Ways of Proof Theory

Wolfram Pohlers

Preface

Wolfram Pohlers is one of the leading researchers in the proof theory of ordinal analysis. On the occasion of his retirement the Institut für Mathematische Logik und Grundlagenforschung of the University of Münster organized a colloquium and a workshop which took place July 17 – 19, 2008. This event brought together proof theorists from many parts of the world who have been acting as teachers, students and collaborators of Wolfram Pohlers and who have been shaping the field of proof theory over the years.

The organizer of the colloquium and workshop gratefully acknowledges financial support from the University of Münster, the DVMLG (the German Logic Society), and Springer–Verlag. The present volume collects papers by the speakers of the colloquium and workshop; and they produce a documentation of the state of the art of contemporary proof theory.

We thank Martina Pfeifer and Jan–Carl Stegert for helping us organize the colloquium and workshop and produce this volume. We dedicate this volume to Wolfram Pohlers, who has always been an inspiring mathematician, an extraordinary colleague, and a great friend.

Münster, June 01, 2010 Ralf Schindler

Contents

Wolfram Pohlers — Life and Work

Justus Diller

Coming from Munich in 1985, Wolfram Pohlers followed a call to the department of mathematics of the University of Münster. From that time to his retirement in the summer of 2008, he occupied the chair of Heinrich Scholz, the first chair in mathematical logic and foundational research in German speaking Central Europe. During this period, he represented the Bavarian element in our department, by accent as well as by temper. For somebody born in Leipzig to a Saxonian father and a Norwegian mother, that may seem somewhat surprising. He proved his character by energetic engagement in many fields, by stubbornness — which is said to be characteristic for Westfalians, too — showing a definite conservative tendency over the years, and by an ability to compromise resulting out of his respect for his partners in negotiations — which is not a matter of course among mathematicians. We review his activities in administration, scientific organization, and science in due brevity.

Wolfram Pohlers served on several committees of our department. He was our dean for two years, from 1990 to 1992, and he was our first dean to hold that job for the full period of two consecutive years. Until then, we had not requested of each other to carry the dean's burden for so long. In his period of office, he, among other things, resuscitated the deans' conference of our alma mater.

Since then, he represented his colleagues in the senate of our university for about 14 years, and he was speaker of professors in the senate for a considerable number of years. Such a position naturally brings about a close, but also time consuming cooperation with the university administration. It was obviously a consequence of that positive cooperation that the rector of the university of Münster, Prof. Ursula Nelles, seized the opportunity to address the conference in honor of Pohlers' retirement in person.

Membership in the senate brought with it tasks of which most of us have never heard, for example in university sports. The central unit university sport is a large organization which moves considerable amounts of money. The steering of this unit lies in the hands of a steering committee, and Wolfram Pohlers presided over this committee for many years. The last meeting over which he presided must have been a moving farewell party. On the other hand, it seems almost a matter of course that for many years he also worked on the university's central IV– (information processing–) committee. Even more important tasks come up, when the university looks for a new chancellor or a new rector. Such a person is not found spontaneously; rather, she or he is looked for by a search committee. If we add to these the committee for the delicate preselection of candidates for the university council — Hochschulrat, a product of recent legislation of our state — we have the impression that within his last five to eight years on our senate, Wolfram Pohlers has been on every such committee of our university.

Even a science with a small faculty like mathematical logic has its own national scientific union. Wolfram Pohlers has served many years on the board of the DVMLG, the German logic union. He is also a member of the editorial staff of the two journals on mathematical logic that appear in Germany, the Mathematical Logic Quarterly and the Archive for Mathematical Logic. Of the latter he was editor–in–chief until 2008 when he passed that job on to our colleague Ralf Schindler. In recent years, he is also active as scientific area editor for the Journal of Applied Logic which, following trends of our time, appears in Singapore.

A noticeable event was the European summer meeting 2002 of the ASL, the association of symbolic logic, which Pohlers organized in Münster, together with the president of the DVMLG, Professor Koepke from Bonn. With more than 200 participants from all over the world, this summer meeting was a big conference, considering the standards of our department in those days. Our Institute as a whole was for quite a while pretty busy with the preparation and implementation of the conference. With such determined engagement of manpower and resources, it proved to be an advantage that once upon a time Wolfram Pohlers had been officer of our federal army. Smaller workshops on proof theory were also organized under his supervision.

To a large part, these fruitful activities would not have taken place in Münster, if, in 1995, he had followed a call to the university of Vienna. After some inner struggle he turned down this honourable offer. The "old" logic institute of the 1990's maintains deeply felt gratitude to him for his staying in Münster.

All this is only the outer framework for his central activity, which is research and teaching of mathematical logic. Wolfram Pohlers graduated from high school in Munich in 1964 where, after two years of military service, he began his studies of mathematics at the Ludwig-Maximilians University. He married his wife Renate in 1970, and passed his diploma in mathematics in March 1971. The day after he had completed his diploma, he started work as scientific assistant with Kurt Schütte with whom he completed his dissertation in mathematical logic in 1973. The area of research from which the topic of his dissertation was taken was to become his research field for all of his career: it is the proof–theoretic field of ordinal analysis, a central topic in the foundations of mathematics. We cast a quick glance at what ordinal analysis is about, and what Wolfram Pohlers has contributed to this field in the last 39 years.

Gödel's second incompleteness theorem of 1931 showed that the original goal of Hilbert's program was unattainable: a mathematically relevant theory like Peano

Arithmetic (elementary number theory) PA cannot prove its own consistency. Already in 1936, however, Gerhard Gentzen found a way out of this dilemma. He proved the consistency of PA employing a transfinite induction up to the ordinal number ϵ_0 in an otherwise finitary, completely combinatorial proof. Here, ϵ_0 is the limit of iterated ω–powers, the first fixed point of the function ω^α, i.e. the smallest ordinal α such that $\omega^\alpha = \alpha$. (ω designates the first transfinite ordinal.) By his proof, he had isolated transfinite induction up to ϵ_0 as the transfinite feature which transcends the means of PA. He had thus proved the consistency of PA by constructive, though not finitary methods. In short: he had shown ϵ_0 to be the proof–theoretic ordinal of PA. By this proof, Gentzen had started a revised version of Hilbert's program which until today plays a central role in the foundations of mathematics. In the 1950's and 1960's, Kurt Schütte, Gaisi Takeuti, and Solomon Feferman began to tackle stronger mathematical systems. Feferman and Schütte worked in particular on predicative analysis which allows quantification over sets of natural numbers, however, only in a strictly constructive, so–called predicative way. They proved the first strongly critical ordinal Γ_0, the first common fixed point of the Veblen hierarchy, to be the proof–theoretic ordinal of predicative analysis.

After this success, impredicative systems of classical mathematics moved into focus. These were, on the one hand, theories of inductive definitions. In this area, Howard 1972 proved the so–called Howard–Bachmann ordinal to be the proof–theoretic ordinal for one inductive definition, and Pohlers 1978 made an ordinal analysis of iterated inductive definitions. On the other hand, there were subsystems of classical analysis, i.e. of second order number theory. To the ordinal analysis of these, Pohlers' dissertation of 1973 made an important contribution.

The study of both of these areas of research and their interrelations came to some completion, when, in 1980, the authors Buchholz, Feferman, Pohlers, and Sieg published the volume "Iterated inductive definitions and subsystems of analysis: Recent proof–theoretical studies" in the Lecture notes in mathematics. It was coordinated by Solomon Feferman and based on the habilitation theses of Pohlers and Bucholz and the PhD thesis of Sieg. In this volume, Pohlers developed his method of local predicativity which presented an essential progress in the ordinal analysis of stronger and stronger impredicative systems. As Gentzen succeeded in isolating the transfinite element in first order number theory, Pohlers' method of local predicativity allows the isolation of the impredicative elements of strong theories. His method simplifies the still troublesome computations of the corresponding proof theoretic ordinals considerably.

Given fresh impetus by this new method, proof theorists now also attacked systems of set theory. While in 1950, Heinz Bachmann was the first to make use of an uncountable number, i.e. $\aleph_1$, to denote countable ordinals, since 1979 hardly any large cardinal was safe from the grip of ordinal analysts. The hunt was to a large part led by students of Schütte and Pohlers. Gerhard Jäger in Munich, now Bern, was the first to make use of inaccessible cardinals, and around 1990, Michael Rathjen in Münster, now Leeds, proceeded to Mahlo and other large cardinals. Rathjen finally succeeded in an ordinal analysis of Π^1_2–analysis, a theory much stronger than methods used by classical mathematicians outside measure theory.

The ambition of the Münster school of proof theory — that is an established term meanwhile — does not only go to still larger, still more complex systems. It also aims at restructuring the already analyzed terrain, at including constructive theories, at applications to other areas of mathematics, computer science, and logic. Michael Rathjen has meanwhile included constructive systems like Martin–Löf theory and constructive set theory in his proof theoretic analysis. Andreas Weiermann, now in Gent, discovered deep connections between ordinal analysis and pure mathematics. For instance, he proved stunning results in proof theory by methods of analytic number theory. Also lower complexities were analyzed. Theories relevant in bounded arithmetic satisfy the conditions of Gödel's second incompleteness theorem only in a restricted sense, and their classical proof theoretic ordinal is ω^2 in all relevant cases. Arnold Beckmann, now in Swansea, developed so–called dynamic ordinals which allow to distinguish between the proof theoretic strengths of some of these theories. They are not ordinals in the classical sense, they may be viewed as cloudy objects assembled around ω. It would be a triumph for proof theory, if they could be used to separate the right theories of bounded arithmetic according to their proof theoretic strength. For that would shed light on the P/NP–problem, the fundamental problem of theoretical computer science, and it would yield the desired answer $P \neq NP$.

Finally, Michael Möllerfeld developed a recursion theory of Π^1_2-analysis. Thus, after set theory, constructivism, and complexity theory, one more field of mathematical logic could be brought into contact with the proof theoretic subject of ordinal analysis.

Wolfram Pohlers accompanied this drive to new frontiers in many ways, in recent years in particular with systematizing publications. These include his "Proof Theory, An Introduction", but even more so his thorough survey chapter "Subsystems of set theory and second order number theory" in the Handbook of Proof Theory and his recent book "Proof Theory: The First Step into Impredicativity".

Adding up, Wolfram Pohlers as a researcher has contributed substantially to the field of ordinal analysis; as an academic teacher he has founded the Münster school of proof theory. To keep this group and other friends in contact, the Pohlers couple regularly arranges a summer party at their home in Nienberge. Only on rare occasions at these parties, an old double bass comes into appearance which many years ago helped Wolfram Pohlers finance his student days in Munich and which, during a rafting tour down the Isar, is said to have gone down the river part of the way on its own. Well, out of his school of proof theory, there emerged a number of scientists who work in several countries and have produced remarkable results. Ordinal analysis has become more complex under Pohlers' influence, but it is also rather more vital and more applicable to neighbouring fields than could have been expected thirty years ago.

The Proof Theory of Classical and Constructive Inductive Definitions. A Forty Year Saga, 1968 – 2008

Solomon Feferman*

1 Pohlers and The Problem

I first met Wolfram Pohlers at a workshop on proof theory organized by Walter Felscher that was held in Tübingen in early April, 1973. Among others at that workshop relevant to the work surveyed here were Kurt Schütte, Wolfram's teacher in Munich, and Wolfram's fellow student Wilfried Buchholz. This is not meant to slight in the least the many other fine logicians who participated there.[1] In Tübingen I gave a couple of survey lectures on results and problems in proof theory that had been occupying much of my attention during the previous decade. The following was the central problem that I emphasized there:

> *The need for an ordinally informative, conceptually clear, proof-theoretic reduction of classical theories of iterated arithmetical inductive definitions to corresponding constructive systems.*

As will be explained below, meeting that need would be significant for the then ongoing efforts at establishing the constructive foundation for and proof-theoretic ordinal analysis of certain impredicative subsystems of classical analysis. I also spoke in Tübingen about possible methods to tackle the central problem, including both cut-elimination applied to (prima-facie) uncountably infinite derivations

*This is a somewhat revised text of a lecture that I gave for a general audience at the PohlersFest, Münster, 18 July 2008 in honor of Wolfram Pohlers, on the occasion of his retirement from the Institute for Mathematical Logic at the University of Münster. Wolfram was an invited participant at a conference in my honor at Stanford in 1998, and it was a pleasure, in reciprocation, to help celebrate his great contributions as a researcher, teacher and expositor. In my lecture I took special note of the fact that the culmination of Wolfram's expository work with his long awaited *Proof Theory* text was then in the final stages of production; it has since appeared as Pohlers (2009). In that connection, one should mention the many fine expositions of proof theory that he had previously published, including Pohlers (1987, 1989, 1992, and 1998).

[1] That meeting was organized by Walter Felscher under the sponsorship of the Volkswagen Stiftung; there were no published proceedings. It is Pohlers' recollection that besides him and Felscher, of course, the audience included Wilfried Buchholz, Justus Diller, Ulrich Felgner, Wolfgang Maas, Gert Müller, Helmut Pfeiffer, Kurt Schütte and Helmut Schwichtenberg. By the way, Felscher passed away in the year 2000.

and functional interpretation on the one hand, and the use of naturally developed systems of ordinal notation on the other. I recall that my wife and I had driven to Tübingen that morning from Oberwolfach after an unusually short night's sleep, and that I was going on pure adrenalin, so that my lectures were particularly intense. Perhaps this, in addition to the intrinsic interest of the problems that I raised, contributed to Wolfram's excited interest in them. Within a year or so he made the first breakthrough in this area (Pohlers 1975), which was to become the core of his Habilitationsschrift with Professor Schütte (Pohlers 1977). The 1975 breathrough was the start of a five year sustained effort in developing a variety of approaches to the above problem by Wolfram Pohlers, Wilfried Buchholz and my student Wilfried Sieg. The results of that work were jointly reported in the Lecture Notes in Mathematics volume 897, *Iterated Inductive Definitions and Subsystems of Analysis. Recent proof-theoretical studies* (Buchholz et al. 1981). In the next section I will give a brief review of what led to posing the above problem in view of several results by Harvey Friedman, William Tait and me at the 1968 Buffalo conference on intuitionism and proof-theory, with some background from a 1963 seminar on the foundations of analysis led by Georg Kreisel at Stanford in which formal theories of "generalized" inductive definitions (i.e., with arithmetical closure conditions) were first formulated.

The goals of proof-theoretic reduction and of proof-theoretic ordinal analysis in one form or another of the relativized Hilbert program (not only for theories of inductive definitions) are here taken at face value, though I have examined both critically; see Feferman (1988, 1993, 2000). In addition to meeting those aims in the problem formulated above are the demands that the solutions be informative and conceptually clear in short, perspicuous. Granted that these are subjective criteria, nevertheless in practice we are able to make reasonably objective judgments of comparison. For example, we greatly valued Schütte's extension of Gentzen's cut-elimination theorem for the predicate calculus to "semi-formal" systems with infinitary rules of inference, because it exhibited a natural and canonical role for ordinals as lengths of derivations and bounds of cut-rank (cf. Schütte 1977) in the case of arithmetic and its extensions to ramified analysis. To begin with, the Cantor ordinal $\varepsilon 0$ emerged naturally as the upper bound of the lengths of cut-free derivations in the semi-formal system of arithmetic with ω-rule, obtained by eliminating cuts from the (translations into that system of) proofs in Peano Arithmetic PA; by comparison the role of $\varepsilon 0$ in Gentzen's consistency proof of PA still had an ad hoc appearance.[2] And the determination by Schütte and me in the mid 1960s of $\Gamma 0$ as the upper bound for the ordinal of predicativity simply fell out of his ordinal anal-

[2]That role became less mysterious as a result of the work of Buchholz (1997, 2001) explaining Gentzen-style and Takeuti-style reduction steps in infinitary terms.

ysis of the systems of ramified analysis translated into infinitary rules of inference when one added the condition of autonomy. Incidentally, because of the connection with predicativity, these kinds of proof-theoretical methods due to Schütte — of ordinal analysis via cut-elimination theorems for semi-formal systems with countably infinitary rules of inference — have come to be referred to as predicative.

The proof-theoretical work on systems of single and (finitely or transfinitely) iterated arithmetical inductive definitions were the first challenges to obtaining perspicuous ordinal analyses and constructive reductions of impredicative theories. The general problem was both to obtain exact bounds on the provably recursive ordinals and to reduce inductive definitions described ”from above” as the least sets satisfying certain arithmetical closure conditions to those constructively generated “from below”. In the event, the work on these systems took us only a certain way into the impredicative realm, but the method of local predicativity for semi-formal systems with uncountably infinitary rules of inference that Pohlers developed to deal with them turned out to be of wider application. What I want to emphasize in the following is, first of all, that ordinal analysis and constructive reduction are separable goals and that in various cases, each can be done without the other, and, secondly, that the aim to carry these out in ever more perspicuous ways has led to recurrent methodological innovations. The most recent of these is the application of a version of the method of functional interpretation to theories of inductive definitions by Avigad and Towsner (2008), following a long period in which cut-elimination for various semi-formal systems of uncountably infinitary derivations had been the dominant method, and which itself evolved methodologically with perspicuity as the driving force. It is not possible in a survey of this length — and at the level of detail dictated by that — to explain or state results in full; for example, I don’t state conservation results that usually accompany theorems on proof-theoretical reduction. Nor is it possible to do justice to all the contributions along the way, let alone all the valuable work on related matters. For example, except for a brief mention in sec. 7 below, I don’t go into the extensive proof-theoretical work on iterated fixed point theories. I hope the interested reader will find this survey useful both as an informative overview and as a point of departure to pursue in more detail not only the topics discussed but also those that are only indicated in passing. Finally, this survey offers an opportunity to remind one of open questions and to raise some interesting new ones.

2 From 1968 to 1981, with some prehistory

In my preface, Feferman (1981), to Buchholz et al. (1981), I traced the developments that led up to that work; in this section I’ll give a brief summary of that material.

The consideration of formal systems of "generalized" inductive definitions originated with Georg Kreisel (1963) in a seminar that he led on the foundations of analysis held at Stanford in the summer of 1963.[3] Kreisel's aim there was to assess the constructivity of Spector's consistency proof of full second-order analysis (Spector 1962) by means of a functional interpretation in the class of so-called bar recursive functionals. The only candidate for a constructive foundation of those functionals would be the hereditarily continuous functionals given by computable representing functions in the sense of (Kleene 1959) or (Kreisel 1959). So Kreisel asked whether the intuitionistic theory of inductive definitions given by monotonic arithmetical closure conditions, denoted $ID_1(mon)^i$ below, serves to generate the class of (indices of) representing functions of the bar recursive functionals. Roughly speaking, $ID_1(mon)$, whether classical or intuitionistic, has a predicate PA for each arithmetic $A(P, x)$ (with a placeholder predicate symbol P) which has been proved to be monotonic in P, together with axioms expressing that PA is the least predicate definable in the system that satisfies the closure condition $\forall x(A(P, x) \rightarrow P(x))$. In the event, Kreisel showed that the representing functions for bar recursive functionals of types ≤ 2 can be generated in an $ID_1(mon)^i$ but not in general those of type ≥ 3.

Because of this negative result, Kreisel did not personally pursue the study of theories of arithmetical inductive definitions any further, but he did suggest consideration of theories of finitely and transfinitely iterated such definitions as well as special cases involving restrictions on the form of the closure conditions $A(P, x)$. For example, those A in which the predicate symbol P has only positive occurrences are readily established to be monotonic in P. And of special interest among such A are those that correspond to the accessible (i.e., well-founded part) of an arithmetical relation. And, finally, paradigmatic for those are the classes of recursive ordinal number classes O_α introduced in Church and Kleene (1936) and continued in Kleene (1938). The corresponding formal systems for α times iterated inductive definitions (α an ordinal) are denoted (in order of decreasing generality) $ID_\alpha(mon), ID_\alpha(pos), ID_\alpha(acc)$ and $ID_\alpha(O)$ in both classical and intuitionistic logic, where the restriction to the latter is signalled with a superscript 'i'.[4] For limit ordinals λ we shall also be dealing with $ID_{<\lambda}(-)$, the union of the $ID_\lambda(-)$ for $\alpha < \lambda$, of each of these kinds, whether classical or intuitionistic. Finally, when no qualification of ID_α or $ID_{<\lambda}(-)$ is given, it is meant that we are dealing with the corresponding $ID_\alpha(pos)$ or $ID_{<\lambda}(pos)$, since — as will be explained in sec. 5 below — there is a relatively easy reduction of the monotonic case to the positive case.

[3]The notes for that seminar are assembled in the unpublished volume Seminar on the Foundations of Analysis, Stanford University 1963. Reports, of which only a few mimeographed copies were made; one copy is available in the Mathematical Sciences Library of Stanford University.

[4]The positivity requirement has to be modified in the case of intuitionistic systems.

