in K of size $\aleph_1$, we write each as the union of an increasing, continuous ω_1-chain of countable models: $\mathcal{A} = \bigcup_{\alpha<\omega_1} \mathcal{A}_\alpha$, $\mathcal{B} = \bigcup_{\alpha<\omega_1} \mathcal{B}_\alpha$, as in decomposition of condition B. Then we consider the forcing $P(\vec{\mathcal{A}}, \vec{\mathcal{B}})$ whose conditions are finite partial isomorphisms p from $\mathcal{A}$ to $\mathcal{B}$ which preserve rank, i.e., such that for x in the domain of p, x belongs to A_α iff $p(x)$ belongs to B_α, for each $\alpha < \omega_1$. This forcing has the countable chain condition, and therefore by $\mathrm{MA}_{\aleph_1}$ there is a compatible set H of conditions in it which meets the $\aleph_1$-many dense sets which require that each element of A belongs to the domain and each element of B belongs to the range of some condition in H. Then the union of the conditions in H is an isomorphism of $\mathcal{A}$ onto $\mathcal{B}$.

The key observation is the following. We say that a forcing P is *almost bounding* iff whenever G is P-generic and $f\colon \omega \to \omega$ belongs to $V[G]$ there is $g\colon \omega \to \omega$ in V such that, for every infinite $X \subseteq \omega$ in V, $g(n) > f(n)$ for infinitely many n in X.

Lemma 4.2 *For any $\mathcal{A}$, $\mathcal{B}$ of size $\aleph_1$, the forcing $P(\vec{\mathcal{A}}, \vec{\mathcal{B}})$ is almost bounding.*

Proof. Suppose that G is $P(\vec{\mathcal{A}}, \vec{\mathcal{B}})$-generic and $f\colon \omega \to \omega$ is a function in $V[G]$. For any countable α let P_α denote the suborder of $P(\vec{\mathcal{A}}, \vec{\mathcal{B}})$ consisting of conditions with domain in A_α. Then $G_\alpha = G \cap P_\alpha$ is P_α-generic over V, as by triviality any condition p is compatible with any extension of $p \restriction A_\alpha$ in P_α and therefore any maximal antichain in P_α is also a maximal antichain in $P(\vec{\mathcal{A}}, \vec{\mathcal{B}})$. And as $P(\vec{\mathcal{A}}, \vec{\mathcal{B}})$ has the countable chain condition, f in fact belongs to $V[G_\alpha]$ for some countable α and therefore it suffices to prove that P_α is almost bounding for each countable α. But P_α is a countable forcing and is therefore equivalent to the forcing that adds one Cohen real. It is easy to check that the latter forcing is almost bounding (see [**1**]). □

We now use the following general lemma, which can be found in [**1**]. A forcing P is *weakly bounding* iff whenever G is P-generic and $f\colon \omega \to \omega$ belongs to $V[G]$ there is $g\colon \omega \to \omega$ in V such that $g(n) > f(n)$ for infinitely many n.

Lemma 4.3 *The countable support iteration of proper, almost bounding forcings is weakly bounding.*

Now, to finish our proof, perform a countable support iteration of length ω_2 over L, at each stage forcing with $P(\vec{\mathcal{A}}, \vec{\mathcal{B}})$ for some choice of $\vec{\mathcal{A}}$, $\vec{\mathcal{B}}$. Using a bookkeeping function we can ensure that if G is generic for this iteration, then every pair $\vec{\mathcal{A}}$, $\vec{\mathcal{B}}$ that exists in $V[G]$ will have been considered at some stage of the iteration. The result is a model in which K is $\aleph_1$-categorical. By Lemma 4.3, the iteration is weakly bounding, and therefore there is no $f\colon \omega \to \omega$ in $V[G]$ which *eventually dominates* each $g\colon \omega \to \omega$ in L, i.e., such that, for each $g\colon \omega \to \omega$ in L, $f(n) > g(n)$ for sufficiently large n. Therefore $\mathrm{MA}_{\aleph_1}$ fails in $V[G]$, by the following observation.

Lemma 4.4 *$MA_{\aleph_1}$ implies that some $f\colon \omega \to \omega$ eventually dominates every $g\colon \omega \to \omega$ in L.*

Proof. Consider *Hechler forcing* in L, whose conditions are pairs (s, g) where $s\colon |s| \to 2$ has domain a natural number and $g\colon \omega \to \omega$ belongs to L. Extension is defined by: $(s^*, g^*) \leq (s, g)$ iff s^* extends s, $g^*(n) > g(n)$ for all n and $s^*(n) > g(n)$ for all n in $|s^*| \setminus |s|$. This forcing is ccc because any two conditions with the same first component are compatible and there are only countably many first components. And for each $h\colon \omega \to \omega$ in L the set $D(s, g)$ of conditions (s, g) such that $g(n) > h(n)$ for all n is dense. It follows that if $f\colon \omega \to \omega$ is the generic function added by Hechler forcing, i.e., the union of the

s such that (s, g) belongs to the generic for some g, then f eventually dominates each $g\colon \omega \to \omega$ in L. The latter only requires that the $\aleph_1$ many dense sets $D(s, g)$ are met, so $\mathrm{MA}_{\aleph_1}$ implies that there is a such a function. □

In summary, with a countable support iteration of almost bounding forcings we produce a model where K is $\aleph_1$-categorical yet $\mathrm{MA}_{\aleph_1}$ fails.

Remark 4.5 We could do better and actually find a model of ZFC in which $\mathrm{MA}_{\aleph_1}$ fails, and in which *all* AEC's satisfying condition B are $\aleph_1$-categorical. The idea would be to apply the described forcings to all pairs of models of size $\aleph_1$ (in all countable signatures) with distinguished filtrations by countable models, for which the corresponding poset of finite partial filtration-preserving isomorphisms has the ccc and for which that forcing is almost bounding. In the procedure of iterating those forcings, we may create new instances of such pairs of models for which we can apply the forcing, but by bookkeeping, we will have taken care of them in an ω_2 long chain of iterated forcings. The resulting universe satisfies our requirement: if (A, B) is a pair of structures of size $\aleph_1$ (with filtrations) of an AEC satisfying condition B, we know by absoluteness of condition B that this instance occurred in our chain of forcings (use Lemma 4.2) and therefore A and B have been forced to be isomorphic. Thus any AEC satisfying condition B in the resulting universe is $\aleph_1$-categorical.

On the other hand, it is not clear whether our universe failing WCH in which a particular AEC with condition A has many models in $\aleph_1$ has the property that *all* such AEC's have many models in $\aleph_1$. The problem is that although we do not add subsets of $\aleph_1$, we *do* add subsets of the continuum (which is $\aleph_2$) and may create new AEC's satisfying condition A. Still, all AEC's with condition A whose restriction to countable models is H_{ω_2} definable will have many models in $\aleph_1$, which is the case for example for AEC's axiomatizable by an $L_{\omega_1,\omega}(Q)$ sentence with a natural notion of substructure.

Question Does there exist an AEC satisfying conditions A and B which is defined by an $L_{\omega_1,\omega}$-sentence?

Question Condition B is sufficient to show $\aleph_1$-categoricity under $\mathrm{MA}_{\aleph_1}$. To what extent is it also a necessary condition? For example, does every potentially (i.e., in some generic extension) $\aleph_1$-categorical AEC have to satisfy decomposition? It is not very difficult to show that, for a first-order theory, decomposition is *equivalent* to $\aleph_1$-categoricity ($\aleph_1$-categoricity is an *absolute* property for first-order theories because it is characterized by ω-stability plus "there are no Vaughtian pairs". Both properties follow directly from decomposition). Also, clearly, triviality is a very strong condition, as it is easy to find $\aleph_1$-categorical first-order theories where it fails (e.g., take an equivalence relation with exactly two classes and a binary relation defining a bijection between those two classes). Is there a way to weaken triviality and get the same results?

References

[1] Avraham, U., *Proper forcing*, in *Handbook of Set Theory*, vol. 1, Foreman, M., Kanamori, A. (eds.), Springer, 2010.

[2] Baldwin, J. T., *Categoricity*, AMS University Lecture Series, vol. 50, 2009.

[3] Easton, W., *Powers of regular cardinals*, Annals of Mathematical Logic 1, 1970.

[4] Shelah, S., *Abstract elementary classes near* $\aleph_1$ (sh88r), revision of Classification of nonelementary classes II, Abstract elementary classes; on the Shelah archive.

Two examples concerning $\aleph_1$-categoricity in abstract elementary classes

Martin Koerwien[*], **Stevo Todorcevic**[†]

[*] Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
koerwien@math.uic.edu

[†] CNRS FRE 3233, UFR de Mathématiques, Université Paris 7, France
stevo@logique.jussieu.fr

Abstract. We investigate two abstract elementary classes coding families of pairwise disjoint countable dense sets of the Cantor space and of the real line respectively. While the first one absolutely has many models in $\aleph_1$, there is evidence that the second one may be $\aleph_1$-categorical under PFA. We show that it consistently has many models under $\mathrm{MA}_{\aleph_1}$, which distinguishes it from an example exposed in [**6**] which is $\aleph_1$-categorical under $\mathrm{MA}_{\aleph_1}$ and has many models in $\aleph_1$ assuming the weak continuum hypothesis.

Introduction

Shelah gave in [**6**] an example of an $L(Q)$ axiomatizable abstract elementary class (where Q is the quantifier expressing that "there are uncountably many") which is $\aleph_1$-categorical under $\mathrm{MA}_{\aleph_1}$ and has the maximum number of models in $\aleph_1$ under the weak continuum hypothesis, thus showing that the notion of categoricity is not absolute for the logic $L(Q)$.

In the present paper, we give another example which is again $L(Q)$ axiomatizable, but may have many models under $\mathrm{MA}_{\aleph_1}$, while we believe that the Proper Forcing Axiom (PFA) implies that it is $\aleph_1$-categorical. We have to admit however that at present we are unable to give the formal proof of the latter. The example is coding families of pairwise disjoint countable dense sets of the real line. The argument that PFA implies categoricity should be a variation of the proof of Theorem 4.2 in [**8**]. We will give a pseudo-proof (containing a hole which we were unable to fix at the time of submission) which should still give a reasonably good idea of what techniques are involved. The proof of the consistency with $\mathrm{MA}_{\aleph_1}$ of having many models uses the model of Avraham and Shelah [**1**] originally constructed for showing that $\mathrm{MA}_{\aleph_1}$ does not imply Baumgartner's axiom [**3**].

Before we give this example, we introduce another similar one which codes families of countable dense sets of the "affine" Cantor space with the natural metric topology. The fact that this example absolutely has many models in $\aleph_1$ makes clear that the particular properties of the real line example are essentially due to the presence of an ordering, not only to its topology.

Preliminaries and notation

We make no notational distinction between a structure and its underlying set. If M and N are L-structures, $M \subset N$ means that M is a substructure of N. The notation

[*] Supported by the John Templeton Foundation under Grant #13152, *The Myriad Aspects of Infinity.*
[†] Research supported by CNRS and NSERC.

$M \prec N$ may have a different meaning than M being a first order elementary substructure of N if we redefine the symbol $\prec$. Recall that, for a signature L, a class of L-models K together with a distinguished partial ordering $\prec$ on K is called an *abstract elementary class* (AEC) if:

(1) $A \prec B$ implies $A \subset B$;
(2) Both K and $\prec$ are closed under isomorphism;
(3) $A \prec C$, $B \prec C$ and $A \subset B$ imply $A \prec B$;
(4) For any continuous $\prec$-chain $(A_\alpha)_{\alpha<\lambda}$,
  (a) $A = \bigcup_{\alpha<\lambda} A_\alpha \in K$;
  (b) for all $\alpha < \lambda$, $A_\alpha \prec A$;
  (c) if $A_\alpha \prec B$ for some B and all $\alpha < \lambda$, then $A \prec B$;
(5) There is a cardinal $\mathrm{LS}(K)$ such that for all $A \in K$ and any *subset* $X \subset A$, there is $B \prec A$ containing X with $|B| \leq |X| + \mathrm{LS}(K)$.

1 The Cantor space

Let $L = \{E, E_i\}_{i<\omega}$ where E and E_i are binary relations. Let σ be the $L_{\omega_1,\omega}(Q)$-sentence stating:

- E and E_i are all equivalence relations, E has infinitely many classes, all countable, E_i has 2^{i+1} classes, E_{i+1} refines E_i binary (every E_i-class is the union of exactly two E_{i+1}-classes).
- $\bigwedge_{i<\omega} E_i(x,y)$ implies $x = y$.
- Any E-class intersects all E_i-classes non-trivially for all $i < \omega$.

Using the notation $[x]_E^N$ for the E-class of an element x in a model M, we define a strong substructure relation which turns the models of σ into an AEC:

$$M \prec N \text{ if and only if } M \subset N \text{ and, for all } x \in M,\ [x]_E^N = [x]_E^M.$$

An easy back-and-forth argument shows that σ is $\aleph_0$-categorical.

Proposition 1.1 *The sentence σ fails amalgamation for countable models.*

Proof. Let $M_0 \models \sigma$ be any countable model and let M_1, M_2 be extensions of M_0 such that

- $M_i \setminus M_0$ consists of exactly one new E-class ($i = 1, 2$);
- there are elements a_i, b ($i = 1, 2$) such that
  - $a_i \in M_i \setminus M_0$, $b \in M_1 \setminus M_0$;
  - for all $x \in M_0$ and $j < \omega$, $E_j(x, a_1) \leftrightarrow E_j(x, a_2)$;
  - for all $y \in M_1$ and $x \in M_0$, there is some $j < \omega$ such that $\neg(E_j(x,b) \leftrightarrow E_j(x,y))$.

If there were an amalgam of M_1 and M_2 over M_0, then a_1 and a_2 would have to be identified in it; but b cannot be identified with any element in $M_2 \setminus M_0$. □

Remark 1.2 In the terminology of [**4**], this AEC satisfies condition A (which implies many models in $\aleph_1$ under ZFC + WCH) and "almost" satisfies condition B (which would imply $\aleph_1$-categoricity under ZFC + $\mathrm{MA}_{\aleph_1}$); $(\mathrm{mod}(\sigma), \prec)$ satisfies (decomposition), witnessed by a continuous chain $(M_\alpha)_{\alpha<\omega_1}$ where M_0 is any countable model and $(M_{\alpha+1} \setminus M_\alpha)$ consists of exactly one E-class. The same back-and-forth argument as above also shows (homogeneity). However we do not have (triviality) because any two elements have the same quantifier-free type over the empty set, but a pair of such can have countably many different types because of the E_i-equivalence relations. In the following, we show that already ZFC implies the existence of many models in $\aleph_1$.

Definition 1.3 Let $M \models \sigma$.

- $A \subset M$ is a *transversal* if it contains exactly one element from any E-class.
- For $S \subset \omega$, a transversal $A \subset M$ has S *type* if $\{i_{x,y} \mid x, y \in A, x \neq y\} \subset S$, where $i_{x,y}$ is defined as the unique i such that $\neg E_i(x,y)$ and $E_j(x,y)$ for all $j < i$.

Proposition 1.4 *Let $S \subset \omega$ be infinite. There is some $M \models \sigma$ of size $\aleph_1$ that can be written as the countable union of transversals of S type.*

Proof. Let $M_0 \models \sigma$ be countable with $M_0 = \bigcup_{k<\omega} T_k$, where each T_k is an S type transversal of M_0 and the T_k are pairwise disjoint. We extend M_0 by a new E-class $A = \{t_k \mid k < \omega\}$ such that each $T_k \cup \{t_k\}$ is of S type.

First, choose any E-class $B = \{b_k \mid k < \omega\}$ of M_0 (where $b_k \in T_k$ for each $k < \omega$) and enumerate all E_n-classes of M_0 as $(B_k)_{k<\omega}$ in such a way that $b_k \in B_k$. Now, choose as t_k any element that is E_k-equivalent to B_k (this guarantees that A will eventually be dense) making inductively sure that $i_{t_k,x} \in S$ for all $x \in T_k$. □

Lemma 1.5 *Suppose that $S, U \subset \omega$ are infinite and almost disjoint. Then $M_S \ncong M_U$ (denoting by M_S the model constructed in the previous proposition).*

Proof. Let $M_S = \bigcup_{k<\omega} T_k^S$ and $M_U = \bigcup_{k<\omega} T_k^U$, where T_k^S and T_k^U are transversals of S and U type respectively, and suppose that $f\colon M_S \to M_U$ is an isomorphism. Then $f[T_0^S]$ is a transversal in M_U and must have an infinite (actually $\aleph_1$ big) intersection X with some T_k^U. Then X is of $S \cap U$ type, which is impossible since $S \cap U$ is finite. □

Corollary 1.6 (ZFC) *There are $2^{\aleph_1}$ pairwise non-isomorphic models of σ of size $\aleph_1$.*

Proof. If $2^{\aleph_1} = 2^{\aleph_0}$, then this follows from the preceding Lemma and the fact that there are almost disjoint families of subsets of ω of size continuum.

If $2^{\aleph_1} > 2^{\aleph_0}$ (which means that the weak continuum hypothesis is true), then this follows from $\aleph_0$-categoricity and the failure of amalgamation for countable models, as Shelah shows in [**6**]. □

2 The real line

Let $L = \{P, Q, R, <\}$ with P and Q unary predicates and R, $<$ binary relations. Let σ be the $L(Q)$-sentence stating:

- P and Q partition the universe;
- $<$ is a dense linear ordering without endpoints on P;
- $R \subset P \times Q$ and writing $A_q = \{p \in P \mid R(p,q)\}$ for all $q \in Q$ we have that all A_q are pairwise disjoint countable dense sets in $(P, <)$, the union of which equals P.

If we denote by A_q^M the set A_q defined in a model M, then the strong substructure relation which turns the models of σ into an AEC is defined as follows:

$$M \prec N \text{ if and only if } M \subset N \text{ and, for all } q \in Q^M,\ A_q^N = A_q^M.$$

Again, a straightforward back-and-forth argument shows $\aleph_0$-categoricity.

Proposition 2.1 *The sentence σ fails amalgamation for countable models.*

Proof. Let M_0 be any countable model and let C_i $(i < 3)$ be three different non-rational cuts in M_0. Let M_1, M_2 be extensions of M_0 adding exactly one new dense set each, such that M_1 realizes C_0 and C_1 but not C_2 and M_2 realizes C_0 and C_2 but not C_1. In an amalgam, the new dense sets must coincide because they intersect on the realization of C_0, which is impossible. □

So, under WCH, σ will have the maximum number of models in $\aleph_1$. Now we will see that this is even consistent with $\mathrm{ZFC} + \mathrm{MA}_{\aleph_1}$.

Lemma 2.2 *Let $M \models \sigma$. Then $(P^M, <)$ can be embedded into $(\mathbb{R}, <)$.*

Proof. This is an easy consequence of $(P^M, <)$ being a separable ordering and $(\mathbb{R}, <)$ being complete: Enumerate $Q^M = \{q_\alpha \mid \alpha < \kappa\}$ and let us write P_α for $\bigcup_{\beta<\alpha} A_{q_\alpha}$. Now we show by induction that $(P_\alpha, <)$ embeds into $(\mathbb{R}, <)$.

Let $f \colon P_\alpha \to \mathbb{R}$ be an embedding and enumerate $P_{\alpha+1} \setminus P_\alpha$ as $\{a_i \mid i < \omega\}$. Suppose that we already extended f to $P' = P_\alpha \cup \{a_0, \ldots, a_{k-1}\}$. Now let $C = \{x \in P' \mid x < a_k\}$ and $D = \{x \in P' \mid x > a_k\}$ (both are non-empty!). Then $f[C]$ has no maximal element and $f[D]$ has no minimal element. For example, suppose that $y \in f[D]$ were minimal. Then there could not be any element of P' between a_k and $f^{-1}(y)$, which would imply that $a_k = f^{-1}(y) \in D$, contrary to the definition of D.

We thus can set $f(a_k)$ to be any element of $\mathbb{R}$ strictly above $f[C]$ and below $f[D]$. □

Theorem 2.3 *It is consistent with $\mathrm{MA}_{\aleph_1}$ that σ has $2^{\aleph_1}$ many models in $\aleph_1$.*

Proof. An uncountable set A of real numbers is called 2-*entangled* if no injective function from an uncountable subset of A into A with disjoint domain and range is order-preserving or order-reversing (see Definition 20 in [**1**]).

As shown in Section 5 of [**1**], the existence of a 2-entangled set A of reals is consistent with $\mathrm{MA}_{\aleph_1}$. Now take any family $(A_\alpha)_{\alpha<2^{\aleph_1}}$ of subsets of A of size $\aleph_1$, such that the symmetric differences between any two of them is uncountable. Use any one of these sets A_α as the P-part of a model $M_\alpha \models \sigma$ of size $\aleph_1$ (defining the family of countable dense subsets arbitrarily).

For any $\alpha < \beta < 2^{\aleph_1}$, we must have $M_\alpha \ncong M_\beta$ because otherwise we would get an order-preserving bijection $f \colon A_\alpha \to A_\beta$ and thus (because of the uncountability of $A_\alpha \triangle A_\beta$) an uncountable $g \subset f$ whose domain and range are disjoint. □

The following lemma (which we shall use in a possible argument that PFA implies $\aleph_1$-categoricity of σ) is well known (see, for example, [**5**, § 31]) and it is also a special case of a result found in [**7**]. However, for the reader's convenience, we provide a simple proof of it.[1]

Lemma 2.4 *Let $A \subset \mathbb{R}^n$ be* any *set and let $f \colon A \to \mathbb{R}$ be a continous function. Then f extends to a continous function on a G_δ-subset of $\mathbb{R}^n$.*

Proof. We first introduce some notation. Given $y \in \mathbb{R}^n$, $\epsilon > 0$, and $X \subset \mathbb{R}^n$, we write $B_\epsilon(y) = \{x : ||x - y|| < \epsilon)\}$ and $\mathrm{diam}(X) = \sup\{||x - y|| : x, y \in X\}$.

For any $a \in A$ and $n < \omega$, let

- $U_{a,n}$ be any open neighborhood (in $\mathbb{R}$) of a such that $\mathrm{diam}(f[U_{a,n} \cap A]) < \frac{1}{n}$;
- $U_n = \bigcup_{a\in A} U_{a,n}$;
- $U = \bigcap_{n<\omega} U_n$.

[1] The first author would like to thank Tapani Hyttinen for explaining the idea of this proof to him.

Clearly, $U_{a,n}$, U_n are open sets, so U is G_δ and contains A. Now we set $U' = U \cap \overline{A}$, which is a G_δ-set and still contains A, and show that f has a (unique) continuous extension to U'.

Let $x \in U'$ and let $(a_i)_{i<\omega}$ be a sequence in A converging to x. Clearly, $(f(a_i))_{i<\omega}$ is a Cauchy sequence: since $x \in U$, for any $n < \omega$ there is some $a \in A$ such that $x \in U_{a,n}$ and the variations of f in $U_{a,n}$ are smaller than $\frac{1}{n}$. Now take a final segment of $(a_i)_{i<\omega}$ contained in $U_{a,n}$.

Define f at x to be the limit of $(f(a_i))_{i<\omega}$. Then f thus extended will be well-defined for the same reason as before: if $(b_i)_{i<\omega}$ is another sequence approaching x, the differences between the $f(b_i)$ and the $f(a_i)$ will become arbitrary small, so the limit will be the same. □

Finally we give an idea of what a proof of $\aleph_1$-categoricity may look like. As pointed out earlier, the main arguments already appear in the proof of Theorem 4.2 in **[8]**. We give a “proof” that has a mistake in it, for which we see no “trivial“ fix.

Conjecture 2.5 PFA implies that σ is $\aleph_1$-categorical.

False proof. We use the fact that the composition of forcing CH and a forcing which provides an isomorphism for two given models of size $\aleph_1$ (which we show is ccc) is proper. So we can start by assuming that CH holds. Let $(A_\alpha)_{\alpha<\omega_1}$ and $(B_\alpha)_{\alpha<\omega_1}$ be families of countable dense sets (of reals) coded by two models of σ and enumerate all continous functions from G_δ-sets of finite Cartesian products of the reals into the reals as $(g_\alpha)_{\alpha<\omega_1}$. We define a continuous increasing sequence of countable ordinals $(\delta_\alpha)_{\alpha<\omega_1}$ inductively as follows. Let $\delta_0 = \omega$ and for $\alpha < \omega_1$ let δ_α be any number greater than $\sup_{\beta<\alpha} \delta_\beta$ such that, writing $A_{<\gamma} = \bigcup_{\beta<\gamma} A_\beta$, $B_{<\gamma} = \bigcup_{\beta<\gamma} B_\beta$ and $C_{<\gamma} = A_{<\gamma} \cup B_{<\gamma}$, we have that $C_{<\delta_\alpha} \cup \bigcup_{\beta<\alpha} g_\beta[C_{<\delta_\alpha}]$ is disjoint from $C_{<\omega_1} \setminus C_{<\delta_\alpha}$. Now let $\rho\colon \omega_1 \to \omega_1$ be any bijection such that, for all $\alpha < \omega_1$, the interval $[\delta_\alpha, \delta_{\alpha+1})$ is mapped onto $[\delta_{\alpha+1}, \delta_{\alpha+2})$.

We consider the familiy $\mathcal{F}$ of order-preserving finite partial functions from $\mathbb{R}$ to $\mathbb{R}$ such that, for all $f \in \mathcal{F}$, $\beta < \omega_1$ and $x \in \operatorname{dom}(f)$, $x \in A_\beta$ if and only if $f(x) \in B_{\rho(\beta)}$, and finish by showing that it has ccc, i.e., any uncountable $X \subset \mathcal{F}$ contains two elements whose union is still a member of $\mathcal{F}$ (a generic of $\mathcal{F}$ ordered by reverse inclusion provides an isomorphism of the two models we started with).

Thus let $X \subset \mathcal{F}$ be uncountable. We may assume that $|\operatorname{dom}(f)| = n$ for all $f \in X$, and our proof of ccc will be by induction over that n (so we suppose ccc holds for such families with $n-1$). We may suppose as well that, for any $f, g \in \mathcal{F}$, if

$$f = \{(a_1, b_1), \dots, (a_n, b_n)\}, \quad g = \{(c_1, d_1), \dots, (c_n, d_n)\}$$

with $a_1 < a_2 < \cdots < a_n$, $c_1 < c_2 < \cdots < c_n$ and $(a_i, b_i) = (c_i, d_i)$ for all $i = 1 \dots (n-1)$, $a_n = c_n$ implies $b_n = d_n$. Finally we assume that there are $2n$ fixed rational intervals I_k, J_k $(k = 1 \dots n)$ such that

- the I_k are pairwise disjoint and in increasing order, and the same is true for J_k;
- any $f \in \mathcal{F}$ maps for each k exactly one element from I_k to one element from J_k.

Now we define a function g on a subset D of $\mathbb{R}^{2n-1}$ into $\mathbb{R}$ as follows: if $f \in \mathcal{F}$ is of the form $f = \{(a_1, b_1), \dots, (a_n, b_n)\}$ with $a_1 < a_2 < \cdots < a_n$, then, and only then, we add $(a_1, b_1, a_2, b_2, \dots, a_{n-1}, b_{n-1}, a_n)$ to D and set $g(a_1, b_1, a_2, b_2, \dots, a_{n-1}, b_{n-1}, a_n) = b_n$.

We show that g must be discontinuous in uncountably many points of D. Suppose not and let D' be the result of removing the countably many discontinuities from D.

Lemma 2.4 yields a G_δ-set containing D' to which g extends continuously. Hence the original g coincides with some g_α on all but possibly countably many points of D.

Now let $\overline{a} = (a_1, b_1, a_2, b_2, \ldots, a_{n-1}, b_{n-1}, a_n) \in D'$ such that $a_n \in A_{<\delta_{\gamma+1}} \setminus A_{<\delta_\gamma}$ for some $\gamma > \alpha$. By definition of ρ, we must have $g(\overline{a}) \in [\delta_{\gamma+1}, \delta_{\gamma+2})$, which contradicts the definition of the ordinal $\delta_{\gamma+1}$ (which is supposed to be "closed under applications of g_α", since $\alpha < \gamma$).

We shrink $\mathcal{F}$ further to an uncountable $\mathcal{F}'$ such that g^*, defined analogously to g, has the following properties:

(i) $\overline{a} \in \mathrm{dom}(g^*)$ implies that $\overline{a}$ is a point of discontinuity of g (not necessarily of g^*).

(ii) There is a rational d such that for all $\overline{a} \in \mathrm{dom}(g^*)$ and all neighborhoods N of $\overline{a}$, we can find some $\overline{a}' \in N$ such that $g^*(\overline{a}) < d < g(\overline{a}')$.

By induction, there are some $f_1, f_2 \in \mathcal{F}'$ such that the union of the first $n-1$ elements of f_1 and of f_2 form an increasing function. Without loss of generality we assume that $x < y$, where $(x, f_1(x))$ and $(y, f_2(y))$ are the n-th elements of those functions. If $f_1(x) < f_2(y)$, then $f \cup g$ is already an increasing function and we are finished (using the intervals I_i and J_i introduced above to be sure that the n-th elements of f_1 and f_2 do not interfere with the $n-1$ first elements of those functions). But now, if $f_1(x) > f_2(y)$, we use (ii) and make a slight perturbation of the domain of f_2 to get some f_3 in the original $\mathcal{F}$ which still has the property that the union of its $n-1$ first elements and those of f_1 form an increasing function, and moreover $f_3(y') > d$, where y' is the result of perturbing y. This implies that $f_1 \cup f_3$ is increasing because $f_1(x) < d$, again by (ii). □

As the reader may have noticed, there is a problematic step in the above "proof". In the proof of the claim that g must be discontinuous at uncountably many points, we pick some $\overline{a} = (a_1, b_1, a_2, b_2, \ldots, a_{n-1}, b_{n-1}, a_n) \in D'$ such that $a_n \in A_{<\delta_{\gamma+1}} \setminus A_{<\delta_\gamma}$ for some $\gamma > \alpha$ and then claim that, by definition of ρ, we must have $g(\overline{a}) \in [\delta_{\gamma+1}, \delta_{\gamma+2})$ contradicting the definition of the ordinal $\delta_{\gamma+1}$, which is supposed to be "closed under applications of $g = g_\alpha$" since $\alpha < \gamma$.

We cannot make that conclusion, since we could have $a_{n-1} \in A_{<\delta_{\gamma+1}} \setminus A_{<\delta_\gamma}$, for example, which would imply that $b_{n-1} \in [\delta_{\gamma+1}, \delta_{\gamma+2})$. We would need that *all* of the arguments to which we apply g, including all the b_i, belong to $A_{<\delta_{\gamma+1}}$.

The second author suggests that the argument can be fixed by applying the above diagonalization argument to multivalued maps g_α instead of single-valued ones. The details however are still left to be worked out.

References

[1] Avraham, U., Shelah, S., *Martin's axiom does not imply that every two $\aleph_1$-dense sets of reals are isomorphic*, Israel J. Math. 38, 1981, pp. 161–176.

[2] Baldwin, J. T., *Categoricity*, AMS University Lecture Series vol. 50, 2009.

[3] Baumgartner, J. E., *All $\aleph_1$-dense sets of reals can be isomorphic*, Fund. Math. 79, 1973, no. 2, pp. 101–106.

[4] Friedman S. D., Koerwien, M., *On absoluteness of categoricity in AEC's*, to appear in the Notre Dame Journal of Formal Logic.

[5] Kuratowski, K., *Topology I*, Academic Press, 1966.

[6] Shelah, S., *Abstract elementary classes near $\aleph_1$* (sh88r), revision of Classification of nonelementary classes II, Abstract elementary classes; on the Shelah archive.

[7] Todorcevic, S., *Remarks on chain conditions in products*, Compositio Math. 55, 1985, pp. 295–302.

[8] Todorcevic, S., *Partition Problems in Topology*, Contemporary Math. vol. 84, AMS, 1989.

Borel reductions on the generalized Cantor space

Vadim Kulikov[†]

[†] Department of Mathematics and Statistics, University of Helsinki, Finland
vadim.kulikov@helsinki.fi

Abstract. It is shown that the power set of κ ordered by the subset relation modulo various versions of the non-stationary ideal can be embedded into the partial order of Borel equivalence relations on 2^κ under Borel reducibility. Here κ is uncountable regular cardinal with $\kappa^{<\kappa} = \kappa$.

Introduction

It is shown that the partial order of Borel equivalence relations on the generalized Baire spaces (2^κ for $\kappa > \omega$) under Borel reducibility has high complexity already at low levels (below E_0). This extends an answer stated in [**4**] to an open problem stated in [**5**] and in particular solves open problems 7 and 9 from [**4**].

The developement of the theory of generalized Baire and Cantor spaces dates back to 1990's, when A. Mekler and J. Väänänen published the paper *Trees and* Π^1_1*-subsets of* ${}^{\omega_1}\omega_1$ [**13**] and A. Halko published *Negligible subsets of the generalized Baire space* $\omega_1^{\omega_1}$. More recently, equivalence relations and Borel reducibility on these spaces and their applications to model theory have been under focus; see my latest joint work with S. Friedman and T. Hyttinen [**5**].

Suppose κ is an infinite cardinal and let E^B_κ be the collection of all Borel equivalence relations on 2^κ. (For definitions in the case $\kappa > \omega$, see next section.) For equivalence relations E_0 and E_1, let us denote $E_0 \le_B E_1$ if there exists a Borel function $f\colon 2^\kappa \to 2^\kappa$ such that $(\eta, \xi) \in E_0 \Leftrightarrow (f(\eta), f(\xi)) \in E_1$. The relation $\le_B$ defines a quasiorder on E^B_κ, i.e., it induces a partial order on $\mathrm{E}^B_\kappa/\sim_B$ where $\sim_B$ is the equivalence relation of bireducibility: $E_0 \sim_B E_1 \Leftrightarrow (E_0 \le_B E_1) \wedge (E_1 \le_B E_0)$.

In the case $\kappa = \omega$ there are many known results that describe the order $\langle \mathrm{E}^B_\kappa, \le_B \rangle$. Two of them are:

Theorem (Louveau–Velickovic [**12**]) *The partial order* $\langle \mathcal{P}(\omega), \subset_* \rangle$ *can be embedded into the partial order* $\langle \mathrm{E}^B_\omega, \le_B \rangle$*, where* $A \subset_* B$ *if* $A \setminus B$ *is finite.*

Theorem (Adams–Kechris [**1**]) *The partial order* $\langle \mathfrak{B}, \subset \rangle$ *can be embedded into the partial order* $\langle \mathrm{E}^B_\omega, \le_B \rangle$*, where* $\mathfrak{B}$ *is the collection of all Borel subsets of the real line* $\mathbb{R}$*. In fact, the embedding is into the suborder of* $\langle \mathrm{E}^B_\omega, \le_B \rangle$ *consisting of the countable Borel equivalence relations, i.e., those Borel equivalence relations each of whose equivalence classes is countable.*

Our aim is to generalize these results to uncountable κ with $\kappa^{<\kappa} = \kappa$, and it is proved that $\langle \mathcal{P}(\kappa), \subset_{\mathrm{NS}(\omega)} \rangle$ can be embedded into $\langle \mathrm{E}^B_\kappa, \le_B \rangle$, where $A \subset_{\mathrm{NS}(\omega)} B$ means that $A \setminus B$ is not ω-stationary. This is proved in ZFC. However under mild additional assumptions on κ or on the underlying set theory, it is shown that $\langle \mathcal{P}(\kappa), \subset_{\mathrm{NS}} \rangle$ can be embedded into $\langle \mathrm{E}^B_\kappa, \le_B \rangle$, where $A \subset_{\mathrm{NS}} B$ means that $A \setminus B$ is non-stationary and that $\langle \mathcal{P}(\kappa), \subset_* \rangle$ can be embedded into $\langle \mathrm{E}^B_\kappa, \le_B \rangle$, where $A \subset_* B$ means that $A \setminus B$ is bounded.

Assumption Everywhere in this article it is assumed that κ is a cardinal which satisfies $|\kappa^\alpha| = \kappa$ for all $\alpha < \kappa$. This requirement is briefly denoted by $\kappa^{<\kappa} = \kappa$.

1 Background in generalized descriptive set theory

Definition 1.1 Consider the function space 2^κ (all functions from κ to $\{0,1\}$) equipped with the topology generated by the sets

$$N_p = \{\eta \in 2^\kappa \mid \eta \restriction \alpha = p\}$$

for $\alpha < \kappa$ and $p \in 2^\alpha$. Borel sets on this space are obtained by closing the topology under unions and intersections of length $\leq \kappa$, and complements.

An equivalence relation E on 2^κ is *Borel reducible* to an equivalence relation E' on 2^κ if there exists a Borel function $f\colon 2^\kappa \to 2^\kappa$ (inverse images of open sets are Borel) such that $\eta E \xi \Leftrightarrow f(\eta) E' f(\xi)$. This is denoted by $E \leq_B E'$.

The descriptive set theory of these spaces, of equivalence relations on them and of their reducibility properties for $\kappa > \omega$, has been developed at least in [**5, 7, 13**]. For $\kappa = \omega$ this is the field of standard descriptive set theory.

By id_X we denote the identity relation on X: $(\eta, \xi) \in \mathrm{id}_X \Leftrightarrow (\eta, \xi) \in X^2 \wedge \eta = \xi$, and by E_0 the equivalence relation on 2^κ (or on κ^κ as in the proof of Theorem 3.20) such that $(\eta, \xi) \in E_0 \Leftrightarrow \{\alpha \mid \eta(\alpha) \neq \xi(\alpha)\}$ is bounded.

Notation Let E^B_κ denote the set of all Borel equivalence relations on 2^κ (i.e., equivalence relations $E \subset (2^\kappa)^2$ such that E is a Borel set). If $X, Y \subset \kappa$ and $X \setminus Y$ is non-stationary, let us denote it by $X \subset_{\mathrm{NS}} Y$. If $X \setminus Y$ is not λ-stationary for some regular $\lambda < \kappa$, it is denoted by $X \subset_{\mathrm{NS}(\lambda)} Y$.

The set of all ordinals below κ which have cofinality λ is denoted by S^κ_λ, and $\lim(\kappa)$ denotes the set of all limit ordinals below κ. Also $\operatorname{reg}\kappa$ denotes the set of regular cardinals below κ and

$$S^\kappa_{\geq\lambda} = \Cup_{\substack{\mu \geq \lambda \\ \mu \in \operatorname{reg}\kappa}} S^\kappa_\mu,$$

$$S^\kappa_{\leq\lambda} = \Cup_{\substack{\mu \leq \lambda \\ \mu \in \operatorname{reg}\kappa}} S^\kappa_\mu.$$

If $A \subset \alpha$ and α is an ordinal, then $\mathrm{OTP}(A)$ is the order type of A in the ordering induced on it by α.

For ordinals $\alpha < \beta$, let us adopt the following abbreviations:

$$(\alpha, \beta) = \{\gamma \mid \alpha < \gamma < \beta\};$$
$$[\alpha, \beta] = \{\gamma \mid \alpha \leq \gamma \leq \beta\};$$
$$(\alpha, \beta] = \{\gamma \mid \alpha < \gamma \leq \beta\};$$
$$[\alpha, \beta) = \{\gamma \mid \alpha \leq \gamma < \beta\}.$$

If η and ξ are functions in 2^κ, then $\eta \operatorname{sd} \xi$ is the function $\zeta \in 2^\kappa$ such that $\zeta(\alpha) = 1 \Leftrightarrow \eta(\alpha) \neq \xi(\alpha)$ for all $\alpha < \kappa$, and $\overline{\eta} = 1 - \eta$ is the function $\zeta \in 2^\kappa$ such that $\zeta(\alpha) = 1 - \eta(\alpha)$ for all $\alpha < \kappa$. If A and B are sets, then $A \operatorname{sd} B$ is just the symmetric difference.

For any set X, 2^X denotes the set of all functions from X to $2 = \{0, 1\}$. If $p \in 2^{[0,\alpha)}$ and $\eta \in 2^{[\alpha,\kappa)}$, then $p^\frown\eta \in 2^\kappa$ is the catenation: $(p^\frown\eta)(\beta) = p(\beta)$ for $\beta < \alpha$ and $(p^\frown\eta)(\beta) = \eta(\beta)$ for $\beta \geq \alpha$.

Definition 1.2 A *co-meager* subset of X is a set which contains an intersection of length $\leq \kappa$ of dense open subsets of X. Co-meager sets are always non-empty and form a filter on 2^κ; cf. [**13**]. A set X has the *property of Baire* if there exists an open set A such that $X \operatorname{sd} A$ is meager, i.e., a complement of a co-meager set. As in standard descriptive set theory, Borel sets have the property of Baire (proved in [**7**]). For a Borel function $f\colon 2^\kappa \to 2^\kappa$, denote by $C(f)$ one of the co-meager sets restricted to which f is continuous (such a set is not unique, but we can always pick one using the property of Baire of Borel sets; see [**5**]).

Lemma 1.3 *Let D be a co-meager set in 2^κ and let $p, q \in 2^\alpha$ for some $\alpha < \kappa$. Then there exists $\eta \in 2^{[\alpha,\kappa)}$ such that $p^\frown\eta \in D$ and $q^\frown\eta \in D$. Also there exists $\eta \in 2^{[\alpha,\kappa)}$ such that $p^\frown\overline{\eta} \in D$ and $q^\frown\eta \in D$ where $\overline{\eta} = 1 - \eta$.*

Proof. Let h be the homeomorphism $N_p \to N_q$ defined by $p^\frown\eta \mapsto q^\frown\eta$. Then $h[N_p \cap D]$ is co-meager in N_q, so $N_q \cap D \cap h[N_p \cap D]$ is non-empty. Pick η' from that intersection and let $\eta = \eta' \mid [\alpha, \kappa)$. This will do. For the second part, take for h the homeomorphism defined by $p^\frown\eta \mapsto q^\frown\overline{\eta}$. □

2 On cub-games and GC_λ-characterization

The notion of cub-games is a useful way to treat certain properties of subsets of cardinals. They generalize closed unbounded sets and are related to combinatorial principles such as $\square_\kappa$. Under mild set-theoretic assumptions, they give characterizations of CUB-filters in different cofinalities. Treatments of this subject can be found for example in [**8, 9, 10**].

Definition 2.1 Let $A \subset \kappa$. The game $\mathrm{GC}_\lambda(A)$ is played between players **I** and **II** as follows. There are λ moves and at the i-th move player **I** picks an ordinal α_i which is greater than any ordinal picked earlier in the game and then **II** picks an ordinal $\beta_i > \alpha_i$. Player **II** wins if $\sup_{i<\lambda} \alpha_i \in A$. Otherwise player **I** wins.

Definition 2.2 A set $C \subset \kappa$ is λ-*closed* for a regular cardinal $\lambda < \kappa$ if, for all increasing sequences $\langle \alpha_i \in C \mid i < \lambda \rangle$, the limit $\sup_{i<\lambda} \alpha_i$ is in C. A set $C \subset \kappa$ is *closed* if it is λ-closed for all regular $\lambda < \kappa$. A set is λ-*cub* if it is λ-closed and unbounded, and *cub* if it is closed and unbounded. A set is λ-stationary if it intersects all λ-cub sets, and *stationary* if it intersects all cub sets.

Definition 2.3 We say that GC_λ-*characterization* holds for κ if

$$\{A \subset \kappa \mid \textbf{II} \text{ has a winning strategy in } \mathrm{GC}_\lambda(A)\} = \{A \subset \kappa \mid A \text{ contains a } \lambda\text{-cub set}\},$$

and we say that GC-*characterization* holds for κ if GC_λ-characterization holds for κ for all regular $\lambda < \kappa$.

Definition 2.4 Assume $\kappa = \lambda^+$ and $\mu \leq \lambda$ a regular uncountable cardinal. The *square principle on κ for μ*, denoted $\square^\kappa_\mu$, defined by Jensen in case $\lambda = \mu$, is the statement that there exists a sequence $\langle C_\alpha \mid \alpha \in S^\kappa_{\leq\mu} \rangle$ with the following properties:

(1) $C_\alpha \subset \alpha$ is closed and unbounded in α;
(2) if $\beta \in \lim C_\alpha$, then $C_\beta = \beta \cap C_\alpha$;
(3) if $\operatorname{cf}(\alpha) < \mu$, then $|C_\alpha| < \mu$.

Remark 2.5 For $\omega < \mu < \lambda$ in the definition above, it was proved by Shelah in [**14**] that $\square^\kappa_\mu$ holds (this can be proved in ZFC —for a proof, see [**2**, Lemma 7.7]). If $\mu = \lambda$, then $\square^\kappa_\mu = \square^{\mu^+}_\mu$ is denoted by $\square_\mu$ and can be easily forced or, on the other hand, it holds if $V = L$. The failure of $\square_\mu$ implies that μ^+ is Mahlo in L, as pointed out by Jensen; see [**11**].

Definition 2.6 For $\kappa > \omega$, the set $I[\kappa]$ consists of those $S \subset \kappa$ that have the following property: there exists a cub set C and a sequence $\langle D_\alpha \mid \alpha < \kappa\rangle$ such that

(1) $D_\alpha \subset \mathcal{P}(\alpha)$, $|D_\alpha| < \kappa$;
(2) $D_\alpha \subset D_\beta$ for all $\alpha < \beta$;
(3) for all $\alpha \in C \cap S$ there exists $E \subset \alpha$ unbounded in α and of order type $\mathrm{cf}(\alpha)$ such that, for all $\beta < \alpha$, $E \cap \beta \in D_\gamma$ for some $\gamma < \alpha$.

Remark 2.7 The following is known:

(1) $I[\kappa]$ is a normal ideal and contains the non-stationary sets.
(2) If $\lambda < \kappa$ is regular and $S^\kappa_\lambda \in I[\kappa]$, then GC_λ-characterization holds for κ.
(3) If μ is regular and $\kappa = \mu^+$, then $S^\kappa_{<\mu} \in I[\kappa]$; cf. [**14**]. This follows also from (4) and Remark 2.5.
(4) When $\lambda > \omega$, then $\square^\kappa_\lambda$ implies that $S^\kappa_\lambda \in I[\kappa]$ (take $D_\alpha = \{C_\alpha \cap \beta \mid \beta < \alpha\}$).
(5) $S^\kappa_\omega \in I[\kappa]$.
(6) If $\kappa^{<\lambda} = \kappa = \lambda^+$, then GC_λ-characterization holds for κ if and only if $\kappa \in I[\kappa]$, if and only if $S^\kappa_\lambda \in I[\kappa]$; see [**8**, Corollary 2.4] and [**14**].
(7) The existence of $\lambda < \kappa$ such that GC_λ-characterization does not hold for κ is equiconsistent with the existence of a Mahlo cardinal.[1] Briefly this is because the failure of the characterization implies the failure of $\square_\lambda$, which implies that λ^+ is Mahlo in L as discussed above. On the other hand, in the Mitchell model, obtained from $S_{\mathrm{in}} = \{\delta < \lambda \mid \delta \text{ is inaccessible}\}$ where $\lambda > \kappa$ is Mahlo, it holds that $S_{\mathrm{in}} \notin I[\kappa^+]$; see [**8**, Lemma 2.6].
(8) If κ is regular and for all regular $\mu < \kappa$ we have $\mu^{<\lambda} < \kappa$, then $\kappa \in I[\kappa]$.

Remark 2.8 As Remark 2.7 shows, the assumption that GC_λ-characterization holds for κ is quite weak. For instance, GC_ω-characterization holds for all regular $\kappa > \omega$ and GCH implies that GC_λ-characterization holds for κ for all regular $\lambda < \kappa$.

3 Main results

Theorems 3.1 and 3.2 constitute the goal of this work. They are stated below but proved in the end of this section, starting at pages 589 and 592 respectively.

Theorem 3.1 *Assume that $\lambda < \kappa$ are regular and GC_λ-characterization holds for κ. Then the order $\langle \mathcal{P}(\kappa), \subset_{\mathrm{NS}(\lambda)}\rangle$ can be embedded into $\langle \mathrm{E}^B_\kappa, \le_B\rangle$ strictly between id_{2^κ} and E_0. More precisely, there exists a one-to-one map $F\colon \mathcal{P}(\kappa) \to \mathrm{E}^B_\kappa$ such that for all $X, Y \in \mathcal{P}(\kappa)$ we have $\mathrm{id}_{2^\kappa} \lneq_B F(X) \lneq_B E_0$ and*

$$X \subset_{\mathrm{NS}(\lambda)} Y \iff F(X) \le_B F(Y).$$

Theorem 3.2 *Assume either $\kappa = \omega_1$ or $\kappa = \lambda^+ > \omega_1$ and $\square_\lambda$. Then the partial order $\langle \mathcal{P}(\kappa), \subset_{\mathrm{NS}}\rangle$ can be embedded into $\langle \mathrm{E}^B_\kappa, \le_B\rangle$.*

[1] A good exposition of this result can be found in Lauri Tuomi's Master's thesis (University of Helsinki, 2009).

3.1 Corollaries

Corollary 3.3 *Assume that $\lambda < \kappa$ is regular. Additionally assume one of the following:*

(1) *$\kappa = \mu^+$, μ is regular and $\lambda < \mu$;*
(2) *$\kappa = \lambda^+$ and $\square_\lambda$ holds;*
(3) *for all regular $\mu < \kappa$, $\mu^{<\lambda} < \kappa$ (e.g., κ is ω_1 or inaccessible).*

Then the partial order $\langle \mathcal{P}(\kappa), \subset_{\mathrm{NS}(\lambda)} \rangle$ can be embedded into $\langle \mathrm{E}^B_\kappa, \leq_B \rangle$.

Proof. Any of the assumptions (1)–(3) is sufficient to obtain GC_λ-characterization for κ by Remarks 2.7 and 2.5, so the result follows from Theorem 3.1. □

Corollary 3.4 *The partial order $\langle \mathcal{P}(\kappa), \subset_{\mathrm{NS}(\omega)} \rangle$ can be embedded into $\langle \mathrm{E}^B_\kappa, \leq_B \rangle$. In particular, $\langle \mathcal{P}(\omega_1), \subset_{\mathrm{NS}} \rangle$ can be embedded into $\langle \mathrm{E}^B_{\omega_1}, \leq_B \rangle$ assuming CH.*

Proof. By Remark 2.7, GC_ω-characterization holds for κ for any regular $\kappa > \omega$, so the result follows from Theorem 3.1. □

Definition 3.5 Let $S \subset \kappa$. Then the combinatorial principle $\diamondsuit_\kappa(S)$ states that there exists a sequence $\langle D_\alpha \mid \alpha \in S \rangle$ such that for every $A \subset \kappa$ the set $\{\alpha \mid A \cap \alpha = D_\alpha\}$ is stationary.

Theorem 3.6 (Shelah [**15**]) *If $\kappa = \lambda^+ = 2^\lambda$ and $S \subset \kappa \setminus S^\kappa_{\mathrm{cf}(\lambda)}$ is stationary, then $\diamondsuit_\kappa(S)$ holds.* □

Corollary 3.7

(1) *The ordering $\langle \mathcal{P}(\kappa), \subset \rangle$ can be embedded into $\langle \mathrm{E}^B_\kappa, \leq_B \rangle$.*
(2) *Assume that $\kappa = \omega_1$ and $\diamondsuit_{\omega_1}$ holds or that κ is not a successor of an ω-cofinal cardinal. Then also the ordering $\langle \mathcal{P}(\kappa), \subset_* \rangle$ can be embedded into $\langle \mathrm{E}^B_\kappa, \leq_B \rangle$, where $\subset_*$ is inclusion modulo bounded sets.*

Proof. For the first part it is sufficient to show that the partial order $\langle \mathcal{P}(\kappa), \subset \rangle$ can be embedded into $\langle \mathcal{P}(\kappa), \subset_{\mathrm{NS}(\omega)} \rangle$. Let $G(A) = \Cup_{i \in A} S_i$ where $\{S_i \subset S^\kappa_\omega \mid i < \kappa\}$ is a collection of disjoint stationary sets. Then $A \subset B \Leftrightarrow G(A) \subset_{\mathrm{NS}} G(B)$, so this proves the first part.

For the second part, let us show that if $\diamondsuit_\kappa(S^\kappa_\lambda)$ holds, then $\langle \mathcal{P}(\kappa), \subset_* \rangle$ can be embedded into $\langle \mathcal{P}(\kappa), \subset_{\mathrm{NS}(\lambda)} \rangle$. Then the result follows. If $\kappa = \omega_1$ and $\diamondsuit_{\omega_1}$ holds, then it follows by Corollary 3.4. On the other hand, if κ is not a successor of an ω-cofinal cardinal, then from Theorem 3.6 it follows that $\diamondsuit_\kappa(S^\kappa_\omega)$ holds and the result follows again from Corollary 3.4.

Suppose that $\langle D_\alpha \mid \alpha \in S^\kappa_\lambda \rangle$ is a $\diamondsuit_\kappa(S^\kappa_\lambda)$-sequence. If $X, Y \subset \alpha$ for $\alpha \leq \kappa$, let $X \subset_* Y$ denote that there is $\beta < \alpha$ such that $X \setminus \beta \subset Y \setminus \beta$, i.e., X is a subset of Y on a final segment of α. Note that this coincides with the earlier defined $\subset_*$ when $\alpha = \kappa$. For $A \subset \kappa$, let

$$H(A) = \{\alpha < \kappa \mid D_\alpha \subset_* A \cap \alpha\}.$$

If $A \subset_* B$ then there is $\gamma < \kappa$ such that $A \setminus \gamma \subset B \setminus \gamma$ and if $\beta > \gamma$ is in $H(A)$, then $D_\beta \subset_* A \cap \beta$ and since $A \cap \beta \subset_* B \cap \beta$, we have $D_\beta \subset_* B \cap \beta$, so $H(A) \subset_* H(B)$ which finally implies $H(A) \subset_{\mathrm{NS}(\omega)} H(B)$.

Assume now that $A \not\subset_* B$ and let $C = A \setminus B$. Let S' be the stationary set such that for all $\alpha \in S'$, $C \cap \alpha = D_\alpha$. Let S be the λ-stationary set $S' \cap \{\alpha \mid C$ is unbounded below $\alpha\}$. S is stationary, because it is the intersection of S' and a cub set. Now for all $\alpha \in S$ we have $D_\alpha = C \cap \alpha \subset A \cap \alpha$, so $S \subset H(A)$. On the other hand, if $\alpha \in S$, then

$$D_\alpha \setminus (B \cap \alpha) = (C \cap \alpha) \setminus (B \cap \alpha) = ((A \setminus B) \cap \alpha) \setminus (B \cap \alpha) = C \cap \alpha$$

is unbounded in α, so $D_\alpha \not\subset_* B\cap\alpha$ and therefore $S\subset H(A)\setminus H(B)$, whence we conclude that $H(A)\not\subset_{\mathrm{NS}(\lambda)} H(B)$. □

Corollary 3.8 *There are 2^κ equivalence relations between* id *and E_0 that form a linear order with respect to $\lneq_B$.*

Proof. Let $K=\{\eta\in 2^\kappa \mid (\exists\beta)(\forall\gamma>\beta)(\eta(\gamma)=0)\}$, let $f\colon K\to\kappa$ be a bijection and for $\eta,\xi\in 2^\kappa$ define $\eta\lessdot\xi$ if and only if

$$\eta(\min\{\alpha\mid\eta(\alpha)\neq\xi(\alpha)\})<\xi(\min\{\alpha\mid\eta(\alpha)\neq\xi(\alpha)\}).$$

For $\eta\in 2^\kappa$, let $A_\eta=\{f(\xi)\mid \xi\lessdot\eta\wedge\xi\in K\}$. Clearly $A_\eta\subsetneq A_\xi$ if and only if $\eta\lessdot\xi$ and the latter is a linear order. The statement now follows from Corollary 3.7. □

3.2 Preparing for the proofs

Definition 3.9 For each $S\subset\lim\kappa$ let us define equivalence relations E^*_S, E_S and $E^*_S(\alpha)$, $\alpha\le\kappa$, on the space 2^κ as follows. Suppose $\eta,\xi\in 2^\delta$ for some $\delta\le\kappa$ and let $\zeta=\eta\,\mathrm{sd}\,\xi$. Let us define η and ξ to be $E^*_S(\delta)$-equivalent if and only if for all ordinals $\alpha\in S\cap\delta$ there exists $\beta<\alpha$ such that $\zeta(\gamma)$ has the same value for all $\gamma\in(\beta,\alpha)$. Let $E^*_S=E^*_S(\kappa)$ and $E_S=E^*_S\cap E_0$, where E_0 is the equivalence modulo bounded sets.

Remark If $S=\varnothing$, then $E_S=E_\varnothing=E_0$. If $S=\lim\kappa$ or equivalently if $S=\lim_\omega\kappa=S^\kappa_\omega$ (ω-cofinal limit ordinals), then $E_S=E_0'$, where E_0' is defined in [**4**].

Theorem 3.10 *For any $S\subset\lim\kappa$ the equivalence relations E_S and E^*_S are Borel.*

Proof. This is obvious by writing out the definitions:

$$E^*_S=\Cap_{\alpha\in S}\Cup_{\beta<\alpha}\Big(\Cap_{\beta<\gamma<\alpha}\{(\eta,\xi)\mid\eta(\gamma)\neq\xi(\gamma)\}\cup\Cap_{\beta<\gamma<\alpha}\{(\eta,\xi)\mid\eta(\gamma)=\xi(\gamma)\}\Big);$$
$$E_0=\Cup_{\alpha<\kappa}\Cap_{\alpha<\beta<\kappa}\{(\eta,\xi)\mid\eta(\beta)=\xi(\beta)\};$$
$$E_S=E^*_S\cap E_0.$$
□

The ideas of the following proofs are simple, but are repeated many times in this article in one way or another.

Theorem 3.11 *For all $S\subset\lim\kappa$, $E_S\not\le_B\mathrm{id}_{2^\kappa}$ and $E^*_S\le_B\mathrm{id}_{2^\kappa}$.*

Proof. For the first part suppose f is a Borel reduction from E_S to id_{2^κ}. Let η be a function such that η and $\overline{\eta}=1-\eta$ are both in $C(f)$ (see Definition 1.2, page 581). This is possible by Lemma 1.3, page 581. Then $(\eta,\overline{\eta})\notin E_S$. Let α be so large that $f(\eta)\restriction\alpha\neq f(\overline{\eta})\restriction\alpha$ and pick β so that

$$f[N_{\eta\restriction\beta}\cap C(f)]\subset N_{f(\eta)\restriction\alpha}$$

and

$$f[N_{\overline{\eta}\restriction\beta}\cap C(f)]\subset N_{f(\eta)\restriction\alpha}.$$

This is possible by the continuity of f on $C(f)$. By Lemma 1.3 pick now a $\zeta\in 2^{[\beta,\kappa)}$ so that $\eta\restriction\beta^\frown\zeta\in C(f)$ and $\overline{\eta}\restriction\beta^\frown\zeta\in C(f)$ which provides us with a contradiction, since

$$\big(\eta\restriction\beta^\frown\zeta,\overline{\eta}\restriction\beta^\frown\zeta\big)\in E_S,\ \text{but}\ f(\eta\restriction\beta^\frown\zeta)\neq f(\overline{\eta}\restriction\beta^\frown\zeta).$$

To prove the second part it is sufficient to construct a reduction from E^*_S to $\mathrm{id}_{\kappa^\kappa}$, since $\mathrm{id}_{\kappa^\kappa}$ and id_{2^κ} are bireducible (see [**5**]). Let us define an equivalence relation $\sim$ on $2^{<\kappa}$ such that $p\sim q$ if and only if $\operatorname{dom}p=\operatorname{dom}q$ and $p\,\mathrm{sd}\,q$ is eventually constant, i.e., for some $\alpha < \operatorname{dom} p$, $(p \operatorname{sd} q)(\gamma)$ is the same for all $\gamma \in [\alpha, \operatorname{dom} p)$. Let $s\colon 2^{<\kappa} \to \kappa$ be a map such that $p \sim q \Leftrightarrow s(p) = s(q)$. Suppose $\eta \in 2^\kappa$ and let us define $\xi = f(\eta)$ as follows. Let β_γ denote the γ-th element of S and let $\xi(\gamma) = s(\eta \restriction \beta_\gamma)$. Now we have $\eta E^*_S \xi$ if and only if $\eta \restriction \beta_\gamma = \xi \restriction \beta_\gamma$ for all $\gamma \in \kappa$, if and only if $f(\eta) = f(\xi)$. □

Corollary 3.12 *Let $S \subset \kappa$. If $p \in 2^{<\kappa}$ and $C \subset N_p$ is any co-meager subset of N_p, then there is no continuous function $C \to 2^\kappa$ such that $(\eta, \xi) \in E_S \cap C^2 \Leftrightarrow f(\eta) = f(\xi)$.*

Proof. Apply the same proof as for the first part of Theorem 3.11; take C instead of $C(f)$ and work inside N_p, e.g., instead of $\eta, \overline{\eta}$ take $p^\frown\eta, p^\frown\overline{\eta}$ for suitable $\eta \in 2^{[\operatorname{dom} p, \kappa)}$. □

Definition 3.13 A set $A \subset \kappa$ *does not reflect* to an ordinal α if the set $\alpha \cap A$ is non-stationary in α, i.e., there exists a closed unbounded subset of α outside of $A \cap \alpha$.

Theorem 3.14 *If $\kappa = \lambda^+ > \omega_1$ and $\square^\kappa_\mu$ holds, $\mu \le \lambda$, then for every stationary $S \subset S^\kappa_\omega$ there exists a set $B^\mu_{\mathrm{nr}}(S) \subset S$ (*nr *for non-reflecting) such that $B^\mu_{\mathrm{nr}}(S)$ does not reflect to any $\alpha \in S^\kappa_{\le\mu} \cap S^\kappa_{\ge\omega_1}$ and the sets $\lim C_\alpha$ witness that, where $\langle C_\alpha \mid \alpha \in S^\kappa_{\le\mu}\rangle$ is the $\square_\lambda$-sequence, i.e., $\lim C_\alpha \subset \alpha \setminus B^\mu_{\mathrm{nr}}(S)$ for $\alpha \in S^\kappa_{\le\mu} \cap S^\kappa_{\ge\omega_1}$. Since* $\operatorname{cf}(\alpha) > \omega$, *$\lim C_\alpha$ is cub in α.*

Proof. This is a well known argument and can be found in [**11**]. Let $g\colon S \to \kappa$ be the function defined by $g(\alpha) = \mathrm{OTP}(C_\alpha)$. By the definition of $\square_\mu$, $\mathrm{OTP}(C_\alpha) < \mu$ for $\alpha \in S^\kappa_\omega$, so for $\alpha > \mu$ we have $g(\alpha) < \alpha$. By Fodor's lemma there exists a stationary $B^\mu_{\mathrm{nr}}(S) \subset S$ such that $\mathrm{OTP}(C_\alpha) = \mathrm{OTP}(C_\beta)$ for all $\alpha, \beta \in B^\mu_{\mathrm{nr}}(\mu)$. If $\alpha \in \lim C_\beta$, then $C_\alpha = C_\beta \cap \alpha$ and therefore $\mathrm{OTP}(C_\alpha) < \mathrm{OTP}(C_\beta)$. Hence $\lim C_\beta \subset \beta \setminus B^\mu_{\mathrm{nr}}(S)$. □

Definition 3.15 Let E_i be equivalence relations on $2^{\kappa\times\{i\}}$ for all $i < \alpha$ where $\alpha < \kappa$. Let $E = \bigotimes_{i<\alpha} E_i$ be an equivalence relation on the space $2^{\kappa\times\alpha}$ such that $(\eta, \xi) \in E$ if and only if for all $i < \alpha$, $(\eta \restriction (\kappa \times \{i\}), \xi \restriction (\kappa \times \{i\})) \in E_i$.

Naturally, if $\alpha = 2$, we denote $\bigotimes_{i<2} E_i$ by just $E_0 \otimes E_1$ and we constantly identify $2^{\kappa\times\{i\}}$ with 2^κ.

Definition 3.16 Given equivalence relations E_i on $2^{\kappa\times\{i\}}$ for $i < \alpha < \kappa^+$, let $\bigoplus_{i\in I} E_i$ be an equivalence relation on $\Cup_{i<\alpha} 2^{\kappa\times\{i\}}$ such that η and ξ are equivalent if and only if for some $i < \alpha$, $\eta, \xi \in 2^{\kappa\times\{i\}}$ and $(\eta, \xi) \in E_i$.

Intuitively the operation $\oplus$ is taking disjoint unions of the equivalence relations. As above, if say $\alpha = 2$, we denote $\bigoplus_{i<2} E_i$ by just $E_0 \oplus E_1$ and we identify $2^{\kappa\times\{i\}}$ with 2^κ.

Theorem 3.17 *Assume that $\lambda \in \operatorname{reg} \kappa$ and* GC_λ*-characterization holds for κ.*

(1) *Suppose that $S_1, S_2 \subset S^\kappa_{\ge\lambda}$ and that $(S_2 \setminus S_1) \cap S^\kappa_\lambda$ is stationary. Then the following holds:*
   (a) $E_{S_1} \not\le_B E_{S_2}$.
   (b) *If $p \in 2^{<\kappa}$ and $C \subset N_p$ is any co-meager subset of N_p, then there is no continuous function $C \to 2^\kappa$ such that $(\eta, \xi) \in E_{S_1} \cap C^2 \Leftrightarrow (f(\eta), f(\xi)) \in E_{S_2}$.*

(2) *Assume that $\kappa = \lambda^+ > \omega_1$, $\mu \in \operatorname{reg}(\kappa) \setminus \{\omega\}$ and $\square^\kappa_\mu$ holds. Let $S \subset S^\kappa_\omega$ be any stationary set and let $B^\mu_{\mathrm{nr}}(S)$ be the set defined by Theorem* 3.14*. Then the following holds:*
   (a) *Suppose that $S_1, S_2 \subset S^\kappa_\mu$, $B \subset B^\mu_{\mathrm{nr}}(S)$ and let $S'_1 = S_1 \cup B$, $S'_2 = S_2 \cup B$. If $(S'_2 \setminus S'_1) \cap S^\kappa_\mu$ is stationary, then $E_{S'_1} \not\le_B E_{S'_2}$.*
   (b) *Let S_1, S_2, B, S'_1 and S'_2 be as above. If $(S'_2 \setminus S'_1) \cap S^\kappa_\mu$ is stationary, $p \in 2^{<\kappa}$ and $C \subset N_p$ is any co-meager subset of N_p, then there is no continuous function $C \to 2^\kappa$ such that $(\eta, \xi) \in E_{S'_1} \cap C^2 \Leftrightarrow (f(\eta), f(\xi)) \in E_{S'_2}$.*

(3) *Let $S_1, S_2, A_1, A_2 \subset S^\kappa_\omega$ be either such that $S_2 \setminus S_1$ and $A_2 \setminus S_1$ are stationary* or *such that $S_2 \setminus A_1$ and $A_2 \setminus A_1$ are stationary. Then the following holds:*
  (a) $E_{S_1} \otimes E_{A_1} \not\le_B E_{S_2} \otimes E_{A_2}$.
  (b) *If $C \subset (2^\kappa)^2$ (we identify $2^{\kappa\times 2}$ with $(2^\kappa)^2$) is a set which is co-meager in some $N_r = \{\eta \in (2^\kappa)^2 \mid \eta \restriction \operatorname{dom} r = r\}$, $r \in (2^\alpha)^2$, $\alpha < \kappa$, then there is no continuous function f from $C \cap N_r$ to $(2^\kappa)^2$ such that $(\eta, \xi) \in (E_{S_1} \otimes E_{A_1}) \cap C^2 \Leftrightarrow (f(\eta), f(\xi)) \in E_{S_2} \otimes E_{A_2}$.*

(4) *Assume that $S_1, S_2, A_2 \subset \kappa$ are such that $A_2 \setminus S_1$ and $S_2 \setminus S_1$ are ω-stationary. Then:*
  (a) $E_{S_1} \not\le_B E_{S_2} \otimes E_{A_2}$.
  (b) *If $p \in 2^{<\kappa}$ and $C \subset N_p$ is any co-meager subset of N_p, there is no continuous function $C \to (2^\kappa)^2$ such that $(\eta, \xi) \in E_{S_1} \cap C^2 \Leftrightarrow (f(\eta), f(\xi)) \in E_{S_2} \otimes E_{A_2}$.*

(5) *Assume that $S_1, A_1, S_2, A_2 \subset \kappa$ are such that $A_2 \setminus A_1$ is ω-stationary. Then:*
  (a) $E_{S_1} \otimes E_{A_1} \not\le_B E_{S_2 \cup A_2}$.
  (b) *If $p \in (2^{<\kappa})^2$ and $C \subset N_p$ is any co-meager subset of N_p, there is no continuous function $C \to 2^\kappa$ such that $(\eta, \xi) \in (E_{S_1} \otimes E_{A_1}) \cap C^2 \Leftrightarrow (f(\eta), f(\xi)) \in E_{S_2 \cup A_2}$.*

Proof. Item (1b) of the theorem implies item (1a), and all (b)-parts imply the corresponding (a)-parts, because if $f\colon 2^\kappa \to 2^\kappa$ is a Borel function, then it is continuous on the co-meager set $C(f)$ (see Definition 1.2). Let us start by proving (1b).

Assume that $S_2 \setminus S_1$ is λ-stationary, $p \in 2^{<\kappa}$, $C \subset N_p$, and assume that $f\colon C \to 2^\kappa$ is a continuous function as described in the Theorem. Let us derive a contradiction. Define a strategy for player **II** in the game $\mathrm{GC}_\lambda(\kappa \setminus (S_2 \setminus S_1))$ as follows.

Denote the i-th move of player **I** by α_i and the i-th move of player **II** by β_i. During the game, at the i-th move, $i < \lambda$, player **II** secretly defines functions $p^0_i, p^1_i, q^0_i, q^1_i \in 2^{<\kappa}$ in such a way that for all i and all $j < i$ we have

(a) $\operatorname{dom} p^0_j = \operatorname{dom} p^1_j = \beta_j$ and $\alpha_j \le \operatorname{dom} q^0_{j+1} = \operatorname{dom} q^1_{j+1} \le \alpha_j$, and if j is a limit, then $\sup_{i<j} \alpha_i \le \operatorname{dom} q^0_j = \operatorname{dom} q^1_j \le \beta_j$;
(b) $p^0_j \subset p^0_{j+1}$, $p^1_i \subset p^1_{i+1}$, $q^0_i \subset q^0_{i+1}$ and $q^1_i \subset q^1_{i+1}$;
(c) $f[C \cap N_{p^0_i}] \subset N_{q^0_i}$ and $f[C \cap N_{p^1_i}] \subset N_{q^1_i}$.

Suppose it is i-th move and $i = \gamma + 2k$ for some $k < \omega$ and γ which is either 0 or a limit ordinal, and suppose that the players have picked the sequences $(\alpha_j)_{j \le i}$ and $(\beta_j)_{j<i}$. Additionally **II** has secretly picked the sequences

$$(p^0_i)_{i<j},\ (p^1_i)_{i<j},\ (q^0_i)_{i<j},\ (q^1_i)_{i<j},$$

which satisfy conditions (a)–(c). Assume first that i is a successor. In this case, if q^0_{i-1} is not $E^*_{S_2}(\operatorname{dom} q^0_{i-1})$-equivalent to q^1_{i-1}, then player **II** plays arbitrarily. Otherwise, to decide her next move, player **II** uses Lemma 1.3 (page 581) to find $\eta \in 2^{[\beta_{i-1}, \kappa)}$ and $\xi = 1 - \eta$, such that $p^0_{i-1} {}^\frown \eta \in C$ and $p^1_{i-1} {}^\frown \xi \in C$. Then she finds $\beta'_i > \alpha_i$ such that $f(p^0_{i-1} {}^\frown \eta)(\delta) \neq f(p^1_{i-1} {}^\frown \xi)(\delta)$ for some $\delta \in [\alpha_i, \beta'_i)$. This is possible since f is a reduction and $(q^0_{i-1}, q^1_{i-1}) \in E^*_{S_2}$. Then she picks $\beta_i > \beta'_i$ so that

$$f[C \cap N_{(p^0_{i-1} {}^\frown \eta) \restriction \beta_i}] \subset N_{f(p^0_{i-1} {}^\frown \eta) \restriction \beta'_i}$$

and

$$f[C \cap N_{(p^1_{i-1} {}^\frown \xi) \restriction \beta_i}] \subset N_{f(p^1_{i-1} {}^\frown \xi) \restriction \beta'_i}.$$

This choice is possible by the continuity of f. Then she (secretly) sets $p_i^0 = (p_{i-1}^0 {}^\frown \eta) \restriction \beta_i$, $p_i^1 = (p_{i-1}^1 {}^\frown \xi) \restriction \beta_i$, $q_i^0 = f(p_{i-1}^0 {}^\frown \eta) \restriction \beta_i'$ and $q_i^1 = f(p_{i-1}^1 {}^\frown \xi) \restriction \beta_i'$. Note that the new partial functions secretly picked by **II** satisfy conditions (a)–(c).

If i is a limit, then player **II** proceeds as above but instead of p_{i-1}^n she uses $\Cup_{i'<i} p_{i'}^n$, $n \in \{0,1\}$, and instead of β_{i-1} she uses $\sup_{i'<i} \beta_{i'}$. If i is 0, then proceed in the same way assuming $p_{-1}^0 = p_{-1}^1 = q_{-1}^0 = q_{-1}^1 = \varnothing$ and $\alpha_{-1} = \beta_{-1} = 0$.

Suppose $i = \gamma + 2k + 1$ where γ is again a limit or zero and $k < \omega$. Then the moves go in the same way, except that she sets $\eta = \xi$ instead of $\eta = 1 - \xi$ and requires $f(p_{i-1}^0 {}^\frown \eta)(\delta) = f(p_{i-1}^1 {}^\frown \xi)(\delta)$ for some $\delta \in [\alpha_{i-1}, \beta_i')$ instead of $f(p_{i-1}^0 {}^\frown \eta)(\delta) \neq f(p_{i-1}^1 {}^\frown \xi)(\delta)$ for some $\delta \in [\alpha_{i-1}, \beta_i')$. Denote this strategy by σ.