The $ID_\alpha(O)$ theories, or similar ones for constructive tree classes, are of particular interest, because the elements of those classes wear their build-up on their sleeves, i.e. can be retraced constructively; some of the $ID_\alpha(acc)$ classes considered below share that significant feature.

Kreisel's initiative led one to study the relationship between such theories to subsystems of classical analysis considered independently of Spector's approach and as the subject of proof-theoretical investigation in their own right. The first such result was obtained by William Howard some time around 1965, though it was not published until 1972. He showed in Howard (1972) that the proof-theoretic ordinal of $ID_1(acc)^i$ is $\varphi\varepsilon_{(\Omega+1)}0$, as measured in the hierarchy of normal functions introduced in Bachmann (1950). Howard's method of proof proceeded via an extension of Gödel's functional interpretation. This was the first ordinally informative characterization of an impredicative system using a system of ordinal notation based on a natural system of ordinal functions. What was left open by Howard's work was whether one could obtain a reduction of the general classical ID_1 to $ID_1(acc)^i$ (and even better to $ID_1(O)^i$) and thus show that the proof-theoretic ordinal is the same, and similarly for the systems of iterated inductive definitions more generally.[5]

Turning now to the 1968 Buffalo Conference on Intuitionism and Proof Theory, here, in brief, is what was done in the three papers I mentioned above.

1. (Friedman 1970) proved that system the $\Sigma^1_{n+1} - AC$ is of the same strength as $\Delta^1_{n+1} - AC$ and is conservative over $(\Pi^1_n - CA)_{<\varepsilon_0}$ for suitable classes of sentences. For $n = 1$ this tied up with the following two results:

2. (Feferman 1970) gave an interpretation of $(\Pi^1_1 - CA)_\alpha$ in ID_α for various α, including $\alpha = \omega$, and of $(\Pi^1_1 - CA)_{<\lambda}$ in $ID_{<\lambda}$ for various limit λ, including $\lambda = \varepsilon_0$.[6]

3. (Tait 1970) established the consistency of $\Sigma^1_2 - AC$ via a certain theory of inductive definitions by informally constructive cut-elimination methods applied to uncountably long propositional derivations.

These results and the prior work of Takeuti (1967) containing constructive proofs of consistency of $(\Pi^1_1 - CA)$ and $(\Pi^1_1 - CA) + BI$ gave hope that one could obtain a constructive reduction of some of the above second order systems via a reduction of classical theories of iterated inductive definitions to their intuitionistic counterparts.[7] For, among the results of my Buffalo conference article was that the

[5] As will be explained in sec. 6, below, Zucker (1971, 1973) showed the ordinals to be the same without a reduction argument and by a method that did not evidently extend to the iterated case.

[6] Actually, the interpretation took one into iterated classical accessibility IDs.

[7] BI is the scheme of Bar Induction, i.e. the implication from well-foundedness to transfinite induction.

system $(\Pi^1_1 - CA) + BI$ is prooftheoretically equivalent to ID_ω. What Takeuti had done was to carry out his consistency proofs by an extension of Gentzen's methods with cut-reduction steps measured in certain partially ordered systems that Takeuti called ordinal diagrams; these are not based on natural systems of ordinal functions such as those in the Bachmann hierarchy. Takeuti proved the well-foundedness of the ordering of ordinal diagrams by constructive arguments that could be formulated in suitable intuitionistic iterated accessible IDs. These methods were later extended to $(\Delta^1_2 - CA) + BI$ in Takeuti and Yasugi (1973).

Before proceeding, a few words are necessary about the systems of ordinal functions involved in proof-theoretic ordinal analysis at that time and in subsequent work. Bachmann had extended the classical Veblen hierarchy φ_α (or $\lambda\alpha, \beta.\varphi_\alpha(\beta)$) of critical functions of countable ordinals by use of indices α to certain uncountable ordinals — including those up to the first ϵ-number greater than Ω — by diagonalizing at α of cofinality Ω, e.g. defining $\varphi\Omega\beta$ to be $\varphi\beta 0$. This method was carried out systematically by Helmut Pfeiffer (1964) by reference to the finite ordinal number classes whose initial ordinals are the Ω_n for $n < \omega$, and then by David Isles (1970) via the number classes up to the first inaccessible ordinal. Each such extension required more and more complicated assignment of fundamental sequences to the ordinals actually drawn from each number class. In 1970, in informal discussions with Peter Aczel, I proposed an alternative method of generating the requisite ordinals and associated functions θ_α in place of the φ_α without any appeal to fundamental sequences and in a uniform way from the function enumerating the initial ordinals Ω_ν of the number classes. Aczel quickly worked out the idea in unpublished notes in a preliminary way; this was then developed systematically by Jane Bridge in her 1972 Oxford dissertation, the results of which were published in Bridge (1975). She showed how to match up the notations obtained in this way with those obtained by the Bachmann-Pfeiffer-Isles procedures, and she initiated work to show that the countable ordinals generated by these means are recursive. The latter verification was carried out systematically and in full in Buchholz (1975); a detailed exposition of the definition and properties of the θ functions was later given in Schütte (1977) in the first sections of Ch. IX. (We'll return below to a much later simplification leading to the ψ functions in Buchholz (1992).)

The first successful results on ordinal analysis for theories of iterated inductive definitions were obtained only on the intuitionistic side by Per Martin-Löf (1971) via normalization theorems for the $ID_n(acc)^i$ systems as formulated in calculi of natural deduction. He conjectured the bounds $\varphi\varepsilon_{(\Omega_n+1)}0$ in the Bachmann-Pfeiffer hierarchies for these and proved that their supremum is the ordinal of $ID_{<\omega}(acc)^i$ by use of Takeuti (1967).

The first breakthrough on the problems of ordinal analysis for the classical systems was made by Pohlers (1975) to give ordinal upper bounds for the finite ID_n

also by an adaptation of the methods of Takeuti (1967); this was extended later in his Habilitationsschrift, Pohlers (1977), to arbitrary α, with the result that

$$|ID_\alpha| \leq \theta\varepsilon_{(\Omega_\alpha+1)}0$$

as measured in the modified hierarchies described above. In addition, Buchholz and Pohlers (1978) showed this to be best possible by verification of

$$\theta\varepsilon_{(\Omega_\alpha+1)}0 < |ID_\alpha(acc)^i|$$

using a constructive well-ordering proof of each proper initial segment of a natural recursive ordering of order type $\theta\varepsilon_{(\Omega_\alpha+1)}0$. These results lent further hope to the solution of the reductive problem posed above. Independently of their work, in his Stanford dissertation, Sieg (1977) adapted and extended the method of Tait (1970) followed by a formalization of the cut-elimination argument to reduce ID_α to $ID_{\alpha+1}(O)^i$, and thence $ID_{<\lambda}$ to $ID_{<\lambda}(O)^i$, for limit λ, without requiring any involvement of ordinal bounds.

In view of these results, it was decided to exposit all this work together, with the addition of suitable background material, in a *Lecture Notes in Mathematics* volume. As it turned out, the resulting joint publication Buchholz et al. (1981) contained important new contributions to the basic problems about theories of iterated inductive definitions, and though that volume has been superseded in various respects by later work, it still has much of value and I would recommend it as a starting point to the reader interested in studying this subject in some depth. In particular, my preface (Feferman 1981) to the volume fills out the historical picture to that point. Then the first chapter, Feferman and Sieg (1981a), goes over reductive relationships between various subsystems of $\Sigma^1_2 - AC$, systems of iterated inductive definitions, and subsystems of the system T_0 of explicit mathematics from Feferman (1975). The second chapter, Feferman and Sieg (1981b) showed how to obtain the reductions of $\Sigma^1_{n+1} - AC$ to $(\Pi^1_n - AC)_{<\varepsilon_0}$ by proof-theoretic arguments (based on a method called Herbrand analysis by Sieg), in place of the modeltheoretic arguments that had been used by Friedman. Following that, Sieg (1981) presented the work of his thesis in providing the reductions of ID_α to $ID_{\alpha+1}(O)^i$ and of $ID_{<\lambda}$ to $ID_{<\lambda}(O)^i$ for limit λ, without the intervention of ordinal analysis. In the next two chapters Buchholz (1981a, 1981b) introduced uncountably infinitary semi-formal systems making use of a special new $\Omega_{\alpha+1}$-rule in order, in the first of these to obtain the proof-theoretical reduction of the ID_α to suitable $ID_\alpha(acc)^i$ and in the second to reestablish the ordinal bounds previously obtained by Pohlers. Finally, in the last chapter, Pohlers (1981) presented a new approach called the *method of local predicativity*, to accomplish the very same results in a different way. This dispensed with the earlier dependence on the methods of

Takeuti's (1967); the more perspicuous method of local predicativity, in its place, utilizes a kind of extension to uncountably branching proof trees of the methods of predicative proof theory. But both Buchholz' and Pohlers' work in the Buchholz et al. (1981) volume required the use of certain syntactically defined collapsing functions, in order to reduce prima-facie uncountable derivations to countable ones in a way that allows one to obtain the recursive ordinal bounds. As will be described in sec. 4, this was superseded a decade later by the work of Buchholz (1992) showing how to obtain the same bounds without the use of such collapsing functions.

3 Admissible proof theory

Insofar as the work in Buchholz et al. (1981) settled the basic problem posed at the beginning, it could be considered the end of the story. But the aim to develop conceptually still clearer methods had already been underway, beginning with the dissertation of Gerhard Jäger (1979), also under Schütte's direction, but in that case with Pohlers' assistance. The novel element there was to embed various of the subsystems of analysis, both predicative and impredicative, in theories of admissible sets, and to carry out the ordinal analysis of the latter by means of a cut-elimination theorem for associated semi-formal systems of ramified set theory. The connection is that one can identify the minimal models of the theories of admissible sets in question as natural initial segments of the constructible hierarchy. This method was further elaborated in Jäger's Habilitationsschrift (1986) (though that relies on the earlier publication for certain prooftheoretic results about ramified set theory).

The systems of admissible set theory considered by Jäger are taken to have a set of urelements interpreted as the set N of natural numbers given with its successor relation. KPN has the usual axioms for Kripke-Platek set theory with urelements (e.g. from Barwise (1975)), including the full induction scheme (IND_N) on the natural numbers and $(IND_\in)$ on the membership relation. KPN^w is the system obtained from KPN by replacing the $\in$-induction scheme by the corresponding set induction axiom, KPN^r is obtained by further replacing the N-induction scheme by the corresponding set induction axiom, and, finally, KPN^0 is obtained by completely dropping induction on the membership relation. We may also represent KPN^w as $KPN^r + IND_N$. Also considered are the extensions KPL and KPI of KPN, obtained by adding the axioms that the universe is a limit of admissible sets, and that the universe is an admissible limit of admissible sets, respectively; these are also considered in the 'w', 'r', '0' restricted versions as for KPN.[8] The

[8] Jäger (1986) uses KPu, KPl and KPi for what is here denoted by KPN, KPL and KPI, resp. NB: the system denoted KPN in Jäger (1979, 1980) is the same as $KPu^r + IND_N$ in the notation of Jäger (1986), and of KPN^w, or alternatively $KPN^r + IND_N$, in the notation used here. KPN is equivalent in strength to the system often denoted as $KP\omega$.

minimal constructible model L_α of KPI is that for which α is the least recursively inaccessible ordinal.

Among the results of Jäger (1986) is that KPI^0 is a kind of universal theory for systems having Γ_0 as their proof theoretic ordinal, in the sense that all such systems (up to that point) have natural embeddings in KPI^0. Among these is Friedman's theory ATR_0, which also has Γ_0 as a lower bound. The proof theoretic treatement of KPI^0 via ramified set theory takes the place of the earlier proof by Friedman, McAloon and Simpson (1982) of Γ_0 as the ordinal of ATR_0 via model-theoretic arguments. Incidentally, ATR_0 is already embeddable in KPL^0, so KPI^0 is no stronger than that. Moving on to impredicative systems, ID_1 is embedded in KPN, which was shown to have the Howard ordinal as upper bound in Jäger (1979). The strongest system considered in Jäger (1986) is KPI, and among the further notable results for restricted subsystems of that are:

$$(\Sigma^1_2 - AC)_0 \equiv KPI^r, \text{ and } \Sigma^1_2 - AC \equiv KPN^r + IND_N \equiv KPI^r + IND_N,$$

where $\equiv$ is the relation of proof-theoretical equivalence; in both cases, the ordinal analysis of the set-theoretic side is obtained via cut-elimination via the semi-formal system of ramified set theory. The main upper bound result for the full KPI was obtained in Jäger and Pohlers (1983) using the method of local predicativity to establish the ordinal upper bound, while (as explained below) the lower bound follows from the work of Jäger (1983):

$$\Sigma^1_2 - AC + BI \equiv KPI \text{ and } |KPI| = \psi_\Omega(\varepsilon_{I+1}),$$

where, for simplicity, I am using the notation introduced later by Buchholz (1992) for the ψ functions in place of the θ functions. For example, the ordinal of ID_α in these terms is $\psi_\Omega(\varepsilon_{\Omega_\alpha+1})$ in place of $\theta\varepsilon_{(\Omega_\alpha+1)}0$.

In the survey article Pohlers (1998) it is shown how various subsystems of KPI match up both with subsystems of $\Sigma^1_2 - AC + BI$ and with theories of iterated inductive definitions, and their proof-theoretic ordinals are identified in terms of the ψ functions; an informative table is given op. cit. p. 333. For example, we have $ID_\omega \equiv \Pi^1_1 - AC + BI \equiv KPL$. Among these are systems lying between $\Sigma^1_2 - AC$ and $\Sigma^1_2 - AC + BI$ in strength (alternatively described, between KPI^w and KPI) studied by Michael Rathjen in his dissertation (1988) at Münster under Pohlers' direction, including autonomously iterated theories of inductive definitions and corresponding systems of autonomously iterated $\Pi^1_1 - CA$ and of admissible sets; see Pohlers (1998) sec. 3.3.5 for a partial account, since the work of Rathjen(1988) has otherwise not yet been published.

The work on admissible proof theory has also been useful in dealing with systems of explicit mathematics that were formulated and studied in Feferman (1975, 1979). These systems have notions of operations $f, g, \ldots$ and classes (a.k.a. classifications, properties, or [variable] types) $A, B, C, \ldots$, both objects in a universe V of individuals; relations $R, S, \ldots$ are treated as classes of pairs, using a basic pairing operation on V. Operations are in general partial, but may apply to any element of V, including operations and classes. The strongest system of explicit mathematics dealt with op. cit. in which the operations have an interpretation as partial recursive functions is denoted T_0. For present purposes, I want only to concentrate on one axiom group of T_0, concerning a general operation i of inductive generation. Given any A and (binary) $R, i(A, R)$ is always defined and its value is a class I that satisfies:

$$\forall x \in A[\forall y((y, x) \in R \to y \in I) \to x \in I]$$

In addition we have induction on I, which is either taken in the restricted class-induction form

$$\forall x \in A[\forall y((y, x) \in R \to y \in X) \to x \in X] \to I \subseteq X.$$

or as a scheme obtained by substituting for X all formulas of the language of T_0. The system $T_0(res - IG)$ assumes only class-induction, while full T_0 includes the full scheme; the latter does not follow from the former since classes are only assumed to satisfy predicative comprehension in T_0. Informally, $i(A, R)$ is the well-founded part of the relation R, hereditarily in A.

It is easily seen that $ID_{<\varepsilon(0)}(acc)^i$ is contained in $T_0(res - IG)^i$. Moreover, $T_0(res - IG)$ is interpretable in $\Delta^1_2 - CA$. So, by the results described in the preceding section we have

$$ID_{<\varepsilon_0}(acc)^i \equiv T_0(res - IG)^i \equiv T_0(res - IG) \equiv \Sigma^1_2 - AC.$$

Turning next to full T_0, what Jäger showed in his 1983 paper was that by use of a primitive recursive ordering $\preceq$ of order type $\psi_\Omega(\varepsilon_{I+1})$, the well-ordering of each initial segment of the $\preceq$ relation can be established in T^i_0. I had given an (easy) interpretation of T_0 in $\Delta^1_2 - CA + BI$. So that combined with the (much, much harder) work of Jäger and Pohlers (1983) and Jäger (1983) established

$$T^i_0 \equiv T_0 \equiv \Sigma^1_2 - AC + BI.$$

In analogy to the above, I conjecture that there is a suitable system $ID_{<\lambda}(acc)^i$ in some sense that can be added to the left of these equivalences.

4 A simplified version of local predicativity

That is the title of Buchholz (1992), the next main methodological improvement in this approach. As he writes at the beginning of that paper:

> The method of local predicativity as developed by Pohlers ... and extended to subsystems of set theory by Jäger ... is a very powerful tool for the ordinal analysis of strong impredicative theories. But up to now it suffers considerably from the fact that it is based on a large amount of very special ordinal theoretic prerequisites. ... The purpose of the present paper is to expose a simplified and conceptually improved version of local predicativity which ... requires only amazingly little ordinal theory. ... The most important feature of our new approach however seems to be its conceptual clarity and flexibility, and in particular the fact that its basic concepts (i.e. the infinitary system RS^∞ and the notion of an $\mathcal{H}$-controlled RS^∞ derivation) are in no way related to any system of ordinal notations or collapsing functions. (Buchholz 1992, p. 117).

Buchholz there goes on to show how to carry out the ordinal analysis of KPI by this new method in full, absorbable detail. Thenceforth, this simplified method of local predicativity became the gold standard for admissible proof theory. It was continued by Rathjen (1994) in a revised treatment of his 1991 ordinal analysis of KPM, i.e. KP with an axiom saying that the universe is at the level of a Mahlo-admissible ordinal. As he writes (op. cit.) p. 139, KPM is "somewhat at the verge [i.e., upper margin] of admissible proof theory ... Roughly speaking the central scheme of KPM falls under the heading of 'Π_2-reflection with constraints'." The first steps in moving beyond admissible proof theory to systems of analysis like $\Pi^1_2 - CA$, required dealing with Π_n-reflection for arbitrary n, as discussed op. cit., pp. 142ff. For more recent progress — going far beyond our principal concerns here — see Rathjen (2006).

5 Monotone inductive definitions

Though the formal theories of generalized inductive definitions as originally proposed by Kreisel (1963) were of the form $ID_n(mon)^i$, their relationship to the systems $ID_n(acc)^i$ was left unsettled by the work of Buchholz et al. (1981), as was the relationship for the corresponding classical systems.[9] This was first taken up in my paper Feferman (1982a) for the 1981 Brouwer Centenary Symposium. I showed there that, at least on the classical side, $ID_n(mon)$ is a conservative extension of $ID_n(O)$ for all n. The method of proof is via an interpretation of $ID_n(mon)$ in a

[9] At first sight, one could obtain a simple reduction of the $ID(mon)$ theories to the $ID(pos)$ theories (whether classical or intuitionistic) by an application of Lyndon's interpolation theorem to formulas of the form $A(Q, P, x) \wedge \forall u[P(u) \rightarrow P'(u)] \rightarrow A(Q, P', x)$, derived from prior axiom schemes. This was indeed stated in Sieg (1977); however, Buchholz pointed out to Sieg soon after that there is a gap in the argument, since one should allow both P and P' to be used together in those schemes. There is no obvious way to get around this obstacle.

predicative second order extension $ID_n(O)^{(2)}$ which is easily shown to be a conservative extension of $ID_n(O)$. The main work goes into showing that if $A(P,x)$ is an arithmetical formula such that $ID_{n-1}(O)^{(2)}$ proves the monotonicity condition $\forall X \forall Y \forall x[A(X,x) \wedge X \subseteq Y \rightarrow A(X,x)]$ then one can define a predicate P_A in $ID_n(O)^{(2)}$ to provably satisfy the required closure and induction scheme axioms. In the same paper I also sketched how to generalize these arguments and results to the case of 'α' in place of 'n'. It follows from the work of Buchholz and Pohlers described in sec. 2 that in general $ID_\alpha(mon)$ is proof-theoretically reducible to $ID_\alpha(acc)^i$ and the proof-theoretic ordinals are the same. Incidentally, as noted by Kreisel in 1963, there is no obvious informal argument for the constructivity of $ID_1(mon)^i$ short of quantification over species in the intuitionistic sense.