Since $S_2 \setminus S_1$ is stationary and GC$_\lambda$-characterization holds for κ, player **I** is able to play against this strategy in such a way that $\sup_{i<\lambda} \alpha_i \in S_2 \setminus S_1$. Suppose they have played the game to the end, so that player **II** used σ, player **I** has won and they have picked the sequence $\langle \alpha_i, \beta_i \mid i < \lambda \rangle$. Let

$$\alpha_\lambda = \sup_{i<\lambda} \alpha_i = \sup_{i<\lambda} \beta_i = \sup_{i<\lambda} \operatorname{dom} p_i = \sup_{i<\lambda} \operatorname{dom} q_i$$

and

$$p_\lambda^0 = \Cup_{i<\lambda} p_i^0,\ p_\lambda^1 = \Cup_{i<\lambda} p_i^1,\ q_\lambda^0 = \Cup_{i<\lambda} q_i^0 \text{ and } q_\lambda^1 = \Cup_{i<\lambda} q_i^1.$$

By continuity, p_λ^0, p_λ^1, q_λ^0 and q_λ^1 satisfy condition (c) above and $\operatorname{dom} p_\lambda^0 = \operatorname{dom} p_\lambda^1 = \operatorname{dom} q_\lambda^0 = \operatorname{dom} q_\lambda^1 = \sup_{i<\lambda} \alpha_i = \sup_{i<\lambda} \beta_i$, so α_λ is well defined.

On one hand q_λ^0 and q_λ^1 cannot be extended in an E_{S_2}-equivalent way, since either they cofinally get same and different values below $\alpha_\lambda \in S_2$, or they are not $E_{S_2}^*(\gamma)$-equivalent already for some $\gamma < \alpha_\lambda$. On the other hand p_λ^0 and p_λ^1 can be extended in an E_{S_1}-equivalent way, since α_λ is not in S_1 and for all $\gamma < \lambda$, $\sup_{i<\gamma} \alpha_\gamma$ is not μ-cofinal for any $\mu \geq \lambda$, so cannot be in S_1 either $(*)$.

Let $\eta, \xi \in 2^\kappa$ be extensions of p_λ^0 and p_λ^1 respectively such that $(\eta, \xi) \in E_{S_1} \cap C^2$. Now $f(\eta)$ and $f(\xi)$ cannot be E_{S_2}-equivalent, since by condition (c), they must extend q_λ^0 and q_λ^1 respectively.

Now let us prove (2b), which implies (2a). Let $\langle C_\alpha^\mu \mid \alpha \in S_{\leq\mu}^\kappa \rangle$ be the $\square_\mu^\kappa$-sequence and denote by t^μ the function $\alpha \mapsto C_\alpha^\mu$.

Let player **II** define her strategy in the game $\mathrm{GC}(\kappa \setminus (S_2' \setminus S_1'))$ exactly as in the proof of (1b). Note that $S_2' \setminus S_1' = S_2 \setminus S_1$ since $\mu > \omega$. Denote this strategy by σ. We know that, as above, player **I** is able to beat σ. However, now it is not enough, because in order to be able to extend p_μ^0 and p_μ^1 in an $E_{S_1'}$-equivalent way, he needs to ensure that

$$(**) \qquad S_1' \cap \lim\nolimits_\omega(\{\alpha_i \mid i < \mu\}) = \varnothing,$$

where $\lim_\omega X$ is the set of ω-limits of elements of X, i.e., we cannot rely on the sentence followed by $(*)$ above. On the other hand $(**)$ is sufficient, because $S_1' \subset S_\mu^\kappa \cup S_\omega^\kappa$.

Let us show that it is possible for player **I** to play against σ as required.

Let $\nu > \kappa$ be a sufficiently large cardinal and let M be an elementary submodel of $\langle H_\nu, \sigma, \kappa, t^\mu \rangle$ such that $|M| < \kappa$ and $\alpha = \kappa \cap M$ is an ordinal in $S_2' \setminus S_1'$.

In the game, suppose that the sequence $d = \langle \alpha_j, \beta_j \mid j < i \rangle$ has been played before move i and suppose that this sequence is in M. Player **I** will now pick α_i to be the smallest element in C_α^μ which is above $\sup_{j<i} \beta_j$. Since $C_\alpha^\mu \cap \beta = C_\beta^\mu$ for any $\beta \in \lim C_\alpha^\mu$ and $C_\beta^\mu \in M$, this element is definable in M from the sequence d and t^μ. This guarantees that the sequence obtained on the following move is also in M. At limits the sequence is

in M, because it is definable from t^μ and σ. Since $\mathrm{OTP}(C^\mu_\alpha) = \mu$, the game ends at α and player **I** wins. Also the requirement $(**)$ is satisfied because he picked elements only from C^μ_α and so $\lim_\omega\{\alpha_i \mid i < \mu\} \subset \lim_\omega(C^\mu_\alpha) \subset \alpha \setminus B$ which gives the result.

Next let us prove (3b), which again implies (3a). The proofs of (4) and (5) are very similar to that of (3) and are left to the reader.

So, let S_1, A_1, S_2, A_2, C and r be as in the statement of (3) and suppose that there is a counterexample f. Assume that $S_2 \setminus S_1$ and $A_2 \setminus S_1$ are stationary, the other case being symmetric. Let us define property P:

P: There exist $p, p' \in (2^\alpha)^2$, $p = (p_1, p_2)$ and $p' = (p'_1, p'_2)$ such that
(a) $r \subset p \cap p'$;
(b) $p_2 = p'_2$, $(p_1, p'_1) \in E^*_{S_1}(\alpha + 1)$ (see Definition 3.9, page 584);
(c) for all $\eta \in C \cap N_p$ and $\eta' \in C \cap N_{p'}$, $\eta = (\eta_1, \eta_2)$, $\eta' = (\eta'_1, \eta'_2)$, if $\eta_2 = \eta'_2$ and $(\eta_1, \eta'_1) \in E^*_{S_1}$, then $f(\eta)_1 \operatorname{sd} f(\eta')_1 \subset \operatorname{dom} p_1$ where $f(\eta) = (f(\eta)_1, f(\eta)_2)$.

We will show that both P and $\neg P$ lead to a contradiction. Assume first $\neg P$. Now the argument is similar to the proof of (1b). Player **II** defines her strategy in the same way but this time she chooses the elements p^n_i and q^n_i from $(2^\alpha)^2$ instead of 2^α so that $p^n_i = (p^n_{i,1}, p^n_{i,2})$, $q^n_i = (q^n_{i,1}, q^n_{i,2})$ and, for all $i < \lambda$, $p^0_{i,2} = p^1_{i,2}$. In building the strategy she looks only at $q^n_{i,1}$ and ignores $q^n_{i,2}$. In other words she pretends that the game is for E_{S_1} and E_{S_2} in the proof of (1). At the even moves she extends $p^0_{i,1}$ and $p^1_{i,1}$ by η and η' which witness the failure of item (c) (but not of (a) and (b)) of property P for p^0_i and p^1_i. Then there is $\alpha \in f(\eta)_1 \operatorname{sd} f(\eta')_1$, $\alpha > \operatorname{dom} p^0_{i,1}$. And then she chooses $q^0_{i,1}$ and $q^1_{i,1}$ to be initial segments of $f(\eta)_1$ and $f(\eta')_1$ respectively.

At the odd moves she just extends $p^0_{i,1}$ and $p^1_{i,1}$ in an E_{S_1}-equivalent way, so that she finds an $\alpha > \operatorname{dom} p^0_{i,1}$, $q^0_{i,1}$ and $q^1_{i,1}$ such that $q^0_{i,1}(\alpha) = q^1_{i,1}(\alpha)$ and $f[N_{p^0_i} \cap C] \subset N_{q^0_i}$.

As in the proof of (1), **I** responses by playing towards an ordinal in $S_2 \setminus S_1$. During the game they either hit a point at which $q^0_{i,2}$ and $q^1_{i,2}$ cannot be extended to be E_{A_2}-equivalent or else they play the game to the end whence $q^0_{\lambda,1}$ and $q^1_{\lambda,1}$ cannot be extended in a E_{S_2}-equivalent way but p^0_λ and p^1_λ can be extended to $E_{S_1} \otimes E_{A_1}$-equivalent way.

Assume that P holds. Fix p and p' which witness that. Now player **II** builds her strategy as if they were playing between E_{S_1} and E_{A_2}. This time she concentrates on $q^0_{i,2}$ and $q^1_{i,2}$ instead of $q^0_{i,1}$ and $q^1_{i,1}$. At the even moves she extends $p^0_{i,1}$ and $p^1_{i,1}$ by η and $\overline{\eta}$ respectively for some η. Also, as above, $p^0_{i,2}$ and $p^1_{i,2}$ are extended in the same way. By item (c), $f(\eta)_1 \operatorname{sd} f(\eta')_1$ is bounded by $\operatorname{dom} p^0_{i,1}$, but $f(\eta)$ and $f(\eta')$ cannot be $E_{S_2} \otimes E_{A_2}$-equivalent, because f is assumed to be a reduction. Hence there must exist $\alpha > \operatorname{dom} p^0_{i,1}$, $q^0_{i,2}$ and $q^0_{i,2}$ such that $q^0_{i,2}(\alpha) \neq q^1_{i,2}(\alpha)$. The rest of the argument goes similarly as above. □

Corollary 3.18 *If* GC_λ*-characterization holds for* κ *and* $S \subset \kappa$ *is* λ*-stationary, then* $E_0 \not\leq E_S$*. In particular, if* S *is* ω*-stationary, then* $E_0 \not\leq E_S$*.*

Proof. This follows from Theorem 3.17(1a) by taking $S_1 = \varnothing$, since $E_\varnothing = E_0$ and GC_ω-characterization holds for κ. □

Corollary 3.19 *There is an antichain*[2] *of Borel equivalence relations on* 2^κ *of length* 2^κ*.*

[2] By an antichain I refer here to a family of pairwise incomparable elements unlike e.g. in forcing context.

Proof. Take disjoint ω-stationary sets S_i, $i < \kappa$. Let $f\colon \kappa \times 2 \to \kappa$ be a bijection. For each $\eta \in 2^\kappa$ let $A_\eta = \{(\alpha, n) \in \kappa \times 2 \mid (n = 0 \wedge \eta(\alpha) = 1) \vee (n = 1 \wedge \eta(\alpha) = 0)\}$. For each $\eta \neq \xi$ clearly $A_\eta \setminus A_\xi \neq \varnothing \neq A_\xi \setminus A_\eta$. Let

$$S_\eta = \Cup_{i \in f[A_\eta]} S_i.$$

Now $\{E_{S_\eta} \mid \eta \in 2^\kappa\}$ is an antichain by Theorem 3.17(1b). □

Let us show that all these relations are below E_0. It is already shown that they are not above it (Corollary 3.18), provided GC_λ-characterization holds for κ. Again, similar ideas will be used in the proof of Theorems 3.1 and 3.2.

Theorem 3.20 *For all S, $E_S \leq_B E_0$.*

Proof. Let us show that E_S is reducible to E_0 on κ^κ, which is in turn bireducible with E_0 on 2^κ (see [**5**]). Let us define an equivalence relation $\sim$ on $2^{<\kappa}$ as on page 585, such that $p \sim q$ if and only if $\operatorname{dom} p = \operatorname{dom} q$ and $p \operatorname{sd} q$ is eventually constant, i.e., for some $\alpha < \operatorname{dom} p$, $(p \operatorname{sd} q)(\gamma)$ is the same for all $\gamma \in [\alpha, \operatorname{dom} p)$. Let $s\colon 2^{<\kappa} \to \kappa$ be a map such that $p \sim q \Leftrightarrow s(p) = s(q)$. Let $\{A_i \mid i \in S\}$ be a partition of $\lim \kappa$ into disjoint unbounded sets. Suppose $\eta \in 2^\kappa$ and define $f(\eta) = \xi \in \kappa^\kappa$ as follows:

- If α is a successor, $\alpha = \beta + 1$, then $\xi(\alpha) = \eta(\beta)$.
- If α is a limit, then $\alpha \in A_i$ for some $i \in S$. Let $\xi(\alpha) = s(\eta \mid i)$.

Let us show that f is the desired reduction from E_S to E_0. Assume that η and ξ are E_S-equivalent. If α is a limit and $\alpha \in A_i$, then, since η and ξ are E_S-equivalent, we have $\eta \mid i \sim \xi \mid i$, so $s(\eta \mid i) = s(\xi \mid i)$ and so $f(\eta)(\alpha) = f(\xi)(\alpha)$. There is β such that $\eta(\gamma) = \xi(\gamma)$ for all $\gamma > \beta$. This implies that for all successors $\gamma > \beta$ we also have $f(\eta)(\gamma) = f(\xi)(\gamma)$. Hence $f(\eta)$ and $f(\xi)$ are E_0-equivalent. Assume now that η and ξ are not E_S-equivalent. Then there are two cases:

(1) $\eta \operatorname{sd} \xi$ is unbounded. Now $f(\eta)(\beta + 1) = \eta(\beta)$ and $f(\xi)(\beta + 1) = \xi(\beta)$ for all β, so we have

$$\{\beta \mid \eta(\beta) \neq \xi(\beta)\} = \{\beta \mid f(\eta)(\beta + 1) \neq \xi(\beta + 1)\}.$$

If the former is unbounded, then so is the latter.

(2) For some $i \in S$, $\eta \mid i \not\sim \xi \mid i$. This implies that $f(\eta)(\alpha) \neq f(\xi)(\alpha)$ for all $\alpha \in A_i$. and we get that $\{\beta \mid f(\eta)(\beta) \neq \xi(\beta)\}$ is again unbounded.

It is easy to check that f is continuous. □

3.3 Proofs of the main theorems

Proof of Theorem 3.1. The subject of the proof is that for a regular $\lambda < \kappa$, if GC_λ-characterization holds for κ, then the order $\langle \mathcal{P}(\kappa), \subset_{\mathrm{NS}(\lambda)} \rangle$ can be embedded into $\langle \mathrm{E}^B_\kappa, \leq_B \rangle$ strictly below E_0 and above id_{2^κ}.

Let $h\colon \omega \times \kappa \to \kappa$ be a bijection. Let $\tilde{h}\colon 2^{\omega \times \kappa} \to 2^\kappa$ be defined by $\tilde{h}(\eta)(\alpha) = \eta(h^{-1}(\alpha))$. We define the topology on $2^{\omega \times \kappa}$ to be generated by the sets $\{\tilde{h}^{-1} V \mid V \text{ is open in } 2^\kappa\}$. Then $\tilde{h}$ is a homeomorphism between $2^{\omega \times \kappa}$ and 2^κ. If $g\colon \kappa \times \kappa \to \kappa$ is a bijection, we similarly get a topology onto $2^{\kappa \times \kappa}$ and a homeomorphism $\tilde{g}$ from $2^{\kappa \times \kappa}$ onto 2^κ. By combining these two we get a homeomorphism between $2^{\omega \times \kappa} \times 2^\kappa$ and 2^κ, and so without loss of generality we can consider equivalence relations on these spaces.

For a given equivalence relation E on 2^κ, let $\overline{E}$ be the equivalence relation on $2^{\omega \times \kappa} \times 2^\kappa$ defined by

$$((\eta, \xi), (\eta', \xi')) \in \overline{E} \iff \eta = \eta' \wedge (\xi, \xi') \in E.$$

Essentially $\overline{E}$ is the same as $\mathrm{id}\otimes E$, since $2^{\omega\times\kappa}\approx 2^\kappa$.

Remark 3.21 Corollary 3.12, Theorem 3.17 and Corollary 3.18 hold even if E_S is replaced everywhere by $\overline{E_S}$ for all $S\subset\kappa$.

Proof. Let us show this for Theorem 3.17(1). The proof goes exactly as the proof of Theorem 3.17(1), but player **I** now picks the functions p^n_k from $\Cup_{\alpha<\kappa}2^{\omega\times\alpha}\times 2^\alpha$ instead of $2^{<\kappa}$, $p^n_k=(p^n_{k,1},p^n_{k,2})$, and requires that at each move $p^0_{k,1}=p^1_{k,1}$. Otherwise the argument proceeds in the same manner. Similarly for 3.17(2), 3.17(3), 3.17(4) and 3.17(5).

Modify the proof of the first part of Theorem 3.11 in a similar way to obtain the result for Corollary 3.12. Corollary 3.18 follows from the modified version of Theorem 3.17. □

For $S\subset\kappa$, let

$$G(S)=\overline{E_{S^\kappa_\lambda\setminus S}}.$$

Let us show that $G\colon\mathcal{P}(\kappa)\to\mathrm{E}^B_\kappa$ is the desired embedding. Without loss of generality let us assume that G is restricted to $\mathcal{P}(S^\kappa_\lambda)$, whence stationary is the same as λ-stationary and non-stationary is the same as not λ-stationary. For arbitrary $S_1,S_2\subset S^\kappa_\lambda$ we have to show:

(1) If $S_2\setminus S_1$ is stationary, then $\overline{E_{S_1}}\not\leq_B\overline{E_{S_2}}$.
(2) If $S_2\setminus S_1$ is non-stationary, then $\overline{E_{S_1}}\leq_B\overline{E_{S_2}}$.
(3) $\mathrm{id}_{2^\kappa}\lneq_B\overline{E_{S_1}}\lneq_B E_0$.

If $\eta\in 2^{\omega\times\kappa}$, denote $\eta_i(\alpha)=\eta(i,\alpha)$ and $(\eta_i)_{i<\omega}=\eta$.

Claim 1 If $S_2\setminus S_1$ is stationary, then $\overline{E_{S_1}}\not\leq_B\overline{E_{S_2}}$. Also $E_0\not\leq\overline{E_S}$.

Proof. It follows from Theorem 3.17(1a) and Remark 3.21. □

Claim 2 If $S_2\setminus S_1$ is non-stationary, then $\overline{E_{S_1}}\leq_B\overline{E_{S_2}}$.

Proof. Let us split this into two parts according to the stationarity of S_2. Assume first that S_2 is non-stationary. Let C be a cub set outside S_2. Let $f\colon 2^\kappa\to 2^{\omega\times\kappa}\times 2^\kappa$ be the function defined as follows. For $\eta\in 2^\kappa$, let $f(\eta)=\langle(\eta_i)_{i<\omega},\xi\rangle$ be such that $\eta_i(\alpha)=0$ for all $\alpha<\kappa$ and $i<\omega$, and $\xi(\alpha)=0$ for all $\alpha\notin C$. If $\alpha\in C$, then let $\xi(\alpha)=\eta(\mathrm{OTP}(\alpha\cap C))$. This is easily verified to be a reduction from E_0 to $\overline{E_{S_2}}$. By the following Claim 3, $\overline{E_{S_1}}\leq_B E_0$, so we are done.

Assume now that S_2 is stationary. Note that then S_1 is also stationary. Let C be a cub set such that $S_2\cap C\subset S_1$. Assume that $\langle(\eta_i)_{i<\omega},\xi\rangle\in 2^{\omega\times\kappa}\times 2^\kappa$ and let us define

$$f(\langle(\eta_i)_{i<\omega},\xi\rangle)=\langle(\eta'_i)_{i<\omega},\xi'\rangle\in 2^{\omega\times\kappa}\times 2^\kappa$$

as follows. For $i\geq 0$, let

$$\eta'_{i+1}=\eta_i.$$

For all $\alpha<\kappa$, let $\xi'(\alpha)=\xi(\min(C\setminus\alpha))$. Then let s be the function defined in the proof of Theorem 3.11 (on page 585) and for all $\alpha<\kappa$ let $\beta(\alpha)$ be the α-th element of $S_1\setminus S_2$. For all $\alpha<\kappa$, let

$$\eta'_0(\alpha)=s(\xi\restriction\beta(\alpha)).$$

Let us show that this defines a continuous reduction.

Suppose $\langle(\eta^0_i)_{i<\omega},\xi^0\rangle$ and $\langle(\eta^1_i)_{i<\omega},\xi^1\rangle$ are $\overline{E_{S_1}}$-equivalent. Denote their images under f by $\langle(\rho^0_i)_{i<\omega},\zeta^0\rangle$ and $\langle(\rho^1_i)_{i<\omega},\zeta^1\rangle$ respectively. Since $\eta^0_i=\eta^1_i$ for all $i<\omega$, we have $\rho^0_i=\rho^1_i$ for all $0<i<\omega$. Since for all $\alpha\in S_1$ we have that $\xi^0\restriction\alpha$ and $\xi^1\restriction\alpha$ are $\sim$-equivalent (as in the definition of s), we have that $\rho^0_0(\beta)=\rho^1_0(\beta)$ for all $\beta<\kappa$.

Suppose now that $\alpha \in S_2$. The aim is to show that $\zeta^0 \mid \alpha \sim \zeta^1 \mid \alpha$. If $\alpha \notin C$, then there is $\beta < \alpha$ such that $C \cap (\beta, \alpha) = \varnothing$, because C is closed. This implies that for all $\beta < \gamma < \gamma' < \alpha$, $\min(C \setminus \gamma') = \min(C \setminus \gamma)$, so by the definition of f, $\zeta^0(\gamma) = \zeta^0(\gamma')$ and $\zeta^1(\gamma) = \zeta^1(\gamma')$. Now by fixing γ_0 between β and α we deduce that $\zeta^0 \mid (\beta, \alpha)$ is constant and $\zeta^1 \mid (\beta, \alpha)$ is constant, since for all $\gamma < \alpha$ we have $\zeta^0(\gamma) = \zeta^0(\gamma_0)$ and $\zeta^1(\gamma) = \zeta^1(\gamma_0) = \zeta^1(\gamma)$. Hence $(\zeta^0 \operatorname{sd} \zeta^1) \mid (\beta, \alpha)$ is constant, which by the definition of $\sim$ implies that $\zeta^0 \mid \alpha \sim \zeta^1 \mid \alpha$.

If $\alpha \in C$, then, since α is also in S_2, we have by the definition of C that $\alpha \in S_1$. Thus, there is $\beta < \alpha$ such that $(\xi^0 \operatorname{sd} \xi^1) \mid (\beta, \alpha)$ is constant, which implies that for some $k \in \{0, 1\}$ we have $(\zeta^0 \operatorname{sd} \zeta^1)(\gamma) = k$ for all $\gamma \in (\beta, \alpha) \cap C$. But if $\gamma \in (\beta, \alpha) \setminus C$, then, again by the definition of f, we have $(\zeta^0 \operatorname{sd} \zeta^1)(\gamma) = (\zeta^0 \operatorname{sd} \zeta^1)(\gamma')$ for some $\gamma \in (\beta, \alpha) \cap C$, so $(\zeta^0 \operatorname{sd} \zeta^1)(\gamma)$ also equals k.

This shows that ζ^0 and ζ^1 are $E^*_{S_2}$-equivalent. It remains to show that they are E_0-equivalent. But since ξ^0 and ξ^1 are E_0-equivalent, the number $k \in \{0, 1\}$ referred above equals 0 for all α large enough and we are done.

Next let us show that if $\langle (\eta^0_i)_{i<\omega}, \xi^0 \rangle$ and $\langle (\eta^1_i)_{i<\omega}, \xi^1 \rangle$ are not $\overline{E_{S_1}}$-equivalent, then $\langle (\rho^0_i)_{i<\omega}, \zeta^0 \rangle$ and $\langle (\rho^1_i)_{i<\omega}, \zeta^1 \rangle$ are not $\overline{E_{S_2}}$-equivalent. This is just reversing implications of the above argument. If $\eta^0_i \neq \eta^1_i$ for some $i < \omega$, then $\rho^0_{i+1} \neq \rho^1_{i+1}$, so we can assume that $(\xi^0, \xi^1) \notin E_{S_1}$. If ξ^0 and ξ^1 are not $E^*_{S_1}$-equivalent, then $\rho^0(\alpha) \neq \rho^1(\alpha)$ for some $\alpha < \kappa$.

The remaining case is that ξ^0 and ξ^1 are $E^*_{S_1}$-equivalent but not E_0-equivalent. But then in fact $\xi^0 \operatorname{sd} \xi^1$ is eventually equal to 1, since otherwise the sets

$$C_1 = \{\alpha \mid \{\beta < \alpha \mid (\xi^0 \operatorname{sd} \xi^1)(\beta) = 1\} \text{ is unbounded in } \alpha\}$$

and

$$C_2 = \{\alpha \mid \{\beta < \alpha \mid (\xi^0 \operatorname{sd} \xi^1)(\beta) = 0\} \text{ is unbounded in } \alpha\}$$

are both cub and, by the stationarity of S_1, there exists a point $\alpha \in C_1 \cap C_2 \cap S_1$ which contradicts the fact that ξ_0 and ξ_1 are $E^*_{S_1}$-equivalent. So $\xi^0 \operatorname{sd} \xi^1$ is eventually equal to 1 and this finally implies that also ζ^0 and ζ^1 cannot be E_0-equivalent. □

Claim 3 Let $S \subset S^\kappa_\lambda$. Then $\mathrm{id} \lneq_B \overline{E_S} <_B E_0$. If S is stationary, then also $E_0 \not\leq_B \overline{E_S}$.

Proof. If $\eta \in 2^\kappa$, let $\eta_0 = \eta$ and $\eta_i(\alpha) = \xi(\alpha) = 0$ for all $\alpha < \kappa$. Then $\eta \mapsto \langle (\eta_i)_{i<\omega}, \xi \rangle$ defines a reduction from id to $\overline{E_S}$. On the other hand $\overline{E_S}$ is not reducible to id by Remark 3.21.

Let $u\colon 2^{\omega\times\kappa} \to 2^\kappa$ be a reduction from $\mathrm{id}_{2^{\omega\times\kappa}}$ to E_0. Let $v\colon 2^\kappa \to 2^\kappa$ be a reduction from E_S to E_0 which exists by 3.20. Let $\{A, B\}$ be a partition of κ into two disjoint unbounded subsets. Let $(\eta, \eta') \in 2^{\omega\times\kappa} \times 2^\kappa$ and let us define $\xi = f(\eta, \eta') \in 2^\kappa$. If $\alpha \in A$, then let $\xi(\alpha) = u(\eta)(\mathrm{OTP}(\alpha \cap A))$. If $\alpha \in B$, then let $\xi(\alpha) = v(\eta')(\mathrm{OTP}(\alpha \cap B))$. (See page 580 for notation.)

Now if $((\eta_0, \eta'_0), (\eta_1, \eta'_1)) \in (2^{\omega\times\kappa} \times 2^\kappa)^2$ are $\overline{E_S}$-equivalent, then $u(\eta_0) \operatorname{sd} u(\eta_1)$ and $v(\eta'_0) \operatorname{sd} v(\eta'_1)$ are eventually equal to zero, which clearly implies that $f(\eta_0, \eta'_0) \operatorname{sd} f(\eta_1, \eta'_1)$ is eventually zero, and so $f(\eta_0, \eta'_0)$ and $f(\eta_1, \eta'_1)$ are E_0-equivalent. Similarly, if (η_0, η'_0) and (η_1, η'_1) are not $\overline{E_S}$-equivalent, then either $u(\eta_0) \operatorname{sd} u(\eta_1)$ or $v(\eta'_0) \operatorname{sd} v(\eta'_1)$ is not eventually zero, and so $f(\eta_0, \eta'_0)$ and $f(\eta_1, \eta'_1)$ are not E_0-equivalent.

If S is stationary, then $E_0 \not\leq_B \overline{E_S}$ by Corollary 3.18 and Remark 3.21. □

This completes the proof of Theorem 3.1. □

Proof of Theorem 3.2. Let us review the statement of the theorem: assuming $\kappa = \omega_1$, or $\kappa = \lambda^+$ and $\Box_\lambda$, the partial order $\langle \mathcal{P}(\kappa), \subset_{\mathrm{NS}} \rangle$ can be embedded into $\langle \mathrm{E}^B_\kappa, \leq_B \rangle$.

If $\kappa = \omega_1$, then this is just the second part (a special case) of Corollary 3.7 on page 583 and follows from Theorem 3.1.

Recall Definition 3.16 on page 585. Let us see that if $\alpha < \kappa$, then $\Cup_{i<\alpha} 2^{\kappa \times \{i\}}$ is homeomorphic to 2^κ and so the domains of the forthcoming equivalence relations can be thought without loss of generality to be 2^κ. So fix $\alpha < \kappa$. For all $\beta{+}1 < \alpha$ let $\zeta_\beta \colon \beta{+}1 \to 2$ be the function $\zeta_\beta(\gamma) = 0$ for all $\gamma < \beta$ and $\zeta_\beta(\beta) = 1$, and let $\zeta_\alpha \colon \alpha \to 2$ be the constant function with value 0. Clearly $(\zeta_\beta)_{\beta \leq \alpha}$ is a maximal antichain. By rearranging the indexation we can assume that $(\zeta_\beta)_{\beta < \alpha}$ is a maximal antichain. If $\eta \in 2^{\kappa \times \{i\}}$, $i < \alpha$, let $\xi = \eta + i$ be the function with $\operatorname{dom} \xi = [i+1, \kappa)$ and $\xi(\gamma) = \eta(\mathrm{OTP}(\gamma \setminus i))$ and let

$$f(\eta) = \zeta_i {}^\frown (\eta + i).$$

Then f is a homeomorphism $\Cup_{i<\alpha} 2^{\kappa \times \{i\}} \to 2^\kappa$.

Assume $S \subset \kappa$ and let us construct the equivalence relation H_S. Denote for short $r = \operatorname{reg} \kappa$, the set of regular cardinals below κ. Since κ is not inaccessible, $|r| < \kappa$. Let $\{K_\mu \subset S^\kappa_\omega \mid \mu \in r\}$ be a partition of S^κ_ω into disjoint stationary sets. For each $\mu \in r \setminus \{\omega\}$, let $A_\mu = B^\mu_{\mathrm{nr}}(K_\mu)$ be the set given by Theorem 3.14. Additionally let $\{A^0_\omega, A^1_\omega, A^2_\omega, A^3_\omega\}$ be a partition of K_ω into disjoint stationary sets.

Let

$$\begin{aligned} H_S = {} & \big(\mathrm{id}_{2^\kappa} \otimes E_{A^2_\omega \cup ((S \cap S^\kappa_\omega) \setminus A^0_\omega)} \otimes E_{A^0_\omega}\big) \\ & \oplus \big(\mathrm{id}_{2^\kappa} \otimes E_{A^3_\omega \cup ((S \cap S^\kappa_\omega) \setminus A^1_\omega)} \otimes E_{A^1_\omega}\big) \\ & \oplus \bigoplus_{\mu \in r,\, \mu > \omega} \big(\mathrm{id}_{2^\kappa} \otimes E_{(S \cap S^\kappa_\mu) \cup A_\mu}\big). \end{aligned}$$

This might require a bit of explanation. H_S is a disjoint union of the equivalence relations listed in the equation. The final part of the equation lists all the relations obtained by splitting the set S into pieces of fixed uncountable cofinality and coupling them with the non-reflecting ω-stationary sets A_μ. The operation $E \mapsto \mathrm{id}_{2^\kappa} \otimes E$ is the same as the operation $E \mapsto \overline{E}$ in the proof of Theorem 3.1 above after the identification $2^{\omega \times \kappa} \approx 2^\kappa$. The first two lines of the equation deal with the ω-cofinal part of S. It is trickier, because the "coding sets" A_μ also consist of ω-cofinal ordinals. The way we have built up the relations makes it possible to use Theorem 3.17 to prove that $S \mapsto H_{\kappa \setminus S}$ is the desired embedding.

In order to make the sequel a bit more readable, let us denote

$$\begin{aligned} \mathfrak{B}^0_\omega(S) &= \big(\mathrm{id}_{2^\kappa} \otimes E_{A^2_\omega \cup ((S \cap S^\kappa_\omega) \setminus A^0_\omega)} \otimes E_{A^0_\omega}\big), \\ \mathfrak{B}^1_\omega(S) &= \big(\mathrm{id}_{2^\kappa} \otimes E_{A^3_\omega \cup ((S \cap S^\kappa_\omega) \setminus A^1_\omega)} \otimes E_{A^1_\omega}\big), \\ \mathfrak{B}_\mu(S) &= \big(\mathrm{id}_{2^\kappa} \otimes E_{(S \cap S^\kappa_\mu) \cup A_\mu}\big), \end{aligned}$$

for $\mu \in r \setminus \{\omega\}$. With this notation we have

$$H_S = \mathfrak{B}^0_\omega(S) \oplus \mathfrak{B}^1_\omega(S) \oplus \bigoplus_{\mu \in r,\, \mu > \omega} \mathfrak{B}_\mu(S).$$

Let us show that $S \mapsto H_{\kappa \setminus S}$ is an embedding from $\langle \mathcal{P}(\kappa), \subset_{\mathrm{NS}} \rangle$ into $\langle \mathrm{E}^B_\kappa, \leq_B \rangle$. Suppose $S_2 \setminus S_1$ is non-stationary. Then for each $\mu \in r \setminus \{\omega\}$ the set

$$\big((S^\kappa_\mu \cap S_2) \cup A_\mu\big) \setminus \big((S^\kappa_\mu \cap S_1) \cup A_\mu\big)$$

is non-stationary as well as are the sets

$$\big(A^2_\omega \cup ((S_2 \cap S^\kappa_\omega) \setminus A^0_\omega)\big) \setminus \big(A^2_\omega \cup ((S_1 \cap S^\kappa_\omega) \setminus A^0_\omega)\big)$$

and

$$\big(A^3_\omega \cup ((S_2 \cap S^\kappa_\omega) \setminus A^1_\omega)\big) \setminus \big(A^3_\omega \cup ((S_1 \cap S^\kappa_\omega) \setminus A^1_\omega)\big),$$

so by Claim 2 of the proof of Theorem 3.1 (page 590) we have for all $\mu \in r \setminus \{\omega\}$ that

$$(\mathrm{id}_{2^\kappa} \otimes E_{(S_1 \cap S^\kappa_\mu) \cup A_\mu}) \leq_B (\mathrm{id}_{2^\kappa} \otimes E_{(S_2 \cap S^\kappa_\mu) \cup A_\mu}),$$

$$(\mathrm{id}_{2^\kappa} \otimes E_{A^2_\omega \cup ((S_1 \cap S^\kappa_\omega) \setminus A^0_\omega)}) \leq_B (\mathrm{id}_{2^\kappa} \otimes E_{A^2_\omega \cup ((S_2 \cap S^\kappa_\omega) \setminus A^0_\omega)}),$$

and

$$(\mathrm{id}_{2^\kappa} \otimes E_{A^3_\omega \cup ((S_1 \cap S^\kappa_\omega) \setminus A^1_\omega)}) \leq_B (\mathrm{id}_{2^\kappa} \otimes E_{A^3_\omega \cup ((S_2 \cap S^\kappa_\omega) \setminus A^1_\omega)}).$$

Of course this implies that, for all $\mu \in r \setminus \{\omega\}$,

$$(\mathrm{id}_{2^\kappa} \otimes E_{A^2_\omega \cup ((S_1 \cap S^\kappa_\omega) \setminus A^0_\omega)} \otimes E_{A^0_\omega}) \leq_B (\mathrm{id}_{2^\kappa} \otimes E_{A^2_\omega \cup ((S_2 \cap S^\kappa_\omega) \setminus A^0_\omega)} \otimes E_{A^0_\omega})$$

and that

$$(\mathrm{id}_{2^\kappa} \otimes E_{A^3_\omega \cup ((S_1 \cap S^\kappa_\omega) \setminus A^1_\omega)} \otimes E_{A^1_\omega}) \leq_B (\mathrm{id}_{2^\kappa} \otimes E_{A^3_\omega \cup ((S_2 \cap S^\kappa_\omega) \setminus A^1_\omega)} \otimes E_{A^1_\omega}),$$

which precisely means that $\mathfrak{B}^0_\omega(S_1) \leq_B \mathfrak{B}^0_\omega(S_2)$, $\mathfrak{B}^1_\omega(S_1) \leq_B \mathfrak{B}^1_\omega(S_2)$ and $\mathfrak{B}_\mu(S_1) \leq_B \mathfrak{B}_\mu(S_2)$ for all $\mu \in r \setminus \{\omega\}$. Combining these reductions, we get a reduction from H_{S_1} to H_{S_2}.

Assume that $S_2 \setminus S_1$ is stationary. We want to show that $H_{S_1} \not\leq_B H_{S_2}$. Here H_{S_1} is a disjoint union of the equivalence relations $\mathfrak{B}^0_\omega(S_1)$, $\mathfrak{B}^1_\omega(S_1)$ and $\mathfrak{B}_\mu(S_1)$ for $\mu \in r \setminus \{\omega\}$. Let us call these equivalence relations the *building blocks* of H_{S_1}, and similarly for H_{S_2}.

Each building block of H_{S_1} can be easily reduced to H_{S_1} via inclusion, so it is sufficient to show that there is one block that cannot be reduced to H_{S_2}. We will show that if μ_1 is the least cardinal such that $S^\kappa_{\mu_1} \cap (S_2 \setminus S_1)$ is stationary, then

- that building block is $\mathfrak{B}_{\mu_1}(S_1)$ if $\mu_1 > \omega$;
- that building block is either $\mathfrak{B}^0_\omega(S_1)$ or $\mathfrak{B}^1_\omega(S_1)$ if $\mu_1 = \omega$.

Such a cardinal μ_1 exists because κ is not inaccessible and $|r| < \kappa$.

Suppose that f is a reduction from a building block of H_{S_1}, call it $\mathfrak{B}$, to H_{S_2}. H_{S_2} is a disjoint union of less than κ building blocks whose domains' inverse images decompose $\operatorname{dom} f$ into less than κ disjoint pieces and one of them, say C, is not meager. By the property of Baire one can find a basic open set U such that $C \cap U$ is co-meager in U. Let $C(f)$ be a co-meager set in which f is continuous. Now $f \restriction (U \cap C \cap C(f))$ is a continuous reduction from $\mathfrak{B}$ restricted to $(U \cap C \cap C(f))^2$ to a building block of H_{S_2}. Thus it is sufficient to show that this correctly chosen building block of H_{S_1} is not reducible to any of the building blocks of H_{S_2} on any such $U \cap C \cap C(f)$. This will follow from Theorem 3.17 and Remark 3.21 once we go through all the possible cases. So the following lemma concludes the proof.

Lemma 3.22 *Assume that $\mu_1 \in r$ is the least cardinal such that $(S_2 \setminus S_1) \cap S^\kappa_{\mu_1}$ is stationary. If $\mu_1 > \omega$, then*

(i) *for all $\mu_2 > \omega$, $\mathfrak{B}_{\mu_1}(S_1) \not\leq_B \mathfrak{B}_{\mu_2}(S_2)$,*

(ii) $\mathfrak{B}_{\mu_1}(S_1) \not\leq_B \mathfrak{B}^0_\omega(S_2)$,

(iii) $\mathfrak{B}_{\mu_1}(S_1) \not\leq_B \mathfrak{B}^1_\omega(S_2)$,

and if $\mu_1 = \omega$ then

(i*) *for all $\mu_2 > \omega$, $\mathfrak{B}^0_\omega(S_1) \not\leq_B \mathfrak{B}_{\mu_2}(S_2)$,* (ii*) *for all* $\mu_2 > \omega$, $\mathfrak{B}^1_\omega(S_1) \not\leq_B \mathfrak{B}_{\mu_2}(S_2)$,

(iii*) *either*

$$\mathfrak{B}^0_\omega(S_1) \not\leq_B \mathfrak{B}^0_\omega(S_2) \textit{ and } \mathfrak{B}^0_\omega(S_1) \not\leq_B \mathfrak{B}^1_\omega(S_2) \tag{3.1}$$

or

$$\mathfrak{B}^1_\omega(S_1) \not\leq_B \mathfrak{B}^0_\omega(S_2) \textit{ and } \mathfrak{B}^1_\omega(S_1) \not\leq_B \mathfrak{B}^1_\omega(S_2). \tag{3.2}$$

Proof of the Lemma. First we assume $\mu_1 > \omega$.

(i) There are two cases:

Case 1: $\mu_2 = \mu_1$. Write $B = A_{\mu_1} = A_{\mu_2}$ and $S'_1 = (S_1 \cap S^\kappa_{\mu_1}) \cup B$ and $S'_2 = (S_2 \cap S^\kappa_{\mu_2}) \cup B$. Now $\mathfrak{B}_{\mu_1}(S_1) = \mathrm{id} \otimes E_{S'_1}$, $\mathfrak{B}_{\mu_2}(S_2) = \mathrm{id} \otimes E_{S'_2}$. Since, by definition, $B = B^\mu_{\mathrm{nr}}(K_\mu)$ where $K_\mu \subset S^\kappa_\omega$ is stationary, and $(S_2 \setminus S_1) \cap S^\kappa_{\mu_1}$ is stationary, the sets S'_1 and S'_2 satisfy the assumptions of Theorem 3.17(2b), so the statement follows from Theorem 3.17(2b) and Remark 3.21.

Case 2: $\mu_2 \neq \mu_1$. Let $S'_1 = (S_1 \cap S^\kappa_{\mu_1}) \cup A_{\mu_1}$ and $S'_2 = (S_2 \cap S^\kappa_{\mu_2}) \cup A_{\mu_2}$ whence $B_{\mu_1}(S_1) = \mathrm{id} \otimes E_{S'_1}$ and $B_{\mu_2}(S_2) = \mathrm{id} \otimes E_{S'_2}$. Now $S'_1 \subset S^\kappa_{\geq\omega}$ and $S'_2 \subset S^\kappa_{\geq\omega}$ and since $A_{\mu_1} \cap A_{\mu_2} = \varnothing$, the result follows from Theorem 3.17(1b) and Remark 3.21.

(ii) Let $S'_1 = (S_1 \cap S^\kappa_{\mu_1}) \cup A_{\mu_1}$, $S'_2 = A^2_\omega \cup ((S_2 \cap S^\kappa_\omega) \setminus A^0_\omega)$, and $A'_2 = A^0_\omega$. By definition,

$$B^0_\omega(S_2) = \mathrm{id}_{2^\kappa} \otimes E_{S'_2} \otimes E_{A'_2}$$

and $B_{\mu_1}(S_1) = E_{S'_1}$. Since $A_{\mu_1} \cap A^2_\omega = \varnothing$, $S'_1 \cap S^\kappa_\omega = A_{\mu_1}$ and $A^2_\omega \subset S'_2$, we have that $S'_2 \setminus S'_1$ is ω-stationary, because it contains A^2_ω. Also $A^0_\omega \setminus S'_1 = A^0_\omega$, because $S'_1 \cap A^0_\omega = \varnothing$, so $A'_2 \setminus S'_1$ is ω-stationary. Now the result follows from Theorem 3.17(4b) and Remark 3.21.

(iii) Similar to (ii).

Then we assume $\mu_1 = \omega$.

(i*) Let $S'_1 = A^2_\omega \cup ((S_1 \cap S^\kappa_\omega) \setminus A^0_\omega)$, $A'_1 = A^0_\omega$ $A'_2 = A_{\mu_2}$ and $S'_2 = (S_2 \cap S^\kappa_{\mu_2})$. Since $A^0_\omega \cap A_{\mu_2} = \varnothing$, we have that $A'_2 \setminus A'_1$ is ω-stationary, so by Theorem 3.17(5) and Remark 3.21,

$$\mathrm{id} \otimes E_{S'_1} \otimes E_{A'_1} \not\leq_B \mathrm{id} \otimes E_{S'_2 \cup A'_2},$$

which by definitions is exactly the subject of the proof.

(ii*) Similar to (i*).

(iii*) The situation is split into two cases, the latter of which is split into two subcases:

Case 1: $((S_2 \setminus S_1) \cap S^\kappa_\omega) \setminus (A^2_\omega \cup A^0_\omega)$ *is stationary.* Let $S'_1 = A^2_\omega \cup ((S_1 \cap S^\kappa_\omega) \setminus A^0_\omega)$, $A'_1 = A^0_\omega$, $S'_2 = A^2_\omega \cup ((S_2 \cap S^\kappa_\omega) \setminus A^0_\omega)$ and $A'_2 = A^0_\omega$. Now $A'_2 \setminus S'_1$ is obviously ω-stationary, since it is equal to A^0_ω. Also $S'_2 \setminus S'_1$ is stationary, because it equals to $((S_2 \setminus S_1) \cap S^\kappa_\omega) \setminus (A^2_\omega \cup A^0_\omega)$ which is stationary by the assumption. Now the first part of (1) follows from Theorem 3.17(3b) and Remark 3.21, because $\mathfrak{B}^0_\omega(S_1) = \mathrm{id} \otimes E_{S'_1} \otimes E_{A'_1}$ and $\mathfrak{B}^0_\omega(S_2) = \mathrm{id} \otimes E_{S'_2} \otimes E_{A'_2}$. On the other hand let $S''_2 = A^3_\omega \cup ((S_2 \cap S^\kappa_\omega) \setminus A^1_\omega)$ and $A''_2 = A^1_\omega$. Now $S''_2 \setminus A'_1$ is stationary, because $A^3_\omega \subset S''_2$ but $A^3_\omega \cap A'_1 = A^3_\omega \cap A^0_\omega = \varnothing$. Also $A''_2 \setminus A'_1$ is stationary since $A''_2 \cap A'_1 = A^1_\omega \cap A^0_\omega = \varnothing$. Now also the second part of (1) follows from Theorem 3.17(3b) and Remark 3.21, because $B^0_1(S_1) = \mathrm{id} \otimes E_{S'_1} \otimes E_{A'_1}$ and $B^1_1(S_2) = \mathrm{id} \otimes E_{S''_2} \otimes E_{A''_2}$.

Case 2: $((S_2 \setminus S_1) \cap S^\kappa_\omega) \setminus (A^2_\omega \cup A^0_\omega)$ *is non-stationary.*

Case 2a: $((S_2 \setminus S_1) \cap S^\kappa_\omega) \setminus (A^3_\omega \cup A^1_\omega)$ *is stationary.* Now (2) follows from Theorem 3.17(3b) and Remark 3.21 in a similar way as (1) followed in Case 1.

Case 2b: $((S_2 \setminus S_1) \cap S^\kappa_\omega) \setminus (A^3_\omega \cup A^1_\omega)$ *is non-stationary.* We have both that

$$((S_2 \setminus S_1) \cap S^\kappa_\omega) \setminus (A^2_\omega \cup A^0_\omega) \text{ is non-stationary} \tag{3.3}$$

and that

$$((S_2 \setminus S_1) \cap S^\kappa_\omega) \setminus (A^3_\omega \cup A^1_\omega) \text{ is non-stationary.} \tag{3.4}$$

Now from (3.3) it follows that $S_2 \setminus S_1 \subset_{\mathrm{NS}(\omega)} A^2_\omega \cup A^0_\omega$, and from (3.4) it follows that $S_2 \setminus S_1 \subset_{\mathrm{NS}(\omega)} A^3_\omega \cup A^1_\omega$. This is a contradiction, because $S_2 \setminus S_1$ is ω-stationary and $(A^2_\omega \cup A^0_\omega) \cap (A^3_\omega \cup A^1_\omega) = \varnothing$. □

This completes the proof of Theorem 3.2. □

4 On chains in $\langle \mathrm{E}^B_\kappa, \leq_B \rangle$

There are chains of order type κ^+ in Borel equivalence relation on 2^κ:

Theorem 4.1 *Let $\kappa > \omega$. There are equivalence relations $R_i \in \mathrm{E}^B_\kappa$, for $i < \kappa^+$, such that $i < j \Leftrightarrow R_i \lneq_B R_j \lneq E_0$.*

Remark 4.2 In many cases there are κ^+-long chains in the power set of κ ordered by inclusion modulo the non-stationary ideal whence a weak version of this theorem could be proved using Theorem 3.2. Namely if the ideal I^κ_{NS} of non-stationary subsets of κ is *not* κ^+-saturated, then there are κ^+-long chains. In this case being *not* κ^+-saturation means that there exists a sequence $\langle A_i \mid i < \kappa^+ \rangle$ of subsets of κ such that A_i is stationary for all i but $A_i \cap A_j$ is non-stationary for all $i \neq j$. Now let f_α be a bijection from κ to α for all $\alpha < \kappa^+$ and let

$$B_\alpha = \mathop{\triangledown}_{i<\alpha} A_i = \big\{\alpha \mid \text{for some } i < \alpha,\ \alpha \in A_{f_\alpha(i)}\big\}.$$

It is not difficult to see that $\langle B_\alpha \mid \alpha < \kappa^+ \rangle$ is a chain. On the other hand, the existence of such a chain implies that I^κ_{NS} is not κ^+-saturated.

By a theorem of Gitik and Shelah [**11**, Theorem 23.17], I^κ_{NS} is not κ^+-saturated for all $\kappa \geq \aleph_2$. By a result of Shelah [**11**, Theorem 38.1], it is consistent relative to the consistency of a Woodin cardinal that $I^{\aleph_1}_{\mathrm{NS}}$ *is* $\aleph_2$-saturated in which case there are no chains of length ω_2 in $\langle \mathcal{P}(\omega_1), \subset_{\mathrm{NS}} \rangle$. On the other hand, in the model provided by Shelah, CH fails. According to Jech [**3**] it is an open question whether CH implies that $I^{\aleph_1}_{\mathrm{NS}}$ is not $\aleph_2$-saturated.

However, as the following shows, it follows from ZFC that there are κ^+-long chains in $\langle \mathrm{E}^B_\kappa, \leq_B \rangle$ for any uncountable κ.

Proof of Theorem 4.1. By the proof of Corollary 3.19, page 588, one can find ω-stationary sets S_i for $i < \kappa^+$ such that $S_i \setminus S_j$ and $S_j \setminus S_i$ are stationary whenever $i \neq j$. For all $j \in [1, \kappa^+)$, let

$$R_j = \bigoplus_{i<j} E_{S_i},$$

where the operation $\oplus$ is from Definition 3.16, page 585.

Let us denote $P_A = \Cup_{i \in A} 2^{\kappa \times \{i\}}$ for $A \subset \kappa^+$, i.e., for example $P_j = \Cup_{i<j} 2^{\kappa \times \{i\}}$. Let us show that

(1) if $i < j$, then $R_i \leq_B R_j$;

(2) if $i < j$, then $R_j \not\leq_B R_i$;

(3) for all $i < \kappa^+$, $R_i \lneqq_B E_0$.

Item (1) is simple: let $f\colon P_i \to P_j$ be the inclusion map (as $P_i \subset P_j$). Then f is clearly a reduction from R_i to R_j.

Suppose then that $i < j$ and that $i \leq k < j$. To prove (2) it is sufficient to show that there is no reduction from E_{S_k} to R_j. Let us assume that $f\colon 2^\kappa \to P_j$ is a Borel reduction from E_{S_k} to R_j. Now

$$2^\kappa = \Cup_{\alpha<i} f^{-1}[P_{\{\alpha\}}],$$

so one of the sets $f^{-1}[P_{\{\alpha\}}]$ is not meager; let α_0 be an index witnessing this. Note that $\alpha_0 < k$, because $\alpha_0 < i \leq k$. Because f is a Borel function and Borel sets have the property of Baire, we can find a $p \in 2^{<\kappa}$ such that $C = N_p \cap C(f) \cap f^{-1}[P_{\{j\}}]$ is co-meager in N_p. But now $f \restriction C$ is a continuous reduction from $E_{S_k} \cap C^2$ to E_{S_α} which contradicts Theorem 3.17(1b).

To prove (3) we will show first that $R_i \leq_B \bigoplus_{j<i} E_0$ and then that $\bigoplus_{j<i} E_0 \leq_B E_0$, after which we will show that $E_0 \not\leq_B R_i$ for all i.

Let f_j be a reduction from E_{S_j} to E_0 for all $j < i$ given by Claim 3 of the proof of Theorem 3.1. Then combine these reductions to get a reduction from R_i to $\bigoplus_{j<i} E_0$. To be more precise, for each $\eta \in P_{\{j\}}$ let $f(\eta)$ be ξ such that $\xi \in P_{\{j\}}$ and $\xi = f_j(\eta)$.

Let $\{A_k \mid k \leq i\}$ be a partition of κ into disjoint unbounded sets. Let $\eta \in P_i$. By definition, $\eta \in P_{\{k\}}$ for some $k < i$. Define $\xi = F(\eta)$ as follows. Let $f\colon A_i \to \kappa$ be a bijection.

- If $\alpha \in A_i$, then let $\xi(\alpha) = \eta(f(\alpha))$.
- If $\alpha \in A_j$ and $j \neq k$, then let $\xi(\alpha) = 0$.
- If $\alpha \in A_k$, then let $\xi(\alpha) = 1$.

It is easy to see that F is a continuous reduction.

Assume for a contradiction that $E_0 \leq_B R_i$ for some $i < \kappa^+$. Then by (1) and transitivity, $E_0 \leq_B R_j$ for all $j \in [i, \kappa^+)$. By the above also $R_j \leq_B E_0$ for all $j \in [i, \kappa^+)$ which, again by transitivity, implies that the relations R_j for $j \in [i, \kappa^+)$ are mutually bireducible to each other, which contradicts (2). □

Acknowledgements

This work is part of my Ph.D. thesis, defended in Fall 2011. I wish to express my gratitude to my supervisor Tapani Hyttinen for his careful attention towards this work, his valuable advice and some technical help regarding the proof of Theorem 3.2.

I am indebted to the Finnish National Graduate School in Mathematics and its Applications for supporting my post-graduate studies during the preparation of this article.

This work is part of the project *The Myriad Aspects of Infinity* supported by the John Templeton Foundation (Grant #13152).

References

[1] S. Adams, A. S. Kechris: *Linear algebraic groups and countable Borel equivalence relations*, J. Amer. Math. Soc. 13 (2000), 909–943.

[2] M. Burke and M. Magidor: *Shelah's pcf theory and its applications*, Ann. Pure Appl. Logic, 1990.

[3] *Handbook of Set Theory*, Foreman, Matthew and Kanamori, Akihiro (eds.), 1st edition, 2010.

[4] S. D. Friedman, T. Hyttinen: *On Borel equivalence relations in the generalized Baire space*, submitted to Arch. Math. Logic.

[5] S. D. Friedman, T. Hyttinen, V. Kulikov: *Generalized descriptive set theory and classification theory*, CRM preprint no. 999, 2011.

[6] J. Gregory: *Higher Souslin trees and the generalized continuum hypothesis*, J. Symbolic Logic 41, no. 3 (1976), 663–671.

[7] A. Halko: *Negligible subsets of the generalized Baire space* $\omega_1^{\omega_1}$, Ann. Acad. Sci. Fenn. Ser. A Diss. Math. 108 (1996).

[8] T. Huuskonen, T. Hyttinen, M. Rautila: *On the* κ*-cub game on* λ *and* $I[\lambda]$, Arch. Math. Logic 38 (1999), 589–557.

[9] T. Hyttinen: Games and infinitary languages, Ann. Acad. Sci. Fenn. Ser. A Diss. Math. 64 (1987), 1–32.

[10] T. Hyttinen, S. Shelah and H. Tuuri: Remarks on strong nonstructure theorems, Notre Dame J. Formal Logic 34, no. 2 (1993), 157–168.

[11] T. Jech: *Set Theory*, Springer-Verlag, Berlin Heidelberg New York, 2003.

[12] A. Louveau, B. Velickovic: *A note on Borel equivalence relations*, Proc. Amer. Math. Soc. 120, no. 1 (1994), 255–259.

[13] A. Mekler, J. Väänänen: *Trees and* Π^1_1*-subsets of* ${}^{\omega_1}\omega_1$, J. Symbolic Logic 58, no. 3 (1993), 1052–1070.

[14] S. Shelah: *Reflecting stationary sets and successors of singular cardinals*, Arch. Math. Logic 31 (1991), 25–53.

[15] S. Shelah: *Diamonds*, Proc. Amer. Math. Soc. 138 (2010), 2151–2161.

On ultrafilter extensions of models

Denis I. Saveliev[†]

[†] Department of Mathematical Logic and Theory of Algorithms, Faculty of Mechanics and Mathematics, Moscow M. V. Lomonosov State University, Russia
d.i.saveliev@gmail.com

Abstract. We show that any model $\mathfrak{A}$ can be extended, in a canonical way, to a model $\beta\mathfrak{A}$ consisting of ultrafilters over it. The extension preserves relationships between models: any homomorphism of $\mathfrak{A}$ into $\mathfrak{B}$ extends to a continuous homomorphism of $\beta\mathfrak{A}$ into $\beta\mathfrak{B}$. Moreover, if a model $\mathfrak{C}$ carries a compact Hausdorff topology compatible with its structure, then any homomorphism of $\mathfrak{A}$ into $\mathfrak{C}$ extends to a continuous homomorphism of $\beta\mathfrak{A}$ into $\mathfrak{C}$. This is also true for embeddings.

Introduction

In this paper, we present a new area in general model theory. Following [**13**], we show that an arbitrary first-order model can be extended, in a canonical way, to the model (of the same language) consisting of all ultrafilters over it, with a model-theoretic behavior similar to the topological behavior of the largest compactification.

Recall standard facts concerning topology of ultrafilters (see [**5, 7, 11**]). The set βX of ultrafilters over a set X carries a natural topology generated by elementary (cl)open sets of the form

$$\tilde{S} = \{u \in \beta X : S \in u\}$$

for all $S \subseteq X$. The space βX is compact Hausdorff, extremally disconnected (the closure of any open set is open), and it is, in fact, the Stone–Čech (and also Wallman) compactification of the discrete space X, i.e., its *largest* compactification. This means that X is dense in βX (one lets $X \subseteq \beta X$ by identifying each $x \in X$ with the principal ultrafilter $\hat{x}$), and any continuous mapping h of X into any compact space Y can be uniquely extended to a continuous mapping $\tilde{h}$ of βX into Y. There exists a one-to-one correspondence between filters over X and closed subsets of βX (a filter D corresponds to $\{u \in \beta X : D \subseteq u\}$ while a closed $C \subseteq \beta X$ corresponds to $\bigcap C$). The cardinality of βX is $2^{2^{|X|}}$ for all infinite X. Some of these facts require a dose of AC. Actually, the compactness of βX is equivalent to PI (the claim that $\{u \in \beta X : D \subseteq u\}$ is nonempty for each filter D) and can partly recover the cardinality evaluation (see [**15**]). PI is weaker than AC but stronger than the existence of non-principal ultrafilters; the latter is still unprovable in ZF alone (see [**1**]).

Here we deal with the case when X is endowed with a first-order structure, i.e., there are some operations $F, \ldots$ and relations $P, \ldots$ on X. In Section 1 we describe a canonical way to extend them to operations $\tilde{F}, \ldots$ and relations $\tilde{P}, \ldots$ on βX, thus extending the model $\mathfrak{A} = (X, F, \ldots, P, \ldots)$ to the model $\beta\mathfrak{A} = (\beta X, \tilde{F}, \ldots, \tilde{P}, \ldots)$. In Section 2 we prove

This paper is a slightly revised and expanded version of [**13**].

[†]Partially supported by INFTY grant of ESF, RFBR grants 11-01-00958-a and 11-01-93107-a, and grant NSh-5593.2012.1. The author thanks the John Templeton Foundation for its support through Project #13152, Myriad Aspects of Infinity.

the First Extension Theorem, showing that the extension procedure preserves model-theoretic relationships: if h is a homomorphism of $\mathfrak{A}$ into $\mathfrak{B}$, then $\tilde{h}$ is a homomorphism of $\beta\mathfrak{A}$ into $\beta\mathfrak{B}$, and the same holds for embeddings and some other relationships between models. In Section 3 we study topological properties of extended models. In Section 4 we prove the Second Extension Theorem: if $\mathfrak{C}$ carries a compact Hausdorff topology, its first-order structure has the same topological properties as ultrafilter extensions, and h is a homomorphism of $\mathfrak{A}$ into $\mathfrak{C}$, then $\tilde{h}$ is a homomorphism of $\beta\mathfrak{A}$ into $\mathfrak{C}$, and similarly for embeddings etc. This shows that the construction provides a right generalization of the Stone–Čech (or Wallman) compactification of a discrete space X to the case when X is endowed with a first-order structure. Finally, in Section 5 we mark directions for further investigations, fix several problems, and mention a few results without proofs.

Historical remarks The largest compactification of Tychonoff spaces was discovered independently by Čech and Stone in 1937; in a year, Wallman did the same for T_1 spaces (by using ultrafilters on lattices of closed sets); see [**7**] for more information. Our ultrafilter extension procedure extends unary relations to elementary open sets, and unary mappings to continuous mappings; thus, in the unary casef it gives classical concepts known in the 30s. As for mappings and relations of greater arities, several instances of their ultrafilter extensions were discovered only in the 60s. We isolate three areas where such instances arose.

The first area concerns *iterated ultrapowers*. Frayne, Morel, and Scott [**8**] showed that finite iteration of ultrapowers gives ultrapowers by using (in our terms) ultrafilter extensions of taking n-tuples. The general construction of iterated ultrapowers, invented by Gaifman and elaborated by Kunen, has become standard in model theory and set theory (see [**4, 12**]).

The second area concerns *ultrafilter extensions of semigroups*. Such structures appeared in the 60s as subspaces of function spaces; the first explicit construction of the ultrafilter extension of a group is due to Ellis [**6**]. In the 70s Galvin and Glazer applied them to give an easy proof of Hindman's Finite Sums Theorem; the key idea was to use idempotent ultrafilters. The method was used then by Hindman, van Douwen, Blass, Protasov, and many others, and gave numerous applications of Ramsey theory in number theory, algebra, topological dynamics, and ergodic theory. A partial case of the Second Extension Theorem in which models are semigroups appeared in [**3**]. The book [**11**] is a comprehensive treatise on this area, with an historical information.

The third area concerns *modal logic*. Characterizing modal definability, van Benthem [**16**] extended binary relations of frames to ultrafilters. Goldblatt [**9**] and then Goranko [**10**] generalized this construction to relations of arbitrary arity. Their extensions coincide with our extensions only for unary relations; in general, our extensions are smaller. Goranko defines there also extensions of operations; moreover, he extends models to arbitrary *filters*. However his filter extension of operations does not work for ultrafilters, and he defines this case separately, in the same way as here. Goranko proves a theorem analogous to the First Extension Theorem —they coincide for ultrafilter extensions of operations.

It seems that the three areas have a very little knowledge about each other, if any.

Our construction of ultrafilter extensions of models, together with the mentioned basic results, has appeared in [**13**]; here we present a slightly revised and expanded version.

1 Basic definitions

Here we describe our extensions of models by ultrafilters.

To extend a model $\mathfrak{A} = (X, F, \ldots, P, \ldots)$, we extend operations $F, \ldots$ on X, i.e., mappings of Cartesian products of X into X itself, and relations $P, \ldots$ on X, i.e., subsets of such products. Let us provide a slightly more general definition involving n-ary mappings of $X_1 \times \cdots \times X_n$ into Y, and n-ary relations that are subsets of $X_1 \times \cdots \times X_n$. We shall use it, in particular, when we shall show that any mapping h of a certain kind between models (e.g., a homomorphism) extends to $\tilde{h}$ of the same kind.

We start by defining ultrafilter extensions of mappings.

Definition 1.1 For an n-ary map $F\colon X_1 \times \cdots \times X_n \to Y$, let $\tilde{F}\colon \beta X_1 \times \cdots \times \beta X_n \to \beta Y$ be defined as follows:

$$\begin{aligned}&\tilde{F}(u_1, \ldots, u_n) = \\ &\qquad \bigl\{S \subseteq Y : \{x_1 \in X_1 : \ldots \{x_n \in X_n : F(x_1, \ldots, x_n) \in S\} \in u_n \ldots\} \in u_1\bigr\}\end{aligned}$$

for all $u_1 \in \beta X_1, \ldots, u_n \in \beta X_n$.

Lemma 1.2 *For all $z_1 \in X_1$ and $u_2 \in \beta X_2, \ldots, u_n \in \beta X_n$,*

$$\begin{aligned}&\bigl\{x_1 : \bigl\{x_2 : \ldots \{x_n : F(x_1, x_2, \ldots, x_n) \in S\} \in u_n \ldots\bigr\} \in u_2\bigr\} \in \hat{z}_1 \\ &\textit{iff } \bigl\{x_2 : \ldots \{x_n : F(z_1, x_2, \ldots, x_n) \in S\} \in u_n \ldots\bigr\} \in u_2.\end{aligned}$$

Proof. Clear. □

Proposition 1.3 *If $F\colon X_1 \times \cdots \times X_n \to Y$, then $\tilde{F}\colon \beta X_1 \times \cdots \times \beta X_n \to \beta Y$. Moreover, the restriction of $\tilde{F}$ on* $\mathrm{dom}(F)$ *is F.*

Proof. By definition, $\mathrm{dom}(\tilde{F}) = \beta X_1 \times \cdots \times \beta X_n$, and a standard argument shows that values of $\tilde{F}$ are ultrafilters. It follows from Lemma 1.2 that for all $z_1 \in X_1, \ldots, z_n \in X_n$,

$$\bigl\{x_1 : \ldots \{x_n : F(x_1, \ldots, x_n) \in S\} \in \hat{z}_n \ldots\bigr\} \in \hat{z}_1 \ \leftrightarrow\ F(z_1, \ldots, z_n) \in S.$$

Therefore,

$$\tilde{F}(\hat{z}_1, \ldots, \hat{z}_n) = \hat{y} \ \text{ whenever } \ F(z_1, \ldots, z_n) = y,$$

and thus $\tilde{F}$ extends F up to identification of x and $\hat{x}$. □

Let us comment on the construction. First, in the unary case, an $F\colon X \to Y$ extends to $\tilde{F}\colon \beta X \to \beta Y$ by

$$\tilde{F}(u) = \bigl\{S \subseteq Y : \{x \in X : F(x) \in S\} \in u\bigr\}.$$

This gives the standard unique continuous extension of F. Indeed, it is easy to see that $\tilde{F}$ is continuous, and continuous extensions agreeing on a dense subset coincide.

Next, consider the binary case. $F\colon X_1 \times X_2 \to Y$ extends to $\tilde{F}\colon \beta X_1 \times \beta X_2 \to \beta Y$ by

$$\tilde{F}(u_1, u_2) = \bigl\{S \subseteq Y : \{x_1 \in X_1 : \{x_2 \in X_2 : F(x_1, x_2) \in S\} \in u_2\} \in u_1\bigr\}.$$

This can be considered as the extension fulfilled in two steps: first one extends left translations, then right ones. In the extended F, all right translations are continuous; in terms of [**11**], the groupoid $(\beta X, \tilde{F})$ is *right topological*. Moreover, all left translations by *principal* ultrafilters are continuous, and such an extension is unique.

The extensions of mappings of arbitrary arity have analogous topological properties: If $F\colon X_1 \times \dots \times X_n \to Y$ and $1 \le i \le n$, then for all $x_1 \in X_1, \dots, x_{i-1} \in X_{i-1}$ and $u_{i+1} \in \beta X_{i+1}, \dots, u_n \in \beta X_n$, the mapping

$$u \longmapsto \tilde{F}(\hat{x}_1, \dots, \hat{x}_{i-1}, u, u_{i+1}, \dots, u_n)$$

of βX_i into βY is continuous. Moreover, $\tilde{F}$ is a unique such extension of F. A proof of this fact will be given in the next section (Lemma 3.3).

Now we define ultrafilter extensions of relations.

Definition 1.4 Given $P \subseteq X_1 \times \dots \times X_n$, let $\tilde{P}$ be defined as follows:

$$\langle u_1, \dots, u_n \rangle \in \tilde{P} \text{ iff } \big\{x_1 \in X_1 : \dots \{x_n \in X_n : \langle x_1, \dots, x_n \rangle \in P\} \in u_n \dots \big\} \in u_1,$$

for all $u_1 \in \beta X_1, \dots, u_n \in \beta X_n$.

Proposition 1.5 *If $P \subseteq X_1 \times \dots \times X_n$, then $\tilde{P} \subseteq \beta X_1 \times \dots \times \beta X_n$. Moreover, the intersection $\tilde{P} \cap (X_1 \times \dots \times X_n)$ is equal to P.*

Proof. This is true by Lemma 1.2. □

Let us comment on the construction. If P is a unary relation on X, $P \subseteq X$, one has

$$u \in \tilde{P} \text{ iff } P \in u.$$

(The definition involves n-tuples; a 1-tuple $\langle x \rangle$ is just x.) Thus $\tilde{P}$ is an elementary open set of βX; the extensions of all unary relations on X form the standard open basis of the topology of βX. As we noted, the $\tilde{P}$ are in fact clopen.

If P is a binary relation, $P \subseteq X_1 \times X_2$, one has

$$\langle u_1, u_2 \rangle \in \tilde{P} \text{ iff } \big\{x_1 \in X_1 : \{x_2 \in X_2 : \langle x_1, x_2 \rangle \in P\} \in u_2 \big\} \in u_1.$$

There is an easier way to say the same. Let $\langle\,\rangle\tilde{}$ denote the extension of the pairing function $\langle\,\rangle$ (cf. Definition 11.1 in **[11]** —there $\langle\,\rangle\tilde{}$ is denoted by $\otimes$ and refered as a "tensor product"; another name that is used is a "Fubini product"). Then

$$\langle u_1, u_2 \rangle \in \tilde{P} \text{ iff } P \in \langle u_1, u_2 \rangle\tilde{}.$$

This formula displays a similarity to the formula with unary P explicitly.

As for topological properties of extended binary relations, it is easy to see that for any $x_1 \in X_1$ and $u_2 \in \beta X_2$, the set $\{u_1 \in \beta X_1 : \langle u_1, u_2 \rangle \in \tilde{P}\}$ is clopen in βX_1, and the set $\{u_2 \in \beta X_2 : \langle \hat{x}_1, u_2 \rangle \in \tilde{P}\}$ is clopen in βX_2.

Likewise, if $\langle\,\rangle\tilde{}$ denotes the extension of taking n-tuples, one gets the following redefinition:

Proposition 1.6 *Let $P \subseteq X_1 \times \dots \times X_n$. Then for all $u_1 \in \beta X_1, \dots, u_n \in \beta X_n$,*

$$\langle u_1, \dots, u_n \rangle \in \tilde{P} \textit{ iff } P \in \langle u_1, \dots, u_n \rangle\tilde{}.$$

Proof. Clear. □

The extensions of relations of arbitrary arity have analogous topological properties: If $P \subseteq X_1 \times \dots \times X_n$ and $1 \le i \le n$, then for every $x_1 \in X_1, \dots, x_{i-1} \in X_{i-1}$ and $u_{i+1} \in \beta X_{i+1}, \dots, u_n \in \beta X_n$, the subset

$$\{u \in \beta X_i : \langle \hat{x}_1, \dots, \hat{x}_{i-1}, u, u_{i+1}, \dots, u_n \rangle \in \tilde{P}\}$$

of βX_i is clopen. A proof of this fact is also postponed to the next section (Lemma 3.7).

Now we are ready to define ultrafilter extensions of arbitrary models.

Definition 1.7 Given a model $\mathfrak{A} = (X, F, \ldots, P, \ldots)$, let $\beta\mathfrak{A}$ denote the extended model $(\beta X, \tilde{F}, \ldots, \tilde{P}, \ldots)$, called the *ultrafilter extension* of $\mathfrak{A}$.

Remarks

(1) Our use of the symbol $\tilde{}$ is ambiguous in two respects. First, the same relation or function can have distinct extensions depends on its implicit arity. Say, let $P \subseteq X \times X$. If P is regarded as a binary relation on X, then $\tilde{P}$ is a binary relation on βX:

$$\tilde{P} \subseteq \beta X \times \beta X.$$

If P is considered as a unary relation on $X \times X$, then $\tilde{P}$ is a unary relation on $\beta(X \times X)$:

$$\tilde{P} \subseteq \beta(X \times X).$$

Note that for discrete X the spaces $\beta(X \times X)$ and $\beta X \times \beta X$ are not homeomorphic. Similarly for extensions of mappings. Second, the same set can have distinct extensions when regarded as a function or as a relation. Say, let P be a binary relation that is a function, and let F_P denote this unary function. If F_P is an injection, then $\tilde{P}$ and $\tilde{F}_P$ do not coincide: $\tilde{P} = P$, while $\tilde{F}_P \neq F_P$ whenever $\beta X \neq X$ (as $\mathrm{dom}(F_P) = \beta X$). Nevertheless, the context usually leaves no doubts, and so we adopt the notation.

(2) The example above suggests to ask about relations that coincide with their extensions. Actually, this example near characterizes them. Let us say that a relation P is *almost injective* iff for any i and all fixed $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$, the set

$$P_{x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n} = \{x_i : \langle x_1, \ldots, x_n \rangle \in P\}$$

is finite. Note that a unary relation is almost injective iff it is finite. Then it can be shown that $\tilde{P} = P$ iff P is almost injective. The 'only if' part assumes that any infinite set carries a non-principal ultrafilter, which is weaker than the compactness of βX but still inprovable in ZF. (The fact can be restated in ZF alone if we redefine almost injective relations by replacing "is finite" with "carries no ultrafilter"; in ZFC the definitions coincide, while in models without ultrafilters all relations are almost injective.)

(3) Proposition 1.6 shows that extensions of n-ary relations can be defined via the extension of taking n-tuples. Let us note that the latter extensions can be defined iteratively via the extension of the pairing function only. Moreover, extensions of an n-ary mapping F can be decomposed into a combination of two extensions: of taking n-tuples and of F *regarded as a unary mapping*, i.e., with the domain $\beta(X \times \cdots \times X)$. Thus the whole construction of ultrafilter extensions can be reduced to two simple basic cases: extensions of unary mappings and the extension of pairing.

2 The First Extension Theorem

In this section, we establish our first main result showing that the extension procedure preserves homomorphisms and some other model-theoretic relationships (Theorem 2.7).

The following lemma is crucial; it states that extensions of compositions are compositions of extensions (e.g., $(G \circ h)\tilde{} = \tilde{G} \circ \tilde{h}$ if G and h are unary).

Lemma 2.1 *Let $h_1\colon X_1 \to Y_1, \dots, h_n\colon X_n \to Y_n$, and $G\colon Y_1 \times \dots \times Y_n \to Z$. For all $S \subseteq Z$ and $u_1 \in \beta X_1, \dots, u_n \in \beta X_n$, the following are equivalent:*

(a) $S \in \tilde{G}(\tilde{h}_1(u_1), \dots, \tilde{h}_n(u_n))$;

(b) $\big\{y_1 \in Y_1 : \dots \{y_n \in Y_n : G(y_1, \dots, y_n) \in S\} \in \tilde{h}_n(u_n) \dots\big\} \in \tilde{h}_1(u_1)$;

(c) $\big\{x_1 \in X_1 : \dots \{x_n \in X_n : G(h_1(x_1), \dots, h_n(x_n)) \in S\} \in u_n \dots\big\} \in u_1$.

Proof. The first and the second formulas are equivalent by definition of $\tilde{F}$.