At the conclusion of Feferman (1982a) I brought attention to the formulation of monotonic inductive definitions in the much more general setting of explicit mathematics. By an operation f from classes to classes, in symbols $Cl - Op(f)$, we mean one such that $\forall X \exists Y(fX = Y)$; then by $Mon(f)$ we mean $C1 - Op(f) \wedge \forall X \forall Y[X \subseteq Y \rightarrow fX \subseteq fY]$. The assertion $ELFP(f)$ that f has a least fixed point is expressed as $\exists X[fX \subseteq X \wedge \forall Y(fY \subseteq Y \rightarrow X \subseteq Y)]$. I suggested adding the following axiom MID for Monotone Inductive Definitions to $T_0 : \forall f[Mon(f) \rightarrow ELFP(f)]$, i.e. the statement that every monotonic operation from classes to classes has a least fixed point. And finally, I raised the question whether $T_0 + MID$ is any stronger than T_0, since as I wrote: "[it] includes all constructive formulations of the iteration of monotone inductive definitions of which I am aware, while T_0 (in its IG axiom) is based squarely on the general iteration of accessibility inductive definitions. Thus it would be of great interest for the present subject to settle the relationship between these theories." At the time I thought that my interpretation of T_0 in $\Sigma^1_2 - AC + BI$ could somehow be extended to one for $T_0 + MID$, and thus give a general reduction of monotone to accessibility inductive definitions. But as I said loc. cit., I did not succeed in doing this. In fact, it was not obvious how to produce any model of $T_0 + MID$, let alone one bounding its strength by that of T_0.

The first progress on these questions was made by my student Shuzuo Takahashi in his PhD dissertation at Stanford, published as Takahashi (1989). He proved that $T_0 + MID$ is interpretable in $\Pi^1_2 - CA + BI$; this required a surprisingly difficult model construction, while no lower bound in strength was revealed by Takahashi's work. Meanwhile I had raised the question of the status of a uniform version $UMID$ of the MID axiom, obtained by adding a constant lfp to the language of T_0 with the statement that for any f, if $Mon(f)$ then $lfp(f)$ is a least fixed point of f; the consistency of $T_0 + UMID$ was unsettled by Takahashi's interpretation. These questions of strength were later addressed in a series of papers by Michael Rathjen (1996, 1998, 1999) and a joint one with Thomas Glass and Andreas Schlüter (1997), all surveyed with some further extensions in Rathjen (2002). Here, briefly, are some of the results.

First of all, it was shown in Rathjen (1996) that $T_0 + MID$ is in fact stronger than T_0; in fact $T_0(res - IG) + MID$ proves the existence of a model of T_0. Then in Glass, Rathjen and Schlüter (1997) it was shown that

$$T_0(res - IG) + MID \equiv (\Sigma^1_2 - AC)^- + (\Pi^1_2 - CA)^-, \text{ and}$$
$$T_0(res - IG) + IND_N + MID \equiv \Sigma^1_2 - AC + (\Pi^1_2 - CA)^-,$$

where the minus sign superscript on a scheme indicates that there are no class parameters (i.e. free class variables). Following that, Rathjen (2002) proved that $T_0 + MID$ is bounded in strength by a theory $\mathcal{K}$ that is slightly stronger than $\Sigma^1_2 - AC + (\Pi^1_2 - CA)^- + BI$.

Rathjen (1999, 2002) also obtained results about the strength of $UMID_N$ (which is the $UMID$ principle relativized to subclasses of N), including the following:

$$T_0(res - IG) + UMID_N \equiv (\Pi^1_2 - CA)_0,$$

while

$$\Pi^1_2 - CA < T_0 + UMID_N \leq \Pi^1_2 - CA + BI.$$

Rathen conjectured (2002), p. 339, that the $\leq$ here can be replaced by $\equiv$ and that $UMID$ gives no stronger theory than $UMID_N$. Finally, it is shown there that

$$T_0 + MID < T_0 + UMID_N.$$

All these results are for the systems of explicit mathematics as based on classical logic. About the intuitionistic side of these various theories, Rathjen wrote (loc. cit.) that virtually nothing is known. However, subsequently, Sergei Tupailo (2004) established that the classical and intuitionistic versions of $T_0(res-IG)+UMID_N$ are of the same strength, by an indirect argument via the so-called μ-calculus.[10]

A number of problems about the MID and $UMID$ principles in explicit mathematics are still left open by this work, especially on the intuitionistic side.

6 The method of functional interpretation, 1968-2008

All of the proof-theoretical analyses of classical theories of iterated inductive definitions surveyed above made use of cut-elimination arguments applied to suitable uncountably infinitary sequent-style systems. But for the purely reductive part

[10] Michael Rathjen has informed me that there is an alternative more direct argument to obtain Tupailo's result via an application of the double negation translation to the operator theory $T^{OP}_{<\omega}$ of Rathjen (1998), which is of the same strength as $T_0(res - IG) + UMID_N$; moreover the same method applies to $T^{OP}_{<\varepsilon(0)}$ which is of the same strength as $T_0(res - IG) + IND_N + UMID_N$ and thence of its intuitionistic version.

of the problem, it seemed to me from the beginning that an extension of Gödel's method of functional interpretation could serve to establish the expected results using finite formulas throughout. In an unpublished lecture that I gave at the 1968 Buffalo conference — though circulated in mimeographed notes Feferman (1968) — I obtained a semi-constructive functional interpretation of ID_1 in the classical system $ID_1(T)$, where the set T of constructive countable tree ordinals is a variant of O. The hope was to then reduce $ID_1(T)$ to a suitable $ID_1(acc)^i$ and thereby show that $|ID_1|$ is the Howard ordinal, but I did not see how to get around the obstacle of essential use of numerical quantification (in its guise as the non-constructive minimum operator μ) in doing so. The next attempts to approach this and the iterated case via functional interpretation were made by my student Jeffery Zucker in his dissertation (1971), the work from which was published in Zucker (1973). Interestingly, Zucker showed that $|ID_1| = |ID_1(acc)^i|$ by application of Howard's majorization technique to my functional interpretation with the μ-operator. However, he did not see a way to extend this to the iterated case. What he *was* able to do was give a Kreisel-style modified realizability functional interpretation of $ID_n(acc)^i$ in a theory of constructive tree classes up to level n for each $n < \omega$ and show that they have the same provably recursive ordinals; he also sketched how this could be extended to transfinite α.

My notes Feferman (1968) and questions about its approach did not see the general light of day until they were outlined in sec. 9 of my survey with Jeremy Avigad in the *Handbook of Proof Theory* of Gödel's functional interpretation, Avigad and Feferman (1998); I included that section there in the hopes that someone would see how to overcome the obstacle that I had met. To my great satisfaction, that was finally achieved by Avigad with his student Henry Towsner in 2008 by a variant functional interpretation; the fact that this took place in the year of celebration of Wolfram Pohlers' retirement is the reason why I subtitled this piece a forty year long saga. Since this is relatively new and unfamiliar material, I want to sketch how the approach in Avigad and Towsner (2008) proceeds.

As background, let's look briefly at Gödel's original *Dialectica* (or $D-$) interpretation (1958, 1972) and its consequences; subsequent work follows a broadly similar pattern. Gödel applied the D-interpretation to Heyting Arithmetic HA to reduce it to a quantifier-free theory of primitive recursive functionals of finite type over N that he simply denoted by 'T'. This is carried out via an intermediate translation which sends each formula A of arithmetic into a formula A^D of the form $\exists z \forall x A_D(z, x)$ where z, x are sequences of variables of finite type (possibly empty) and A_D is a quantifier free formula of the language of T. The main theorem was that if $HA \vdash A$ then $T \vdash A_D(t, x)$ for some sequence t of terms of the same type as z; this gives the reduction $HA \leq T$. A is equivalent to A^D under the assumption of the Axiom of Choice, which in this setting is constructively accepted, plus the non-constructive Markov's Principle and a principle called Independence of Premises. But the interpretation of A by A^D can be applied in combination with the double negation translation of PA into HA to show that these systems have the same provably recursive functions and that, moreover, they are the same as the functions of type 1 generated by the terms of T. For if $PA \vdash \forall x \exists y R(x, y)$ with R primitive recursive then $HA \vdash \forall x \neg\neg \exists y R(x, y)$ and so by Markov's Principle and the Axiom of Choice we have $\exists z \forall x R(x, z(x))$; finally, by the D-interpretation, there is a closed term of type 1 such that $T \vdash R(x, t(x))$. The set of functions of type 1 generated by the primitive recursive functionals of finite type is called the 1-section of T. So this result can be summarized by the equations

$$Prov - Rec(PA) = Prov - Rec(HA) = 1 - sec(T).$$

Further work must be done if one wants to use this to recapture the result of Kreisel (1952) that the provably recursive functions of PA and HA are just those obtained by recursion on ordinals $\alpha < \varepsilon_0$. This can be obtained via the normalization of the terms of T using an assignment to them of ordinals $< \varepsilon_0$. That was first carried out by Tait (1965) and later by Howard (1970) in ways akin to the use of ordinals $< \varepsilon_0$ in the cut elimination arguments for PA by Schütte and Gentzen, respectively.

The details for the functional interpretation of theories of inductive definition are only given in full for ID_1 in Avigad and Towsner (2008) and sketched for arbitrary ID_n in their final section, though they say it can be extended to transfinite iterations. The first step, for a given arithmetical $A(P, x)$, is to translate ID_1 into the classical theory OR_1 of abstract countable tree ordinals extended by axioms (I) for a predicate $I(x, \alpha)$ of natural numbers and (tree) ordinals, interpreted as $x \in I_\alpha$ in the approximations from below to the least fixed point of A. The functional interpretation is then used to obtain a reduction of $OR_1 + (I)$ to an $ID_1(acc)^i$ via a quantifier-free theory T_Ω of primitive recursive functionals of finite type over the tree ordinals and two of its extensions, QT_Ω, which allows quantifiers over all finite type variables, and Q_0T_Ω, which allows only numerical quantification; unless otherwise indicated both are in classical logic. Avigad and Towsner show that $OR_1 + (I) \leq Q_0T_\Omega$ by the Diller-Nahm-Shoenfield variant of the D-interpretation. The problem then is to get rid of Q_0 and pass to intuitionistic logic, which was essentially the obstacle that I and Zucker had met. The novel key step is to establish the reduction $Q_0T_\Omega \leq (QT_\Omega)^i$, using an adaptation of the argument in Sieg (1981) to formalize cut-elimination for a semi-formal version of Q_0T_Ω in $(QT_\Omega)^i$. Finally, the model of T_Ω and thence of $(QT_\Omega)^i$ in the hereditarily recursive operations over the recursive countable tree ordinals may be formalized in $ID_1(O)^i$. Chaining together these successive reductions, Avigad and Towsner obtain:

$$ID_1 \leq ID_1(O)^i, |ID_1| = |ID_1(O)^i|, \text{ and}$$
$$Prov - Rec(ID_1) = Prov - Rec(ID_1(O)^i) = 1 - Sec(T_\Omega).$$

As I said, they assert that the same methods serve to establish $ID_\alpha \leq ID_\alpha(O)^i$ and $|ID_\alpha| = |ID_\alpha(O)^i|$ in general; it would be good to see the details of that presented in full. But assuming that is the case, on the basis of present evidence this work of Avigad and Towsner is an improvement on both Sieg (1977, 1981), which only obtained $ID_\alpha \leq ID_{\alpha+1}(O)^i$, and Buchholz (1981a), which only obtained $ID_\alpha \leq ID_\alpha(acc)^i$. In addition, their functional interpretation has the advantage of giving a mathematical characterization of the provable recursive functions of a given ID theory in terms of the 1-section of a natural class of functionals. Of course, one would need to use something like the methods of local predicativity with ordinal analysis in order to further describe those functions in terms of suitable ordinal recursions.

7 Conclusion

All the work surveyed here illustrates how the initial aim to use the constructive reduction and ordinal analysis of theories of iterated inductive definitions for the extension of Hilbert's program to impredicative systems of analysis became transmuted into a subject of interest in its own right. In addition, the continuing desire for conceptually clear arguments led to successive methodological improvements, which in turn proved useful in other applications. Though the proof theory of iterated inductive definitions as first order systems falls far short of serving to deal with the next level of impredicative systems of analysis such as $\Pi^1_2 - CA$, the work described in sec. 5 on classical and constructive theories of monotonic inductive definitions suggests that suitable second order theories of such may be useful for that purpose.

To conclude, here are some questions suggested by the work that has been surveyed above.

1. One does not have to be a devotee of purity of method to ask whether an alternative, more purely functional interpretation approach might be possible to arrive at the reduction $ID_\alpha \leq ID_\alpha(O)^i$ in general. Recall that Zucker (1973) showed that the proof theoretic ordinals of ID_1 and T_Ω are the same by applying the majorization argument of Howard (1973) to the semi-constructive functional interpretation of my 1968 notes. For me, this is reminiscent of the use by Kohlenbach (1992) of his method of monotone functional interpretation

to eliminate numerical quantification in the reduction of the system WKL to PRA.

So the question is whether the appeal to cut-elimination in the final step of the Avigad and Towsner work both for ID_1 and in general for ID_α can be avoided by an application of the monotone functional interpretation or one of its variants, such as the bounded functional interpretation of Ferreira and Oliva (2005). Incidentally, I was misled by the work of Avigad and Towsner (2008) into thinking that they had somehow refined Sieg's argument to replace '$\alpha+1$' by 'α' in the target system. But it seems that that was only possible in combination with their use of functional interpretation. So if a purely functional interpretation approach does not succeed to obtain a proof-theoretic reduction of ID_α to $ID_\alpha(O)^i$, it is still a question whether a refinement of Sieg's arguments using cut elimination can achieve the same result.

2. What part of mathematics can be carried out in ID_1? A recent interesting case study is provided by Avigad and Towsner (2009) (cf. also Avigad (2009) sec. 5): a version of the structure theorem in combinatorial ergodic theory due to Furstenberg (1977) can be formalized in ID_1, via the interpretation in $Q_0T_\Omega + (I)$ described in the preceding section. That theorem was used by Furstenberg to prove by conceptually high level means the famous theorem of Szemerédi (1975), whose original combinatorial proof was very difficult. The work of Beleznay and Foreman (1996) suggests that the full Furstenberg structure theorem is equivalent to the Π^1_1 comprehension axiom. But the work of Avigad and Towsner shows that the full strength of the structure theorem is far from necessary for the ergodic-theoretic proof of the Szemerédi theorem. As this example shows, it may be that the pursuit of what other mathematics can be formalized in ID_1 is more conveniently examined in proof-theoretically equivalent systems in which ordinals play an explicit role, such as the theory $OR_1 + (I)$ or its functional interpretations in the preceding section.

3. What about what can be done in iterated IDs?

4. Ordinal analysis only tells us something about the provably countable ordinals of a theory. In the case of the ID_αs, it would seem to make sense to talk about their provably uncountable ordinals. How would that be defined, and what can be established about them?

5. ID_1 is similar to Peano Arithmetic in various respects. In Feferman (1996) I introduced the general notion of an open-ended schematic axiom system and its unfolding, to explain the idea of what we ought to accept if we have accepted given notions and given principles concerning them. In Feferman and Strahm

(2000) we showed that the full unfolding of a very basic schematic system NFA for non-finitist arithmetic is proof-theoretically equivalent to predicative analysis. There is a natural formulation of a basic schematic system NFI which stands to ID_1 as NFA stands to PA. What is its unfolding?

6. A side development of the work on theories of iterated inductive definitions is that on theories of iterated fixed point theories $ID_{\alpha}^{\wedge}$, whose basic axiom for a given A takes the form $\forall x[A(P_A, x) \leftrightarrow P_A(x)]$. Building on work of Aczel characterizing the strength of $ID_1^{\wedge}$, I showed in Feferman (1982b) that the union of the finitely iterated fixed point theories is equivalent in strength to predicative analysis. That work was continued into the transfinite by Jäger, Kahle, Setzer and Strahm (1999) who showed that even though one thereby goes beyond predicativity in strength, the methods of predicative proof theory can still be applied. They thus introduced the term metapredicativity for the study of systems that can be treated by such means. In unpublished work by Jäger and Strahm, that even goes beyond ID_1. One should try to characterize the domain of metapredicativity in terms analogous to those used at the outset to characterize predicativity as the limit of the autonomous progression of ramified systems. Assuming that, I would conjecture that the full unfolding of the schematic system NFI suggested above is proof-theoretically equivalent to the union of the metapredicative systems.

7. The set-theoretical treatment of least fixed points of monotonic operator apply to operators on subsets of arbitrary sets M. Are there reasonable theories of IDs over other sets than the natural numbers, e.g. the real numbers? What can be said about their strength?

8 Acknowledgements

I wish to thank Jeremy Avigad, Wilfried Buchholz, Gerhard Jäger, Grigori Mints, Michael Rathjen, Wilfried Sieg, Thomas Strahm and Henry Towsner for their useful comments on a draft of this article.

References

[1] P. Aczel, H. Simmons and S. Wainer, eds. (1992), Proof Theory, Cambridge University Press, Cambridge.

[2] J. Avigad (2009), The metamathematics of ergodic theory, Annals of Pure and Applied Logic 157, 64-76.

[3] J. Avigad and S. Feferman (1998), Gödel's functional ("Dialectica") interpretation, in Buss (1998), 337-405.

[4] J. Avigad and H. Towsner (2008), Functional interpretation and inductive definitions, `http://arxiv.org/abs/0802.1938`; to appear in J. Symbolic Logic.

[5] J. Avigad and H. Towsner (2009), Metastability in the Furstenberg-Zimmer tower, `http://arxiv.org/abs/0902.0356`.

[6] H. Bachmann (1950), Die Normalfunktionen und das Problem der ausgezeichneten Folgen von Ordnungzahlen, Vierteljahresschr. Nat. Ges., Zürich 95, 5-37.

[7] J. Barwise (1975), Admissible Sets and Structures, Springer-Verlag, Berlin.

[8] F. Beleznay and M. Foreman (1996), The complexity of the collection of measure-distal transformations, Ergodic Theory and Dynamical Systems 16, 929-962.

[9] J. Bridge (1975), A simplification of the Bachmann method for generating large countable ordinals, J. Symbolic Logic 40, 171-185.

[10] W. Buchholz (1975), Normalfunktionen und konstruktiven Systeme von Ordinalzahlen, in Proof Theory Symposion, Kiel 1974, Lecture Notes in Mathematics 500, 4-25.

[11] W. Buchholz (1981a), The $\Omega_{\mu+1}$-rule, in Buchholz et al. (1981), 188-233.

[12] W. Buchholz (1981b), Ordinal analysis of ID_{ν}, in Buchholz et al. (1981), 234-260.

[13] W. Buchholz (1992), A simplified version of local predicativity, in Aczel et al. (1992), 115-147.

[14] W. Buchholz (1997) Explaining Gentzen's consistency proof within infinitary proof theory, in G. Gottlob, A. Leitsch and D. Mundici (eds.), Computational Logic and Proof Theory. KGC '97, Lecture Notes in Computer Science 1298.

[15] W. Buchholz (2001) Explaining the Gentzen-Takeuti reduction steps: A second-order system, Archive for Mathematical Logic 40,

[16] W. Buchholz, S. Feferman, W. Pohlers and W. Sieg (1981), Iterated Inductive Definitions and Subsystems of Analysis: Recent proof-theoretical studies, Lecture Notes in Mathematics 897.

[17] W. Buchholz and W. Pohlers (1978), Provable well-orderings of formal theories for transfinitely iterated inductive definitions, J. Symbolic Logic 43, 118-125.

[18] S. R. Buss, ed. (1998), Handbook of Proof Theory, Elsevier, Amsterdam.

[19] A. Church and S. C. Kleene (1936), Formal definitions in the theory of ordinal numbers, Fundamenta Mathematicae 28, 11-21.

[20] S. Feferman (1968), Ordinals associated with theories for one inductively defined set. (Unpublished notes.)

[21] S. Feferman (1970) Formal theories for transfinite iteration of generalized inductive definitions and some subsystems of analysis, in Kino et al. (1970), 303-326.

[22] S. Feferman (1975), A language and axioms for explicit mathematics, in J. N. Crossley (ed.) Algebra and Logic, Lecture Notes in Mahtematics 450, 87-139.

[23] S. Feferman (1979), Constructive theories of functions and classes, in M. Boffa et al. (eds.) Logic Colloquium '78, North-Holland, Amsterdam, 159-224.

[24] S. Feferman (1981), How we got from there to here, in Buchholz et al. (1981), 1-15.

[25] S. Feferman (1982a), Iterated inductive fixed-point theories: application to Hancock's conjecture, in G. Metakides (ed.), Patras Logic Symposion, North-Holland, Amsterdam, 171-196.

[26] S. Feferman (1982b), Monotone inductive definitions, in A. S. Troelstra and D. van Dalen (eds.), The L.E.J. Brouwer Centenary Symposium, North-Holland, Amsterdam, 77- 89.

[27] S. Feferman (1988), Hilbert's program relativized: proof-theoretical and foundational reductions, J. Symbolic Logic 53, 364-384.

[28] S. Feferman (1993), What rests on what? The proof-theoretical and foundational analysis of mathematics, in J. Czermak (ed.), Philosophy of Mathematics, Vol.I, Verlag-Hölder- Pichler-Tempsky, Vienna, 147-171. Reprinted in Feferman 1998, 187-208.

[29] S. Feferman (1996), Gödel's program for new axioms: Why, where, how and what?, in P. Hájek (ed.), Gödel '96, Lecture Notes in Logic 6, 3-22.