That the second and the third formulas are equivalent can be proved by a straightforward induction on n. First one gets

$$\begin{aligned}&\big\{y_1 : \dots \{y_n : G(y_1, \dots, y_n) \in S\} \in \tilde{h}_1(u_n) \dots\big\} \in \tilde{h}_1(u_1)\\ &\text{iff } \big\{x_1 : h_1(x_1) \in \{y_1 : \dots \{y_n : G(y_1, \dots, y_n) \in S\} \in \tilde{h}_n(u_n) \dots\}\big\} \in u_1\\ &\text{iff } \big\{x_1 : \{y_2 : \dots \{y_n : G(h_1(x_1), y_2, \dots, y_n) \in S\} \in \tilde{h}_n(u_n) \dots\} \in \tilde{h}_2(u_2)\big\} \in u_1.\end{aligned}$$

Then similarly

$$\begin{aligned}&\big\{y_2 : \dots \{y_n : G(h_1(x_1), y_2, \dots, y_n) \in S\} \in \tilde{h}_n(u_n) \dots\big\} \in \tilde{h}_2(u_2)\\ &\text{iff } \big\{x_2 : \dots \{y_n : G(h_1(x_1), h_2(x_2), \dots, y_n) \in S\} \in \tilde{h}_n(u_n) \dots\big\} \in u_2,\end{aligned}$$

etc. After n steps we obtain the required equivalence. □

Remark An analogous statement holds as well if $h_1, \dots, h_n$ are of arbitrary arities. Actually, a similar formula holds for any open formula; we discuss this elsewhere.

Corollary 2.2 *The following are equivalent:*

(a) $\langle \tilde{h}_1(u_1), \dots, \tilde{h}_n(u_n)\rangle \in \tilde{P}$;

(b) $P \in \langle \tilde{h}_1(u_1), \dots, \tilde{h}_n(u_n)\rangle\tilde{}$;

(c) $\{x_1 : \dots \{x_n : \langle x_1, \dots, x_n\rangle \in P\} \in \tilde{h}_n(u_n) \dots\} \in \tilde{h}_1(u_1)$;

(d) $\{x_1 : \dots \{x_n : \langle h_1(x_1), \dots, h_n(x_n)\rangle \in P\} \in u_n \dots\} \in u_1$.

Proof. The first and the second formulas are equivalent by Proposition 1.6, while the second and two last formulas are equivalent by Lemma 2.1 with $\langle\,\rangle$ as G. □

Before going further, let us fix "model-theoretic relationships" we shall consider. Besides common and important concepts of homomorphism and isomorphic embedding, we shall consider less common and more general concepts of homotopy and isotopy. These concepts are customarily used for groupoids, especially in quasigroup theory. Let us give a general definition suitable for arbitrary models.

Definition 2.3 Let F and G be n-ary operations on X and Y respectively. A collection of mappings $h, h_1, \dots, h_n$ of X into Y form a *homotopy* of (X, F) into (Y, G) iff

$$h(F(x_1, \dots, x_n)) = G(h_1(x_1), \dots, h_n(x_n))$$

for all $x_1, \dots, x_n \in X$. The homotopy is an *isotopy* iff all the $h, h_1, \dots, h_n$ are bijective.

Definition 2.4 Let P and Q be n-ary relations on X and Y respectively. A collection of mappings $h_1, \dots, h_n$ of X into Y are a *homotopy* of (X, P) into (Y, Q) iff

$$P(x_1, \dots, x_n) \text{ implies } Q(h_1(x_1), \dots, h_n(x_n))$$

for all $x_1, \dots, x_n \in X$. The homotopy is an *isotopy* iff all the $h_1, \dots, h_n$ are bijective and

$$P(x_1, \dots, x_n) \text{ iff } Q(h_1(x_1), \dots, h_n(x_n)).$$

Note that when all the $h, h_1, \ldots, h_n$ coincide, then the homotopy is a homomorphism (and the isotopy is an isomorphism). In particular, homotopies of unary relations are homomorphisms.

If $\mathfrak{A}$ and $\mathfrak{B}$ have more than one operation or relation, there are various ways to define homotopies (and isotopies) between them, the weakest of which is as follows.

Definition 2.5 A family H of mappings of X into Y form a *homotopy* of $\mathfrak{A}$ into $\mathfrak{B}$ iff for any m-ary operation F in $\mathfrak{A}$ there are mappings $h, h_1, \ldots, h_m$ in H forming a homotopy of (X, F) into (Y, G) with the corresponding operation G in $\mathfrak{B}$, and for any n-ary relation P in $\mathfrak{A}$ there are mappings $h_1, \ldots, h_n$ in H forming a homotopy of (X, P) into (Y, Q) with the corresponding relation Q in $\mathfrak{B}$. The homotopy H is an *isotopy* iff all mappings in H are bijective.

Obviously, a homotopy H is a homomorphism iff $|H| = 1$. In general, the size of H can be regarded as a degree of its dissimilarity to a homomorphism. Similarly for isotopies and isomorphisms.

We need the following auxiliary result.

Proposition 2.6 *Let $F \colon X \to Y$.*

(i) *If F is surjective, then so is $\tilde{F}$.*
(ii) *If F is injective, then so is $\tilde{F}$. Moreover, $(\tilde{F})^{-1} = (F^{-1})\tilde{}$.*
(iii) *If F is bijective, then $\tilde{F}$ is a homeomorphism of βX onto βY.*

Proof. (i) We must show that for any $v \in \beta Y$ there is $u \in \beta X$ such that $\tilde{F}(u) = v$, i.e.,

$$S \in v \;\text{ iff }\; \{x : F(x) \in S\} \in u$$

for all $S \subseteq Y$. Given v, let

$$D = \big\{\{x : F(x) \in S\} : S \in v\big\}.$$

D has the finite intersection property: Given $S', S'' \in v$, we have $\{x : F(x) \in S'\} \cap \{x : F(x) \in S''\} = \{x : F(x) \in S' \cap S''\}$, so this set is in D (since $S' \cap S''$ is in v).

Let u be any ultrafilter that extends D. Then u is as required: The 'only if' part holds by definition of u. To verify the 'if' part, notice that if $S \notin v$ then $Y \setminus S \in v$, and so $\{x : F(x) \in Y \setminus S\} \in u$, whence it follows $\{x : F(x) \in S\} \notin u$ (as preimages of disjoint sets are disjoint).

(ii) We must show that if $u', u'' \in \beta X$ are distinct, then so are $\tilde{F}(u'), \tilde{F}(u'') \in \beta Y$, i.e., there is $T \in \tilde{F}(u') \setminus \tilde{F}(u'')$, and thus $\{x : F(x) \in T\} \in u' \setminus u''$. As $u' \neq u''$, there is $S \in u' \setminus u''$. Since F is injective, we have $\{x : F(x) \in F``S\} = S$, so we can put $T = F``S$.

The equality $(\tilde{F})^{-1} = (F^{-1})\tilde{}$ follows immediately.

(iii) This follows from (i) and (ii). □

Remark Clause (i) uses the assumption that any filter extends to an ultrafilter, which is, as we mentioned above, equivalent to the compactness of βX.

Now all is fixed.

Theorem 2.7 (First Extension Theorem) *Let $\mathfrak{A}$ and $\mathfrak{B}$ be two models.*

(i) *If h is a homomorphism of $\mathfrak{A}$ into $\mathfrak{B}$, then $\tilde{h}$ is a homomorphism of $\beta\mathfrak{A}$ into $\beta\mathfrak{B}$. Similarly for embeddings.*

(ii) *If H is a homotopy of $\mathfrak{A}$ into $\mathfrak{B}$, then $\{\tilde{h} : h \in H\}$ is a homotopy of $\beta\mathfrak{A}$ into $\beta\mathfrak{B}$. Similarly for isotopies.*

Proof. Obviously, clause (i) is a partial case of clause (ii), so it suffices to prove the latter. Let $\mathfrak{A} = (X, F, \ldots, P, \ldots)$ and $\mathfrak{B} = (Y, G, \ldots, Q, \ldots)$.

Operations. Let $h, h_1, \ldots, h_n$ form a homotopy of (X, F) into (Y, G). We have, for all $x_1, \ldots, x_n \in X$,

$$h(F(x_1, \ldots, x_n)) = G(h_1(x_1), \ldots, h_n(x_n)).$$

Then, by Lemma 2.1, for all $u_1, \ldots, u_n \in \beta X$,

$$\begin{aligned}
&\tilde{h}(\tilde{F}(u_1, \ldots, u_n))\\
&\qquad = \big\{S : \{x_1 : \ldots \{x_n : h(F(x_1, \ldots, x_n)) \in S\} \in u_n \ldots\} \in u_1\big\}\\
&\qquad = \big\{S : \{x_1 : \ldots \{x_n : G(h_1(x_1), \ldots, h_n(x_n)) \in S\} \in u_n \ldots\} \in u_1\big\}\\
&\qquad = \tilde{G}(\tilde{h}_1(u_1), \ldots, \tilde{h}_n(u_n)),
\end{aligned}$$

thus $\tilde{h}, \tilde{h}_1, \ldots, \tilde{h}_n$ form a homotopy of $(\beta X, \tilde{F})$ into $(\beta Y, \tilde{G})$.

If $h, h_1, \ldots, h_n$ form an isotopy, then $\tilde{h}, \tilde{h}_1, \ldots, \tilde{h}_n$ form an isotopy by Proposition 2.6.

Relations. Let $h_1, \ldots, h_n$ form a homotopy of (X, P) into (Y, Q). Then we have, for all $x_1, \ldots, x_n \in X$, that

$$\langle x_1, \ldots, x_n \rangle \in P \ \text{ implies } \ \langle h_1(x_1), \ldots, h_n(x_n) \rangle \in Q.$$

We must verify that, for all $u_1, \ldots, u_n \in \beta X$,

$$\langle u_1, \ldots, u_n \rangle \in \tilde{P} \ \text{ implies } \ \langle \tilde{h}_1(u_1), \ldots, \tilde{h}_n(u_n) \rangle \in \tilde{Q},$$

thus $\big\{x_1 : \ldots \{x_n : \langle x_1, \ldots, x_n \rangle \in P\} \in u_n\} \ldots\big\} \in u_1$ implies that

$$\big\{x_1 : \ldots \{x_n : \langle x_1, \ldots, x_n \rangle \in Q\} \in \tilde{h}_n(u_n) \ldots\big\} \in \tilde{h}_1(u_1).$$

By Corollary 2.2, the latter formula is equivalent to

$$\big\{x_1 : \ldots \{x_n : \langle h_1(x_1), \ldots, h_n(x_n) \rangle \in Q\} \in u_n \ldots\big\} \in u_1.$$

That $h_1, \ldots, h_n$ form a homotopy just means that

$$P \subseteq \{\langle x_1, \ldots, x_n \rangle : \langle h_1(x_1), \ldots, h_n(x_n) \rangle \in Q\}.$$

Therefore, the implication holds since $u_1, \ldots, u_n$ are filters, thus $\tilde{h}_1, \ldots, \tilde{h}_n$ form a homotopy of $(\beta X, \tilde{P})$ into $(\beta Y, \tilde{Q})$.

If $h, h_1, \ldots, h_n$ form an isotopy, we prove that $\tilde{h}, \tilde{h}_1, \ldots, \tilde{h}_n$ form an isotopy in the same way but with "iff" instead of "implies" and apply Proposition 2.6. □

3 Topological properties of extensions

Topology provides a natural language describing the structure of ultrafilter extensions of models. In this section, we establish specific topological properties of extended mapping and relations, then we isolate them in abstracto, for mapping and relations on topological spaces; this results to a certain class of models endowed with topologies (which is larger than the classes of usual topological, or even semitopological models). The main aim of these studies lies in the next section; then we shall show that ultrafilter extensions are *largest* extensions belonging to that class.

We start from an explicit description of extensions of (unary) mappings to arbitrary compact Hausdorff spaces.

Definition 3.1 If $F\colon X \to Y$ where Y is a compact Hausdorff topological space, let $\tilde{F}\colon \beta X \to Y$ be defined as follows:

$$\tilde{F}(u) = v \ \text{ iff } \ \{v\} = \bigcap_{A \in u} \operatorname{cl}_Y(F``A).$$

It is routine to check that the intersection consists of a single point (it has at most one point as Y is compact, and at least one point as u is ultra), so the definition is correct, and that $\tilde{F}$ is a continuous extension of F, unique since Y is Hausdorff.

If the compact space is βY, the ultrafilter $\tilde{F}(u)$ can be rewritten in a form closer to that we knew already.

Lemma 3.2 *If $F\colon X \to \beta Y$, then*

$$\tilde{F}(u) = \bigl\{S \subseteq Y : \{x \in X : F(x) \in \tilde{S}\} \in u\bigr\}.$$

Proof. It easily follows from the definition that

$$\tilde{F}(u) = \{S \subseteq Y : (\forall A \in u)\ (\exists x \in A)\ \ F(x) \in \tilde{S}\}.$$

It remains to verify that

$$(\forall A \in u)\ (\exists x \in A)\ \ F(x) \in \tilde{S} \ \text{ iff } \ \{x \in X : F(x) \in \tilde{S}\} \in u.$$

'If' uses the fact that u is a filter, while 'only if' uses that u is ultra. □

In particular, if $F\colon X \to Y \subseteq \beta Y$ with Y discrete, then $\tilde{F}$ in the sense of the former definition coincides with $\tilde{F}$ in the sense of the new definition, thus witnessing that we do not abuse notation.

Now we establish topological properties of extended mappings of arbitrary arity.

Lemma 3.3 *Let $F\colon X_1 \times \cdots \times X_n \to Y$. For each i, $1 \le i \le n$, and for all $x_1 \in X_1, \ldots, x_{i-1} \in X_{i-1}$ and $u_{i+1} \in \beta X_{i+1}, \ldots, u_n \in \beta X_n$, the mapping $\tilde{F}_{x_1,\ldots,x_{i-1},u_{i+1},\ldots,u_n}$ of βX_i into βY defined by*

$$u \longmapsto \tilde{F}(x_1, \ldots, x_{i-1}, u, u_{i+1}, \ldots, u_n)$$

is continuous. Moreover, $\tilde{F}$ is the only such extension of F.

Proof. We shall show that $\tilde{F}$ can be constructed by fixing successively all but one arguments and extending the resulting unary functions. First we describe the construction and verify that the constructed extension has the required continuity properties. Then we verify that it coincides with $\tilde{F}$.

Step 1. Fix all but the last arguments: $x_1 \in X_1, \dots, x_{n-1} \in X_{n-1}$, and put

$$f_{x_1,\dots,x_{n-1}}(x) = F(x_1, \dots, x_{n-1}, x).$$

Thus $f_{x_1,\dots,x_{n-1}} \colon X_n \to Y$. We extend it to $\tilde{f}_{x_1,\dots,x_{n-1}} \colon \beta X_n \to \beta Y$ and put

$$F_1(x_1, \dots, x_{n-1}, u) = \tilde{f}_{x_1,\dots,x_{n-1}}(u).$$

Thus $F_1 \colon X_1 \times \cdots \times X_{n-1} \times \beta X_n \to \beta Y$. It is obvious from the construction that F_1 is continuous in its last argument (since then it coincides with $\tilde{f}_{x_1,\dots,x_{n-1}}$). And it is continuous in any other of its arguments (since then its domain is discrete).

Step 2. Fix all but the $(n-1)$th arguments: $x_1 \in X_1, \dots, x_{n-2} \in X_{n-2}$, $u_n \in \beta X_n$, and put

$$f_{x_1,\dots,x_{n-2},u_n}(x) = F_1(x_1, \dots, x_{n-2}, x, u_n).$$

Thus $f_{x_1,\dots,x_{n-2},u_n} \colon X_{n-1} \to \beta Y$. We extend it to $\tilde{f}_{x_1,\dots,x_{n-2},u_n} \colon \beta X_{n-1} \to \beta Y$ and put

$$F_2(x_1, \dots, x_{n-2}, u, u_n) = \tilde{f}_{x_1,\dots,x_{n-2},u_n}(u).$$

Hence $F_2 \colon X_1 \times \cdots \times \beta X_{n-2} \times \beta X_n \to \beta Y$. The mapping F_2 is continuous in its $(n-1)$th argument (since then it coincides with $\tilde{f}_{x_1,\dots,x_{n-2},u_n}$). Moreover, it is continuous in its nth argument whenever the fixed $(n-1)$th argument is in X_{n-1} (since then it coincides with F_1).

Arguing so, after $n-1$ steps we get $F_{n-1} \colon X_1 \times \beta X_2 \times \cdots \times \beta X_n \to \beta Y$, which is continuous in its ith argument whenever any jth fixed argument is in X_j, for all i, $1 \le i \le n$, and all $j < i$.

Step n. Fix all but the first arguments: $u_2 \in \beta X_2, \dots, u_n \in \beta X_n$, and put

$$f_{u_2,\dots,u_n}(x) = F_{n-1}(x, u_2, \dots, u_n).$$

Thus $f_{u_2,\dots,u_n} \colon X_1 \to \beta Y$. We extend it to $\tilde{f}_{u_2,\dots,u_n} \colon \beta X_1 \to \beta Y$ and put

$$F_n(u, u_2, \dots, u_n) = \tilde{f}_{u_2,\dots,u_n}(u).$$

Thus $F_n \colon \beta X_1 \times \cdots \times \beta X_n \to \beta Y$. The mapping F_n is continuous in its first argument (since then it coincides with $\tilde{f}_{u_2,\dots,u_n}$). Moreover, it is continuous in its ith argument whenever any jth fixed argument is in X_j, for all i, $1 \le i \le n$, and all $j < i$.

The uniqueness of such an extension follows from the uniqueness of continuous extensions of unary mappings by induction.

It remains to verify that F_n coincides with $\tilde{F}$. We have:

$$\begin{aligned} F_1(x_1, \dots, x_{n-1}, u_n) &= \tilde{f}_{x_1,\dots,x_{n-1}}(u_n) \\ &= \bigl\{S : \{x : f_{x_1,\dots,x_{n-1}}(x) \in S\} \in u_n\bigr\} \\ &= \tilde{F}(\hat{x}_1, \dots, \hat{x}_{n-1}, u_n). \end{aligned}$$

Then

$$\begin{aligned}
F_2(x_1,\dots,x_{n-2},u_{n-1},u_n) &= \tilde{f}_{x_1,\dots,x_{n-2},u_n}(u_{n-1})\\
&= \bigl\{S : \{x_{n-1} : f_{x_1,\dots,x_{n-2},u_n}(x_{n-1}) \in \tilde{S}\} \in u_{n-1}\bigr\}\\
&= \bigl\{S : \{x_{n-1} : F_1(x_1,\dots,x_{n-2},x_{n-1},u_n) \in \tilde{S}\} \in u_{n-1}\bigr\}\\
&= \bigl\{S : \{x_{n-1} : \tilde{f}_{x_1,\dots,x_{n-1}}(u_n) \in \tilde{S}\} \in u_{n-1}\bigr\}\\
&= \bigl\{S : \{x_{n-1} : \{T : \{x : f_{x_1,\dots,x_{n-1}}(x) \in T\} \in u_n\} \in \tilde{S}\} \in u_{n-1}\bigr\}\\
&= \bigl\{S : \{x_{n-1} : S \in \{T : \{x : f_{x_1,\dots,x_{n-1}}(x) \in T\} \in u_n\}\} \in u_{n-1}\bigr\}\\
&= \bigl\{S : \{x_{n-1} : \{x_n : f_{x_1,\dots,x_{n-1}}(x_n) \in S\} \in u_n\} \in u_{n-1}\bigr\}\\
&= \tilde{F}(\hat{x}_1,\dots,\hat{x}_{n-2},u_{n-1},u_n).
\end{aligned}$$

Likewise we get $F_n(u_1,\dots,u_n) = \tilde{F}(u_1,\dots,u_n)$, as required. □

Remark This description of continuity of extended mappings cannot be improved. If some of $u_1,\dots,u_{i-1}$ is non-principal, then the mapping $\tilde{F}_{u_1,\dots,u_{i-1},u_{i+1},\dots,u_n}$ of βX_i into βY defined by

$$u \longmapsto \tilde{F}(u_1,\dots,u_{i-1},u,u_{i+1},\dots,u_n)$$

is not necessarily continuous. For example, let F be the usual (binary) addition of natural numbers; then the mapping $u \mapsto u_1 \tilde{+} u$ is discontinuous. Also for fixed only $x_1 \in X_1,\dots,x_{i-1} \in X_{i-1}$, the $(n-i+1)$-ary mapping $\tilde{F}_{x_1,\dots,x_{i-1}}$ of $\beta X_i \times \cdots \times \beta X_n$ into βY defined by

$$\langle u_i,\dots,u_n\rangle \longmapsto \tilde{F}(x_1,\dots,x_{i-1},u_i,\dots,u_n)$$

is not necessarily continuous. For instance, let $F(x_1,x_2,x_3) = x_2 + x_3$ and use the previous observation.

To name the established topological property of $\tilde{F}$ shortly, let us introduce a terminology.

Definition 3.4 Let $X_1,\dots,X_n,Y$ be topological spaces, and let $C_1 \subseteq X_1,\dots,C_{n-1} \subseteq X_{n-1}$. We shall say that an n-ary function $F\colon X_1 \times \cdots \times X_n \to Y$ is *right continuous* with respect to $C_1,\dots,C_n$ iff for each i, $1 \le i \le n$, and all $c_1 \in C_1,\dots,c_{i-1} \in C_{i-1}$ and $x_{i+1} \in X_{i+1},\dots,x_n \in X_n$, the mapping

$$x \longmapsto F(c_1,\dots,c_{i-1},x,x_{i+1},\dots,x_n)$$

of X_i into Y is continuous. If all the C_i coincide with, say C, we shall say that F is right continuous with respect to C.

In particular, F is right continuous with respect to the empty set if and only if, for all $x_2 \in X_2,\dots,x_n \in X_n$, the mapping

$$x \longmapsto F(x,x_2,\dots,x_n)$$

of X_1 into Y is continuous. Clearly, a unary F is right continuous iff it is continuous. If the operation is binary, right continuity with respect to the empty set means that all right translations are continuous, and usually referred to as "right continuity"; see e.g. **[11]**. If F is right continuous with respect to the whole $X_1,\dots,X_n$, then it is called *separately continuous*.

The following proposition notes obvious properties of compositions of right continuous functions.

Proposition 3.5

(i) *Let $F\colon X_1 \times \cdots \times X_n \to Y$ be right continuous with respect to $C_1, \ldots, C_n$, and let $g\colon Y \to Z$ be continuous. Then $H\colon X_1 \times \cdots \times X_n \to Z$ defined by*

$$H(x_1, \ldots, x_n) = g(F(x_1, \ldots, x_n))$$

is right continuous with respect to $C_1, \ldots, C_n$.

(ii) *Let $f_1\colon X_1 \to Y_1, \ldots, f_n\colon X_n \to Y_n$ be continuous, and let $G\colon Y_1 \times \cdots \times Y_n \to Z$ be right continuous with respect to $D_1, \ldots, D_n$. Then $H\colon X_1 \times \cdots \times X_n \to Z$ defined by*

$$H(x_1, \ldots, x_n) = F(h_1(x_1), \ldots, h_n(x_n))$$

is right continuous with respect to $f_1^{-1}D_1, \ldots, f_n^{-1}D_n$.

Proof. Clear. □

Definition 3.6 We shall say that an algebra is *right topological* with C a *right topological center* iff all its operations are right continuous with respect to C.

In these terms, Lemma 3.3 states that for any algebra $\mathfrak{A} = (X, F, \ldots)$, its extension $\beta\mathfrak{A} = (\beta X, \tilde{F}, \ldots)$ is right topological with X a right topological center.

Further, we establish topological properties of extended relations.

Lemma 3.7 *Let $P \subseteq X_1 \times \cdots \times X_n$. For all i, $1 \le i \le n$, and all $x_1 \in X_1, \ldots, x_{i-1} \in X_{i-1}$ and $u_{i+1} \in \beta X_{i+1}, \ldots, u_n \in \beta X_n$, the subset*

$$\tilde{P}_{x_1,\ldots,x_{i-1},u_{i+1},\ldots,u_n} = \{u \in \beta X_i : \langle \hat{x}_1, \ldots, \hat{x}_{i-1}, u, u_{i+1}, \ldots, u_n\rangle \in \tilde{P}\}$$

of βX_i is clopen.

Proof. Let

$$f_{x_1,\ldots,x_{i-1},u_{i+1},\ldots,u_n}(u) = \langle \hat{x}_1, \ldots, \hat{x}_{i-1}, u, u_{i+1}, \ldots, u_n\rangle^{\sim}.$$

The mapping $f_{x_1,\ldots,x_{i-1},u_{i+1},\ldots,u_n}$ of βX_i into $\beta(X_1 \times \cdots \times X_n)$ is continuous by the previous lemma. Hence

$$\begin{aligned}
\tilde{P}_{x_1,\ldots,x_{i-1},u_{i+1},\ldots,u_n} &= \{u \in \beta X_i : \langle \hat{x}_1, \ldots, \hat{x}_{i-1}, u, u_{i+1}, \ldots, u_n\rangle \in \tilde{P}\} \\
&= \{u \in \beta X_i : P \in \langle \hat{x}_1, \ldots, \hat{x}_{i-1}, u, u_{i+1}, \ldots, u_n\rangle^{\sim}\} \\
&= \{u \in \beta X_i : P \in f_{x_1,\ldots,x_{i-1},u_{i+1},\ldots,u_n}(u)\} \\
&= \{u \in \beta X_i : f_{x_1,\ldots,x_{i-1},u_{i+1},\ldots,u_n}(u) \in \tilde{Q}\},
\end{aligned}$$

where Q is P considered as a unary relation on $X_1 \times \cdots \times X_n$, thus $\tilde{Q}$ is a unary relation on $\beta(X_1 \times \cdots \times X_n)$. Since Q is clopen, so is its preimage $\tilde{P}_{x_1,\ldots,x_{i-1},u_{i+1},\ldots,u_n}$ under the continuous mapping $f_{x_1,\ldots,x_{i-1},u_{i+1},\ldots,u_n}$. □

Remark One can also derive Lemma 3.7 from Lemma 3.3 by replacing the relation P with its characteristic function.

To name shortly the established topological property of $\tilde{P}$, let us introduce notation.

Definition 3.8 Let $X_1, \ldots, X_n$ be topological spaces, and $C_1 \subseteq X_1, \ldots, C_{n-1} \subseteq X_{n-1}$. We shall say that an n-ary relation $P \subseteq X_1 \times \cdots \times X_n$ is *right open* with respect to $C_1, \ldots, C_n$ iff for every i, $1 \le i \le n$, for all $c_1 \in C_1, \ldots, c_{i-1} \in C_{i-1}$, and for all $x_{i+1} \in X_{i+1}, \ldots, x_n \in X_n$, the subset

$$P_{c_1,\ldots,c_{i-1},x_{i+1},\ldots,x_n} = \{x \in X_i : \langle c_1, \ldots, c_{i-1}, x, x_{i+1}, \ldots, x_n\rangle \in P\}$$

of X_i is open. That a relation is *right closed* (or *right clopen*, etc.) is defined likewise.

In particular, P is right open with respect to the empty set if and only if, for all $x_2 \in X_2, \dots, x_n \in X_n$, the subset

$$P_{x_2,\dots,x_n} = \{x \in X_1 : \langle x, x_2, \dots, x_n \rangle \in P\}$$

of X_1 is open. Clearly, a unary P is right open iff it is open. Likewise for right closed (right clopen, etc.) relations.

The following proposition notes an obvious interplay of right open (right closed, right clopen, etc.) relations and right continuous functions.

Proposition 3.9

(i) *Let $F\colon X_1 \times \cdots \times X_n \to Y$ be right continuous with respect to $C_1, \dots, C_n$, and let $Q \subseteq Y$ be open. Then*

$$P = \{\langle x_1, \dots, x_n \rangle \in X_1 \times \cdots \times X_n : F(x_1, \dots, x_n) \in Q\}$$

is right open with respect to $C_1, \dots, C_n$.

(ii) *Let $F_1\colon X_1 \to Y_1, \dots, F_n\colon X_n \to Y_n$ be continuous, and let $Q \subseteq Y_1 \times \cdots \times Y_n$ be right open with respect to $D_1, \dots, D_n$. Then*

$$P = \{\langle x_1, \dots, x_n \rangle \in X_1 \times \cdots \times X_n : \langle F_1(x_1), \dots, F_n(x_n) \rangle \in Q\}$$

is right open with respect to $F_1^{-1}D_1, \dots, F_n^{-1}D_n$.

Both clauses also hold for right closed (right clopen, etc.) relations.

Proof. Clear. □

Now we are ready to isolate the class of models having “the same” topological properties as ultrafilter extensions.

Definition 3.10 Let $\mathfrak{A} = (X, F, \dots, P, \dots)$ be a model equipped with a topology, and let $C \subseteq X$. We shall say that $\mathfrak{A}$ is *right open*, and C is its *right topological center*, iff all its operations are right continuous with respect to C and all its relations are right open with respect to C. Likewise for *right closed* (*right clopen*, etc.) models.

Note that if the model is an algebra (i.e., does not have relations), each of these properties means that the algebra is right topological with C a right topological center.

In these terms, two last lemmas state the following.

Theorem 3.11 *For any model $\mathfrak{A}$, its extension $\beta\mathfrak{A}$ is right clopen with $\mathfrak{A}$ a right topological center.*

Proof. Lemmas 3.3 and 3.7. □

4 The Second Extension Theorem

Here we prove the main result of the paper, which establishes that ultrafilter extensions are largest extensions in the class of compact Hausdorff right-closed models (Theorem 4.2). This result confirms that the construction provides a right generalization of the largest compactification of a discrete space X, i.e., its Stone–Čech (or Wallman) compactification, to the situation when X is endowed with a first-order structure.

We start with an “abstract extension theorem”, which concerns rather arbitrary right open and right closed models with dense right topological centers than ultrafilter extensions.

Theorem 4.1 *Let $\mathfrak{A}$ be a right open model, $\mathfrak{B}$ a Hausdorff right closed model, and $\mathfrak{C} \subseteq \mathfrak{A}$ a dense submodel and a right topological center of $\mathfrak{A}$.*

(i) *If h is a continuous mapping of $\mathfrak{A}$ into $\mathfrak{B}$ such that $h \restriction \mathfrak{C}$ is a homomorphism and $h``\mathfrak{C}$ is a right topological center of $\mathfrak{B}$, then h is a homomorphism of $\mathfrak{A}$ into $\mathfrak{B}$. Similarly for embeddings.*

(ii) *If H is a family of continuous mappings of $\mathfrak{A}$ into $\mathfrak{B}$ such that $\{h \restriction \mathfrak{C} : h \in H\}$ is a homotopy and $h``\mathfrak{C}$ is a right topological center of $\mathfrak{B}$ for each $h \in H$, then H is a homotopy of $\mathfrak{A}$ into $\mathfrak{B}$. Similarly for isotopies.*

Proof. Likewise Theorem 2.7, it suffices to prove clause (ii). To simplify notation, however, below we prove clause (i) —it is easy to see that the proof of clause (ii) is essentially the same (typically, $h(x_i)$ should be replaced by $h_i(x_i)$).

Let $\mathfrak{A} = (X, F, \ldots, P, \ldots)$ and $\mathfrak{B} = (Y, G, \ldots, Q, \ldots)$.

Operations. We argue by induction on the arity of F (and G).

Step 1. Fix $c_1, \ldots, c_{n-1} \in C$ and put, for all $x \in X$ and $y \in Y$,

$$f_{c_1,\ldots,c_{n-1}}(x) = F(c_1, \ldots, c_{n-1}, x),$$
$$g_{h(c_1),\ldots,h(c_{n-1})}(y) = G(h(c_1), \ldots, h(c_{n-1}), y).$$

The functions $f_{c_1,\ldots,c_{n-1}}$ and $g_{h(c_1),\ldots,h(c_{n-1})}$ are continuous (since $c_1, \ldots, c_{n-1}$ are in C, C is a right topological center of $\mathfrak{A}$, and $h``C$ is a right topological center of $\mathfrak{B}$). Therefore the functions $h \circ f_{c_1,\ldots,c_{n-1}}$ and $g_{h(c_1),\ldots,h(c_{n-1})} \circ h$ (both of X to Y) are continuous too (as compositions of continuous functions). Moreover, they agree on the dense subset C of X (since $\mathfrak{C}$ is a subalgebra and $h \restriction C$ is a homomorphism), i.e., for all $c \in C$,

$$h(f_{c_1,\ldots,c_{n-1}}(c)) = g_{h(c_1),\ldots,h(c_{n-1})}(h(c)).$$

Hence (as Y is Hausdorff) they coincide, i.e., for all $x \in X$,

$$h(f_{c_1,\ldots,c_{n-1}}(x)) = g_{h(c_1),\ldots,h(c_{n-1})}(h(x)).$$

Thus we proved that, for all $c_1, \ldots, c_{n-1} \in C$ and $x_n \in X$,

$$h(F(c_1, \ldots, c_{n-1}, x_n)) = G(h(c_1), \ldots, h(c_{n-1}), h(x_n)).$$

Step 2. Fix $c_1, \ldots, c_{n-2} \in C$ and $x_n \in X$, and put, for all $x \in X$ and $y \in Y$,

$$f_{c_1,\ldots,c_{n-2},x_n}(x) = F(c_1, \ldots, c_{n-2}, x, x_n),$$
$$g_{h(c_1),\ldots,h(c_{n-2}),h(x_n)}(y) = G(h(c_1), \ldots, h(c_{n-2}), y, h(x_n)).$$

Again, the functions $f_{c_1,\ldots,c_{n-2},x_n}$ and $g_{h(c_1),\ldots,h(c_{n-2}),h(x_n)}$ are continuous (as $c_1, \ldots, c_{n-2}$ are in C, C is a right topological center of $\mathfrak{A}$, and $h``C$ is a right topological center of $\mathfrak{B}$). Therefore the compositions $h \circ f_{c_1,\ldots,c_{n-2},x_n}$ and $g_{h(c_1),\ldots,h(c_{n-2}),h(x_n)} \circ h$ (both of X to Y) are continuous too. Moreover, they agree on the dense subset C of X (by Step 1), i.e., for all $c \in C$,

$$h(f_{c_1,\ldots,c_{n-2},x_n}(c)) = g_{h(c_1),\ldots,h(c_{n-2}),h(x_n)}(h(c)).$$

Hence they coincide, i.e., for all $x \in X$,

$$h(f_{c_1,\ldots,c_{n-2},x_n}(x)) = g_{h(c_1),\ldots,h(c_{n-2}),h(x_n)}(h(x)).$$

Thus we proved that, for all $c_1, \ldots, c_{n-2} \in C$ and $x_{n-1}, x_n \in X$,

$$h(F(c_1, \ldots, c_{n-2}, x_{n-1}, x_n)) = G(h(c_1), \ldots, h(c_{n-2}), h(x_{n-1}), h(x_n)).$$

After n steps, we get $h(F(x_1, \ldots, x_n)) = G(h(x_1), \ldots, h(x_n))$ for all $x_1, \ldots, x_n \in X$, thus showing that h is a homomorphism of (X, F) into (Y, G), as required.

Relations. Assuming $\langle x_1, \ldots, x_n\rangle \in P$, we shall show that $\langle h(x_1), \ldots, h(x_n)\rangle \in Q$ by induction on n.

Step 1. First we suppose $c_1, \ldots, c_{n-1} \in C$. Pick an arbitrary neighborhood V of $h(x_n)$. Since h is continuous, there exists a neighborhood U of x_n such that $h``U \subseteq V$. The set $U \cap P_{c_1,\ldots,c_{n-1}}$ is open ($P_{c_1,\ldots,c_{n-1}}$ is open as $c_1, \ldots, c_{n-1}$ are in the right topological center C) and nonempty (x_n belongs to it), and so there is $c \in C \cap U \cap P_{c_1,\ldots,c_{n-1}}$ (since C is dense). Therefore, we have $\langle c_1, \ldots, c_{n-1}, c\rangle \in P$, and so $\langle h(c_1), \ldots, h(c_{n-1}), h(c)\rangle \in Q$ (since $h \restriction C$ is a homomorphism).

Hence, every neighborhood of $h(x_n)$ has a point y with $\langle h(c_1), \ldots, h(c_{n-1}), y\rangle \in Q$. Since the set

$$Q_{h(c_1),\ldots,h(c_{n-1})} = \{y : \langle h(c_1), \ldots, h(c_{n-1}), y\rangle \in Q\}$$

is closed (as $h(c_1), \ldots, h(c_{n-1})$ are in the right topological center $h``C$), it contains the point $h(x_n)$. Thus we proved that whenever $c_1, \ldots, c_{n-1} \in C$ and $\langle c_1, \ldots, c_{n-1}, x_n\rangle \in P$, then $\langle h(c_1), \ldots, h(c_{n-1}), h(x_n)\rangle \in Q$.

Step 2. Now we suppose $c_1, \ldots, c_{n-2} \in C$ and $x_n \in X$. Pick an arbitrary neighborhood V of $h(x_{n-1})$. Since h is continuous, there exists a neighborhood U of x_{n-1} such that $h``U \subseteq V$. Again, the set $U \cap P_{c_1,\ldots,c_{n-2},x_n}$ is open and nonempty, so there exists $c \in C \cap U \cap P_{c_1,\ldots,c_{n-2},x_n}$. Hence, $\langle c_1, \ldots, c_{n-2}, c, x_n\rangle \in P$, and so

$$\langle h(c_1), \ldots, h(c_{n-2}), h(c), h(x_n)\rangle \in Q$$

(by Step 1). Consequently, we infer that every neighborhood of $h(x_{n-1})$ has a point y with $\langle h(c_1), \ldots, h(c_{n-2}), y, h(x_n)\rangle \in Q$. Since the set

$$Q_{h(c_1),\ldots,h(c_{n-2}),h(x_n)} = \{y : \langle h(c_1), \ldots, h(c_{n-2}), y, h(x_n)\rangle \in Q\}$$

is closed, it has the point $h(x_{n-1})$. Thus we proved that whenever $c_1, \ldots, c_{n-2} \in C$ and $\langle c_1, \ldots, c_{n-2}, x_{n-1}, x_n\rangle \in P$, then

$$\langle h(c_1), \ldots, h(c_{n-2}), h(x_{n-1}), h(x_n)\rangle \in Q.$$

After n steps, we conclude that whenever $\langle x_1, \ldots, x_n\rangle \in P$, then $\langle h(x_1), \ldots, h(x_n)\rangle \in Q$, thus h is a homomorphism of (X, P) into (Y, Q), as required.

For embeddings we use the same argument, with the only difference in the part about relations where implications should be replaced by logical equivalences, and apply Proposition 2.6. □

After all this, we are able to obtain our main result.

Theorem 4.2 (Second Extension Theorem) *Let $\mathfrak{A}$ and $\mathfrak{B}$ be two models, and let $\mathfrak{B}$ be compact Hausdorff right closed (while $\mathfrak{A}$ is considered discrete).*

(i) *If h is a homomorphism of $\mathfrak{A}$ into $\mathfrak{B}$ such that $h``\mathfrak{A}$ is a right topological center of $\mathfrak{B}$, then $\tilde{h}$ is a homomorphism of $\beta\mathfrak{A}$ into $\mathfrak{B}$. Similarly for embeddings.*

(ii) *If H is a homotopy of $\mathfrak{A}$ into $\mathfrak{B}$ such that $h``\mathfrak{A}$ is a right topological center of $\mathfrak{B}$ for each $h \in H$, then $\{\tilde{h} : h \in H\}$ is a homotopy of $\beta\mathfrak{A}$ into $\mathfrak{B}$. Similarly for isotopies.*

Proof. Theorem 3.11 and Theorem 4.1. □

Note that the First Extension Theorem (Theorem 2.7) follows from this one.

5 Problems

Here we mark some tasks and directions for further investigations and also mention a few related results without proofs.

We have seen that under ultrafilter extensions, certain model-theoretic properties of mappings (or systems of mappings) are preserved; this is the case for homomorphisms and embeddings (and more generally, homotopies and isotopies). It is natural to ask whether that holds for other kinds of model-theoretic mappings, e.g. elementary embeddings or strong homomorphisms. Furthermore, one can inquire about other relationships between models, such as elementary equivalence or bisimulations.

Problem 5.1 Characterize relationships between models satisfying the extension theorems.

The next problem, besides a pure model-theoretic interest, has a motivation in Ramsey theory. A cornerstone of its numerous applications in number theory, algebra, topological dynamics, etc. is the existence of non-principal ultrafilters that are *idempotents* of ultrafilter extensions of semigroups. The proof of the existence is based on two facts. One is that ultrafilter extensions of semigroups are semigroups themselves, thus associativity is *stable* under the ultrafilter extension procedure (and the second one is that compact Hausdorff right-topological semigroups contain idempotents).

Problem 5.2 Characterize formulas that are stable under ultrafilter extensions.

In general, equational theories of ultrafilter extensions quite differ from the equational theories of underlying models and are highly complicated (e.g., the "simple" question whether some three non-principal u, v, w in $\beta\mathbb{N}$ satisfy $u \,\tilde{\cdot}\, (v \mathbin{\tilde{+}} w) = u \,\tilde{\cdot}\, v \mathbin{\tilde{+}} u \,\tilde{\cdot}\, w$ was posed decades ago and still remains open; see also [**11**]). In [**14**] we produce the following sufficient condition (but we do not know at the moment whether it is necessary).

Theorem 5.3 *Let an identity $s_1 = s_2$ be equivalent to some identity $t_1 = t_2$ such that the common variables of t_1 and t_2 appear in these terms in the same ordering, and any common variable occurs in each of the terms only once. Then the identity $s_1 = s_2$ is stable under ultrafilter extensions.* □

For example, it is easy to check that the following identities (in the language of groupoids) satisfy the condition of Theorem 5.3 and so are stable under ultrafilter extensions: $xy = (xy)z$, $xy = xx$, $xy = (yx)z$, $(xy)(zw) = x(y(zw))$. On the other hand, it can be shown that neither commutativity nor idempotency are stable.

The literature devoted to various types of ultrafilters is too vast for describing all its directions; let us mention a few of them. First, one has special ultrafilters over $\mathbb{N}$: p-points, q-points, selective ultrafilters, etc. Second, certain ultrafilters are important for model-theoretic constructions: regular, good, etc. Third, concepts of σ-completeness and normality of ultrafilters are at the heart of the theory of large cardinals.

Problem 5.4 Study the role of specific ultrafilters in ultrafilter extensions.

The following result shows that ultrafilters of certain types form submodels of extended models ([**15**]). An n-ary operation F is *κ-cancellative* iff for any i and all fixed $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n, y$, we have $|\{x_i : F(x_1, \ldots, x_n) = y\}| < \kappa$.

Theorem 5.5 *Let $\mathfrak{A}$ be a model and κ a cardinal.*

(i) *The set $\{u \in \beta\mathfrak{A} : u$ is κ-complete$\}$ forms a submodel of $\beta\mathfrak{A}$.*

(ii) *The set* $\{u \in \beta\mathfrak{A} : u$ *is* κ*-uniform*$\}$ *forms a closed submodel of* $\beta\mathfrak{A}$ *whenever all operations in* $\beta\mathfrak{A}$ *are* λ*-cancellative with some* $\lambda < \kappa$ *or* κ*-cancellative and* κ *regular.* □

Further, let us treat ultrafilters as quantifiers (as mentioned in [**2**]). For each $u \in \beta X$ we can provide a new symbol Q_u and interpret $\mathsf{Q}_u x\ \varphi(x)$ in a model with the universe X by $\{x \in X : \varphi(x)\} \in u$. Clearly, each Q_u satisfies $\mathsf{Q}_u x\ \varphi(x) \to \exists x\ \varphi(x)$ and commutes with all Boolean connectives but not with another Q_v.

Problem 5.6 Develop a theory of ultrafilter extensions treating ultrafilters as quantifiers.

Perhaps this approach could simplify the cumbersome notation used here.

Problem 5.7 Study ultrafilter extensions without the axiom of choice or with alternative axioms, e.g. AD.

This is the subject of [**15**].

Problem 5.8 Study connections of ultrafilter extensions with ultraproducts.

Let us give a simple related result.

Theorem 5.9 *If* $\tilde{F}(u_1, \ldots, u_n) = v$ *then* $j\colon \prod_v \mathfrak{A} \prec \prod_u \mathfrak{A}$, *where* $u = \langle u_1, \ldots, u_n\rangle\tilde{}$ *and* j *is defined by* $j = [f \circ F]_u$ *for all* f. □

In particular, $v \leq_{\mathrm{RK}} u$ implies $\prod_v \mathfrak{A} \prec \prod_u \mathfrak{A}$.

Problem 5.10 Generalize the ultrafilter extension procedure from discrete models to larger classes of models with topologies.

Perhaps, Wallman compactifications of T_1 right open or right closed models can be endowed with the extended first-order structure essentially in the same way as this has been done here.

A motivation of the following task is to find *effective* analogs of proofs that use ultrafilters. Recall that ultrafilters on Boolean algebras of definable subsets are complete types.

Problem 5.11 Construct extensions of models by ultrafilters on Boolean algebras of definable subsets.

Ultrafilters *encode* complex properties of their underlying models; applications are based on this phenomen. As the simplest example, Ramsey's Theorem corresponds to *non-principal* ultrafilters; whenever we have one, then we can easily get the theorem. A harder example, Hindman's Finite Sums Theorem, corresponds to *idempotent non-principal* ultrafilters, in the same sense. Other Ramsey-theoretic results may use non-principal idempotents of some special type (see [**11**] for various further examples). Is not there here a general principle?

Problem 5.12 Find a translation of statements about models to statements about their ultrafilter extensions et vice versa.

References

[1] A. Blass, *A model without ultrafilters*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys., 25, 4 (1977), 329–331.

[2] A. Blass, *Ultrafilters: where topological dynamics = algebra = combinatorics*, Topology Proc., 18 (1993), 33–56.

[3] J. Berglund, H. Junghenn, P. Milnes, *Analysis on Semigroups*, Wiley, New York, 1989.

[4] C. C. Chang, H. J. Keisler, *Model Theory*, North-Holland, Amsterdam, London, New York, 1973.

[5] W. Comfort, S. Negrepontis, *The theory of ultrafilters*, Springer, Berlin, 1974.

[6] R. Ellis, *Lectures on Topological Dynamics*, Benjamin, New York, 1969.

[7] R. Engelking, *General Topology*, Monogr. Matem. 60, Warszawa, 1977.

[8] T. E. Frayne, A. C. Morel, D. S. Scott, *Reduced direct products*, Fund. Math., 51 (1962), 195–228. Abstract: Notices Amer. Math. Soc., 5 (1958), 673–675.

[9] R. Goldblatt, *Varieties of complex algebras*, Ann. Pure Appl. Logic, 44 (1989), 173–242.

[10] V. Goranko, *Filter and ultrafilter extensions of structures: universal-algebraic aspects*, manuscript, 2007.

[11] N. Hindman, D. Strauss, *Algebra in the Stone–Čech compactification*, W. de Gruyter, Berlin, New York, 1998.

[12] A. Kanamori, *The Higher Infinite: Large Cardinals in Set Theory from Their Beginnings*, Springer, Berlin, 2005 (2nd ed.).

[13] D. I. Saveliev, *Ultrafilter extensions of models*, Logic and Its Applications, Lecture Notes in Computer Science, 6521 (2011), 162–177.

[14] D. I. Saveliev, *Identities stable under ultrafilter extensions*, in progress.

[15] D. I. Saveliev, *On ultrafilters without the axiom of choice*, in progress. A preliminary report (2011) is available at `http://www.crm.cat/cinfinity/Saveliev_Dennis.pdf`.

[16] J. van Benthem, *Notes on modal definability*, Notre Dame J. Formal Logic, 30, 1 (1988), 20–35.

Part VI

Proofs and Sets

Some results on PA-provably recursive functions

Sy-David Friedman[†], **Michael Rathjen**[‡], **Andreas Weiermann**[§]

[†] Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
sdf@logic.univie.ac.at

[‡] Department of Pure Mathematics, University of Leeds, UK
rathjen@maths.leeds.ac.uk

[§] Vakgroep Zuivere Wiskunde en Computeralgebra, Universiteit Gent, Belgium
Andreas.Weiermann@UGent.be

Abstract. We provide some results which emerged from joint research carried out during the CRM Infinity Project. The theorems are inspired by analogies with forcing.

Analogies with forcing

In set theory one has a "ground model" with a given set of functions from ω to ω. Then three things can happen when passing to a larger model, as exemplified by:

1. Sacks forcing: One adds new functions but any new function is dominated by an old (ground model) function.
2. Cohen forcing: One adds a new function that cannot be dominated by a ground model function but no single function dominating (mod finite) all ground model functions. If one adds two such functions f, g using "Cohen times Cohen" forcing, then in addition any function added by both f and g is in the ground model.
3. Hechler forcing: One adds a new function that dominates (mod finite) all ground model functions. If one adds two such functions f, g using "Hechler times Hechler" forcing, then again any function added by both f and g is in the ground model.

The analogy in proof theory is the following: Fix a theory T like PA (Peano Arithmetic). Let us take the "T-provably recursive" functions to be those with primitive recursive graph (the honest functions) such that for some choice of primitive recursive representation of that graph, totality of the function is T-provable.

Our Theorem 1.1 says the following: There is a "natural" total recursive function f with primitive recursive graph which is not PA-provably recursive (via any primitive recursive representation of its graph) and such that no provably recursive function of $PA + \mathrm{Tot}(f)$ (where f is expressed using any primitive recursive graph representation and $\mathrm{Tot}(f)$ expresses the fact that f is total) dominates (mod finite) all provably recursive functions of PA. This is a proof-theoretic analogue of Cohen forcing.

The situation is similar for Theorem 1.2. It says that there are "natural" functions f_0, f_1 with primitive recursive graphs which are not provably recursive in PA (via any primitive recursive graph representation), yet any function which is provably recursive

The authors wish to thank the John Templeton Foundation for its generous support of this research through the CRM Infinity Project (Grant ID 13152).

in both $PA + \mathrm{Tot}(f_0)$ and $PA + \mathrm{Tot}(f_1)$ (where the latter are expressed using primitive recursive graph representations) is in fact provably recursive in PA. This is a proof-theoretic analogue of Cohen times Cohen forcing.

1 Results

Theorem 1.1 *There exists an honest number-theoretic function f such that f is not provably recursive in PA and such that any g which is provably total in $PA + \mathrm{Tot}(f)$ does not eventually dominate every PA-recursive function.*

Proof. (The proof is inspired by [**1**].) We construct f in stages. Set $d_0 := 0$. Assume we are at stage $s = 2i$ and that d_s is defined. Assume that $f(x)$ is defined for $x < d_s$. Then set $d_{s+1} := H_{\varepsilon_0}(d_s)$. We extend f by $f(x) := H_{\varepsilon_0}(x)$ for $d_s \le x < d_{s+1}$.

Assume that $s = 2i + 1$. Let $d'_{s+1} := H_{\varepsilon_0}(d_s)$ and $d_{s+1} := H_{\varepsilon_0}(d'_{s+1})$. We extend f by $f(x) := d'_{s+1} + x$ for $d_s \le x < d_{s+1}$. Moreover let $\overline{f}_s(x) := f(x)$ for $x < d_s$ and $\overline{f}_s(x) := d'_{s+1} + x$ for $d_s \le x$.

Since $f(x) = H_{\varepsilon_0}(x)$ for infinitely many x, we see that f is not provably recursive in PA. Assume now that $PA + \mathrm{Tot}(f)$ proves $\mathrm{Tot}(g)$. Then there exists an $\alpha < \varepsilon_0$ such that for all x we have $g(x) < f^\alpha(x)$, where

$$f^\alpha(x) := \max(\{f(x)\} \cup \{f^\beta(f^\beta(x)) : \beta < \alpha \wedge N\beta \le f(N\alpha + x)\}).$$

Here $N\alpha$ is defined via $N0 := 0$ and $N\alpha := N\beta + N\gamma + 1$ if α has the normal form $\omega^\beta + \gamma$.

By construction we know that for each odd s we have $\overline{f}_s(x) \le H_{d'_{s+1}}(x)$. Choose an odd stage s with $2 \cdot N\alpha + 12 \le d'_{s+1}$. Then $\overline{f}^\alpha_s(d'_{s+1}) \le H_{\varepsilon_0}(d'_{s+1})$.

For a proof we apply Theorem 4 from [**2**] (and tacitly Lemma 9 from the same article). With this we obtain from $\overline{f}_s(x) \le H_{d'_{s+1}}(x)$ (which holds for all x) that

$$\begin{aligned}
\overline{f_s}^\alpha(d'_{s+1}) &\le H_{\omega^{\alpha + d'_{s+1} + 1}+8}(d'_{s+1}) \\
&\le H_{\omega^{\alpha + d'_{s+1} + 1}+8}(d'_{s+1} + N\alpha + 10) \\
&\le H_{\omega^{\alpha+\omega}}(d'_{s+1} + N\alpha + 10) \\
&\le H_{\omega^{\alpha+\omega}}(H_{\alpha+10}(d'_{s+1})) \\
&\le H_{\omega^{\alpha+\omega}+\alpha+10}(d'_{s+1}) \le H_{\varepsilon_0}(d'_{s+1}).
\end{aligned}$$

Now we prove:

$$\overline{f_s}^\alpha(y) \le H_{\varepsilon_0}(d'_{s+1}) \Rightarrow f^\alpha(y) = \overline{f_s}^\alpha(y) \tag{1.1}$$

by induction on $\alpha < \varepsilon_0$. Assume that $\underline{f^\alpha(y) = f^\beta(f^\beta(y))}$ for some $\beta < \alpha$ with $N\beta \le f(N\alpha + y)$. We know that $\overline{f_s}(N\alpha + y) \le H_{\varepsilon_0}(d'_{s+1})$; hence $\overline{f_s}(N\alpha + y) = f(N\alpha + y) \ge N\beta$. Thus $\overline{f_s}^\beta\Big(\overline{f_s}^\beta(y)\Big) \le \overline{f_s}^\alpha(y) \le H_{\varepsilon_0}(d'_{s+1})$. So $\overline{f_s}^\beta(y) \le H_{\varepsilon_0}(d'_{s+1})$ and the induction hypothesis yields $\overline{f_s}^\beta(y) = f^\beta(y)$ and hence $\overline{f_s}^\beta\Big(\overline{f_s}^\beta(y)\Big) = f^\beta\Big(\overline{f_s}^\beta\Big)(y)) = f^\beta\big(f^\beta(y)\big)$. Since $f^\alpha(y) = f^\beta\big(f^\beta(y)\big) = \overline{f_s}^\beta\Big(\overline{f_s}^\beta(y)\Big) \le \overline{f_s}^\alpha(y)$, we are done with the proof of (1.1).

Putting things together we obtain for large enough s that

$$g(d'_{s+1}) < f^\alpha(d'_{s+1}) = \overline{f_s}^\alpha(d'_{s+1}) \le H_{\omega^{\alpha \# \omega}+\alpha+10}(d'_{s+1}).$$

So the function $H_{\omega^{\alpha \# \omega}+\alpha+10}$ is not eventually dominated by g. □

Theorem 1.2 *There are two honest recursive functions f_0, f_1 which are not provably recursive in PA such that if a function g is provably recursive in $PA + \mathrm{Tot}(f_0)$ and $PA + \mathrm{Tot}(f_1)$ then g is provably recursive in PA.*

Proof. (The proof is inspired by [**1**].) We construct f_0, f_1 in stages. Set $d_0 := 0$. Assume we are at stage $s = 2i$ and d_s is defined. Assume that $f_i(x)$ are defined for $x < d_s$. Put $\hat{d}_{s+1} := H_{\varepsilon_0}(d_s)$. Define $\overline{f_{1,s}}$ by $\overline{f_{1,s}}(x) := f_1(x)$ for $x < d_s$ and $\overline{f_{1,s}}(x) := f_1(d_s - 1) + x$ for $x \geq d_s$. Put

$$d'_{s+1} := \mu n : \exists x < n[x \geq \hat{d}_{s+1} \wedge \overline{f_{1,s}}^{\omega_i}(x) < H_{\varepsilon_0}(x) \leq n]$$

and $d_{s+1} := H_{\varepsilon_0}(d'_{s+1})$.

Extend f_0 by $f_0(x) := f_0(x)$ for $x < d_s$, $f_0(x) := \hat{d}_{s+1} + x$ for $\hat{d}_{s+1} > x \geq d_s$ and $f_0(x) := H_{\varepsilon_0}(x)$ for $d'_{s+1} > x \geq \hat{d}_{s+1}$ and $f_0(x) := d_{s+1} + x$ for $d_{s+1} > x \geq d'_{s+1}$. Define $\overline{f_{0,s}}$ by $\overline{f_{0,s}}(x) := f_0(x)$ for $x < d'_{s+1}$ and $\overline{f_{0,s}}(x) := d_{s+1} + x$ for $x \geq d'_{s+1}$. Extend f_1 by $f_1(x) := f_1(x)$ for $x < d_s$ and $f_1(x) := \overline{f_{1,s}}(x)$ for $d_{s+1} > x \geq d_s$. Assume we are at stage $s = 2i + 1$ and d_s is defined. We interchange the roles of f_0 and f_1. That means: Assume that $f_i(x)$ are defined for $x < d_s$. Put $\hat{d}_{s+1} := H_{\varepsilon_0}(d_s)$. Define $\overline{f_{0,s}}$ by $\overline{f_{0,s}}(x) := f_0(x)$ for $x < d_s$ and $\overline{f_{0,s}}(x) := f_0(d_s - 1) + x$ for $x \geq d_s$. Put

$$d'_{s+1} := \mu n : \exists x < n[x \geq \hat{d}_{s+1} \wedge \overline{f_{0,s}}^{\omega_i}(x) < H_{\varepsilon_0}(x) \leq n]$$

and $d_{s+1} := H_{\varepsilon_0}(d'_{s+1})$. Extend f_1 by $f_1(x) := f_1(x)$ for $x < \hat{d}_{s+1}$ and $f_1(x) := H_{\varepsilon_0}(x)$ for $d'_{s+1} > x \geq \hat{d}_{s+1}$ and $f_1(x) := d_{s+1} + x$ for $d_{s+1} > x \geq d'_{s+1}$.

Define $\overline{f_{1,s}}$ by $\overline{f_1}(x) := f_1(x)$ for $x < d'_{s+1}$ and $\overline{f_{1,s}}(x) := d_{s+1} + x$ for $x \geq d'_{s+1}$. Extend f_0 by $f_0(x) := f_0(x)$ for $x < d_s$ and $f_0(x) := \overline{f_{0,s}}(x)$ for $d_{s+1} > x \geq d_s$.

We write $f_0 \wedge f_1$ for $x \mapsto \min\{f_0(x), f_1(x)\}$. Then $(f_0 \wedge f_1)(x) \leq 2 \cdot x$.

Assume that $PA + \mathrm{Tot}(f_1) \vdash \mathrm{Tot}(f_0)$. Then, by [**2**], there is an $\alpha < \varepsilon_0$ such that for all x we have $f_0(x) < f_1^{\alpha}(x)$. Choose i_1 such that $\alpha < \omega_{i_1}$. Then for $x \geq N\alpha$ we have

$$f_0(x) < f_1^{\omega_{i_1}}(x). \tag{1.2}$$

Assume that s is $2i + 1$ with $i > \max\{N\alpha, i_1\}$. Then there is an $x < d'_{s+1}$ with $x \geq d_s$ such that $\overline{f_1}^{\omega_i}(x) < H_{\varepsilon_0}(x) \leq d'_{s+1}$. Since f_1 and $\overline{f_1}$ agree on $x < d_{s+1}$ we obtain

$$f_1^{\omega_i}(x) = \overline{f_1}^{\omega_i}(x) < H_{\varepsilon_0}(x) = f_0(x),$$

contradicting (1.2). By a symmetric argument, $PA + \mathrm{Tot}(f_0)$ does not prove $\mathrm{Tot}(f_1)$.

Now assume that $PA + \mathrm{Tot}(f_0) \vdash \mathrm{Tot}(g)$ and $PA + \mathrm{Tot}(f_1) \vdash \mathrm{Tot}(g)$. Then there exist α_i such that for all x we have $g(x) < f_i^{\alpha_i}(x)$. Our idea is now to show that

$$(f_0^{\alpha_0} \wedge f_1^{\alpha_1})(x) \leq (f_0 \wedge f_1)^{\alpha_0 \# \alpha_1}(x).$$

Since $(f_0 \wedge f_1)(x) \leq 2 \cdot x$, this would yield the claim. But the obvious verification does not seem to work.

For this purpose, let us introduce another iteration hierarchy:

$$f_{\alpha}(x) := \max(\{f(x)\} \cup \{f_{\beta}(f_{\beta}(x)) : \beta < \alpha \wedge N\beta \leq N\alpha + x\}).$$

This hierarchy behaves better with respect to the $\wedge$ operator.

Indeed for two increasing functions f, h we have

$$(f_{\alpha} \wedge h_{\beta})(x) \leq (f \wedge h)_{\alpha \# \beta}(x).$$

This is proved by induction on $\alpha\#\beta$. Assume first that $\beta = 0$. Then $h_\beta(x) = h(x)$. If $\alpha = 0$ then $f_\alpha(x) = f(x)$ and the assertion is clear. If $\alpha > 0$ then $f_\alpha(x) = f_\gamma(f_\gamma(x))$ for some $\gamma < \alpha$ with $N\gamma \leq N\alpha + x$. We have $(f \wedge h)_\alpha(x) \geq (f \wedge h)_\gamma((f \wedge h)_\gamma(x))$. The induction hypothesis yields $(f \wedge h)_\gamma(x) \geq (f_\gamma \wedge h)(x)$. If $(f_\gamma \wedge h)(x) = h(x)$ then

$$(f \wedge h)_\alpha(x) \geq (f \wedge h)_\gamma(h(x)) \geq h(x)$$

and the assertion follows. So assume $(f_\gamma \wedge h)(x) = f_\gamma(x)$. We then have

$$(f_\gamma \wedge h)(f_\gamma \wedge h) = (f_\gamma \wedge h)(f_\gamma(x)).$$

If $(f_\gamma \wedge h)(f_\gamma(x)) = h(f_\gamma(x))$ then the assertion follows from $h(f_\gamma(x)) \geq h(x)$. We may assume that $(f_\gamma \wedge h)(f_\gamma(x)) = f_\gamma(f_\gamma(x)) = f_\alpha(x)$. Since $f_\alpha(x) \geq h(x)$, the assertion also holds in this case.

Assume (by symmetry) for the induction step that $(f_\alpha \wedge h_\beta)(x) = h_\beta(x) = h_\gamma(h_\gamma(x))$ for some $\gamma < \beta$ with $N\gamma \leq N\beta + x$. The inequality $f_\alpha(x) \geq h_\beta(x)$ yields

$$h_\gamma(x) \leq (f_\alpha \wedge h_\gamma)(x) \leq (f \wedge h)_{\alpha\#\gamma}(x)$$

by the induction hypothesis, and the inequality $f_\alpha(h_\gamma(x)) \geq f_\alpha(x) \geq h_\beta(x)$ yields

$$h_\gamma(h_\gamma(x)) \leq f_\alpha(h_\gamma(x)) \leq (f_\alpha \wedge h_\gamma)(h_\gamma(x)) \leq (f \wedge h)_{\alpha\#\gamma}(h_\gamma(x))$$

by the induction hypothesis.

Putting things together we obtain $h_\gamma(h_\gamma(x)) \leq (f \wedge h)_{\alpha\#\gamma}((f \wedge h)_{\alpha\#\gamma}(x))$. Now we have $N(\alpha\#\gamma) \leq N(\alpha\#\beta) + x$. Hence $(f \wedge h)_{\alpha\#\gamma}((f \wedge h)_{\alpha\#\gamma}(x)) \leq (f \wedge h)_{\alpha\#\beta}(x)$ and we are done.

To prove the theorem it suffices to show that $f^\alpha(x) \leq f_{\omega^\alpha+1}(x)$. This is proved by induction on α. Assume that $f^\alpha(x) = f^\beta(f^\beta(x))$ with $\beta < \alpha$ and $N\beta \leq f(N\alpha + x)$. The induction hypothesis yields $f^\beta(f^\beta(x)) \leq f_{\omega^\beta+1}(f_{\omega^\beta+1}(x))$. Then

$$\begin{aligned} f^\beta(f^\beta(x)) &\leq f_{\omega^\beta+1}(f_{\omega^\beta+1}(x)) \\ &\leq f_{\omega^\beta+1}(f_{\omega^\beta+1}(f(N\alpha + x))) \\ &\leq f_{\omega^\alpha}(f(N\alpha + x)) \\ &\leq f_{\omega^\alpha}(f_{\omega^\alpha}(x)) \\ &\leq f_{\omega^\alpha+1}(x), \end{aligned}$$

where we made use of $f(N\alpha + x) \leq f_\alpha(x)$. The last claim is again proved by induction on α. For the induction step for proving this claim, note that $f_\alpha(x) = f_\beta(f_\beta(x))$ for some $\beta < \alpha$ with $N\beta = N\alpha + x$. (It is easily seen that here = has to hold.) Then $f_\beta(f_\beta(x)) \geq f_\beta(x) \geq f(N\beta + x) \geq f(N\alpha + x + x) \geq f(N\alpha + x)$. □

References

[1] Lars Kristiansen. Subrecursive degrees and fragments of Peano arithmetic. *Archive for Mathematical Logic*, 40:365–397, 2001.

[2] Andreas Weiermann. Classifying the provably total functions of PA. *Bulletin of Symbolic Logic*, 12(2):177–190, 2006.

Slow consistency

Sy-David Friedman[*], **Michael Rathjen**[†], **Andreas Weiermann**[‡]

[*] Kurt Gödel Research Center for Mathematical Logic, Universität Wien, Austria
sdf@logic.univie.ac.at

[†] Department of Pure Mathematics, University of Leeds, UK
rathjen@maths.leeds.ac.uk

[‡] Vakgroep Zuivere Wiskunde en Computeralgebra, Universiteit Gent, Belgium
Andreas.Weiermann@ugent.be

Abstract. The fact that "natural" theories, i.e., theories which have something like an "idea" to them, are almost always linearly ordered with regard to logical strength has been called one of the great mysteries of the foundation of mathematics. However, one easily establishes the existence of theories with incomparable logical strengths using self-reference (Rosser-style). As a result, **PA** + Con(**PA**) is not the least theory whose strength is greater than that of **PA**. But still we can ask: is there a sense in which **PA** + Con(**PA**) is the least "natural" theory whose strength is greater than that of **PA**? In this paper we exhibit natural theories in strength strictly between **PA** and **PA** + Con(**PA**) by introducing a notion of slow consistency.

1 Preliminaries

PA is Peano Arithmetic. $\mathbf{PA}\restriction_k$ denotes the subtheory of **PA** usually denoted by $\mathrm{I}\Sigma_k$. It consists of a finite base theory $\mathbf{P}^-$ (which are the axioms for a commutative discretely ordered semiring) together with a single Π_{k+2} axiom which asserts that induction holds for Σ_k formulae. For functions $F\colon \mathbb{N}\to\mathbb{N}$ we use exponential notation $F^0(x)=x$ and $F^{k+1}(x)=F(F^k(x))$ to denote repeated compositions of F.

In what follows we require an ordinal representation system for ε_0. Moreover, we assume that these ordinals come equipped with specific fundamental sequences $\lambda[n]$ for each limit ordinal $\lambda \le \varepsilon_0$. Their definition springs forth from their representation in Cantor normal form (to base ω). For an ordinal α such that $\alpha > 0$, α has a unique representation

$$\alpha = \omega^{\alpha_1}\cdot n_1 + \cdots + \omega^{\alpha_k}\cdot n_k,$$

where $0 < k, n_1, \dots, n_k < \omega$, and $\alpha_1, \dots, \alpha_k$ are ordinals such that $\alpha_1 > \cdots > \alpha_k$.

If the Cantor normal form of $\beta > 0$ is $\omega^{\beta_1}\cdot m_1 + \cdots + \omega^{\beta_l}\cdot m_l$, we write $\alpha \gg \beta$ if $\alpha > \beta$ and $\alpha_k \ge \beta_1$.

Definition 1.1 For α an ordinal and n a natural number, let ω_n^α be defined inductively by $\omega_0^\alpha := \alpha$ and $\omega_{n+1}^\alpha := \omega^{\omega_n^\alpha}$.

We also write ω_n for ω_n^1. In particular, $\omega_0 = 1$ and $\omega_1 = \omega$.

[*]The research of all authors was supported by Templeton Foundation Grant #13152, the CRM Infinity Project. The first author also wishes to thank the Austrian Science Fund for its support through research project P22430-N13.

[†]The second author's research was also supported by U.K. EPSRC grant No. EP/G029520/1.

Definition 1.2 For each limit ordinal $\lambda \le \varepsilon_0$, define a strictly monotone sequence $\lambda[n]$ of ordinals converging to λ from below. We use the fact, following from the Cantor normal form representation, that if $0 < \alpha < \varepsilon_0$ then there are unique $\beta, \gamma < \varepsilon_0$ and $0 < m < \omega$ such that

$$\alpha = \beta + \omega^\gamma \cdot m,$$

and either $\beta = 0$ or β has normal form $\omega^{\beta_1} \cdot m_1 + \cdots + \omega^{\beta_l} \cdot m_l$ with $\beta_l > \gamma$.

The definition of $\lambda[n]$ proceeds by recursion on this representation of λ.

Case 1: $\lambda = \beta + \omega^\gamma \cdot m$ and $\gamma = \delta + 1$. Put $\lambda[n] = \beta + \omega^\gamma \cdot (m-1) + \omega^\delta \cdot (n+1)$. (Remark: In particular, $\omega[n] = n+1$.)

Case 2: $\lambda = \beta + \omega^\gamma \cdot m$, and $\gamma < \lambda$ is a limit ordinal. Put $\lambda[n] = \beta + \omega^\gamma \cdot (m-1) + \omega^{\gamma[n]}$.

Case 3: $\lambda = \varepsilon_0$. Put $\varepsilon_0[0] = \omega$ and $\varepsilon_0[n+1] = \omega^{\varepsilon_0[n]}$. (Remark: Thus $\varepsilon_0[n] = \omega_{n+1}$.)

It will be convenient to have $\alpha[n]$ defined for non-limit α. We set $(\beta+1)[n] = \beta$ and $0[n] = 0$.

Definition 1.3 By "a fast growing" hierarchy we simply mean a transfinitely extended version of the Grzegorczyk hierarchy, i.e., a transfinite sequence sequence of number-theoretic functions $F_\alpha \colon \mathbb{N} \to \mathbb{N}$ defined recursively by iteration at successor levels and diagonalization over fundamental sequences at limit levels. We use the hierarchy

$$\begin{aligned} F_0(n) &= n+1 \\ F_{\alpha+1}(n) &= F_\alpha^{n+1}(n) \\ F_\alpha(n) &= F_{\alpha[n]}(n) \quad \text{if } \alpha \text{ is a limit.} \end{aligned}$$

It is closely related to the Hardy hierarchy:

$$\begin{aligned} H_0(n) &= n \\ H_{\alpha+1}(n) &= H_\alpha(n+1) \\ H_\alpha(n) &= H_{\alpha[n]}(n) \quad \text{if } \alpha \text{ is a limit.} \end{aligned}$$

Their relationship is as follows:

$$H_{\omega^\alpha} = F_\alpha \tag{1.1}$$

for every $\alpha < \varepsilon_0$. If $\alpha = \omega^{\alpha_1} \cdot n_1 + \cdots + \omega^{\alpha_k} \cdot n_k$ is in Cantor normal form and $\beta < \omega^{\alpha_k + 1}$, then

$$H_{\alpha+\beta} = H_\alpha \circ H_\beta. \tag{1.2}$$

Ketonen and Solovay [**6**] found an interesting combinatorial characterization of the H_α's. Call an interval $[k, n]$ 0-*large* if $k \le n$, $(\alpha+1)$-*large* if there are $m, m' \in [k, n]$ such that $m \ne m'$ and $[m, n]$ and $[m', n]$ are both α-large; and λ-*large* (where λ is a limit) if $[k, n]$ is $\lambda[k]$-large.

Theorem 1.4 (Ketonen–Solovay [**6**]) *Let $\alpha < \varepsilon_0$. Then*

$$\begin{aligned} H_\alpha(n) &= \textit{least } m \textit{ such that } [n, m] \textit{ is } \alpha\textit{-large;} \\ F_\alpha(n) &= \textit{least } m \textit{ such that } [n, m] \textit{ is } \omega^\alpha\textit{-large.} \end{aligned}$$

The order of growth of F_{ε_0} is essentially the same as that of the Paris–Harrington function F_{PH}.

Definition 1.5 Let X be a finite set of natural numbers and $|X|$ be the number of elements in X. Then X is *large* if X if X is non-empty, and, letting s be the least element of X, X has at least s elements. If $d \in \mathbb{N}$ then $[X]^d$ denotes the set of all subsets of X of cardinality d. If $g\colon [X]^d \to Y$, a subset Z of Y is *homogeneous* for g if g is constant on $[Z]^d$. Identify $n \in \mathbb{N}$ with the set $\{0, \ldots, n-1\}$.

Let $a, b, c \in \mathbb{N}$. Then $a \to (\text{large})^b_c$ if for every map $g\colon [a]^b \to c$ there is a large homogeneous set for g of cardinality greater than b.

Let $\sigma(b, c)$ be the least integer a such that $a \to (\text{large})^b_c$ and $f_{PH}(n) = \sigma(n, n)$.

Theorem 1.6

(i) (Harrington–Paris [**13**]) *The function* f_{PH} *dominates all* **PA**-*provably recursive functions.*

(ii) (Ketonen–Solovay [**6**]) *For* $n \geq 20$,

$$F_{\varepsilon_0}(n-3) \leq \sigma(n, 8) \leq F_{\varepsilon_0}(n-2);$$
$$f_{PH}(n) \leq F_{\varepsilon_0}(n-1).$$

2 Capturing the F_α's in PA

In [**6**] many facts about the functions F_α, as befits their definition, are proved by transfinite induction on the ordinals $\leq \varepsilon_0$. In [**6**] there is no attempt to determine whether they are provable in **PA** (let alone in weaker theories). In what follows we will have to assume that some of the properties of the F_α's hold in all models of **PA**. As a consequence, we will revisit some parts of [**6**], especially section 2, and recast them in such a way that they become provable in **PA**. Statements shown by transfinite on the ordinals in [**6**] will be proved by ordinary induction on the term complexity of ordinal representations, adding extra assumptions.

Definition 2.1 The computation of $F_\alpha(x)$ is closely connected with the step-down relations of [**6**] and [**14**]. For $\alpha < \beta \leq \varepsilon_0$ we write $\beta \underset{n}{\longrightarrow} \alpha$ if for some sequence of ordinals $\gamma_0, \ldots, \gamma_r$ we have $\gamma_0 = \beta$, $\gamma_{i+1} = \gamma_i[n]$, for $0 \leq i < r$, and $\gamma_r = \alpha$. If we also want to record the number of steps r, we shall write $\alpha \overset{r}{\underset{n}{\longrightarrow}} \beta$.