[30] S. Feferman (1998), In the Light of Logic, Oxford University Press, New York.

[31] S. Feferman (2000) Does reductive proof theory have a viable rationale?, Erkenntnis 53, 63-96.

[32] S. Feferman and W. Sieg (1981a), Iterated inductive definitions and subsystems of analysis, in Buchholz et al. (1981), 16-77.

[33] S. Feferman and W. Sieg (1981b) Proof theoretic equivalences between classical and constructive theories for analysis, in Buchholz et al. (1981) 78-142.

[34] S. Feferman and T. Strahm (2000), The unfolding of non-finitist arithmetic, Annals of Pure and Applied Logic 104, 75 - 96.

[35] F. Ferreira and P. Oliva (2005), Bounded functional interpretation, Annals of Pure and Applied Logic 135, 73-112.

[36] H. M. Friedman (1970) Iterated inductive definitions and $\Sigma^1_2 - AC$, in Kino, Myhill and Vesley 1970, 435-442.

[37] H. M. Friedman, K. McAloon, and S. G. Simpson (1982), A finite combinatorial principle which is equivalent to the 1-consistency of predicative analysis, in G. Metakides (ed.), Patras Logic Symposion, North-Holland, Amsterdam, 197-230.

[38] H. Furstenberg (1977), Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions, J. d'Analyse Mathematique 31, 204-256.

[39] T. Glass, M. Rathjen and A. Schlüter (1997), The strength of monotone inductive definitions in explicit mathematics, Annals of Pure and Applied Logic 85, 1-46.

[40] K. Gödel (1958), Über eine bisher noch nicht benüzte Erweiterung des finiten Standpunktes, Dialectica 12, 280-287. Reproduced with English translation in Gödel (1990), 241-251.

[41] K. Gödel (1972), On an extension of finitary methods which has not yet been used, in Gödel (1990), 271-280.

[42] K. Gödel (1990), Collected Works,Vol. II. Publications 1938-1974 (S. Feferman, et al., eds.), Oxford University Press, New York.

[43] A. Heyting, ed. (1959), Constructivity in Mathematics, North-Holland, Amsterdam.

[44] W. A. Howard (1970), Assignment of ordinals to terms for primitive recursive functionals of finite type, in Kino et al. (1970), 453-468.

[45] W. A. Howard (1972), A system of abstract constructive ordinals, J. Symbolic Logic 37, 355-374.

[46] W. A. Howard (1973), Hereditarily majorizable functionals of finite type, in Troelstra (1973), 454-461.

[47] D. Isles (1970), Regular ordinals and normal forms, in Kino et al. (1970), 339-361. 25

[48] G. Jäger (1979), Die Konstruktible Hierarchie als Hilfsmittel zur beweistheoretischen Untersuchung von Teilsystemen der Mengenlehre und Analysis, Dissertation, Ludwig-Maximilians-Universität, Munich.

[49] G. Jäger (1980), Beweistheorie von KPN, Archiv für Mathematische Logik und Grundlagenforschung 20,

[50] G. Jäger (1983), A well-ordering proof for Feferman's theory T_0, Archive for Mathematical Logic 23, 65-77.

[51] G. Jäger (1986), Theories for Admissible Sets. A unifying approach to proof theory, Bibliopolis, Naples.

[52] G. Jäger, R. Kahle, A. Setzer and T. Strahm (1999), The proof-theoretic analysis of transfinitely iterated fixed point theories, J. Symbolic Logic 64, pp. 53 - 67.

[53] G. Jäger and W. Pohlers (1983), Eine beweistheoretische Untersuchung von $(\Delta^1_2 - CA) + (BI)$ und verwandter Systeme, Bayerische Akademie der Wissenschaften, Sitzungsberichte 1982, 1-28.

[54] A. Kino, J. Myhill and R. Vesley, eds. (1970), Intuitionism and Proof Theory, North-Holland, Amsterdam.

[55] S. C. Kleene (1938), On notation for ordinal numbers, J. Symbolic Logic 3, 150-155.

[56] S. C. Kleene (1959), Countable functionals, in Heyting (1959), 81-100.

[57] U. Kohlenbach (1992), Effective bounds from ineffective proofs in analysis: an application of functional interpretation and majorization, J. Symbolic Logic 57, 1239-1273.

[58] G. Kreisel (1952), On the interpretation of non-finitist proofs, part II: interpretation of number theory, applications, J. Symbolic Logic17, 43-58.

[59] G. Kreisel (1959) Interpretation of analysis by means of constructive functionals of finite type, in Heyting (1959), 101-128.

[60] G. Kreisel (1963), Generalized inductive definitions, in Seminar on the Foundations of Analysis, Stanford 1963. Reports (mimeographed), Mathematical Sciences Library, Stanford University, 3.1-3.25.

[61] P. Martin-Löf (1971), Hauptsatz for the intuitionistic theory of iterated inductive definitions, in J. E. Fenstad (ed.), Proceedings of the Second Scandinavian Logic Symposium, North-Holland, Amsterdam, 179-216.

[62] H. Pfeiffer (1964), Ausgezeichnete Folgen für gewisse Abschnitte der zweiten und weiteren Zahlklassen, Dissertation, Technische Hochschule Hannover.

[63] W. Pohlers (1975), An upper bound for the provability of transfinite induction in systems with N-times iterated inductive definitions, in J. Diller and G. H. Müller (eds.), ISILC Proof Theory Symposium, Lecture Notes in Mathematics 500, 271-289.

[64] W. Pohlers (1977), Beweistheorie der iterierten induktiven Definitionen, Habilitationsschrift, Ludwig-Maximilians-Universität, Munich.

[65] W. Pohlers (1981) Proof-theoretical analysis of ID_ν by the method of local predicativity, in Buchholz et al. (1981), 261-357.

[66] W. Pohlers (1987), Contributions of the Schütte school in Munich to proof theory, in Takeuti (1987), 406-431.

[67] W. Pohlers (1989) Proof Theory. An introduction, Lecture Notes in Mathematics, v. 1407.

[68] W. Pohlers (1992) A short course in ordinal analysis, in Aczel et al. (1992), 27-78.

[69] W. Pohlers (1998) Subsystems of set theory and second order number theory, in Buss (1998), 209-335.

[70] W. Pohlers (2009) Proof Theory. The first step into impredicativity, Springer-Verlag, Berlin.

[71] M. Rathjen (1988), Untersuchungen zu Teilsystemen der Zahlentheorie zweiter Stufe und der Mengenlehre mit einer zwischen $\Delta^1_2 - CA$ und $\Delta^1_2 - CA + BI$ liegenden Beweisstärke, Doctoral thesis, University of Münster.

[72] M. Rathjen (1994), Admissible proof theory and beyond, in D. Prawitz, B. Skyrms, and D. Westerstahl (eds.), Logic, Methodology and Philosophy of Science IX, Elsevier, Amsterdam.

[73] M. Rathjen (1996), Monotone inductive definitions in explicit mathematics, J. Symbolic Logic 61, 125-146

[74] M. Rathjen (1998), Explicit mathematics with the monotone fixed point principle, J. Symbolic Logic 63, 509-542.

[75] M. Rathjen (1999), Explicit mathematics with the monotone fixed point principle. II: Models, J. Symbolic Logic 64, 517-550.

[76] M. Rathjen (2002), Explicit mathematics with monotone inductive definitions: A survey, in Sieg, et al. (2002), 329-346.

[77] M. Rathjen (2006), The art of ordinal analysis, in Proceedings of the International Congress of Mathematicians, Vol. II, European Mathematical Society, 45-69.

[78] K. Schütte (1977), Proof Theory, Springer-Verlag, Berlin.

[79] W. Sieg (1977), Trees in Metamathematics. Theories of inductive definitions and subsystems of analysis, PhD Dissertation, Stanford University.

[80] W. Sieg (1981), Inductive definitions, constructive ordinals, and normal derivations, in Buchholz et al. (1981), 143-187.

[81] W. Sieg, R. Sommer and C. Talcott, eds. (2002), Reflections on the Foundations of Mathematics. Essays in honor of Solomon Feferman, Lecture Notes in Logic 15, Assoc. for Symbolic Logic, A.K. Peters, Ltd., Natick.

[82] C. Spector (1962), Provably recursive functions of analysis: a consistency proof of analysis by an extension of principles formulated in current intuitionistic mathematics, in J. C. E. Dekker (ed.), Recursive Function Theory. Proc. Symposia in Pure Mathematics 5, AMS, Rhode Island, 1-27.

[83] E. Szemerédi (1975), On sets of integers containing no k elements in arithmetic progression, Acta Arithmetica 27, 199-245.

[84] W. W. Tait (1965), Infinitely long terms of transfinite type, in J. N. Crossley and M. A. E. Dummett (eds.), Formal Systems and Recursive Functions, North-Holland, Amsterdam, 176-185.

[85] W. W. Tait (1970) Applications of the cut elimination theorem to some subsystems of classical analysis, in Kino et al. (1970), 475-488.

[86] S. Takahashi (1989), Monotone inductive definitions in a constructive theory of functions and classes, Annals of Pure and Applied Logic 42, 255-279.

[87] G. Takeuti (1967), Consistency proofs of subsystems of classical analysis, Annals of Mathematics 86, 299-348.

[88] G. Takeuti (1987), Proof Theory, North-Holland, Amsterdam, 2nd edn, (with appendices by G. Kreisel, W. Pohlers, S. Simpson and S. Feferman).

[89] G. Takeuti and M. Yasugi (1973), The ordinals of the system of second order arithmetic with the provably Δ^1_2-comprehension axiom and with the Δ^1_2-comprehension axiom respectively, Japanese J. of Mathematics 41, 1-67.

[90] A. S. Troelstra, ed. (1973), Metamathematical Investigation of Intuitionistic Arithmetic and Analysis, Lecture Notes in Mathematics 344.

[91] S. Tupailo (2004), On the intuitionistic strength of monotone inductive definitions, J. Symbolic Logic 69, 790-798.

[92] J. I. Zucker (1971), Proof Theoretic Studies of Systems of Iterated Inductive Definitions and Subsystems of Analysis, PhD Dissertation, Stanford University. J. I. Zucker (1973), Iterated inductive definitions, trees and ordinals, in Troelstra (1973) 392-453. Stanford University.

A New Approach to Predicative Set Theory

Arnon Avron

School of Computer Science, Tel Aviv University, Israel
aa@cs.tau.ac.il

Abstract We suggest a new basic framework for the Weyl-Feferman predicativist program by constructing a formal predicative *set theory* PZF which resembles ZF. The basic idea is that the predicatively acceptable instances of the comprehension schema are those which determine the collections they define in an absolute way, independent of the extension of the "surrounding universe". This idea is implemented using syntactic safety relations between formulas and sets of variables. These safety relations generalize both the notion of domain-independence from database theory, and Gödel notion of absoluteness from set theory. The language of PZF is type-free, and it reflects real mathematical practice in making an extensive use of statically defined abstract set terms. Another important feature of PZF is that its underlying logic is ancestral logic (i.e. the extension of FOL with a transitive closure operation).

1 Introduction

The predicativist program for the foundations of mathematics, initiated by Poincaré in [35, 36] [1], and first seriously developed by Weyl in [50], seeks to establish certainty in mathematics without revolutionizing it (as the intuitionistic program does). The program as is usually conceived nowadays (following Weyl and Feferman) is based on the following two basic principles:

(PRE) Higher order constructs, such as sets or functions, are acceptable only when introduced through definitions. These definitions cannot be circular. Hence in defining a new construct one can only refer to constructs which were introduced by previous definitions.

(NAT) The natural-numbers sequence is a basic well understood mathematical concept, and as a totality it constitutes a set.

[1] Though its kernel can be found in Richard's discussion of his paradox [38].

The first of these principles, (PRE), was interpreted by Russell according to his philosophical views of logic, [39], [40], and incorporated as the *ramified type theory* (RTT) in *Principia Mathematica* ([51]). In RTT objects are divided into types, and each higher-order type is further divided into levels. However, the use of levels makes it impossible to develop mathematics in RTT, and so Russell had to add a special axiom of reducibility which practically destroyed the predicative nature of his system ([37]). The principle was then taken again by Weyl in [50], but instead of Russell's ramified hierarchy, Weyl adopted the second principle, (NAT), which also goes back to Poincaré. Weyl's predicativist program was later extensively pursued by Feferman, who in a series of papers (see e.g. [15, 17, 19, 20]) developed proof systems for predicative mathematics. Feferman's systems are less complex than RTT, and he has shown that a very large part of classical analysis can be developed within them. He further conjectured that predicative mathematics in fact suffices for developing all the mathematics that is actually indispensable to present-day natural sciences.

Despite this success, Feferman's systems failed to receive in the mathematical community the interest they deserve. Unlike constructive mathematics, they were also almost totally ignored in the computer science community. The main reason for this seems to be the fact that on the one hand Feferman's systems are not "revolutionary" (since they allow the use of *classical* logic), but on the other hand they are still rather complicated in comparison to the impredicative formal set theory ZF, which provides the standard foundations and framework for developing mathematics. In particular: Feferman's systems still use complicated systems of types, and both functions and classes are taken in them as independent primitives. Therefore working within Feferman's systems is not easy for someone used to ZF (or something similar).

The main goal of this paper is to suggest a new framework for the Weyl-Feferman predicativist program by constructing an absolutely (at least in our opinion) reliable predicative *set theory* PZF which is suitable for mechanization, and has the following properties:

1. Its language is type-free, and it reflects real mathematical practice by making an extensive use of *abstract set terms* (i.e. terms of the form $\{x \mid \varphi\}$). [2].

2. Like ZF, it is a *pure* set theory, in which everything (including functions) is assumed to be a set. Moreover: from a platonic point of view, the universe V of ZF (whatever this universe is) is a model of it.

[2] The use of such terms, albeit in a somewhat cumbersome form, more complicated than that actually used in mathematical texts, is also a major feature of the systems developed in [8, 9].

3. ZF itself (or each intuitively true extension of it) is obtainable from it in a straightforward way.

2 The Main Ideas

2.1 Interpreting and Implementing Principle (PRE)

According to our approach, a predicative set theory need *not* exclude the possibility that "arbitrary (undefinable) sets of integers", or "real numbers", or even "arbitrary sets of reals", do exist in some sense, and that propositions about them might be meaningful. However, it cannot be committed to the existence of such entities. Accordingly, one may formulate and use in such a theory propositions that refer to all sets. However, only those of them which are true independently of the exact extension of "the true universe V of sets" may be theorems. Therefore classical logic is acceptable, but there should be restrictions on principles that entail the *existence* "in the universe" of certain objects. Now the major existence principle of naive set theory is given by the comprehension scheme, and so it is this principle that should be restricted. We suggest that principle (PRE) means that the predicatively acceptable instances of the comprehension scheme are those which determine the collections they define in an absolute way, independently of any "surrounding universe". In other words: according to our interpretation of (PRE) in the context of set theory, a formula ψ is predicative (with respect to x) if the collection $\{x \mid \psi(x, y_1, \ldots, y_n)\}$ is completely and uniquely determined by the identity of the parameters $y_1, \ldots, y_n$, and the identity of other objects referred to in the formula (all of which should be well-determined beforehand). [3] Next we translate this idea into an exact definition. For simplicity of presentation, we assume in our definition the "platonic" cumulative universe V of ZF.

Notation. We denote by $Fv(exp)$ the set of free variables of exp, and by $\varphi\{t_1/x_1, \ldots, t_n\}$ the result of simultaneously substituting the term t_i for the free occurrences of x_i in φ ($i = 1, \ldots, n$).

Definition 2.1. Let T be a set theory, and let $Fv(\varphi) = \{y_1, \ldots, y_n, x_1, \ldots, x_k\}$. We say that φ is *predicative* in T for $\{x_1, \ldots, x_k\}$ if $\{\langle x_1, \ldots, x_k\rangle \mid \varphi\}$ is a set for all values of the parameters $y_1, \ldots, y_n$, and the following is true (in V) for every transitive model $\mathcal{M}$ of T:

$$\forall y_1 \ldots \forall y_n . y_1 \in \mathcal{M} \wedge \ldots \wedge y_n \in \mathcal{M} \rightarrow [\varphi \leftrightarrow (x_1 \in \mathcal{M} \wedge \ldots \wedge x_k \in \mathcal{M} \wedge \varphi_{\mathcal{M}})]$$

[3] Our notion of predicativity of formulas seems to be less restrictive than that used by Weyl and Feferman, since it makes the l.u.b. principle valid for predicatively acceptable sets of reals.

Thus a formula $\varphi(x)$ is predicative (in T) for x if it has the same extensions in all transitive models of T which contains the values of its other parameters. Note on the other hand that φ is predicative for $\emptyset$ iff it is absolute in the usual sense of set theory. (see e.g. [33]).

The main problem in formulating a predicative, type-free, set theory is how to syntactically impose this predicativity property on formulas without introducing syntactic types or levels. The solution suggested here to this problem comes from the observation that this is an instance of a more general task, not peculiar only to set Theory. In fact, in [3] and [6] an appropriate purely logical framework that can be used for this task has been introduced. This framework unifies different notions of "safety" of formulas, coming from different areas of mathematics and computer science, like: domain independence in database theory ([1, 48]), decidability of arithmetical formulas in computability theory and metamathematics, and absoluteness in set theory. In the next definition we review (an improved version of) this framework.

Definition 2.2.

1. Let $Fv(\varphi) = \{x_1, \dots, x_n, y_1, \dots, y_m\}$, and let S_1 and S_2 be two structures. φ is d.i. (domain-independent) for S_1 and S_2 with respect to $\{x_1, \dots, x_n\}$ (notation: $\varphi \succ^{S_1;S_2} \{x_1, \dots, x_n\}$), if for all $b_1 \dots, b_m \in S_1 \cap S_2$: [4]
$$\{\overrightarrow{x} \in S_2^n \mid S_2 \models \varphi(\overrightarrow{x}, \overrightarrow{b})\} = \{\overrightarrow{x} \in S_1^n \mid S_1 \models \varphi(\overrightarrow{x}, \overrightarrow{b})\}$$

2. A *safety-signature* is a pair (σ, F), where σ is an ordinary first-order signature with equality and *no function symbols*, and F is a function which assigns to every n-ary predicate symbol from σ (other than equality) a subset of $\mathcal{P}(\{1, \dots, n\})$.

3. Let (σ, F) be a safety-signature, and let S_1 and S_2 be structures for σ. S_1 and S_2 are $(\sigma, F)-$*compatible* if:
   a) Constants are interpreted identically in S_1 and S_2.
   b) $p(x_1, \dots, x_n) \succ^{S_1;S_2} \{x_{i_1}, \dots, x_{i_k}\}$ in case p is n-ary, $x_1, \dots, x_n$ are distinct, and $\{i_1, \dots, i_k\} \in F(p)$.

[4]Below we use the informal notation $S \models \varphi(a_1, \dots, a_n)$ (or even just $\varphi(a_1, \dots, a_n)$, in case S is the "universe of sets") instead of the more precise, but cumbersome, "$S, V \models \varphi$, where $Fv(\varphi) = \{x_1, \dots, x_n\}$, and V is an assignment in S such that $V(x_i) = a_i$ $(i = 1, \dots, n)$". This notation should not be confused with the notation $\varphi\{t_1/x_1, \dots, t_n/x_n\}$ for substituting terms of a language for variables. The informal notation $\{\overrightarrow{x} \in S^n \mid S \models \varphi(\overrightarrow{x}, \overrightarrow{b})\}$ has a similar obvious meaning. Note also that for convenience we use the same name (e.g. S) for a structure and for its domain.

4. S_2 is a $(\sigma, F)-$*extension* of S_1 if S_1 and S_2 are $(\sigma, F)-$compatible, and $S_1 \subseteq S_2$.

5. Let (σ, F) be a safety signature.
   - A formula φ is called $(\sigma, F)-$*safe w.r.t.* X ($\varphi \succ_{(\sigma,F)} X$) if $\varphi \succ^{S_1;S_2} X$ whenever S_1 and S_2 are $(\sigma, F)-$compatible.
   - φ is $(\sigma, F)-$*d.i.* if $\varphi \succ_{(\sigma,F)} Fv(\varphi)$.
   - φ is $(\sigma, F)-$*absolute* if $\varphi \succ_{(\sigma,F)} \varnothing$.

Examples.

- Let $\sigma_{\overrightarrow{P}} = \{P_1, \ldots, P_k\}$. Assume that the arity of P_i is n_i, and define $F_{\overrightarrow{P}}(P_i) = \{\{1, \ldots, n_i\}\}$. Then φ is $(\sigma_{\overrightarrow{P}}, F_{\overrightarrow{P}})-$d.i. iff it is domain-independent in the sense of database theory (see [1, 48]).

- Let $\sigma_{\mathcal{N}} = \{0, <, P_+, P_\times\}$, where 0 is a constant, $<$ is binary, and $P_+, P_\times$ are ternary. Define $F_{\mathcal{N}}(<) = \{\{1\}\}$, $F_{\mathcal{N}}(P_+) = F_{\mathcal{N}}(P_\times) = \{\varnothing\}$. Then the standard structure $\mathcal{N}$ for $\sigma_{\mathcal{N}}$ (with the usual interpretations of 0 and $<$, and the (graphs of the) operations $+$ and $\times$ on N as the interpretations of P_+ and $P_\times$, respectively) is a $(\sigma_{\mathcal{N}}, F_{\mathcal{N}})$-extension of a structure S for $\sigma_{\mathcal{N}}$ iff the domain of S is an initial segment of $\mathcal{N}$ (where the interpretations of the relation symbols are the corresponding reductions of the interpretations of those symbols in $\mathcal{N}$). It was shown in [6] that every Δ_0-formula of $\sigma_{\mathcal{N}}$ is $(\sigma_{\mathcal{N}}, F_{\mathcal{N}})$-absolute, that every $(\sigma_{\mathcal{N}}, F_{\mathcal{N}})$-absolute formula defines a decidable relation on the set of natural numbers, and that a relation on the natural numbers is r.e. iff it is definable by a formula of the form $\exists y_1, \ldots, y_n \psi$, where the formula ψ is $(\sigma_{\mathcal{N}}, F_{\mathcal{N}})$-absolute.

- Let $\sigma_{ZF} = \{\in\}$ and let $F_{ZF}(\in) = \{\{1\}\}$. Then S_2 is a $(\sigma_{ZF}, F_{ZF})-$extension of S_1 iff $S_1 \subseteq S_2$, and $x_1 \in x_2 \succ^{S_1;S_2} \{x_1\}$. The latter condition means that S_1 is a transitive substructure of S_2 (In particular, the universe V is a $(\sigma_{ZF}, F_{ZF})-$extension of the transitive sets and classes). Accordingly, $\varphi(x_1, \ldots, x_n, y_1, \ldots, y_k) \succ_{(\sigma_{ZF}, F_{ZF})} \{x_1, \ldots, x_n\}$ iff the following holds whenever $S_1 \cap S_2$ is *transitive*, and $y_1, \ldots, y_k$ are assigned values from $S_1 \cap S_2$:

$$\{\langle x_1, \ldots, x_n\rangle \mid S_1 \models \varphi\} = \{\langle x_1, \ldots, x_n\rangle \mid S_2 \models \varphi\}$$

  In particular, a formula is (σ_{ZF}, F_{ZF})-absolute iff it is absolute in the usual sense this notion is used in set theory.

Obviously, "domain independence" and "predicativity" in the sense of "universe independence" are very close relatives. Accordingly, a plausible interpretation of principle (PRE) is that φ is predicative with respect to x iff $\varphi \succ_{(\sigma_{ZF},F_{ZF})} \{x\}$. However, it follows from results in [6] that the relation $\succ_{(\sigma_{ZF},F_{ZF})}$ is undecidable. Therefore in order to base predicative formal systems on this interpretation of principle (PRE) we should replace the semantic relation of (σ, F)-safety by a useful syntactic approximation. Now the most natural way to define a syntactic approximation of a semantic logical relation concerning formulas is by a structural induction. Such an inductive definition should be based on the behavior with respect to the original semantic relation of the atomic formulas and of the logical connectives and quantifiers. The next theorem from [6] lists the most obvious and useful relevant properties that every relation $\succ_{(\sigma,F)}$ has in the first-order framework:

Theorem 2.3. *$\succ_{(\sigma,F)}$ has the following properties:*

1. *$p(t_1,\dots,t_n) \succ_{(\sigma,F)} X$ in case p is an n-ary predicate symbol of σ, and there is $I \in F(p)$ such that:*
   a) *For every $x \in X$ there is $i \in I$ such that $x = t_i$.*
   b) *$X \cap Fv(t_j) = \emptyset$ for every $j \in \{1,\dots,n\} - I$.*
2. a) *$\varphi \succ_{(\sigma,F)} \{x\}$ if $\varphi \in \{x \neq x, x = t, t = x\}$, and $x \notin Fv(t)$.*
   b) *$t = s \succ_{(\sigma,F)} \emptyset$.*
3. *$\neg\varphi \succ_{(\sigma,F)} \emptyset$ if $\varphi \succ_{(\sigma,F)} \emptyset$.*
4. *$\varphi \vee \psi \succ_{(\sigma,F)} X$ if $\varphi \succ_{(\sigma,F)} X$ and $\psi \succ_{(\sigma,F)} X$.*
5. *$\varphi \wedge \psi \succ_{(\sigma,F)} X \cup Y$ if $\varphi \succ_{(\sigma,F)} X$, $\psi \succ_{(\sigma,F)} Y$, and $Y \cap Fv(\varphi) = \emptyset$.*
6. *$\exists y\varphi \succ_{(\sigma,F)} X - \{y\}$ if $y \in X$ and $\varphi \succ_{(\sigma,F)} X$.*
7. *If $\varphi \succ_{(\sigma,F)} \{x_1,\dots,x_n\}$, and $\psi \succ_{(\sigma,F)} \emptyset$, then $\forall x_1 \dots x_n(\varphi \to \psi) \succ_{(\sigma,F)} \emptyset$.*

By a "safety relation" we shall henceforth mean a relation $\succ$ between formulas of σ_{ZF} and finite sets of variables which satisfies the clauses in Theorem 2.3 with respect to F_{ZF}[5]. The least safety relation is a plausible syntactic approximation of predicativity. However, a better approximation is obtained if greater power is given to the first two clauses by providing a much more extensive set of terms than that

[5]Property 7 is easily derivable from the others. Hence if $\forall$ and $\to$ are taken as defined in terms of the other logical constants, then the same relation is obtained if we omit property 7 from the list in Theorem 2.3.

provided by σ_{ZF} (the only terms of which are its variables). This is achieved by allowing $\{x \mid \psi\}$ to be a legal term whenever $\psi \succ \{x\}$. Note that this is in full coherence with our intended meaning of $\succ$. Moreover, this move is still justified by Theorem 2.3, since its proof remains valid also for languages which include complex terms (not just variables and constants), as long as $x = t \succ_{(\sigma,F)} \{x\}$ whenever $x \notin Fv(t)$.

2.2 Interpreting and Implementing Principle (NAT)

First we note that by "acceptance of the set N of natural numbers" we understand here also acceptance of principles and ideas implicit in the construction of N. This includes proofs by mathematical induction, as well as the idea of iterating (an operation or a relation) an arbitrary (finite) number of times. Hence finitary inductive definitions of sets, relations, and functions are accepted. In particular, the ability to form the transitive closure of a given relation (like forming the notion of an ancestor from the notion of a parent) should be taken as a major ingredient of our logical abilities (even prior to our understanding of the natural numbers). In fact, in [2] it was argued that this concept is the key for understanding finitary inductive definitions and reasoning, and evidence was provided for the thesis that systems which are based on it provide the right framework for the formalization and mechanization of mathematics. This suggestion will be used as our main tool for implementing (NAT). Hence in addition to allowing the use of set terms we shall also go beyond FOL (First-Order Logic) by introducing an operation TC for transitive closure[6]. The corresponding language and semantics are defined as follows (see, e.g., [13, 28–30, 34, 47]):

Definition 2.4. Let σ be a signature for a first-order language with equality. The language $L^1_{TC}(\sigma)$ is defined like the usual first-order language which is based on σ, but with the addition of the following clause: If φ is a formula, x, y are distinct variables, and t, s are terms, then $(TC_{x,y}\varphi)(s,t)$ is a formula (in which all occurrences of x and y in φ are bound). The intended meaning of $(TC_{x,y}\varphi)(s,t)$ is the following "infinite disjunction": (where $w_1, w_2, \ldots,$ are all new variables):

$$\varphi\{s/x, t/y\} \vee \exists w_1(\varphi\{s/x, w_1/y\}) \wedge \varphi\{w_1/x, t/y\})\vee$$
$$\vee\, \exists w_1 \exists w_2(\varphi\{s/x, w_1/y\} \wedge \varphi\{w_1/x, w_2/y\} \wedge \varphi\{w_2/x, t/y\}) \vee \ldots$$

The most important relevant facts shown in [2] concerning TC are:

[6] It is well known (see [47]) that the language of FOL enriched with TC is equivalent in its expressive power to the language of weak SOL. So taking "transitive closure" as primitive is equivalent to taking "finite set" as primitive (which is the approach of [23], though the system presented there is essentially first-order). We prefer the former as primitive, because it allows a very natural treatment of induction as a logical rule, as well as a neat extension of the safety relation - see below.

1. If σ contains a constant 0 and a (symbol for a) pairing function, then all types of finitary inductive definitions of relations and functions (as defined by Feferman in [21]) are available in $L^1_{TC}(\sigma)$. This result, in turn, allows for presenting a simple version of Feferman's framework FS_0, demonstrating that TC-logics provide an excellent framework for mechanizing formal systems.

2. Let V_0 be the smallest set including 0 and closed under the operation of pairing. Then a subset S of V_0 is recursively enumerable iff there exists a formula $\varphi(x)$ of $\mathcal{P}TC^+$ such that $S = \{x \in V_0 \mid \varphi(x)\}$, where the language $\mathcal{P}TC^+$ is defined as follows:

   Terms of $\mathcal{P}TC^+$

   a) The constant 0 is a term.
   b) Every (individual) variable is a term.
   c) If t and s are terms then so is (t, s).

   Formulas of $\mathcal{P}TC^+$

   a) If t and s are terms then $t = s$ is a formula.
   b) If φ and ψ are formulas then so are $\varphi \vee \psi$ and $\varphi \wedge \psi$.
   c) If φ is a formula, x, y are two different variables, and t, s are terms, then $(TC_{x,y}\varphi)(t, s)$ is a formula.

3. By generalizing a particular case which has been used by Gentzen in [26], mathematical induction can be presented as a logical rule of languages with TC. Indeed, Using a Gentzen-type format, a general form of this principle can be formulated as follows:

$$\frac{\Gamma, \psi, \varphi \Rightarrow \Delta, \psi\{y/x\}}{\Gamma, \psi\{s/x\}, (TC_{x,y}\varphi)(s, t) \Rightarrow \Delta, \psi\{t/x\}}$$

   where x and y are not free in Γ, Δ, and y is not free in ψ.

Now in order to combine the two central ideas described above, a clause concerning TC should be added to the list of clauses in Theorem 2.3. Such a clause was suggested in [2]. To understand it, let us look at the first three disjuncts in the infinite disjunction θ which corresponds to $(TC_{x,y}\varphi)(x, y)$:

$$\varphi(x, y) \vee \exists w_1(\varphi(x, w_1) \wedge \varphi(w_1, y)) \vee \exists w_1 \exists w_2(\varphi(x, w_1) \wedge \varphi(w_1, w_2) \wedge \varphi(w_2, y))$$

Call this finite disjunction ψ. From the clauses in Theorem 2.3 concerning $\wedge, \exists$ and $\vee$ it follows that if $\varphi \succ_{(\sigma,F)} X$ and $y \in X$ (or $x \in X$) then $\psi \succ_{(\sigma,F)} X$. This remains true for every finite subdisjunction of θ. Hence every such finite subdisjunction is safe with respect to X, and this easily implies that so is the whole disjunction. This observation leads to the following new condition (in which the variables x and y may be elements of X):

- $(TC_{x,y}\varphi)(x,y) \succ_{(\sigma,F)} X$ if either $\varphi \succ_{(\sigma,F)} X \cup \{x\}$ or $\varphi \succ_{(\sigma,F)} X \cup \{y\}$.

3 PZF and Its Formal Counterparts

In this section we use the ideas described in the previous section for introducing a family of systems for predicative set theory. All these systems share the same language and the same axioms. They differ only with respect to the strength of their formal underlying apparatus. We shall denote by PZF the strongest (and non-axiomatizable) system in this family.

3.1 Language

We define the terms and formula of the language $\mathcal{L}_{PZF}$, as well as the safety relation $\succ_{PZF}$ between formulas and finite sets of variables, by simultaneous recursion as follows (where $Fv(exp)$ denotes the set of free variables of exp):

Terms:

- Every variable is a term.
- If x is a variable, and φ is a formula such that $\varphi \succ_{PZF} \{x\}$, then $\{x \mid \varphi\}$ is a term (and $Fv(\{x \mid \varphi\}) = Fv(\varphi) - \{x\}$).[7]

Formulas:

- If t and s are terms than $t = s$ and $t \in s$ are atomic formulas.
- If φ and ψ are formulas, and x is a variable, then $\neg\varphi$, $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$, and $\exists x\varphi$ are formulas (where $Fv(\exists x\varphi) = Fv(\varphi) - \{x\}$).

[7] Note that for convenience, we use in this paper the notation $\{x \mid \varphi\}$ in the formal language $\mathcal{L}_{PZF}$ as well as in our metalanguage. This should not cause a confusion.

- If φ is a formula, t and s are terms, and x and y are distinct variables then $(TC_{x,y}\varphi)(t,s)$ is a formula, and

$$Fv((TC_{x,y}\varphi)(t,s)) = (Fv(\varphi) - \{x,y\}) \cup Fv(t) \cup Fv(s)$$

The Safety Relation $\succ_{PZF}$:

1. a) $\varphi \succ_{PZF} \varnothing$ if φ is atomic.
   b) $\neg\varphi \succ_{PZF} \varnothing$ if $\varphi \succ_{PZF} \varnothing$.
2. $\varphi \succ_{PZF} \{x\}$ if $\varphi \in \{x \in x, x = t, t = x, x \in t\}$, and $x \notin Fv(t)$.
3. $\varphi \vee \psi \succ_{PZF} X$ if $\varphi \succ_{PZF} X$ and $\psi \succ_{PZF} X$.
4. $\varphi \wedge \psi \succ_{PZF} X \cup Y$ if $\varphi \succ_{PZF} X, \psi \succ_{PZF} Y$ and either $Y \cap Fv(\varphi) = \varnothing$ or $X \cap Fv(\psi) = \varnothing$.
5. $\exists y\varphi \succ_{PZF} X - \{y\}$ if $y \in X$ and $\varphi \succ_{PZF} X$.
6. $(TC_{x,y}\varphi)(x,y) \succ_{PZF} X$ if $\varphi \succ_{PZF} X \cup \{x\}$, or $\varphi \succ_{PZF} X \cup \{y\}$.

Note 3.1. The intended *intuitive* meaning of "$\varphi \succ_{PZF} \{y_1,\ldots,y_k\}$", where $Fv(\varphi) = \{x_1,\ldots,x_n,y_1,\ldots,y_k\}$, is that for every "accepted" sets $a_1,\ldots,a_n$, the collection of all tuples $\langle y_1,\ldots,y_k\rangle$ such that $\varphi(a_1,\ldots,a_n,y_1,\ldots,y_k)$ is a *set* which is constructed in an absolute, "universe independent" way from previously "accepted" sets and from (elements in the transitive closure of) $a_1,\ldots,a_n$. Since this is an imprecise explanation, it cannot be proved in the strict sense of the word. However, it is not difficult to convince oneself that $\succ_{PZF}$ indeed has this property. For example, assume that $\theta = \varphi \wedge \psi$, where $Fv(\varphi) = \{x,z\}, Fv(\psi) = \{x,y,z\}, \varphi \succ_{PZF} \{x\}$, and $\psi \succ_{PZF} \{y\}$. Given some absolute set c, by induction hypothesis the collection $Z(c)$ of all x such that $\varphi(x,c)$ is an absolute set. Again by induction hypothesis, for every d in this set the collection $W(c,d)$ of all y such that $\psi(d,y,c)$ is an absolute set. Now the collection of all $\langle x,y\rangle$ such that $\theta(x,y,c)$ is the union for $d \in Z(c)$ of the sets $\{d\} \times W(c,d)$. Hence it is a set containing only previously accepted, absolute collections, and its identity is obviously absolute too. This is exactly what $\theta \succ_{PZF} \{x,y\}$ (which holds in this case by the clause concerning conjunction in the definition of $\succ_{PZF}$) intuitively means.

Note 3.2. Officially, the language we use does not include the universal quantifier $\forall$ and the implication connective $\rightarrow$. Below they are taken therefore as defined (in the usual way) in terms of the official connectives and $\exists$.

Note 3.3. It is not difficult to show that $\succ_{PZF}$ has the following properties:

- If $\varphi \succ_{PZF} X$ then $X \subseteq Fv(\varphi)$.
- If $\varphi \succ_{PZF} X$ and $Z \subseteq X$, then $\varphi \succ_{PZF} Z$.
- If $\varphi \succ_{PZF} \{x_1, \ldots, x_n\}$, $v_1, \ldots v_n$ are n distinct variables not occurring in φ, and φ' is obtained from φ by replacing *all* (not only the free) occurrences of x_i by v_i ($i = 1, \ldots, n$), then $\varphi' \succ_{PZF} \{v_1, \ldots, v_n\}$.
- If $x \notin Fv(t)$, and $\varphi \succ_{PZF} \varnothing$, then both $\forall x(x \in t \rightarrow \varphi) \succ_{PZF} \varnothing$, and $\exists x(x \in t \land \varphi) \succ_{PZF} \varnothing$. Hence $\varphi \succ_{PZF} \varnothing$ for every Δ_0 formula φ in $\mathcal{L}_{ZF}$.

The following proposition can easily be proved:

Proposition 3.4. *There is an algorithm which given a string of symbols E determines whether E is a term of $\mathcal{L}_{PZF}$, a formula of $\mathcal{L}_{PZF}$, or neither, and in case E is a formula it returns the set of all X such that $E \succ_{PZF} X$.*

3.2 Axioms

We turn to the axioms of PZF and its formal counterparts. The basic idea here is to use a version of the "ideal calculus" ([14]) for naive set theory, in which the comprehension schema is applicable only to safe formulas. In addition we include also $\in$-induction, which seems to be quite natural within a predicative framework. Here is the resulting list of axioms:

Extensionality:

- $\forall z(z \in x \leftrightarrow z \in y) \rightarrow x = y$

The Comprehension Schema: [8]

- $\forall x(x \in \{x \mid \varphi\} \leftrightarrow \varphi)$

The Regularity Schema ($\in$-induction):

- $(\forall x(\forall y(y \in x \rightarrow \varphi\{y/x\}) \rightarrow \varphi)) \rightarrow \forall x\varphi$

[8] This name is justified here because for φ which is predicative with respect to x (i.e. $\varphi \succ_{PZF} \{x\}$) it easily entails the usual formulation: $\exists Z \forall x(x \in Z \leftrightarrow \varphi)$.

3.3 Logic

The logic which underlies PZF is TC-logic (transitive closure logic, also called ancestral logic): the logic which corresponds to ordinary first-order logic (with equality) augmented with TC, the operator which produces the transitive closure of a given binary relation. Now the set of valid formulas of this logic is not r.e. (or even arithmetical). Hence no sound and complete *formal* system for it exists. It follows that PZF, our version of predicative set theory, cannot be fully formalized. The problem whether the above set of axioms is sound and complete for predicative set theory should therefore be understood as being relative to this underlying logic. This means that according to our approach, *no single formal system can capture the whole of predicative mathematics*. It also follows that the problem of producing formal systems for actually using PZF (for making formal deductions in predicative mathematics) reduces to finding appropriate formal approximations of this underlying *logic*. Hence what we introduce here together with PZF is really a family of formal systems.

One crucial logical rule that should be available in any such approximation is the general rule of induction formulated in subsection 2.2:

$$\frac{\Gamma, \psi, \varphi \Rightarrow \Delta, \psi\{y/x\}}{\Gamma, \psi\{s/x\}, (TC_{x,y}\varphi)(s,t) \Rightarrow \Delta, \psi\{t/x\}}$$

(where x and y are not free in Γ, Δ, and y is not free in ψ). Two other obvious rules introduce TC on the right hand side of sequents: [9]

$$\frac{\Gamma \Rightarrow \Delta, \varphi\{t/x, s/y\}}{\Gamma \Rightarrow \Delta, (TC_{x,y}\varphi)(t,s)}$$

$$\frac{\Gamma \Rightarrow \Delta, (TC_{x,y}\varphi)(r,s) \qquad \Gamma \Rightarrow \Delta, (TC_{x,y}\varphi)(s,t)}{\Gamma \Rightarrow \Delta, (TC_{x,y}\varphi)(r,t)}$$

Henceforth we denote by PZF_0 the formal approximation of PZF in which the underlying formal logic is the extension of first-order logic with these three rules for TC. PZF_0 suffices for everything we do below, and we believe (but this remains to be confirmed) that it should in fact suffice for (most of) applicable mathematics. Now PZF_0 is relatively a week system. Thus it can easily be interpreted in Kripke-Platek set theory KP together with the infinity axiom (see [7, 11, 31])[10]. However, it should again be emphasized that PZF as a whole is open-ended, and transcends any given formal system.