The definition of the functions F_α for $\alpha \leq \varepsilon_0$ employs transfinite recursion on α. It is therefore not immediately clear how we can speak about these functions in arithmetic. Later on we shall need to refer to a definition of $F_\alpha(x) = y$ in an arbitrary model of **PA**. As it turns out, this can be done via a formula of low complexity.

Lemma 2.2 *There is a* Δ_0*-formula expressing* $F_\alpha(x) = y$ *(as a predicate of* α, x, y*).*

Proof. This is shown in [**17**, 5.2]. The main idea is that the computation of $F_\alpha(x)$ can be described as a rewrite systems, that is, as a sequence of manipulations of expressions of the form

$$F^{n_1}_{\alpha_1}(F^{n_2}_{\alpha_2}(\ldots(F^{n_k}_{\alpha_k}(n))\ldots)),$$

where $n_1, \ldots, n_k \in \omega - \{0\}$ and $\alpha_1 > \ldots > \alpha_k \geq 0$. □

Let IΔ_0 be the subsystem of Peano Arithmetic in which induction applies only to formulas with bounded quantifiers (Δ_0-formulas). If we add to IΔ_0 the axiom

$$\exp = \forall x > 1\, \forall y\, \exists z\, E_0(x, y, z),$$

saying that the exponential function is total, then the resulting theory will be denoted by IΔ_0(exp). IΔ_0(exp) is strong enough to prove all of the results of elementary number theory. For example, Matijasevic's Theorem is provable in it.

Lemma 2.3 *We use $F_\alpha(x)\downarrow$ to denote $\exists y\, F_\alpha(x) = y$, and $F_\alpha\downarrow$ stands for $\forall x\, F_\alpha(x)\downarrow$. The following are provable in* IΔ_0(exp)*:*

(i) *If $\beta \,\mathrm{sd}\, x\alpha$ and $F_\beta(x)\downarrow$, then $F_\alpha(x)\downarrow$ and $F_\beta(x) \geq F_\alpha(x)$.*
(ii) *If $F_\beta(x)\downarrow$ and $x > y$, then $F_\beta(y)\downarrow$ and $F_\beta(x) \geq F_\beta(y)$.*
(iii) *If $\alpha > \beta$ and $F_\alpha\downarrow$, then $F_\beta\downarrow$.*
(iv) *If $i > 0$ and $F^i_\alpha(x)\downarrow$ then $x < F^i_\alpha(x)$.*

Proof. Part (i) follows by induction on the length r of the sequence $\gamma_0, \ldots, \gamma_r$ with $\gamma_0 = \beta$, $\gamma_{i+1} = \gamma_i[n]$, for $0 \leq i < r$, and $\gamma_r = \alpha$. In the proof one uses the fact that '$F_\delta(x) = y$' is Δ_0 as a relation with arguments δ, x, y, and also uses [**17**, Theorem 5.3] (or rather Claim 1 in Appendix A of [**16**]).

Part (ii) follows from [**17**, Proposition 5.4(v)]; (iii) follows from [**17**, Proposition 5.4(iv)], and (iv) is [**17**, Proposition 5.4(i)]. □

There is an additional piece of information that is provided by the particular coding and Δ_0 formula denoting $F_\alpha(x) = y$ used in [**17**, 5.2], namely that there is a fixed polynomial P in one variable such that for all $\alpha \leq \varepsilon_0$, the number of steps it takes to compute $F_\alpha(x)$ is always bounded by $P(F_\alpha(x))$. This has a useful consequence that we are going to exploit in the next lemma.

Lemma 2.4 *The following is provable in* IΔ_0(exp)*: Let $\alpha \leq \varepsilon_0$, and suppose $F_\alpha(n)\downarrow$. Then $\alpha \xrightarrow[n]{r} 0$ for some $r \leq P(F_\alpha(n))$.*

Proof. We clearly have that the number of steps it takes to compute $F_\alpha(n)$ is a bound for any sequence of ordinals $\gamma_0, \ldots, \gamma_s$ with $\gamma_0 = \alpha$, $\gamma_s > 0$, and $\gamma_{i+1} = \gamma_i[n]$ for $0 \leq i < s$. Hence $s < P(F_\alpha(n))$ and thus $\alpha \xrightarrow[n]{r} 0$ for some $r \leq P(F_\alpha(n))$. □

Convention For the remainder of this section we will be working in the background theory **PA**; thus all statements are formally provable in **PA**. A cursory glance would reveal that the fragment IΣ_1 is certainly capacious enough, and very likely IΔ_0(exp) would suffice, too.

Lemma 2.5

(i) *Let $\alpha \xrightarrow[n]{} \beta$, $\alpha \xrightarrow[n]{} \gamma$, $\beta > \gamma$. Then $\beta \xrightarrow[n]{} \gamma$.*
(ii) *Let $\alpha \xrightarrow[n]{} \beta$, $\beta \xrightarrow[n]{} \gamma$. Then $\alpha \xrightarrow[n]{} \gamma$. Then $\alpha \xrightarrow[n]{} \gamma$.*

Proof. This is evident from the definition. □

Definition 2.6 Let α, β be ordinals. Say that α *meshes with* β if, for some ordinals γ, δ, we have $\alpha = \omega^\gamma \cdot \delta$ and $\beta < \omega^{\gamma+1}$.

Note that if α and β have Cantor normal forms $\alpha = \omega^{\alpha_1} \cdot n_1 + \cdots + \omega^{\alpha_k} \cdot n_k$, $\beta = \omega^{\beta_1} \cdot m_1 + \cdots + \omega^{\beta_l} \cdot m_l$, respectively, then the condition that α meshes with β is precisely that $\alpha_k \geq \beta_1$.

Lemma 2.7 *Let $\alpha, \beta < \varepsilon_0$. Let α mesh with $\beta > 0$. Then $(\alpha + \beta)[n] = \alpha + \beta[n]$. Thus if $\beta \xrightarrow[n]{} \gamma$, then $\alpha + \beta \xrightarrow[n]{} \alpha + \gamma$.*

Proof. That α meshes with β implies that the Cantor normal form of $\alpha+\beta$ is basically the concatenation of those for α, β. The first claim thus follows from the way that the definition of $\delta[n]$ focuses on the rightmost term of the Cantor normal form of δ, provided $\delta < \varepsilon_0$. The second claim reduces to the special case when $\gamma = \beta[n]$, using the transitivity of $\xrightarrow[n]{}$. This special claim is evident by the first claim. □

Lemma 2.8 *Let $k < l < \omega$, $\alpha < \varepsilon_0$, and suppose that $\omega^\alpha \cdot l \xrightarrow[n]{} 0$. Then $\omega^\alpha \cdot l \xrightarrow[n]{} \omega^\alpha \cdot k$.*

Proof. This holds by assumption if $k = 0$. So suppose that $n > 0$. Let $\omega^\alpha \cdot k < \delta \leq \omega^\alpha \cdot l$. Then δ can be uniquely written as $\delta = \omega^\alpha \cdot k + \gamma$ for some $\gamma > 0$, and $\omega^\alpha \cdot k$ and γ mesh. Thus it follows from Lemma 2.7 that $\delta[n] = \omega^\alpha \cdot k + \gamma[n]$ and hence $\delta[n] \geq \omega^\alpha \cdot k$. Since $\omega^\alpha \cdot l \xrightarrow[n]{} 0$, we conclude that $\omega^\alpha \cdot l \xrightarrow[n]{} \omega^\alpha \cdot k$. □

Lemma 2.9 *Let $n \geq 1$. Let $\delta < \varepsilon_0$. Suppose $\omega^{\delta+1} \xrightarrow[n]{} 0$. Then $\omega^{\delta+1} \xrightarrow[n]{} \omega^\delta$.*

Proof. $\omega^{\delta+1} \xrightarrow[n]{} \omega^{\delta+1}[n] = \omega^\delta \cdot (n+1)$. Now apply Lemma 2.8 and Lemma 2.5(ii). □

Lemma 2.10 *Let $\alpha_1 < \varepsilon_0$. Let $n \geq 1$. Suppose that $\alpha_1 \xrightarrow[n]{} \alpha_2$ and $\omega^{\alpha_1} \xrightarrow[n]{} 0$. Then $\omega^{\alpha_1} \xrightarrow[n]{} \omega^{\alpha_2}$.*

Proof. Let $\alpha_1 \xrightarrow[n]{x} \alpha_2$. By induction on x we show that $\omega^{\alpha_1} \xrightarrow[n]{} \omega^{\alpha_2}$.

If $x = 0$ this is trivial. Suppose $x > 0$. If α_1 is a successor $\alpha_0 + 1$, then $\alpha_1[n] = \alpha_0 \xrightarrow[n]{x-1} \alpha_2$ and thus $\alpha^{\alpha_0} \xrightarrow[n]{} \omega^{\alpha_2}$ by the induction hypothesis. Also $\omega^{\alpha_1}[n] = \omega^{\alpha_0} \cdot (n+1)$ and $\omega^{\alpha_0} \cdot (n+1) \xrightarrow[n]{} \omega^{\alpha_0}$ owing to Lemma 2.8. Consequently, $\omega^{\alpha_1} \xrightarrow[n]{} \omega^{\alpha_2}$.

Now let α_1 be a limit. Then $\omega^{\alpha_1}[n] = \omega^{\alpha_1[n]}$. Inductively, as $\alpha_1[n] \xrightarrow[n]{x-1} \alpha_2$, we have that $\omega^{\alpha_1[n]} \xrightarrow[n]{} \omega^{\alpha_2}$. Hence $\omega^{\alpha_1} \xrightarrow[n]{} \omega^{\alpha_2}$. □

Lemma 2.11 *Let $\alpha < \varepsilon_0$. Suppose $\omega^\alpha \xrightarrow[n]{x} 0$. Then $\alpha \xrightarrow[n]{y} 0$ for some $y < x$.*

Proof. We proceed by induction on x. If $\alpha = 0$ then this is obvious. Let $\alpha = \alpha_0 + 1$. Then $\omega^\alpha[n] = \omega^{\alpha_0} \cdot n + \omega^{\alpha_0} \xrightarrow[n]{x-1} 0$. In light of Lemma 2.7 we conclude that $\omega^{\alpha_0} \xrightarrow[n]{u} 0$ for some $u \leq x-1$. Thus, by the inductive assumption, $\alpha_0 \xrightarrow[n]{v} 0$ for some $v < x-1$. Therefore $\alpha \xrightarrow[n]{v+1} 0$ with $v+1 < x$.

Now let α be a limit. Then $\omega^\alpha[n] = \omega^{\alpha[n]} \xrightarrow[n]{x-1} 0$. Inductively we thus have $\alpha[n] \xrightarrow[n]{u} 0$ for some $u < x-1$, and hence $\alpha \xrightarrow[n]{u+1} 0$ where $u+1 < x$. □

Proposition 2.12 *Let λ be a limit $\leq \varepsilon_0$. Suppose $i < j < \omega$ and $\lambda[j] \xrightarrow[n]{} 0$. Then $\lambda[j] \xrightarrow[n]{} \lambda[i]$.*

Proof. We proceed by induction on the (term) complexity of λ.

Case 1: $\lambda = \beta + \omega^{\alpha+1} \cdot m$. Then $\lambda[k] = \beta + \omega^{\alpha+1} \cdot (m-1) + \omega^\alpha \cdot (k+1)$. As $\lambda[j] \xrightarrow[n]{} 0$ entails that $\omega^\alpha \cdot (j+1) \xrightarrow[n]{} 0$, it follows from Lemma 2.8 that $\omega^\alpha \cdot (j+1) \xrightarrow[n]{} \omega^\alpha \cdot (i+1)$. But then, by Lemma 2.7,

$$\lambda[j] = \beta + \omega^{\alpha+1} \cdot (m-1) + \omega^\alpha \cdot (j+1) \xrightarrow[n]{} \beta + \omega^{\alpha+1} \cdot (m-1) + \omega^\alpha \cdot (i+1) = \lambda[i].$$

Case 2: $\lambda = \beta + \omega^\gamma \cdot m$, and γ is a limit ordinal. Then $\lambda[k] = \beta + \omega^\gamma \cdot (m-1) + \omega^{\gamma[k]}$. $\lambda[j] \xrightarrow[n]{} 0$ implies that $\omega^{\gamma[j]} \xrightarrow[n]{} 0$, and hence, by Lemma 2.11, $\gamma[j] \xrightarrow[n]{} 0$. Since the term complexity of γ is smaller than that of λ, the inductive assumption yields $\gamma[j] \xrightarrow[n]{} \gamma[i]$, and hence $\omega^{\gamma[j]} \xrightarrow[n]{} \omega^{\gamma[i]}$ by Lemma 2.10. As a result, by Lemma 2.7,

$$\lambda[j] = \beta + \omega^\gamma \cdot (m-1) + \omega^{\gamma[j]} \xrightarrow[n]{} \beta + \omega^\gamma \cdot (m-1) + \omega^{\gamma[i]} = \lambda[i].$$

Case 3: $\lambda = \varepsilon_0$. Then $\lambda[j] = \omega_{j+1} = \omega^{\omega_j}$. From the assumption $\lambda[j] \xrightarrow[n]{} 0$, applying Lemma 2.11 iteratively, one deduces that $\omega_k \xrightarrow[n]{} 0$ holds for all $k \leq j+1$. Obviously, $\omega \xrightarrow[n]{} 1$. Thus, by Lemma 2.10, $\omega_2 = \omega^\omega \xrightarrow[n]{} \omega^1 = \omega = \omega_1$. Iterating this procedure we have $\omega_{l+1} \xrightarrow[n]{} \omega_l$ for all $l \leq j$. By transitivity of $\xrightarrow[n]{}$ we thus arrive at $\lambda[j] = \omega_{j+1} \xrightarrow[n]{} \omega_{i+1} = \lambda[j]$. □

Lemma 2.13 *Let $n, k < \omega$ and $n > 0$. Suppose $\omega_{k+1} \xrightarrow[n]{} 0$. Then $\omega_{k+1} \xrightarrow[n]{} \omega_k + 1$.*

Proof. From the proof of Proposition 2.12, Case 3, we infer that $\omega_{u+1} \xrightarrow[n]{} 0$ for all $u \leq k$. Now use induction on $u \leq k$ to show that $\omega_{u+1} \xrightarrow[n]{} \omega_u + 1$. If $u = 0$ then $\omega_u = 1$ and $\omega_{u+1} = \omega$, and $\omega \xrightarrow[n]{} 2$ holds since $n \geq 1$. Now suppose $u = v+1$ and $\omega_{v+1} \xrightarrow[n]{} \omega_v + 1$. Then, as $\omega_{u+1} \xrightarrow[n]{} 0$, we have

$$\omega_{u+1} = \omega^{\omega_{v+1}} \xrightarrow[n]{} \omega^{\omega_v + 1} \tag{2.1}$$

by applying Lemma 2.10. In particular, $\omega^{\omega_v+1} \xrightarrow[n]{} 0$, and therefore

$$\omega^{\omega_v+1}[n] = \omega^{\omega_v} \cdot (n+1) = \omega_{v+1} \cdot (n+1) \xrightarrow[n]{} \omega_{v+1} + \omega_{v+1} \tag{2.2}$$

since $n > 0$. Since we also have $\omega_{v+1} \xrightarrow[n]{} \omega_0 = 1$ by Proposition 2.12, (2.2) implies

$$\omega^{\omega_v+1} \xrightarrow[n]{} \omega_{v+1} + 1. \tag{2.3}$$

Combining (2.1) and (2.3) yields $\omega_{u+1} \xrightarrow[n]{} \omega_u + 1$. □

Corollary 2.14 *Let $k, n < \omega$ and $n > 0$.*

(i) *Suppose $\varepsilon_0[k+1] \xrightarrow[n]{} 0$. Then $\varepsilon_0[k+1] \xrightarrow[n]{} \varepsilon_0[k] + 1$.*
(ii) *Suppose $F_{\varepsilon_0[k+1]}(n)\downarrow$. Then $F_{\varepsilon_0[k+1]}(n) \geq F_{\varepsilon_0[k]}(F_{\varepsilon_0[k]}(n))$.*

Proof. As $\varepsilon_0[u] = \omega_{u+1}$, (i) is a consequence of Lemma 2.13. We next prove (ii). By Lemma 2.4, $F_{\varepsilon_0[k+1]}(n)\downarrow$ implies that $\varepsilon_0[k+1] \xrightarrow[n]{} 0$. Thus, using (i), we may infer that $\varepsilon_0[k+1] \xrightarrow[n]{} \varepsilon_0[k] + 1$. Hence, by Lemma 2.3(i),

$$F_{\varepsilon_0[k+1]}(n) \geq F_{\varepsilon_0[k]+1}(n) = F^{n+1}_{\varepsilon_0[k]}(n) \geq F_{\varepsilon_0[k]}(F_{\varepsilon_0[k]}(n)),$$

where the last inequality is a consequence of Lemma 2.3(iv). □

3 Slow consistency

To motivate our notion of slow consistency we recall the concept of interpretability of one theory in another theory. Let S and S' be arbitrary theories. S' is *interpretable* in S or S *interprets* S' (in symbols $S' \lhd S$) "if, roughly speaking, the primitive concepts and the range of the variables of S' are defined in such a way as to turn every theorem of S' into a theorem of S" (quoted from [**10**, p. 96]; for details, see [**10**, Section 6]).

To simplify matters, we restrict attention to theories T formulated in the language of **PA** which contain the axioms of **PA** and have a primitive recursive axiomatization, i.e., the axioms are enumerated by such a function. For an integer $k \geq 0$, we denote by $T \restriction_k$ the theory consisting of the first k axioms of T. Let $\mathrm{Con}(T)$ be the arithmetized statement that T is consistent.

A theory T is *reflexive* if it proves the consistency of all its finite subtheories, i.e., $T \vdash \mathrm{Con}(T\restriction_k)$ for all $k \in \mathbb{N}$. Note that theories satisfying the conditions spelled out above will always be reflexive.

Another interesting relationship between theories we shall consider is $T_1 \subseteq_{\Pi^0_1} T_2$, i.e., every Π^0_1 theorem of T_1 is also a theorem of T_2.

Theorem 3.1 *Let S, T be theories that satisfy the conditions spelled out above. Then:*

$$S \lhd T \text{ if and only if } T \vdash \mathrm{Con}(S\restriction_n) \text{ holds for all } n \in \mathbb{N} \tag{3.1}$$

$$\text{if and only if } S \subseteq_{\Pi^0_1} T. \tag{3.2}$$

Proof. Fact (3.1) seems to be due to Orey [**11**]. Another easily accessible proof of (3.1) can be found in [**10**, Section 6, Theorem 5]. Fact (3.2) was first stated in [**5**] and [**9**]. A proof can also be found in [**10**, Section 6, Theorem 6]. □

We know that

$$\mathrm{Con}(\mathbf{PA}) \leftrightarrow \forall x\, \mathrm{Con}(\mathbf{PA}\restriction_x).$$

Given a function $f\colon \mathbb{N} \to \mathbb{N}$ (say provably total in **PA**) we are thus led to the following consistency statement:

$$\mathrm{Con}_f(\mathbf{PA}) := \forall x\, \mathrm{Con}(\mathbf{PA}\restriction_{f(x)}). \tag{3.3}$$

It is perhaps worth pointing out that the exact meaning of $\mathrm{Con}_f(\mathbf{PA})$ depends on the representation that we choose for f.

Statements of the form (3.3) are interesting only if the function f grows extremely slowly, though still has an infinite range but **PA** cannot prove that fact.

Definition 3.2 Define

$$F^{-1}_{\varepsilon_0}(n) = \max(\{k \leq n \mid \exists y \leq n\, F_{\varepsilon_0}(k) = y\} \cup \{0\}).$$

Note that, by Lemma 2.2, the graph of $F^{-1}_{\varepsilon_0}$ has a Δ_0 definition. Thus it follows that $F^{-1}_{\varepsilon_0}$ is a provably recursive function of **PA**.

Let $\mathrm{Con}^*(\mathbf{PA})$ be the statement $\forall x\, \mathrm{Con}\Big(\mathbf{PA}\restriction_{F^{-1}_{\varepsilon_0}(x)}\Big)$. Of course, in the definition of $\mathrm{Con}^*(\mathbf{PA})$ we have in mind some standard representation of F_{ε_0} referred to in Lemma 2.2. Note that $\mathrm{Con}^*(\mathbf{PA})$ is equivalent to the statement

$$\forall x\, [F_{\varepsilon_0}(x)\downarrow \rightarrow \mathrm{Con}(\mathbf{PA}\restriction_x)].$$

Proposition 3.3 $\mathbf{PA} \nvdash \mathrm{Con}^*(\mathbf{PA})$.

Proof. Aiming at a contradiction, suppose $\mathbf{PA} \vdash \mathrm{Con}^*(\mathbf{PA})$. Then $\mathbf{PA} \restriction_k \vdash \mathrm{Con}^*(\mathbf{PA})$ for all sufficiently large k. As $\mathbf{PA} \restriction_k \vdash F_{\varepsilon_0}(k) \downarrow$ on account of $F_{\varepsilon_0}(k) \downarrow$ being a true Σ_1 statement, we arrive at $\mathbf{PA} \restriction_k \vdash \mathrm{Con}(\mathbf{PA} \restriction_k)$, contradicting Gödel's second incompleteness theorem. □

Proposition 3.3 holds in more generality.

Corollary 3.4 *If T is a recursive consistent extension of* $\mathbf{PA}$ *and f is a total recursive function with unbounded range, then*

$$T \nvdash \forall x \, \mathrm{Con}(T \restriction_{f(x)})$$

where $f(x) \downarrow$ is understood to be formalized via some Σ_1 representation of f.

Proof. Basically the same proof as for Proposition 3.3. □

It is quite natural to consider another version of slow consistency where the function $f\colon \mathbb{N} \to \mathbb{N}$, rather than acting as a bound on the fragments of $\mathbf{PA}$, restricts the lengths of proofs. Let $\perp$ be a Gödel number of the canonical inconsistency and let $\mathrm{Proof}_{\mathbf{PA}}(y, z)$ be the primitive recursive predicate expressing the concept that "*y is the Gödel number of a proof in* $\mathbf{PA}$ *of a formula with Gödel number z*".

$$\mathrm{Con}^{\ell}_{f}(\mathbf{PA}) := \forall x \, \forall y < f(x) \, \neg \mathrm{Proof}_{\mathbf{PA}}(y, \perp) \tag{3.4}$$

Let $\mathrm{Con}^{\#}(\mathbf{PA})$ be the statement $\mathrm{Con}^{\ell}_{F^{-1}_{\varepsilon_0}}(\mathbf{PA})$.

Note that $\mathrm{Con}^{\#}(\mathbf{PA})$ is equivalent to the following formula:

$$\forall u \, [F_{\varepsilon_0}(u) \downarrow \rightarrow \ \forall y < u \, \neg \mathrm{Proof}_{\mathbf{PA}}(y, \perp)].$$

As it turns out, by contrast with $\mathrm{Con}^*(\mathbf{PA})$, $\mathrm{Con}^{\#}(\mathbf{PA})$ is not very interesting.

Lemma 3.5 $\mathbf{PA} \vdash \mathrm{Con}^{\#}(\mathbf{PA})$.

Proof. Recall that Gentzen showed how to effectively transform an alleged $\mathbf{PA}$-proof of an inconsistency (the empty sequent) in his sequent calculus into another proof of the empty sequent such that the latter gets assigned a smaller ordinal than the former. More precisely, there is a reduction procedure $\mathcal{R}$ on proofs P of the empty sequent together with an assignment ord of representations for ordinals $< \varepsilon_0$ to proofs such that $\mathrm{ord}(\mathcal{R}(P)) < \mathrm{ord}(P)$. Here $<$ denotes the ordering on ordinal representations induced by the ordering of the pertaining ordinals. The functions $\mathcal{R}$ and ord and the relation $<$ are primitive recursive (when viewed as acting on codes for the syntactic objects). With $g(n) = \mathrm{ord}(\mathcal{R}^n(P))$, the n-fold iteration of $\mathcal{R}$ applied to P, one has

$$g(0) < g(1) < g(2) < \cdots < g(n)$$

for all n, which is absurd as the ordinals are well-founded.

We will now argue in $\mathbf{PA}$. Suppose that $F_{\varepsilon_0}(u) \downarrow$. Aiming at a contradiction, assume that there is a $p < u$ such that $\mathrm{Proof}_{\mathbf{PA}}(p, \perp)$. We have not said anything about the particular proof predicate $\mathrm{Proof}_{\mathbf{PA}}$ we use; however, whatever proof system is assumed, p will be larger than the Gödel numbers of all formulae occurring in the proof. The proof that p codes can be primitive recursively transformed into a sequent calculus proof P of the empty sequent in such a way that $\mathrm{ord}(P) < \omega_p$ since p is larger than the number of logical symbols occurring in any cut or induction formulae featuring in P (for details see [**18**, Ch. 2]). Inspection of Gentzen's proof, as presented e.g. in [**18**, 2.12.8], shows that there is a primitive recursive function ℓ such that the number of steps it takes to get from $\mathrm{ord}(P)$ to 0 by applying the reduction procedure $\mathcal{R}$ is majorized by $\ell(F_{\varepsilon_0}(u))$. As a result we have a contradiction, since there is no proof P_0 of the empty sequent with ordinal $\mathrm{ord}(P_0) = 0$.

The authors realize that the foregoing proof is merely a sketch. An alternative proof can be obtained by harking back to [**1**]. The reader will be assumed to have access to [**1**]. That paper uses an infinitary proof system with the ω-rule (of course). But this system is also quite peculiar in that the ordinal assignment adhered to is very rigid and, crucially, it has a so-called accumulation rule. To deal with infinite proofs in **PA**, though, one has to use primitive recursive proof trees instead of arbitrary ones (for details, see [**3**]). The role of the repetition rule (or trivial rule; cf. [**3**]) is of central importance to capturing the usual operations on proofs, such as inversion and cut elimination, by primitive recursive functions acting on their codes. In the proof system of [**1**] the accumulation rule takes over this role. Now assume that everything in [**1**] has been recast in terms of primitive recursive proof trees. Then the cut elimination for infinitary proofs with finite cut rank (as presented in [**3**, Theorem 2.19]) can be formalized in **PA**. Working in **PA**, suppose that $F_{\varepsilon_0}(u)\downarrow$. Aiming at a contradiction, assume there is a $p < u$ such that $\mathrm{Proof}_{\mathbf{PA}}(p, \perp)$. As above, the proof that p codes, can be primitive recursively transformed into a proof P of $\perp$ in the sequent calculus of [**1**] with ordinal ω_p and cut-degree 0 (in the sense of [**1**, Definition 5]). The plan is to reach a contradiction by constructing an infinite descending sequence of ordinals $(\alpha_i)_{i\in\mathbb{N}}$ such that $\alpha_0 = \omega_p$, $\alpha_{i+1} < \alpha_i$ and $\alpha_{i+1} <_{l_{i+1}} \alpha_i$ for some $l_{i+1} < F_{\omega_p}(2)$. It remains to determine $(\alpha_i)_{i\in\mathbb{N}}$. To this end, we construct a branch of the proof-tree P with $\vdash^{\alpha_i} \Delta_i, \Gamma_i$ being the i-th node of the branch (bottom-up). The sequent Γ_i contains only closed elementary prime formulas and formulas of the form $n \in N$, whereas Δ_i is of the form $\{n_1 \notin N, \ldots, n_r \notin N\}$ or $\emptyset$. We set

$$k_{\Delta_i} := \max\big(\{2\} \cup \{3\cdot n_1, \ldots, 3\cdot n_r\}\big)$$

in the former and $k_{\Delta_i} := 2$ in the latter case. We say that Γ_i is true in m if Γ_i is true when N is interpreted as the finite set $\{n \mid 3\cdot n < m\}$. Let $\Gamma_0 = \{0 = 1\}$ and $\Delta_0 = \emptyset$. Clearly, Γ_0 is false in $F_{\alpha_0}(2)$. Now assume $\vdash^{\alpha_i} \Delta_i, \Gamma_i$ has been constructed in such a way that $F_{\alpha_i}(k_{\Delta_i})\downarrow$ and Γ_i is false in $F_{\alpha_i}(k_{\Delta_i})$ and $F_{\alpha_i}(k_{\Delta_i}) \le F_{\alpha_0}(2)$. Since Γ_i is false in $F_{\alpha_i}(k_{\Delta_i})$ and $F_{\alpha_i}(k_{\Delta_i}) > k_{\Delta_i}$, it follows that Δ_i, Γ_i is not an axiom. Thus $\vdash^{\alpha_i} \Delta_i, \Gamma_i$ is not an end-node in P and therefore it is the result of an application of an inference rule. As the cut-rank of P is 0, the only possible rules are a cut of rank 0, an N-rule, and Accumulation.

If it is an N-rule, Γ_i contains "$Sn \in N$" for some n and $\vdash^{\beta} \Delta_i, \Gamma'_i, n \in N$ will be a node in P immediately above $\vdash^{\alpha_i} \Delta_i, \Gamma_i$ with $\Gamma'_i \subseteq \Gamma_i$ and $\beta + 1 = \alpha_i$. We let $\alpha_{i+1} = \beta$, $l_{i+1} = 1$, $\Delta_{i+1} = \Delta_i$ and $\Gamma_{i+1} = \Gamma_i, n \in N$. Since Γ_i is false in $F_{\alpha_i}(k_{\Delta_i})$ and $F_{\alpha_{i+1}}(k_{\Delta_i}) + 3 \le F_{\alpha_i}(k_{\Delta_i})$ it follows that Γ_{i+1} is false in $F_{\alpha_i}(k_{\Delta_{i+1}})$.

If the last rule is Accumulation, $\vdash^{\beta} \Delta_i, \Gamma_i$ will be a node in P immediately above $\vdash^{\alpha_i} \Delta_i, \Gamma_i$ for some $\beta <_{k_{\Delta_i}} \alpha_i$. Then let $\Delta_{i+1} = \Delta_i$, $\Gamma_{i+1} = \Gamma_i$, $\alpha_{i+1} = \beta$, and $l_{i+1} = k_{\Delta_i}$. Since $F_\beta(k_{\Delta_i}) \le F_{\alpha_i}(k_{\Delta_i})$, Γ_{i+1} is false in $F_{\alpha_{i+1}}(k_{\Delta_{i+1}})$, too. Inductively we also have $F_{\alpha_i}(k_{\Delta_i}) \le F_{\alpha_0}(2)$, and hence $l_{i+1} < F_{\alpha_0}(2)$.

If the last rule is a cut with a closed elementary prime formula A, the immediate nodes above $\vdash^{\alpha_i} \Delta_i, \Gamma_i$ in P are of the form $\vdash^{\beta} \Delta_i, \Gamma_i, A$ and $\vdash^{\beta} \Delta_i, \Gamma_i, \neg A$, respectively, where $\beta + 1 = \alpha_i$. Let $\Delta_{i+1} = \Delta_i$, $\alpha_{i+1} = \beta$, and $l_{i+1} = 1$. If A is false let $\Gamma_{i+1} = \Gamma_i, A$.

If A is true, let $\Gamma_{i+1} = \Gamma_i, \neg A$. Clearly, Γ_{i+1} will be false in $F_{\alpha_{i+1}}(k_{\Delta_{i+1}})$ since this value is smaller than $F_{\alpha_i}(k_{\Delta_i})$.

Finally suppose the last rule is a cut with cut formula "$n \in N$". Then the immediate nodes above $\vdash^{\alpha_i} \Delta_i, \Gamma_i$ in P are of the form $\vdash^{\beta} \Delta_i, n \in N, \Gamma_i$ and $\vdash^{\beta} \Delta_i, n \notin N, \Gamma_i$, respectively, where $\beta + 1 = \alpha_i$. Set $\alpha_{i+1} = \beta$ and and $l_{i+1} = 1$. If $F_\beta(k_{\Delta_i}) \leq 3 \cdot n$, then "$n \in N$" will be false in $F_\beta(k_{\Delta_i})$, and hence, as $F_\beta(k_{\Delta_i}) < F_{\alpha_i}(k_{\Delta_i})$, it follows that $n \in N, \Gamma_i$ will be false in $F_\beta(k_{\Delta_i})$ as well. So in this case let $\Delta_{i+1} = \Delta_i$ and $\Gamma_{i+1} = n \in N, \Gamma_i$.

If on the other hand $3 \cdot n < F_\beta(k_{\Delta_i})$, we compute that

$$F_\beta(k_{\Delta_i, n\notin N}) < F_\beta(F_\beta(k_{\Delta_i})) \leq F_{\alpha_i}(k_{\Delta_i}).$$

Hence Γ_i will be false in $F_\beta(k_{\Delta_i, n\notin N})$, and we put $\Delta_{i+1} = \Delta_i, n \notin N$ and $\Gamma_{i+1} = \Gamma_i$.

This finishes the definition of the $(\alpha_i)_{i\in\mathbb{N}}$. Their construction also guarantees that $F_{\alpha_i}(l_{i+1}) \downarrow$ and $F_{\alpha_{i+1}}(l_{i+1}) \leq F_{\alpha_i}(l_{i+1}) \leq F_{\omega_p}(2)$. Note also that whenever the inference involving $\vdash^{\alpha_{i+1}} \Delta_{i+1}, \Gamma_{i+1}$ as a premiss and $\vdash^{\alpha_i} \Delta_i, \Gamma_i$ as its conclusion was an application of a rule other than the Accumulation rule, then we have $\alpha_i = \alpha_{i+1} + 1$ and $l_{i+1} = 1$, and hence $F_{\alpha_{i+1}}(l_{i+1}) < F_{\alpha_i}(l_{i+1})$. As a result, there can only be finitely many of those. Hence there exists x_0 such that for $i \geq x_0$ the inference from $\vdash^{\alpha_{i+1}} \Delta_{i+1}, \Gamma_{i+1}$ to $\vdash^{\alpha_i} \Delta_i, \Gamma_i$ is always an instance of Accumulation. Furthermore, this entails that $\Delta_i, \Gamma_i = \Delta_j, \Gamma_j$ and $l_i = l_j$ for all $i, j > x_0$. Hence $\alpha_{i+1} <_k \alpha_i$ for all $i \geq x_0$ where $k = l_{x_0+1}$. However, this is absurd in view of Lemma 2.4 since then the computation of $F_{\alpha_{x_0}}(k)$ (i.e., $F_{\alpha_{x_0}}(l_{x_0+1})$) would never halt. □

The next goal will be to show that Con(**PA**) is not derivable in **PA** + Con*(**PA**). We need some preparatory definitions.

Definition 3.6 Let E denote the "stack of two's" function, $E(0) = 0$, $E(n+1) = 2^{E(n)}$. Given two elements a and b of a non-standard model $\mathfrak{M}$ of **PA**, we say that *'b is much larger than a'* if for every standard integer k we have $E^k(a) < b$.

If $\mathfrak{M}$ is a model of **PA** and $\mathfrak{I}$ is a substructure of $\mathfrak{M}$ we say that $\mathfrak{I}$ is an *initial segment* of $\mathfrak{M}$ if for all $a \in |\mathfrak{I}|$ and $x \in |\mathfrak{M}|$, $\mathfrak{M} \models x < a$ implies $x \in |\mathfrak{I}|$. We will write $\mathfrak{I} < b$ to mean $b \in |\mathfrak{M}| \setminus |\mathfrak{I}|$. Sometimes we write $a < \mathfrak{I}$ to indicate $a \in |\mathfrak{I}|$.

Theorem 3.7 *Let $\mathfrak{N}$ be a non-standard model of* **PA** *(or $\Delta_0(\exp)$), n be a standard integer, and $e, d \in |\mathfrak{N}|$ be non-standard such that $\mathfrak{N} \models F_{\omega_n^e}(e) = d$. Then there is an initial segment $\mathfrak{I}$ of $\mathfrak{N}$ such $e < \mathfrak{I} < d$ and $\mathfrak{I}$ is a model of Π_{n+1}-induction.*

Proof. This follows e.g. from [**17**, Theorem 5.25], letting $\alpha = 0$, $c = e$, $a = e$ and $b = d$. The technique used to prove Theorem 5.25 in [**17**] is a variation of techniques used by Paris in [**12**]. □

Corollary 3.8 *Let $\mathfrak{N}$ be a non-standard model of* **PA**, *$a, e, c \in |\mathfrak{N}|$ be non-standard such that $\mathfrak{N} \models F_{\varepsilon_0}(a) = e$ and $\mathfrak{N} \models F_{\varepsilon_0}(a+1) = c$. Then for every standard n there is an initial segment $\mathfrak{I}$ of $\mathfrak{N}$ such $e < \mathfrak{I} < c$ and $\mathfrak{I}$ is a model of Π_{n+1}-induction.*

Proof. We argue in $\mathfrak{N}$. From $F_{\varepsilon_0}(a+1) = F_{\varepsilon_0[a+1]}(a+1) = c$ we conclude with the help of Corollary 2.14 that

$$c \geq F_{\varepsilon_0[a]}(F_{\varepsilon_0[a]}(a+1)) \geq F_{\varepsilon_0[a]}(F_{\varepsilon_0[a]}(a)) = F_{\varepsilon_0[a]}(e) > e.$$

In view of the previous theorem we just have to ensure that $F_{\omega_n^e}(e) = d$ for some d with $d \leq c$. From $F_{\varepsilon_0[a]}(e) \downarrow$ we get $\varepsilon_0[a] \underset{e}{\rightarrow} 0$ by Lemma 2.4. Proposition 2.12 guarantees that $\varepsilon_0[p] \underset{e}{\rightarrow} e$ holds for all $p \leq a$. In particular, $\varepsilon_0[a-n] \underset{e}{\rightarrow} e$. Applying Lemma 2.10 n times, we arrive at

$$\varepsilon_0[a] = \omega_n^{\varepsilon_0[a-n]} \underset{e}{\rightarrow} \omega_n^e.$$

In view of Lemma 2.3(i) the latter implies that $F_{\omega_n^e}(e)\downarrow$ and $F_{\varepsilon_0[a]}(e) \geq F_{\omega_n^e}(e)$. □

Definition 3.9 Below we shall need the notion of two models $\mathfrak{M}$ and $\mathfrak{N}$ of **PA** 'agreeing up to e'. For this to hold, the following conditions must be met:

(1) e belongs to both models.
(2) e has the same predecessors in both $\mathfrak{M}$ and $\mathfrak{N}$.
(3) If d_0, d_1, and c are $\leq e$ (in one of the models $\mathfrak{M}$ and $\mathfrak{N}$), then $\mathfrak{M} \models d_0 + d_1 = c$ iff $\mathfrak{N} \models d_0 + d_1 = c$.
(4) If d_0, d_1, and c are $\leq e$ (in one of the models $\mathfrak{M}$ and $\mathfrak{N}$), then $\mathfrak{M} \models d_0 \cdot d_1 = c$ iff $\mathfrak{N} \models d_0 \cdot d_1 = c$.

If $\mathfrak{M}$ and $\mathfrak{N}$ agree up to e, $d \leq e$ and $\theta(x)$ is a Δ_0 formula, it follows that $\mathfrak{M} \models \theta(d)$ iff $\mathfrak{N} \models \theta(d)$ (cf. [**2**, Proposition 1]).

Theorem 3.10 $\mathbf{PA} + \mathrm{Con}^*(\mathbf{PA}) \nvdash \mathrm{Con}(\mathbf{PA})$.

Proof. Let $\mathfrak{M}$ be a countable non-standard model of $\mathbf{PA} + F_{\varepsilon_0}$ *is total.* Let M be the domain of $\mathfrak{M}$ and $a \in M$ be non-standard. Moreover, let $e = F_{\varepsilon_0}^{\mathfrak{M}}(a)$. As a result of the standing assumption, $\mathfrak{M} \models \mathrm{Con}(\mathbf{PA}\restriction_a)$. Owing to a result of Solovay's [**15**, Theorem 1.1] (or similar results in [**7**]), there exists a countable model $\mathfrak{N}$ of **PA** such that:

(a) $\mathfrak{M}$ and $\mathfrak{N}$ agree up to e (in the sense of Definition 3.9).
(b) $\mathfrak{N}$ thinks that $\mathbf{PA}\restriction_a$ is consistent.
(c) $\mathfrak{N}$ thinks that $\mathbf{PA}\restriction_{a+1}$ is inconsistent. In fact there is a proof of $0 = 1$ from $\mathbf{PA}\restriction_{a+1}$ whose Gödel number is less than 2^{2^e} (as computed in $\mathfrak{N}$).

In actuality, to be able to apply [**15**, Theorem 1.1] we have to ensure that e is much larger than a, i.e., $E^k(a) < e$ for every standard number k. It is a standard fact (provable in **PA**) that $E(x) \leq F_3(x)$ holds for all sufficiently large x (cf. [**6**, p. 269]). In particular this holds for all non-standard elements s of $\mathfrak{M}$ and hence

$$E^k(s) \leq F_3^k(s) \leq F_3^s(s) \leq F_4(s) < F_{\varepsilon_0}(s),$$

so that $E^k(a) < e$ holds for all standard k, leading to e being much larger than a.

We will now distinguish two cases.

Case 1: $\mathfrak{N} \models F_{\varepsilon_0}(a+1)\uparrow$. Then also $\mathfrak{N} \models F_{\varepsilon_0}(d)\uparrow$ for all $d > a$ by Lemma 2.3(ii). Hence, in light of (b), $\mathfrak{N} \models \mathrm{Con}^*(\mathbf{PA})$. As (c) yields $\mathfrak{N} \models \neg\mathrm{Con}(\mathbf{PA})$, we have

$$\mathfrak{N} \models \mathbf{PA} + \mathrm{Con}^*(\mathbf{PA}) + \neg\mathrm{Con}(\mathbf{PA}). \tag{3.5}$$

Case 2: $\mathfrak{N} \models F_{\varepsilon_0}(a+1)\downarrow$. We then also have $e = F_{\varepsilon_0}^{\mathfrak{N}}(a)$, for $\mathfrak{M}$ and $\mathfrak{N}$ agree up to e and the formula '$F_{\varepsilon_0}(x) = y$' is Δ_0 by Lemma 2.2. Let $c := F_{\varepsilon_0}^{\mathfrak{N}}(a+1)$. By Corollary 3.8, for every standard n there is an initial segment $\mathfrak{I}$ of $\mathfrak{N}$ such $e < \mathfrak{I} < c$ and $\mathfrak{I}$ is a model of Π_{n+1}-induction. Moreover, it follows from the properties of $\mathfrak{N}$ and the fact $2^{2^e} < \mathfrak{I}$ that

(1) $\mathfrak{I}$ thinks that $\mathbf{PA}\restriction_a$ is consistent.
(2) $\mathfrak{I}$ thinks that $\mathbf{PA}\restriction_{a+1}$ is inconsistent.
(3) $\mathfrak{I}$ thinks that $F_{\varepsilon_0}(a+1)$ is not defined.

Consequently, $\mathfrak{I} \models \mathrm{Con}^*(\mathbf{PA}) + \neg\mathrm{Con}(\mathbf{PA}) + \Pi_{n+1}$-induction. Since n was arbitrary, this shows that $\mathbf{PA} + \mathrm{Con}^*(\mathbf{PA}) + \neg\mathrm{Con}(\mathbf{PA})$ is a consistent theory. □

Proposition 3.3 and Theorem 3.10 can be extended to theories $\mathbf{T} = \mathbf{PA} + \psi$ where ψ is a true Π^0_1 statement.

Theorem 3.11 *Let* $\mathbf{T} = \mathbf{PA} + \psi$ *where* ψ *is a* Π_1 *statement such that* $T +$ 'F_{ε_0} is total' *is a consistent theory. Let* $\mathbf{T}\restriction_k$ *be the theory* $\mathbf{PA}\restriction_k + \psi$ *and* $\mathrm{Con}^*(\mathbf{T}) := \forall x \mathrm{Con}(\mathbf{T}\restriction_{F^{-1}_{\varepsilon_0}(x)})$. *Then the strength of* $\mathbf{T} + \mathrm{Con}^*(T)$ *is strictly between* $\mathbf{T}$ *and* $\mathbf{T} + \mathrm{Con}(\mathbf{T})$, *i.e.,*

(i) $\mathbf{T} \nvdash \mathrm{Con}^*(\mathbf{T})$.
(ii) $\mathbf{T} + \mathrm{Con}^*(\mathbf{T}) \nvdash \mathrm{Con}(\mathbf{T})$.
(iii) $\mathbf{T} + \mathrm{Con}(\mathbf{T}) \vdash \mathrm{Con}^*(\mathbf{T})$.

Proof. For (i) the same proof as in Proposition 3.3 works with **PA** replaced by **T**, while (iii) is obvious. For (ii) note that Solovay's Theorem also works for **T**, so that the proof of case 1 of Theorem 3.10 can be copied. To deal with case 2, observe that $\mathfrak{I} \models \psi$ since ψ is Π_1, $\mathfrak{N} \models \psi$ and $\mathfrak{I}$ is an initial segment of $\mathfrak{N}$. □

The methods of Theorem 3.10 can also be used to produce two 'natural' slow growing functions f and g such that the theories $\mathbf{PA} + \mathrm{Con}_f(\mathbf{PA})$ and $\mathbf{PA} + \mathrm{Con}_g(\mathbf{PA})$ are mutually non-interpretable in each other.

Definition 3.12 The even and odd parts of F_{ε_0} are defined as follows:

$$F^{\mathrm{even}}_{\varepsilon_0}(2n) = F_{\varepsilon_0}(2n), \qquad F^{\mathrm{even}}_{\varepsilon_0}(2n+1) = F_{\varepsilon_0}(2n) + 1,$$

$$F^{\mathrm{odd}}_{\varepsilon_0}(2n+1) = F_{\varepsilon_0}(2n+1), \qquad F^{\mathrm{odd}}_{\varepsilon_0}(2n+2) = F_{\varepsilon_0}(2n+1) + 1, \qquad F^{\mathrm{odd}}_{\varepsilon_0}(0) = 1,$$

$$f(n) = \max(\{k \le n \mid \exists y \le n\, F^{\mathrm{even}}_{\varepsilon_0}(k) = y\} \cup \{0\}),$$

$$g(n) = \max(\{k \le n \mid \exists y \le n\, F^{\mathrm{odd}}_{\varepsilon_0}(k) = y\} \cup \{0\}).$$

By Lemma 2.2, the graphs of f and g are Δ_0 and both functions are provably recursive functions of **PA**.

Remark 3.13 In a much more elaborate form, the method of defining variants of given computable functions (such as F_{ε_0}) in a piecewise manner has been employed in [**8**] to obtain results about degree structures of computable functions and in [**4**] to obtain forcing-like results about provably recursive functions.

Theorem 3.14

(i) $\mathbf{PA} + \mathrm{Con}_f(\mathbf{PA}) \nvdash \mathrm{Con}_g(\mathbf{PA})$.
(ii) $\mathbf{PA} + \mathrm{Con}_g(\mathbf{PA}) \nvdash \mathrm{Con}_f(\mathbf{PA})$.

Proof. The proof of (i) is a variant of that of Theorem 3.10. Let $\mathfrak{M}$ be a countable non-standard model of $\mathbf{PA} + F_{\varepsilon_0}$ *is total.* Let M be the domain of $\mathfrak{M}$ and $a \in M$ be non-standard such that $\mathfrak{M}$ thinks that a is *odd.* Let $e = F^{\mathfrak{M}}_{\varepsilon_0}(a)$. As before, there exists a countable model $\mathfrak{N}$ of **PA** such that:

(a) $\mathfrak{M}$ and $\mathfrak{N}$ agree up to e.
(b) $\mathfrak{N}$ thinks that $\mathbf{PA}\restriction_a$ is consistent.
(c) $\mathfrak{N}$ thinks that $\mathbf{PA}\restriction_{a+1}$ is inconsistent. In fact there is a proof of $0 = 1$ from $\mathbf{PA}\restriction_{a+1}$ whose Gödel number is less than 2^{2^e} (as computed in $\mathfrak{N}$).

Again we distinguish two cases.

Case 1: $\mathfrak{N} \models F_{\varepsilon_0}(a+1)\uparrow$. Then also $\mathfrak{N} \models F_{\varepsilon_0}(d)\uparrow$ for all $d > a$ by Lemma 2.3(ii). Since $\mathfrak{M}$ thinks that $a+1$ is even, so does $\mathfrak{N}$, as both models agree up to e. Thus $\mathfrak{N} \models F^{\mathrm{even}}_{\varepsilon_0}(d)\uparrow$ for all $d > a$. As a result, $\mathfrak{N} \models \forall x\, f(x) \le a$, and hence, $\mathfrak{N} \models \mathrm{Con}_f(\mathbf{PA})$. On the other hand, since $\mathfrak{N} \models F^{\mathrm{odd}}_{\varepsilon_0}(a+1) = e+1$ and $\mathfrak{N}$ thinks that $\mathbf{PA}\restriction_{a+1}$ is inconsistent, it follows that $\mathfrak{N} \not\models \mathrm{Con}_g(\mathbf{PA})$.

Case 2: $\mathfrak{N} \models F_{\varepsilon_0}(a+1)\downarrow$. As in the proof of Theorem 3.10, letting $c := F^{\mathfrak{N}}_{\varepsilon_0}(a+1)$, for each n we find an initial segment $\mathfrak{I}$ of $\mathfrak{N}$ such $e < \mathfrak{I} < c$ and $\mathfrak{I}$ is a model of Π_{n+1}-induction. Moreover, it follows from the properties of $\mathfrak{N}$ and the fact that $2^{2^e} < \mathfrak{I}$, that

(1) $\mathfrak{I}$ thinks that $\mathbf{PA}\restriction_a$ is consistent.
(2) $\mathfrak{I}$ thinks that $\mathbf{PA}\restriction_{a+1}$ is inconsistent.
(3) $\mathfrak{I}$ thinks that $F_{\varepsilon_0}(a+1)$ is not defined.

Consequently as $\mathfrak{I}$ thinks that $a+1$ is even, $\mathfrak{I} \models \forall x\, f(x) \le a$, whence $\mathfrak{I} \models \mathrm{Con}_f(\mathbf{PA})$. On the other hand, since $\mathfrak{I} \models F^{\mathrm{odd}}_{\varepsilon_0}(a+1) = e+1$, we also have that $\mathfrak{N} \not\models \mathrm{Con}_g(\mathbf{PA})$. Since n was arbitrary, this shows that $\mathbf{PA} + \mathrm{Con}_f(\mathbf{PA}) + \neg\mathrm{Con}_g(\mathbf{PA})$ is a consistent theory.

For (ii), the argument is completely analogous, the only difference being that we start with a non-standard $a \in M$ such that $\mathfrak{M}$ thinks that a is even. □

Corollary 3.15 *Neither is* $\mathbf{PA} + \mathrm{Con}_f(\mathbf{PA})$ *interpretable in* $\mathbf{PA} + \mathrm{Con}_g(\mathbf{PA})$ *nor* $\mathbf{PA} + \mathrm{Con}_g(\mathbf{PA})$ *interpretable in* $\mathbf{PA} + \mathrm{Con}_f(\mathbf{PA})$.

Proof. This follows from Theorem 3.14 and Theorem 3.1. □

3.1 A natural Orey sentence

A sentence φ of $\mathbf{PA}$ is called an *Orey sentence* if both $\mathbf{PA}+\varphi \lhd \mathbf{PA}$ and $\mathbf{PA}+\neg\varphi \lhd \mathbf{PA}$ hold.

Corollary 3.16 *The sentence* $\exists x\,(F_{\varepsilon_0}(x)\uparrow \wedge \forall y < x\, F_{\varepsilon_0}(y)\downarrow \wedge\ x \textit{ is even})$ *is an Orey sentence.*

Proof. Let ψ be the foregoing sentence. In view of Theorem 3.1, it suffices to show that $\mathbf{PA} \vdash \mathrm{Con}(\mathbf{PA}\restriction_k +\psi)$ and $\mathbf{PA} \vdash \mathrm{Con}(\mathbf{PA}\restriction_k +\neg\psi)$ hold for all k. Fix $k > 0$.

First we show that $\mathbf{PA} \vdash \mathrm{Con}(\mathbf{PA}\restriction_k +\psi)$. Note that $\mathbf{PA}$ proves the consistency of $\mathbf{PA}\restriction_k + \forall x\, F_{\omega_{k+1}}(x)\downarrow + \exists x F_{\varepsilon_0}(x)\uparrow$. Arguing in $\mathbf{PA}$ we thus find a non-standard model $\mathfrak{N}$ such that

$$\mathfrak{N} \models \mathbf{PA}\restriction_k + \forall x\, F_{\omega_{k+1}}(x)\downarrow + \exists x F_{\varepsilon_0}(x)\uparrow .$$

In particular there exists a least $a \in |\mathfrak{N}|$ in the sense of $\mathfrak{N}$ such that $\mathfrak{N} \models F_{\varepsilon_0}(a)\uparrow$. If $\mathfrak{N}$ thinks that a is even, then $\mathfrak{N} \models \psi$, which entails that $\mathrm{Con}(\mathbf{PA}\restriction_k +\psi)$. If $\mathfrak{N}$ thinks that a is odd, we define a cut $\mathfrak{I}$ such that $\mathfrak{I} \models \mathbf{PA}\restriction_k$ and $F^{\mathfrak{N}}_{\varepsilon_0}(a-2) < \mathfrak{I} < F^{\mathfrak{N}}_{\varepsilon_0}(a-1)$, applying Theorem 3.7. Then $\mathfrak{I} \models \psi$ which also entails $\mathrm{Con}(\mathbf{PA}\restriction_k +\psi)$.

Next we show that $\mathbf{PA} \vdash \mathrm{Con}(\mathbf{PA}\restriction_k +\neg\psi)$. As $\mathbf{PA}$ proves

$$\mathrm{Con}(\mathbf{PA}\restriction_k + \forall x\, F_{\omega_{k+1}}(x)\downarrow),$$

we can argue in $\mathbf{PA}$ and assume that we have a model $\mathfrak{M} \models \mathbf{PA}\restriction_k + \forall x\, F_{\omega_{k+1}}(x)\downarrow$. If $\mathfrak{M} \models \forall x F_{\varepsilon_0}(x)\downarrow$ then $\mathfrak{M} \models \neg\psi$, and $\mathrm{Con}(\mathbf{PA}\restriction_k +\neg\psi)$ follows. Otherwise there is a least a in the sense of $\mathfrak{M}$ such that $F^{\mathfrak{M}}_{\varepsilon_0}(a)\uparrow$. If $\mathfrak{M}$ thinks that a is odd we have $\mathfrak{M} \models \neg\psi$, too. If $\mathfrak{M}$ thinks that a is even we introduce a cut $F^{\mathfrak{M}}_{\varepsilon_0}(a-2) < \mathfrak{I}' < F^{\mathfrak{M}}_{\varepsilon_0}(a-1)$ such that $\mathfrak{I}' \models \mathbf{PA}\restriction_k$. Since $\mathfrak{I}' \models F_{\varepsilon_0}(a-1)\uparrow$ we have $\mathfrak{I}' \models \neg\psi$, whence $\mathrm{Con}(\mathbf{PA}\restriction_k +\neg\psi)$. □

References

[1] W. Buchholz, S. Wainer: *Provably computable functions and the fast growing hierarchy*, in: S. Simpson (ed.): Logic and Combinatorics, Contemporary Mathematics 65 (AMS, Providence, 1987), 179–198.

[2] C. Dimitracopoulos, J. B. Paris: *Truth definitions for Δ_0 formulae*, in: Logic and Algorithmic, L'Enseignement Mathématique 30 (Univ. Genève, Geneva, 1982), 317–329.

[3] H. Friedman, S. Sheard: *Elementary descent recursion and proof theory*, Annals of Pure and Applied Logic 71 (1995), 1–45.

[4] S.-D. Friedman, M. Rathjen, A. Weiermann: *Some results on PA-provably recursive functions*, preprint, 2011.

[5] D. Guaspari: *Partially conservative extensions of arithmetic*, Transactions of the American Mathematical Society 254 (1979), 47–68.

[6] J. Ketonen, R. M. Solovay: *Rapidly growing Ramsey functions*, Annals of Mathematics 113 (1981), 267–314.

[7] J. Krajíček, P. Pudlák: *On the structure of initial segments of models of arithmetic*, Archive for Mathematical Logic 28 (1989), 91–98.

[8] L. Kristiansen: *Subrecursive degrees and fragments of Peano arithmetic*, Archive for Mathematical Logic 40 (2001), 365–397.

[9] P. Lindström: *Some results on interpretability*, in: Proceedings of the 5th Scandinavian Logic Symposium 1979 (Aalborg University Press, Aalborg, 1979), 329–361.

[10] P. Lindström: *Aspects of Incompleteness*, Lecture Notes in Logic 10, second edition (Association for Symbolic Logic, 2003).

[11] S. Orey: *Relative interpretations*, Zeitschrift für mathematische Logik 7 (1961), 146–153.

[12] J. B. Paris: *A hierarchy of cuts in models of arithmetic*, in: Lecture Notes in Mathematics, vol. 834 (Springer, Berlin, 1980), 312–337.

[13] J. Paris, L. Harrington: *A mathematical incompleteness in Peano arithmetic*, in: J. Barwise (ed.): Handbook of Mathematical Logic (North-Holland, Amsterdam, 1977), 1133–1142.

[14] D. Schmidt: *Built-up systems of fundamental sequences and hierarchies of number-theoretic functions*, Archive for Mathematical Logic 18 (1976), 47–53.

[15] R. M. Solovay; *Injecting inconsistencies into models of PA*, Annals of Pure and Applied Logic 44 (1989), 101–132.

[16] R. Sommer: *Transfinite induction and hierarchies generated by transfinite recursion within Peano arithmetic*, PhD thesis, U. C. Berkeley, 1990.

[17] R. Sommer: *Transfinite induction within Peano arithmetic*, Annals of Pure and Applied Logic 76 (1995), 231–289.

[18] G. Takeuti: *Proof Theory*, 2nd edition (North-Holland, Amsterdam, 1987).

Relativized ordinal analysis: The case of Power Kripke-Platek set theory

Michael Rathjen

Department of Pure Mathematics, University of Leeds, UK
rathjen@maths.leeds.ac.uk

Abstract. The paper relativizes the method of ordinal analysis developed for Kripke–Platek set theory to theories which have the power set axiom. We show that it is possible to use this technique to extract information about Power Kripke–Platek set theory, $\mathbf{KP}(\mathcal{P})$.

Introduction

Ordinal analyses of ever stronger theories have been obtained over the last 20 years [**1, 2, 3, 19, 20, 23, 24**]. The strongest theories for which proof-theoretic ordinals have been determined are subsystems of second-order arithmetic with comprehension restricted to Π^1_2-comprehension (or even Δ^1_3-comprehension; see [**26, 27, 28**]). Thus it appears that it is currently impossible to furnish an ordinal analysis of any set theory which has the power set axiom among its axioms as such a theory would dwarf the strength of second-order arithmetic. Notwithstanding the foregoing, the current paper relativizes the techniques of ordinal analysis developed for Kripke–Platek set theory, **KP**, to obtain very useful information about Power Kripke–Platek set theory, $\mathbf{KP}(\mathcal{P})$, crystallizing in a bound for the transfinite iterations of the power set operation that are provable in the latter theory.

Technically we draw on tools that were developed more than 30 years ago. With the work of Jäger and Pohlers [**13, 14**] the forum of ordinal analysis switched from subsystems of second-order arithmetic to set theory, shaping what is called *admissible proof theory*, after the standard models of **KP**. We also draw on the framework of operator controlled derivations developed by Buchholz [**22**] that allows one to express the uniformity of infinite derivations and to carry out their bookkeeping in an elegant way.

The results and techniques of this paper have important applications. The characterization of the strength of $\mathbf{KP}(\mathcal{P})$ in terms of the von Neumann hierarchy is used in [**31**, Theorem 1.1] to calibrate the strength of the calculus of construction with one type universe (which is an intuitionistic type theory). Another application is made in connection with the so-called *existence property*, **EP**, that intuitionistic set theories may or may not have. Full intuitionistic Zermelo–Fraenkel set theory, **IZF**, does not have the existence property, where **IZF** is formulated with Collection (cf. [**12**]). By contrast, an ordinal analysis of intuitionistic $\mathbf{KP}(\mathcal{P})$ similar to the one given in this paper together with results from [**30**] can be utilized to show that **IZF** with only bounded separation has the **EP**.

Research supported by Templeton Foundation Grant #13152, the CRM Infinity Project, and by U.K. EPSRC grant No. EP/G029520/1. The author is also grateful for a Mercator Professorship of the German Science Foundation (DFG) which allowed him to work on some of the ideas of this article.

1 Power Kripke–Platek set theory

A particularly interesting (classical) subtheory of **ZF** is Kripke–Platek set theory, **KP**. Its standard models are called *admissible sets.* One of the reasons that this is an important theory is that a great deal of set theory requires only the axioms of **KP**. An even more important reason is that admissible sets have been a major source of interaction between model theory, recursion theory and set theory (cf. [**5**]). **KP** arises from **ZF** by completely omitting the power set axiom and restricting separation and collection to bounded formulae. These alterations are suggested by the informal notion of 'predicative'.

To be more precise, quantifiers of the forms $\forall x \in a$, $\exists x \in a$ are called *bounded. Bounded* or Δ_0*-formulae* are the formulae wherein all quantifiers are bounded. The axioms of **KP** consist of *Extensionality, Pair, Union, Infinity, Bounded Separation*

$$\exists x \, \forall u \, [u \in x \leftrightarrow (u \in a \, \wedge \, A(u))]$$

for all bounded formulae $A(u)$, *Bounded Collection*

$$\forall x \in a \, \exists y \, G(x,y) \, \rightarrow \, \exists z \, \forall x \in a \, \exists y \in z \, G(x,y)$$

for all bounded formulae $G(x,y)$, and *Set Induction*

$$\forall x \, \left[(\forall y \in x \, C(y)) \rightarrow C(x) \right] \, \rightarrow \, \forall x \, C(x)$$

for all formulae $C(x)$.

A transitive set A such that $(A, \in)$ is a model of **KP** is called an *admissible set.* Of particular interest are the models of **KP** formed by segments of Gödel's *constructible hierarchy* **L**. The constructible hierarchy is obtained by iterating the definable powerset operation through the ordinals

$$\begin{aligned} \mathbf{L}_0 &= \emptyset, \\ \mathbf{L}_\lambda &= \bigcup \{ \mathbf{L}_\beta : \beta < \lambda \}, \ \ \lambda \text{ limit}, \\ \mathbf{L}_{\beta+1} &= \big\{ X : \ X \subseteq \mathbf{L}_\beta; \ X \text{ definable over } \ \langle \mathbf{L}_\beta, \in \rangle \big\}. \end{aligned}$$

So any element of **L** of level α is definable from elements of **L** with levels $< \alpha$ and the parameter $\mathbf{L}_\alpha$. An ordinal α is *admissible* if the structure $(\mathbf{L}_\alpha, \in)$ is a model of **KP**.

If the power set operation is considered as a definite operation, but the universe of all sets is regarded as an indefinite totality, we are led to systems of set theory having Power Set as an axiom but only Bounded Separation axioms and intuitionistic logic for reasoning about the universe at large. The study of subsystems of **ZF** formulated in intuitionistic logic with Bounded Separation but containing the Power Set axiom was apparently initiated by Pozsgay [**17, 18**] and then pursued more systematically by Tharp [**33**], Friedman [**10**] and Wolf [**35**]. These systems are actually semi-intuitionistic as they contain the law of excluded middle for bounded formulae.

In the classical context, weak subsystems of **ZF** with Bounded Separation and Power Set have been studied by Thiele [**34**], Friedman [**11**] and more recently at great length by Mathias [**16**]. Mac Lane has singled out and championed a particular fragment of **ZF**, especially in his book *Form and Function* [**15**]. Mac Lane Set Theory, christened **MAC** in [**16**], comprises the axioms of Extensionality, Null Set, Pairing, Union, Infinity, Power Set, Bounded Separation, Foundation, and Choice. **MAC** is naturally related to systems derived from topos-theoretic notions and, moreover, to type theories.

Definition 1.1 We use subset bounded quantifiers $\exists x \subseteq y \ldots$ and $\forall x \subseteq y \ldots$ as abbreviations for $\exists x(x \subseteq y \wedge \ldots)$ and $\forall x(x \subseteq y \rightarrow \ldots)$, respectively.

The $\Delta_0^{\mathcal{P}}$ formulae are the smallest class of formulae containing the atomic formulae closed under $\wedge, \vee, \rightarrow, \neg$ and the quantifiers

$$\forall x \in a,\ \exists x \in a,\ \forall x \subseteq a,\ \exists x \subseteq a.$$

Definition 1.2 $\mathbf{KP}(\mathcal{P})$ has the same language as **ZF**. Its axioms are the following: Extensionality, Pairing, Union, Infinity, Powerset, $\Delta_0^{\mathcal{P}}$-Separation and $\Delta_0^{\mathcal{P}}$-Collection.

The transitive models of $\mathbf{KP}(\mathcal{P})$ have been termed *power admissible* sets in [**11**].

Remark 1.3 Alternatively, $\mathbf{KP}(\mathcal{P})$ can be obtained from **KP** by adding a function symbol $\mathcal{P}$ for the powerset function as a primitive symbol to the language and the axiom

$$\forall y\,[y \in \mathcal{P}(x) \leftrightarrow y \subseteq x]$$

and extending the schemes of Δ_0 Separation and Collection to the Δ_0 formulae of this new language.

Lemma 1.4 $\mathbf{KP}(\mathcal{P})$ *is* not *the same theory as* $\mathbf{KP} + \mathbf{Pow}$. *Indeed,* $\mathbf{KP} + \mathbf{Pow}$ *is a much weaker theory than* $\mathbf{KP}(\mathcal{P})$ *in which one cannot prove the existence of* $V_{\omega+\omega}$.

Proof. Note that in the presence of full Separation and Infinity there is no difference between our system **KP** and Mathias' [**16**] **KP**. It follows from [**16**, Theorem 14] that $\mathbf{Z} + \mathbf{KP} + \mathbf{AC}$ is conservative over $\mathbf{Z} + \mathbf{AC}$ for stratifiable sentences. **Z** and $\mathbf{Z} + \mathbf{AC}$ are of the same proof-theoretic strength as the constructible hierarchy can be simulated in **Z**; a stronger statement is given in [**16**, Theorem 16]. As a result, **Z** and $\mathbf{Z} + \mathbf{KP}$ are of the same strength. As $\mathbf{KP} + \mathbf{Pow}$ is a subtheory of $\mathbf{Z} + \mathbf{KP}$, we have that $\mathbf{KP} + \mathbf{Pow}$ is not stronger than **Z**. If $\mathbf{KP} + \mathbf{Pow}$ could prove the existence of $V_{\omega+\omega}$ it would prove the consistency of **Z**. On the other hand $\mathbf{KP}(\mathcal{P})$ proves the existence of V_α for every ordinal α and hence proves the existence of arbitrarily large transitive models of **Z**. □

Remark 1.5 Our system $\mathbf{KP}(\mathcal{P})$ is not quite the same as the theory $\mathbf{KP}^{\mathcal{P}}$ in Mathias' paper [**16**, 6.10]. The difference between $\mathbf{KP}(\mathcal{P})$ and $\mathbf{KP}^{\mathcal{P}}$ is that in the latter system set induction only holds for $\Sigma_1^{\mathcal{P}}$ formulae, or what amounts to the same, $\Pi_1^{\mathcal{P}}$ foundation ($A \neq \emptyset \rightarrow \exists x \in A\ x \cap A = \emptyset$ for $\Pi_1^{\mathcal{P}}$ classes A).

Friedman [**11**] includes only Set Foundation in his formulation of a formal system $\mathbf{PAdm}^s$ appropriate to the concept of recursion in the power set operation $\mathcal{P}$.

2 A Tait-style formalization of $\mathbf{KP}(\mathcal{P})$

For technical reasons we shall use a Tait-style sequent calculus version of $\mathbf{KP}(\mathcal{P})$ in which finite sets of formulae can be derived. In addition, formulae have to be in negation normal form (cf. [**32**]). The language consists of: free variables $a_0, a_1, \ldots$, bound variables $x_0, x_1, \ldots$; the predicate symbol $\in$; the logical symbols $\neg, \vee, \wedge, \forall, \exists$. One peculiarity will be that we treat bounded quantifiers and subset bounded quantifiers as quantifiers in their own right.

We will use $a, b, c, \ldots,\ x, y, z, \ldots,\ A, B, C, \ldots$ as metavariables whose domains are the domain of the free variables, bound variables, formulae, respectively.

The *atomic formulae* are those of the form $(a \in b)$, $\neg(a \in b)$. *Formulae* are defined inductively as follows:

(i) Atomic formulae are formulae.

(ii) If A and B are formulae, then so are $(A \land B)$ and $(A \lor B)$.

(iii) If $A(b)$ is a formula in which x does not occur, then $\forall x A(x)$, $\exists x A(x)$, $(\forall x \in a)$ $A(x)$, $(\exists x \in a)A(x)$, $(\forall x \subseteq a)A(x)$, and $(\exists x \subseteq a)A(x)$ are formulae.

The quantifiers $\exists x$, $\forall x$ will be called *unrestricted*, whereas the other quantifiers will be referred to as *restricted quantifiers.* A $\Delta_0^{\mathcal{P}}$*-formula* is a formula which contains no unrestricted quantifiers. The Δ_0-formulae are those $\Delta_0^{\mathcal{P}}$-formulae that do not contain subset bounded quantifiers.

The *negation* $\neg A$ of a formula A is defined to be the formula obtained from A by (i) putting $\neg$ in front of any atomic formula; (ii) replacing $\land$, $\lor$, $\forall x$, $\exists x$, $(\forall x \in a)$, $(\exists x \in a)$, $(\forall x \subseteq a)$, $(\exists x \subseteq a)$ by $\lor$, $\land$, $\exists x$, $\forall x$, $(\exists x \in a)$, $(\forall x \in a)$, $(\exists x \subseteq a)$, $(\forall x \subseteq a)$, respectively, and (iii) dropping double negations. $A \to B$ stands for $\neg A \lor B$.

$\vec{a}, \vec{b}, \vec{c}, \ldots$ and $\vec{x}, \vec{y}, \vec{z}, \ldots$ will be used to denote finite sequences of free and bound variables, respectively.

We use $F[a_1, \ldots, a_n]$ (by contrast with $F(a_1, \ldots, a_n)$) to denote a formula the free variables of which are among $a_1, \ldots, a_n$. We will write $a = \{x \in b : G(x)\}$ for

$$(\forall x \in a)[x \in b \land G(x)] \land (\forall x \in b)[G(x) \to x \in a].$$

$a = b$ stands for $(\forall x \in a)(x \in b) \land (\forall x \in b)(x \in a)$, and $a \subseteq b$ stands for $(\forall x \in a)(x \in b)$. However, as part of a subset bounded quantifier $(\forall x \subseteq a)$ or $(\exists x \subseteq b)$, $\subseteq$ is considered to be a primitive symbol.

Definition 2.1 The sequent-style version of $\mathbf{KP}(\mathcal{P})$ derives finite sets of formulae denoted by $\Gamma, \Delta, \Theta, \Xi, \ldots$. The intended meaning of Γ is the disjunction of all formulae of Γ. We use the notation Γ, A for $\Gamma \cup \{A\}$, and Γ, Ξ for $\Gamma \cup \Xi$.

The *axioms of* $\mathbf{KP}(\mathcal{P})$ are the following:

Logical axioms:	$\Gamma, A, \neg A$ for every $\Delta_0^{\mathcal{P}}$-formula A
Extensionality:	$\Gamma, a = b \land B(a) \to B(b)$ for every $\Delta_0^{\mathcal{P}}$-formula $B(a)$
Pair:	$\Gamma, \exists x[a \in x \land b \in x]$
Union:	$\Gamma, \exists x(\forall y \in a)(\forall z \in y)(z \in x)$
$\Delta_0^{\mathcal{P}}$*-Separation:*	$\Gamma, \exists y(y = \{x \in a : G(x)\})$ for every $\Delta_0^{\mathcal{P}}$-formula $G(b)$
Set Induction:	$\Gamma, \forall u\,[(\forall x \in u)\, G(x) \to G(x)] \to \forall u\, G(u)$ for every formula $G(b)$
Infinity:	$\Gamma, \exists x\,[(\exists y \in x)\, y \in x \ \land\ (\forall y \in x)(\exists z \in x)\, y \in z]$
Power Set:	$\Gamma, \exists z\,(\forall x \subseteq a) x \in z$

The *logical rules of inference* are:

$(\land)$	$\vdash \Gamma, A$ and $\vdash \Gamma, B$	$\Rightarrow$	$\vdash \Gamma, A \land B$
$(\lor)$	$\vdash \Gamma, A_i$ for $i \in \{0, 1\}$	$\Rightarrow$	$\vdash \Gamma, A_0 \lor A_1$
$(b\forall)$	$\vdash \Gamma, a \in b \to F(a)$	$\Rightarrow$	$\vdash \Gamma, (\forall x \in b)F(x)$
$(pb\forall)$	$\vdash \Gamma, a \subseteq b \to F(a)$	$\Rightarrow$	$\vdash \Gamma, (\forall x \subseteq b)F(x)$
$(\forall)$	$\vdash \Gamma, F(a)$	$\Rightarrow$	$\vdash \Gamma, \forall x F(x)$
$(b\exists)$	$\vdash \Gamma, a \in b \land F(a)$	$\Rightarrow$	$\vdash \Gamma, (\exists x \in b)F(x)$
$(pb\exists)$	$\vdash \Gamma, a \subseteq b \land F(a)$	$\Rightarrow$	$\vdash \Gamma, (\exists x \subseteq b)F(x)$
$(\exists)$	$\vdash \Gamma, F(a)$	$\Rightarrow$	$\vdash \Gamma, \exists x F(x)$
(Cut)	$\vdash \Gamma, A$ and $\vdash \Gamma, \neg A$	$\Rightarrow$	$\vdash \Gamma$

In the foregoing rules, $F(a)$ is an arbitrary formula. Of course, it is demanded that in $(b\forall)$, $(pb\forall)$ and $(\forall)$ the free variable a is not to occur in the conclusion; a is called the *eigenvariable* of that inference.

The *non-logical rule of inference* is

$$(\Delta_0^{\mathcal{P}}\text{-COLLR}) \quad \vdash \Gamma, (\forall x \in a)\exists y H(x,y) \;\Rightarrow\; \vdash \Gamma, \exists z(\forall x \in a)(\exists y \in z)H(x,y)$$

for every $\Delta_0^{\mathcal{P}}$-formula $H(b,c)$.