[9] The resulting system is equivalent to Myhill's system for ancestral logic in [34].

[10] KP itself includes the Δ_0-collection schema, which is not predicatively justified.

Note 3.5. In addition to having TC (which is the major difference between our underlying logic and FOL), one should also note that the language of PZF provides a class of terms which is much richer than those allowed in orthodox first-order systems. In particular: a variable can be bound in it within a term. The notion of a term being free for substitution should be generalized accordingly (also for substitutions within terms!). As usual this amounts to avoiding the capture of free variables within the scope of an operator which binds them. Otherwise the rules/axioms concerning the quantifiers and terms remain unchanged (for example: $\forall x\varphi \rightarrow \varphi\{t/x\}$ is valid for *every* term t which is free for x in φ). We also assume α-conversion to be a part of the logic[11].

For simplicity of presentation and understanding, we again assume in the rest of this paper the platonic cumulative universe V (although its exact extension is irrelevant). Predicatively meaningful counterparts of our various claims can be formulated and proved, but we leave this task to another opportunity.

The straightforward proof of the following proposition was practically given in Note 3.1 (see [5] for a proof of a stronger claim):

Proposition 3.6. *V is a model of PZF.*

4 The Expressive Power of PZF

4.1 Some Standard Notations for Sets

In $\mathcal{L}_{PZF}$ we can introduce as *abbreviations* most of the standard notations for sets used in mathematics. Note that all these abbreviations can be introduced in a purely static way: unlike in the extension by definition procedure (see [46]), no formal proofs within the system (of corresponding justifying existence and uniqueness propositions) are needed before introducing them.

- $\emptyset =_{Df} \{x \mid x \in x\}$.
- $\{t_1, \ldots, t_n\} =_{Df} \{x \mid x = t_1 \vee \ldots \vee x = t_n\}$ (where x is new).
- $\langle t, s\rangle =_{Df} \{\{t\}, \{t, s\}\}$
- $\{x \in t \mid \varphi\} =_{Df} \{x \mid x \in t \wedge \varphi\}$, provided $\varphi \succ_{PZF} \emptyset$. (where $x \notin Fv(t)$).
- $\{t \mid x \in s\} =_{Df} \{y \mid \exists x.x \in s \wedge y = t\}$ (where y is new, and $x \notin Fv(s)$).

[11] Other rules, like substitution of equals for equals within any context (under the usual conditions concerning bound variables) are derivable from the usual first-order axioms for equality by using the axioms of PZF.

- $s \times t =_{Df} \{x \mid \exists a \exists b. a \in s \wedge b \in t \wedge x = \langle a, b\rangle\}$ (where x, a and b are new).
- $s \cap t =_{Df} \{x \mid x \in s \wedge x \in t\}$ (where x is new).
- $s \cup t =_{Df} \{x \mid x \in s \vee x \in t\}$ (where x is new).
- $S(x) =_{Df} x \cup \{x\}$
- $\bigcup t =_{Df} \{x \mid \exists y. y \in t \wedge x \in y\}$ (where x and y are new).
- $\bigcap t =_{Df} \{x \mid x \in \bigcup t \wedge \forall y(y \in t \rightarrow x \in y)\}$ (where x, y are new).
- $\iota x\varphi =_{Df} \bigcap\{x \mid \varphi\}$ (provided $\varphi \succ_{PZF} \{x\}$).
- $P_1(z) =_{Df} \iota x. \exists v \exists y(v \in z \wedge x \in v \wedge y \in v \wedge z = \langle x, y\rangle)$ ($\vdash_{PZF_0} P_1(\langle t, s\rangle) = t$).
- $P_2(z) =_{Df} \iota y. \exists v \exists x(v \in z \wedge x \in v \wedge y \in v \wedge z = \langle x, y\rangle)$ ($\vdash_{PZF_0} P_2(\langle t, s\rangle) = s$).
- $\omega =_{Df} \{x \mid x = \varnothing \vee \exists y. y = \varnothing \wedge (TC_{x,y}(x = S(y)))(x, y)\}$
- $TH(x) =_{Df} x \cup \{y \mid (TC_{x,y} y \in x)(x, y)\}$ (the transitive hull of x).

Our term above for $\bigcap t$ is valid (and so denotes a set) whenever t is valid. It is easy to see that if t denotes a non-empty set A, then $\bigcap t$ indeed denotes the intersection of all the elements of A. On the other hand, if the set denoted by t is empty, then the set denoted by the term $\bigcap t$ is empty as well. With the help of the extensionality axiom this in turn implies that our term above for $\iota x\varphi$ is meaningful for *every* φ such that $\varphi \succ_{PZF} \{x\}$. This term denotes $\varnothing$ if there is no set which satisfies φ, and it denotes the intersection of all the sets which satisfy φ otherwise. In particular: if there is exactly one set which satisfies φ then $\iota x\varphi$ denotes this unique set. All these facts are theorems of PZF_0. In particular we have:

Proposition 4.1. *If* $\varphi \succ_{PZF} \{x\}$ *then* $\vdash_{PZF_0} \exists! x\varphi \rightarrow \forall x(\varphi \leftrightarrow x = \iota x\varphi)$.

From Proposition 4.1 it follows that if a formula $\varphi(y_1, \ldots, y_n, x)$ implicitly defines in PZF a function f_φ such that for all $y_1, \ldots, y_n$, $f_\varphi(y_1, \ldots, y_n)$ is the unique x such that $\varphi(y_1, \ldots, y_n, x)$, and if $\varphi \succ_{PZF} \{x\}$, then there is a term in PZF which explicitly denotes f_φ, and no extension by definitions of the language is needed for introducing it. Moreover: in PZF we can introduce as abbreviations the terms used in the λ-calculus for handling explicitly defined functions (except that our terms for functions should specify the domains of these functions, which should be explicitly definable sets):

- $\lambda x \in s.t =_{Df} \{\langle x, t\rangle \mid x \in s\}$ (where $x \notin Fv(s)$)
- $f(x) =_{Df} \iota y.\exists z \exists v(z \in f \wedge v \in z \wedge y \in v \wedge z = \langle x, y\rangle)$
- $Dom(f) =_{Df} \{x \mid \exists z \exists v \exists y(z \in f \wedge v \in z \wedge y \in v \wedge x \in v \wedge z = \langle x, y\rangle\}$
- $Rng(f) =_{Df} \{y \mid \exists z \exists v \exists x(z \in f \wedge v \in z \wedge y \in v \wedge x \in v \wedge z = \langle x, y\rangle\}$
- $f \upharpoonright s =_{Df} \{\langle x, f(x)\rangle \mid x \in s\}$ (where x is new).

Identifying $\perp$ from domain theory with $\varnothing$, we can easily check now that rules β and η obtain in PZF:

- $\vdash_{PZF_0} u \in s \to (\lambda x \in s.t)u = t\{u/x\}$ (if u is free for x in t).
- $\vdash_{PZF_0} u \notin s \to (\lambda x \in s.t)u = \perp$ (if u is free for x in t).
- $\vdash_{PZF_0} \lambda x \in s.tx = t \upharpoonright s$ (in case $x \notin Fv(t)$).

4.2 RST and Rudimentary Functions

Let $\mathcal{L}_{RST}$ and $\succ_{RST}$ be defined like $\mathcal{L}_{PZF}$ and $\succ_{PZF}$ (respectively), but without using the TC operator. Let RST be the first-order system in $\mathcal{L}_{RST}$ which is based on the three axioms of PZF (and with a suitable version of ordinary first-order logic as the underlying logic). It should be noted that with the exception of ω and $TH(x)$, all the constructions above have actually been done in the framework of $\mathcal{L}_{RST}$ (and can be justified in RST). Now HF, the set of hereditarily finite sets, is a model of RST. Hence ω is not definable in $\mathcal{L}_{RST}$, and so TC is indeed necessary for its definition. [12]

Note 4.2. RST can be shown to be equivalent to Gandy's basic set theory ([25]), and to the system called BST_0 in [43].

The following theorem and its two corollaries determine the expressive power of $\mathcal{L}_{RST}$, and connect it (and $\succ_{RST}$) with the class of rudimentary set functions — a refined version of Gödel basic set functions (from [27]) which was independently introduced by Gandy in [25] and Jensen in [32] (See also [10]).

[12] It is known (see e.g. [25]) that the property of being a finite ordinal is definable by a Δ_0-formula $\varphi(x)$, but this φ does not satisfy $\varphi \succ_{PZF} \{x\}$ (it only satisfies $\varphi \succ_{RST} \varnothing$, like any other Δ_0-formula). Hence $\{x \mid \varphi\}$ is not a legal term of RST.

Theorem 4.3.

1. *If F is an n-ary rudimentary function, then there exists a formula φ_F with the following properties:*
   a) $Fv(\varphi_F) \subseteq \{y, x_1, \ldots, x_n\}$
   b) $\varphi_F \succ_{RST} \{y\}$
   c) $F(x_1, \ldots, x_n) = \{y \mid \varphi_F\}$.
2. *If φ is a formula of $\mathcal{L}_{RST}$ such that:*
   a) $Fv(\varphi) \subseteq \{y_1, \ldots, y_k, x_1, \ldots, x_n\}$
   b) $\varphi \succ_{RST} \{y_1, \ldots, y_k\}$

   then there exists a rudimentary function F_φ such that:

$$F_\varphi(x_1, \ldots, x_n) = \{\langle y_1, \ldots, y_k\rangle \mid \varphi\}$$
$$(= \{x \mid \exists y_1, \ldots, y_k.x = \langle y_1, \ldots, y_k\rangle \wedge \varphi\}).$$

3. *If t is a term of $\mathcal{L}_{RST}$ such that $Fv(t) \subseteq \{x_1, \ldots, x_n\}$, then there exists a rudimentary function F_t such that $F_t(x_1, \ldots, x_n) = t$ for every $x_1, \ldots, x_n$.*

Proof: We prove part (1) by induction, following the definition of the rudimentary functions given in [10]:

- If $F(x_1, \ldots, x_n) = x_i$ then φ_F is $y \in x_i$. Here $\varphi_F \succ_{RST} \{y\}$ by clause (2) of the definition of $\succ_{RST}$.
- If $F(x_1, \ldots, x_n) = \{x_i, x_j\}$ then φ_F is $y = x_i \vee y = x_j$. Here $\varphi_F \succ_{RST} \{y\}$ by clauses (2) and (3) of the definition of $\succ_{RST}$.
- If $F(x_1, \ldots, x_n) = x_i - x_j$ then φ_F is $y \in x_i \wedge \neg(y \in x_j)$. Here $\varphi_F \succ_{RST} \{y\}$ by clause (2), (1a), (1b), and (4) of the definition of $\succ_{RST}$.
- Suppose $F(x_1, \ldots, x_n) = H(G_1(x_1, \ldots, x_n), \ldots, G_k(x_1, \ldots, x_n))$, where H and $G_1, \ldots, G_k$ are rudimentary. Let $w_1, \ldots, w_k$ be new variables. Then φ_F is $\exists w_1 \ldots w_k(w_1 = \{y \mid \varphi_{G_1}\} \wedge \ldots \wedge w_k = \{y \mid \varphi_{G_k}\} \wedge \varphi_H(y, w_1, \ldots, w_k))$. Here $\varphi_F \succ_{RST} \{y\}$ by clauses (2), (4), and (5) of the definition of $\succ_{RST}$.
- Suppose $F(x_1, \ldots, x_n) = \bigcup_{z \in x_1} G(z, x_2, \ldots, x_n)$, where G is rudimentary. Then φ_F is $\exists z(z \in x_1 \wedge \varphi_G(y, z, x_2, \ldots, x_n))$. Here again $\varphi_F \succ_{RST} \{y\}$ by clauses (2), (4), and (5) of the definition of $\succ_{RST}$.

Next we prove parts (2) and (3) together by induction on the complexity of φ and t.

- If t is x_i then $F_t(x_1, \ldots, x_n) = x_i$.
- If t is $\{y \mid \varphi\}$, where $\varphi \succ_{RST} \{y\}$, then $F_t = F_\varphi$.
- If φ is $t = s$ and $k = 0$ then

$$F_\varphi(x_1, \ldots, x_n) = \begin{cases} \{\varnothing\} & F_t(x_1, \ldots, x_n) = F_s(x_1, \ldots, x_n) \\ \varnothing & F_t(x_1, \ldots, x_n) \neq F_s(x_1, \ldots, x_n) \end{cases}$$

  The case in which φ is $t \in s$ and $k = 0$ is treated similarly.
- If φ is $\neg\psi$ and $k = 0$ then $F_\varphi(x_1, \ldots, x_n) = \{\varnothing\} - F_\psi(x_1, \ldots, x_n)$.
- If φ is $y_1 \neq y_1$ (and $k = 1$), then $F_\varphi(x_1, \ldots, x_n) = \varnothing$.
- If φ is $y_1 = t$ or $t = y_1$, where $y_1 \notin Fv(t)$ (and $k = 1$), then $F_\varphi(x_1, \ldots, x_n) = \{F_t(x_1, \ldots, x_n)\}$.
- If φ is $y_1 \in t$, where $y_1 \notin Fv(t)$ (and $k = 1$), then $F_\varphi(x_1, \ldots, x_n) = F_t(x_1, \ldots, x_n)$.
- If φ is $\psi_1 \vee \psi_2$ then $F_\varphi(x_1, \ldots, x_n) = F_{\psi_1}(x_1, \ldots, x_n) \cup F_{\psi_2}(x_1, \ldots, x_n)$.
- If φ is $\psi \wedge \theta$, where $\psi \succ_{RST} \{y_1, \ldots, y_l\}$ $(l \leq k)$, $\theta \succ_{RST} \{y_{l+1}, \ldots, y_k\}$, and $Fv(\psi) \cap \{y_{l+1}, \ldots, y_k\} = \varnothing$, then

$$F_\varphi(x_1, \ldots, x_n) = \bigcup_{\langle y_1, \ldots, y_l \rangle \in F_\psi(x_1, \ldots, x_n)} \; \bigcup_{\langle y_{l+1}, \ldots, y_k \rangle \in F_\theta(x_1, \ldots, x_n, y_1, \ldots, y_l)} \{\langle y_1, \ldots, y_k \rangle\}$$

- Suppose $\varphi = \exists z \psi$, where $\psi \succ_{RST} \{z, y_1, \ldots, y_k\}$. Then $F_\varphi(x_1, \ldots, x_n) = \bigcup_{\langle z, y_1, \ldots, y_k \rangle \in F_\psi(x_1, \ldots, x_n)} \{\langle y_1, \ldots, y_k \rangle\}$.

It is not difficult to see that all functions defined above are indeed rudimentary.

Corollary 4.4. *Every term of $\mathcal{L}_{RST}$ with n free variables explicitly defines an n-ary rudimentary function. Conversely, every rudimentary function is defined by some term of $\mathcal{L}_{RST}$.*

Corollary 4.5. *If $Fv(\varphi) = \{x_1, \ldots, x_n\}$, and $\varphi \succ_{RST} \varnothing$, then φ defines a rudimentary predicate P. Conversely, if P is a rudimentary predicate, then there is a formula φ such that $\varphi \succ_{RST} \varnothing$, and φ defines P.*

4.3 Recursion and Inductive Definitions

The inclusion of the operation TC in $\mathcal{L}_{PZF}$ strongly extends its expressive power. As a simple example of this power we take primitive recursion on ω:

Proposition 4.6. *Assume g is a function on ω^2 which is definable by a (closed) term of $\mathcal{L}_{PZF}$. Let f be a function on ω defined by $f(0) = a$, $f(n+1) = g(n, f(n))$ (where a is definable in $\mathcal{L}_{PZF}$). Then f is definable (as a set of pairs) by a closed term of $\mathcal{L}_{PZF}$.*

Proof: Assume t_g is a term which defines g in $\mathcal{L}_{PZF}$. Let $\psi_1(z, w)$ be the formula $(TC_{z,w} w = \langle S(P_1(z)), t_g(z)\rangle)(z, w)$ (note that we use here the notation for function application which was introduced in subsection 4.1). Let ψ_2 be the formula $z = \langle 0, a\rangle \wedge \psi_1(z, w) \wedge P_1(w) = n \wedge P_2(w) = x$, and φ the formula $\exists z \exists w \psi_2$. Since $w = \langle S(P_1(z)), t_g(z)\rangle \succ_{PZF} \{w\}$, also $\psi_1 \succ_{PZF} \{w\}$ (by the clause concerning TC in the definition of $\succ_{PZF}$). Hence $\psi_2 \succ_{PZF} \{z, w, n, x\}$ (by the clauses concerning $\wedge$ and $=$ in the definition of $\succ_{PZF}$). It follows that $\varphi \succ_{PZF} \{n, x\}$, and so $\iota x \varphi$ is defined. Since it is easy to prove by induction that $\vdash_{PZF_0} \forall n \in \omega \exists! x \varphi$, Proposition 4.1 entails that $\lambda n \in \omega . \iota x \varphi$ is a term as required.

Proposition 4.6 is only a special case of the following much more general theorem, which implies that all types of *finitary inductive* definitions (as characterized in [21]) are available in $\mathcal{L}_{PZF}$. Its proof is similar to the proof of Theorem 15 in [2]:

Theorem 4.7. *For $1 \leq j \leq p$, let $\varphi_1(y, x_1, \ldots, x_{n_1}), \ldots, \varphi_p(y, x_1, \ldots, x_{n_p})$ be p formulas such that $\varphi_j \succ_{PZF} \{y\}$, and let $k_1(j), \ldots, k_{n_j}(j)$ and $o(j)$ be (not necessarily distinct) natural numbers between 1 and m. Assume that $A_1, \ldots, A_m$ are sets, and that $B_1, \ldots, B_m$ are the least $X_1, \ldots, X_m$ which satisfy the following conditions (for $1 \leq i \leq m$ and $1 \leq j \leq p$):*

(1) $A_i \subseteq X_i$

(2) If $a_1 \in X_{k_1(j)}, \ldots, a_{n_j} \in X_{k_{n_j}(j)}$ *and* $\varphi_j(b, a_1, \ldots, a_{n_j})$ *then* $b \in X_{o(j)}$

Then $B_1, \ldots, B_m$ are definable by terms of $\mathcal{L}_{PZF}$ with parameters $A_1, \ldots, A_m$.

Example: The set HF of hereditarily finite sets is the least X such that $\{\emptyset\} \subseteq X$, and $y \in X$ whenever $a \in X$, $b \in X$, and $y = a \cup \{b\}$. Hence HF is defined by a closed term of $\mathcal{L}_{PZF}$.

5 The Predicativity of PZF

The following theorem implies that PZF indeed satisfies condition (PRE):

Theorem 5.1.

1. *If* $\varphi \succ_{PZF} X$ *then* φ *is predicative (see Definition 2.1) in* PZF *for* X.
2. *If* t *is a valid term of* PZF *then* t *is predicative in the sense that it satisfies the following condition: If* $Fv(t) = \{y_1, \ldots, y_n\}$ *then the following is true (in* V*) for every transitive model* $\mathcal{M}$ *of* PZF*:*

$$\forall y_1 \ldots \forall y_n . y_1 \in \mathcal{M} \wedge \ldots \wedge y_n \in \mathcal{M} \rightarrow t_{\mathcal{M}} = t$$

where $t_{\mathcal{M}}$ *denotes the relativization of* t *to* $\mathcal{M}$.

Proof: By a simultaneous induction on the complexity of t and φ.
Discussion. By Theorem 5.1, every term t of $\mathcal{L}_{PZF}$ has the same interpretation in all transitive models of PZF which contains the values of its parameters. Thus the identity of the set denoted by t is independent of the exact extension of the assumed universe of sets. This already justifies seeing PZF as predicative. However, we want to argue that the predicativity of PZF intuitively goes deeper than this. The argument will necessarily be less exact (and on a more intuitive level) than that given by Theorem 5.1.
The problem with Theorem 5.1 is that it is a theorem of platonistic mathematics, and so it assumes an all-encompassing collection V which includes all potential “sets” and contains all “universes”, but is itself a universe too (meaning that classical logic holds within it). This assumption is doubtful from a predicativist point of view[13]. To see how we can do without it, call two universes $\mathcal{M}_1$ and $\mathcal{M}_2$ *strongly compatible* if the following conditions are satisfied:

1. Suppose a is an object in both $\mathcal{M}_1$ and $\mathcal{M}_2$. then

$$\{x \in \mathcal{M}_1 \mid \mathcal{M}_1 \vDash x \in a\} = \{x \in \mathcal{M}_2 \mid \mathcal{M}_2 \vDash x \in a\}$$

2. Suppose a and b are objects in $\mathcal{M}_1$ and $\mathcal{M}_2$ (respectively), and that the collections $\{x \in \mathcal{M}_1 \mid \mathcal{M}_1 \vDash x \in a\}$ and $\{x \in \mathcal{M}_2 \mid \mathcal{M}_2 \vDash x \in b\}$ are identical. Then a and b are identical. (Note that here we again use the notation $\{x \mid A\}$ *in the metalanguage* to denote classes of objects.)