We shall conceive of axioms as inferences with an empty set of premisses. The *minor formulae* (m.f.) of an inference are those formulae which are rendered prominently in its premises. The *principal formulae* (p.f.) of an inference are the formulae rendered prominently in its conclusion. (Cut) has no p.f. So any inference has the form

$$\text{For all } i < k \;\vdash \Gamma, \Xi_i \;\Rightarrow\; \vdash \Gamma, \Xi \tag{2.1}$$

$(0 \le k \le 2)$, where Ξ consists of the p.f. and Ξ_i is the set of m.f. in the i-th premise. The formulae in Γ are called *side formulae* (s.f.) of (2.1). *Derivations* are defined inductively, as usual. $\mathcal{D}, \mathcal{D}', \mathcal{D}_0, \ldots$ range as syntactic variables over derivations. All this is completely standard, and we refer to [**32**] for notions like "*length of a derivation* $\mathcal{D}$" (abbreviated by $|\mathcal{D}|$), "*last inference of* $\mathcal{D}$", or "*direct subderivation of* $\mathcal{D}$". We write $\mathcal{D} \vdash \Gamma$ if $\mathcal{D}$ is a derivation of Γ.

3 A representation system for the Bachmann–Howard ordinal

Definition 3.1 Let Ω be a "big" ordinal, e.g., $\Omega = \aleph_1$ or ω_1^{ck}. By recursion on α we define sets $C^{\Omega}(\alpha,\beta)$ and the ordinal $\psi_\Omega(\alpha)$ as follows:

$$C^{\Omega}(\alpha,\beta) = \begin{cases} \text{closure of } \beta \cup \{0,\Omega\} \text{ under:} \\ +, (\xi \mapsto \omega^{\xi}) \\ (\xi \longmapsto \psi_\Omega(\xi))_{\xi<\alpha}; \end{cases} \tag{3.1}$$

$$\psi_\Omega(\alpha) \simeq \min\{\rho < \Omega \,:\, C^{\Omega}(\alpha,\rho) \cap \Omega = \rho\}. \tag{3.2}$$

It can be shown that $\psi_\Omega(\alpha)$ is always defined and that

$$\psi_\Omega(\alpha) < \Omega.$$

In the case of Ω being ω_1^{ck}, this follows from [**22**]. Moreover,

$$[\psi_\Omega(\alpha), \Omega] \;\cap\; C^{\Omega}(\alpha, \psi_\Omega(\alpha)) = \emptyset.$$

Thus the order-type of the ordinals below Ω which belong to the set $C^{\Omega}(\alpha,\psi_\Omega(\alpha))$ is $\psi_\Omega(\alpha)$. $\psi_\Omega(\alpha)$ is also a countable ordinal. In more pictorial terms, $\psi_\Omega(\alpha)$ is the α-th collapse of Ω.

Let $\varepsilon_{\Omega+1}$ be the least ordinal $\alpha > \Omega$ such that $\omega^{\alpha} = \alpha$. The set of ordinals $C^{\Omega}(\varepsilon_{\Omega+1}, 0)$ gives rise to an elementary computable ordinal representation system (cf. [**7, 13, 22, 25**]). In what follows, $C^{\Omega}(\varepsilon_{\Omega+1}, 0)$ will sometimes be denoted by $\mathcal{T}(\Omega)$.

In point of fact,

$$C^{\Omega}(\varepsilon_{\Omega+1}, 0) \;\cap\; \Omega = \psi_\Omega(\varepsilon_{\Omega+1}).$$

The ordinal $\psi_\Omega(\varepsilon_{\Omega+1})$ is known as the *Bachmann–Howard ordinal.* Its relation to **KP** is that it is the proof-theoretic ordinal of this theory as was shown by Jäger [**13**]. Moreover it is the smallest ordinal such that $L_{\psi_\Omega(\varepsilon_{\Omega+1})}$ is a Π_2-model of **KP** (see [**21**, Theorem 2.1] or [**29**, Theorem 4.3]), i.e., whenever **KP** proves a Π_2 sentence C of set theory, then $L_{\psi_\Omega(\varepsilon_{\Omega+1})} \models C$.

4 The infinitary proof system $\boldsymbol{RS}_{\Omega}^{\mathcal{P}}$

Henceforth all ordinals will be assumed to belong to $C^{\Omega}(\varepsilon_{\Omega+1}, 0)$.

The problem of "naming" sets will be solved by building a formal von Neumann hierarchy using the ordinals $< \Omega$ belonging to this set (i.e., ordinals $< \psi_{\Omega}(\varepsilon_{\Omega+1})$).

Definition 4.1 We define the $RS_{\Omega}^{\mathcal{P}}$-terms. To each $RS_{\Omega}^{\mathcal{P}}$-term t we also assign its *level*, $|t|$.

1. For each $\alpha < \Omega$, $\mathbb{V}_{\alpha}$ is an $RS_{\Omega}^{\mathcal{P}}$-term with $|\mathbb{V}_{\alpha}| = \alpha$.
2. For each $\alpha < \Omega$, we have infinitely many free variables $a_1^{\alpha}, a_2^{\alpha}, a_3^{\alpha}, \ldots$ which are $RS_{\Omega}^{\mathcal{P}}$-terms with $|a_i^{\alpha}| = \alpha$.
3. If $F(x, \vec{y})$ is a $\Delta_0^{\mathcal{P}}$ formula (whose free variables are exactly those indicated) and $\vec{s} \equiv s_1, \ldots, s_n$ are $RS_{\Omega}^{\mathcal{P}}$-terms, then the formal expression
$$\{x{\in}\mathbb{V}_{\alpha} \mid F(x, \vec{s})\}$$
is an $RS_{\Omega}^{\mathcal{P}}$-term with $|\{x{\in}\mathbb{V}_{\alpha} \mid F(x, \vec{s})\}| = \alpha$.

The $RS_{\Omega}^{\mathcal{P}}$*-formulae* are the expressions of the form $F(s_1, \ldots, s_n)$, where $F[a_1, \ldots, a_n]$ is a formula of $\mathbf{KP}(\mathcal{P})$ and $s_1, \ldots, s_n$ are $RS_{\Omega}^{\mathcal{P}}$-terms. We set
$$|F(s_1, \ldots, s_n)| = \{|s_1|, \ldots, |s_n|\}.$$
If $F[a_1, \ldots, a_n]$ is in $\Delta_0^{\mathcal{P}}$, then $F(s_1, \ldots, s_n)$ is also called a $\Delta_0^{\mathcal{P}}$ formula (of $RS_{\Omega}^{\mathcal{P}}$).

As in the case of the Tait-style version of $\mathbf{KP}(\mathcal{P})$, we let $\neg A$ be the formula which arises from A by (i) putting $\neg$ in front of each atomic formula; (ii) replacing $\wedge$, $\vee$, $(\forall x{\in}s)$, $(\exists x{\in}s)$, $(\forall x \subseteq s)$, $(\exists x \subseteq s)$, $\forall x$, $\exists x$ by $\vee$, $\wedge$, $(\exists x{\in}s)$, $(\forall x{\in}s)$, $(\exists x \subseteq s)$, $(\forall x \subseteq s)$, $\exists x$, $\forall x$, respectively, and (iii) dropping double negations. $A \to B$ stands for $\neg A \vee B$.

Convention In the sequel, $RS_{\Omega}^{\mathcal{P}}$-formulae will simply be referred to as formulae. The same usage applies to $RS_{\Omega}^{\mathcal{P}}$-terms.

We denote by upper case Greek letters $\Gamma, \Delta, \Lambda, \ldots$ finite sets of $RS_{\Omega}^{\mathcal{P}}$-formulae. The intended meaning of $\Gamma = \{A_1, \ldots, A_n\}$ is the disjunction $A_1 \vee \cdots \vee A_n$. Γ, Ξ stands for $\Gamma \cup \Xi$ and Γ, A stands for $\Gamma \cup \{A\}$.

Definition 4.2 The *axioms* of $RS_{\Omega}^{\mathcal{P}}$ are:

(A1) $\Gamma,\, A,\, \neg A$ for A in $\Delta_0^{\mathcal{P}}$.

(A2) $\Gamma,\, t = t$.

(A3) $\Gamma,\, s_1 \neq t_1, \ldots, s_n \neq t_n, \neg A(s_1, \ldots, s_n),\, A(t_1, \ldots, t_n)$
for $A(s_1, \ldots, s_n)$ in $\Delta_0^{\mathcal{P}}$.

(A4) $\Gamma,\, s \in \mathbb{V}_{\alpha}$ if $|s| < \alpha$.

(A5) $\Gamma,\, s \subseteq \mathbb{V}_{\alpha}$ if $|s| \leq \alpha$.

(A6) $\Gamma, t \notin \{x \in \mathbb{V}_{\alpha} \mid F(x, \vec{s})\}, F(t, \vec{s})$

whenever $F(t, \vec{s})$ is $\Delta_0^{\mathcal{P}}$ and $|t| < \alpha$.

(A7) $\Gamma, \neg F(t, \vec{s}), t \in \{x \in \mathbb{V}_{\alpha} \mid F(x, \vec{s})\}$

whenever $F(t, \vec{s})$ is $\Delta_0^{\mathcal{P}}$ and $|t| < \alpha$.

The *inference rules* of $RS_\Omega^{\mathcal{P}}$ are:

$$(\wedge) \qquad \frac{\Gamma, A \quad \Gamma, A'}{\Gamma, A \wedge A'}$$

$$(\vee) \qquad \frac{\Gamma, A_i}{\Gamma, A_0 \vee A_1} \quad \text{if } i = 0 \text{ or } i = 1$$

$$(b\forall)_\infty \qquad \frac{\Gamma,\ s{\in}t \to F(s) \ \text{ for all } |s| < |t|}{\Gamma,\ (\forall x \in t)F(x)}$$

$$(b\exists) \qquad \frac{\Gamma,\ s{\in}t \wedge F(s)}{\Gamma, (\exists x \in t)F(x)} \quad \text{if } |s| < |t|$$

$$(pb\forall)_\infty \qquad \frac{\Gamma, s \subseteq t \to F(s) \ \text{ for all } |s| \le |t|}{\Gamma, (\forall x \subseteq t)F(x)}$$

$$(pb\exists) \qquad \frac{\Gamma, s \subseteq t \wedge F(s)}{\Gamma, (\exists x \subseteq t)F(x)} \quad \text{if } |s| \le |t|$$

$$(\forall)_\infty \qquad \frac{\Gamma,\ F(s) \ \text{ for all } s}{\Gamma, \forall x\, F(x)}$$

$$(\exists) \qquad \frac{\Gamma,\ F(s)}{\Gamma, \exists x\, F(x)}$$

$$(\notin)_\infty \qquad \frac{\Gamma, r \in t \to r \neq s \ \text{ for all } |r| < |t|}{\Gamma, s \notin t}$$

$$(\in) \qquad \frac{\Gamma,\ r{\in}t \ \wedge \ r = s}{\Gamma, s \in t} \quad \text{if } |r| < |t|$$

$$(\not\subseteq)_\infty \qquad \frac{\Gamma, r \subseteq t \to r \neq s \ \text{ for all } |r| \le |t|}{\Gamma, s \not\subseteq t}$$

$$(\subseteq) \qquad \frac{\Gamma,\ s = r \ \wedge \ r \subseteq t}{\Gamma, s \subseteq t} \quad \text{if } |r| \le |s|$$

$$(\text{Cut}) \qquad \frac{\Gamma, A \qquad \Gamma, \neg\, A}{\Gamma}$$

$$(\Sigma^{\mathcal{P}}\text{-Ref}) \qquad \frac{\Gamma, A}{\Gamma,\ \exists z\, A^z} \quad \text{if } A \text{ is a } \Sigma^{\mathcal{P}}\text{-formula,}$$

where a formula is said to be in $\Sigma^{\mathcal{P}}$ if all its unbounded quantifiers are existential.

A^z results from A by restricting all unbounded quantifiers to z.

4.1 $\mathcal{H}$-controlled derivations

In general, in $RS_\Omega^{\mathcal{P}}$ we cannot remove cuts that have $\Delta_0^{\mathcal{P}}$ cut formulae. What is more, the rule ($\Sigma^{\mathcal{P}}$-Ref) poses an obstacle to removing cuts involving $\Sigma_1^{\mathcal{P}}$ formulae. Notwithstanding that, it will turn out that cuts of a complexity higher than $\Delta_0^{\mathcal{P}}$ can be removed from derivations of $\Sigma^{\mathcal{P}}$ formulae if they are of a very uniform kind.

For the presentation of infinitary proofs we draw on **[7]**. Buchholz developed a very elegant and flexible setting for describing uniformity in infinitary proofs, called *operator controlled derivations.*

Definition 4.3 Let

$$P(ON) = \{X : X \text{ is a set of ordinals}\}.$$

A class function $\mathcal{H}\colon P(ON) \to P(ON)$ will be called *operator* if $\mathcal{H}$ is a closure operator, i.e., monotone, inclusive, and idempotent, and satisfies the following conditions for all $X \in P(ON)$:

(1) $0 \in \mathcal{H}(X)$ and $\Omega \in \mathcal{H}(X)$.

(2) If α has Cantor normal form $\omega^{\alpha_1} + \cdots + \omega^{\alpha_n}$, then

$$\alpha \in \mathcal{H}(X) \iff \alpha_1, \ldots, \alpha_n \in \mathcal{H}(X).$$

The latter ensures that $\mathcal{H}(X)$ will be closed under $+$ and $\sigma \mapsto \omega^{\sigma}$, and decomposition of its members into additive and multiplicative components.

For a sequent $\Gamma = \{A_1, \ldots, A_n\}$ we define

$$|\Gamma| := |A_1| \cup \ldots \cup |A_n|.$$

If s is an $RS_{\Omega}^{\mathcal{P}}$-term, the operator $\mathcal{H}[s]$ is defined by

$$\mathcal{H}[s](X) = \mathcal{H}(X \,\cup\, \{|s|\}).$$

Likewise, if $\mathfrak{X}$ is a formula or a sequent we define

$$\mathcal{H}[\mathfrak{X}](X) = \mathcal{H}(X \,\cup\, |\mathfrak{X}|).$$

If $\mathfrak{Y}_i$ is a term, or a formula, or a sequent for $1 \le i \le n$, we let $\mathcal{H}[\mathfrak{Y}_1, \mathfrak{Y}_2] = (\mathcal{H}[\mathfrak{Y}_1])[\mathfrak{Y}_2]$, $\mathcal{H}[\mathfrak{Y}_1, \mathfrak{Y}_2, \mathfrak{Y}_3] = (\mathcal{H}[\mathfrak{Y}_1, \mathfrak{Y}_2])[\mathfrak{Y}_3]$, etc.

Lemma 4.4 *Let $\mathcal{H}$ be an operator. Let s be a term and $\mathfrak{X}$ be a formula or a sequent.*

(i) $\forall X, X' {\in} P(On)[X' \subseteq X \Longrightarrow \mathcal{H}(X') \subseteq \mathcal{H}(X)]$.

(ii) $\mathcal{H}[s]$ *and* $\mathcal{H}[\mathfrak{X}]$ *are operators.*

(iii) $|\mathfrak{X}| \subseteq \mathcal{H}[\emptyset] \implies \mathcal{H}[\mathfrak{X}] = \mathcal{H}$.

(iv) $|s| \in \mathcal{H}[\emptyset] \implies \mathcal{H}[s] = \mathcal{H}$.

Since we also want to keep track of the complexity of cuts appearing in derivations, we endow each formula with an ordinal rank.

Definition 4.5 The *rank* of a formula is determined as follows:

(1) $rk(s{\in}t) := rk(s{\notin}t) := \max\{|s|+1, |t|+1\}$.

(2) $rk((\exists x{\in}t)F(x)) := rk((\forall x{\in}t)F(x)) := \max\{|t|, rk(F(\mathbb{V}_0)) + 2\}$.

(3) $rk((\exists x \subseteq t)F(x)) := rk((\forall x \subseteq t)F(x)) := \max\{|t|, rk(F(\mathbb{V}_0)) + 2\}$.

(4) $rk(\exists x\, F(x)) := rk(\forall x\, F(x)) := \max\{\Omega, rk(F(\mathbb{V}_0)) + 2\}$.

(5) $rk(A \wedge B) := rk(A \vee B) := \max\{rk(A), rk(B)\} + 1$.

Note that for a $\Delta_0^{\mathcal{P}}$ formula A we have $rk(A) < \Omega$.

There is plenty of leeway in designing the actual rank of a formula.

Definition 4.6 Let $\mathcal{H}$ be an operator and let Λ be a finite set of $RS_{\Omega}^{\mathcal{P}}$-formulae. Then $\mathcal{H} \vdash^{\alpha}_{\rho} \Lambda$ is defined by recursion on α.

If Λ is an *axiom* and $|\Lambda| \cup \{\alpha\} \subseteq \mathcal{H}(\emptyset)$, then $\mathcal{H} \vdash^{\alpha}_{\rho} \Lambda$.

Moreover, we have inductive clauses pertaining to the inference rules of $RS_{\Omega}^{\mathcal{P}}$, which come with the additional requirement that $|\Lambda| \cup \{\alpha\} \subseteq \mathcal{H}(\emptyset)$, where Λ is the sequent of the conclusion. We shall not repeat this requirement below. The clauses are the following:

$$(\wedge) \qquad \frac{\mathcal{H} \vdash^{\alpha_0}_{\rho} \Gamma, A_0 \qquad \mathcal{H} \vdash^{\alpha_0}_{\rho} \Gamma, A_1}{\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, A_0 \wedge A_1} \qquad \alpha_0 < \alpha$$

$$(\vee) \qquad \frac{\mathcal{H} \vdash^{\alpha_0}_{\rho} \Lambda, A_i}{\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, A_0 \vee A_1} \qquad \begin{array}{c} \alpha_0 < \alpha \\ i \in \{0,1\} \end{array}$$

$$(b\forall)_{\infty} \qquad \frac{\mathcal{H}[s] \vdash^{\alpha_s}_{\rho} \Gamma, s{\in}t \to F(s) \text{ for all } |s| < |t|}{\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, (\forall x{\in}t)F(x)} \qquad |s| \leq \alpha_s < \alpha$$

$$(b\exists) \qquad \frac{\mathcal{H} \vdash^{\alpha_0}_{\rho} \Gamma, s{\in}t \wedge F(s)}{\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, (\exists x{\in}t)F(x)} \qquad \begin{array}{c} \alpha_0 < \alpha \\ |s| < |t| \\ |s| < \alpha \end{array}$$

$$(pb\forall)_{\infty} \qquad \frac{\mathcal{H}[s] \vdash^{\alpha_s}_{\rho} \Gamma, s \subseteq t \to F(s) \text{ for all } |s| \leq |t|}{\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, (\forall x \subseteq t)F(x)} \qquad |s| \leq \alpha_s < \alpha$$

$$(pb\exists) \qquad \frac{\mathcal{H} \vdash^{\alpha_0}_{\rho} \Gamma, s \subseteq t \wedge F(s)}{\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, (\exists x \subseteq t)F(x)} \qquad \begin{array}{c} \alpha_0 < \alpha \\ |s| \leq |t| \\ |s| < \alpha \end{array}$$

$$(\forall)_{\infty} \qquad \frac{\mathcal{H}[s] \vdash^{\alpha_s}_{\rho} \Gamma, F(s) \text{ for all } s}{\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, (\forall x F(x)} \qquad |s| \leq \alpha_s + 1 < \alpha$$

$$(\exists) \qquad \frac{\mathcal{H} \vdash^{\alpha_0}_{\rho} \Gamma, F(s)}{\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, \exists x F(x)} \qquad \begin{array}{c} \alpha_0 < \alpha \\ |s| < \alpha \end{array}$$

$$(\notin)_{\infty} \qquad \frac{\mathcal{H}[r] \vdash^{\alpha_r}_{\rho} \Gamma, r{\in}t \to r \neq s \text{ for all } |r| < |t|}{\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, s \notin t} \qquad |r| \leq \alpha_r < \alpha$$

$$(\in) \qquad \frac{\mathcal{H} \vdash^{\alpha_0}_{\rho} \Gamma, r{\in}t \wedge r = s}{\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, s{\in}t} \qquad \begin{array}{c} \alpha_0 < \alpha \\ |r| < |t| \\ |r| < \alpha \end{array}$$

$$(\not\subseteq)_\infty \qquad \frac{\mathcal{H}[r] \vdash^{\alpha_r}_{\rho} \Gamma, r \subseteq t \to r \neq s \text{ for all } |r| \le |t|}{\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, s \not\subseteq t} \qquad |r| \le \alpha_r < \alpha$$

$$(\subseteq) \qquad \frac{\mathcal{H} \vdash^{\alpha_0}_{\rho} \Gamma, r \subseteq t \wedge r = s}{\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, s \subseteq t} \qquad \begin{array}{r} \alpha_0 < \alpha \\ |r| \le |t| \\ |r| < \alpha \end{array}$$

$$(\text{Cut}) \qquad \frac{\mathcal{H} \vdash^{\alpha_0}_{\rho} \Lambda, B \qquad \mathcal{H} \vdash^{\alpha_0}_{\rho} \Lambda, \neg B}{\mathcal{H} \vdash^{\alpha}_{\rho} \Lambda} \qquad \begin{array}{r} \alpha_0 < \alpha \\ rk(B) < \rho \end{array}$$

$$(\Sigma^{\mathcal{P}}\text{-Ref}) \qquad \frac{\mathcal{H} \vdash^{\alpha_0}_{\rho} \Gamma, A}{\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, \exists z\, A^z} \qquad \begin{array}{r} \alpha_0, \Omega < \alpha \\ A \in \Sigma^{\mathcal{P}} \end{array}$$

Remark 4.7 Suppose $\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma(s_1, \ldots, s_n)$, where $\Gamma(a_1, \ldots, a_n)$ is a sequent of $\mathbf{KP}(\mathcal{P})$ and $s_1, \ldots, s_n$ are $RS^{\mathcal{P}}_{\Omega}$-terms. Then we have that $|s_1|, \ldots, |s_n| \in \mathcal{H}(\emptyset)$. Standing in sharp contrast to the ordinal analysis of **KP** (cf. [**7, 13**]), however, the terms s_i may and often will contain subterms that the operator $\mathcal{H}$ does *not* control, that is, subterms t with $|t| \notin \mathcal{H}(\emptyset)$.

The following observation is easily established by induction on α.

Lemma 4.8 (Weakening)

$$\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma \ \wedge\ \alpha \le \alpha' \in \mathcal{H} \ \wedge\ \rho \le \rho' \ \wedge\ |\Lambda| \subseteq \mathcal{H}(\emptyset) \implies \mathcal{H} \vdash^{\alpha'}_{\rho'} \Gamma, \Lambda.$$

Lemma 4.9 (Inversion)

(i) *If* $\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, A \vee B$ *and* $rk(A \vee B) \ge \Omega$*, then* $\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, A, B$.

(ii) *If* $\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, A_0 \wedge A_1$, $i \in \{0, 1\}$ *and* $rk(A_0 \wedge A_1) \ge \Omega$*, then* $\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, A_i$.

(iii) $\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, \forall x\, F(x) \ \wedge\ \gamma \in \mathcal{H}(\emptyset) \ \wedge\ \gamma < \Omega \implies \mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, (\forall x \in \mathbb{V}_\gamma) F(x)$.

(iv) *If* $\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, (\forall x \in t)\, F(x)$ *and* $rk(F(\mathbb{V}_0)) \ge \Omega$*, then* $\mathcal{H}[s] \vdash^{\alpha}_{\rho} \Gamma, s \in t \to F(s)$ *for all* $|s| < |t|$.

(v) *If* $\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, (\forall x \subseteq t)\, F(x)$ *and* $rk(F(\mathbb{V}_0)) \ge \Omega$*, then* $\mathcal{H}[s] \vdash^{\alpha}_{\rho} \Gamma, s \subseteq t \to F(s)$ *for all* $|s| \le |t|$.

Proof. All proofs are by induction on α. Note that if a formula C of $rk(C) \ge \Omega$ is active in a derivation then it must have been the principal formula of an inference.

We show (iii). Suppose that $\forall x\, F(x)$ was the principal formula of the last inference. Then we have $\mathcal{H}[s] \vdash^{\alpha_s}_{\rho} \Gamma, \forall x\, F(x), F(s)$ for all terms s, using weakening (Lemma 4.8) if $\forall x\, F(x)$ was not a side formula of the inference. Moreover, $|s| \le \alpha_s + 1 < \alpha$ holds for all s. Inductively we have $\mathcal{H}[s] \vdash^{\alpha_s}_{\rho} \Gamma, (\forall x \in \mathbb{V}_\gamma) F(x), F(s)$ for all $|s| < \gamma$. As

$$\mathcal{H}[s] \vdash^{\alpha_s}_{\rho} \Gamma, (\forall x \in \mathbb{V}_\gamma) F(x), s \in \mathbb{V}_\gamma$$

holds for $|s| < \gamma$ on account of being an axiom, we get

$$\mathcal{H}[s] \vdash^{\alpha_s+1}_{\rho} \Gamma, (\forall x{\in}\mathbb{V}_\gamma)F(x), s \in \mathbb{V}_\gamma \wedge F(s)$$

for all $|s| < \gamma$ via an inference $(\wedge)$, and hence $\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, (\forall x{\in}\mathbb{V}_\gamma)F(x)$ via an inference $(b\forall)$.

If $\forall x\, F(x)$ is not the principal formula of the last inference, then the assertion follows by using the induction hypothesis to its premises and re-applying the same inference. □

5 Embedding

To connect $\mathbf{KP}(\mathcal{P})$ with the infinitary system $RS^{\mathcal{P}}_{\Omega}$, we show that $\mathbf{KP}(\mathcal{P})$ can be embedded into $RS^{\mathcal{P}}_{\Omega}$. Indeed, the finite $\mathbf{KP}(\mathcal{P})$-derivations give rise to very uniform infinitary derivations.

Definition 5.1 For $\Gamma = \{A_1, \ldots, A_n\}$, let

$$no(\Gamma) := \omega^{rk(A_1)} \# \cdots \# \omega^{rk(A_n)}.$$

We define

$$\Vdash \Gamma \;:\Longleftrightarrow\; \text{for all operators } \mathcal{H},\; \mathcal{H}[\Gamma] \vdash^{no(\Gamma)}_{0} \Gamma$$

and

$$\Vdash^{\xi}_{\rho} \Gamma \;:\Longleftrightarrow\; \text{for all operators } \mathcal{H},\; \mathcal{H}[\Gamma] \vdash^{no(\Gamma)\#\xi}_{\rho} \Gamma.$$

Lemma 5.2 *For all formulae A,*

$$\Vdash A, \neg A.$$

Proof. We proceed by induction on the syntactic complexity of A. For A in $\Delta^{\mathcal{P}}_0$, this is an axiom of $RS^{\mathcal{P}}_{\Omega}$. Suppose that A is of the form $\forall x F(x)$. Let $\mathcal{H}$ be an arbitrary operator. Let $\alpha_s := no(\{F(s), \neg F(s)\})$ and $\alpha := no(\{\forall x F(x), \exists x \neg F(x)\})$. Inductively we have $\mathcal{H}[F(s)] \vdash^{\alpha_s}_{0} F(s), \neg F(s)$ for all terms s. Using an inference $(\exists)$ we get $\mathcal{H}[F(s)] \vdash^{no(\{F(s), \exists x \neg F(x)\})}_{0} F(s), \exists x \neg F(x)$. Hence, via an inference $(\forall)$, we arrive at $\mathcal{H}[\forall x F(x)] \vdash^{\alpha}_{0} \forall x F(x), \exists x \neg F(x)$, noting that $\mathcal{H}[F(s)\}] \subseteq (\mathcal{H}[\forall x \neg F(x)])[s]$.

The other cases are similar. □

Lemma 5.3 (Equality and Extensionality)

$$\Vdash s_1 \neq t_1, \ldots, s_n \neq t_n, \neg A(s_1, \ldots, s_n), A(t_1, \ldots, t_n).$$

Proof. We proceed by induction on the buildup of $A(\vec{s})$. If $A(\vec{s})$ is $\Delta^{\mathcal{P}}_0$ then this is an axiom.

Suppose $A(\vec{s})$ is a formula $\forall x\, F(x, \vec{s})$. Let $\vec{s} \neq \vec{t}$ stand for $s_1 \neq t_1, \ldots, s_n \neq t_n$. Let $\Gamma_r := \{\vec{s} \neq \vec{t}, \neg F(r, \vec{s}), F(r, \vec{t})\}$ and $\alpha_r := no(\Gamma_r)$. Let $\mathcal{H}$ be an arbitrary operator. Inductively we have

$$\mathcal{H}[\Gamma_r] \vdash^{\alpha_r}_{0} \Gamma_r$$

for all terms r. Using an inference $(\exists)$ we obtain $\mathcal{H}[\tilde{\Gamma}_r] \vdash^{\Gamma}_{0} \tilde{\Gamma}_r$, where

$$\tilde{\Gamma}_r := \{\vec{s} \neq \vec{t}, \exists x \neg F(x, \vec{s}), F(r, \vec{t})\}$$

and $\tilde{\alpha}_r := no(\tilde{\Gamma}_r)$, noting that $|r| < \Omega \leq no(\exists x \neg F(x, \vec{s}))$. Thus, using an inference $(\forall)$, we have

$$\mathcal{H}[\Gamma] \mid\!\frac{no(\Gamma)}{0}\, \Gamma,$$

where $\Gamma := \{\vec{s} \neq \vec{t}, \exists x \neg F(x, \vec{s}), \forall x F(x, \vec{t})\}$. In the latter we used the fact that $\mathcal{H}[\tilde{\Gamma}_r] \subseteq (\mathcal{H}[\Gamma])[r]$.

The other cases are similar. □

Lemma 5.4 (Set Induction)

$$\Vdash \forall x\, [(\forall y{\in}x) F(y) \to F(x)] \;\longrightarrow\; \forall x F(x).$$

Proof. Fix an operator $\mathcal{H}$. Let $A \equiv (\forall x\, [(\forall y{\in}x) F(y) \to F(x)]$. First, we show, by induction on $|s|$, that

$$(+) \qquad \mathcal{H}[A, s] \mid\!\frac{\omega^{rk(A)} \# \omega^{|s|+1}}{0}\, \neg A, F(s).$$

So assume that

$$\mathcal{H}[A, t] \mid\!\frac{\omega^{rk(A)} \# \omega^{|t|+1}}{0}\, \neg A, F(t)$$

holds for all $|t| < |s|$. Using $(\vee)$, this yields

$$\mathcal{H}[A, s, t] \mid\!\frac{\omega^{rk(A)} \# \omega^{|t|+1} + 1}{0}\, \neg A, t{\in}s \to F(t)$$

for all $|t| < |s|$, and hence

$$(5.1) \qquad \mathcal{H}[A, s] \mid\!\frac{\omega^{rk(A)} \# \omega^{|s|} + 2}{0}\, \neg A, (\forall x{\in}s) F(x)$$

via $(\forall)_\infty$. Set $\eta_s := \omega^{rk(A)} \# \omega^{|s|} + 2$. By Lemma 5.2 we have $\mathcal{H}[A, s] \mid\!\frac{\eta_s}{0}\, \neg F(s), F(s)$. Therefore, using (5.1) and $(\wedge)$,

$$\mathcal{H}[A, s] \mid\!\frac{\eta_s + 1}{0}\, \neg A, (\forall y{\in}s) F(y) \wedge \neg F(s), F(s).$$

From the latter we obtain

$$\mathcal{H}[A, s] \mid\!\frac{\eta_s + 2}{0}\, \neg A, \exists x\, [(\forall y{\in}x) F(y) \wedge \neg F(x)], F(s)$$

via $(\exists)$. This shows (+).

Finally, (+) enables us to deduce, via $(\forall)_\infty$, that

$$\mathcal{H}[A, s] \mid\!\frac{\omega^{rk(A)} + \Omega}{0}\, \neg A, \forall x F(x).$$

From this the assertion follows by applying $(\vee)$ twice. □

Lemma 5.5 (Infinity Axiom) *For any operator $\mathcal{H}$ we have*

$$\mathcal{H} \mid\!\frac{\omega + 2}{0}\, \exists x\, [(\exists y{\in}x)\, y \in x \;\wedge\; (\forall y \in x)(\exists z{\in}x)\, y \in z].$$

Proof. Let s be a term with $|s| = n < \omega$. Then $\mathcal{H} \mid\!\frac{0}{0}\, s{\in}\mathbb{V}_{n+1}$ and $\mathcal{H} \mid\!\frac{0}{0}\, \mathbb{V}_{n+1} \in \mathbb{V}_\omega$ since these formulae are axioms. Via $(\wedge)$ we deduce $\mathcal{H} \mid\!\frac{1}{0}\, \mathbb{V}_{n+1} \in \mathbb{V}_\omega \wedge s{\in}\mathbb{V}_{n+1}$ and hence $\mathcal{H} \mid\!\frac{n+2}{0}\, (\exists z \in \mathbb{V}_\omega) s{\in}z$, using $(b\exists)$. An inference $(\vee)$ yields

$$\mathcal{H} \mid\!\frac{n+3}{0}\, s{\in}\mathbb{V}_\omega \to (\exists z \in \mathbb{V}_\omega) s{\in}z.$$

Since this holds for all terms s with $|s| < \omega$, we conclude that

$$(5.2) \qquad \mathcal{H} \mid\!\frac{\omega}{0}\, (\forall y{\in}\mathbb{V}_\omega)(\exists z \in \mathbb{V}_\omega) y{\in}z.$$

Since $\mathbb{V}_0 \in \mathbb{V}_\omega$ is an axiom we have $\mathcal{H} \vdash^{1}_{0} \mathbb{V}_0 \in \mathbb{V}_\omega \ \wedge \ \mathbb{V}_0 \in \mathbb{V}_\omega$ via $(\wedge)$ and thus

$$\mathcal{H} \vdash^{2}_{0} (\exists z \in \mathbb{V}_\omega) z \in \mathbb{V}_\omega, \tag{5.3}$$

using $(b\exists)$. Combining (5.2) and (5.3) we arrive at

$$\mathcal{H} \vdash^{\omega+1}_{0} (\exists z \in \mathbb{V}_\omega) z \in \mathbb{V}_\omega \ \wedge \ (\forall y {\in} \mathbb{V}_\omega)(\exists z \in \mathbb{V}_\omega) y {\in} z.$$

Thus an inference $(b\exists)$ furnishes us with

$$\mathcal{H} \vdash^{\omega+2}_{0} \exists x \, [(\exists z \in x) z \in x \ \wedge \ (\forall y {\in} x)(\exists z {\in} x) y {\in} z]. \qquad \square$$

Lemma 5.6 ($\Delta_0^{\mathcal{P}}$-Separation) *Let $A(a, b, c_1, \ldots, c_n)$ be a $\Delta_0^{\mathcal{P}}$-formula of $\mathcal{L}$ with all free variables among the exhibited. Let $r, s_1, \ldots, s_n$ be $RS_\Omega^{\mathcal{P}}$-terms. Let $\mathcal{H}$ be an arbitrary operator. Then:*

$$\mathcal{H}[r, \vec{s}\,] \vdash^{s}_{\rho} \exists y \, [(\forall x {\in} y)(x {\in} r \wedge A(x, r, \vec{s}\,) \ \wedge \ (\forall x {\in} r)(A(x, r, \vec{s}\,) \rightarrow x {\in} y)],$$

where $\alpha = |r|$ and $\rho = \max\{|r|, |s_1|, \ldots, |s_n|\} + \omega$.

Proof. Define the $RS_\Omega^{\mathcal{P}}$-term p by

$$p := \{x \in \mathbb{V}_\alpha \mid x {\in} r \ \wedge \ A(x, r, \vec{s}\,)\}.$$

Then $|p| = \alpha$. Let $\tilde{\mathcal{H}} := \mathcal{H}[r, \vec{s}\,]$. We have $\tilde{\mathcal{H}}[t] \vdash^{H}_{0} t \notin p, t \in r \ \wedge \ A(t, r, \vec{s}\,)$ for all $|t| < \alpha$ since this is an axiom. Hence $\tilde{\mathcal{H}}[t] \vdash^{H}_{0} t {\in} p \rightarrow t \in r \ \wedge \ A(t, r, \vec{s}\,)$ using $(\vee)$ twice, and therefore

$$\tilde{\mathcal{H}} \vdash^{H}_{0} (\forall x {\in} p)(x \in r \ \wedge \ A(x, r, \vec{s}\,)) \tag{5.4}$$

by applying $(b\forall)_\infty$. We also have $\tilde{\mathcal{H}}[t] \vdash^{H}_{0} t \notin r, t {\in} r$ and $\tilde{\mathcal{H}}[t] \vdash^{H}_{0} \neg A(t, r, \vec{s}\,), A(t, r, \vec{s}\,)$ as these sequents are axioms. Using $(\wedge)$ and weakening (Lemma 4.8) we conclude that

$$\tilde{\mathcal{H}}[t] \vdash^{H}_{0} t \notin r, \neg A(t, r, \vec{s}\,), t {\in} r \ \wedge \ A(t, r, \vec{s}\,). \tag{5.5}$$

Since $\tilde{\mathcal{H}}[t] \vdash^{H}_{0} \neg(t {\in} r \ \wedge \ A(t, r, \vec{s}\,)), t {\in} p$ holds on account of being an axiom, a cut applied to (5.5) and the latter yields

$$\tilde{\mathcal{H}}[t] \vdash^{H}_{\rho} t \notin r, \neg A(t, r, \vec{s}\,), t {\in} p, \tag{5.6}$$

since $rk(t {\in} r \ \wedge \ A(t, r, \vec{s}\,)) < \rho$ holds for terms t with $|t| < \alpha$. Now use $(\vee)$ four times to arrive at

$$\tilde{\mathcal{H}}[t] \vdash^{H}_{\rho} t {\in} r \rightarrow (A(t, r, \vec{s}\,) \rightarrow t {\in} p). \tag{5.7}$$

Applying $(b\forall)_\infty$ to (5.7) yields

$$\tilde{\mathcal{H}} \vdash^{H}_{\rho} (\forall x \in r)(A(x, r, \vec{s}\,) \rightarrow x \in p). \tag{5.8}$$

Combining (5.4) and (5.8) via $(\wedge)$ we have

$$\tilde{\mathcal{H}} \vdash^{H}_{\rho} (\forall x {\in} p)(x \in r \ \wedge \ A(x, r, \vec{s}\,)) \ \wedge \ (\forall x \in r)(A(x, r, \vec{s}\,) \rightarrow x {\in} p).$$

Consequently, by means of $(b\exists)$,

$$\tilde{\mathcal{H}} \vdash^{H}_{\rho} \exists y[(\forall x{\in}y)(x \in r \wedge A(x,r,\vec{s})) \ \wedge\ (\forall x \in r)(A(x,r,\vec{s}) \to x{\in}y)]. \qquad \square$$

Lemma 5.7 (Pair and Union) *For any operator $\mathcal{H}$, the following hold:*

(i) $\mathcal{H}[s,t] \vdash^{\alpha+1}_{0} \exists z\,(s{\in}z \wedge t{\in}z)$ *, where* $\alpha = \max(|s|,|t|)+1$.

(ii) $\mathcal{H}[s] \vdash^{\beta+7}_{0} \exists z\,(\forall y{\in}s)(\forall x{\in}y)(x{\in}z)$ *, where* $\beta = |s|$.

Proof. For (i), $s \in \mathbb{V}_\alpha$ and $t \in \mathbb{V}_\alpha$ are axioms. Thus $\mathcal{H}[s,t] \vdash^{1}_{0} s \in \mathbb{V}_\alpha \wedge t \in \mathbb{V}_\alpha$, and hence $\mathcal{H}[s,t] \vdash^{\alpha+2}_{0} \exists z\,(s{\in}z \wedge t{\in}z)$ by means of $(b\exists)$.

For (ii), let r and t be terms of levels $< \beta$. Since $r \in \mathbb{V}_\beta$ is an axiom, we have

$$\mathcal{H}[s] \vdash^{0}_{0} t \notin s, r \notin t, r \in \mathbb{V}_\beta.$$

Thus we get

$$\begin{aligned}
&\mathcal{H}[s] \vdash^{2}_{0} t \notin s, r{\in}t \to r \in \mathbb{V}_\beta\\
&\mathcal{H}[s] \vdash^{\beta+3}_{0} t \notin s, (\forall x{\in}t)x \in \mathbb{V}_\beta\\
&\mathcal{H}[s] \vdash^{\beta+5}_{0} t{\in}s \to (\forall x{\in}t)x \in \mathbb{V}_\beta\\
&\mathcal{H}[s] \vdash^{\beta+6}_{0} (\forall y{\in}s)(\forall x{\in}t)x \in \mathbb{V}_\beta\\
&\mathcal{H}[s] \vdash^{\beta+7}_{0} \exists z\,(\forall y{\in}s)(\forall x{\in}t)x{\in}z. \qquad \square
\end{aligned}$$

Lemma 5.8 (Power Set) *For any operator $\mathcal{H}$, the following holds:*

$$\mathcal{H}[s] \vdash^{\alpha+1}_{0} \exists z\,(\forall x \subseteq s)\, x \in z,$$

where $\alpha = |s|$.

Proof. Let t be a term with $|t| \leq \alpha$. Then $t \in \mathbb{V}_{\alpha+1}$ is an axiom. Whence, using $(\vee)$ (twice), $(pb\forall)_\infty$, and $(\exists)$, we have

$$\begin{aligned}
&\mathcal{H}[s] \vdash^{0}_{0} t \not\subseteq s, t \in \mathbb{V}_{\alpha+1}\\
&\mathcal{H}[s] \vdash^{2}_{0} t \subseteq s \to t \in \mathbb{V}_{\alpha+1}\\
&\mathcal{H}[s] \vdash^{\alpha+3}_{0} (\forall x \subseteq s)x \in \mathbb{V}_{\alpha+1}\\
&\mathcal{H}[s] \vdash^{\alpha+4}_{0} \exists z\,(\forall x \subseteq s)x \in z. \qquad \square
\end{aligned}$$

Theorem 5.9 *If* $\mathbf{KP}(\mathcal{P}) \vdash \Gamma(a_1,\ldots,a_l)$, *then there exist* $m,n < \omega$ *such that*

$$\mathcal{H}[s_1,\ldots,s_l\,] \vdash^{\omega^{\Omega+m}}_{\Omega+n} \Gamma(s_1,\ldots,s_l)$$

holds for all $RS^{\mathcal{P}}_{\Omega}$*-terms* $s_1,\ldots,s_l$ *and operators* $\mathcal{H}$*. Thus* m *and* n *depend only on the* $\mathbf{KP}(\mathcal{P})$*-derivation of* $\Gamma(\vec{a})$.

Proof. One proceeds by induction on the length of the $\mathbf{KP}(\mathcal{P})$-derivation of $\Gamma(\vec{a}\,)$. Note that the rank of an $RS^{\mathcal{P}}_{\Omega}$-formula A is always $< \Omega+\omega$ and thus the norms of $RS^{\mathcal{P}}_{\Omega}$-sequents will always be $< \omega^{\Omega+\omega}$.

If $\Gamma(\vec{a}\,)$ is an axiom of $\mathbf{KP}(\mathcal{P})$ then the assertion follows from the earlier results of this section.

As an example of a rule we shall treat $(pb\exists)$. So suppose the last inference of our $\mathbf{KP}(\mathcal{P})$-derivation $\mathcal{D}$ was an instance of $(pb\exists)$. Then $\Gamma(\vec{a}\,)$ contains a formula of the form $(\exists x \subseteq a_i) \wedge F(x, \vec{a}\,)$ and there exists a shorter $\mathbf{KP}(\mathcal{P})$-derivation $\mathcal{D}_0$ whose end sequent is either of the form $\Gamma(\vec{a}\,), c \subseteq a_i \wedge F(c, \vec{a}\,)$ with c not occurring in $\Gamma(\vec{a}\,)$ or c is a_j for some $1 \leq j \leq l$. In the former case the induction hypothesis supplies us with $n_0, m_0 < \omega$ such that

$$\mathcal{H}[\vec{s}\,] \vdash^{\omega^{\Omega+m_0}}_{\Omega+n_0} \Gamma(\vec{s}\,), \mathbb{V}_0 \subseteq s_i \wedge F(\mathbb{V}_0, \vec{s}\,) \tag{5.9}$$

holds for all terms $\vec{s}$. As $|\mathbb{V}_0| = 0 \leq |s_i|$ we can apply an inference $(pb\exists)$ in the system $RS^{\mathcal{P}}_{\Omega}$, yielding

$$\mathcal{H}[\vec{s}\,] \vdash^{\omega^{\Omega+m_0+2}}_{\Omega+n_0} \Gamma(\vec{s}\,), (\exists x \subseteq s_i) F(x, \vec{s}\,) \tag{5.10}$$

and thus $\mathcal{H}[\vec{s}\,] \vdash^{\omega^{\Omega+m_0+2}}_{\Omega+n_0} \Gamma(\vec{s}\,)$ as $(\exists x \subseteq s_i) F(x, \vec{s}\,)$ belongs to $\Gamma(\vec{s}\,)$.

Now let us turn to the case where c is a_j. Then, by the induction hypothesis, there are $n_0, m_0 < \omega$ such that

$$\mathcal{H}[\vec{s}\,] \vdash^{\omega^{\Omega+m_0}}_{\Omega+n_0} \Gamma(\vec{s}\,), s_j \subseteq s_i \wedge F(s_j, \vec{s}\,) \tag{5.11}$$

holds for all terms $\vec{s}$. Owing to Lemma 5.3 we can find m_1, n_1 such that with $\rho := \omega^{\Omega+m_1}$ we have

$$\mathcal{H}[\vec{s}, r] \vdash^{S}_{\Omega+n_1} s_j \neq r, s_j \not\subseteq s_i, r \subseteq s_i$$

and $\mathcal{H}[\vec{s}, r] \vdash^{S}_{\Omega+n_1} s_j \neq r, \neg F(s_j, \vec{s}\,), F(r, \vec{s}\,)$ hold for all $r, \vec{s}$. By applying weakening and $(\wedge)$ we thus get

$$\mathcal{H}[\vec{s}, r] \vdash^{S}_{\Omega+n_1} r \not\subseteq s_i, s_j \neq r, \neg F(s_j, \vec{s}\,), r \subseteq s_i \wedge F(r, \vec{s}\,)$$

for all r with $|r| \leq |s_i|$. Now apply $(pb\exists)$, $(\vee)$ (twice), $(\not\subseteq)_\infty$, and $(\vee)$ (twice):

$$\begin{aligned}
&\mathcal{H}[\vec{s}, r] \vdash^{S}_{\Omega+n_1} r \not\subseteq s_i, s_j \neq r, \neg F(s_j, \vec{s}\,), (\exists x \subseteq s_i) F(x, \vec{s}\,) \\
&\mathcal{H}[\vec{s}, r] \vdash^{S}_{\Omega+n_1} r \subseteq s_i \to s_j \neq r, \neg F(s_j, \vec{s}\,), (\exists x \subseteq s_i) F(x, \vec{s}\,) \\
&\mathcal{H}[\vec{s}\,] \vdash^{S}_{\Omega+n_1} s_j \not\subseteq s_i, \neg F(s_j, \vec{s}\,), (\exists x \subseteq s_i) F(x, \vec{s}\,) \\
&\mathcal{H}[\vec{s}\,] \vdash^{S}_{\Omega+n_1} \neg(s_j \subseteq s_i \ \wedge\ F(s_j, \vec{s}\,)), (\exists x \subseteq s_i) F(x, \vec{s}\,).
\end{aligned} \tag{5.12}$$

Finally, by applying a cut to (5.11) and (5.12) we have

$$\mathcal{H}[\vec{s}\,] \vdash^{S}_{\Omega+n} \Gamma(\vec{s}\,), (\exists x \subseteq s_i) F(x, \vec{s}\,),$$

i.e., $\mathcal{H}[\vec{s}\,] \vdash^{S}_{\Omega+n} \Gamma(\vec{s}\,)$, where $m = \max(m_0, m_1) + 1$ and n is chosen such that $n > n_0, n_1$ and $rk(s_j \subseteq s_i \ \wedge\ F(s_j, \vec{s}\,) < \Omega + n$ for all $\vec{s}$.

The case of the last inference being $(b\exists)$ is treated in the same vein as $(pb\exists)$. All the other inferences are straightforward as the desired assertion can be obtained immediately from the induction hypothesis applied to the premises followed by the corresponding inference in $RS^{\mathcal{P}}_{\Omega}$. For example, in the case of the ($\Delta^{\mathcal{P}}_0$-COLLR) one inductively finds $m_0, n_0 < \omega$ such that

$$\mathcal{H}[\vec{s}\,] \vdash^{S}_{\Omega+n} \Gamma_0(\vec{s}\,), (\forall x{\in}s_i)\exists y\, H(x, y, \vec{s}\,)$$

holds for all $\vec{s}$, where $H(x, y, \vec{a}\,)$ is $\Sigma^{\mathcal{P}}$. Using ($\Sigma^{\mathcal{P}}$-*Ref*) one obtains

$$\mathcal{H}[\vec{s}\,] \vdash^{S}_{\Omega+n} \Gamma_0(\vec{s}\,), \exists z(\forall x{\in}s_i)(\exists y \in z)\, H(x, y, \vec{s}\,). \qquad \square$$

6 Cut elimination

The usual cut elimination procedure works as long as the cut formulae are not in $\Delta^{\mathcal{P}}_0$ and have not been introduced by an inference ($\Sigma^{\mathcal{P}}$-Ref). As the principal formula of an inference ($\Sigma^{\mathcal{P}}$-Ref) has rank Ω, one gets the following result.

Theorem 6.1 (Cut elimination I)

$$\mathcal{H} \vdash^{\alpha}_{\Omega+n+1} \Gamma \quad \Longrightarrow \quad \mathcal{H} \vdash^{\omega_n(\alpha)}_{\Omega+1} \Gamma$$

where $\omega_0(\beta) := \beta$ *and* $\omega_{k+1}(\beta) := \omega^{\omega_k(\beta)}$.

Proof. The proof is standard. For details, see [**7**, Lemma 3.14]. □

Lemma 6.2 (Boundedness) *Let* A *be a* $\Sigma^{\mathcal{P}}$*-formula,* $\alpha \le \beta < \Omega$*, and* $\beta \in \mathcal{H}(\emptyset)$*. If*

$$\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, A$$

then

$$\mathcal{H} \vdash^{\alpha}_{\rho} \Gamma, A^{\mathbb{V}_\beta}.$$

Proof. Note that the derivation contains no instances of ($\Sigma^{\mathcal{P}}$-*Ref*). The proof is by induction on α. For details, see [**7**, Lemma 3.17]. □

The obstacle to pushing cut elimination further is exemplified by the following scenario:

$$\dfrac{\dfrac{\mathcal{H} \vdash^{\delta}_{\Omega} \Gamma, A}{\mathcal{H} \vdash^{\xi}_{\Omega} \Gamma, \exists z\, A^z}\,(\Sigma^{\mathcal{P}}\text{-Ref}) \qquad \dfrac{\ldots \mathcal{H}[s] \vdash^{\xi_s}_{\Omega} \Gamma, \neg A^s \ldots (s \in \mathcal{T})}{\mathcal{H} \vdash^{\xi}_{\Omega} \Gamma, \forall z\, \neg A^z}\,(\forall)}{\mathcal{H} \vdash^{\alpha}_{\Omega+1} \Gamma}\,(\text{Cut}).$$

Fortunately, it is possible to eliminate cuts in the above situation provided that the side formulae Γ are of complexity $\Sigma^{\mathcal{P}}$. The technique is known as "collapsing" of derivations.

If the length of a derivation of $\Sigma^{\mathcal{P}}$-formulae is $\ge \Omega$, then "collapsing" results in a shorter derivation, however, at the cost of a much more complicated controlling operator.

Definition 6.3 $\mathcal{H}_\delta(X) = \bigcap\{C^{\Omega}(\alpha, \beta) :\ X \subseteq C^{\Omega}(\alpha, \beta) \ \wedge\ \delta < \alpha\}$.

Theorem 6.4 (Collapsing Theorem) *Let* Γ *be a set of* $\Sigma^{\mathcal{P}}$*-formulae for which we have* $|\Gamma| \subseteq C^{\Omega}(\eta + 1, \psi_{\Omega}(\eta + 1))$*. Suppose also that* $\eta \in \mathcal{H}_\eta[\Gamma](\emptyset)$*. Then*

$$\mathcal{H}_\eta[\Gamma] \vdash^{\alpha}_{\Omega+1} \Gamma \quad \Longrightarrow \quad \mathcal{H}_{\hat{\alpha}}[\Gamma] \vdash^{\psi_{\Omega}(\hat{\alpha})}_{\psi_{\Omega}(\hat{\alpha})} \Gamma,$$

where $\hat{\alpha} = \eta + \omega^{\Omega+\alpha}$.

Proof. By induction on α. Suppose $\mathcal{H}_\eta[\Gamma] \mathrel{\frac{\alpha}{\Omega+1}} \Gamma$. We shall distinguish cases according to the last inference of $\mathcal{H}_\eta[\Gamma] \mathrel{\frac{\alpha}{\Omega+1}} \Gamma$. Firstly, note that $\eta \in \mathcal{H}_\eta[\Gamma](\emptyset)$ implies $\eta \in \mathcal{H}_{\hat{\alpha}}[\Gamma](\emptyset)$, and therefore

$$\alpha \in \mathcal{H}_\eta[\Gamma](\emptyset) \implies \psi_\Omega(\hat{\alpha}) \in \mathcal{H}_{\hat{\alpha}}[\Gamma](\emptyset). \tag{6.1}$$

Case 0: Suppose Γ is an axiom. Then $\mathcal{H}_{\hat{\alpha}}[\Gamma] \mathrel{\frac{\psi_\Omega(\hat{\alpha})}{\psi_\Omega(\hat{\alpha})}} \Gamma$ follows immediately by (6.1).

Case 1: Suppose the last inference was $(pb\forall)_\infty$. Then there is an $A \in \Gamma$ of the form $(\forall x \subseteq t)F(x)$ and $\mathcal{H}_\eta[\Gamma][s] \mathrel{\frac{\alpha_s}{\Omega+1}} \Gamma, s \subseteq t \to F(s)$ and $\alpha_s < \alpha$ hold for all s with $|s| < |t|$. Since $|t| \in C^\Omega(\eta + 1, \psi_\Omega(\eta + 1)) \cap \Omega$ we have $|t| < \psi_\Omega(\eta + 1)$ and hence $|s| < \psi_\Omega(\eta + 1)$ whenever $|s| < |t|$. As a result, $|s| \in C^\Omega(\eta+1, \psi_\Omega(\eta+1))$ holds for all $|s| < |t|$. Therefore, by the induction hypothesis,

$$\mathcal{H}_{\hat{\alpha}_s}[\Gamma][s] \mathrel{\frac{\psi_\Omega(\hat{\alpha}_s)}{\psi_\Omega(\hat{\alpha}_s)}} \Gamma, s \subseteq t \to F(s) \tag{6.2}$$

for all $|s| < |t|$. Let $|s| < |t|$. Since $|s| < \psi_\Omega(\eta + 1)$, one computes that $\psi_\Omega(\hat{\alpha}_s) < \psi_\Omega(\hat{\alpha})$. Therefore, an inference $(pb\forall)_\infty$ applied to (6.2) yields $\mathcal{H}_{\hat{\alpha}}[\Gamma] \mathrel{\frac{\psi_\Omega(\hat{\alpha})}{\psi_\Omega(\hat{\alpha})}} \Gamma$.

The cases were the last inference is an instance of $(b\forall)_\infty$, $(\notin)_\infty$, $(\not\subseteq)_\infty$, or $(\wedge)$ are dealt with in a similar manner.

Case 2: Suppose the last inference was $(\exists)$. Then there is a formula $A \in \Gamma$ of the form $\exists x\, F(x)$ such that $\mathcal{H}_\eta[\Gamma] \mathrel{\frac{\alpha_0}{\Omega+1}} \Gamma, F(s)$ holds for some term s and $\alpha_0 < \alpha$. The induction hypothesis yields

$$\mathcal{H}_{\hat{\alpha}_0}[\Gamma] \mathrel{\frac{\psi_\Omega(\hat{\alpha}_0)}{\psi_\Omega(\hat{\alpha}_0)}} \Gamma, F(s).$$

Since $\alpha_0, |s| \in \mathcal{H}_\eta[\Gamma](\emptyset)$ and $|\Gamma| \subseteq C^\Omega(\eta + 1, \psi_\Omega(\eta + 1))$ we see that

$$\alpha_0, |s| \in C^\Omega(\eta + 1, \psi_\Omega(\eta + 1)).$$

Consequently, $|s|, \psi_\Omega(\hat{\alpha}_0) < \psi_\Omega(\hat{\alpha})$. Thus, via $(\exists)$ we conclude that $\mathcal{H}_{\hat{\alpha}}[\Gamma] \mathrel{\frac{\psi_\Omega(\hat{\alpha})}{\psi_\Omega(\hat{\alpha})}} \Gamma$.

The cases were the last inference is an instance of $(b\exists)$, $(\in)$, $(\subseteq)$, or $(\vee)$ are dealt with in a similar manner.

Case 3: Suppose $\exists z\, A^z \in \Gamma$ and $\mathcal{H}_\eta[\Gamma] \mathrel{\frac{\alpha_0}{\Omega+1}} \Gamma, A$ with $\alpha_0 < \alpha$. This means that the last inference was ($\Sigma^{\mathcal{P}}$-Ref). Note that $|A| = |\exists z\, A^z|$, and hence $\mathcal{H}_\eta[\Gamma, A] = \mathcal{H}_\eta[\Gamma]$. The induction hypothesis therefore yields $\mathcal{H}_{\hat{\alpha}_0}[\Gamma] \mathrel{\frac{\psi_\Omega(\hat{\alpha}_0)}{\psi_\Omega(\hat{\alpha}_0)}} \Gamma, A$ and therefore, as A is a $\Sigma^{\mathcal{P}}$-formula, we get $\mathcal{H}_{\hat{\alpha}_0}[\Gamma] \mathrel{\frac{\psi_\Omega(\hat{\alpha}_0)}{\psi_\Omega(\hat{\alpha}_0)}} \Gamma, A^{\mathbb{V}_{\psi_\Omega(\hat{\alpha}_0)}}$ by Lemma 6.2. Since $\psi_\Omega(\hat{\alpha}_0) \in \mathcal{H}_{\hat{\alpha}}$ and $\psi_\Omega(\hat{\alpha}_0) < \psi_\Omega(\hat{\alpha})$, an inference $(\exists)$ yields $\mathcal{H}_{\hat{\alpha}}[\Gamma] \mathrel{\frac{\psi_\Omega(\hat{\alpha})}{\psi_\Omega(\hat{\alpha})}} \Gamma, \exists z\, A^z$, i.e., $\mathcal{H}_{\hat{\alpha}}[\Gamma] \mathrel{\frac{\psi_\Omega(\hat{\alpha})}{\psi_\Omega(\hat{\alpha})}} \Gamma$.

Case 4: Suppose the last inference is (Cut). Then

$$\mathcal{H}_\eta[\Gamma] \mathrel{\frac{\alpha_0}{\Omega+1}} \Gamma, A \quad \text{and} \quad \mathcal{H}_\eta[\Gamma] \mathrel{\frac{\alpha_0}{\Omega+1}} \Gamma, \neg A,$$

where $\alpha_0 < \alpha$ and A is a formula with $rk(A) \leq \Omega$.

Since $|A| \subseteq \mathcal{H}_\eta[\Gamma](\emptyset)$ and $|\Gamma| \subseteq C^\Omega(\eta + 1, \psi_\Omega(\eta + 1))$, this implies

$$|A| \subseteq C^\Omega(\eta + 1, \psi_\Omega(\eta + 1))$$

and

$$\mathcal{H}_{\eta'}[\Gamma, A] = \mathcal{H}_{\eta'}[\Gamma]$$

for all $\eta' \geq \eta$.

Case 4.1: Suppose that $rk(A) < \Omega$. This implies $rk(A) \in C^{\Omega}(\eta+1, \psi_\Omega(\eta+1))$ and hence $rk(A) < \psi_\Omega(\eta+1) < \psi_\Omega(\hat{\alpha})$. Inductively we have

$$\mathcal{H}_{\hat{\alpha}_0}[\Gamma] \vdash^{\psi_\Omega(\hat{\alpha}_0)}_{\psi_\Omega(\hat{\alpha}_0)} \Gamma, A \text{ and } \mathcal{H}_{\hat{\alpha}_0}[\Gamma] \vdash^{\psi_\Omega(\hat{\alpha}_0)}_{\psi_\Omega(\hat{\alpha}_0)} \Gamma, \neg A.$$

Thus $\mathcal{H}_{\hat{\alpha}} \vdash^{\psi_\Omega(\hat{\alpha})}_{\psi_\Omega(\hat{\alpha})} \Gamma$ by means of (Cut).

Case 4.2: Suppose that $rk(A) = \Omega$. Then A or $\neg A$ will be of the form $\exists z\, F(z)$ with $F(\mathbb{V}_0)$ being $\Delta_0^{\mathcal{P}}$. We may assume that the former is the case. Then the induction hypothesis applied to $\mathcal{H}_\eta[\Gamma] \vdash^{\alpha_0}_{\Omega+1} \Gamma, A$ yields $\mathcal{H}_{\hat{\alpha}_0}[\Gamma] \vdash^{\psi_\Omega(\hat{\alpha}_0)}_{\psi_\Omega(\hat{\alpha}_0)} \Gamma, A$. Since $\psi_\Omega(\hat{\alpha}_0) \in \mathcal{H}_{\hat{\alpha}_0}(\emptyset)$, we can apply the Boundedness Lemma 6.2, obtaining

$$\mathcal{H}_{\hat{\alpha}_0}[\Gamma] \vdash^{\psi_\Omega(\hat{\alpha}_0)}_{\psi_\Omega(\hat{\alpha}_0)} \Gamma, A^{\mathbb{V}_{\psi_\Omega(\hat{\alpha}_0)}}. \tag{6.3}$$

By applying inversion (Lemma 4.9(iii)) to $\mathcal{H}_{\hat{\alpha}_0}[\Gamma] \vdash^{\alpha_0}_{\Omega+1} \Gamma, \neg A$ we also get

$$\mathcal{H}_{\hat{\alpha}_0}[\Gamma] \vdash^{\alpha_0}_{\Omega+1} \Gamma, \neg A^{\mathbb{V}_{\psi_\Omega(\hat{\alpha}_0)}}. \tag{6.4}$$

Observing that $\Gamma, \neg A^{\mathbb{V}_{\psi_\Omega(\hat{\alpha}_0)}}$ is a set of $\Sigma^{\mathcal{P}}$-formulae, we can apply the induction hypothesis to (6.4), yielding

$$\mathcal{H}_{\alpha_1}[\Gamma] \vdash^{\psi_\Omega(\alpha_1)}_{\psi_\Omega(\alpha_1)} \Gamma, \neg A^{\mathbb{V}_{\psi_\Omega(\hat{\alpha}_0)}}, \tag{6.5}$$

where $\alpha_1 = \hat{\alpha}_0 + \omega^{\Omega+\alpha_0} = \eta + \omega^{\Omega+\alpha_0} + \omega^{\Omega+\alpha_0} < \eta + \omega^{\Omega+\alpha} = \hat{\alpha}$. Moreover, we have $\psi_\Omega(\alpha_1) < \psi_\Omega(\hat{\alpha})$. Therefore (Cut) applied to (6.3) and (6.5) furnishes $\mathcal{H}_{\hat{\alpha}}[\Gamma] \vdash^{\psi_\Omega(\hat{\alpha})}_{\psi_\Omega(\hat{\alpha})} \Gamma$. □

Note that the Collapsing Theorem produces a derivation in which all instances of ($\Sigma^{\mathcal{P}}$-Ref) have been removed.

Also note that we cannot eliminate cuts with $\Delta_0^{\mathcal{P}}$-formulae since we do not have predicative cut elimination as in the case **KP**.

Corollary 6.5 *Let A be a $\Sigma^{\mathcal{P}}$-sentence of* **KP**$(\mathcal{P})$*. Suppose that* **KP**$(\mathcal{P}) \vdash A$*. Then there exists an operator $\mathcal{H}$ and an ordinal $\rho < \psi_\Omega(\varepsilon_{\Omega+1})$ such that*

$$\mathcal{H} \vdash^{\rho}_{\rho} A.$$

Proof. Let $\mathcal{H}_0$ be defined as in Definition 6.3. By Theorem 5.9 we have

$$\mathcal{H}_0 \vdash^{\omega^{\Omega+m}}_{\Omega+m+1} A$$

for some $0 < m < \omega$. Applying ordinary cut elimination, Theorem 6.1, we get

$$\mathcal{H}_0 \vdash^{\omega_m(\omega^{\Omega+m})}_{\Omega+1} A.$$

Finally, using the Collapsing Theorem 6.4 we arrive at

$$\mathcal{H}_{\omega_{m+1}(\omega^{\Omega+m})} \vdash^{\rho}_{\rho} A$$

with $\rho := \psi_\Omega(\omega_{m+1}(\omega^{\Omega+m}))$. □

7 Soundness

For the main theorem of this paper, we want to show that derivability in $RS_\Omega^\mathcal{P}$ entails truth. Since $RS_\Omega^\mathcal{P}$-formulae contain variables we need the notion of assignment. Let *VAR* be the set of free variables of $RS_\Omega^\mathcal{P}$. A variable assignment ℓ is a function

$$\ell\colon\ VAR \longrightarrow V_{\psi_\Omega(\varepsilon_{\Omega+1})}$$

satisfying $\ell(a^\alpha) \in V_{\alpha+1}$, where as per usual V_α denotes the α-th level of the von Neumann hierarchy.

The function ℓ can be canonically lifted to all $RS_\Omega^\mathcal{P}$-terms as follows:

$$\ell(\mathbb{V}_\alpha) = V_\alpha$$
$$\ell(\{x \in \mathbb{V}_\alpha \mid F(x, s_1, \ldots, s_n)\}) = \{x \in V_\alpha \mid F(x, \ell(s_1), \ldots, \ell(s_n))\}.$$

Note that $\ell(s) \in V_{\psi_\Omega(\varepsilon_{\Omega+1})}$ holds for all $RS_\Omega^\mathcal{P}$-terms s. Moreover, $\ell(s) \in V_{|s|+1}$.

Theorem 7.1 (Soundness) *Let $\mathcal{H}$ be an operator and $\alpha, \rho < \psi_\Omega(\varepsilon_{\Omega+1})$. Let $\Gamma(s_1, \ldots, s_n)$ be a sequent consisting only of $\Sigma^\mathcal{P}$-formulae. Suppose*

$$\mathcal{H} \overset{\alpha}{\underset{\rho}{\vdash}} \Gamma(s_1, \ldots, s_n).$$

Then, for all variable assignments ℓ,

$$V_{\psi_\Omega(\varepsilon_{\Omega+1})} \models \Gamma(\ell(s_1), \ldots, \ell(s_n)).$$

Proof. The proof proceeds by induction on α. Note that, owing to $\alpha, \rho < \Omega$, the proof tree pertaining to $\mathcal{H} \overset{\alpha}{\underset{\rho}{\vdash}} \Gamma(s_1, \ldots, s_n)$ neither contains any instances of ($\Sigma^\mathcal{P}$-Ref) nor of $(\forall)_\infty$, and that all cuts are with $\Delta_0^\mathcal{P}$-formulae. The proof is straightforward as all the axioms of $RS_\Omega^\mathcal{P}$ are true under the interpretation and all other rules are truth preserving with respect to this interpretation. Observe that we make essential use of the free variables when showing the soundness of $(b\forall)_\infty$, $(pb\forall)_\infty$, $(\notin)_\infty$ and $(\nsubseteq)_\infty$. □

Combining Theorem 7.1 and Corollary 6.5 we have the following:

Theorem 7.2 *If A is a $\Sigma^\mathcal{P}$-sentence and*

$$\mathbf{KP}(\mathcal{P}) \vdash A$$

then

$$V_{\psi_\Omega(\varepsilon_{\Omega+1})} \models A.$$

The bound of this corollary is actually sharp, that is, $\psi_\Omega(\varepsilon_{\Omega+1})$ is the first ordinal with that property. This follows immediately from [**21**, Theorem 4.9].

The previous result can be extended to $\Pi_2^\mathcal{P}$ sentences, basically by the same proof as for [**21**, Theorem 2.1].

Theorem 7.3 *Let A be a $\Pi_2^\mathcal{P}$-sentence. Then $\mathbf{KP}(\mathcal{P}) \vdash A$ implies $V_{\psi_\Omega(\varepsilon_{\Omega+1})} \models A$.*

Proof. Assume $\mathbf{KP}(\mathcal{P}) \vdash \forall u \exists w H(u, w)$ with $H(u, w)$ being $\Delta_0^\mathcal{P}$. Let $\sigma := \psi_\Omega(\varepsilon_{\Omega+1})$. Let $b \in V_\sigma$. We have to verify that $V_\sigma \models \exists w H(b, w)$. Since σ is a limit, there is $\xi < \sigma$ such that $b \in V_\xi$. Since V_ξ does not satisfy all $\Sigma^\mathcal{P}$-sentences provable in $\mathbf{KP}(\mathcal{P})$, we have $\mathbf{KP}(\mathcal{P}) \vdash B$ and $V_\xi \models \neg B$ for some $\Sigma^\mathcal{P}$-sentence B. Since $\Sigma^\mathcal{P}$-reflection is provable in $\mathbf{KP}(\mathcal{P})$, we also get $\mathbf{KP}(\mathcal{P}) \vdash \exists \alpha \exists x(x = V_\alpha \wedge B^x)$. Then, using $\Delta_0^\mathcal{P}$-Collection, we obtain

$$\mathbf{KP}(\mathcal{P}) \vdash \exists z \exists \alpha \exists x[x = V_\alpha \wedge B^x \wedge (\forall u \in x)(\exists w \in z) H(u, w)].$$

Since this formula is equivalent to a $\Sigma^{\mathcal{P}}$-formula in $\mathbf{KP}(\mathcal{P})$, we get

$$V_\sigma \models \exists\alpha\exists x[x = V_\alpha \wedge B^x \wedge (\forall u \in x)\exists w H(u, w)].$$

As the formula "$x = V_\alpha$" has the same meaning in V_σ as it has in V, there exists $\alpha < \sigma$ such that $V_\alpha \models B$ and $(\forall u \in V_\alpha)(\exists w \in V_\sigma)H(u, w)$. By the choice of B, this implies $\xi < \alpha$, hence $b \in V_\alpha$, thus $V_\sigma \models \exists w H(b, w)$. □

References

[1] T. Arai: *Ordinal diagrams for* PI3*-reflection*, Journal of Symbolic Logic 65 (2000), 3, 1375–1394.

[2] T. Arai: *Proof theory for theories of ordinals I: Recursively Mahlo ordinals*, Annals of Pure and Applied Logic 122 (2003), 1–85.

[3] T. Arai: *Proof theory for theories of ordinals II:* Π_3*-reflection*, Annals of Pure and Applied Logic 129 (2004), 39–92.

[4] H. Bachmann: *Die Normalfunktionen und das Problem der ausgezeichneten Folgen von Ordinalzahlen*, Vierteljahresschrift Naturforsch. Ges. Zürich 95 (1950), 115–147.

[5] J. Barwise: *Admissible Sets and Structures* (Springer, Berlin, 1975).

[6] W. Buchholz: *Eine Erweiterung der Schnitteliminationsmethode*, Habilitationsschrift (München, 1977).

[7] W. Buchholz: *A simplified version of local predicativity*, in: Aczel, Simmons, Wainer (eds.), Leeds Proof Theory 1991 (Cambridge University Press, Cambridge, 1993), 115–147.

[8] W. Buchholz, S. Feferman, W. Pohlers, W. Sieg: *Iterated inductive definitions and subsystems of analysis* (Springer, Berlin, 1981).

[9] W. Buchholz and K. Schütte: *Proof theory of impredicative subsystems of analysis* (Bibliopolis, Naples, 1988).

[10] H. Friedman: *Some applications of Kleene's method for intuitionistic systems,* in: A. R. D. Mathias and H. Rogers Jr. (eds.): Cambridge Summer School in Mathematical Logic, Lecture Notes in Mathematics, vol. 337 (Springer, Berlin, 1973), 113–170.

[11] H. Friedman: *Countable models of set theories*, in: A. R. D. Mathias and H. Rogers Jr. (eds.), Cambridge Summer School in Mathematical Logic, Lecture Notes in Mathematics, vol. 337 (Springer, Berlin, 1973), 539–573.

[12] H. Friedman, S. Ščedrov: *The lack of definable witnesses and provably recursive functions in intuitionistic set theory*, Advances in Mathematics 57 (1985), 1–13.

[13] G. Jäger: *Zur Beweistheorie der Kripke–Platek Mengenlehre über den natürlichen Zahlen*, Archiv für mathematische Logik 22 (1982), 121–139.

[14] G. Jäger and W. Pohlers: *Eine beweistheoretische Untersuchung von* $\mathbf{\Delta}^1_2$-$\mathbf{CA} + \mathbf{BI}$ *und verwandter Systeme*, Sitzungsberichte der Bayerischen Akademie der Wissenschaften, Mathematisch-Naturwissenschaftliche Klasse (1982).

[15] S. Mac Lane: *Form and Function* (Springer, Berlin, 1992).

[16] A. R. D. Mathias: *The strength of Mac Lane set theory*, Annals of Pure and Applied Logic 110 (2001), 107–234.

[17] L. Pozsgay: *Liberal intuitionism as a basis for set theory*, in: Axiomatic Set Theory, Proc. Symp. Pure Math. XIII, Part 1 (1971), 321–330.

[18] L. Pozsgay: *Semi-intuitionistic set theory*, Notre Dame Journal of Formal Logic 13 (1972), 546–550.

[19] M. Rathjen: *Ordinal notations based on a weakly Mahlo cardinal*, Archive for Mathematical Logic 29 (1990), 249–263.

[20] M. Rathjen: *Proof-theoretic analysis of KPM*, Archive for Mathematical Logic 30 (1991), 377–403.

[21] M. Rathjen: *Fragments of Kripke–Platek set theory*, in: P. Aczel, S. Wainer, H. Simmons (eds.), Proof Theory (Cambridge University Press, 1992), 251–273.

[22] M. Rathjen: *How to develop proof-theoretic ordinal functions on the basis of admissible sets*, Mathematical Quarterly 39 (1993), 47–54.

[23] M. Rathjen: *Collapsing functions based on recursively large ordinals: A well-ordering proof for KPM*, Archive for Mathematical Logic 33 (1994), 35–55.

[24] M. Rathjen: *Proof theory of reflection*, Annals of Pure and Applied Logic 68 (1994), 181–224.

[25] M. Rathjen: *The realm of ordinal analysis*, S. B. Cooper and J. K. Truss (eds.), Sets and Proofs (Cambridge University Press, 1999), 219–279.

[26] M. Rathjen: *Recent advances in ordinal analysis:* Π^1_2*-CA and related systems*, Bulletin of Symbolic Logic 1 (1995), 468–485.

[27] M. Rathjen: *An ordinal analysis of stability*, Archive for Mathematical Logic 44 (2005), 1–62.

[28] M. Rathjen: *An ordinal analysis of parameter-free* Π^1_2 *comprehension*, Archive for Mathematical Logic 44 (2005), 263–362.

[29] M. Rathjen: *Theories and ordinals in proof theory*, Synthese 148 (2006), 719–743.

[30] M. Rathjen: *From the weak to the strong existence property,* Annals of Pure and Applied Logic (2012), doi:10.1016/j.apal.2012.01.012.

[31] M. Rathjen: *Constructive Zermelo–Fraenkel Set Theory, Power Set, and the Calculus of Constructions*, 2010. To appear in 2012, in: P. Dybjer, S. Lindström, E. Palmgren and G. Sundholm: Epistemology versus Ontology: Essays on the Philosophy and Foundations of Mathematics in Honour of Per Martin-Löf (Logic, Epistemology and the Unity of Science Series, NewYork/Dordrecht: Springer Verlag).

[32] H. Schwichtenberg: *Some applications of cut elimination*, in J. Barwise (ed.), Handbook of Mathematical Logic (North-Holland, Amsterdam, 1977), 868–895.

[33] L. Tharp: *A quasi-intuitionistic set theory*, Journal of Symbolic Logic 36 (1971), 456–460.

[34] E. J. Thiele: *Über endlich axiomatisierbare Teilsysteme der Zermelo–Fraenkel'schen Mengenlehre*, Zeitschrift für mathematische Logik und Grundlagen der Mathematik 14 (1968), 39–58.

[35] R. S. Wolf: *Formally Intuitionistic Set Theories with Bounded Predicates Decidable*, PhD thesis (Stanford University, 1974).

Uniform density in Lindenbaum algebras

Vladimir Yu. Shavrukov[*], **Albert Visser**[†]

[*] Nijenburg 24, Amsterdam, The Netherlands
`v.yu.shavrukov@gmail.com`

[†] Department of Philosophy, Universiteit Utrecht, The Netherlands
`albert.visser@phil.uu.nl`

Abstract. In this paper we prove that the preordering $\lesssim$ of provable implication over any recursively enumerable theory T containing a modicum of arithmetic is uniformly dense. This means that we can find a recursive extensional density function F for $\lesssim$. A recursive function F is a density function if it computes, for A and B with $A \lnsim B$, an element C such that $A \lnsim C \lnsim B$. The function is extensional if it preserves T-provable equivalence.

Secondly, we prove a general result that implies that, for extensions of Elementary Arithmetic, the ordering $\lesssim$ restricted to Σ_n-sentences is uniformly dense.

In the last section we provide historical notes and background material.

Introduction

It is well known that the Lindenbaum algebras of theories that contain a modicum of arithmetic are dense with respect to the implication ordering. In this paper we will study a property that is stronger than *density*, to wit *uniform density*. We prove that the Lindenbaum algebras of these theories are uniformly dense with respect to the implication ordering. We first provide the necessary definitions to formulate the result.

Consider any recursively enumerable theory T that interprets the theory R introduced by Tarski, Mostowski and Robinson in [**19**]. We define:

- $A \lesssim_T B$ iff $T + A \vdash B$.
- $A \lnsim_T B$ iff $A \lesssim_T B$ and not $B \lesssim_T A$.
- $A \sim_T B$ iff $A \lesssim_T B$ and $B \lesssim_T A$.

Here $\lesssim_T$ is the "provable implication" ordering on $\mathfrak{L}_T$, the Lindenbaum sentence algebra of T. It is well known that $\lesssim_T$ is dense. We say that $\mathfrak{L}_T$ (or $\lesssim_T$) is *uniformly dense* if there is a recursive function F such that:

(i) F is *a density function*, i.e., we have $A \lnsim_T F(A,B) \lnsim_T B$, whenever $A \lnsim_T B$, and if $A \sim_T B$, then $A \sim_T F(A,B) \sim_T B$;
(ii) F is *extensional*, i.e., if $A \sim_T A'$ and $B \sim_T B'$, then $F(A,B) \sim_T F(A',B')$.

We show that $\mathfrak{L}_T$ is uniformly dense for recursively enumerable theories T that interpret R. Moreover, we can take the function F to be elementary and, in some specific cases, even p-time computable.

[†] The question concerning the uniform density of Lindenbaum algebras was raised by Sy Friedman in conversation. Sy's question was inspired by the idea that perhaps the consistency operator is, in some sense, the least way to strengthen theories. This is not so, as is corroborated by the results of this paper. I acknowledge the support of the Templeton Foundation through my participation in the CRM Infinity Project (Templeton Foundation project #13152).

We present our proof in Section 2. It consists of two stages. First we prove the desired result for Peano Arithmetic PA, or, more generally, for essentially reflexive theories.[1] Our construction delivers a p-time computable density function. This result is then generalized to all r.e. consistent theories that interpret R although the new density functions fall outside polynomial time —we do not know if this can be rectified.

A variant of the density question is obtained by imposing a restriction to a prescribed formula class. We explore this variant in Section 3. We prove a general result which implies e.g. that, for extensions T of Elementary Arithmetic, the ordering $\lnsim_T$ restricted to Σ_n-sentences is uniformly dense.

The basic idea and the ingredients of our construction for PA come with a history. The sentences we produce are certain unique Rosser sentences of a kind studied by Craig Smoryński ([**17**]) and they are the unique Gödel sentences of a certain Feferman predicate studied in [**15**]. Finally, they are Orey sentences. We will explain this background in Section 4. The reader who wants to just see the solution may, of course, skip Section 4.

1 The usual proof of density

Our proof of uniform density for $\mathfrak{L}_{\mathsf{PA}}$ is a specific instance of the usual proof of the density of $\mathfrak{L}_T$, where T is a recursively enumerable theory that interprets R. We first present this usual proof.

Suppose $A \lnsim_T B$. It follows that $T + \neg A + B$ is consistent. Let C be any arithmetical sentence that is independent of $T + \neg A + B$, i.e., $T + \neg A + B \nvdash C$ and $T + \neg A + B \nvdash \neg C$.

The essential ingredients of the proof that, for every consistent recursively enumerable theory U that interprets R, there exists a sentence R that is independent of T were provided by J. Barkley Rosser in his classical paper [**14**].

We consider $D := A \vee (C \wedge B)$. We claim: $A \lnsim_T D \lnsim_T B$. First, we clearly have $A \lesssim_T D \lesssim_T B$. Suppose $T + B \vdash A \vee (C \wedge B)$. Then, by propositional logic, it follows that $T + \neg A + B \vdash C$. Quod non. Suppose $T + A \vee (C \wedge B) \vdash A$. Then, by propositional logic, we find that $T + \neg A + B \vdash \neg C$. Quod non.

We can squeeze a little bit more information out of our construction. If C is independent, then $\neg C$ is also independent. So we can construct $E := A \vee (\neg C \wedge B)$. We find that $A \lnsim_T E \lnsim_T B$. Moreover, we have $T \vdash (D \wedge E) \leftrightarrow A$ and $T \vdash (D \vee E) \leftrightarrow B$. So we have two sentences D and E strictly between A and B such that B is the supremum with respect to $\lesssim_T$ of D and E and A is the infimum with respect to $\lesssim_T$ of D and E.

In the light of the proof given above, to prove p-time uniform density for PA, it is, modulo some simple details, sufficient to give a p-time construction $A \mapsto C_A$, where C_A is independent over $\mathsf{PA} + A$, provided that $\mathsf{PA} + A$ is consistent and, if $\mathsf{PA} \vdash A_0 \leftrightarrow A_1$, then $\mathsf{PA} \vdash C_{A_0} \leftrightarrow C_{A_1}$.

In the next section, we will provide a mapping $A \mapsto C_A$ that satisfies the desiderata.

[1] A theory is *essentially reflexive* (*uniformly essentially reflexive*) if it proves reflection (resp. uniform reflection) for each of its finitely axiomatized subtheories. Here (uniform) reflection concerns a proof predicate that is formalized with respect to an interpretation of a weak arithmetic, like S^1_2, in the given theory. Uniform essential reflexivity implies full induction with respect to the designated interpretation of the numbers. Conversely, a theory that satisfies full induction and is sequential is uniformly essentially reflexive. (For the definition of *sequential*, see [**4**] or [**23**].) If we drop uniformity, essentially reflexive theories can be much weaker. For example, the minimal essentially reflexive extension of elementary arithmetic EA (also known as $\mathsf{I}\Delta_0 + \mathsf{Exp}$) is both a subtheory of PA and of EA plus all true Π_1-sentences.

2 Uniform density

The first order of business for this section is:

Theorem 2.1 $\mathfrak{L}_{\mathsf{PA}}$ *is uniformly dense via a p-time computable density function. More generally, this result holds for all essentially reflexive sequential r.e. theories.*

This theorem will later be extended to more theories.

Since we are interested in getting our density function p-time, we will use efficiently coded syntax and base 2 numerals. See e.g. [**1**, 7.3] or [**4**, V.3].

We consider the sequence of theories Ar_n, where Ar_0 is EA, also known as $\mathsf{I}\Delta_0 + \mathsf{Exp}$, and $\mathsf{Ar}_{n+1} := \mathsf{I}\Sigma_{n+1}$. These theories have the following important property.

Theorem 2.2 ($\mathsf{I}\Sigma_1$ *proves that) for all* n*,* Ar_{n+1} *proves uniform* Π_{n+2}*-reflection for* Ar_n*.*

For a proof, see e.g. [**16**] or [**10**]. We note that we have as an immediate consequence:

Corollary 2.3 ($\mathsf{I}\Sigma_1$ *proves that) for any sentence* A *in* Σ_{n+2}*,* $\mathsf{Ar}_{n+1} + A$ *proves uniform* Π_{n+2}*-reflection for* $\mathsf{Ar}_n + A$*.*

We give the following definitions:

- $\Box_{A,x} B$ stands for $\mathsf{prov}_{\mathsf{Ar}_x + A}(\underline{\ulcorner B \urcorner})$.
- $\Diamond_{A,x} B$ stands for $\neg\, \mathsf{prov}_{\mathsf{Ar}_x + A}(\underline{\ulcorner \neg B \urcorner})$, i.e., $\neg\, \Box_{A,x} \neg B$.
- $C_A :\leftrightarrow A \wedge \forall x\, (\Box_{A,x}\Box_{A,x}\bot \to \Box_{A,x}\bot)$.

We have the following useful lemma:

Lemma 2.4 *Suppose* $A \in \Sigma_n$ *and* $n \geq 1$*. Then,*

$$\mathsf{Ar}_n \vdash C_A \leftrightarrow (A \wedge \forall x \geq n\, (\Box_{A,x}\Box_{A,x}\bot \to \Box_{A,x}\bot)).$$

Proof. By Corollary 2.3, $\mathsf{Ar}_n + A$ proves Σ_1-reflection for $\Box_{A,k}$ with $k < n$. Ergo,

$$\mathsf{Ar}_n + A \vdash \forall x < n\, (\Box_{A,x}\Box_{A,x}\bot \to \Box_{A,x}\bot).$$

The desired result is immediate. □

We have:

Lemma 2.5 *Suppose* $\mathsf{PA} \vdash A_0 \leftrightarrow A_1$*. Then* $\mathsf{PA} \vdash C_{A_0} \leftrightarrow C_{A_1}$*.*

Proof. Suppose $\mathsf{PA} \vdash A_0 \leftrightarrow A_1$. Then, for some n, $\mathsf{Ar}_n \vdash A_0 \leftrightarrow A_1$. We can pick n so large that $n \geq 1$ and $A_0, A_1 \in \Sigma_n$. We have:

$$\begin{aligned} \mathsf{Ar}_n \vdash\ C_{A_0} &\leftrightarrow A_0 \wedge \forall x \geq n\, (\Box_{A_0,x}\Box_{A_0,x}\bot \to \Box_{A_0,x}\bot) \\ &\leftrightarrow A_1 \wedge \forall x \geq n\, (\Box_{A_1,x}\Box_{A_1,x}\bot \to \Box_{A_1,x}\bot) \\ &\leftrightarrow C_{A_1}. \end{aligned}$$

Hence, $\mathsf{PA} \vdash C_{A_0} \leftrightarrow C_{A_1}$. □

We show that, if $\mathsf{PA} + A$ is consistent, then C_A is independent over $\mathsf{PA} + A$.

Lemma 2.6 *Suppose* $\mathsf{PA} + A$ *is consistent. Then* C_A *is independent over* $\mathsf{PA} + A$*.*

Proof. Suppose that $\mathsf{PA} + A \vdash C_A$. Then, for some n, $\mathsf{Ar}_n + A \vdash C_A$. We may assume that $n \geq 1$ and $A \in \Sigma_n$. It follows that $\mathsf{Ar}_n + A \vdash \Box_{A,n}\Box_{A,n}\bot \to \Box_{A,n}\bot$. Hence, by Löb's Theorem, $\mathsf{Ar}_n + A \vdash \Box_{A,n}\bot$. We may conclude that $\mathsf{PA} + A \vdash \bot$. Quod non.

Suppose that $\mathsf{PA} + A \vdash \neg\, C_A$. Then, for some n, $\mathsf{Ar}_n + A \vdash \neg\, C_A$. We may assume that $n \geq 1$ and $A \in \Sigma_n$. We find, using Lemma 2.4, that

$$\mathsf{Ar}_n + A \vdash \exists x \geq n\, (\Box_{A,x}\Box_{A,x}\bot \wedge \Diamond_{A,x}\top).$$

But then $\mathsf{Ar}_n + A \vdash \Diamond_{A,n}\top$, contradicting the Second Incompleteness Theorem. □

We put $F(A,B) := A \vee (C_{\neg A \wedge B} \wedge B)$. We note that this is defined for any arithmetical sentences A and B. Moreover, we always have: $A \lesssim_{\mathsf{PA}} F(A,B) \lesssim_{\mathsf{PA}} A \vee B$. By Lemma 2.5, F is extensional. By Lemma 2.6, we have that, if $A \lnsim_{\mathsf{PA}} B$, then $A \lnsim_{\mathsf{PA}} F(A,B) \lnsim_{\mathsf{PA}} B$.

This concludes the proof of Theorem 2.1. We note that F is p-time in A and B. □

Remark 2.7 We can put:

- $F_0(A,B) := A \vee (C_{\neg A \wedge B} \wedge B)$,
- $F_1(A,B) := A \vee (\neg C_{\neg A \wedge B} \wedge B)$.

We construct an infinite $\lnsim_{\mathsf{PA}}$-antichain between A and B by considering, e.g., $D_{A,B,0} := F_0(A,B)$, $D_{A,B,1} := F_0(A, F_1(A,B))$, $D_{A,B,2} := F_0(A, F_1(A, F_1(A,B)))$, ... The mapping $H\colon A, B, n \mapsto D_{A,B,n}$ need not be p-time, but since this mapping is elementary we can represent it in PA. Since the $D_{A,B,n}$ have complexity (in the sense of the arithmetical hierarchy) bounded by the maximum of 2, the compexity of A and that of B —say the complexity is $k(A,B)$— we can, using efficient numerals, replace $D_{A,B,n}$ by $E_{A,B,n} := \mathsf{True}_{k(A,B)}(H(A,B,n))$. If we use a reasonable version of the definition of True_k, the mapping $A, B, n \mapsto E_{A,B,n}$ becomes p-time. (Note that we do not need to worry about the length of the verifications of the usual properties of the True_k. We are only interested in the size of the formulas.)

Our proof can be immediately adapted to any essentially reflexive theory —like ZF: all the ingredients of the construction of C are also present in such a theory.

Let $C^\circ_A := \forall x\, (\Box_{A,x}\Box_{A,x}\bot \to \Box_{A,x}\bot)$. Note that C°_A is Δ_2 over $\mathsf{I}\Sigma_1$ because it is $\mathsf{I}\Sigma_1$-provably equivalent to $\Diamond_A\top \vee \exists x\, (\Box_{A,x+1}\bot \wedge \Diamond_{A,x}\Diamond_{A,x}\top)$.

Let us consider the relationship between C°_A and $\mathsf{con}(\mathsf{PA} + A)$.

Proposition 2.8 $\mathsf{PA} + \mathsf{con}(\mathsf{PA} + A) \vdash \forall x\, (\Box_{A,x}\Box_{A,x}\bot \to \Box_{A,x}\bot)$.

Proof. Suppose A is Σ_{n+2}. We reason in $\mathsf{PA} + \mathsf{con}(\mathsf{PA} + A)$: Suppose that $\Box_{A,x}\Box_{A,x}\bot$. Then $\Box_{A,\max\{x+1,n\}}\bot$ and, hence, $\Box_A\bot$. Quod non. We may conclude $\neg\,\Box_{A,x}\Box_{A,x}\bot$, and, a fortiori, C°_A. □

Since, as we will show in Section 4, C°_A is an Orey sentence of $\mathsf{PA} + A$ and, provided that $\mathsf{PA} + A$ is consistent, $\mathsf{con}(\mathsf{PA} + A)$ is not an Orey sentence, C_A is strictly between $A + \mathsf{con}(\mathsf{PA} + A)$ and A over PA. In other words, C°_A is a reflection principle that is strictly between $\mathsf{con}(\mathsf{PA} + A)$ and $\top$ over $\mathsf{PA} + A$.

Theorem 2.1 generalizes to theories containing R thanks to

Theorem 2.9 (Pour-El & Kripke [**12**, Theorem 2]) *The Lindenbaum sentence algebras of all recursively enumerable, consistent theories that interpret* R *are effectively isomorphic.*

For us, 'effective isomorphism' means a recursive function from sentences of one theory to those of the other theory that, through provable equivalence, quotients down to an isomorphism between the two Lindenbaum algebras. The functions constructed in [**12**] however possess further nice properties.

Pulling the density function of Theorem 2.1 off $\mathfrak{L}_{\mathsf{PA}}$ back to $\mathfrak{L}_T$ along an effective isomorphism $\mathfrak{L}_T \to \mathfrak{L}_{\mathsf{PA}}$, we obtain

Corollary 2.10 *The Lindenbaum sentence algebras of all recursively enumerable consistent theories that interpret* R *enjoy uniform density.*

It is seen from the proof of Theorem 2 in [**12**] that the isomorphisms of Theorem 2.9 together with their inverses can be given by elementary functions (aka ones from Grzegorczyk class $\mathcal{E}^3$).

Accordingly, we are only able to claim elementarity rather than polynomial time for the second-hand density functions obtained via (the intended proof of) Corollary 2.10. Our proof also forfeits the ability to have uniform density achieved by just mixing in an appropriate Δ_2 sentence.

3 Orderings of Σ_n sentences and precomplete lattices

We address the question of uniform density for restricted classes of formulas in a somewhat more general setting.

An *r.e. lattice* L is a pair of recursive functions $\vee$ and $\wedge$ defined on an r.e. subset $\mathsf{field}\, L$ of ω together with an r.e. equivalence relation $\sim$ on $\mathsf{field}\, L$ which is a congruence for $\vee$ and $\wedge$ and such that the quotient is a lattice. If that lattice is Boolean then L is called an *r.e. Boolean algebra* and, as is easily seen, has a recursive negation function. Lindenbaum sentence algebras of r.e. theories provide typical examples.

A *density function* for L is a function $D\colon (\mathsf{field}\, L)^2 \to \mathsf{field}\, L$ such that, if $a \precnsim b$ then $a \precnsim D(a,b) \precnsim b$, and $D(a,b) \sim a \sim b$ whenever $a \sim b$. D is *extensional* (with respect to $\sim$ or L) if $\sim$ is a congruence for D. L is *uniformly dense* if it admits an effective extensional density function —note that for $\mathfrak{L}_T$ this agrees with our earlier definition.

Montagna & Sorbi [**9**, Proposition 3.1(b)] extend Theorem 2.9 to all *effectively inseparable* r.e. Boolean algebras, i.e., algebras where the $\sim$-equivalence classes of (Boolean) 0 and 1 are effectively inseparable within $\mathsf{field}\, L$. Hence Corollary 2.10 also holds for all e.i. r.e. Boolean algebras.

When the proof of Theorem 2.1 works for a theory T, it works equally well for the sublattice Σ_n/T of $\mathfrak{L}_T$ determined by Σ_n sentences provided $n > 1$ because $D(a,b)$ is a lattice polynomial in a, b, and a Δ_2 sentence. In this section we handle Σ_1/T using a different approach which starts with the definition of a precomplete numeration/equivalence.

A non-trivial equivalence relation $\sim$ on an r.e. subset $\mathsf{field}\sim$ of ω is *precomplete* if to every partial recursive $f\colon \omega \to \mathsf{field}\sim$ there is a total recursive $F : \omega \to \mathsf{field}\sim$ that *makes f total modulo $\sim$*, i.e., $F(n) \sim f(n)$ whenever $f(n)$ converges. Reducing f to a universal ($\mathsf{field}\sim$)-valued partial recursive function, we see that an index for F can be found effectively in one for f.

An r.e. lattice L is *precomplete* if its associated (r.e.) equivalence relation $\sim$ is. By $\precsim$ we denote the corresponding (r.e.) preorder on $\mathsf{field}\, L$.

Example 3.1 (Visser [**21**, 1.6.6]) Σ_n/T is r.e. and precomplete whenever T is a consistent r.e. extension of EA.

Hint 3.2 The mapping that assigns to k the Σ_n sentence $\exists y\,(y = f(\overline{k}) \wedge \mathsf{True}_n(y))$ makes f total modulo T-provable equivalence.

It is an open question whether Σ_1/S^1_2 or $\exists\Sigma^{\mathsf{b}}_1/\mathsf{S}^1_2$ is precomplete. (See [**1**] for definitions of $\exists\Sigma^{\mathsf{b}}_1$ and S^1_2.)

Mutual interpretability for finitely axiomatized sequential theories is also r.e. precomplete, since the interpretability ordering on finitely axiomatized sequential theories (modulo mutual interpretability) is p-time anti-isomorphic to Π_1/EA. This uses the Friedman characterization of interpretability between finitely axiomatized theories (see e.g. [**24**], Theorem 3.6). Thus, the lattice of finitely axiomatized sequential interpretability degrees is (p-time) isomorphic to Σ_1/EA.

Sentences of the form $\exists x \mathsf{T}_0(\underline{n}, x)$, where T_0 is Kleene's T-predicate for the 0-ary case, form an example of a class Γ, such that Γ/S^1_2 is r.e., precomplete, and is, modulo S^1_2-provability, a sublattice of $\mathfrak{L}_{\mathsf{S}^1_2}$.

The r.e. extensions of PA in the language of PA modulo interpretability give us under an appropriate indexing an example of a precomplete numeration that is *not* recursively enumerable.

Theorem 3.3 *Any r.e. precomplete lattice is uniformly dense.*

Note that the theorem needs neither distributivity nor boundedness.

Here is the plan: given a recursive F, we are going to craft a partial recursive f. In other words, the Construction below will effectively associate to an index e for F an index $c(e)$ for f. An index $t(c(e))$ for some F' making f total modulo $\sim$ is then effective in $c(e)$. By the 2nd Recursion Theorem there is an e_0 indexing the same function as $t(c(e_0))$. For that e_0 we have $F' \simeq F$. We may therefore assume from the outset that F makes f total modulo $\sim$.

Lastly, we put $D(a, b) = a \vee (F(a, b) \wedge b)$ which will be the desired extensional density function for L.

We fix effective enumerations $(\sim_n)_{n\in\omega}$ and $(\precsim_n)_{n\in\omega}$ of $\sim$ and $\precsim$ resp. that satisfy the following:

- for each $n \in \omega$, $\mathsf{field} \sim_n = \mathsf{field} \precsim_n$ is a finite non-empty subset of $\mathsf{field}\, L$;
- $\sim_n$ is an equivalence relation;
- $\sim_n \subseteq \sim_{n+1}$ and $\precsim_n \subseteq \precsim_{n+1}$;
- $\sim = \bigcup_{n\in\omega} \sim_n$ and $\precsim = \bigcup_{n\in\omega} \precsim_n$.

Construction 3.4 The construction of f proceeds in stages. The following happens at stage n:

(C1) Suppose $a, b \in \mathsf{field} \sim_n$ and $f(a, b)$ has not yet been defined.
Let $a_0, b_0 \in \mathsf{field} \sim_n$ be the minimal such that $a_0 \sim_n a$ and $b_0 \sim_n b$, and put $f(a, b) = F(a_0, b_0)$ unless $(a, b) = (a_0, b_0)$.

(C2) Suppose $a \precsim_n b$, $f(a, b)$ has not yet been defined and $a \vee (F(a, b) \wedge b) \precsim_n a$.
Put $f(a, b) = b$.

(C3) Suppose $a \precsim_n b$, $f(a, b)$ has not yet been defined and $b \precsim_n a \vee (F(a, b) \wedge b)$.
Put $f(a, b) = a$.

Claim 1 If $a \precsim b$ then $a \precsim D(a, b) \precsim b$. In particular, $a \sim b$ implies $a \sim D(a, b) \sim b$.

Proof. This holds by virtue of the definition $D(a, b) = a \vee (F(a, b) \wedge b)$ regardless of the value of $F(a, b)$. □

Claim 2 $f(a,b)$ is defined unless a and b are minima of their respective $\sim$-equivalence classes.

Proof. Clause (C1) takes care of this. □

Claim 3 If $f(a,b)$ is defined via clause (C2) or (C3) then $a \sim b$.

Proof. Suppose $f(a,b)$ is defined via clause (C2). We may conclude that $a \precsim b$, and $a \vee (F(a,b) \wedge b) \precsim a$, and $F(a,b) \sim f(a,b) = b$, so $a \vee b \precsim a$, hence $a \sim b$.

Clause (C3) is treated similarly. □

Claim 4 D is extensional with respect to $\sim$.

Proof. That the $\sim$-equivalence class of $D(a,b)$ only depends on those of a and b follows from Claim 1 for the case $a \sim b$. We may therefore assume $a \nsim b$. This implies, by Claim 3, that the only way to define $f(a',b')$ for $a' \sim a$ and $b' \sim b$ is via clause (C1).

Assume a_0 and b_0 are the minima of the $\sim$-equivalence classes of a and b respectively. We show by induction on $a+b$ that $F(a,b) \sim F(a_0,b_0)$ for all $a \sim a_0$ and $b \sim b_0$. Suppose $(a,b) \neq (a_0,b_0)$. By Claim 2, $f(a,b)$ is defined —via clause (C1). So $f(a,b) = F(a',b')$ where $a' \sim a$, $b' \sim b$, and $a' + b' < a + b$. Accordingly,

$$F(a,b) \sim f(a,b) = F(a',b') \sim F(a_0,b_0)$$

with the last equivalence holding by i.h. Hence $D(a,b) \sim D(a_0,b_0)$. □

Claim 5 If $a \precnsim b$ then $a \precnsim D(a,b) \precnsim b$.

Proof. In view of Claim 1, it suffices to exclude the situations $a \sim D(a,b)$ and $D(a,b) \sim b$.

Suppose $a \sim D(a,b) = a \vee (F(a,b) \wedge b)$. Let a_0, b_0 be the minima of the $\sim$-equivalence classes of a and b. Then a_0, b_0 also are minima of any $\sim_n$-equivalence classes they belong to. Thus clause (C1) cannot define $f(a_0,b_0)$. By Claim 3 neither can (C2) nor (C3). Yet clause (C2) will sooner or later define $f(a_0,b_0)$ if nothing else does. Contradiction.

$D(a,b) \sim b$ is outruled in a similar fashion. □

Claims 1, 4, and 5 amount to a proof of Theorem 3.3.

Corollary 3.5 *For r.e. consistent T extending EA the lattice Σ_n/T is uniformly dense.*

Remark 3.6 Using $\mathsf{True}_n(\cdots)$ as in Remark 2.7, one can bring down to p-time the complexity of any recursive function with values in Σ_n/T. The density functions for Σ_n/T obtained through Theorem 3.3 however are already polynomial time because in Σ_n/T totalization works by substitution (see the hint to Example 3.1) as does, for that matter, the 2nd Recursion Theorem.

Corollary 3.7 *The finitely axiomatized sequential theories are uniformly dense with respect to the interpretability preordering $\lhd$. The density function can be taken to be p-time.*

Open Question 3.8 Are Σ_1/S^1_2 and/or $\exists\Sigma^{\mathsf{b}}_1/\mathsf{S}^1_2$ uniformly dense?

Harvey Friedman shows in his Tarski lectures that the interpretability preordering on arbitrary finitely axiomatized theories of predicate logic is dense. Is this ordering uniformly dense? ▽

4 Archaeology

In this section we provide various background materials that make our construction of C_A meaningful. We will sketch how the main ingredient of our formula C_A, to wit the formula $C_A^\circ := \forall x\,(\Box_{A,x}\Box_{A,x}\bot \to \Box_{A,x}\bot)$, can be viewed as either a unique Rosser or Gödel fixed point. We first discuss the Rosser construction.

A standard way to produce independent sentences is the Rosser construction, invented by J. Barkley Rosser. The original paper is [**14**]. Rosser's construction has some extra good properties. The construction is verifiable in PA, and, after some careful inspection, even in EA.[2] A second point is that Rosser's argument works for a very wide class of theories including the recursively enumerable extensions of the Tarski–Mostowski–Robinson theory R. Finally, the sentence produced by his construction, the Rosser sentence, is Σ_1 or Π_1, more specifically: $\exists\Pi_1^{\mathsf{b}}$ or $\forall\Sigma_1^{\mathsf{b}}$.

Can we use the original Rosser construction to obtain independent sentences in a uniform way? This does not look very promising: the sentences delivered by that construction are quite sensitive to implementation details. E.g., suppose we use a standard fixed point construction to obtain a Rosser sentence R_A for $\mathsf{PA}+A$ and a Rosser sentence $R_{A'}$ for $\mathsf{PA}+A'$. Suppose further that A and A' are PA-provably equivalent. Then, R_A and $R_{A'}$ need not be PA-provably equivalent. The *intensionality* of the Rosser construction has, for example, been studied in [**3**]. However, several variants of the Rosser construction have been considered in the literature and among these we find one that is sufficiently uniform. This Rosser construction was introduced by Craig Smoryński. As we will see this Rosser construction can also be viewed as a Gödel construction.

Consider an r.e. extension T of PA in the same language. Let $\tau := (T_n)_{n\in\omega}$ be a recursive sequence of theories so that $\mathsf{I}\Sigma_1$ proves that:

i. for all n and k, if $n<k$, then T_n is a subtheory of T_k;
ii. the union of the T_n is T;
iii. for each n, $T_{n+1} \vdash \mathsf{con}(T_n)$.

We need the following definitions:

- $\Box^\star_\tau B$ stands for $\exists x\,\Box_{T_x} B$. Note that $\Box^\star_\tau B$ is provably equivalent to $\Box_T B$.
- If C and D are of the respective forms $\exists x\, C_0(x)$ and $\exists y\, D_0(y)$, then
$$C < D := \exists x\,(C_0(x) \wedge \forall y \leq x\,\neg\, D_0(y))$$
and
$$C \leq D := \exists x\,(C_0(x) \wedge \forall y < x\,\neg\, D_0(y)).$$

We note that $\Box^\star_\tau B < \Box^\star_\tau C$ is $\mathsf{I}\Sigma_1$-provably equivalent to $\Box_T B \wedge \neg\,(\Box^\star_\tau C \leq \Box^\star_\tau B)$. Thus, the formula $\Box^\star_\tau B < \Box^\star C$ is Δ_2 over $\mathsf{I}\Sigma_1$. Similarly, for $\Box^\star_\tau B \leq \Box^\star C$.

The formula $\Box^\star_\tau B < \Box^\star_\tau C$ is equivalent over PA to $\exists x\,(\Box_{T_x} B \wedge \neg\,\Box_{T_x} C)$. It follows that the formula $\Box^\star_\tau B < \Box^\star_\tau \neg\, B$ is equivalent to $\Box^\star_\tau B < \Box^\star_\tau \bot$ which coincides with the *Feferman predicate* for τ defined as

- $\triangle_\tau A := \Box^\star_\tau A < \Box^\star_\tau \bot$.

We note that, over EA, $\triangle_\tau A$ is equivalent to $\exists x\,(\Box_{T_x} A \wedge \Diamond_{T_x} \top)$.

The Feferman predicate was introduced by Solomon Feferman in his classical paper [**2**]. It is a sort of self correcting provability predicate, which is related to trial-and-error predicates as studied in [**5**] and [**13**]. Feferman's aim in introducing it was not just the

[2] A modified argument even works in $\mathsf{I}\Delta_0 + \Omega_1$. The basic idea of this argument is due to Švejdar (see [**18**]). For the verification that Švejdar's assumptions are fulfilled, see [**20**].

study of ways to escape the second incompleteness theorem, but also applications to the study of relative interpretability.

Here is the central insight. We write

- $\triangledown_\tau A$ for $\neg\triangle_\tau\neg A$;
- $U \rhd V$ for: there is a relative interpretation of V in U.

See e.g. [**2, 6, 19, 24**] for basic definitions concerning interpretations. We have:

Theorem 4.1 $(T + \triangledown_\tau A) \rhd (T + A)$.

See [**2**] for the main ingredients of the proof. The basic idea of the result is that $\triangledown_\tau A$ is a consistency statement of $T + A$. We can use the Henkin construction to build the desired interpretation.

We now consider the specialized sequences $\nu_A := (\mathsf{Ar}_n + A)_{n\in\omega}$ for the theories $\mathsf{PA}+A$. We simply write $\Box^\star_A$ for $\Box^\star_{\nu_A}$, etcetera.

By the Gödel Fixed Point Lemma, we can find a sentence R_A such that

$$\mathsf{PA} \vdash R_A \leftrightarrow \neg\,(\Box^\star_A R_A < \Box^\star_A \neg\, R_A).$$

Thus R_A is a Rosser sentence for the $\Box^\star_A$. By our previous remarks, the sentence R_A is also a Gödel sentence for the Feferman predicate $\triangle_A$, that is:

$$\mathsf{PA} \vdash R_A \leftrightarrow \neg\,\triangle_A R_A.$$

Smoryński gave the construction of R_A in his paper [**17**]. This paper was inspired by a study of a variant of the Rosser construction in the context of set theory by Kenneth McAloon [**7**].

Theorem 2.1 in [**17**] implies that R_A is, up to provable equivalence, unique over $\mathsf{PA} + A$. By a minor addition to Smoryński's argument, one can show that the mapping $A \mapsto (A \wedge R_A)$ preserves PA-provable equivalence. It is shown in [**15**] that uniqueness can fail under a choice of stratification sequence different from $(\mathsf{Ar}_n)_{n\in\omega}$.

Smoryński also shows that R_A is independent over $\mathsf{PA} + A$, provided that $\mathsf{PA} + A$ is consistent. As we will see, Smoryński's Rosser sentence R_A is $(\mathsf{PA} + A)$-provably equivalent to the sentence $\forall x\,(\Box_{A,x}\Box_{A,x}\bot \to \Box_{A,x}\bot)$. So the independence of R_A also follows from our Lemma 2.6.

Since the Feferman–Smoryński predicate explicates a notion of provability, it can be studied modally. This study was taken up in [**8, 15, 22**]. The latter paper studies the Feferman predicate over PA based on the sequence Ar_n with conclusions translatable to the hierarchy $\mathsf{Ar}_n + A$. Thus, [**15**] contains an alternative, modal, proof of the uniqueness of R_A.

Recall $C^\circ_A := \forall x\,(\Box_{A,x}\Box_{A,x}\bot \to \Box_{A,x}\bot)$. We show that C°_A is a Gödel sentence for $\triangle_A$.

Theorem 4.2 ([**15**, Exercise 2.7]) *C°_A is a Gödel sentence of $\triangle_A$ over* $\mathsf{PA} + A$.

Proof. We have, using Löb's theorem in the third step:

$$\begin{aligned}\mathsf{PA} + A \;\vdash\; \triangle_A C^\circ_A &\to \exists x\,(\Box_{A,x}\forall y\,(\Box_{A,y}\Box_{A,y}\bot \to \Box_{A,y}\bot) \wedge \Diamond_{A,x}\top)\\ &\to \exists x\,(\Box_{A,x}(\Box_{A,x}\Box_{A,x}\bot \to \Box_{A,x}\bot) \wedge \Diamond_{A,x}\top)\\ &\to \exists x\,(\Box_{A,x}\Box_{A,x}\bot \wedge \Diamond_{A,x}\top)\\ &\to \neg C^\circ_A.\end{aligned}$$

We treat the other direction. Suppose $A \in \Sigma_n$, where $n \geq 1$. We work in $\mathsf{PA}+A$. Suppose $\neg C^\circ_A$, i.e., $\exists x\,(\Box_{A,x}\Box_{A,x}\bot \wedge \Diamond_{A,x}\top)$. Clearly, it follows that

$$\exists x\,(\Box_{A,x}\forall y \geq x\,\Box_{A,y}\bot \wedge \Diamond_{A,x}\top),$$

and hence

$$\exists x\,(\Box_{A,x}\forall y \geq x\,(\Box_{A,y}\Box_{A,y}\bot \to \Box_{A,y}\bot) \wedge \Diamond_{A,x}\top). \tag{4.1}$$

We note that we may assume that $x \geq n$, since for any standardly finite k, $\Box_{A,k}\Box_{A,k}\bot$ implies $\bot$. Hence, by the fact that $\mathsf{Ar}_x + A$ proves Σ_1-reflection for $\mathsf{Ar}_y + A$, for $y < x$, we find

$$\exists x\,(\Box_{A,x}\forall y < x\,(\Box_{A,y}\Box_{A,y}\bot \to \Box_{A,y}\bot) \wedge \Diamond_{A,x}\top). \tag{4.2}$$

Combining (4.1) and (4.2), we find

$$\exists x\,(\Box_{A,x}\forall y\,(\Box_{A,y}\Box_{A,y}\bot \to \Box_{A,y}\bot) \wedge \Diamond_{A,x}\top). \tag{4.3}$$

Of course (4.3) is $\triangle_A C^\circ_A$.

Thus, we have shown that $\mathsf{PA} + A \vdash C^\circ_A \leftrightarrow \neg \triangle_A C^\circ_A$. □

So, C°_A is modulo PA-provable equivalence Smoryński's Rosser sentence for $\mathsf{PA} + A$.

Open Question 4.3 Our proof of the extensionality of $A \mapsto C_A$ as well as that of unprovability of C_A go through, with minor modifications, for any stratification sequence τ for PA satisfying our conditions. The consistency of C°_A with $\mathsf{PA} + A$ is the only element of Theorem 2.1 that ostensibly depends on $\tau = (\mathsf{Ar}_n)_{n\in\omega}$ (or, more generally, on the "fast-growing" property of τ that each level proves enough reflection for the previous ones).

This makes us wonder if there exists a consistent theory of the form $\mathsf{PA} + A$ together with a stratification sequence $\tau = (T_n)_{n\in\omega}$ such that $\mathsf{PA} + A$ refutes:

$$\forall x\,(\Box_{A,T_x}\Box_{A,T_x}\bot \to \Box_{A,T_x}\bot).$$

A similar question can be asked of Theorem 4.2. ▽

We end this section by showing that C°_A is an Orey sentence of $\mathsf{PA} + A$.

Consider any theory T. A sentence O in the language of T is an *Orey sentence* of T if $T \rhd (T + O)$ and $T \rhd (T + \neg O)$. Note that the negation of an Orey sentence is an Orey sentence. An Orey sentence O of T is clearly independent of T. Neither an Orey sentence nor its negation add interpretability strength to the given theory.

The idea of Orey sentences was introduced by Orey ([**11**]), who also provided the first known Orey sentence for PA. There are many salient natural Orey sentences. Two well-known examples are the Parallel Axiom over a suitable version of neutral geometry and the Continuum Hypothesis over ZFC. For essentially reflexive sequential theories, the Gödel sentence of a Feferman predicate for the theory is an Orey sentence —see below.

We will show that C°_A is an Orey sentence of $\mathsf{PA} + A$. This sentence is still metamathematical and does involve coding, but it is, at least, self-reference-free. Since C°_A is a Gödel sentence of $\triangle_A$, the desired insight is immediate by the following theorem. (This theorem was also proved in [**22**].)

Theorem 4.4 *Consider a consistent theory T given by a sequence τ satisfying the conditions given above. Then any Gödel sentence of $\triangle_\tau$ is an Orey sentence for T.*

Proof. Let G be a Gödel sentence of $\triangle_\tau$. We have:

$$\begin{aligned} T + G &\vdash T + \triangledown_\tau \neg\, G \\ &\rhd T + \neg G; \end{aligned}$$

$$\begin{aligned} T + \neg\, G &\vdash T + \triangle_\tau G \\ &\vdash T + \triangledown_\tau G \\ &\rhd T + G. \end{aligned}$$

In the second step of the second proof, we use $T \vdash \triangle_\tau \neg A \to \neg \triangle_\tau A$. Thus, we have both $(T + G) \rhd (T + G)$, by the identity interpretation, and $(T + \neg\, G) \rhd (T + G)$. So, using a disjunctive interpretation, we find $T \rhd (T + G)$. Similarly, $T \rhd (T + \neg\, G)$. □

References

[1] S. R. Buss. *Bounded Arithmetic.* Bibliopolis, Napoli, 1986.

[2] S. Feferman. Arithmetization of metamathematics in a general setting. *Fundamenta Mathematicae*, 49:35–92, 1960.

[3] D. Guaspari and R. M. Solovay. Rosser sentences. *Annals of Mathematical Logic*, 16:81–99, 1979.

[4] P. Hájek and P. Pudlák. *Metamathematics of First-Order Arithmetic.* Perspectives in Mathematical Logic. Springer, Berlin, 1991.

[5] R. G. Jeroslow. Experimental logics and Δ^0_2-theories. *Journal of Philosophical Logic*, 4:253–267, 1975.

[6] P. Lindström. *Aspects of Incompleteness, 2nd ed.*, vol. 10 of *Lecture Notes in Logic.* ASL / A. K. Peters, Natick, Massachusetts, 2003.

[7] K. McAloon. Formules de Rosser pour ZF. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences*, Série A, 281(16):669–672, 1975.

[8] F. Montagna. On the algebraization of a Feferman's predicate. *Studia Logica*, 37:221–236, 1978.

[9] F. Montagna and A. Sorbi. Universal recursion theoretic properties of r.e. preordered structures. *The Journal of Symbolic Logic*, 50:395–406, 1985.

[10] H. Ono. Reflection principles in fragments of Peano arithmetic. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 33:317–333, 1987.

[11] S. Orey. Relative interpretations. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 7:146–153, 1961.

[12] M. B. Pour-El and S. Kripke. Deduction-preserving "Recursive Isomorphisms" between theories. *Fundamenta Mathematicae*, 61:141–163, 1967.

[13] H. Putnam. Trial and error predicates and the solution to a problem of Mostowski. *The Journal of Symbolic Logic*, 30:49–57, 1965.

[14] J. B. Rosser. Extensions of some theorems of Gödel and Church. *The Journal of Symbolic Logic*, 1:87–91, 1936.

[15] V. Yu. Shavrukov. A smart child of Peano's. *Notre Dame Journal of Formal Logic*, 35:161–185, 1994.

[16] W. Sieg. Fragments of arithmetic. *Annals of Pure and Applied Logic*, 28:33–71, 1985.

[17] C. Smoryński. Arithmetic analogues of McAloon's unique Rosser sentences. *Archive for Mathematical Logic*, 28:1–21, 1989.

[18] V. Švejdar. Modal analysis of generalized Rosser sentences. *The Journal of Symbolic Logic*, 48:986–999, 1983.

[19] A. Tarski, A. Mostowski, and R. M. Robinson. *Undecidable Theories.* North-Holland, Amsterdam, 1953.

[20] R. Verbrugge and A. Visser. A small reflection principle for bounded arithmetic. *The Journal of Symbolic Logic*, 59:785–812, 1994.

[21] A. Visser. Numerations, λ-calculus & arithmetic. In J. P. Seldin and J. R. Hindley, editors, *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, pages 259–284. Academic Press, London, 1980.

[22] A. Visser. Peano's smart children: A provability logical study of systems with built-in consistency. *Notre Dame Journal of Formal Logic*, 30:161–196, 1989.

[23] A. Visser. What is the right notion of sequentiality? Logic Group Preprint Series 288, Department of Philosophy, Utrecht University, 2010, `http://www.phil.uu.nl/preprints/lgps/`.

[24] A. Visser. Can we make the Second Incompleteness Theorem coordinate free? *Journal of Logic and Computation*, 21(4):543–560, 2011.

Derivation lengths classification of Gödel's T extending Howard's assignment

Gunnar Wilken[†], **Andreas Weiermann**[‡]

[†] Structural Cellular Biology Unit, OIST Graduate University, Okinawa, Japan
wilken@oist.jp

[‡] Department of Mathematics, Universiteit Gent, Belgium
weierman@cage.ugent.be

Abstract. Let T be Gödel's system of primitive recursive functionals of finite type in the lambda formulation. We define by constructive means using recursion on nested multisets a multivalued function I from the set of terms of T into the set of natural numbers such that if a term A reduces to a term B and if a natural number I(A) is assigned to A then a natural number I(B) can be assigned to B such that I(A) is greater than I(B). The construction of I is based on Howard's 1970 ordinal assignment for T and Weiermann's 1998 treatment of T in the combinatory logic version. As a corollary we obtain an optimal derivation lengths classification for the lambda formulation of T and its fragments. Compared with Weiermann's 1998 exposition this article yields solutions to several non-trivial problems arising from dealing with lambda terms instead of combinatory logic terms. It is expected that the methods developed here can be applied to other higher order rewrite systems resulting in new powerful termination orderings since T is a paradigm for such systems.

Introduction

This article is part of a general program of investigations on subrecursive complexity classes via derivation lengths classifications of term rewriting systems. Quite often, an equationally defined subrecursive complexity class $\mathcal{C}$ of number-theoretic functions can be defined in terms of a corresponding rewrite system $R_{\mathcal{C}}$ which computes the functions from $\mathcal{C}$. Appropriate bounds on the $R_{\mathcal{C}}$-derivation lengths then yield intrinsic information on the computational complexity of $\mathcal{C}$. Successful examples of this program have been documented, for example, in [**1, 5**]. Having such applications in mind, it seems desirable to have a large variety of powerful methods for establishing bounds on derivation lengths in general.

A common and very convenient tool for proving termination of a reduction system consists in defining an interpretation function I from the set of terms in question into the set of natural numbers such that if a term A rewrites to a term B then $I(A) > I(B)$. A rewriting sequence of terms $A_1 \to \cdots \to A_n$ then yields a strictly descending chain of natural numbers $I(A_1) > \cdots > I(A_n)$. The number $I(A_1)$ is thus an upper bound for n and hence the assignment function I provides a termination proof plus a non-trivial upper bound on resulting lengths of longest possible reductions. In this paper we apply a generalization of this method —the non-unique assignment technique— to GT^{λ}, the λ-formulation of Gödel's GT, which is the prototype for a higher order rewrite system. For $\mathrm{GT}^{\mathcal{L}}$, the combinatory logic formulation of GT, a corresponding

This paper has been published in *Logical Methods in Computer Science* under the creative commons license; see [**16**].

[†]This work was partially supported by the Mathematical Biology Unit at OIST.

interpretation has already been constructed in [**15**]. In this article we solve the technically more involved problem of classifying the derivation lengths for GT^λ via a multivalued interpretation function. The extra complications when compared with the treatment in [**15**] are due to the need for a variable concept underlying the assignment technique and to the simultaneous treatment of recursion, β-conversion, and the ξ-rule.

For a recent and extensive exposition of the history of termination proofs for Gödel's GT we refer the reader to Section 8.2 in [**4**]. In fact, [**4**] covers the history of λ-calculus and combinatory logic in general. Unlike the present paper, the majority of termination proofs for Gödel's GT mentioned in [**4**] does not yield non-trivial upper bounds on the lengths of reductions.

An alternative approach for proving termination of Gödel's GT which yields non-trivial upper bounds on the lengths of reductions was suggested in [**2**]. There the lengths of derivations were classified by proof-theoretic investigations on head-reduction trees.

The current approach is more direct, and as a possible benefit for future work we expect the extraction of powerful (syntactic) termination orderings for higher order reduction systems which generalize the recursive path ordering. We conjecture that such an ordering for Gödel's GT will have the Bachmann–Howard ordinal as order type.

Preliminary remarks

Frequent mention of Howard's work [**8**] does not mean that we presuppose its knowledge. In fact, we have intended this paper to become as much self-contained as possible. Nevertheless, we have adopted much of the notation used in [**8**] and [**15**]. Knowledge of those works together with [**3, 13, 14**] is certainly useful in order to understand this article in greater depth, but it is not required.

Section 1 introduces Gödel's GT in the typed λ-calculus version together with several well-known notions for its analysis. In addition, Subsection 1.4 will prove useful for the particular purposes of this paper, and Subsection 1.6 gives a heuristic explanation and motivation of the method of non-unique assignment. We believe that this facilitates the understanding of the method considerably. Nevertheless, the results of Subsection 1.6 are not needed later on, hence Subsections 1.6.2 and 1.6.3 can be skipped at first reading. Subsection 3.3 will clarify how 1.6 relates to our argumentation in the central section of this paper, Section 3. At that stage we expect the benefits from Subsections 1.6 and 3.3 to become fully convincing. The ordinal theoretic means applied in Section 3 are introduced in Section 2. Section 4 provides a thorough analysis of the assignments given in Section 3, showing that we obtain an exact classification of derivation lengths of GT and its fragments.

The paper is organized in a way that allows for linear reading. However, detailed technical proofs required in our argumentation are given in the appendix in order to increase readability for an audience less familiar with ordinal theory.

1 Typed λ-calculus with recursion

We give a short description of typed λ-calculus extended by recursors and case distinction functionals for each type. We will call this variant $\lambda\beta\mathsf{R}$-*calculus*. The theory GT^λ is Gödel's GT based on $\lambda\beta\mathsf{R}$-calculus. We are going to slightly deviate from the notation of [**8**].

1.1 Types and levels

The set of *finite types* is defined inductively: it contains the type 0 and the type $(\sigma\tau)$ whenever $\sigma\tau$ are finite types. Type 0 is intended to consist of the natural numbers whereas $(\sigma\tau)$ for given finite types σ and τ denotes the type of functions $f\colon \sigma \to \tau$. We will denote finite types by the letters ρ, σ, τ, etc. In cases where ambiguity is unlikely, instead of $(\sigma\tau)$ we will simply write $\sigma\tau$, and we identify $\rho\sigma\tau$ with $\rho(\sigma\tau)$. As in [**8**] the *level* of a finite type is defined recursively by

$$\mathrm{lv}(0) := 0, \quad \mathrm{lv}(\sigma\tau) := \max\{\mathrm{lv}(\sigma)+1, \mathrm{lv}(\tau)\}.$$

1.2 Terms, subterms, parse trees, substitution, α-equivalence

Let $\mathcal{V}$ be a countably infinite set of variables which we denote by X, Y, Z, etc. Following the Church-style convention, we choose a type for each of those variables in such a way that $\mathcal{V}$ contains infinitely many variables of each type. This choice is fixed throughout the paper, and we will sometimes indicate the type, say σ, of a variable X by the notation X^σ.

The set $\mathfrak{T}$ of *typed terms* is defined inductively. It contains

- all variables in $\mathcal{V}$,
- a constant $\mathsf{0}$ of type 0,
- a constant S of type 00,
- case distinction functionals[1] D_τ of type $0\tau\tau\tau$,
- primitive recursion functionals R_τ of type $0(0\tau\tau)\tau\tau$ for each τ,

and is closed under

- application, that is, (AB) is a term of type τ whenever A is a term of type $\sigma\tau$ and B is a term of type σ, and
- abstraction, that is, $\lambda X.B$ is a term of type $\rho\sigma$ whenever X is a variable of type ρ and B is a term of type σ.

We will suppress parentheses whenever ambiguity is unlikely to arise, e.g., we write AB instead of (AB). We further identify ABC with $(AB)C$ for any terms A, B, C of suitable types. If a term A is of type τ we communicate this sometimes by writing A^τ. Conversely, instead of writing $\mathsf{D}_\tau, \mathsf{R}_\tau$ we sometimes simply write D, R, respectively. We continue to denote terms of $\mathfrak{T}$ by Roman capital letters with the exception of recursion arguments in terms of the form $\mathsf{R}t$, where we sometimes use lower case Roman letters s, t, etc.

The set $\mathrm{BV}(A)$ of *bound variables* of a term A consists of all variables X for which λX occurs in A, whereas the set $\mathrm{FV}(A)$ of *free variables* of A consists of all variables X occurring in A outside the scope of λX.

The set of *subterms* of a given term A is defined as usual by induction on the buildup of A, including A itself. The *direct subterms* of terms of a form AB are A and B; the only direct subterm of $\lambda X.A$ is A —all other terms do not have any direct subterm.

The *parse tree* of a term A is the labeled tree whose root is labeled with A, and whose immediate subtrees are the parse trees of the immediate subterms of A (if any). The set of all labels of nodes of the parse tree of a term A is therefore equal to the set of subterms of A. The parse tree, however, distinguishes between possibly various occurrences of the same subterm of A and provides their contexts within A. The set of nodes of the parse tree of a term A can be identified with the *addresses* of A, which are finite strings

[1] Case distinction functionals are needed to the full extent in the functional interpretation of the fragments $\mathrm{I}\Sigma_{n+1}$ of Peano Arithmetic in the fragments GT_n of GT; cf. [**9**].

over $\{1,2\}$ according to the following definition, where ϵ denotes the empty string. The *subterm of A at address a*, written $A_{|a}$, is defined inductively by

- A if $a=\epsilon$,
- $B_{|b}$ if $A=\lambda X.B$ and $a=1b$,
- $B_{|b}$ if $A=BC$ and $a=1b$,
- $C_{|c}$ if $A=BC$ and $a=2c$.

Substitution $A\{X:=B\}$ is defined as usual by induction on the buildup of A, replacing any free occurrence of the variable X in A by the term B. The variable condition

$$\mathrm{BV}(\lambda X.A)\cap\mathrm{FV}(B)=\emptyset$$

for β-conversion (see below) avoids variable capture. In other words, it makes sure that none of the free variables in B becomes bound within $A\{X:=B\}$. Moreover, the abstraction variable X does not occur in $A\{X:=B\}$.

A term A is called *well-named* iff for each subterm B of A we have

$$\mathrm{FV}(B)\cap\mathrm{BV}(B)=\emptyset,$$

and furthermore, for each variable X, the binder λX occurs at most once in B.

Let A be a term with subterm $\lambda X.B$ at address a. If Y is a variable of the same type as X such that $Y\notin\mathrm{FV}(B)$, then we say that A *α-converts* to the term resulting from A by the replacement of $A_{|a}$ by $\lambda Y.(B\{X:=Y\})$. If a term C is obtained from A via a finite number of α-conversions, then we call A and C *α-equivalent*, $A=_\alpha C$. Notice that every term is α-equivalent to some well-named term.

1.3 Equivalence and $\lambda\beta$R-reduction

The *equivalence relation* is the reflexive, symmetric and transitive closure of α-equivalence and the *one-step reduction* $\rhd$, defined inductively as least binary relation on terms such that

$$\begin{array}{llll}
(D_0) & \mathsf{D}0AB\rhd A & (D_{\mathsf{S}}) & \mathsf{D}(\mathsf{S}t)AB\rhd B\\
(R_0) & \mathsf{R}0AB\rhd B & (R_{\mathsf{S}}) & \mathsf{R}(\mathsf{S}t)AB\rhd At(\mathsf{R}tAB)\\
(App_r) & A\rhd B\Rightarrow AC\rhd BC & (App_l) & B\rhd C\Rightarrow AB\rhd AC\\
(\beta) & \multicolumn{3}{l}{(\lambda X.A)B\rhd A\{X:=B\} \text{ where } \mathrm{BV}(\lambda X.A)\cap\mathrm{FV}(B)=\emptyset}\\
(\xi) & \multicolumn{3}{l}{A\rhd B\Rightarrow\lambda X.A\rhd\lambda X.B.}
\end{array}$$

Clearly, the variable condition for β-conversion is satisfied if the term $(\lambda X.A)B$ is well-named. Note that $\rhd$ does not preserve well-namedness. However, as mentioned above, well-namedness can always be restored: if $A\rhd B'$ then there exists a well-named term B such that $A\rhd B'=_\alpha B$. On well-named terms we define $A\rhd_\alpha B$ iff there exists (the unique) B' such that $A\rhd B'=_\alpha B$.

The calculus $\lambda\beta$R enjoys the Church–Rosser property (confluence) and strong normalization, and every functional of type 0 reduces to a numeral (cf. [7]).

Note that extending $\lambda\beta$R by η-contractions would cause loss of the Church–Rosser property —consider for example the term $\lambda X.\mathsf{R}0AX$.

1.4 Extension to $\mathfrak{T}'$ and $\lambda\beta$R$'$

In order to prepare a separate treatment of R-reductions and β-reductions, which otherwise would be incompatible within the framework of our intended non-unique assignment, we introduce the following definitional extension of $\lambda\beta\mathsf{R}$-calculus. We extend the clauses in 1.2 defining the terms of $\mathfrak{T}$ so that

- R^t is a term of type $(0\tau\tau)\tau\tau$ for any R_τ and any term t of type 0.

We call the extended set of terms $\mathfrak{T}'$. The terms R^t are special forms of application terms and thus allow for a clean way to give two different assignments to terms resulting from the application of some t to a recursor R. The effect will be that assignments to $\mathsf{R}t$ allow for a treatment of β-conversion while assignments to R^t allow for a treatment of R-reduction. This smoothly fits with our treatment of the ξ-rule by the method of non-unique assignment according to [**8**], which takes advantage from the additional information given by the reduction history of terms; see Subsection 1.6.

The subterms of R^t are R^t itself, R, and the subterms of t. The direct subterms are R and t. Parse trees for the extension are defined accordingly; $(\mathsf{R}^t)_{|a}$ is defined by $(\mathsf{R}t)_{|a}$. Substitution for the new terms is defined by

$$\mathsf{R}^t\{X := A\} :\equiv \mathsf{R}^{t\{X:=A\}},$$

and α-equivalence is expanded to all terms of $\mathfrak{T}'$. The one-step reduction relation is then modified to include additional rules (R) and (App_R), and (R_0) and (R_{S}) are replaced by (R^0) and (R^{S}), respectively, as follows:

$$\begin{array}{llll} (R) & \mathsf{R}t \rhd \mathsf{R}^t & (App_R) & s \rhd t \Rightarrow \mathsf{R}^s \rhd \mathsf{R}^t \\ (R^0) & \mathsf{R}^0 AB \rhd B & (R^{\mathsf{S}}) & \mathsf{R}^{\mathsf{S}t}AB \rhd At(\mathsf{R}^t AB). \end{array}$$

We denote the modified calculus for the terms of $\mathfrak{T}'$ by $\lambda\beta\mathsf{R}'$. The properties and definitions regarding $\lambda\beta\mathsf{R}$ mentioned in the previous subsection are easily seen to carry over to $\lambda\beta\mathsf{R}'$.

We define the *length* of a term A, $\mathrm{lh}(A)$, as follows:

- $\mathrm{lh}(A) := 1$ if A is a variable or constant,
- $\mathrm{lh}(BC) := \mathrm{lh}(B) + \mathrm{lh}(C)$,
- $\mathrm{lh}(\mathsf{R}^t) := 1 + \mathrm{lh}(t)$, and
- $\mathrm{lh}(\lambda X.G) := \mathrm{lh}(G) + 1$.

It is a trivial observation that by identifying any R^t with $\mathsf{R}t$ and omitting all (R)-reductions we recover the original $\lambda\beta\mathsf{R}$-calculus where (App_R)-reductions turn into (App_l)-reductions and (R^0), (R^{S}) turn into (R_0), (R_{S}), respectively. Clearly, reduction sequences can only become shorter in this process.

On the other hand, given any reduction sequence in $\lambda\beta\mathsf{R}$, we obtain a corresponding reduction sequence in $\lambda\beta\mathsf{R}'$ (of at most double length) by straightforward insertion of (R)-reductions as needed, from which we recover the original sequence by the process described above.

Another trivial observation is that given a reduction sequence in $\lambda\beta\mathsf{R}'$, starting from a term $A \in \mathfrak{T}'$ we can straightforwardly find a term $A' \in \mathfrak{T}$ and a reduction sequence which transforms A' into A in at most $\mathrm{lh}(A)$-many steps, only using (R)-reduction. It is therefore sufficient to perform a derivation lengths classification for the class of reduction sequences in $\lambda\beta\mathsf{R}'$ which start from $\mathfrak{T}$-terms, as the resulting bounds will turn out to be sharp for $\lambda\beta\mathsf{R}$.

1.5 Reduction trees

The rules $(App_l), (App_r), (\xi)$, and (App_R) imply that the one-step reduction $\rhd$ applied to some term A consists of the reduction of one redex of A. A redex of A is the occurrence of a subterm of A (corresponding to a unique node in the parse tree of A) that matches any left-hand side of the reduction clauses for D-, R-, or β-reduction. Inside A the redex is then replaced by the (properly instantiated) right-hand side of the corresponding clause. We call the redex chosen for a particular one-step reduction $\rhd$ the *working redex.* The *reduction tree* of a (well-named) term A is then given by the exhaustive application of $\rhd_\alpha$ in order to pass from a parent to a child node, starting from A. We do not need to be any more specific about the arrangement of nodes in reduction trees.

1.6 Assignment of ordinals to terms of $\mathfrak{T}'$

1.6.1 Overview

The aim of the present paper is to give exact bounds on the heights of reduction trees in the usual sense of derivation lengths classification for term rewriting systems. This is achieved by the assignment of strictly decreasing natural numbers to the terms of reduction sequences. The natural numbers assigned to terms are computed along vectors of ordinal terms (from upper down to lower components) which in turn are built over variable vectors that correspond to typed variables.

The assignment is not unique for terms because it is dependent on the respective reduction history. The main reason for this dependency on reduction histories is the same as already encountered in [**8**]. The treatment of β-reduction involves an operator on ordinal vectors which is not monotone with respect to the ξ-rule. The other reason is the incompatibility of our treatments of R- and β-reduction, as already mentioned in Subsection 1.4. The solution outlined there is based on the fact that when considering $\lambda\beta\mathsf{R}'$-reduction sequences which start from terms in $\mathfrak{T}$, we can trace back abstraction subterms in the reduction history, obtaining *corresponding* subterms, cf. Definition 1.1, in which the respective abstraction variable does not occur in subterms of the form R^t. This observation is crucial for our simultaneous treatment of β-reductions, the ξ-rule, and *arbitrary* R-reductions, which were excluded in [**8**]. The assignment to such corresponding terms earlier in the reduction history is then the key to the handling of β-reductions that may occur much later in the reduction sequence.

Now, given the particular reduction history of a term A occurring in a fixed reduction sequence, the assignment is determined uniquely (we will introduce the crucial notion of *assignment derivation* in our formal argumentation) and built up from the assignments to the nodes of the parse tree of A and terms occurring earlier in the reduction history of A. Additionally, the assembly of new assignments along the reduction sequence involves (iterated) substitutions of variable vectors by already defined assignments, generating terms which do *not* occur as subterms of terms in the reduction history of A. For this reason, besides the obvious reason in the treatment of β-reduction, the assignment method has to be designed so as to naturally commute with substitution, as was done already in [**8**]. Another essential property of our assignment method is its invariance under α-equivalence, as in [**8**]. This will enable us to treat $\rhd_\alpha$-reductions in the same way as $\rhd$-reductions.

Our construction of assignments to terms will start from unique assignments to all variables and constants, using Howard's operator $\Box$ in its refined form of [**15**] to compute an assignment to a term BC from assignments to B and C, a specific treatment of terms R^t as used in [**15**], where it was used for terms of the form $\mathsf{R}t$, and a refinement of Howard's operator δ, cf. [**8**], to be applied in the treatment of abstraction terms, which causes the non-uniqueness of the assignment method.

1.6.2 Basic considerations

Given a reduction $A \rhd B$ we begin with describing how the parse tree for B is obtained from the parse tree for A in a uniform way. This will be crucial for the construction of our assignment. Focusing on the working redex of the reduction $A \rhd B$ we consider the path $\mathcal{P}$ from the root, labeled with A, to the working redex at node r, labeled with F. Clearly, F is a subterm of any term labeling a node of $\mathcal{P}$. Assume the working redex is reduced (via D-, R-, or β-reduction) to the term G at node s in the parse tree of B. The subtree with root r of the parse tree of A, that is the parse tree for F, is replaced with the parse tree for G, and along the path $\mathcal{P}$ the labels are modified by the replacement of the working redex F by the reduct G. If the meaning is clear from the context we will sometimes denote such a replacement by $H[s/r]$ where H is a label of a node on $\mathcal{P}$. All remaining nodes of the parse tree of A are preserved in the parse tree of B with the same labeling term. In other words, the tree structure is modified by the replacement of the parse tree of F at node r by the parse tree of G with the corresponding labeling, while the modification of labels additionally involves the labels along $\mathcal{P}$ in form of the replacement $[s/r]$ of the working redex F by G.

In the case of D- and R-reductions the transformation of the parse tree of F to the parse tree of G is clear, and we can uniquely identify subtrees of redex and reduct, including their labels. For example in the case where F is a term $\mathsf{R}^{\mathsf{S}t}CD$ and G is the term $Ct(\mathsf{R}^t CD)$, we can trace the parse tree of F until we reach the parse trees of t, C, and D, and do the same with the parse tree of G, identifying the corresponding subtrees. We are going to use this identification of subtrees in our assignment in order to carry over already defined assignments.

Now consider the case where the working redex is a β-redex, say F is $(\lambda X.C)D$ and G is $C\{X := D\}$. The immediate subtrees of the parse tree of F are then the parse trees of $\lambda X.C$ and D, the immediate subtree of the former being the parse tree $\mathcal{S}$ of C. Consider the tree $\mathcal{S}\{X := D\}$ which is obtained as follows. The tree structure is obtained from $\mathcal{S}$ by replacing every leaf with label X by the parse tree of D with the corresponding labels. The remaining labels are modified by the substitution $\{X := D\}$. The subtrees substituted for the leaves with label X can be identified with the immediate subtree of F which is the parse tree of D.

Having discussed the transition from the parse tree of A to the parse tree of B, where $A \rhd B$, let us now assume that $Y \in \mathrm{FV}(B)$ and that C is a term of the same type as Y such that $\mathrm{BV}(B) \cap \mathrm{FV}(C) = \emptyset$. Modulo α-congruence we may assume that also $\mathrm{BV}(A) \cap \mathrm{FV}(C) = \emptyset$. We then clearly have $Y \in \mathrm{FV}(A)$, and $A\{Y := C\} \rhd B\{Y := C\}$. The parse trees of $A\{Y := C\}$ and $B\{Y := C\}$ are obtained from those for A and B, respectively, by replacing every leaf with label Y by the parse tree of C and by the modification of all remaining labels by the substitution $\{Y := C\}$.

1.6.3 Precise motivation

Bearing the above preparation in mind we proceed with a precise explanation of the method of non-unique assignment that was used in Section 4 of [**8**] in order to handle the unrestricted ξ-rule. As mentioned above, the approach of non-unique assignment had to be refined so as to manage arbitrary R-reductions.

Definition 1.1 Let A, B be terms such that $A \rhd B$ with working redex and reduct at address w, respectively. Let b be an address in B such that $b = q1$ and $B_{|q}$ is an abstraction. We define a unique address a and say that *a corresponds to b* with respect to the pair (A, B) as follows.

(1) If b and w are incomparable, then we have $A_{|b} = B_{|b}$ and set $a := b$.
(2) If b is a prefix of w (written as $b \subseteq w$), then we have $A_{|b} \rhd B_{|b}$ and set $a := b$.
(3) If w is a proper prefix of b, that is, $w \subsetneq b$.
  (3.1) $w \neq \epsilon$. Then let b' be such that $b = wb'$, let a' be such that a' corresponds to b' with respect to $(A_{|w}, B_{|w})$, and set $a := wa'$.
  (3.2) $w = \epsilon$. Then we distinguish between the following cases:
    (a) $A = (\lambda X.C)D$, $B = C\{X := D\}$.
      – If b is of a form cd such that $C_{|c} = X$ then $a := 2d$,
      – otherwise $a := 11b$.
    (b) $A = \mathsf{R}t$, $B = \mathsf{R}^t$. Then set $a := b$.
    (c) $A = \mathsf{R}^0 CB$. Set $a := 2b$.
    (d) $A = \mathsf{R}^{\mathsf{S}t} CD$, $B = Ct(\mathsf{R}^t CD)$.
      – If b is of a form $22d$, then $a := 2d$.
      – If b is of a form either $11c$ or $212c$, then $a := 12c$.
      – If b is of a form either $12e$ or $2112e$, then $a := 1122e$.
    (e) $A = D0BC$. Set $a := 12b$.
    (f) $A = \mathsf{D}(\mathsf{S}t)CB$. Set $a := 2b$.

In case (2) we call the reduction $A_{|b} \rhd B_{|b}$ the *associated reduction* with respect to (A, B) and b, otherwise an associated reduction is not defined.

If the working redex is a β-redex, say $A_{|w} = (\lambda X.C)D$, and $w \subsetneq b$, b not of a form $b = wcd$ with $C_{|c} = X$, then we call the substitution $\{X := D\}$ the *associated substitution* with respect to (A, B) and b, otherwise an associated substitution is not defined.

For well-named terms A, B such that $A \rhd_\alpha B$ and addresses a, b in A, B, respectively, let B' be the unique term such that $A \rhd B' =_\alpha B$. Then we say that a corresponds to b with respect to the pair (A, B) iff a corresponds to b with respect to the pair (A, B'). The associated reduction (respectively, associated substitution) with respect to (A, B) and b is defined iff it is defined for (A, B') and b, in which case they are the same. ◇

Note that for any A, B, a, b as in the above definition, a is of the form $p1$, and $A_{|p}$ is an abstraction. Notice further that exactly one of the following holds:

(1) The associated reduction is defined.
(2) The associated substitution is defined.
(3) Neither the associated reduction nor the associated substitution is defined.

Notice that if $A \rhd B$, in case (I) we have $A_{|a} \rhd B_{|b}$, in case (II) we have $A_{|a}\{X := D\} = B_{|b}$ where $\{X := D\}$ is the associated substitution, and in case (III) we have $A_{|a} = B_{|b}$.

However, if we have $A \rhd_\alpha B$, say $A \rhd B' =_\alpha B$, in general the terms $B_{|b}$ and $B'_{|b}$ do not only differ due to the renaming of bound variables but also due to renaming of free variables as they might be subterms of abstractions that have undergone α-conversion. In particular, it does not make sense to keep track of the abstraction variable of $B_{|q}$. The following lemma addresses this technical issue.

Lemma 1.2 *Let A, B, C be terms such that $A \rhd B =_\alpha C$ and suppose that A, C are well-named and that the abstraction variables used in order to α-convert B to C do neither occur free nor bound in A or B. Let $b = q1$ be an address such that $C_{|q}$ is an abstraction term, say $\lambda X.H$, and let $a = p1$ be the corresponding address in A. Then there exists a well-named term A^* such that*

(1) $A^* =_\alpha A$,
(2) $A^* \rhd B^* =_\alpha C$,
(3) $B^*_{|q} = \lambda X.(B^*_{|b})$ *where* $B^*_{|b} =_\alpha H$, *and*
(4) *exactly one of the following holds:*
  (a) $A^*_{|a} \rhd H$,
  (b) $A^*_{|a}\{Y := D\} =_\alpha H$ *where* $\{Y := D\}$ *is the associated substitution with respect to* (A^*, C),
  (c) $A^*_{|a} =_\alpha H$.

Proof. Notice first of all that (2) follows from (1) and B^* is determined by (2). Suppose $\{v_1, \ldots, v_m\}$ is the set of addresses in B at which α-conversion has to be performed in order to obtain C (the order of those α-conversions is of course irrelevant), with corresponding new variables $V_1, \ldots, V_m$. Let w be the address of the working redex in A and hence also the address of the reduct in B. In the case $m = 0$ we choose $A^* := A$ and are done.

Now suppose that $m > 0$ and consider $v \in \{v_1, \ldots, v_m\}$ with corresponding new abstraction variable V. We investigate whether an α-conversion in A becomes necessary in order to satisfy the lemma. The collection of all α-conversions that have to be performed in A will then determine the term A^*.

If v is not a prefix of q, then the α-conversion at v in B does not cause any change in $C_{|q}$ and hence does not require any additional α-conversion in A.

Suppose now that v is a prefix of q, that is, $v \subseteq q$. We then consider three cases regarding the addresses v and w.

Case 1: v and w are incomparable. Then we have $p = q$, $v1$ corresponds to $v1$ with respect to (A, B), and we draw the α-conversion at v in B back to an α-conversion at v in A with the same new variable V.

Case 2: $v \subsetneq w$. Then $A_{|v}$ is of a form $\lambda Z.G$ and $A_{|v} \rhd B_{|v} = \lambda Z.G'$. Hence

$$C_{|v} = \lambda V.G'\{Z := V\}.$$

Again $v1$ corresponds to $v1$ with respect to (A, B), and we draw the α-conversion at v in B back to an α-conversion at v in A with the same new variable V. The subterms of $A_{|v}$ are now subject to the variable substitution $\{Z := V\}$, which includes the working redex and the term $A_{|p}$.