It is now not difficult to check that if t is a term of $\mathcal{L}_{PZF}$ then the value of t for the assignment $x_1 := a_1, \ldots, x_n := a_n$ (where $Fv(t) = \{x_1, \ldots, x_n\}$) is the same in any two strongly compatible universes which include $\{a_1, \ldots, a_n\}$. This is what we really had in mind when we talked above about “universe independence” (note

[13] Thus both Sanchis in [42] and Weaver in [49] argue that classical logic is unsuitable for dealing with the whole of V, and intuitionistic logic should be used for it instead.

that if the platonic universe V exists, then every two transitive subcollections of V are compatible according to the definition above).

Turning next to Principle (NAT), we first of all note again that the set of all natural numbers is available in PZF in the form of ω. This easily implies that PA, the first-order Peano's Arithmetics, has a natural interpretation in PZF_0 (see Proposition 4.6 for a partial proof). However, the availability of ω alone is not sufficient for getting the full power of mathematical induction, since the full separation schema is not available in PZF. Nevertheless, the fact that the underlying logic is the TC-logic implies that the following induction *schema* is available (alternatively, this schema can be derived from the availability of ω with the help of $\in$-induction):

$$\vdash_{PZF_0} \varphi\{\varnothing/x\} \wedge \forall x(\varphi \rightarrow \varphi\{S(x)/x\}) \rightarrow \forall x.x \in \omega \rightarrow \varphi$$

No less crucial than the ability to use induction is the ability to use inductive definitions. Theorem 4.7 (see also Proposition 4.6) entails that the most important form of using such definitions is available in $\mathcal{L}_{PZF}$.

Note 5.2. Unlike in the case of proofs by induction (where $\in$-induction would do), the TC-machinery is essential for the ability to use in PZF inductive definitions. Now in previous systems for predicative mathematics, recursion in ω was obtained using Δ-comprehension (or Δ-collection). The explanation was that a Δ-formula φ is both upward absolute and downward absolute, and so it is absolute. This argument implicitly assumes the platonic universe V, and so it is doubtful in view of the discussion of (PRE) in this section (without V as a maximal universe, or some other doubtful assumptions concerning universes, I do not see why the combination of upward absoluteness and of downward absoluteness entails absoluteness).

6 Relations with the Axioms of ZF

The definability of $\{t, s\}$, $\bigcup t$, and ω means that the axioms of pairing, union, and infinity are provable in PZF. On the other hand $\{x \in t \mid \varphi\}$ is a valid term only if $\varphi \succ_{PZF} \varnothing$. Hence we do not have in PZF the full power of the other comprehension axioms of ZF. Instead we have the following counterparts:

The predicative separation schema: If $\varphi \succ_{PZF} \varnothing$; ψ is equivalent in PZF_0 to φ; x, w, Z are distinct variables and $Z \notin Fv(\psi)$, then:

$$\vdash_{PZF_0} \forall w \exists Z \forall x(x \in Z \leftrightarrow x \in w \wedge \psi)$$

The predicative replacement schema: If x, y, w, Z are distinct variables, and $Z, x \notin Fv(t)$ then

$$\vdash_{PZF_0} \forall w \exists Z \forall x(x \in Z \leftrightarrow \exists y.y \in w \wedge x = t)$$

The predicative collection schema: If $\varphi \succ_{PZF} \{x\}$; ψ is equivalent in PZF_0 to φ; x, y, w, Z are distinct variables, and $Z \notin Fv(\psi)$, then:

$$\vdash_{PZF_0} \forall w \exists Z \forall x(x \in Z \leftrightarrow \exists y.y \in w \wedge \psi)$$

The predicative powerset schema: If $\varphi \succ_{PZF} \{x\}$; ψ is equivalent in PZF_0 to φ; x, w, Z are distinct variables, and $Z \notin Fv(\psi)$, then:

$$\vdash_{PZF_0} \forall w \exists Z \forall x(x \in Z \leftrightarrow x \subseteq w \wedge \psi)$$

Thus although $P(\omega)$, the powerset of ω, is not available in PZF (This easily follows from Theorem 5.1, and the fact that $P(\omega)$ is not absolute), every set of the form $\{x \in P(\omega) \mid \varphi\}$ where $\varphi \succ_{PZF} \{x\}$ is available nevertheless.
At this point it is interesting to note that TZF, a system similar to PZF_0 which is intuitively sound (from a platonistic point of view), and does have the full power of ZF (though not ZFC), can be defined in a way similar to PZF_0, but using another relation $\succ_{TZF}$, instead of $\succ_{PZF}$. $\succ_{TZF}$ is the relation obtained by adding to the definition of $\mathcal{L}_{PZF}$ the following three conditions:

1. $\varphi \succ_{TZF} \emptyset$ for *every* formula φ.
2. $x \subseteq t \succ_{TZF} \{x\}$ if $x \notin Fv(t)$.
3. $\exists y\varphi \wedge \forall y(\varphi \to \psi) \succ_{TZF} X$ if $\psi \succ_{TZF} X$, and $X \cap Fv(\varphi) = \emptyset$.

In [4,5] it was shown that a first-order system which is equivalent to ZF (but more natural and easier to mechanize than the usual presentation of ZF) is obtained from TZF if the underlying logic is changed to classical first-order logic (in a first-order language enriched with abstract terms), and instead of using TC, a special constant for ω is added to the language, together with Peano's axioms for it. This shows that ZF and PZF are indeed close in spirit.

7 The Minimal Model of PZF

7.1 The Basic Universe

Next we show that in the spirit of (PRE), we may take our universe to be the collection of predicatively definable sets.

Definition 7.1. PD_0 (for "predicatively Definable") is the set (in V) of all sets (in V) which are defined by closed terms of $\mathcal{L}_{PZF}$.

Lemma 7.2. *Let s be a term of $\mathcal{L}_{PZF}$.*

1. *If s is free for y in the term t of $\mathcal{L}_{PZF}$, then $t\{s/y\}$ is a term of $\mathcal{L}_{PZF}$.*

2. *If s is free for y in the formula φ of $\mathcal{L}_{PZF}$, $\varphi \succ_{PZF} X$, $y \notin X$, and $Fv(s) \cap X = \emptyset$, then $\varphi\{s/y\} \succ_{PZF} X$.*

The proof is by a simultaneous induction on the complexity of t and φ.

Notation. 1. If t is a term of $\mathcal{L}_{PZF}$, and v is an assignment in V, we denote by $\|t\|_v$ the value (in V) that t gets under v. In case t is closed we denote by $\|t\|$ the value of t in V.

2. Let φ be a formula of $\mathcal{L}_{PZF}$, and let v be an assignment in V. $v \models \varphi$ denotes that v satisfies φ in V.

3. If φ is a formula of $\mathcal{L}_{PZF}$, $X \subseteq Fv(\varphi)$, and v is an assignment in V, we denote by $\|\varphi\|_v^X$ the class of all $a \in V$ for which there exists an assignment v' such that $a = v'(x)$ for some $x \in X$, $v'(y) = v(y)$ for $y \notin X$, and $v' \models \varphi$.

Lemma 7.3. *Let $Fv(t) = \{x_1, \ldots, x_n\}$, and let $s_1, \ldots, s_n$ be closed terms of $\mathcal{L}_{PZF}$. Suppose v is an assignment such that $v(x_i) = \|s_i\|$ for $i = 1, \ldots, n$. Then $\|t\|_v = \|t\{s_1/x_1, \ldots, s_n/x_n\}\|$.*

Theorem 7.4. *PD_0 is transitive (in other words: all elements of a predicatively definable set are themselves predicatively definable).*

Proof: Denote by HPD_0 (for "Hereditarily Predicatively Definable") the set of all sets $a \in V$ such that $TC(\{a\}) \subseteq PD_0$. Obviously, HPD_0 is a transitive subset of PD_0. Hence it suffices to show that $PD_0 \subseteq HPD_0$ (implying that $PD_0 = HPD_0$). For this we prove the following by a simultaneous induction on the complexity of t and φ:

1. $\|t\|_v \in HPD_0$ if t is a term of $\mathcal{L}_{PZF}$, and v is an assignment in HPD_0.

2. $\|\varphi\|_v^X \subseteq HPD_0$ in case $\varphi \succ_{PZF} X$, and v is an assignment in HPD_0 (Equivalently: if $\varphi \succ_{PZF} X$, $v \models \varphi$, and $v(x) \in HPD_0$ for $x \notin X$, then $v(x) \in HPD_0$ also for $x \in X$).

- The case where t is a variable is trivial.

- Suppose t is $\{x \mid \varphi\}$. Then $\|t\|_v \in PD_0$ by Lemma 7.3. Obviously $a \in \|t\|_v$ iff $a \in \|\varphi\|_v^{\{x\}}$. Hence $\|t\|_v \subseteq HPD_0$ by the I.H. for φ. It follows that $\|t\|_v \in HPD_0$.

- The cases where $\varphi \succ_{PZF} \varnothing$ and $X = \varnothing$, or φ is $x \in x$ and $X = \{x\}$ are trivial.

- Suppose φ is $x \in t$ where $x \notin Fv(t)$, and $X = \{x\}$. Then $\|\varphi\|_v^X = \|t\|_v$. Hence $\|\varphi\|_v^X \subseteq HPD_0$ by the I.H. concerning t and the transitivity of HPD_0.

- Suppose φ is $x = t$ (or $t = x$) where $x \notin Fv(t)$, and $X = \{x\}$. Then $\|\varphi\|_v^X = \{\|t\|_v\}$. Hence $\|\varphi\|_v^X \subseteq HPD_0$ by the I.H. concerning t.

- Suppose φ is $\varphi_1 \vee \varphi_2$, where $\varphi_1 \succ_{PZF} X$ and $\varphi_2 \succ_{PZF} X$. Then $\|\varphi\|_v^X = \|\varphi_1\|_v^X \cup \|\varphi_2\|_v^X$. Hence $\|\varphi\|_v^X \subseteq HPD_0$ by the I.H. concerning φ_1 and φ_2.

- Suppose φ is $\varphi_1 \wedge \varphi_2$, where $\varphi_1 \succ_{PZF} Y$, $\varphi_2 \succ_{PZF} Z$, $X = Y \cup Z$, and $Z \cap Fv(\varphi_1) = \varnothing$. To prove the claim for φ and X, it suffices to show that if $v' \vDash \varphi$, and $v'(w) \in HPD_0$ in case $w \notin X$, then $v'(x) \in HPD_0$ also for $x \in X$. So let v' be such an assignment. Then $v' \vDash \varphi_1$ and $v' \vDash \varphi_2$. Let v_1 be any assignment such that $v_1(x) = v'(x)$ for $x \notin Z$, and $v_1(x) \in HPD_0$ if $x \in Z$. Since $Z \cap Fv(\varphi_1) = \varnothing$, also $v_1 \vDash \varphi_1$. By the induction hypothesis concerning φ_1 and Y, this and the fact that $v_1(x) \in HPD_0$ in case $x \notin Y$ together imply that $v_1(x) \in HPD_0$ also in case $x \in Y$. It follows that $v'(x) \in HPD_0$ in case $x \in Y$, and that v_1 is an assignment in HPD_0. Now v' differs from v_1 only for variables in Z. This and the facts that $v' \vDash \varphi_2$ and $\varphi_2 \succ_{PZF} Z$, together entail that $v'(z) \in \|\varphi_2\|_{v_1}^Z$ for every $z \in Z$. Hence the I.H. for φ_2 implies that $v'(z) \in HPD_0$ in case $z \in Z$. Since we have already shown that $v'(y) \in HPD_0$ in case $y \in Y$, it follows that $v'(x) \in HPD_0$ for every $x \in X$.

- Suppose φ is $\exists z\psi$, where $\psi \succ_{PZF} X \cup \{z\}$. Then $\|\varphi\|_v^X \subseteq \|\psi\|_v^{X\cup\{z\}}$. Hence $\|\varphi\|_v^X \subseteq HPD_0$ by the I.H. concerning ψ.

- Suppose φ is $(TC_{x,y}\psi)(x,y)$, where $\psi \succ_{PZF} X \cup \{y\}$ (say). For $n \geq 0$, let φ_n be $\exists w_1 \ldots \exists w_n.\psi(x,w_1) \wedge \psi(w_1,w_2) \wedge \ldots \wedge \psi(w_{n-1},w_n) \wedge \psi(w_n,y)$ (where $w_1,\ldots,w_n$ are distinct variables not occurring in φ). Then $\|\varphi\|_v^X = \bigcup_{n\geq 0} \|\varphi_n\|_v^X$. Now it is easy to show by induction on n (using the I.H. for ψ and the cases concerning $\wedge$ and $\exists$ already dealt with above) that $\|\varphi_n\|_v^X$ is a subset of HPD_0 for every $n \geq 0$. Hence $\|\varphi\|_v^X \subseteq HPD_0$.

Let now $a \in PD_0$. Then there is a closed term t of $\mathcal{L}_{PZF}$ such that $a = \|t\|$. Hence $a \in HPD_0$ as a special case of (1), and so $a \subseteq PD_0$.

Definition 7.5. Let the language $\mathcal{L}_{PZF}^{\mathcal{M}}$ be defined like $\mathcal{L}_{PZF}$, but with the additional constant $\mathcal{M}$. For every term t and formula φ of $\mathcal{L}_{PZF}$ we define in $\mathcal{L}_{PZF}^{\mathcal{M}}$ the corresponding relativization $t_{\mathcal{M}}$ and $\varphi_{\mathcal{M}}$ (respectively):

- $x_{\mathcal{M}} = \{y \in \mathcal{M} \mid y \in x\}$.
- $\{x \mid \varphi\}_{\mathcal{M}} = \{x \mid x \in \mathcal{M} \wedge \varphi_{\mathcal{M}}\}$
- $(sRt)_{\mathcal{M}} = s_{\mathcal{M}} R t_{\mathcal{M}}$ for R in $\{\in, =\}$.
- $(\neg\varphi)_{\mathcal{M}} = \neg\varphi_{\mathcal{M}}$
- $(\varphi * \psi)_{\mathcal{M}} = \varphi_{\mathcal{M}} * \psi_{\mathcal{M}}$ for $*$ in $\{\vee, \wedge\}$.
- $(\exists x\varphi)_{\mathcal{M}} = \exists x.x \in \mathcal{M} \wedge \varphi_{\mathcal{M}}$.
- $((TC_{x,y}\varphi)(s,t))_{\mathcal{M}} = (TC_{x,y} x \in \mathcal{M} \wedge y \in \mathcal{M} \wedge \varphi_{\mathcal{M}})(s_{\mathcal{M}}, t_{\mathcal{M}})$.

Theorem 7.6. *Suppose the constant $\mathcal{M}$ is interpreted in V as PD_0.*

1. *If t is term of $\mathcal{L}_{PZF}$ and v is an assignment in PD_0 then $\|t_{\mathcal{M}}\|_v = \|t\|_v$.*
2. *Suppose that φ is a formula of $\mathcal{L}_{PZF}$ s. t. $Fv(\varphi) = \{y_1, \ldots, y_n, x_1, \ldots, x_k\}$, and $\varphi \succ_{PZF} \{x_1, \ldots, x_k\}$. Then the following is true in V:*

   $\forall y_1 \ldots \forall y_n . y_1 \in \mathcal{M} \wedge \ldots \wedge y_n \in \mathcal{M} \to [\varphi \leftrightarrow (x_1 \in \mathcal{M} \wedge \ldots \wedge x_k \in \mathcal{M} \wedge \varphi_{\mathcal{M}})]$

Proof: As usual, the proof is by a simultaneous induction on the complexity of t and φ.

- If t is a variable x then $\|t\|_v = \|t_{\mathcal{M}}\|_v$ follows from Theorem 7.4, because in this case $\|x_{\mathcal{M}}\|_v = \|x\|_v \cap PD_0$, and $\|x\|_v \in PD_0$.
- If t is $\{x \mid \varphi\}$ then the claim for t follows from the I.H. concerning φ.
- If φ is $s \in t$ or $s = t$ then the claim for φ immediately follows from the I.H. concerning t and s.
- If φ is $x \in t$, where $x \notin Fv(t)$, then the claim for φ follows from Lemma 7.3, Theorem 7.4, and the I.H. concerning t.
- If φ is $x = t$ or $t = x$, where $x \notin Fv(t)$, then the claim for φ follows from Lemma 7.3, and the I.H. concerning t.

The proofs of the other cases are similar to those given in the proof of the predicativity of $\mathcal{L}_{PZF}$, and are left for the reader.

Theorem 7.7. *PD_0 is a minimal model of PZF.*

Proof: That PD_0 is a model of PZF easily follows from Theorem 7.4 and Theorem 7.6. Minimality is obvious from the fact that every element in PD_0 is denoted by some closed term of $\mathcal{L}_{PZF}$ (and the absoluteness of the interpretations of these closed terms).

7.2 Ordinals in PD_0

Theorem 7.8. *If α is an ordinal and $\alpha < \omega^\omega$ then $\alpha \in PD_0$.*

Proof: We prove that for every $n \in N$ there exists a term t_n of PZF such that $Fv(t_n) = \{a\}$, and for every assignment v in V, if $v(a)$ is an ordinal, then $\|t_n\|_v = v(a) + \omega^n$. Obviously, t_0 is $S(a)$ (see subsection 4.1). Assume that t_n has been constructed, and let t_{n+1} be $\bigcup\{y \mid (TC_{a,y}y = t_n)(a,y)\}$. Given v, from the induction hypothesis concerning t_n it follows that $\|t_{n+1}\|_v$ is $\bigcup_{k\in N} v(a) + \omega^n k$. Hence $\|t_{n+1}\|_v = v(a) + \omega^{n+1}$.

Now let s_n be the closed term obtained from t_n by substituting 0 (i.e. $\emptyset$) for a. From what we have proved it follows that $\|s_n\| = \omega^n$. Hence $\omega^n \in PD_0$ for every $n \in N$. Since for every $\alpha < \omega^\omega$ there exists $n \in N$ such that $\alpha \in \omega^n$, the transitivity of PD_0 implies that $\alpha \in PD_0$ for every $\alpha < \omega^\omega$.

Theorem 7.9. *$\rho(a) < \omega^\omega$ for every $a \in PD_0$ (where $\rho(a)$ is the rank of a).*

Proof: We first show the following two facts:

1. For every term t of $\mathcal{L}_{PZF}$ there exists $n(t) \in N$ such that the following inequality obtains for every assignment v in V:

$$\rho(v(t)) < max\{\rho(v(y)) \mid y \in Fv(t)\} + \omega^{n(t)}$$

2. Let φ be a formula of $\mathcal{L}_{PZF}$ such that $Fv(\varphi) = X \uplus Y$, and $\varphi \succ_{PZF} X$. Then there exists $n(\varphi) \in N$ for which the following inequality obtains for every assignment v in V such that $v \models \varphi$:

$$max\{\rho(v(x)) \mid x \in X\} < max\{\rho(v(y)) \mid y \in Y\} + \omega^{n(\varphi)}$$

The proof is by a simultaneous induction on the complexity of t and φ:

- If t is a variable we take $n(t) = 0$.
- Suppose t is $\{x \mid \varphi\}$. By the induction hypothesis concerning φ, we can take $n(t) = n(\varphi) + 1$.
- The cases where $\varphi \succ_{PZF} \emptyset$ and $X = \emptyset$, or φ is $x \in x$ and $X = \{x\}$ are trivial.
- If φ is $x \in t$ or $x = t$ (and $X = \{x\}$) then we take $n(\varphi) = n(t)$.
- Suppose φ is $\varphi_1 \vee \varphi_2$, where $\varphi_1 \succ_{PZF} X$ and $\varphi_2 \succ_{PZF} X$. Take $n(\varphi) = max\{n(\varphi_1), n(\varphi_2)\}$.

- Suppose φ is $\varphi_1 \wedge \varphi_2$, where $\varphi_1 \succ_{PZF} X_1$, $\varphi_2 \succ_{PZF} X_2$, $X = X_1 \cup X_2$, and $X_2 \cap Fv(\varphi_1) = \varnothing$. By induction hypothesis for φ_1:

  $$max\{\rho(v(x)) \mid x \in X_1\} < max\{\rho(v(y)) \mid y \in Y\} + \omega^{n(\varphi_1)}$$

  While by induction hypothesis for φ_2:

  $$max\{\rho(v(x)) \mid x \in X_2\} < max\{\rho(v(y)) \mid y \in Y \cup X_1\} + \omega^{n(\varphi_2)}$$

  Together these two inequalities imply:

  $$max\{\rho(v(x)) \mid x \in X\} < max\{\rho(v(y)) \mid y \in Y\} + \omega^{n(\varphi_1)} + \omega^{n(\varphi_2)}$$

  It follows that we can take $n(\varphi) = max\{n(\varphi_1), n(\varphi_2)\} + 1$.