Case 3: $w \subseteq v$. Then we have $w \subseteq v \subseteq q$. Let u be the address in A such that $u1$ corresponds to $v1$ with respect to (A, B). We have $v1 \subseteq b$ and $w \subsetneq u$ (abstractions cannot be working redices). We perform the α-conversion at u in A switching to the abstraction variable V. The reduction of A to B might generate further copies of $A_{|u}$ in B whose addresses in B, however, cannot be prefixes of q. In B^* we will have corresponding copies, differing from those in B by the abstraction variable V. Hence α-conversions of those modified copies in B^* to fit the corresponding abstractions in C are possible and determined by C. □

Definition 1.3 Let $F_0 \rhd \cdots \rhd F_n$ be a reduction sequence and let p be the address identified with a node in the parse tree of F_n that is labeled with an abstraction term $\lambda X.H$. The *trace* of p in $F_0 \rhd \cdots \rhd F_n$ is the sequence $(b_0, \ldots, b_n)$ such that $b_n = p1$ and each b_i corresponds to b_{i+1} with respect to (F_i, F_{i+1}). The *associated trace terms*

$H_0, \ldots, H_n$ are the terms $F_{0|b_0}, \ldots, F_{n|b_n}$, thus $H_n = H$. Let a partition I_R, I_S, I_E of the index set $\{0, \ldots, n-1\}$ be defined as follows:

- $i \in I_R$ if the associated reduction with respect to (F_i, F_{i+1}) and b_{i+1} is defined,
- $i \in I_S$ if the associated substitution with respect to (F_i, F_{i+1}) and b_{i+1} is defined, and
- $i \in I_E$ otherwise,

according to the remark following Definition 1.1. The index set I_S therefore gives rise to the *associated partial substitution list.*

For a reduction sequence $F_0 \rhd_\alpha \cdots \rhd_\alpha F_n$ we define the same notions, proceeding as in Definition 1.1, and defining the trace terms H_i by $F_{i|b_i}$. ◇

Given a reduction sequence $F_0 \rhd_\alpha \cdots \rhd_\alpha F_n$ we have unique terms $F'_1, \ldots, F'_n$ such that

$$F_0 \rhd F'_1 =_\alpha F_1 \ldots F_{n-1} \rhd F'_n =_\alpha F_n.$$

Assume that each conversion from F'_{i+1} to F_{i+1} only uses new variables, i.e., variables that have not been used earlier in the reduction sequence. Clearly, this requirement on the reduction sequence can always be obtained via α-equivalence (possibly including renamings of bound variables in F_n). Given an address p identified with a node in the parse tree of F_n that is labeled with an abstraction term $\lambda X.H$, let $(b_0, \ldots, b_n)$ be the trace of p and I_R, I_S, I_E be the partition of the set of indices $\{0, \ldots, n-1\}$ according to Definition 1.3.

Let $F^*_n := F_n$. Iterated application of Lemma 1.2 starting from $F_{n-1} \rhd F'_n =_\alpha F_n$ yields terms $F^*_{n-1}, \ldots, F^*_0$ such that

(1) $F_i =_\alpha F^*_i$ for each i,

(2) $F^*_0 \rhd_\alpha \cdots \rhd_\alpha F^*_n$ with the same trace of p, the same partition I_R, I_S, I_E, and trace terms $H_i := F^*_{i|b_i}$, and

(3) for each $i < n$ we have:

  (a) $H_i \rhd H_{i+1}$ if $i \in I_R$,

  (b) $H_i\{Y_i := D_i\} =_\alpha H_{i+1}$ if $i \in I_S$ where $\{Y_i := D_i\}$ is the associated substitution with respect to (F^*_i, F^*_{i+1}), or

  (c) $H_i =_\alpha H_{i+1}$ if $i \in I_E$.

(4) The associated partial list of substitutions with respect to $F^*_0 \rhd_\alpha \cdots \rhd_\alpha F^*_n$ and p is X*-free*, that is, none of the terms D_i which is defined contains the variable X and none of the defined variables Y_i is identical with X.

(5) Each b_i is of a form $b'_i 1$ and $F^*_{i|b'_i} = \lambda X.H_i$.

Definition 1.4 For a reduction sequence $F_0 \rhd_\alpha \cdots \rhd_\alpha F_n$ with the above specified condition of freshness of abstraction variables introduced by α-conversions and node p in F_n as above, we call the sequence $F^*_0 \rhd_\alpha \cdots \rhd_\alpha F^*_n$ defined above the p*-companion* of $F_0 \rhd_\alpha \cdots \rhd_\alpha F_n$. If $F_0 \rhd_\alpha \cdots \rhd_\alpha F_n$ can be a p-companion of itself, then we call it *nice* with respect to p. ◇

Notice that the crucial property for p-niceness of reduction sequences roots in property 1.2 of Lemma 1.2: the invariance of the free variables which occur in the trace terms.

Definition 1.5 Let $F_0 \rhd_\alpha \cdots \rhd_\alpha F_n$, $n > 0$, be a nice reduction sequence with respect to a node p in the parse tree of F_n that is labeled with an abstraction term $\lambda X.H$. Let further $(b_0, \ldots, b_n)$ be the trace of p in $F_0 \rhd_\alpha \cdots \rhd_\alpha F_n$ with associated trace terms $H_0, \ldots, H_n$ and associated partial list of substitutions $(\{Y_i := D_i\})_{i \in I_S}$.

For each $i \in \{0, \dots, n-1\}$ we define terms $\{H_i^j\}_{i \leq j \leq n}$ such that

- $H_i^i := H_i$,
- $H_i^{j+1} =_\alpha C_i^j\{Y_j := D_j\}$ if $j \in I_S$, where
  - $C_i^j =_\alpha H_i^j$,
  - H_i^{j+1} and $(\lambda Y_j.C_i^j)D_j$ are well-named,

  using only fresh variables for α-conversion,
- H_i^{j+1} otherwise.

Let $(i_j)_{0 \leq j \leq m}$ be the longest sequence such that

- $i_0 = 0$,
- i_{j+1} is the least $i > i_j$ such that $H_{i-1} \rhd H_i$.

We now let $G_m := H_n$; define $G_j := H_{i_j}^n$ for each $j \in \{0, \dots, m-1\}$, and call the sequence $G_0 \rhd_\alpha \cdots \rhd_\alpha G_m$ the *associated reduction sequence* with respect to $F_0 \rhd_\alpha \cdots \rhd_\alpha F_n$ and p.

In the case of the trivial reduction sequence F_0 with node p in the parse tree of F_0 that is labeled with $\lambda X.H$ we define $G_0 := H$ to be the associated reduction sequence with respect to F_0 and p. ◇

Notice that in the above definition $H_{i_m}^n =_\alpha H_n$ which justifies our choice of G_m. We have now carefully shown how, making extensive use of α-conversion, abstraction terms can be traced in reduction sequences. This will be essential in the treatment of the ξ-rule. As mentioned before, our assignments to terms will be invariant modulo α-congruence.

The inductive definition below introduces a binary relation w originating from t on p. 457 of [**8**], which assigns weights to terms of $\mathfrak{T}'$ in a non-unique manner. For the purpose of the present section the assignment of weights serves as a useful illustration. However, the use of weights will not be required later on.

Definition 1.6 Let w be the minimal binary relation on $\mathfrak{T}' \times \mathbb{N}^+$ which satisfies:

- $\mathrm{w}(A, 1)$ if A is a variable or constant,
- $\mathrm{w}(BC, n)$ if $\mathrm{w}(B, n_B)$, $\mathrm{w}(C, n_C)$, and $n = n_B + n_C$,
- $\mathrm{w}(\mathsf{R}^t, n+1)$ if $\mathrm{w}(t, n)$,
- $\mathrm{w}(\lambda X.G, n)$ if there are terms $G_0 \rhd_\alpha \cdots \rhd_\alpha G_m = G$ such that X does not occur in any subterm of G_0 of a form R^t, $\mathrm{w}(G_i, n_i)$ for $i = 0, \dots, m$, and $n = 1 + n_0 + \cdots + n_m$.

Every relation $\mathrm{w}(F, n_F)$ for some $F \in \mathfrak{T}'$ comes with a witnessing *derivation*, which is a tree whose root is labeled with $\mathrm{w}(F, n_F)$, defined as follows:

- If A is a variable or constant, then the tree consisting only of its root is the derivation of $\mathrm{w}(A, 1)$.
- For derivations $\mathcal{R}$ of $\mathrm{w}(B, n_B)$ and $\mathcal{S}$ of $\mathrm{w}(C, n_C)$, the tree with direct subtrees $\mathcal{R}$ and $\mathcal{S}$ is a derivation of $\mathrm{w}(BC, n_B + n_C)$.
- For a derivation $\mathcal{R}$ of $\mathrm{w}(t, n)$, the tree with direct subtree $\mathcal{R}$ is a derivation of $\mathrm{w}(\mathsf{R}^t, n+1)$.
- For derivations $\mathcal{R}_0, \dots, \mathcal{R}_m$ of $\mathrm{w}(G_0, n_0), \dots, \mathrm{w}(G_m, n_m)$ with

  $$G_0 \rhd_\alpha \cdots \rhd_\alpha G_m =: G$$

  such that X does not occur in any subterm of G_0 of a form R^t, the tree with direct subtrees $\mathcal{R}_0, \dots, \mathcal{R}_m$ is a derivation of $\mathrm{w}(\lambda X.G, 1 + n_0 + \cdots + n_m)$. ◇

In Section 3, instead of weights we will assign ordinal vectors to terms, resulting in *assignment derivations*; see Definition 3.1. The notion of assignment derivation will be *more restrictive* than the notion of derivation here. Reduction sequences $G_0 \rhd_\alpha \cdots \rhd_\alpha G_m = G$ as mentioned above for the treatment of an abstraction $\lambda X.G$ will additionally have to have a strictly decreasing sequence of assignments, which is crucial in the treatment of the ξ-rule.

We are going to show how weights can be assigned to terms along any reduction sequence $F_0 \rhd_\alpha \cdots \rhd_\alpha F_n$, in a way that is compatible with Definition 1.5, relying on associated reduction sequences. Recall the observation that, given any reduction sequence $F_0 \rhd_\alpha \cdots \rhd_\alpha F_n$ of terms in $\mathfrak{T}'$, we may prepend a sequence of terms starting with a term in $\mathfrak{T}$ which reduces to F_0 in at most $\mathrm{lh}(F_0)$-many steps and merely involves (R)-redices as working redices. We may therefore assume that $F_0 \in \mathfrak{T}$.

Lemma 1.7 *The relation* w *is invariant modulo α-congruence.*

Proof. Straightforward. □

The above lemma justifies to consider nice reduction sequences without loss of generality. The next definition relates derivations to substitution.

Definition 1.8 Let $A, D \in \mathfrak{T}'$ be such that D is substitutable for Y in A, i.e., $\mathrm{BV}(A) \cap \mathrm{FV}(D) = \emptyset$. For any fixed derivations of $\mathrm{w}(A, n_A)$ and $\mathrm{w}(D, n_D)$ we define a canonical derivation of $\mathrm{w}(A\{Y := D\}, n)$ with suitable n. The definition proceeds by induction along the inductive definition of derivation of $\mathrm{w}(A, n_A)$.

(1) $A = Y$. Then choose $n := n_D$ and the derivation of $\mathrm{w}(D, n_D)$.

(2) $Y \notin \mathrm{FV}(A)$. Then choose $n := n_A$ and the derivation of $\mathrm{w}(A, n_A)$.

(3) $A = BC$ with derivations of $\mathrm{w}(B, n_B)$ and $\mathrm{w}(C, n_C)$, and $n_A = n_B + n_C$. The canonical derivation of $\mathrm{w}((BC)\{Y := D\}, n)$ is then assembled from the already defined canonical derivations $\mathrm{w}(B\{Y := D\}, n_1)$ and $\mathrm{w}(C\{Y := D\}, n_2)$, setting $n := n_1 + n_2$.

(4) $A = \mathsf{R}^t$ with a derivation of $\mathrm{w}(t, n_t)$ and $n_A = 1 + n_t$. We have the canonical derivation of $\mathrm{w}(t\{Y := D\}, n_1)$, from which, setting $n := 1 + n_1$, we define the canonical derivation of $\mathrm{w}(A\{Y := D\}, n)$.

(5) $A = \lambda X.B$, where $X \neq Y$, with a sequence $B_0 \rhd_\alpha \cdots \rhd_\alpha B_m = B$ such that X does not occur in any subterm of B_0 of a form R^t, and derivations of $\mathrm{w}(B_i, n_{B_i})$ for $i = 0, \ldots, m$, and $n_A = 1 + n_{B_0} + \cdots + n_{B_m}$. By the previous lemma we may assume that D is substitutable for Y in every B_i, $i = 0, \ldots, m$. We already have the canonical derivation of each $\mathrm{w}(B_i\{Y := D\}, n_i)$. By assumption, X does not occur in D, hence X does not occur in any subterm of $B_0\{Y := D\}$ of a form R^t. We have

$$B_0\{Y := D\} \rhd_\alpha \cdots \rhd_\alpha B_m\{Y := D\} = B\{Y := D\}$$

and assembly the canonical derivation of $\mathrm{w}(A\{Y := D\}, n)$, $n := 1 + n_0 + \cdots + n_m$.

Definition 1.9 We give an inductive definition of canonical weight assignments along reduction sequences. Suppose $F_0 \rhd_\alpha \cdots \rhd_\alpha F_n$ is a $\lambda\beta\mathsf{R}'$-reduction sequence with $F_0 \in \mathfrak{T}$ and let p be a node in F_n with label A. Assume that derivations of $\mathrm{w}(B, n_B)$ have been specified for all terms B labeling nodes in the parse trees of F_i for each $i < n$ and to nodes descending from p in the parse tree of F_n. In case A is a variable or constant we are done with the unique derivation of $\mathrm{w}(A, 1)$, and if $A = BC$, we have derivations of $\mathrm{w}(B, n_B)$ and $\mathrm{w}(C, n_C)$ for the labels B and C of the direct child nodes of p and accordingly choose the canonical derivation of $\mathrm{w}(A, n_B + n_C)$. The interesting case is where $A = \lambda X.G$. Here, let us assume that $F_0 \rhd_\alpha \cdots \rhd_\alpha F_n$ is nice with respect to p, justified by Lemma 1.7. Let $G_0 \rhd_\alpha \cdots \rhd_\alpha G_m$, where $G_m = G$, be the associated reduction sequence for p in $F_0 \rhd_\alpha \cdots \rhd_\alpha F_n$ according to Definition 1.5. Then Definition 1.8 yields canonical derivations $\mathrm{w}(G_i, n_i)$ for each i, from which we assembly the canonical derivation of $\mathrm{w}(A, 1 + n_0 + \cdots + n_m)$. ◇

We now see that we may give a non-unique assignment to terms of $\mathfrak{T}'$ inductively along a definition similar to Definition 1.6. As illustrated in the case of derivations, our assignment method will be unique with respect to assignment derivations. We can then conveniently prove properties of the assignment method by induction along assignment derivations; cf. Definition 1.8. As illustrated in Definition 1.9, we will have canonical assignment derivations in the context of reduction sequences. The assignment method will thus easily be seen to be constructive.

The central result obtained in this paper will be a constructive procedure that, given terms $A, B \in \mathfrak{T}'$ such that $A \rhd B$ and given an assignment in form of a natural number a to A together with its *derivation*, outputs the derivation of an assignment $b \in \mathbb{N}$ to B such that $a > b$. The uniformity of the procedure then allows for a derivation lengths classification of $\lambda\beta\mathsf{R}'$ and hence also of $\lambda\beta\mathsf{R}$.

2 Ordinal terms and vectors

This section provides the theoretical framework of our assignment of *ordinal vectors* to terms of $\mathfrak{T}'$. The computational complexity of the calculus $\lambda\beta\mathsf{R}$ poses a challenge regarding the determination of appropriate upper bounds on the lengths of reduction sequences. Ordinal terms containing exponential towers occur —see Definition 2.10— and require precise bookkeeping. Due to the presence of recursion with arguments not explicitly known at the beginning of a reduction sequence, the ordinal ω occurs in our assignments, leading to ordinal terms below the ordinal ε_0 which is the least fixed point of exponentiation to base ω. Bookkeeping during the computation of upper bounds will require a sufficiently expressive term algebra, and using ordinal vectors will help keeping track of exponential (sub-)terms of a certain height. Vectors allow us to build up ordinal terms from the upper components down; see Definitions 2.6, 2.9, 2.10, 2.18. The δ-operator will require decomposition of ordinal terms, which is facilitated by the concept of ordinal vectors. The starting component is determined by the type level of the term for which the assignment is being defined; cf. Subsection 1.1. This approach was designed in [**8**] and is used here with some modifications. The innovation, first used by Weiermann in a treatment of the combinatory logic version of GT (see [**15**]) concerns the 0-th component. The 0-th component of the vectors generated will in general result from a collapsing operation below ω and save an ω-power when compared to [**8**];[2] see Definition 2.10 and cf. also the difference between [**14**] and [**15**]. In order to treat arbitrary R-reductions, where the recursion argument has not yet been reduced to a numeral and might even contain variables, we feed collapsed terms back into the process of vector generation; see Definition 3.1. The 0-th component of a vector assigned to such a recursion argument, say t, turns out to be both an upper bound for the height of the reduction tree of t as

[2] In [**8**], the 0-th component introduces another ω-power, whereas our modification (essentially) collapses the component 1.

well as the *value* of t itself; see the proof of Corollary 4.2. Clearly, it is crucial that during a reduction of t (by means of the ξ-rule) we obtain a strictly descending chain of ordinal terms below ω in the 0-th component of the respectively assigned ordinal vectors, while the values of the terms in the reduction sequence remain constant. The information on the derivation length of t is therefore needed for the assignment to the term R^t, and comes genuinely as a natural number which in turn is provided by the collapse in the 0-th component of the vector assigned to t. In summary, collapsing plays the essential role in the treatment of unrestricted recursion, and a more modular procedure, where collapsing is applied in a separate step after the assignment of ordinal terms would be at the expense of sharp upper bounds (cf. [**8**] and [**14**]).

We develop an autonomous theory of ordinal terms and vectors and give an interpretation of closed ordinal terms as ordinals below ε_0. Due to the presence of variables in $\mathfrak{T}'$ we introduce ordinal variables and a notion of comparison on the ordinal terms containing variables. We adopt most of the terminology and conventions introduced in Section 2 of [**8**]; see in particular the Introduction to Section 2 there. However, knowledge of [**8**] is not a prerequisite. The main new ingredient in this paper is the concept of *norm* of an ordinal term and Weiermann's collapsing function ψ which is defined using norms of ordinal notations; see below.

2.1 Ordinal terms

The basic expressions of our assignment are ordinal terms as introduced syntactically in the following definition. The interpretation of ordinal terms is explained in the remainder of this subsection, together with some facts from ordinal theory. First of all, let us adopt the following convention regarding ordinal variables.

Convention Let $\mathrm{ot}_{\mathcal{V}}$ be a set of fresh variables, called *ordinal variables.* We assume the existence of a function mapping each typed variable X^σ from $\mathcal{V}$ to a sequence $(x_0, \ldots, x_{\mathrm{lv}(\sigma)})$ of pairwise distinct ordinal variables, in such a way that each ordinal variable belongs to exactly one such sequence. We will sometimes explicitly indicate the type by writing, e.g., x_i^σ for x_i in the above setting.

Definition 2.1 The set ot of ordinal terms is defined inductively as follows:

- $\mathrm{ot}_{\mathcal{V}} \subseteq \mathrm{ot}$.
- $0, 1, \omega \in \mathrm{ot}$.
- If $f, g \in \mathrm{ot}$, then $f + g$, $2^f \cdot g$, $\psi(\omega \cdot f + g) \in \mathrm{ot}$.

We call $h \in \mathrm{ot}$ *closed* if it does not contain any variable and *x-free* if none of the variables x_i occurs in h. The notion of parse tree for ordinal terms is clear from the above inductive definition, that is, the immediate subterms of $f + g$, $2^f \cdot g$, and $\psi(\omega \cdot f + g)$ are f and g. If g in $\psi(\omega \cdot f + g)$ is itself a sum, we will sometimes drop parentheses. Also, if $h \in \mathrm{ot}$ and $n \in \mathbb{N}$, we sometimes write nh or $n \cdot h$ in order to denote the n-fold summation of h.

As mentioned above we will interpret closed ot-terms as ordinal numbers below the ordinal ε_0, the least fixed point of exponentiation to base ω and hence the proof-theoretic ordinal of Peano arithmetic, as was shown by Gentzen. The interpretation of ot-terms will make use of the natural sum and product of ordinals, exponentiation to base 2, as well as the ψ-function which was introduced in [**13**]. For the readers' convenience we are going to recall these ordinal functions, starting with the natural sum $\oplus$ and the natural product $\otimes$ of ordinals, also called Hessenberg sum and product, respectively, as well as the exponentiation to bases 2 and ω, ω denoting the least infinite ordinal. The natural sum of α and 0 agrees with ordinal addition, $\alpha \oplus 0 = 0 \oplus \alpha = \alpha$, for the natural product of α and 0 we have $\alpha \otimes 0 = 0 \otimes \alpha = 0$ in agreement with ordinal multiplication. Now let

$$\alpha = \omega^{\gamma_0} + \cdots + \omega^{\gamma_m} > \gamma_0 \geq \cdots \geq \gamma_m,\ m \geq 0,$$

and

$$\beta = \omega^{\gamma_{m+1}} + \cdots + \omega^{\gamma_n} > \gamma_{m+1} \geq \cdots \geq \gamma_n,\ n \geq m+1,$$

be the Cantor normal form representations of non-zero ordinals α, β below ε_0, where $+$ is ordinal addition and $\xi \mapsto \omega^\xi$ enumerates the non-zero ordinals which are closed under ordinal addition, also called additive principal numbers.[3] The natural sum of α and β is then defined by

$$\alpha \oplus \beta = \omega^{\gamma_{\pi(0)}} + \cdots + \omega^{\gamma_{\pi(n)}}$$

where π is a permutation of $\{0, \ldots, n\}$ such that $\gamma_{\pi(0)} \geq \cdots \geq \gamma_{\pi(n)}$. The natural product of α and β is then defined by

$$\alpha \otimes \beta = (\omega^{\gamma_0 \oplus \gamma_{m+1}} \oplus \cdots \oplus \omega^{\gamma_0 \oplus \gamma_n}) \oplus \cdots \oplus (\omega^{\gamma_m \oplus \gamma_{m+1}} \oplus \cdots \oplus \omega^{\gamma_m \oplus \gamma_n}).$$

Exponentiation to base 2 is characterized by

$$2^\alpha = \omega^{\alpha_0} \cdot 2^n$$

where $\alpha = \omega \cdot \alpha_0 + n$ and $n < \omega$ (see for example [**11**]). In other words: If α is the n-th successor of the α_0-th limit ordinal, then 2^α is 2^n times the α_0-th additive principal number.

We define the function $\psi\colon \varepsilon_0 \to \omega$, which will be used to interpret ψ-terms of ot, exactly as in [**13, 14, 15**], where the interested reader can find a detailed explanation of the collapsing mechanism and its relation to the theory of subrecursive functions and ordinal recursion. An abstract exposition of the underlying concepts can be found in [**3**], where among other results a comparison of ψ with the classical Hardy functions is given. The Hardy functions provide a fine scale that allows for the comparison of provably recursive functions of fragments of Peano Arithmetic or even the distinction in levels of the Grzegorczyk hierarchy. By directly assigning ψ-terms to terms of $\mathfrak{T}$ we can compare the "run-time" of different terms in $\mathfrak{T}$, varying in the occurrence of, say, R-functionals of various types. The assignment using ψ resolves nested occurrences of recursion in terms of $\mathfrak{T}$ into unnested recursion along corresponding ordinal lengths.

Let $\Phi\colon \omega \to \omega$ be a sufficiently fast growing number theoretic function, for example the function $x \mapsto F_5(x{+}100)$ where $F_0(x) := 2^x$ and $F_{n+1}(x) := F_n^{x+1}(x)$. There is leeway in the choice of Φ; however, it is essential that Φ is bounded by some F_k, $k < \omega$, which guarantees the primitive recursiveness of Φ. Let further the norm function $\mathrm{no}\colon \varepsilon_0 \to \omega$ be defined by $\mathrm{no}(0) := 0$ and

$$\mathrm{no}(\alpha) := n + \mathrm{no}(\alpha_1) + \cdots + \mathrm{no}(\alpha_n)$$

for $\alpha = \omega^{\alpha_1} + \cdots + \omega^{\alpha_n} > \alpha_1 \geq \cdots \geq \alpha_n$. This definition of a norm has the convenient property that $\mathrm{no}(k) = k$ for any $k < \omega$, and we have that for every $m < \omega$ the set

$$\{\beta \mid \mathrm{no}(\beta) \leq m\}$$

is *finite*. This property of the norm makes the following definition of the collapsing function ψ by recursion on ε_0 possible:

$$\psi(\alpha) := \max(\{0\} \cup \{\psi(\beta) + 1 \mid \beta < \alpha \,\&\, \mathrm{no}(\beta) \leq \Phi(\mathrm{no}(\alpha))\}).$$

[3] See [**10**] for a comprehensive introduction to the basics of ordinal arithmetic, the proof theory of Peano arithmetic and further advanced topics of proof theory.

This definition can be carried out in $\mathrm{PRA} + \mathrm{PRWO}(\varepsilon_0)$; cf. [**13, 15**]. We now state two basic propositions concerning the norm and the ψ-function (see [**15**]).

Proposition 2.2 *Let α and β be ordinals less than ε_0. Then we have*

(1) $\mathrm{no}(\alpha \oplus \beta) = \mathrm{no}(\alpha) + \mathrm{no}(\beta)$.
(2) $\mathrm{no}(\alpha) + \mathrm{no}(\beta) - 1 \leq \mathrm{no}(\alpha \otimes \beta) \leq \mathrm{no}(\alpha) \cdot \mathrm{no}(\beta)$ *if* $\alpha \neq 0 \neq \beta$.
(3) $\mathrm{no}(\alpha) \leq 2 \cdot \mathrm{no}(2^\alpha)$ *and* $\mathrm{no}(2^\alpha) \leq 2^{\mathrm{no}(\alpha)}$.

Proof. The proof is given in full detail in the appendix. □

Proposition 2.3 *Let $k < \omega$ and ordinals α, $\beta < \varepsilon_0$ be given. Then we have*

(1) $k = \psi(k)$, $k \leq \psi(\alpha + k)$, $\mathrm{no}(\alpha) \leq \psi(\alpha)$, *and* $\psi(\beta) + k \leq \psi(\beta + k)$.
(2) $\psi(\alpha) + \psi(\beta) \leq \psi(\alpha \oplus \psi(\beta)) \leq \psi(\alpha \oplus \beta)$.
(3) $\alpha < \beta \,\&\, \mathrm{no}(\alpha) \leq \Phi(\mathrm{no}(\beta)) \Rightarrow \psi(\alpha) < \psi(\beta)$.
(4) $\alpha \geq \omega \Rightarrow \Phi(\mathrm{no}(\alpha)) < \psi(\alpha)$.

Proof. The proof is given in full detail in the appendix. □

Our preparations now enable us to introduce a canonical interpretation of closed ot-terms as ordinals below ε_0. This clarifies the notion of the norm $\mathrm{no}(h)$ for closed terms $h \in \mathrm{ot}$ and how closed ot-terms can be compared.

Definition 2.4 Closed terms $h \in \mathrm{ot}$ are interpreted canonically, however, $+$ is interpreted by $\oplus$ and $\cdot$ is interpreted by $\otimes$. ◇

In order to clarify the above definition, consider the example of the closed ot-term

$$\psi\left(\omega \cdot \left(2^{\omega+1} \cdot (\omega + 0)\right) + 2^{2^{\omega+1}\cdot\omega} \cdot 1\right),$$

which is interpreted by the ordinal $\psi(\omega \otimes (2^{\omega \oplus 1} \otimes (\omega \oplus 0)) \oplus 2^{2^{\omega\oplus 1}\otimes\omega} \otimes 1) = \psi(\omega^{\omega\cdot 2} + \omega^3 \cdot 2)$.

Convention When working with (closed) ot-terms we will always assume their interpretation by Definition 2.4 and compare them accordingly. Therefore, instead of using the symbols $\oplus$ and $\otimes$ we are going to simply use the ordinary symbols $+$ and $\cdot$ in order to refer to the natural sum and product, respectively.

2.2 Ordinal vectors

As in [**8**], a *vector of level* n is an $(n+1)$-tuple $\vec{h} = \langle h_0, \ldots, h_n \rangle$ where the h_i are ordinal terms, hence $\vec{h} \in \mathcal{O}^{<\omega}$. As there is no danger of ambiguity we write $\mathrm{lv}(\vec{h}) = n$. For simplicity, we define h_i to be 0 if $i > \mathrm{lv}(\vec{h})$. Thus, the sum $\vec{h} = \vec{f} + \vec{g}$ of vectors $\vec{f}, \vec{g}$ is a vector of level $\max\{\mathrm{lv}(\vec{f}), \mathrm{lv}(\vec{g})\}$ where $h_i = f_i + g_i$.

The comparison of arbitrary ot-terms and their norms poses the problem of how to interpret or substitute variables. We can consider ot-terms and their norms as functions in the variables occurring in them and then define the comparison relation via pointwise domination of functions. These functions are viewed as intensional objects and are given by the buildup of the corresponding ot-term.

Let X be a variable of type σ. The *variable vector* associated with X is the vector

$$\vec{x} := \langle x_0, \ldots, x_{\mathrm{lv}(\sigma)} \rangle$$

where $(x_0, \ldots, x_{\mathrm{lv}(\sigma)})$ is the sequence of ordinal variables associated with X according to the convention at the beginning of the previous subsection. We will sometimes explicitly indicate the type by writing, e.g., $\vec{x}^\sigma$ for $\vec{x}$ or x_i^σ for a component x_i of the vector $\vec{x}$

corresponding to X^σ. By convention we write $\vec{x}$ for the variable vector associated with X; similarly, we write $\vec{y}$ for the variable vector associated with Y, etc.

We will later introduce operations $\Box$ and $\delta^{\vec{x}}$ (where $\vec{x}$ ranges over variable vectors) on vectors that will be applied in our assignment in order to handle application and abstraction, respectively. Suitable domains for $\Box$ and $\delta^{\vec{x}}$, namely classes C and $\mathcal{C}^{\vec{x}}$, respectively, will be defined towards the end of this subsection.

Substitution of variable vectors is defined as follows. For a variable vector $\vec{x}$ of level n and a vector $\vec{a}$ of ordinal terms of the same level we define the substitution $\{\vec{x} := \vec{a}\}$ as the replacement of x_i by a_i for each $i \leq n$. We write $\{\vec{x} := \vec{1}\}$ for the replacement of x_i by 1 for $i \leq n$, etc.

The question arises over which domain the substitutions of variable vectors should vary. The variable vectors involved are those, of which at least one component occurs in one of the ot-terms that are compared or whose norms are compared. We are going to introduce the notion of bounded norm and the class $\mathfrak{B}$ of vectors characterizing the restrictions of $\mathcal{C}^{\vec{x}}$ and C to closed vectors, i.e., vectors whose components are closed ordinal terms. The restriction of $\mathfrak{B}$ to vectors of bounded norm will then serve as the domain of substitutions considered for the comparison relation. The operators $\delta^{\vec{x}}$ will be defined via vectors from $\mathcal{C}^{\vec{x}}$ which in general are not of bounded norm, which is the reason for not integrating this condition into the definitions of $\mathfrak{B}, \mathcal{C}^{\vec{x}}, C$.

Definition 2.5 A closed vector $\vec{f} \in \mathcal{O}^{<\omega}$ is of *bounded norm* if

$$\text{no}(f_i) \leq \text{no}(f_0)$$

for every i.

Definition 2.6 We define sets $\mathcal{B}_i \subseteq \text{ot}$ for $i < \omega$ by simultaneous induction:

- $1 \in \mathcal{B}_i$ for all i.
- $\omega \in \mathcal{B}_i$ for $i \geq 1$.
- If $f, g \in \mathcal{B}_i$, then $f + g \in \mathcal{B}_i$ for all i.
- If $f \in \mathcal{B}_{i+1}$ and $g \in \mathcal{B}_i$, then $2^f \cdot g \in \mathcal{B}_i$ for $i \geq 1$.
- If $f \in \mathcal{B}_1$ and $g \in \mathcal{B}_0$ with[4] $\text{no}(f) \leq F_2(g)$, then $\psi(\omega \cdot f + g) \in \mathcal{B}_0$.
- If $h \in \mathcal{B}_0$, then $h \in \mathcal{B}_i$ for $i \geq 1$.

The class $\mathfrak{B} \subseteq \mathcal{O}^{<\omega}$ consists of all vectors $\vec{h}$ such that $h_i \in \mathcal{B}_i$ for all $i \leq \text{lv}(\vec{h})$. ◇

Notice that 0 does not occur in the parse tree of any term in any $\mathcal{B}_i$. Terms $h \in \mathcal{B}_0$ cannot be of a shape $2^f \cdot g$, and they satisfy $h < \omega$, which plays a crucial role in this paper as the 0-th component of a vector assigned to a term in $\mathfrak{T}'$ is intended to yield an upper bound on the height of the term's reduction tree.

Definition 2.7 Let $f, g \in \text{ot}$. The relation $f \prec g$ holds if and only if $\chi(f) < \chi(g)$ for all substitutions χ satisfying the following conditions:

(1) The domain of χ is the set of elements of all variable vectors $\vec{x}$ such that at least one element of $\vec{x}$ occurs in f or in g.

(2) For each such $\vec{x}$ the vector $\chi(\vec{x}) := \langle \chi(x_0), \ldots, \chi(x_n) \rangle$, where $n = \text{lv}(\vec{x})$, is an element of $\mathfrak{B}$ of bounded norm.

The relation $\preceq$ on ot-terms is defined similarly, as well as (extensional) equality $=$.

The comparison of $\text{no}(f)$ with $\text{no}(g)$ or some term $h \in \mathcal{B}_0$ is defined accordingly, using the same symbols $\prec$, $\preceq$, and $=$.

[4] This condition will be useful later; all ψ-terms involved in our assignment will satisfy this condition.

We are going to use the relations $\prec$, $\preceq$, and $=$ also when comparing expressions containing the functions no and F_i, $i < \omega$. Clearly, for expressions of the form $F_i(f)$ to make sense we must have $f \prec \omega$. $\diamond$

Notice that this definition implies that

$$\mathrm{no}(h) = h \text{ for every } h \in \mathrm{ot} \text{ such that } h \prec \omega,$$

$$0 \prec \mathrm{no}(x_i) \preceq x_0 \prec \omega, \quad \text{and} \quad h \prec \varepsilon_0 \text{ for every } h \in \mathrm{ot}.$$

Propositions 2.2 and 2.3 now generalize to ot-terms and their norms using the above generalized comparison relations.

Definition 2.8 Let $\vec{f}, \vec{g} \in \mathcal{O}^{<\omega}$. We define

$$\vec{f} \prec \vec{g} \ :\Longleftrightarrow\ f_0 \prec g_0 \ \& \ \forall i > 0 \ f_i \preceq g_i$$

and

$$\vec{f} \preceq \vec{g} \ :\Longleftrightarrow\ \forall i \ f_i \preceq g_i.$$

Equality is componentwise equality in the sense of the previous definition.

Thus, $\vec{f}$ is of bounded norm if $\mathrm{no}(f_i) \preceq \mathrm{no}(f_0)$ for every i. $\diamond$

Note that all comparison relations defined in this subsection are transitive, but in general not total. We conjecture that there is a way to make the comparison relations introduced here effective; however, as we do not need such effectiveness in order to achieve our results, we stay with the above elegant comparison notion.

We now adapt the class C, introduced in [**8**] as domain of the operators δ^r. Our version is variable-specific, depending on the abstraction variable, so we will introduce classes $\mathcal{C}^{\vec{x}}$ serving as domains for $\delta^{\vec{x}}$. The classes C_i and C defined here comprise the union of all $\mathcal{C}_i^{\vec{x}}$ and $\mathcal{C}^{\vec{x}}$, respectively, and will become the general domain of ordinal vectors used in this article. We will make use of C-vectors which are not of bounded norm, namely when defining $\delta^{\vec{x}}$ in terms of partial operators $\delta_i^{\vec{x}}$. However, the vectors we are going to assign to terms of $\mathfrak{T}'$ will always be C-vectors of bounded norm.

Definition 2.9 For every $\vec{x}^\sigma \in \mathcal{O}^{<\omega}$, corresponding to some variable $X^\sigma \in \mathcal{V}$, we define sets $\mathcal{C}_i^{\vec{x}} \subseteq \mathrm{ot}$ for $i < \omega$ by simultaneous induction:

- $1 \in \mathcal{C}_i^{\vec{x}}$ for all i.
- $\omega \in \mathcal{C}_i^{\vec{x}}$ for $i \geq 1$.
- $y_i^\rho \in \mathcal{C}_i^{\vec{x}}$ for $i \leq \mathrm{lv}(\rho)$ where $Y^\rho \in \mathcal{V}$.
- If $f, g \in \mathcal{C}_i^{\vec{x}}$, then $f + g \in \mathcal{C}_i^{\vec{x}}$ for all i.
- If $f \in \mathcal{C}_{i+1}^{\vec{x}}$ and $g \in \mathcal{C}_i^{\vec{x}}$, then $2^f \cdot g \in \mathcal{C}_i^{\vec{x}}$ for $i \geq 1$.
- If $f \in \mathcal{C}_1^{\vec{x}}$ and $g \in \mathcal{C}_0^{\vec{x}}$ with $\mathrm{no}(f) \preceq F_2(g)$, then $\psi(\omega \cdot f + g) \in \mathcal{C}_0^{\vec{x}}$.
- If $h \in \mathcal{C}_0^{\vec{x}}$ is x-free, then $h \in \mathcal{C}_i^{\vec{x}}$ for $i \geq 1$.

The class $\mathcal{C}^{\vec{x}} \subseteq \mathcal{O}^{<\omega}$ is defined to consist of all $\vec{h}$ such that $h_i \in \mathcal{C}_i^{\vec{x}}$ for all $i \leq \mathrm{lv}(\vec{h})$.

Classes C_i and C are defined in the same way as $\mathcal{C}_i^{\vec{x}}$ and $\mathcal{C}^{\vec{x}}$ with the only difference that the condition of being x-free in the last clause defining the classes $\mathcal{C}_i^{\vec{x}}$ is dropped. $\diamond$

It is easy to see that the sets of closed $\mathcal{C}^{\vec{x}}$-vectors, closed C-vectors, and $\mathfrak{B}$-vectors coincide. Notice that if $h \in \mathcal{C}_i^{\vec{x}}$, then it does not contain any variable x_j such that $j < i$, cf. Lemma 2.7 of [**8**]. Notice further that the class $\mathcal{C}^{\vec{x}}$ is closed under substitution with x-free $\mathcal{C}^{\vec{x}}$-vectors, cf. Lemma 2.9 of [**8**]. We obviously have $\mathcal{C}_i^{\vec{x}} \subseteq C_i$ and $\mathcal{C}^{\vec{x}} \subseteq C$, and C is closed under substitution with C-vectors.

2.3 The operator □

We now define the operator □ and show its basic properties. The definition originates from Howard's [**8**] and was used by Schütte (see [**11**]) for an analysis of $\mathrm{GT}^{\mathcal{L}}$ which in turn later served as starting point for Weiermann's [**15**] with a refinement for vector level 0 that enabled a derivation lengths classification of Gödel's T in the combinatory logic variant. The modification of the 0-th level has two important effects: firstly, it saves one ω-power, and secondly, by using the collapsing function ψ we obtain an assignment of natural numbers to terms in $\mathfrak{T}'$ instead of ordinal terms below ε_0.

Definition 2.10 Let $\vec{f}, \vec{g} \in \mathcal{O}^{<\omega}$ be such that $m := \mathrm{lv}(\vec{f}) > \mathrm{lv}(\vec{g}) =: n$. We define[5]

$$(\vec{f} \,\square\, \vec{g})_i := \begin{cases} \psi(\omega \cdot (\vec{f} \,\square\, \vec{g})_1 + f_0 + g_0 + n) & \text{if } i = 0, \\ 2^{(\vec{f} \,\square\, \vec{g})_{i+1}} \cdot (f_i + g_i) & \text{if } 1 \le i \le n, \\ f_i & \text{if } n < i \le m, \end{cases}$$

to obtain the vector $\vec{f} \,\square\, \vec{g}$ of level m. For $i \le m$ we define $\vec{f} \,\square\, \vec{g} \restriction_i$ to be the vector of level i whose components are $(\vec{f} \,\square\, \vec{g})_j$ for $j = 0, \ldots, i$. ◇

Lemma 2.11 *Let $\vec{f}, \vec{g} \in \mathcal{O}^{<\omega}$ be such that $m := \mathrm{lv}(\vec{f}) > \mathrm{lv}(\vec{g}) =: n$ and let $k \in [1, m]$. Suppose expressions $f, g \prec \omega$ satisfy $\mathrm{no}(f_i) \preceq f$ for $k \le i \le m$ and $\mathrm{no}(g_i) \preceq g$ for $k \le i \le n$ (in the sense explained at the end of Definition 2.7). Then we have*

$$\mathrm{no}\left((\vec{f} \,\square\, \vec{g})_i\right) \preceq F_2(f + g + n)$$

for $k \le i \le m$.

Proof. The proof is given in the appendix. □

By the above lemma it follows that $\mathfrak{B}$ is closed under □, as well as are $\mathcal{C}^{\vec{x}}$ and C, cf. Lemma 2.8 of [**8**]. If $f_1 + g_1 \succ 0$ in the case $\mathrm{lv}(\vec{g}) > 0$, then, due to Proposition 2.3, part (4), $\vec{f} \,\square\, \vec{g}$ is of bounded norm. We therefore have the following:

Corollary 2.12 *The restriction of □ to C-vectors maintains bounded norm.* □

We now provide several lemmas which establish crucial properties of □ and are adapted from [**8**], extended to work for component 0.

Lemma 2.13 *Let $\vec{a}, \vec{b} \in C$ such that $m := \mathrm{lv}(\vec{a}) > \mathrm{lv}(\vec{b})$. Then we have*

$$\vec{a} + \vec{b} \preceq \vec{a} \,\square\, \vec{b}.$$

Proof. By induction on $m \dot{-} i$ it is easily shown that $a_i + b_i \preceq (\vec{a} \,\square\, \vec{b})_i$. □

Lemma 2.14 *Let $\vec{a}, \vec{b}, \vec{c} \in C$.*

(1) *Suppose $m := \mathrm{lv}(\vec{a}) = \mathrm{lv}(\vec{b}) > \mathrm{lv}(\vec{c}) =: n$ and let $i \in [1, m]$.*
If $a_j \preceq b_j$ for $i \le j \le m$, then

$$(\vec{a} \,\square\, \vec{c})_i \preceq (\vec{b} \,\square\, \vec{c})_i,$$

where "$\prec$" holds if additionally $a_k \prec b_k$ for some $k \in [i, n+1]$.

[5] Compared to [**8**] we have chosen an asymmetric (non-commutative) definition in order to more directly fit the intended application of □ and facilitate the syntactic fit with our version of C, e.g. avoiding occurrences of 0 in the parse trees of components of $\vec{f} \,\square\, \vec{g}$ for $\vec{f}, \vec{g} \in C$.

If the vectors $\vec{a}$, $\vec{b}$, and $\vec{c}$ are of bounded norm and $\vec{a} \preceq \vec{b}$, then we have

$$\vec{a} \,\Box\, \vec{c} \preceq \vec{b} \,\Box\, \vec{c},$$

where "$\prec$" holds if $\vec{a} \prec \vec{b}$.

(2) *Suppose* $\mathrm{lv}(\vec{a}) > \mathrm{lv}(\vec{b}), \mathrm{lv}(\vec{c}) =: n$ *and let* $i \in [1, n]$.
If $b_j \preceq c_j$ for $i \leq j \leq n$, then

$$(\vec{a} \,\Box\, \vec{b})_i \preceq (\vec{a} \,\Box\, \vec{c})_i,$$

where "$\prec$" holds if additionally $b_k \prec c_k$ for some $k \in [i, n]$.
If the vectors $\vec{a}$, $\vec{b}$, and $\vec{c}$ are of bounded norm and $\vec{b} \preceq \vec{c}$, then we have

$$\vec{a} \,\Box\, \vec{b} \preceq \vec{a} \,\Box\, \vec{c},$$

where "$\prec$" holds if $\vec{b} \prec \vec{c}$.

Proof. The proof of part 2.14 is by straightforward induction on $m \dot{-} i$ for the claims concerning components i, $1 \leq i \leq m$. The claim for component 0 then follows by straightforward application of Proposition 2.3, part (3), and Lemma 2.11, whose assumptions are satisfied since we have assumed bounded norms. The proof of part 2.14 is analogous. □

Lemma 2.15 *Let $\vec{a}, \vec{b}, \vec{c}, \vec{d} \in C$ be such that $m := \mathrm{lv}(\vec{a}) = \mathrm{lv}(\vec{b}) = \mathrm{lv}(\vec{c}) > \mathrm{lv}(\vec{d}) =: n$.*

(1) *If $a_i + b_i \preceq c_i$ for $1 \leq i \leq n+1$, then*

$$(\vec{a} \,\Box\, \vec{d})_i + (\vec{b} \,\Box\, \vec{d})_i \prec (\vec{c} \,\Box\, \vec{d})_i$$

for $1 \leq i \leq n$.

(2) *If* $\mathrm{no}(a_i) \preceq \mathrm{no}(b_i)$ *for $1 \leq i \leq m$, then*

$$\mathrm{no}((\vec{a} \,\Box\, \vec{d})_i) \preceq F_3 \left(\mathrm{no}((\vec{b} \,\Box\, \vec{d})_i) + n \right)$$

for $1 \leq i \leq m$.

(3) *If $a_i + b_i \prec c_i$ and* $\mathrm{no}(a_i), \mathrm{no}(b_i) \preceq \mathrm{no}(c_i)$ *for $i \leq m$, then*

$$(\vec{a} \,\Box\, \vec{d})_i + (\vec{b} \,\Box\, \vec{d})_i \prec (\vec{c} \,\Box\, \vec{d})_i$$

for $i \leq m$.

Proof. For the detailed proof the reader is referred to the appendix. □

The next lemma is an adaptation of the crucial Lemma 2.6 of [**8**]. It is the key to the treatment of the combinatorial complexity of the combinator S, cf. [**11, 15**], in part of the recursor R, and, when combined with the operator δ, of β-reduction. Regarding the latter property, which is essential in Howard's approach, notice the correspondence of the factor 2 occurring in Definition 2.18 in the case of $\delta_i^{\vec{x}} h$ where $h \equiv 2^f \cdot g$ is not x-free (this factor is in fact only necessary for the highest vector component), with the factor 2 in the assumption $2a_{n+1} + b_{n+1} \prec c_{n+1}$ of the following lemma. We regard this correspondence as crucial in order to understand why the operators $\Box$ and $\delta^{\vec{x}}$ model β-reduction. Consider, as an instructive example, β-reduction of terms of the form $(\lambda X.AB)D$.

Lemma 2.16 *Let $\vec{a}, \vec{b}, \vec{c}, \vec{d} \in C$ and $n \in \mathbb{N}$ be such that $\mathrm{lv}(\vec{a}) = \mathrm{lv}(\vec{b}) = \mathrm{lv}(\vec{c}) = n+1$ and $\mathrm{lv}(\vec{d}) = n$. If $a_i + b_i \prec c_i$ for $1 \leq i \leq n$ and $2a_{n+1} + b_{n+1} \prec c_{n+1}$, then setting*

$$\vec{e} := (\vec{a} \,\Box\, \vec{d}) \,\Box\, (\vec{b} \,\Box\, \vec{d}\restriction_n)$$

we have

$$2e_i \prec (\vec{c} \,\square\, \vec{d})_i$$

for $1 \le i \le n+1$.

Proof. See the appendix for a proof in full detail. □

For the treatment of R-reductions we will need estimations of norms of the type stated in the following lemma.

Lemma 2.17 *Let* $\vec{a}, \vec{b}, \vec{c}, \vec{d} \in C$ *be of bounded norm and* $n \in \mathbb{N}$ *such that* $\mathrm{lv}(\vec{a}) = \mathrm{lv}(\vec{c}) = n+1 > \mathrm{lv}(\vec{b}), \mathrm{lv}(\vec{d})$. *Setting*

$$\vec{e} := (\vec{a} \,\square\, \vec{b}) \,\square\, (\vec{c} \,\square\, \vec{d}\restriction_n),$$

we have

$$\mathrm{no}(e_i) \prec F_3(a_0 + b_0 + c_0 + d_0 + n)$$

for $1 \le i \le n+1$.

Proof. The proof is given in the appendix. □

2.4 The operators $\delta^{\vec{x}}$

Here we introduce our refinement of the operators δ^r —see [**8**]— which provide the key to appropriate assignments of ordinal vectors to abstraction terms in order to allow for the treatment of β-contraction. Our modification of δ essentially concerns vector level zero, which ranges over terms for natural numbers instead of ordinals below ε_0 as in the original version. This is made possible by application of the collapsing function ψ. Our refinement is formulated on the basis of the $\mathcal{C}^{\vec{x}}$-classes introduced earlier in order to make the treatment of general R-reductions possible.

Definition 2.18 In order to define $\delta^{\vec{x}} \colon \mathcal{C}^{\vec{x}} \to \mathcal{C}^{\vec{x}}$, let $\vec{h} \in \mathcal{C}^{\vec{x}}$ be of level $m := \mathrm{lv}(\vec{h})$ and set $n := \mathrm{lv}(\vec{x}) + 1$. Then $\delta^{\vec{x}}\vec{h}$ is a vector of level $l := \max\{n, m\}$, defined componentwise by

$$(\,\delta^{\vec{x}}\vec{h})_j := \begin{cases} \mathrm{S}^{\vec{x}}(\vec{h}) \cdot (\,\delta^{\vec{x}}_0 h_0)_0 & \text{if } \ j = 0, \\ \sum_{i=0}^{m}(\,\delta^{\vec{x}}_i h_i)_j & \text{if } \ 0 < j \le n, \\ h_j & \text{if } \ n < j \le l, \end{cases}$$

where $\mathrm{S}^{\vec{x}}$ will be defined below and the $\delta^{\vec{x}}_i \colon \mathcal{C}^{\vec{x}}_i \to \mathcal{C}^{\vec{x}}$ are defined recursively as follows. Let $h \in \mathcal{C}^{\vec{x}}_i$. Then $\delta^{\vec{x}}_i h$ is a $\mathcal{C}^{\vec{x}}$-vector of level n, defined componentwise as follows. If h is x-free, we set

$$(\,\delta^{\vec{x}}_i h)_j := \begin{cases} 1 & \text{if } \ i \ne j \le n, \\ h+1 & \text{if } \ i = j \le n. \end{cases}$$

The following cases apply if h is not x-free.

- $h \equiv x^{\sigma}_i$:

  $\delta^{\vec{x}}_i h := \vec{1}$.

- $h \equiv f + g$ where $f, g \in \mathcal{C}^{\vec{x}}_i$:

  $\delta^{\vec{x}}_i h := \,\delta^{\vec{x}}_i f + \,\delta^{\vec{x}}_i g + \vec{1}$.

- $h \equiv 2^f \cdot g$ where $f \in \mathcal{C}^{\vec{x}}_{i+1}, g \in \mathcal{C}^{\vec{x}}_i$, and $i > 0$:

  $\delta^{\vec{x}}_i h := 2\, \delta^{\vec{x}}_{i+1} f + \delta^{\vec{x}}_i g + \vec{1}$.

- $h \equiv \psi(\omega \cdot f + g)$ where $f \in \mathcal{C}^{\vec{x}}_1, g \in \mathcal{C}^{\vec{x}}_0$ and $i = 0$:

  $$(\delta^{\vec{x}}_i h)_j := \begin{cases} \psi(\omega \cdot f\{\vec{x} := \vec{1}\} + (\delta^{\vec{x}}_0 g)_0) & \text{if } j = 0, \\ (\delta^{\vec{x}}_1 f)_j + (\delta^{\vec{x}}_0 g)_j & \text{if } 1 \le j \le n. \end{cases}$$

The norm controlling factor $\mathrm{S}^{\vec{x}}(\vec{h}) \in \mathbb{N}$ is defined by

$$\mathrm{S}^{\vec{x}}(\vec{h}) := 2^n \cdot \sum_{i=0}^{m} \mathrm{sz}^{\vec{x}}(h_i),$$

where the auxiliary $\mathrm{sz}^{\vec{x}}(h)$ for $h \in \mathrm{ot}$ is defined by

- $\mathrm{sz}^{\vec{x}}(h) := 1$ if h is x-free or $h \equiv x_i$ for some i;
- $\mathrm{sz}^{\vec{x}}(h) := \mathrm{sz}^{\vec{x}}(f) + \mathrm{sz}^{\vec{x}}(g) + 1$ if h is not x-free and either of a form $f + g$ or $\psi(\omega \cdot f + g)$;
- $\mathrm{sz}^{\vec{x}}(h) := 2\mathrm{sz}^{\vec{x}}(f) + \mathrm{sz}^{\vec{x}}(g) + 1$ if h is not x-free and of a form $2^f \cdot g$. ◇

Notice that $\delta^{\vec{x}}\vec{h}$ does not contain any component of $\vec{x}$. In order to see that the above definition is sound, we have to verify that the vectors $\delta^{\vec{x}}_i h$ are indeed $\mathcal{C}^{\vec{x}}$-vectors. We have the following:

Lemma 2.19 *For $h \in \mathcal{C}^{\vec{x}}_0$ we have*

$$h\{\vec{x} := \vec{1}\} \prec (\delta^{\vec{x}}_0 h)_0.$$

Proof. The proof is by induction on the buildup of h. The interesting case is where h is of the form $\psi(\omega \cdot f + g)$ and not x-free. We then use the induction hypothesis for g, obtaining

$$\begin{aligned} h\{\vec{x} := \vec{1}\} &\equiv \psi(\omega \cdot f\{\vec{x} := \vec{1}\} + g\{\vec{x} := \vec{1}\}) \\ &\prec \psi(\omega \cdot f\{\vec{x} := \vec{1}\} + (\delta^{\vec{x}}_0 g)_0) \\ &\equiv (\delta^{\vec{x}}_0 h)_0. \end{aligned}$$

□

The above lemma shows that for terms $h \equiv \psi(\omega \cdot f + g) \in \mathcal{C}^{\vec{x}}_0$ we have

$$\mathrm{no}(f\{\vec{x} := \vec{1}\}) \preceq F_2((\delta^{\vec{x}}_0 g)_0),$$

using that $\mathrm{no}(f\{\vec{x} := \vec{1}\}) \preceq F_2(g\{\vec{x} := \vec{1}\})$. It is then easy to verify that $\delta^{\vec{x}}_i h \in \mathcal{C}^{\vec{x}}$ for $h \in \mathcal{C}^{\vec{x}}_i$ and hence $\delta^{\vec{x}}\vec{h} \in \mathcal{C}^{\vec{x}}$ for $\vec{h} \in \mathcal{C}^{\vec{x}}$.

We call a substitution $\{\vec{y} := \vec{g}\}$ an *x-free substitution* if $\vec{y} \not\equiv \vec{x}$ and $\vec{g}$ is x-free. This notion facilitates an elegant statement of the next lemma, which corresponds to Lemma 2.10 and Corollary of [**8**].

Lemma 2.20 *The operator $\delta^{\vec{x}}$ commutes with x-free substitution: for $\vec{f}, \vec{h} \in \mathcal{C}^{\vec{x}}$, $\vec{f}$ x-free, and $\vec{y} \not\equiv \vec{x}$, we have*

$$(\delta^{\vec{x}}\vec{h})\{\vec{y} := \vec{f}\} = \delta^{\vec{x}}(\vec{h}\{\vec{y} := \vec{f}\}).$$

Proof. Notice that $\mathrm{sz}^{\vec{x}}$ and hence $\mathrm{S}^{\vec{x}}$ are invariant under x-free substitution. It is then straightforward to verify the commutativity of the partial operators $\delta^{\vec{x}}_i$ with x-free substitution and finally conclude the lemma. □

Lemma 2.21 *For any variable vector $\vec{x}$ the operator $\delta^{\vec{x}}$ preserves bounded norm: for every $\vec{h} \in \mathcal{C}^{\vec{x}}$ of bounded norm, $\delta^{\vec{x}}\vec{h}$ is of bounded norm.*

Proof. This is part 5.2 of Lemma 5.2, which is stated and proved in the appendix. □

We conclude this section establishing the interplay of the operators □ and $\delta^{\vec{x}}$, corresponding to Lemma 2.11 of [**8**] and its Corollary.

Lemma 2.22 *Let $\vec{h} \in \mathcal{C}^{\vec{x}}$. We have*

$$\vec{h} \prec \delta^{\vec{x}}\vec{h} \,\Box\, \vec{x}.$$

Proof. The proof is given in the appendix. □

3 Assignment of ordinal vectors to terms

3.1 Assignment derivations

We are now prepared to assign ordinal vectors to terms of $\mathfrak{T}'$. Recall, for illustrative reasons, Definition 1.6 and the notion of derivation along an inductive definition. Definition 1.9 provides canonical derivations of $\mathrm{w}(F, n_F)$ for every $F \in \mathfrak{T}'$ along a given reduction sequence. With the following inductive definition, which is independent of Subsection 1.6, we refine the notion of derivation towards *assignment derivation*, carrying the ordinal vectors assigned to terms as labels.

Definition 3.1 We define *assignment derivations* inductively for terms $A^\sigma \in \mathfrak{T}'$, which assign vectors $[\![A]\!]$ of level $\mathrm{lv}(\sigma)$ to A. The notation $[\![A]\!]$ is therefore only determined uniquely in the context of a fixed assignment derivation.

Assignment to prime terms of $\mathfrak{T}'$. If A is a variable or constant, then it has a unique assignment $[\![A]\!]$ as defined below, and its assignment derivation is a single-node tree which is labeled with $(A, [\![A]\!])$.

$[\![X^\sigma]\!] := \vec{x}^\sigma$.

$[\![0]\!] := \langle 1 \rangle$.

$[\![\mathsf{S}]\!] := \langle 1, 1 \rangle$.

$[\![\mathsf{D}_\tau]\!] := \langle 1, \dots, 1 \rangle$ of level $\mathrm{lv}(\tau) + 1$.

$[\![\mathsf{R}_\tau]\!] := \langle 2, 1, \dots, 1, \omega \rangle$ of level $\mathrm{lv}(\tau) + 2$.

Terms formed by application. For assignment derivations $\mathcal{R}$ of $(B^{\sigma\tau}, [\![B]\!])$ and $\mathcal{S}$ of $(C^\sigma, [\![C]\!])$, the tree with direct subtrees $\mathcal{R}$ and $\mathcal{S}$ is a derivation of $(BC, [\![BC]\!])$, where $[\![BC]\!]$ is defined by

$$[\![BC]\!] := [\![B]\!] \,\Box\, [\![C]\!] \restriction_{\mathrm{lv}(\tau)} .$$

For an assignment derivation $\mathcal{R}$ of $(t, [\![t]\!])$, the tree with direct subtree $\mathcal{R}$ is an assignment derivation of $(\mathsf{R}^t, [\![\mathsf{R}^t]\!])$ where $[\![\mathsf{R}^t_\tau]\!]$ of level $\mathrm{lv}(\tau) + 2$ is defined by

$$[\![\mathsf{R}^t_\tau]\!] := \langle [\![t]\!]_0, 1, \dots, 1, [\![t]\!]_0 \rangle.$$

Terms formed by abstraction. For assignment derivations $\mathcal{R}_i$ of $(G_i, [\![G_i]\!])$, $i \le m$, where $G_0 \rhd_\alpha \cdots \rhd_\alpha G_m =: G$ such that X does not occur in any subterm of G_0 of a form R^t

and $[\![G_0]\!] \succ \cdots \succ [\![G_m]\!]$, the tree with direct subtrees $\mathcal{R}_0, \ldots, \mathcal{R}_m$ in this order is an assignment derivation of $(\lambda X.G, [\![\lambda X.G]\!])$, the label of its root, where

$$[\![\lambda X.G]\!] := \delta^{\vec{x}}[\![G_0]\!] + [\![G_m]\!]\{\vec{x} := \vec{1}\}.$$

Whenever a particular assignment derivation is clear from the context of argumentation, we will use the notation $[\![\cdot]\!]$ as if it were an operator returning a unique ordinal vector.

For a term $A \in \mathfrak{T}$ we define the *canonical assignment for A* by choosing for every subterm of a form $\lambda X.G$ the assignment $[\![\lambda X.G]\!] := \delta^{\vec{x}}[\![G]\!] + [\![G]\!]\{\vec{x} := \vec{1}\}$. This results in a unique assignment $[\![A]\!]$ to the term A. We call the vector $\vec{a}$ resulting from $[\![A]\!]$ by replacing every variable by 1 the *closed canonical assignment for A*. ◇

It is easy to verify that all vectors assigned to terms are C-vectors of bounded norm, cf. Corollary 2.12 and Lemma 2.21. Notice also that in case of an assignment $[\![\lambda X.G]\!] := \delta^{\vec{x}}[\![G_0]\!] + [\![G_m]\!]\{\vec{x} := \vec{1}\}$ the vector $[\![G_0]\!]$ even is a $\mathcal{C}^{\vec{x}}$-vector, as is required for the application of the operator $\delta^{\vec{x}}$. This latter property $[\![G_0]\!] \in \mathcal{C}^{\vec{x}}$ is guaranteed by the fact that the variable X corresponding to $\vec{x}$ does not occur in any subterm of the form R^t of G_0. This is crucial for the compatibility of the original treatment of β-reductions with unrestricted R-reductions and is one of the two reasons why we need this form of non-unique ordinal assignment that depends on the reduction history of terms in a given reduction sequence, as explained in greater detail in Subsection 1.6. The other reason is the same as in [**8**]: the operators $\delta^{\vec{x}}$ are in general not monotonically increasing[6] (see also [**8**], p. 456) and therefore do not allow for a direct treatment of the unrestricted ξ-rule.

Lemma 3.2 *Assignment derivations are invariant modulo α-congruence.*

Proof. Straightforward. □

The following lemma corresponds to Lemma 3.1 of [**8**]; cf. also Definition 1.8.

Lemma 3.3 *The assignment $[\![\cdot]\!]$ commutes with substitution. Suppose that $F, H \in \mathfrak{T}'$ satisfy $\mathrm{FV}(F) \cap \mathrm{BV}(H) = \emptyset$ and let Y be a variable of the same type as F. Given assignment derivations of $(H, [\![H]\!])$ and $(F, [\![F]\!])$, there is a canonical assignment derivation of $(H\{Y := F\}, [\![H]\!]\{\vec{y} := [\![F]\!]\})$, defined straightforwardly in the proof.*

Proof. The proof is by induction along the definition of an assignment derivation of $(H, [\![H]\!])$. The interesting case is where H is an abstraction term, say $\lambda X.G$, whose assignment

$$[\![H]\!] = \delta^{\vec{x}}[\![G_0]\!] + [\![G_m]\!]\{\vec{x} := \vec{1}\}$$

is based on assignments $[\![G_0]\!], \ldots, [\![G_m]\!]$ where $G_0 \rhd_\alpha \cdots \rhd_\alpha G_m = G$ such that X does not occur in any subterm of G_0 of a form R^t and $[\![G_0]\!] \succ \cdots \succ [\![G_m]\!]$.

Since the case $Y = X$ is trivial, we assume $Y \neq X$. By assumption we have $X \notin \mathrm{FV}(F)$, and according to Lemma 3.2 we may further assume without loss of generality that $\mathrm{FV}(F) \cap \mathrm{BV}(G_0) = \emptyset$. The induction hypothesis yields assignment derivations of $(G_i\{Y := F\}, [\![G_i]\!]\{\vec{y} := [\![F]\!]\})$ for $i \leq m$, and we have

$$[\![G_0]\!]\{\vec{y} := [\![F]\!]\} \succ \cdots \succ [\![G_m]\!]\{\vec{y} := [\![F]\!]\}.$$

[6] Consider for example variables x, y of type 00, variables z, u of type 0 and compute the canonical assignments to $\lambda x.((\lambda z.x(yz))u) \rhd \lambda x.(x(yu))$. Setting e.g. $y := \lambda w^0.w^0$ and $u := y(yv^0)$ we see that $\delta^{\vec{x}}$ is not even weakly monotonically increasing.

Clearly,

$$([\![G_m]\!]\{\vec{x} := \vec{1}\})\{\vec{y} := [\![F]\!]\} = ([\![G_m]\!]\{\vec{y} := [\![F]\!]\})\{\vec{x} := \vec{1}\},$$

and by Lemma 2.20 we have

$$(\,\delta^{\vec{x}}[\![G_0]\!])\{\vec{y} := [\![F]\!]\} = \delta^{\vec{x}}([\![G_0]\!]\{\vec{y} := [\![F]\!]\}).$$

We have

$$G_0\{Y := F\} \rhd_\alpha \cdots \rhd_\alpha G_m\{Y := F\},$$

and X does not occur in any subterm of $G_0\{Y := F\}$ of a form R^t. The assignment derivation of $(H\{Y := F\}, [\![H]\!]\{\vec{y} := [\![F]\!]\})$ can therefore be assembled from the assignment derivations of the $(G_i\{Y := F\}, [\![G_i]\!]\{\vec{y} := [\![F]\!]\})$. □

Definition 3.4 By recursion on $\mathrm{lh}(A)$ we define an algorithm which, given two terms $A, B \in \mathfrak{T}'$ such that $A \rhd B$ and given an assignment derivation for A, returns an assignment derivation for B.

Reductions:

- $(\boldsymbol{D_0})$, $(\boldsymbol{D_\mathsf{S}})$, $(\boldsymbol{R})$, and $(\boldsymbol{R^0})$ are trivial, proceeding in the same way as in the following case.
- $(\boldsymbol{R^\mathsf{S}})$ $\mathsf{R}^{\mathsf{S}t}FG \rhd Ft(\mathsf{R}^tFG)$ with $[\![\mathsf{R}^{\mathsf{S}t}FG]\!]$ given via assignments $[\![t]\!], [\![F]\!], [\![G]\!]$. Then
$$[\![Ft(\mathsf{R}^tFG)]\!]$$
is built up from the same assignments $[\![t]\!], [\![F]\!], [\![G]\!]$.
- $(\boldsymbol{\beta})$ $(\lambda X.G)H \rhd G\{X := H\}$ where $\mathrm{BV}(\lambda X.G) \cap \mathrm{FV}(H) = \emptyset$ with $[\![(\lambda X.G)H]\!]$ given via assignments $[\![\lambda X.G]\!], [\![H]\!]$ where the former is in turn given via assignments $[\![G_0]\!], \ldots, [\![G_m]\!]$ from the assignment derivation of $\lambda X.G$, whence $G_0 \rhd_\alpha \cdots \rhd_\alpha G_m = G$. By Lemma 3.3 we obtain an assignment
$$[\![G_m]\!]\{\vec{x} := [\![H]\!]\}$$
to the term $G\{X := H\}$ with the canonical assignment derivation.

Rules:

- $(\boldsymbol{App_r})$, $(\boldsymbol{App_l})$, and $(\boldsymbol{App_R})$ are handled in the same straightforward manner, e.g. in the case where $\mathsf{R}^s \rhd \mathsf{R}^t$ is derived from $s \rhd t$ and the assignment $[\![\mathsf{R}^s]\!]$ given via an assignment $[\![s]\!]$ to s, we let $[\![t]\!]$ be the assignment provided by the algorithm and build $[\![\mathsf{R}^t]\!]$ up from $[\![t]\!]$.
- $(\boldsymbol{\xi})$ $\lambda X.F \rhd \lambda X.G$ derived from $F \rhd G$, with $[\![\lambda X.F]\!]$ given by means of assignments $[\![F_0]\!], \ldots, [\![F_m]\!]$ from the assignment derivation of $\lambda X.F$, whence $F_0 \rhd_\alpha \cdots \rhd_\alpha F_m = F$. We then choose the assignment
$$[\![\lambda X.G]\!] := \delta^{\vec{x}}[\![F_0]\!] + [\![G]\!]\{\vec{x} := \vec{1}\},$$
where $[\![G]\!]$ is the assignment to G provided by the algorithm. ◇

The soundness of the above definition hinges on the verification that in the clause for $(\boldsymbol{\xi})$ we indeed have $[\![F_m]\!] \succ [\![G]\!]$. This is accomplished by our Main Theorem.

3.2 Main theorem

Theorem 3.5 *For $A, B \in \mathfrak{T}'$ such that $A \rhd B$ and a given assignment derivation assigning $[\![A]\!]$ to A, the assignment $[\![B]\!]$ to B that is provided by the algorithm specified in Definition* 3.4 *satisfies*

$$[\![A]\!] \succ [\![B]\!].$$

Proof. The proof is by induction on $\mathrm{lh}(A)$. We use the terminology of Definition 3.4.

$(\boldsymbol{D}_{\mathbf{0}})$, $(\boldsymbol{D}_{\mathsf{S}})$, $(\boldsymbol{R})$, and $(\boldsymbol{R}^{\mathbf{0}})$ are handled straightforwardly using Lemma 2.13.

$(\boldsymbol{R}^{\mathsf{S}})$ $\mathsf{R}^{\mathsf{S}t}FG \rhd Ft(\mathsf{R}^{t}FG)$ where $\mathsf{R} \equiv \mathsf{R}_\tau$ and $n := \mathrm{lv}(\tau)$. Suppose $[\![\mathsf{R}^{\mathsf{S}t}FG]\!]$ is given via assignments $[\![t]\!]$, $[\![F]\!]$, and $[\![G]\!]$. We introduce the following abbreviations:

$$\begin{aligned}
\vec{a} &:= [\![Ft]\!]\\
\vec{b} &:= [\![\mathsf{R}^{t}F]\!]\\
\vec{c} &:= [\![\mathsf{R}^{\mathsf{S}t}F]\!]\\
\vec{d} &:= [\![G]\!]\\
\vec{e} &:= (\vec{a} \Box \vec{d}) \Box (\vec{b} \Box \vec{d}\restriction_n)\\
\vec{f} &:= [\![F]\!]\\
\vec{t} &:= [\![t]\!].
\end{aligned}$$

Notice that we have $\vec{a} = \vec{f} \Box \vec{t}$ and

$$\begin{aligned}
[\![\mathsf{R}^{t}]\!] &= \langle t_0, 1, \ldots, 1, t_0\rangle\\
&\prec \langle \psi(\omega + t_0 + 1), 1, \ldots, 1, \psi(\omega + t_0 + 1)\rangle\\
&= [\![\mathsf{R}^{\mathsf{S}t}]\!],
\end{aligned}$$

hence $b_i = ([\![\mathsf{R}^{t}]\!] \Box \vec{f})_i \prec ([\![\mathsf{R}^{\mathsf{S}t}]\!] \Box \vec{f})_i = c_i$ for $i \leq n+1$ by Lemma 2.14, part 2.14. We further have $[\![\mathsf{R}^{t}FG]\!] = \vec{b} \Box \vec{d}\restriction_n$, and

$$[\![Ft(\mathsf{R}^{t}FG)]\!] = \vec{a} \Box (\vec{b} \Box \vec{d}\restriction_n)\restriction_n \prec \vec{e}$$

by part 2.14 of Lemma 2.14 since $\vec{a} \prec \vec{a} \Box \vec{d}$ by Lemma 2.13. We obtain

$$2e_i \prec (\vec{c} \Box \vec{d})_i$$

for $1 \leq i \leq n+1$ by an application of Lemma 2.16, whose assumptions $a_i + b_i \prec c_i$ for $1 \leq i \leq n$ and $2a_{n+1} + b_{n+1} \prec c_{n+1}$ are easily verified. As $[\![\mathsf{R}^{\mathsf{S}t}FG]\!] = \vec{c} \Box \vec{d}\restriction_n$, we obtain

$$2[\![Ft(\mathsf{R}^{t}FG)]\!]_i \prec [\![\mathsf{R}^{\mathsf{S}t}FG]\!]_i$$

for $1 \leq i \leq n$, and in the case $n > 0$ by Lemma 2.13 we thus have

$$[\![Ft(\mathsf{R}^{t}FG)]\!]_1 + f_1 + [\![\mathsf{R}^{t}FG]\!]_1 \prec [\![\mathsf{R}^{\mathsf{S}t}FG]\!]_1. \tag{3.1}$$

It remains to prove that

$$[\![Ft(\mathsf{R}^{t}FG)]\!]_0 \prec [\![\mathsf{R}^{\mathsf{S}t}FG]\!]_0. \tag{3.2}$$

We begin with the following estimation:

$$\begin{aligned} b_0 + f_0 + t_0 &= \psi(\omega \cdot b_1 + t_0 + f_0 + n + 1) + f_0 + t_0 \\ &\preceq \psi(\omega \cdot b_1 + 2t_0 + 2f_0 + n + 1) \\ &\prec \psi(\omega \cdot c_1 + \psi(\omega + t_0 + 1) + f_0 + n + 1) \\ &= c_0, \end{aligned}$$

which follows by Proposition 2.3, part (3), since $b_1 \prec c_1$ and $\mathrm{no}(b_1) \preceq F_2(t_0 + f_0 + n + 1)$ using Lemma 2.11. In the case $n = 0$ it is easy to verify (3.2). Let us therefore assume that $n > 0$. Using parts (2) and (3) of Proposition 2.3, from (3.1) we then obtain

$$\begin{aligned} [\![Ft(\mathsf{R}^t FG)]\!]_0 &= \psi(\omega \cdot [\![Ft(\mathsf{R}^t FG)]\!]_1 + a_0 + [\![\mathsf{R}^t FG]\!]_0 + n) \\ &= \psi(\omega \cdot [\![Ft(\mathsf{R}^t FG)]\!]_1 + \psi(\omega \cdot f_1 + f_0 + t_0) + \\ &\qquad \psi(\omega \cdot [\![\mathsf{R}^t FG]\!]_1 + b_0 + d_0 + n) + n) \\ &\preceq \psi(\omega \cdot ([\![Ft(\mathsf{R}^t FG)]\!]_1 + f_1 + [\![\mathsf{R}^t FG]\!]_1) + b_0 + f_0 + d_0 + t_0 + 2n) \\ &\prec \psi(\omega \cdot [\![\mathsf{R}^{\mathsf{S}t} FG]\!]_1 + c_0 + d_0 + n) \\ &= [\![\mathsf{R}^{\mathsf{S}t} FG]\!]_0, \end{aligned}$$

since using Lemma 2.17 we may estimate

$$\begin{aligned} \mathrm{no}([\![Ft(\mathsf{R}^t FG)]\!]_1 + f_1 + [\![\mathsf{R}^t FG]\!]_1) &\preceq F_3(f_0 + t_0 + b_0 + d_0 + n) + \\ &\qquad F_2(f_0 + t_0) + F_2(b_0 + d_0 + n) \\ &\prec \Phi(c_0 + d_0 + n). \end{aligned}$$

($\boldsymbol{\beta}$) $(\lambda X.G)H \rhd G\{X := H\}$. With the notations of Definition 3.4, the vector assigned to $\lambda X.G$ is $\delta^{\vec{x}}[\![G_0]\!] + [\![G]\!]\{\vec{x} := \vec{1}\}$. By Lemma 2.22 we have

$$[\![G_0]\!] \prec \delta^{\vec{x}}[\![G_0]\!] \,\Box\, \vec{x},$$

hence

$$\begin{aligned} [\![\lambda X.G]\!] \,\Box\, [\![H]\!] &\succ \delta^{\vec{x}}[\![G_0]\!] \,\Box\, [\![H]\!] \text{ by part (i) of Lemma 2.14} \\ &\succ [\![G_0]\!]\{\vec{x} := [\![H]\!]\}. \end{aligned}$$

We have $[\![G_0]\!] \succ \cdots \succ [\![G_m]\!]$; hence

$$[\![G_0]\!]\{\vec{x} := [\![H]\!]\} \succ \cdots \succ [\![G_m]\!]\{\vec{x} := [\![H]\!]\},$$

and by Lemma 3.3 the $[\![G_i]\!]\{\vec{x} := [\![H]\!]\}$ are vectors assigned to $G_i\{X := H\}$ for $i \leq m$.

($\boldsymbol{App_r}$) $FH \rhd GH$, derived from $F \rhd G$. By the induction hypothesis we have $[\![F]\!] \succ [\![G]\!]$, hence part (i) of Lemma 2.14 yields $[\![FH]\!] \succ [\![GH]\!]$.

($\boldsymbol{App_l}$) $FG \rhd FH$, derived from $G \rhd H$. From the induction hypothesis it follows that $[\![G]\!] \succ [\![H]\!]$, hence part (ii) of Lemma 2.14 yields $[\![FG]\!] \succ [\![FH]\!]$.

($\boldsymbol{App_R}$) $\mathsf{R}^s \rhd \mathsf{R}^t$, derived from $s \rhd t$. By the induction hypothesis we have $[\![s]\!] \succ [\![t]\!]$, so we immediately obtain $[\![\mathsf{R}^s]\!] \succ [\![\mathsf{R}^t]\!]$.

($\boldsymbol{\xi}$) $\lambda X.F \rhd \lambda X.G$, derived from $F \rhd G$. Then $[\![\lambda X.F]\!] \succ [\![\lambda X.G]\!]$ follows directly from the induction hypothesis, which yields $[\![F_m]\!] \succ [\![G]\!]$. □

Corollary 3.6 *Let $A \in \mathfrak{T}$ and $\vec{a}$ be its closed canonical assignment. Then $a_0 \in \mathbb{N}$ is an upper bound of the height of the reduction tree of A. We obtain strong normalization for $\lambda\beta\mathsf{R}$ and $\lambda\beta\mathsf{R}'$.*

Proof. Regarding the relationships between reduction sequences of $\mathfrak{T}$-terms in $\lambda\beta\mathsf{R}$ and $\mathfrak{T}'$-terms in $\lambda\beta\mathsf{R}'$, recall the remarks stated in Subsection 1.4. The corollary then follows from Theorem 3.5. □

3.3 Tying in with Subsection 1.6

For illustrative reasons we establish the link of our assignment with Definition 1.9.

Let $F_0 \rhd_\alpha \cdots \rhd_\alpha F_n$ be a $\lambda\beta\mathsf{R}'$-reduction sequence with $F_0 \in \mathfrak{T}$ and let p be a node in F_n with label A. Suppose that assignment derivations of $(B, [\![B]\!])$ have been specified for all terms B labeled to nodes in the parse trees of F_i for $i = 0, \ldots, n-1$ and to nodes descending from p in the parse tree of F_n. In case A is a variable or constant we are done with $(A, [\![A]\!])$, and if $A^\tau \equiv BC$, we have $(B, [\![B]\!])$ and $(C, [\![C]\!])$ for the labels B and C of the direct child nodes of p and accordingly choose $(A, [\![B]\!] \,\square\, [\![C]\!] \restriction_{\mathrm{lv}(\tau)})$. The interesting case is where $A = \lambda X.G$. We may assume that $F_0 \rhd_\alpha \cdots \rhd_\alpha F_n$ is p-nice, according to Lemma 3.2. Let $(G_0, \ldots, G_m)$, where $G_m = G$, be the associated reduction sequence with respect to $F_0 \rhd_\alpha \cdots \rhd_\alpha F_n$ and p according to Definition 1.5, so using Lemma 3.3 we obtain assignment derivations labeled with $(G_i, [\![G_i]\!])$ for $i \leq m$. If we can show that

$$[\![G_0]\!] \succ \cdots \succ [\![G_m]\!],$$

then we obtain an assignment derivation of $(A,\, \delta^{\vec{x}}[\![G_0]\!] + [\![G_m]\!]\{\vec{x} := \vec{1}\})$. We are going to show that the assignments $[\![G_1]\!], \ldots, [\![G_m]\!]$ are obtained by consecutive application of the algorithm given in Definition 3.4, starting from $[\![G_0]\!]$. Recalling our description of how the parse tree of F_{i+1} is obtained from the parse tree of F_i in 1.6.2, we see that for corresponding terms A in the parse tree of F_i and B in the parse tree of F_{i+1} such that $A \rhd B$, the assignment derivation of $[\![B]\!]$ is obtained from the assignment derivation of $[\![A]\!]$ by the algorithm given in Definition 3.4. Such terms A and B are either the working redex itself and its reduct, corresponding to a D-, R-, or β-reduction, or corresponding terms on the paths leading from the roots of the parse trees of F_i and F_{i+1} to the working redex and its reduct, respectively, corresponding to an *App*- or ξ-rule. Notice that the claimed property $[\![G_0]\!] \succ \cdots \succ [\![G_m]\!]$ for the associated reduction sequence mentioned above then follows after observing that the algorithm in Definition 3.4 commutes with substitution in the sense of Lemma 3.3 and checking the possible cases as outlined in Definition 1.1.

4 Derivation lengths classification

We are going to show that the upper bounds for reduction sequences in Gödel's GT and its fragments T_n are optimal. The set of terms $\mathcal{T}_n$ in the fragment T_n, $n \in \mathbb{N}$, is the restriction of $\mathfrak{T}$ to recursors of type level $\leq n+2$. From our remarks in Subsection 1.4 it follows that we can discuss term reductions in $\mathfrak{T}$ using our results regarding reduction sequences in $\mathfrak{T}'$ via the mutual embeddings of reduction sequences in $\mathfrak{T}$ and $\mathfrak{T}'$.

Definition 4.1 For $G \in \mathfrak{T}$ let $\mathrm{L}(G)$ denote the maximum type level of subterms of G, and let $\mathrm{R}(G)$ denote the maximum type level of recursors occurring in G. We further define

$$\mathrm{D}_{\mathrm{GT}}(m) := \max\{k \mid \exists G_1, \ldots, G_k \in \mathfrak{T}\ G_1 \rhd \cdots \rhd G_k \ \&\ \mathrm{lh}(G_1), \mathrm{L}(G_1) \le m\}$$

$$\mathrm{D}_{\mathrm{T}_n}(m) := \max\{k \mid \exists G_1, \ldots, G_k \in \mathcal{T}_n\ G_1 \rhd \cdots \rhd G_k \ \&\ \mathrm{lh}(G_1), \mathrm{L}(G_1) \le m\}.$$

We are going to use the following common notation for exponential expressions. We set $\omega_0 := 1$ and $\omega_{i+1} := \omega^{\omega_i}$, $2_0(\alpha) := \alpha$, and $2_{i+1}(\alpha) := 2^{2_i(\alpha)}$ where $\alpha < \varepsilon_0$.

Corollary 4.2 *Corollary* 3.6 *gives rise to the following derivation lengths classifications.*

(1) *The functions definable in* GT, *i.e., the provably recursive functions of* PA, *comprise the* $< \varepsilon_0$*-recursive functions. The derivation lengths function* D_{GT} *is* ε_0*-recursive.*

(2) *The functions definable in* T_n*, i.e., the provably recursive functions of* $\mathrm{I}\Sigma_{n+1}$, *comprise the* $<\omega_{n+2}$*-recursive functions. The derivation lengths function* $\mathrm{D}_{\mathrm{T}_n}$ *is* ω_{n+2}*-recursive.*

Proof. We make use of the well-known fact that PA has a functional interpretation in GT (see [**6, 11, 12**]) and the fact that the fragments $\mathrm{I}\Sigma_{n+1}$ have functional interpretations in the T_n (see [**9**]). By results from [**3**] the corollary then follows from Corollary 3.6. The detailed argumentation is given below. It is based on preparations worked out in the appendix (5.4), which we will use in the form of citations of Lemma 5.8, whose purpose it is to extract bounds on the ordinal vectors $[\![G]\!]$ assigned to terms G of $\mathfrak{T}$ which are expressed in terms of maximum type level $\mathrm{L}(G)$ and length $\mathrm{lh}(G)$ of G. Notice that we only need to consider the (unique) canonical assignment for the terms of $\mathfrak{T}$.

Let $C^0 \in \mathfrak{T}$ be closed. There is an $m \in \mathbb{N}$ such that

$$C \rhd_\alpha \cdots \rhd_\alpha \underline{m} :\equiv \mathsf{S}^{(m)}\mathsf{0}.$$

We define

$$\vec{\alpha}(C) := m.$$

Theorem 3.5 shows that $[\![C]\!]_0$ is an upper bound for $\vec{\alpha}(C)$ and the length of any reduction sequence starting from C, since we have

$$m < [\![\underline{m}]\!]_0.$$

Note that

$$[\![\underline{m}]\!]_0 \le \psi(\omega \cdot (m+1) + m + 2) < \psi(\omega^2 + m)$$

using Proposition 2.3, parts (2) and (3).

Let $F^{00} \in \mathcal{T}_n$ be closed. F represents the function

$$m \mapsto \vec{\alpha}(F\underline{m}),$$

and for any $m \in \mathbb{N}$ we have

$$\mathrm{D}_F(m), \vec{\alpha}(F\underline{m}) \le [\![F\underline{m}]\!]_0$$

where

$$\mathrm{D}_F(m) := \max\{k \mid \exists G_1, \ldots, G_k \in \mathfrak{T}\ F\underline{m} \equiv G_1 \rhd \cdots \rhd G_k\}.$$

We have

$$[\![F\underline{m}]\!]_0 = \psi(\omega \cdot [\![F]\!]_1 + [\![F]\!]_0 + [\![\underline{m}]\!]_0) < \psi(\omega \cdot ([\![F]\!]_1 + \omega^2) + [\![F]\!]_0 + m),$$

and the function

$$m \mapsto \psi(\underbrace{\omega \cdot (\llbracket F \rrbracket_1 + \omega^2) + \llbracket F \rrbracket_0}_{<\omega_{n+2} \text{ by Lemma 5.8}} + m)$$

is $<\omega_{n+2}$-recursive (cf. [**3**]), implying that also D_F and $m \mapsto \vec{\alpha}(F\underline{m})$ are $<\omega_{n+2}$-recursive. We therefore obtain that the functions definable in T_n are $<\omega_{n+2}$-recursive and the functions definable in GT are $<\varepsilon_0$-recursive.

Now let some $m \in \mathbb{N}$ and a term $G^\sigma \in \mathfrak{T}$ with $\mathrm{lh}(G)$, $\mathrm{L}(G) \leq m$ and $\mathrm{R}(G) \leq n+2$ be given. Let $\vec{g}$ be the closed canonical assignment to the term G. Then, according to Lemma 5.8,

$$\begin{aligned} g_0 &< \psi(\underbrace{\omega \cdot 2_{n+1}(\omega \cdot 2_{m+1}(2(m+1+\mathrm{lh}(G))))}_{<\omega_{n+2}}) \\ &< \psi(\omega_{n+2} + m) \quad \text{by Proposition 2.3, part (3).} \end{aligned}$$

This implies $\mathrm{D}_{\mathrm{T}_n}(m) < \psi(\omega_{n+2} + m)$, and hence $\mathrm{D}_{\mathrm{T}_n}$ is an ω_{n+2}-recursive function. Omitting the restriction concerning $\mathrm{R}(G)$ it similarly follows that D_{GT} is an ε_0-recursive function. □

5 Appendix: Proofs omitted in Sections 2 and 4

5.1 Proofs in Subsection 2.1

We give the detailed proofs of the two propositions regarding the ψ-function.

Proof of Proposition 2.2. Part (1) is an immediate consequence of the definitions of $\oplus$ and no. For part (ii), suppose

$$\alpha = \omega^{\alpha_1} + \cdots + \omega^{\alpha_n} > \alpha_1 \geq \cdots \geq \alpha_n \text{ and}$$

$$\beta = \omega^{\beta_1} + \cdots + \omega^{\beta_m} > \beta_1 \geq \cdots \geq \beta_m,$$

where n, $m \geq 1$. By definition of $\otimes$ we have

$$\begin{aligned} \mathrm{no}(\alpha \otimes \beta) &= nm + m\sum_{i=1}^{n} \mathrm{no}(\alpha_i) + n\sum_{j=1}^{m} \mathrm{no}(\beta_j) \\ &\leq nm + m\sum_{i=1}^{n} \mathrm{no}(\alpha_i) + n\sum_{j=1}^{m} \mathrm{no}(\beta_j) + \sum_{i=1}^{n} \mathrm{no}(\alpha_i) \cdot \sum_{j=1}^{m} \mathrm{no}(\beta_j) \\ &= \mathrm{no}(\alpha) \cdot \mathrm{no}(\beta). \end{aligned}$$

This estimation of the norm of the natural product holds for all α and β. Equality holds if and only if $\alpha < \omega$ or $\beta < \omega$. We further have

$$\begin{aligned} \mathrm{no}(\alpha) + \mathrm{no}(\beta) - 1 &= n + m - 1 + \sum_{i=1}^{n} \mathrm{no}(\alpha_i) + \sum_{j=1}^{m} \mathrm{no}(\beta_j) \\ &\leq nm + m\sum_{i=1}^{n} \mathrm{no}(\alpha_i) + n\sum_{j=1}^{m} \mathrm{no}(\beta_j) \\ &= \mathrm{no}(\alpha \otimes \beta). \end{aligned}$$

We next prove (3). Since the case $\alpha < \omega$ is trivial, we may assume that $\alpha \geq \omega$, say, $\alpha = \omega \cdot \alpha_0 + m$ where $0 < \alpha_0 \leq \alpha$ and $m < \omega$. Let

$$\alpha_0 = \omega^{\alpha_1} + \cdots + \omega^{\alpha_k} + \cdots + \omega^{\alpha_n} > \alpha_1 \geq \cdots \geq \alpha_k \geq \omega > \alpha_{k+1} \geq \cdots \geq \alpha_n$$

with $0 \leq k \leq n > 0$. Then we have

$$\omega \cdot \alpha_0 = \omega^{\alpha_1} + \cdots + \omega^{\alpha_k} + \omega^{\alpha_{k+1}+1} + \cdots + \omega^{\alpha_n+1},$$

which implies $\mathrm{no}(\alpha) = \mathrm{no}(\alpha_0) + n - k + m$. By definition, $2^\alpha = \omega^{\alpha_0} \cdot 2^m$; hence

$$\mathrm{no}(2^\alpha) = (\mathrm{no}(\alpha_0) + 1) \cdot 2^m.$$

We obtain from these preparations

$$\mathrm{no}(\alpha) \leq 2\mathrm{no}(\alpha_0) + m < 2(\mathrm{no}(\alpha_0) \cdot 2^m + 2^m) = 2 \cdot \mathrm{no}(2^\alpha)$$

as well as

$$\mathrm{no}(2^\alpha) = (\mathrm{no}(\alpha_0) + 1) \cdot 2^m \leq 2^{\mathrm{no}(\alpha_0)} \cdot 2^m \leq 2^{\mathrm{no}(\alpha)}.$$

This concludes the proof of Proposition 2.2. □

Proof of Proposition 2.3. The proof of part (1) is by straightforward induction on k and follows directly from the definition of ψ. Also part (3) immediately follows from the Definition of ψ. For part (4) note that $\Phi(\mathrm{no}(\alpha)) = \psi(\Phi(\mathrm{no}(\alpha)))$ by part (1), and then apply part (3). For $\alpha = 0$ part (2) follows immediately from part (1). In the case $\alpha > 0$ the first $\leq$-relation is immediate by part (1); for the second $\leq$-relation we argue by induction on β:

- $\beta = 0$: Trivial.
- $\beta > 0$: By definition of ψ there exists a $\gamma < \beta$ such that $\mathrm{no}(\gamma) \leq \Phi(\mathrm{no}(\beta))$ and $\psi(\beta) = \psi(\gamma) + 1$. Therefore

  $$\psi(\alpha + \psi(\beta)) \leq \psi(\alpha \oplus \gamma + 1) \leq \psi(\alpha \oplus \beta)$$

  where the first $\leq$-relation follows from the induction hypothesis for $\alpha+1$ and the second $\leq$-relation is verified using part (3). If $\gamma + 1 = \beta$ we are done. Otherwise we have $\alpha \oplus \gamma + 1 < \alpha \oplus \beta$ and

  $$\mathrm{no}(\alpha \oplus \gamma + 1) \leq \mathrm{no}(\alpha) + 1 + \Phi(\mathrm{no}(\beta)) \leq \Phi(\mathrm{no}(\alpha \oplus \beta)),$$

  using that $\alpha > 0$. □

5.2 Proofs in Subsection 2.3

We give detailed proofs of the lemmas regarding the □-operator.

Proof of Lemma 2.11. The proof is essentially the same as in [**15**]. We give the details for the reader's convenience, frequently using Proposition 2.2. We first show the following:

Claim For $k \leq i \leq n$, we have

$$\mathrm{no}((\vec{f} \,\Box\, \vec{g})_i) \prec F_0^{2(n+1 \dot{-} i)}(f + g + 1), \tag{5.1}$$

where $n \dot{-} i := n - i$ if $n \geq i$, and $n \dot{-} i := 0$ otherwise. The claim is shown by induction on $n \dot{-} i$. For $i = n$, we obtain

$$\begin{aligned}
\mathrm{no}((\vec{f} \,\Box\, \vec{g})_n) &= \mathrm{no}(2^{f_{n+1}} \cdot (f_n + g_n)) \\
&\preceq 2^{\mathrm{no}(f_{n+1})} \cdot (\mathrm{no}(f_n) + \mathrm{no}(g_n)) \\
&\preceq 2^f \cdot (f + g) \\
&\prec 2^{2^{f+g}} \cdot (f + g) \\
&\prec 2^{2^{f+g+1}} \\
&= F_0^2(f + g + 1).
\end{aligned}$$

For $k \leq i < n$, we have

$$\begin{aligned}
\mathrm{no}((\vec{f} \,\Box\, \vec{g})_i) &= \mathrm{no}(2^{(\vec{f} \,\Box\, \vec{g})_{i+1}} \cdot (f_i + g_i)) \\
&\preceq 2^{F_0^{2(n+1-i)-2}(f+g+1)} \cdot (f + g) \\
&= F_0^{2(n+1-i)-1}(f + g + 1) \cdot (f + g) \\
&\prec F_0^{2(n+1-i)}(f + g + 1).
\end{aligned}$$

Now we prove the lemma. For $k, n+1 \leq i \leq m$, we have

$$\mathrm{no}((\vec{f} \,\Box\, \vec{g})_i) = \mathrm{no}(f_i) \preceq f \prec F_2(f + g + n).$$

For $k \leq i \leq n$, we obtain

$$\begin{aligned}
\mathrm{no}((\vec{f} \,\Box\, \vec{g})_i) &\prec F_0^{2(n+1 \dot{-} i)}(f + g + 1) \text{ by 5.1} \\
&\preceq F_0^{2n}(f + g + 1) \\
&\preceq F_1(f + g + 2n) \\
&\prec F_2(f + g + n),
\end{aligned}$$

concluding the proof of the lemma. □

Proof of Lemma 2.15. Part (1) is adapted from [**8**], p. 450, and is shown here for the reader's convenience. We argue by induction on $n + 1 \dot{-} i$ for $1 \leq i \leq n$.

$$\begin{aligned}
&(\vec{a} \,\Box\, \vec{d})_i + (\vec{b} \,\Box\, \vec{d})_i \\
&\quad \prec 2^{(\vec{a} \,\Box\, \vec{d})_{i+1}} \cdot (a_i + b_i + d_i) + 2^{(\vec{b} \,\Box\, \vec{d})_{i+1}} \cdot (a_i + b_i + d_i) \\
&\quad \preceq 2^{(\vec{a} \,\Box\, \vec{d})_{i+1} + (\vec{b} \,\Box\, \vec{d})_{i+1}} \cdot (a_i + b_i + d_i) \\
&\quad \preceq 2^{(\vec{c} \,\Box\, \vec{d})_{i+1}} \cdot (c_i + d_i) \\
&\quad = (\vec{c} \,\Box\, \vec{d})_i.
\end{aligned}$$

We now prove part (2). For $1 \leq i \leq m$, we set

$$N_i := \sum_{j=i}^{m} \mathrm{no}(a_j), \quad M_i := \sum_{j=i}^{m} \mathrm{no}(b_j), \quad L_i := \sum_{j=i}^{n} \mathrm{no}(d_j).$$

Lemma 2.11 yields, for $1 \leq i \leq m$,

$$\mathrm{no}((\vec{a} \,\Box\, \vec{d})_i) \preceq F_2(N_i + L_i + n). \tag{5.2}$$

Since the case $i > n$ is trivial, we assume in the sequel that $1 \leq i \leq n$. By Proposition 2.2, part (iii), we have

$$\mathrm{no}((\vec{b} \,\Box\, \vec{d})_{i+1}) \preceq 2\mathrm{no}((\vec{b} \,\Box\, \vec{d})_i).$$

From this we obtain, for $i \leq j \leq n+1$,

$$\mathrm{no}((\vec{b} \,\Box\, \vec{d})_j) \preceq 2^n \cdot \mathrm{no}((\vec{b} \,\Box\, \vec{d})_i).$$

Since $\mathrm{no}(b_j) \preceq \mathrm{no}((\vec{b} \,\Box\, \vec{d})_j)$ for every j and $\mathrm{no}(d_j) \preceq \mathrm{no}((\vec{b} \,\Box\, \vec{d})_j)$ for $1 \leq j \leq n$, we get

$$M_i + L_i \preceq (n+1) \cdot 2^{n+1} \cdot \mathrm{no}((\vec{b} \,\Box\, \vec{d})_i). \tag{5.3}$$

Now (5.2) and (5.3) together with $N_i \preceq M_i$ yield

$$\begin{aligned}\mathrm{no}((\vec{a} \,\Box\, \vec{d})_i) &\preceq F_2(M_i + L_i + n)\\ &\preceq F_2((n+1) \cdot 2^{n+1} \cdot \mathrm{no}((\vec{b} \,\Box\, \vec{d})_i) + n)\\ &\preceq F_3(\mathrm{no}((\vec{b} \,\Box\, \vec{d})_i) + n).\end{aligned}$$

Finally we prove part (3). For $n < i \leq m$ the claim holds by assumption, and for $1 \leq i \leq n$ the claim follows by part (1). Consider the case $i = 0$ (the inequalities are explained below):

$$\begin{aligned}&(\vec{a} \,\Box\, \vec{d})_0 + (\vec{b} \,\Box\, \vec{d})_0\\ &\qquad = \psi(\omega \cdot (\vec{a} \,\Box\, \vec{d})_1 + a_0 + d_0 + n) + \psi(\omega \cdot (\vec{b} \,\Box\, \vec{d})_1 + b_0 + d_0 + n)\\ &\qquad \preceq \psi(\omega \cdot ((\vec{a} \,\Box\, \vec{d})_1 + (\vec{b} \,\Box\, \vec{d})_1) + a_0 + b_0 + 2d_0 + 2n)\\ &\qquad \prec \psi(\omega \cdot (\vec{c} \,\Box\, \vec{d})_1 + c_0 + d_0 + n)\\ &\qquad = (\vec{c} \,\Box\, \vec{d})_0.\end{aligned}$$

The $\preceq$-relationship follows by Proposition 2.3, parts (1) and (2). We next show that the $\prec$-relationship holds, making use of Proposition 2.3, part (3), and part (2) of the present lemma:

$$(\vec{a} \,\Box\, \vec{d})_1 + (\vec{b} \,\Box\, \vec{d})_1 \prec (\vec{c} \,\Box\, \vec{d})_1$$

holds by assumption if $n = 0$, and for $n > 0$ this has already been shown. We obtain

$$\omega \cdot ((\vec{a} \,\Box\, \vec{d})_1 + (\vec{b} \,\Box\, \vec{d})_1) + a_0 + b_0 + 2d_0 + 2n \prec \omega \cdot (\vec{c} \,\Box\, \vec{d})_1 + c_0 + d_0 + n.$$

Part (2) yields

$$\mathrm{no}((\vec{a} \,\Box\, \vec{d})_1), \mathrm{no}((\vec{b} \,\Box\, \vec{d})_1) \preceq F_3(\mathrm{no}(\vec{c} \,\Box\, \vec{d})_1 + n).$$

From this we easily verify the second assumption of Proposition 2.3, part (3). □

Proof of Lemma 2.16. We proceed by induction on $n + 1 \dot{-} i$. If $i = n+1$, then

$$2e_i = 2a_i \prec c_i = (\vec{c} \,\Box\, \vec{d})_i.$$

In the case $i = n$, we have

$$\begin{aligned} e_n &= 2^{a_{n+1}} \cdot (2^{a_{n+1}} \cdot (a_n + d_n) + 2^{b_{n+1}} \cdot (b_n + d_n)) \\ &\prec 2^{2a_{n+1}} \cdot (a_n + b_n + d_n) + 2^{a_{n+1}+b_{n+1}} \cdot (a_n + b_n + d_n) \\ &\preceq 2^{2a_{n+1}+b_{n+1}} \cdot (a_n + b_n + d_n). \end{aligned}$$

This implies

$$2e_n \prec 2^{c_{n+1}} \cdot (c_n + d_n) = (\vec{c} \,\Box\, \vec{d})_n.$$

Now let us assume that $1 \leq i < n$:

$$\begin{aligned} e_i &= 2^{e_{i+1}} \cdot (2^{(\vec{a} \,\Box\, \vec{d})_{i+1}} \cdot (a_i + d_i) + 2^{(\vec{b} \,\Box\, \vec{d})_{i+1}} \cdot (b_i + d_i)) \\ &\prec 2^{e_{i+1}+(\vec{a} \,\Box\, \vec{d})_{i+1}} \cdot (a_i + b_i + d_i) + 2^{e_{i+1}+(\vec{b} \,\Box\, \vec{d})_{i+1}} \cdot (a_i + b_i + d_i) \\ &\preceq 2^{2e_{i+1}} \cdot (c_i + d_i), \end{aligned}$$

where the last $\preceq$-relation holds since $i + 1 \leq n$ and, by Lemma 2.13,

$$(\vec{a} \,\Box\, \vec{d})_{i+1} + (\vec{b} \,\Box\, \vec{d})_{i+1} \preceq e_{i+1}.$$

Using the induction hypothesis, which allows us to estimate $2e_i \prec 2^{2e_{i+1}+1} \cdot (c_i + d_i) \preceq 2^{(\vec{c} \,\Box\, \vec{d})_{i+1}} \cdot (c_i + d_i)$, we finally obtain

$$2e_i \prec (\vec{c} \,\Box\, \vec{d})_i.$$

This concludes the proof of the lemma. □

Proof of Lemma 2.17. For convenience we set $a := a_0$, $b := b_0$, $c := c_0$, and $d := d_0$. We first show the following:

Claim For $1 \leq i \leq n$, setting $e := a + b + c + d + 2(n + 1)$, we have

$$\operatorname{no}(e_i) \preceq F_2^{2(n+1 \dot{-} i)}(e). \tag{5.4}$$

The claim is proved by induction on $n \dot{-} i$. We will make use of the following abbreviations:

$$\begin{aligned} \alpha &:= F_2(a + b + n), \\ \beta &:= F_2(c + d + n), \\ \gamma &:= F_2(a + b + c + d + 2n + 1). \end{aligned}$$

If $i = n$, then we obtain

$$\begin{aligned} \operatorname{no}(e_n) &= \operatorname{no}(2^{a_{n+1}} \cdot ((\vec{a} \,\Box\, \vec{b})_n + (\vec{c} \,\Box\, \vec{d})_n)) \\ &\preceq 2^{\operatorname{no}(a_{n+1})} \cdot (\operatorname{no}((\vec{a} \,\Box\, \vec{b})_n) + \operatorname{no}((\vec{c} \,\Box\, \vec{d})_n)) \\ &\preceq 2^a \cdot (\alpha + \beta) \text{ by Lemma 2.11} \\ &\preceq F_2(\gamma + 1) \\ &\preceq F_2^2(e). \end{aligned}$$

If $1 \leq i < n$, then we have

$$\begin{aligned}
\mathrm{no}(e_i) &\preceq 2^{\mathrm{no}(e_{i+1})} \cdot (\mathrm{no}((\vec{a} \,\Box\, \vec{b})_i) + \mathrm{no}((\vec{c} \,\Box\, \vec{d})_i)) \\
&\preceq 2^{F_2^{2(n+1 \dot{-} i)-2}(e)} \cdot (\alpha + \beta) \text{ by Lemma 2.11 and the induction hypothesis} \\
&\preceq F_2^{2(n+1 \dot{-} i)-1}(e) \cdot \gamma \\
&\prec F_2^{2(n+1 \dot{-} i)}(e).
\end{aligned}$$

Now we prove the lemma from the above claim. The case $i = n + 1$ is trivial, since we then have $e_{n+1} = a_{n+1}$. For $1 \leq i \leq n$, we finally obtain

$$\begin{aligned}
\mathrm{no}(e_i) &\preceq F_2^{2(n+1 \dot{-} i)}(e) \text{ by 5.4} \\
&\preceq F_2^{2n}(e) \\
&\prec F_3(a + b + c + d + n).
\end{aligned}$$

This concludes the proof of Lemma 2.17. □

5.3 Proofs in Subsection 2.4

We provide the proofs regarding the operator δ. Our first goal is to show that the operators $\delta^{\vec{x}}$ preserve bounded norms. To this end, and in preparation of the analysis of our assignment in Section 4, we need to introduce precise notions of subterms.

Definition 5.1 By recursion on the buildup of $h \in \mathcal{C}_i^{\vec{x}}$ we define the set $\mathrm{T}_{i,j}^{\vec{x}}(h)$ of maximal x-free subterms of j-th level of h and the set $\mathrm{Sub}_{i,j}^{\vec{x}}(h)$ of those subterms of j-th level of h which are different from x_i, where x-free subterms are considered atomic. We use the abbreviation

$$\{h\}_{i,j} := \{h \mid i = j\}.$$

If h is x-free, then

$$\mathrm{T}_{i,j}^{\vec{x}}(h) := \{h\}_{i,j} =: \mathrm{Sub}_{i,j}^{\vec{x}}(h),$$

otherwise we distinguish between the following cases:

- If $h \equiv x_i$, then
$$\mathrm{T}_{i,j}^{\vec{x}}(h) := \emptyset =: \mathrm{Sub}_{i,j}^{\vec{x}}(h).$$
- If $h \equiv f + g$, then
$$\mathrm{T}_{i,j}^{\vec{x}}(h) := \mathrm{T}_{i,j}^{\vec{x}}(f) \cup \mathrm{T}_{i,j}^{\vec{x}}(g),$$
$$\mathrm{Sub}_{i,j}^{\vec{x}}(h) := \{h\}_{i,j} \cup \mathrm{Sub}_{i,j}^{\vec{x}}(f) \cup \mathrm{Sub}_{i,j}^{\vec{x}}(g).$$
- If $h \equiv 2^f \cdot g$ or $h \equiv \psi(\omega \cdot f + g)$, then
$$\mathrm{T}_{i,j}^{\vec{x}}(h) := \mathrm{T}_{i+1,j}^{\vec{x}}(f) \cup \mathrm{T}_{i,j}^{\vec{x}}(g),$$
$$\mathrm{Sub}_{i,j}^{\vec{x}}(h) := \{h\}_{i,j} \cup \mathrm{Sub}_{i+1,j}^{\vec{x}}(f) \cup \mathrm{Sub}_{i,j}^{\vec{x}}(g).$$

We further define

$$\mathrm{T}_i^{\vec{x}}(h) := \bigcup_{i \leq j} \mathrm{T}_{i,j}^{\vec{x}}(h) \quad \text{ and } \quad \mathrm{Sub}_i^{\vec{x}}(h) := \bigcup_{i \leq j} \mathrm{Sub}_{i,j}^{\vec{x}}(h).$$

Notice that for $h \in \mathcal{C}_i^{\vec{x}}$ the set $\mathrm{T}_{i,j}^{\vec{x}}(h)$ comprises the x-free terms of $\mathrm{Sub}_{i,j}^{\vec{x}}(h)$, that we have $\mathrm{T}_{i,j}^{\vec{x}}(h) = \mathrm{Sub}_{i,j}^{\vec{x}}(h) = \emptyset$ if $i > j$, and that for every $t \in \mathrm{Sub}_{i,i}^{\vec{x}}(h)$ we have $t \preceq h$.

Lemma 5.2 *Let $\vec{x}$ be a variable vector and set $k := \mathrm{lv}(\vec{x})$, $n := k+1$.*

(1) *For any $h \in \mathcal{C}_i^{\vec{x}}$ and $t \in \mathrm{Sub}_i^{\vec{x}}(h)$, we have*

$$\mathrm{no}(t) \preceq 2^{n \dot{-} i} \cdot \mathrm{no}(h).$$

(2) *Let $\vec{h} \in \mathcal{C}^{\vec{x}}$ be of bounded norm, $m := \mathrm{lv}(\vec{h})$. Then for all $t \in \mathrm{T}_i^{\vec{x}}(h_i)$, $i \le m$, we have*

$$\mathrm{no}(t) \prec 2^{n \dot{-} i} \cdot (\,\delta_0^{\vec{x}} h_0)_0, \tag{5.5}$$

and for $0 < j \le n$ we have

$$\mathrm{no}((\,\delta_i^{\vec{x}} h_i)_j) \preceq \mathrm{sz}^{\vec{x}}(h_i) \cdot 2^n \cdot (\,\delta_0^{\vec{x}} h_0)_0. \tag{5.6}$$

(3) *$\delta^{\vec{x}}$ preserves bounded norm: for every $\vec{h} \in \mathcal{C}^{\vec{x}}$ of bounded norm, $\delta^{\vec{x}}\vec{h}$ is of bounded norm.*

Proof. Part (1) is shown by induction on the buildup of h. If $i > k$, then h is x-free, so $t \equiv h$ and we are done. Suppose $i \le k$. If h is of a form $\psi(\omega \cdot f + g)$, we use Proposition 2.3 to see that $\mathrm{no}(f), g \preceq h$. In the interesting case, where $h \equiv 2^f \cdot g$ and $t \in \mathrm{Sub}_{i+1}^{\vec{x}}(f)$, we have $\mathrm{no}(t) \preceq 2^{n \dot{-} (i+1)} \cdot \mathrm{no}(f)$, and by Proposition 2.2 we have $\mathrm{no}(f) \preceq 2\mathrm{no}(2^f) \preceq 2\mathrm{no}(h)$. Therefore, $\mathrm{no}(t) \preceq 2^{n \dot{-} i} \cdot \mathrm{no}(h)$.

We turn to the proof of part (2). By part (1) we have $\mathrm{no}(t) \preceq 2^{n \dot{-} i} \cdot \mathrm{no}(h_i)$. Since t is x-free and $\vec{h}$ of bounded norm, we even have

$$\mathrm{no}(t) \preceq 2^{n \dot{-} i} \cdot \mathrm{no}(h_i\{\vec{x} := \vec{1}\}) \preceq 2^{n \dot{-} i} \cdot h_0\{\vec{x} := \vec{1}\},$$

which by Lemma 2.19 implies (5.5). In order to show (5.6), set $\nu := 2^n \cdot (\,\delta_0^{\vec{x}} h_0)_0$. We show by induction on the buildup of $h \in \mathcal{C}_i^{\vec{x}}$ that if $\mathrm{no}(t) \prec \nu$ for all $t \in \mathrm{T}_i^{\vec{x}}(h)$, then

$$\mathrm{no}((\,\delta_i^{\vec{x}} h)_j) \preceq \mathrm{sz}^{\vec{x}}(h) \cdot \nu.$$

If h is x-free, we obtain $\mathrm{no}((\,\delta_i^{\vec{x}} h)_j) \preceq \mathrm{no}(h) + 1 \preceq \nu$. Now suppose h is not x-free. The case $h \equiv x_i$ is trivial, the case $h \equiv f + g$ follows directly from the induction hypothesis, and in the remaining cases notice that $\mathrm{T}_{i+1}^{\vec{x}}(f), \mathrm{T}_i^{\vec{x}}(g) \subseteq \mathrm{T}_i^{\vec{x}}(h)$. For $h \equiv 2^f \cdot g$, we obtain

$$\mathrm{no}((\,\delta_i^{\vec{x}} h)_j) \preceq 2\mathrm{no}((\,\delta_{i+1}^{\vec{x}} f)_j) + \mathrm{no}((\,\delta_i^{\vec{x}} g)_j) + 1,$$

which implies the claimed estimate since $\mathrm{sz}^{\vec{x}}(h) = 2\mathrm{sz}^{\vec{x}}(f) + \mathrm{sz}^{\vec{x}}(g) + 1$. The remaining situation $h \equiv \psi(\omega \cdot f + g)$ is handled similarly.

Part (3) is now easy to see. In the case $j > n$, we apply Lemma 2.19 to obtain

$$\mathrm{no}((\,\delta^{\vec{x}}\vec{h})_j) = \mathrm{no}(h_j) \preceq h_0\{\vec{x} := \vec{1}\} \prec (\,\delta_0^{\vec{x}} h_0)_0 \preceq (\,\delta^{\vec{x}}\vec{h})_0.$$

For $0 < j \le n$, we apply (5.6) to obtain

$$\mathrm{no}((\,\delta^{\vec{x}}\vec{h})_j) = \sum_{i=0}^{m} \mathrm{no}((\,\delta_i^{\vec{x}} h_i)_j) \preceq \left(\sum_{i=0}^{m} \mathrm{sz}^{\vec{x}}(h_i)\right) \cdot 2^n \cdot (\,\delta_0^{\vec{x}} h_0)_0 = (\,\delta^{\vec{x}}\vec{h})_0.$$

Thus $\delta^{\vec{x}}\vec{h}$ is of bounded norm. □

Proof of Lemma 2.22. For convenience we set $k := \mathrm{lv}(\vec{x})$, $n := k+1$, $m := \mathrm{lv}(\vec{h})$, and $l := \max\{m, n\}$. We first show that the lemma is a consequence of the following:

Claim For $h \in \mathcal{C}^{\vec{x}}_i$, $i \leq n$, we have

$$h \prec (\,\delta^{\vec{x}}_i h \,\Box\, \vec{x})_i. \tag{5.7}$$

In order to derive the lemma from (5.7), let $i \leq m$. We distinguish between the following three cases.

Case 1: $n < i \leq l$. Then clearly $h_i \equiv (\,\delta^{\vec{x}} \vec{h} \,\Box\, \vec{x})_i$.

Case 2: $1 \leq i \leq n$. We have

$$(\,\delta^{\vec{x}}_i h_i)_j \preceq (\,\delta^{\vec{x}} \vec{h})_j$$

for $i \leq j \leq n$. Thus by Lemma 2.14, part (1),

$$(\,\delta^{\vec{x}}_i h_i \,\Box\, \vec{x})_i \preceq (\,\delta^{\vec{x}} \vec{h} \,\Box\, \vec{x})_i,$$

as we may ignore components of $\delta^{\vec{x}} \vec{h}$ above the n-th. By (5.7), we have $h_i \prec (\,\delta^{\vec{x}}_i h_i \,\Box\, \vec{x})_i$.

Case 3: $i = 0$. Here (5.7) applies, since we have

$$\begin{aligned}(\,\delta^{\vec{x}}_0 h_0 \,\Box\, \vec{x})_0 &= \psi(\omega \cdot (\,\delta^{\vec{x}}_0 h_0 \,\Box\, \vec{x})_1 + (\,\delta^{\vec{x}}_0 h_0)_0 + x_0 + k) \\ &\prec \psi(\omega \cdot (\,\delta^{\vec{x}} \vec{h} \,\Box\, \vec{x})_1 + (\,\delta^{\vec{x}} \vec{h})_0 + x_0 + k) \\ &= (\,\delta^{\vec{x}} \vec{h} \,\Box\, \vec{x})_0\end{aligned}$$

by Proposition 2.3, part (3), whose assumptions are easily checked: For all $j \leq n$, we have $(\,\delta^{\vec{x}}_0 h_0)_j \prec (\,\delta^{\vec{x}} \vec{h})_j$. Lemma 2.14, part (1), yields $(\,\delta^{\vec{x}}_0 h_0 \,\Box\, \vec{x})_1 \prec (\,\delta^{\vec{x}} \vec{h} \,\Box\, \vec{x})_1$. By Lemma 2.15, part (2), $\mathrm{no}((\,\delta^{\vec{x}}_0 h_0 \,\Box\, \vec{x})_1) \preceq F_3(\mathrm{no}((\,\delta^{\vec{x}} \vec{h} \,\Box\, \vec{x})_1) + k)$.

We now prove (5.7) by induction on the definition of $\delta^{\vec{x}}_i$. Assume first that h is x-free. Then clearly $h \prec h + 1 = (\,\delta^{\vec{x}}_i h)_i \preceq (\,\delta^{\vec{x}}_i h \,\Box\, \vec{x})_i$. Otherwise we must have $i \leq k$ and distinguish between the following four cases:

Case 1: $h \equiv x_i$. $h \prec x_i + 1 \preceq (\vec{1} \,\Box\, \vec{x})_i$.

Case 2: $h \equiv f + g$. Then we apply the induction hypothesis and use Lemma 2.15, part (3):

$$\begin{aligned}h \equiv f + g &\prec (\,\delta^{\vec{x}}_i f \,\Box\, \vec{x})_i + (\,\delta^{\vec{x}}_i g \,\Box\, \vec{x})_i \\ &\prec (\,\delta^{\vec{x}}_i h \,\Box\, \vec{x})_i.\end{aligned}$$

Case 3: $h \equiv 2^f \cdot g$. Then $i \geq 1$, and after applying the induction hypothesis we use Lemma 2.16:

$$\begin{aligned}h \equiv 2^f \cdot g &\prec 2^{(\,\delta^{\vec{x}}_{i+1} f \,\Box\, \vec{x})_{i+1}} \cdot (\,\delta^{\vec{x}}_i g \,\Box\, \vec{x})_i \\ &\preceq ((\,\delta^{\vec{x}}_{i+1} f \,\Box\, \vec{x}) \,\Box\, (\,\delta^{\vec{x}}_i g \,\Box\, \vec{x}\restriction_k))_i \\ &\prec (\,\delta^{\vec{x}}_i h \,\Box\, \vec{x})_i.\end{aligned}$$

Case 4: $h \equiv \psi(\omega \cdot f + g)$. Then we have $i = 0$ and obtain

$$\begin{aligned}
h &\prec \psi(\omega \cdot (\, \delta_1^{\vec{x}} f \,\Box\, \vec{x})_1 + (\, \delta_0^{\vec{x}} g \,\Box\, \vec{x})_0) \text{ (see below)} \\
&= \psi(\omega \cdot (\, \delta_1^{\vec{x}} f \,\Box\, \vec{x})_1 + \psi(\omega \cdot (\, \delta_0^{\vec{x}} g \,\Box\, \vec{x})_1 + (\, \delta_0^{\vec{x}} g)_0 + x_0 + k)) \\
&\preceq \psi(\omega \cdot ((\, \delta_1^{\vec{x}} f \,\Box\, \vec{x})_1 + (\, \delta_0^{\vec{x}} g \,\Box\, \vec{x})_1) + (\, \delta_0^{\vec{x}} g)_0 + x_0 + k) \\
&\preceq \psi(\omega \cdot (\, \delta_0^{\vec{x}} h \,\Box\, \vec{x})_1 + (\, \delta_0^{\vec{x}} h)_0 + x_0 + k) \text{ (see below)} \\
&= (\, \delta_0^{\vec{x}} h \,\Box\, \vec{x})_0.
\end{aligned}$$

The strict inequality follows from the induction hypothesis, using that

$$\text{no}(f) \preceq F_2(g) \prec F_2((\, \delta_0^{\vec{x}} g \,\Box\, \vec{x})_0).$$

The last inequality is easily verified: In the case $k > 0$, Lemma 2.15, part (1), yields

$$(\, \delta_1^{\vec{x}} f \,\Box\, \vec{x})_1 + (\, \delta_0^{\vec{x}} g \,\Box\, \vec{x})_1 \prec (\, \delta_0^{\vec{x}} h \,\Box\, \vec{x})_1,$$

and for $k = 0$ both terms are equal. Using Lemma 2.15, part (2), we obtain

$$\text{no}((\, \delta_1^{\vec{x}} f \,\Box\, \vec{x})_1), \text{no}((\, \delta_0^{\vec{x}} g \,\Box\, \vec{x})_1) \preceq F_3(\text{no}((\, \delta_0^{\vec{x}} h \,\Box\, \vec{x})_1) + k).$$

Clearly, we have $(\, \delta_0^{\vec{x}} g)_0 \preceq (\, \delta_0^{\vec{x}} h)_0$. □

5.4 Preparations for the proof of Corollary 4.2 in Section 4

Here we are going to show Lemmata 5.5 and 5.8, where the former will be used in the proof of the latter. As a preparation for the proof of Lemma 5.5, recall the definition of $\text{sz}^{\vec{x}}$ in Definition 2.18. We define the variable independent version sz to serve as another auxiliary function in order to estimate the term complexity of ordinal terms occurring in our assignment. Note that we have $\text{sz}^{\vec{x}} \leq \text{sz}$.

Definition 5.3 $\text{sz}(h)$ for $h \in \text{ot}$ is defined by

- $\text{sz}(h) := 1$ if h is a variable or constant.
- $\text{sz}(h) := \text{sz}(f) + \text{sz}(g) + 1$ if h is of a form either $f + g$ or $\psi(\omega \cdot f + g)$.
- $\text{sz}(h) := 2\text{sz}(f) + \text{sz}(g) + 1$ if h is of a form $2^f \cdot g$.

Lemma 5.4 *Suppose $h \in \mathcal{C}_i^{\vec{x}}$ for some $i \in \mathbb{N}$. We have*

$$\text{sz}((\, \delta_i^{\vec{x}} h)_j) \leq 4\text{sz}(h)$$

for $j \leq \text{lv}(\vec{x}) + 1$.

Proof. The proof is by straightforward induction on the buildup of h, along the definition of $\delta_i^{\vec{x}} h$. □

Lemma 5.5 *Let $G \in \mathfrak{T}$ and $\vec{g} := [\![G]\!]$ be its canonical assignment. Setting $m := \text{lv}(\vec{g})$, we have*

$$\sum_{i=0}^{m} \text{sz}(g_i) < 2_2(\text{L}(G) + 2\text{lh}(G)) =: M(G). \tag{5.8}$$

Setting $n := \text{lv}(\vec{x}) + 1$ and $L_n(G) := \max\{n, \text{L}(G)\}$, we have

$$\text{S}^{\vec{x}}(\vec{g}) < 2_2(L_n(G) + 1 + 2\text{lh}(G)). \tag{5.9}$$

Proof. The proof of (5.8) is by induction on $\mathrm{lh}(G)$. The second claim (5.9) then follows from the first one, since $n \leq L_n(G)$.

Case 1: G is a constant or variable. Then the claim is trivial once we notice that $m \leq \mathrm{L}(G)$.

Case 2: $G \equiv AB$. Let $\vec{a}$ and $\vec{b}$ be the canonical assignments of A and B, respectively, and set $m_A := \mathrm{lv}(\vec{a})$, $m_B := \mathrm{lv}(\vec{b})$. The vector $\vec{c} := \vec{a} \mathbin{\square} \vec{b}$ agrees with $\vec{g}$ up to component m, and for $m_B < i \leq m_A$ we have $c_i = a_i$, hence by the induction hypothesis

$$\mathrm{sz}(c_i) = \mathrm{sz}(a_i) < M(A).$$

By side induction on $m_B + 1 \dot{-} i$ we show that

$$\mathrm{sz}(c_i) < 2^{2(m_B+1\dot{-}i)}(M(A) + M(B)). \tag{5.10}$$

The case $i = m_B + 1$ has already been taken care of. Suppose $i \leq m_B$. If $i > 0$, we have $\mathrm{sz}(c_i) = 2\mathrm{sz}(c_{i+1}) + \mathrm{sz}(a_i) + \mathrm{sz}(b_i) + 2$, while $\mathrm{sz}(c_i) = \mathrm{sz}(c_{i+1}) + \mathrm{sz}(a_i) + \mathrm{sz}(b_i) + 2(m_B + 1)$ for $i = 0$. In any case, we obtain

$$\begin{aligned} \mathrm{sz}(c_i) &< 2^{2(m_B \dot{-} i)+1}(M(A) + M(B)) + M(A) + M(B) \\ &< 2^{2(m_B+1\dot{-}i)}(M(A) + M(B)). \end{aligned}$$

Using (5.10), the formula for the geometric series yields

$$\sum_{i=0}^{m_B} \mathrm{sz}(c_i) < 4^{m_B+2}(M(A) + M(B));$$

hence, together with the induction hypothesis applied to A,

$$\begin{aligned} \sum_{i=0}^{m} \mathrm{sz}(g_i) &< (4^{\mathrm{L}(G)+1} + 1)(M(A) + M(B)) \\ &< 2_2(\mathrm{L}(G) + 2\mathrm{lh}(G) - 1) \cdot (2_2(\mathrm{L}(G) + 2\mathrm{lh}(A)) + 2_2(\mathrm{L}(G) + 2\mathrm{lh}(B))) \\ &\leq (2_2(\mathrm{L}(G) + 2\mathrm{lh}(G) - 1))^2 \\ &= M(G), \end{aligned}$$

showing (5.10) for application terms.

Case 3: $G \equiv \lambda Y.F$. Let $\vec{f}$ be the canonical assignment to F, and set $k := \mathrm{lv}(\vec{f})$ and $l := \mathrm{lv}(\vec{y}) + 1$. We have $\vec{g} = \delta^{\vec{y}} \vec{f} + \vec{f}\{\vec{y} := \vec{1}\}$ and $m = \max\{k, l\}$. Setting

$$M := 2_2(\mathrm{L}(G) + 2\mathrm{lh}(F)),$$

by the induction hypothesis (5.9) applied to F, we have

$$\begin{aligned} \mathrm{S}^{\vec{y}}(\vec{f}) &< 2_2(L_l(F) + 1 + 2\mathrm{lh}(F)) \\ &\leq 2_2(\mathrm{L}(G) + 1 + 2\mathrm{lh}(F)) \\ &= M^2, \end{aligned}$$

since $L_l(F) = \max\{l, \mathrm{L}(F)\} \leq \mathrm{L}(G)$, and by the induction hypothesis (5.8), we have

$$\sum_{j=0}^{k} \mathrm{sz}(f_j) < M(F) \leq M,$$

whence using Lemma 5.4 we obtain

$$\mathrm{sz}(g_i) < \begin{cases} M^2(4\mathrm{sz}(f_0)+1) & \text{if } i=0, \\ 4M+k+\mathrm{sz}(f_i)+2 & \text{if } 1 \leq i \leq l, \\ 2\mathrm{sz}(f_i)+2 & \text{if } l < i \leq m. \end{cases}$$

We may now generously estimate the sum of the above terms:

$$\begin{aligned} \sum_{i=0}^{m} \mathrm{sz}(g_i) &< M^2(4\mathrm{sz}(f_0)+1) + (4\mathrm{L}(G)+2)M + (\mathrm{L}(G)+2)^2 \\ &< M^2(4\mathrm{sz}(f_0)+1) + (4\mathrm{L}(G)+3)M \\ &< M^2(4\mathrm{sz}(f_0)+2) \\ &< M^4 \\ &= M(G), \end{aligned}$$

which concludes the proof of Lemma 5.5. □

The following two lemmas will prepare the proof of Lemma 5.8. Recall Definition 5.1, and for any $h \in \mathrm{ot}$ let $\overline{h}$ be the closure of h by replacing every variable in h with 1.

Lemma 5.6 *Suppose $h \in \mathcal{C}^{\vec{x}}_k$ and $i < j$ where $i \leq n := \mathrm{lv}(\vec{x})+1$. Then we have*

$$\mathrm{Sub}^{\vec{y}}_{i,j}((\,\delta^{\vec{x}}_k h)_i) \subseteq \mathrm{Sub}^{\vec{y}}_{k,j}(h\{\vec{x} := \vec{1}\}).$$

Proof. The proof is by induction on the buildup of h along the definition of the partial operators $\delta^{\vec{x}}_k$, $k \in \mathbb{N}$. If h is x-free, then we have $(\,\delta^{\vec{x}}_i h)_i = h+1$ if $k=i$ while $(\,\delta^{\vec{x}}_k h)_i = 1$ if $k \neq i$, and the claim follows immediately. Now suppose that h is not x-free. We then distinguish between the following cases.

Case 1: h is a variable or constant. Since $i < j$, we then have $\mathrm{Sub}^{\vec{y}}_{i,j}((\,\delta^{\vec{x}}_k h)_i) = \emptyset$, so there is nothing to show.

Case 2: $h \equiv f+g$. Then we have $(\,\delta^{\vec{x}}_k h)_i = (\,\delta^{\vec{x}}_k f)_i + (\,\delta^{\vec{x}}_k g)_i + 1$, and therefore

$$\mathrm{Sub}^{\vec{y}}_{i,j}((\,\delta^{\vec{x}}_k h)_i) = \mathrm{Sub}^{\vec{y}}_{i,j}((\,\delta^{\vec{x}}_k f)_i) \cup \mathrm{Sub}^{\vec{y}}_{i,j}((\,\delta^{\vec{x}}_k g)_i)$$

since $i < j$. Thus we may directly apply the induction hypothesis.

Case 3: $h \equiv 2^f \cdot g$. Then $(\,\delta^{\vec{x}}_k h)_i = 2(\,\delta^{\vec{x}}_{k+1} f)_i + (\,\delta^{\vec{x}}_k g)_i + 1$, and we argue as in the previous case.

Case 4: $h \equiv \psi(\omega \cdot f + g)$. Here the case $k \neq i$ is treated as the previous cases, so assume $k = i = 0$, whence $(\,\delta^{\vec{x}}_0 h)_0 = \psi(\omega \cdot f\{\vec{x} := \vec{1}\} + (\,\delta^{\vec{x}}_0 g)_0)$. We therefore have

$$\mathrm{Sub}^{\vec{y}}_{0,j}((\,\delta^{\vec{x}}_0 h)_0) = \mathrm{Sub}^{\vec{y}}_{1,j}(f\{\vec{x} := \vec{1}\}) \cup \mathrm{Sub}^{\vec{y}}_{0,j}((\,\delta^{\vec{x}}_k g)_i)$$

since $j > 0$, and clearly

$$\mathrm{Sub}^{\vec{y}}_{1,j}(f\{\vec{x} := \vec{1}\}) \cup \mathrm{Sub}^{\vec{y}}_{0,j}(g\{\vec{x} := \vec{1}\}) = \mathrm{Sub}^{\vec{y}}_{0,j}(h\{\vec{x} := \vec{1}\}),$$

applying the induction hypothesis for g if necessary. □

Lemma 5.7

(1) *Let $h \in \mathcal{C}_i^{\vec{x}}$, $n := \mathrm{lv}(\vec{x}) + 1$, $j \in (0, n]$, and $\alpha \in (0, \varepsilon_0)$ such that $\overline{t} < \alpha$ for all $t \in \mathrm{T}_{i,j}^{\vec{x}}(h)$. Then we have*

$$\overline{(\,\delta_i^{\vec{x}} h)_j} \leq \mathrm{sz}^{\vec{x}}(h) \cdot \alpha.$$

(2) *Let $h \in \mathcal{C}_0^{\vec{x}}$, $m \in (0, \omega)$, and $\alpha \in (0, \varepsilon_0)$ such that $\overline{t} < m$ for all $t \in \mathrm{T}_{0,0}^{\vec{x}}(h)$ and $\overline{f} < \alpha$ for all $f \in \mathrm{Sub}_{0,1}^{\vec{x}}(h)$. Then we have*

$$\overline{(\,\delta_0^{\vec{x}} h)_0} < \psi(\omega \cdot \mathrm{sz}^{\vec{x}}(h) \cdot \alpha + \mathrm{sz}^{\vec{x}}(h) \cdot m).$$

Proof. Part (1) is proved by induction on the buildup of $h \in \mathcal{C}_i^{\vec{x}}$. If h is x-free or $h \equiv x_i$, the claim follows immediately. Let us assume that h is not x-free. In the case $h \equiv f + g$ the claim directly follows from the induction hypothesis for f and g. If $h \equiv 2^f \cdot g$ or $h \equiv \psi(\omega \cdot f + g)$, we have $\mathrm{T}_{i,j}^{\vec{x}}(h) = \mathrm{T}_{i+1,j}^{\vec{x}}(f) \cup \mathrm{T}_{i,j}^{\vec{x}}(g)$, and straightforwardly apply the induction hypothesis to f and g.

Part (2) is shown by induction on the buildup of $h \in \mathcal{C}_0^{\vec{x}}$ along the definition of $\delta_0^{\vec{x}} h$. If h is x-free, then $h \in \mathrm{T}_{0,0}^{\vec{x}}(h)$, and we have

$$\overline{(\,\delta_0^{\vec{x}} h)_0} = \overline{h} + 1 \leq m < \psi(\omega \cdot \mathrm{sz}^{\vec{x}}(h) \cdot \alpha + \mathrm{sz}^{\vec{x}}(h) \cdot m).$$

Let us now assume that h is not x-free. The case $h \equiv x_0^{\sigma}$ is trivial. If $h \equiv f + g$, then using Proposition 2.3 and the induction hypothesis we may estimate straightforwardly as follows:

$$\begin{aligned}
\overline{(\,\delta_0^{\vec{x}} h)_0} &= \overline{(\,\delta_0^{\vec{x}} f)_0} + \overline{(\,\delta_0^{\vec{x}} g)_0} + 1 \\
&< \psi(\omega \cdot \mathrm{sz}^{\vec{x}}(f) \cdot \alpha + \mathrm{sz}^{\vec{x}}(f) \cdot m) + \psi(\omega \cdot \mathrm{sz}^{\vec{x}}(g) \cdot \alpha + \mathrm{sz}^{\vec{x}}(g) \cdot m) \\
&\leq \psi(\omega \cdot (\mathrm{sz}^{\vec{x}}(f) + \mathrm{sz}^{\vec{x}}(g)) \cdot \alpha + (\mathrm{sz}^{\vec{x}}(f) + \mathrm{sz}^{\vec{x}}(g)) \cdot m) \\
&< \psi(\omega \cdot \mathrm{sz}^{\vec{x}}(h) \cdot \alpha + \mathrm{sz}^{\vec{x}}(h) \cdot m).
\end{aligned}$$

Finally, suppose $h \equiv \psi(\omega \cdot f + g)$. Since $f \in \mathrm{Sub}_{0,1}^{\vec{x}}(h)$, we have $\overline{f} < \alpha$, and since $h \in \mathcal{C}_0^{\vec{x}}$, we have $\mathrm{no}(f) \preceq F_2(g)$. By Lemma 2.19, we obtain

$$\mathrm{no}\left(\overline{f}\right) \leq F_2\left(\overline{g}\right) < F_2\left(\overline{(\,\delta_0^{\vec{x}} g)_0}\right),$$

and, using the induction hypothesis, Proposition 2.3 yields

$$\begin{aligned}
\overline{(\,\delta_0^{\vec{x}} h)_0} &= \psi\left(\omega \cdot \overline{f} + \overline{(\,\delta_0^{\vec{x}} g)_0}\right) \\
&< \psi(\omega \cdot \alpha + \psi(\omega \cdot \mathrm{sz}^{\vec{x}}(g) \cdot \alpha + \mathrm{sz}^{\vec{x}}(g) \cdot m)) \\
&\leq \psi(\omega \cdot (\mathrm{sz}^{\vec{x}}(g) + 1) \cdot \alpha + (\mathrm{sz}^{\vec{x}}(g) + 1) \cdot m) \\
&\leq \psi(\omega \cdot \mathrm{sz}^{\vec{x}}(h) \cdot \alpha + \mathrm{sz}^{\vec{x}}(h) \cdot m),
\end{aligned}$$

concluding the proof of Lemma 5.7. □

Lemma 5.8 *Let $G \in \mathfrak{T}$, $L := \mathrm{L}(G)$, $R := \mathrm{R}(G)$, and for every subterm H of G set $M_H := 2(L + 1 + \mathrm{lh}(H))$ and define a vector $\vec{\alpha}^H$ of level L by*

$$\alpha_i^H := \begin{cases} 2_{L+2\dot{-}i}(M_H \dot{-} i) & \text{if } R < i \leq L, \\ 2_{R\dot{-}i}(\omega \cdot 2_{L+2\dot{-}i}(M_H \dot{-} i)) & \text{if } 1 \leq i \leq R, \\ \psi(\omega \cdot 2_{R\dot{-}1}(\omega \cdot 2_{L+1}(M_H))) & \text{if } i = 0. \end{cases}$$

Then for every subterm H of G with canonical assignment $\vec{h} := [\![H]\!]$ we have

$$\overline{h}_i < \alpha_i^H \tag{5.11}$$

for $i \leq \mathrm{lv}(\vec{h}) =: m$. If $\vec{h} \in \mathcal{C}^{\vec{y}}$, then for all $i \leq m$, all $j \leq L$, and every $t \in \mathrm{Sub}^{\vec{y}}_{i,j}(h_i)$, we have

$$\overline{t} < \alpha_j^H. \tag{5.12}$$

Proof. The lemma is proved by induction on $\mathrm{lh}(H)$.

Case 1: H is a variable or constant. Then the claims follow immediately. The subcase $H \equiv \mathsf{R}$ is where infinite ordinals enter the picture.

Case 2: $H \equiv A^{\sigma\tau} B^{\sigma}$. Then $\vec{h} = \vec{a} \Box \vec{b} \restriction_m$ where $\vec{a}$ and $\vec{b}$ are the canonical assignments to A and B, respectively. Let $n := \mathrm{lv}(\sigma)$. We first show (5.11). If $i > n$, we have $\overline{h}_i \leq \overline{a}_i$, and the claim follows by the induction hypothesis applied to A. Suppose $i \leq n$, whence $i < L$. We argue by side induction on $n \dot{-} i$.

- $L > i > R$: Then $0 < i \leq n$ and hence

$$\begin{aligned} \overline{h}_i &\leq 2^{\overline{h}_{i+1}} \cdot (\overline{a}_i + \overline{b}_i) \\ &< 2_{L+2\dot{-}i}(M_H \dot{-} (i+1)) \cdot (2_{L+2\dot{-}i}(M_A \dot{-} i) + 2_{L+2\dot{-}i}(M_B \dot{-} i)) \\ &< (2_{L+2\dot{-}i}(M_H \dot{-} (i+1)))^2 \\ &< 2_{L+2\dot{-}i}(M_H \dot{-} i). \end{aligned}$$

- $R \geq i \geq 1$: Suppose first that $i = R$. Then we have

$$\begin{aligned} \overline{h}_i &< 2_{L+2\dot{-}i}(M_H \dot{-} (i+1)) \cdot (\omega \cdot 2_{L+2\dot{-}i}(M_A \dot{-} i) + \omega \cdot 2_{L+2\dot{-}i}(M_B \dot{-} i)) \\ &< 2_{L+2\dot{-}i}(M_H \dot{-} (i+1)) \cdot (\omega \cdot 2_{L+2\dot{-}i}(M_H \dot{-} (i+1))) \\ &< \omega \cdot 2_{L+2\dot{-}i}(M_H \dot{-} i). \end{aligned}$$

  We now consider the case $i < R$, where we estimate as follows.

$$\begin{aligned} \overline{h}_i &< 2_{R\dot{-}i}(\omega \cdot 2_{L+2\dot{-}(i+1)}(M_H \dot{-} (i+1))) \\ &\qquad \cdot (2_{R\dot{-}i}(\omega \cdot 2_{L+2\dot{-}i}(M_A \dot{-} i)) + 2_{R\dot{-}i}(\omega \cdot 2_{L+2\dot{-}i}(M_B \dot{-} i))) \\ &< 2_{R\dot{-}i}(\omega \cdot 2_{L+2\dot{-}(i+1)}(M_H \dot{-} (i+1))) \\ &\qquad \cdot 2_{R\dot{-}i}(\omega \cdot 2_{L+2\dot{-}i}(M_H \dot{-} (i+1))) \\ &< 2_{R\dot{-}i}(\omega \cdot 2_{L+2\dot{-}i}(M_H \dot{-} i)). \end{aligned}$$

- $i = 0$: In case of $R = 0$, we obtain, using Proposition 2.3,
$$\begin{aligned}\overline{h}_0 &< \psi(\omega \cdot 2_{L+1}(M_H \dot{-} 1) + \psi(\omega^2 \cdot 2_{L+1}(M_A)) + \psi(\omega^2 \cdot 2_{L+1}(M_B)) + n)\\ &< \psi(\omega^2 \cdot 2_{L+1}(M_H)).\end{aligned}$$
If $R > 0$, we even have $R \geq 2$, and using again Proposition 2.3 we obtain
$$\begin{aligned}\overline{h}_0 &< \psi(\omega \cdot 2_{R \dot{-} 1}(\omega \cdot 2_{L+1}(M_H \dot{-} 1))\\ &\qquad + \psi(\omega \cdot 2_{R \dot{-} 1}(\omega \cdot 2_{L+1}(M_A))) + \psi(\omega \cdot 2_{R \dot{-} 1}(\omega \cdot 2_{L+1}(M_B))) + n)\\ &< \psi(\omega \cdot 2_{R \dot{-} 1}(\omega \cdot 2_{L+1}(M_H))).\end{aligned}$$

This finishes the verification of (5.11), and we proceed with proving (5.12). If $\vec{h} \in \mathcal{C}^{\vec{y}}$, then we must also have $\vec{a}, \vec{b} \in \mathcal{C}^{\vec{y}}$ and may apply the respective induction hypotheses. Suppose $t \in \mathrm{Sub}^{\vec{y}}_{i,j}(h_i)$, whence $i \leq j$, as the set $\mathrm{Sub}^{\vec{y}}_{i,j}(h_i)$ is empty if $i > j$. In order to show $\overline{t} < \alpha^H_j$, we employ an induction on $j \dot{-} i$. If $i = j$, we clearly have $\overline{t} \leq \overline{h_i} < \alpha^H_i$ by (5.11). Suppose $i < j$. In the case $i > n$, we use the induction hypothesis applied to A. If on the other hand $i \leq n$, then we have
$$t \in \mathrm{Sub}^{\vec{y}}_{i+1,j}(h_{i+1}) \cup \mathrm{Sub}^{\vec{y}}_{i,j}(a_i) \cup \mathrm{Sub}^{\vec{y}}_{i,j}(b_i),$$
and in each case $\overline{t} < \alpha^H_j$ follows using the induction hypothesis since clearly $\alpha^A_j, \alpha^B_j < \alpha^H_j$.

Case 3: $H \equiv \lambda X^\sigma.F^\tau$. Then $\vec{h} = \delta^{\vec{x}}\vec{f} + \vec{f}\{\vec{x} := \vec{1}\}$ where $\vec{f} := [\![f]\!]$, and $m = \max\{n, l\}$ where $n := \mathrm{lv}(\vec{x}) + 1$ and $l = \mathrm{lv}(\vec{f})$. We first show (5.11), where we distinguish between the following three cases.

- $n < i \leq m$. Then we have $h_i = 2f_i$, and the claim follows easily from the induction hypothesis for F.
- $1 \leq i \leq n$. By part (1) of Lemma 5.7 and the induction hypothesis for F, we have
$$\overline{(\delta^{\vec{x}}_k f_k)_i} \leq \mathrm{sz}^{\vec{x}}(f_k) \cdot \alpha^F_i$$
for every $k \leq l$, which yields
$$\overline{h}_i \leq \left(\sum_{k=0}^{l} \mathrm{sz}^{\vec{x}}(f_k) + 1\right) \cdot \alpha^F_i \leq 2_2(L + 2\mathrm{lh}(F)) \cdot \alpha^F_i$$
by (5.8) of Lemma 5.5. In case of $i > R$, we may now estimate
$$\begin{aligned}\overline{h}_i &\leq 2_2(L + 2\mathrm{lh}(F)) \cdot 2_{L+2 \dot{-} i}(M_F \dot{-} i)\\ &< (2_{L+2 \dot{-} i}(M_F \dot{-} i))^2\\ &< 2_{L+2 \dot{-} i}(M_H \dot{-} i)\\ &= \alpha^H_i,\end{aligned}$$
whereas in case of $i \leq R$ we estimate
$$\begin{aligned}\overline{h}_i &\leq 2_2(L + 2\mathrm{lh}(F)) \cdot 2_{R \dot{-} i}(\omega \cdot 2_{L+2 \dot{-} i}(M_F \dot{-} i))\\ &< 2_{R \dot{-} i}(\omega \cdot 2_{L+2 \dot{-} i}(M_H \dot{-} (i+1)))\\ &< \alpha^H_i.\end{aligned}$$

- $i = 0$. Since in the case $R = 0$ the argumentation is easier, let us assume that $R > 0$. By Lemma 5.5, we have

  $$\mathrm{sz}^{\vec{x}}(f_0), \mathrm{S}^{\vec{x}}(\vec{f}) < 2_2(L + 1 + 2\mathrm{lh}(F)) =: K,$$

  and relying on the induction hypothesis for F we obtain, using part (2) of Lemma 5.7,

  $$\overline{(\,\delta_0^{\vec{x}} f_0)_0} < \psi(\omega \cdot \mathrm{sz}^{\vec{x}}(f_0) \cdot \alpha_1^F + \mathrm{sz}^{\vec{x}}(f_0) \cdot \alpha_0^F),$$

  which, using Proposition 2.3, allows for the following estimation:

  $$\begin{aligned} \overline{h}_0 &= \mathrm{S}^{\vec{x}}(\vec{f}) \cdot \overline{(\,\delta_0^{\vec{x}} f_0)_0} + \overline{f}_0 \\ &< K \cdot \psi(\omega \cdot K \cdot \alpha_1^F + K\alpha_0^F) \\ &\leq \psi(\omega \cdot K^2 \cdot \alpha_1^F + K^2\alpha_0^F) \\ &< \psi(\omega \cdot 2K^2 \cdot 2_{R \dot{-} 1}(\omega \cdot 2_{L+1}(M_F))) \\ &< \psi(\omega \cdot 2_{R \dot{-} 1}(\omega \cdot 2_{L+1}(M_H))) \\ &= \alpha_0^H. \end{aligned}$$

In order to verify (5.12), suppose $t \in \mathrm{Sub}_{i,j}^{\vec{y}}(h_i)$. If $i = j$, we obtain $\overline{t} \leq \overline{h}_i < \alpha_i^H$ from (5.11) by the monotonicity properties of $+, \cdot$, and ψ. Assume $i < j$. Then we either have $t \in \mathrm{Sub}_{i,j}^{\vec{y}}(f_i\{\vec{x} := \vec{1}\})$ where $i \leq l$, or $i \leq n$ and

$$t \in \mathrm{Sub}_{i,j}^{\vec{y}}((\,\delta_k^{\vec{x}} f_k)_i)$$

for some $k \leq l$, which is 0 in the case $i = 0$, and Lemma 5.6 yields $t \in \mathrm{Sub}_{k,j}^{\vec{y}}(f_k\{\vec{x} := \vec{1}\})$, hence $k \leq j$. If $\vec{y} \equiv \vec{x}$, we must have $k = j$, $t \equiv f_j\{\vec{x} := \vec{1}\}$, and therefore $\overline{t} \leq \alpha_j^F$ by the induction hypothesis for F. Now assume $\vec{y} \not\equiv \vec{x}$. Then we have

$$\mathrm{Sub}_{k,j}^{\vec{y}}(f_k\{\vec{x} := \vec{1}\}) = \mathrm{Sub}_{k,j}^{\vec{y}}(f_k)\{\vec{x} := \vec{1}\},$$

and $\overline{t} \leq \alpha_j^F$ follows from the induction hypothesis for F. The case $t \in \mathrm{Sub}_{i,j}^{\vec{y}}(f_i\{\vec{x} := \vec{1}\})$ is treated in the same way. □

Acknowledgements

The authors would like to thank Roger Hindley and Jonathan Seldin for informative discussions as well as Pierre-Louis Curien for constructive assistance in the rewrite process of the original contribution to TLCA 2009. We are indebted to one of the referees who pointed out our sloppiness regarding the treatment of bound variables in an earlier version of the paper and gave many constructive suggestions that have led to an improvement of the first section of this paper. We wish to thank William Howard for a number of helpful suggestions and simplifications for the final version of the paper. Finally, we express our gratitude to the John Templeton Foundation for its support through Project #13152, The Myriad Aspects of Infinity.

References

[1] A. Beckmann, A. Weiermann, A term rewriting characterization of the polytime functions and related complexity classes, *Archive for Mathematical Logic* **36** (1996), 11–30.

[2] A. Beckmann, A. Weiermann, Analyzing Gödel's T via expanded head reduction trees, *Mathematical Logic Quarterly* **46** (2000), 517–536.

[3] W. Buchholz, E. A. Cichon, A. Weiermann, A uniform approach to fundamental sequences and hierarchies, *Mathematical Logic Quarterly* **40** (1994) 273–286.

[4] F. Cardone, J. R. Hindley, Lambda-calculus and combinators in the 20th century, *Logic from Russell to Church*, Handbook of the History of Logic, 5, Elsevier, 2009.

[5] E. A. Cichon, A. Weiermann, Term rewriting theory for the primitive recursive functions, *Annals of Pure and Applied Logic* **83** (1997), 199–223.

[6] K. Gödel, Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes, *Dialectica* **12** (1958), 280–287.

[7] J. R. Hindley, J. P. Seldin, *Introduction to Combinators and λ-Calculus*, London Mathematical Society, Cambridge University Press, 1986.

[8] W. A. Howard, Assignment of ordinals to terms for primitive recursive functionals of finite type. *Intuitionism and Proof Theory*, North-Holland, Amsterdam, 1970, 443–458.

[9] C. Parsons, On n-quantifier induction, *The Journal of Symbolic Logic* **37** (1972), 466–482.

[10] W. Pohlers, *Proof Theory. The First Step into Impredicativity*, Springer, Berlin, 2009.

[11] K. Schütte, *Proof Theory*, Springer, 1977.

[12] J. R. Shoenfield, *Mathematical Logic*, Addison-Wesley, New York, 1967.

[13] A. Weiermann, How to characterize provably total functions by local predicativity, *The Journal of Symbolic Logic* **61** (1996), 52–69.

[14] A. Weiermann, A proof of strongly uniform termination for Gödel's T by methods from local predicativity, *Archive for Mathematical Logic* **36** (1997), 445–460.

[15] A. Weiermann, How is it that infinitary methods can be applied to finitary mathematics? Gödel's T: a case study, *The Journal of Symbolic Logic* **63** (1998), 1348–1370.

[16] G. Wilken, A. Weiermann, Derivation lengths classification of Gödel's T extending Howard's assignment, *Logical Methods in Computer Science* **8** (2012), 1–44.