- Suppose φ is $\exists z\psi$, where $\psi \succ_{PZF} X \cup \{z\}$. Then obviously we can take $n(\varphi) = n(\psi)$.

- Suppose φ is $(TC_{z,y}\psi)(z, y)$, where $\psi \succ_{PZF} X \cup \{z\}$ (say, where possibly $z \in X$), and suppose $v \vDash \varphi$. Then for some $k \in N$:

  $$v \vDash \exists w_1 \dots \exists w_n.\psi(z, w_1) \wedge \varphi(w_1, w_2) \wedge \dots \wedge \varphi(w_{n-1}, w_n) \wedge \varphi(w_n, y)$$

  (where $w_1, \dots, w_n$ are distinct variables not occurring in φ). By induction hypothesis for ψ applied k times, this entails:

  $$max\{\rho(v(x)) \mid x \in X\} < max\{\rho(v(y)) \mid y \in Y\} + \omega^{n(\psi)} \cdot k$$

  It follows that we can take $n(\varphi) = n(\psi) + 1$.

This ends the proof of the two facts. Now in case t is a closed term of $\mathcal{L}_{PZF}$ fact (1) implies that $\rho(\|t\|) < \omega^\omega$. From this the theorem is immediate.

Corollary 7.10. $\omega^\omega \notin PD_0$.

Corollary 7.11. *ω^ω is the set of ordinals in PD_0.*

Corollary 7.12. *Ordinal addition (+) is not definable by a term of $\mathcal{L}_{PZF}$*

Proof: Had $+$ been definable, so would have been (using TC) multiplication by ω (since such a multiplication is equivalent to a repeated addition of the same ordinal). Again using TC, this would have made the set $\{\omega^n \mid n \in N\}$ definable, and so its union, ω^ω, would have been definable too, in contradiction to the previous corollary.

Theorem 7.13. *Suppose F is a monotonic set operation definable by some term of $\mathcal{L}_{PZF}$. Define a transfinite sequence of operations $F^{(\alpha)}$ by:*

- $F^{(0)}(a) = a$
- $F^{(\alpha+1)}(a) = F(F^{(\alpha)}(a))$
- $F^{(\alpha)}(a) = \bigcup_{\beta<\alpha} F^{(\beta)}(a)$ *in case α is a limit ordinal.*

Than for every $\alpha < \omega^\omega$, $F^{(\alpha)}$ is definable by some term of $\mathcal{L}_{PZF}$.

Proof: The following two facts can easily be shown:

1. $F^{(\alpha+\beta)} = F^{(\beta)} \circ F^{(\alpha)}$
2. $F^{(\alpha\cdot\beta)} = (F^{(\alpha)})^{(\beta)}$

Since every ordinal $\alpha < \omega^\omega$ can be obtained from 0, 1, and ω using addition and multiplication, it follows from these two facts that it suffices to prove that $F^{(\omega)}$ is definable whenever F is. So let t be a term of $\mathcal{L}_{PZF}$ such that $Fv(t) = \{a\}$, and t defines F. Then the term $a \cup \bigcup\{x \mid (TC_{a,x}x = t)(a, x)\}$ defines $F^{(\omega)}$.

Corollary 7.14. *If F is a monotonic set operation definable by some term of $\mathcal{L}_{PZF}$, and $a \in PD_0$, then $F^{(\alpha)}(a) \in PD_0$ for every $\alpha < \omega^\omega$.*

Note 7.15. Theorem 7.8 is a special case of Corollary 7.14 (take $F = S$).

Corollary 7.16. $J_\alpha \in PD_0$ *for every $\alpha < \omega^\omega$.*

Proof: $J_{\alpha+1}$ is obtained from J_α using a finitary inductive definition (it is the closure of J_α under the 9 operations listed in Lemma 1.11 of Chapter VI of [10]). Hence this monotonic operation is defined by a term of $\mathcal{L}_{PZF}$. The claim follows therefore from Corollary 7.14.

Theorem 7.17. $PD_0 = J_{\omega^\omega}$

Proof: From Corollary 7.16 it follows that $J_{\omega^\omega} \subseteq PD_0$.

For the converse, we first prove the following two facts:

1. For any term t of $\mathcal{L}_{PZF}$ there exists a natural number $n(t)$ and a term t^* of $\mathcal{L}_{RST}$ such that $Fv(t^*) \subseteq Fv(t) \cup \{w\}$ (where $w \notin Fv(t)$)), and the following holds for every ordinal α and valuation v: If $v(x) \in J_\alpha$ for every $x \in Fv(t)$, and $v(w) = J_\beta$ where $\beta \geq \alpha + \omega^{n(t)}$, then $\|t\|_v = \|t^*\|_v$.

2. Let $X = \{x_1, \ldots, x_n\}$. For any formula φ of $\mathcal{L}_{PZF}$ such that $\varphi \succ_{PZF} X$ and $w \notin Fv(\varphi)$, there exist a natural number $n(\varphi)$ and a formula φ^* of $\mathcal{L}_{RST}$ such that $Fv(\varphi^*) \subseteq Fv(\varphi) \cup \{w\}$, and for every ordinal α and valuation v, if $v(y) \in J_\alpha$ for every $y \in Fv(\varphi) - X$, and $v(w) = J_\beta$ where $\beta \geq \alpha + \omega^{n(\varphi)}$, then $\|\{\langle x_1, \ldots, x_n\rangle \mid \varphi\}\|_v = \|\{\langle x_1, \ldots, x_n\rangle \in J_\beta \mid \varphi^*\}\|_v$.

As usual, the proof of these two facts is by induction on the structure of t and φ, and is similar to the proof of Theorem 7.9. The only case which is not straightforward is when φ is $(TC_{y,x}\psi)(y, x)$, where $\psi \succ_{PZF} \{x\}$ (for simplicity, we suppress other variables). In this case $n(\varphi) = n(\psi) + 1$, and φ^* is:

$$\exists f \in w \exists n \in N.F(f) \wedge Dom(f) = n + 1 \wedge f(0) = y \wedge \\ f(n) = x \wedge \forall k < n.\psi^*(f(k), f(k + 1))$$

where $F(f)$ is the Δ_0 formula which says that F is a function.

Suppose now that $a \in PD_0$. Then $a = \|t\|$ for some closed term t of $\mathcal{L}_{PZF}$. By (1) it follows that $a = \|t^*\|_v$, where v is a valuation such that $v(w) = J_{\omega^{n(t)}}$. Since $J_{\omega^{n(t)}} \in J_{\omega^\omega}$, J_{ω^ω} is closed under rudimentary functions, and t^* is a term of $\mathcal{L}_{RST}$ (and so defines a rudimentary function by Corollary 4.4), $\|t^*\|_v \in J_{\omega^\omega}$. Hence $a \in J_{\omega^\omega}$. It follows that $PD_0 \subseteq J_{\omega^\omega}$.

8 Directions for Further Research

8.1 Strengthening PZF

PZF is a rich set theory, which is sufficient for the goals described in the introduction. Still, it is far from capturing the potential of predicative set theory. Thus although ω^n is definable in PZF for each n, and there is an effective procedure to derive a definition of ω^{n+1} from a a definition of ω^n, the set $\{\omega^n \mid n \in N\}$ and the function $\lambda n \in N.\omega^n$ are not definable in $\mathcal{L}_{PZF}$, even though their identity is clearly absolute and predicatively acceptable. There are at least five possible directions to remedy this by extending the definability power of PZF:

New Constants and Autonomous Progressions: A system $RST\omega$ where ω is definable can be obtained from RST by adding to $\mathcal{L}_{RST}$ a constant HF that denotes the set of sets which are defined by terms of RST, and by adding to RST appropriate closure axioms concerning HF. [14] Similarly, it is not difficult to show that by adding to $\mathcal{L}_{PZF}$ a constant denoting J_{ω^ω} with appropriate

[14] A similar analysis to that given above for PZF shows that $\omega \cdot 2$ is the set of ordinals which are definable by some closed term of $RST\omega$.

closure axioms, we get a system in which it is easy to construct closed terms for $\lambda n \in N.\omega^n$ and for ω^ω, and prove their main properties. Obviously this process can be repeated using transfinite recursion, creating by this a transfinite progression of languages and theories. To do so, we need first of all to precisely define the process of passing from a theory $\mathbf{T}_\alpha$ to $\mathbf{T}_{\alpha+1}$, and of constructing $\mathbf{T}_\alpha$ for limit α. Moreover, like in the systems for predicative analysis of Feferman and Schütte (see [15,44]), the progression should be autonomous, in the sense that only ordinals justified in previous systems may be used. Now instead of using indirect systems of (numerical) notations for ordinals, it would be much more natural to use terms of our systems which provably denote in them *von Neumann's ordinals*. We expect that every ordinal less than Γ_0, the Feferman-Schütte ordinal for predicativity ([15, 17, 44, 45]), can be obtained in this way.

Decoding: Although $\{\omega^n \mid n \in N\}$ and $\lambda n \in N.\omega^n$ are not definable in PZF, $\{\ulcorner\omega^n\urcorner \mid n \in N\}$ and $\lambda n \in N.\ulcorner\omega^n\urcorner$ *are* definable, where $\ulcorner\omega^n\urcorner$ is some natural Gödel code in HF for the term of $\mathcal{L}_{PZF}$ that defines ω^n. Now there should exist predicatively acceptable methods for passing from, say, $\{\ulcorner\omega^n\urcorner \mid n \in N\}$ to $\{\omega^n \mid n \in N\}$, and the language and proof system of PZF might be extended using these methods.

Dynamic Safety Relations: The safety relations we used in our 3 basic systems are all *static*, and are prior to the proof system. More power can be gained by allowing dynamic connections between safety and provability. Thus Δ-comprehension is equivalent to the following dynamic condition: $\exists y\varphi(y) \succ \emptyset$ in case $\varphi(y) \succ \emptyset, \psi(z) \succ \emptyset$, and $\vdash_{PZF} \exists y\varphi(y) \leftrightarrow \forall z\psi(z)$.

Inductive Definitions: The use of TC makes it possible to provide inductive definitions of relations and functions which are *sets*. In certain cases it also allows for defining global relations (using formulas of the language). However, its use is quite limited for inductively defining global operations. Take e.g. the ternary operation $G(n, k, a) = a+\omega^n\cdot(k+1)$ (where $n, k \in N$). G can be inductively defined as follows: $G(0, 0, a) = a \cup \{a\}$, $G(n + 1, 0, a) = \bigcup_{k\in N} G(n, k, a)$, $G(n+1, k+1, a) = G(n+1, 0, G(n+1, k, a))$. Intuitively, G should therefore be a predicatively acceptable operation. However, it is not definable in $\mathcal{L}_{PZF}$ by a term $t(n, k, a)$. Another possible direction for extending the power of $\mathcal{L}_{PZF}$ is therefore to allow stronger methods of inductive definitions over the natural numbers, as well as predicatively accepted transfinite recursion.

Introducing Classes Introducing global operations might be done by allowing terms for classes (of the form $[x : \varphi]$ where $\varphi \succ_{PZF} \emptyset$).

8.2 Other Directions

A necessary direction of research is to determine the relations of our framework and systems with previous works concerned with predicative set theory. This includes first of all Feferman's various systems for predicative mathematics, especially his system PS_1E for predicative set theory ([16, 18]), and his system W from [20]. Also relevant are the proof-theoretic investigations of systems of Kripke-Platek set theory by Jäger, Pohlers, and Rathjen (a partial list), as well as the works on constructive set theory by Aczel, Beeson, Friedman, Gambino, Rathjen, and many others. Another work that seems closely related is Weaver's recent work (see e.g. [49]) on predicative mathematics.

Beyond this, a major future project should be to produce concrete formal systems within the framework of PZF (based on valid, sufficiently strong formal systems for TC-logics), to determine their proof-theoretical strength, and to actually developed large portions of classical mathematics in them.

References

[1] Abiteboul, S., Hull, R., Vianu, V.: **Foundations of Databases**, Addison-Wesley, 1995.

[2] Avron A., *Transitive Closure and the mechanization of Mathematics*, In **Thirty Five Years of Automating Mathematics** (F. Kamareddine, ed.), 149-171, Kluwer Academic Publishers, 2003.

[3] Avron A., *Safety Signatures for First-order Languages and Their Applications*, In **First-Order Logic Revisited** (Hendricks et all,, eds.), 37-58, Logos Verlag Berlin, 2004.

[4] Avron A., *Formalizing Set Theory as It Is Actually Used*, In **Proceedings of Mathematical Knowledge Management (MKM 2004)** (A. Asperti, G. Banecerek, and A. Trybulec, eds.), 32-43, LNCS 3119, Springer, 2004.

[5] Avron A., *A Framework for Formalizing Set Theories Based on the Use of Static Set Terms*, In **Pillars of Computer Science**, (A. Avron, N. Dershowitz, and A. Rabinovich, eds.), 87–106, LNCS 4800, Springer, 2008.

[6] Avron A., *Constructibility and Decidability versus Domain Independence and Absoluteness*, Theoretical Computer Science 394 (2008), pp. 144–158 (http://dx.doi.org/10.1016/j.tcs.2007.12.008).

[7] Barwise J., **Admissible Sets and Structures**, Springer, 1975.

[8] Cantone D., Ferro A., and Omodeo E., **Computable Set Theory**, Clarendon Press, Oxford, 1989.

[9] Cantone D., Omodeo E., and Policriti A., **Set Theory for Computing: From Decisions Procedures to Declarative Programming with Sets**, Springer, 2001.

[10] K. J. Devlin, **Constructibility**, Perspectives in Mathematical Logic, Springer-Verlag, 1984.

[11] Y. L. Ershov, **Definability and Computability**, Siberian School of Algebra and Logic, Consultants Bureau, New-York, 1996.

[12] Ewald W., **From Kant to Hilbert**, Clarendon Press, London (1996).

[13] Ebbinghaus H. D., and Flum J., **Finite Model Theory** (2nd ed.), Perspectives in Mathematical Logic, Springer, 1999.

[14] A. Fraenkel, Y. Bar-Hillel, and A. Levy, **Foundations of Set Theory**, North-Holland, Amsterdam, 1973.

[15] Feferman S., *Systems of Predicative Analysis I*, Journal of Symbolic Logic 29 (1964), pp. 1-30.

[16] Feferman S., *Predicative Provability in Set Theory*, Bulletin of the American Mathematical Society 72 (1966), pp. 486-489.

[17] Feferman S., *Systems of Predicative Analysis II*, Journal of Symbolic Logic 29 (1968), pp. 193-220.

[18] Feferman S., *Predicatively Reducible Systems of Set Theory*, in **Axiomatic Set Theory** Proceedings of Symposia in Pure Mathematics, Vol. 13, Part 2, pp. 11-32, American Mathematical Society, Providence, 1974.

[19] Feferman S., *A More Perspicuous Formal System for Predicativity*, In **Konstruktionen versus Positionen, beitraǧe zur Diskussion um die Konstruktive Wissenschaftstheorie** (K. Lorenz, ed.), Walter de Gruyter, Berlin, 1978.

[20] Feferman S., *Weyl Vindicated: Das Kontinuum seventy years later*, In **Remi e prospettive della logica e della scienza contemporanee, vol. I.** (C. Celluci and G. Sambin, eds.), Cooperative Libraria Universitaria Editrice, Bologna (1988); Reprinted in [22].

[21] Feferman S., *Finitary Inductively Presented Logics*, in: **Logic Colloquium 1988** (1989), Amsterdam, North-Holland, pp. 191-220. Reprinted in [24], pp. 297-328.

[22] Feferman S., **In the Light of Logic**, Oxford University Press, Oxford (1998).

[23] Feferman S., and Hellman G., *Predicative Foundations of Arithmetics*, Journal Of Philosophical Logic 24 (1995), pp. 1-17.

[24] Gabbay D., editor, **What is a Logical System?** Oxford Science Publications, Clarendon Press, Oxford, 1994.

[25] Gandy, R. O., *Set-theoretic functions for elementary syntax*, In **Axiomatic set theory, Part 2**, AMS, Providence, Rhode Island, 103-126, 1974.

[26] Gentzen G., *Neue Fassung des Widerspruchsfreiheitsbeweises für die reine Zahlentheorie*, Forschungen zur Logik, N.S., No. 4, pp. 19-44 (English translation in: **The collected work of Gerhard Gentzen**, edited by M.E. Szabo, North-Holland, Amsterdam, (1969)).

[27] K. Gödel, **The Consistency of the Continuum Hypothesis**, Annals of Mathematical Studies, No. 3, Princeton University Press, Princeton, N.J., 1940.

[28] Grädel E., *On Transitive Closure Logic*, in: **Computer Science Logic (Bern 1991)**, Springer LNCS 626, 1992, pp. 149-163.

[29] Gurevich Y., *Logic and the Challenge of Computer Science*, in: Börger E., ed., **Trends in Theoretical Computer Science**, Computer Science Press Inc., Rockville, Maryland, USA (1988), pp. 1-58.

[30] Immerman, N., *Languages which Capture Complexity Classes*, in: 15th Symposium on Theory of Computing, Association for Computing Machinery (1983), pp. 347-354.

[31] Jäger G., **Theories for Admissible Sets: a Unifying Approach to Proof Theory**, Bibliopolis, Naples, 1986.

[32] R. B. Jensen, *The Fine Structure of the Constructible Hierarchy*, Annals of Mathematical Logic 4, pp. 229-308, 1972.

[33] Kunen K., **Set Theory: an Introduction to Independence Proofs**, North-Holland, Amsterdam, 1980.

[34] Myhill J., *A Derivation of Number Theory from Ancestral Theory*, Journal of Symbolic Logic 17, pp. 192-297, 1952.

[35] Poincaré H., *Les Mathématiques et la Logique, II, III*, Revue de Métaphysique et Morale 14 (1906), pp. 17-34, 294-317; translated in [12].

[36] Poincaré H., **Dernière Pensées**, Flammarion, Paris (1913); trans. by J. Bolduc as **Mathematics and Science: Last Essays**, Dover Press, New-York (1963).

[37] Ramsey F., *The foundations of mathematics*, Proceedings of the London Mathematical Society, 2nd series, 25(5), 1925.

[38] Richard, J., *Letter à Monnsieur le rédacteur de la* Revue général de Sciences. Acta Mathematica, 30, pp. 295-96. 1905

[39] Russelll B., *Les Paradoxes de la logique*, Revue de Métaphisique et de Morale, 14, (September 1906) Part of an exchange with Poincaré. English version entitled "On 'Insolubilia' and their Solution by Symbolic Logic" (chapter 9 of [41]).

[40] Russelll B., *Mathematical Logic as based on a theory of logical types*, Am. Journal of Mathematics 30, 1908.

[41] Russelll B. **Essays in Analysis**, edited by D. Lackey, Braziller New York, 1973.

[42] Sanchis L. E., **Set Theory — An Operational Approach**, Gordon and Breach Scientific Publishers, 1996.

[43] V. Y. Sazonov, *On Bounded Set Theory*, Proceedings of the 10th International Congress on Logic, Methodology and Philosophy of Sciences, Florence, August 1995, in Volume I: Logic and Scientific Method, Kluwer Academic Publishers, 85-103, 1997.

[44] Schütte K., *Predicative Well-ordering*, in **Formal Systems and Recursive Functions** (J. Crossley and M. Dummett, eds.), North-Holland, pp. 279-302.

[45] Schütte K., **Proof Theory**, Springer-Verlag, 1977.

[46] Shoenfield J. R., **Mathematical Logic**, Addison-Wesley, 1967.

[47] Shapiro S., **Foundations Without Foundationalism: A Case for Second-order Logic**, Oxford University Press, Oxford, 1991.

[48] Ullman, J.D.: **Principles of database and knowledge-base systems**, Computer Science Press, 1988.

[49] Weaver N., *Mathematical Conceptualism*, unpublished manuscript (available from http://www.math.wustle.edu/nweaver/concept.pdf).

[50] Weyl H., **Das Kontinuum: Kritische Untersuchungen über die Grundlagen der Analysis**, Veit, Leipzig (1918).

[51] Whitehead A., and Russell B., **Principia Mathematica**, vols. I, II, and III, Cambridge University Press, Cambridge (1910-13); 2nd edition, Cambridge University Press, Cambridge (1925-27).

Characterising Definable Search Problems in Bounded Arithmetic via Proof Notations

Arnold Beckmann* and Samuel R. Buss†

[1] Department of Computer Science
Swansea University
Swansea SA2 8PP, UK
a.beckmann@swansea.ac.uk
[2] Department of Mathematics
University of California, San Diego
La Jolla, CA 92093-0112, USA
sbuss@math.ucsd.edu

Abstract The complexity class of Π^{p}_k-Polynomial Local Search (PLS) problems with Π^{p}_ℓ-goal is introduced, and is used to give new characterisations of definable search problems in fragments of Bounded Arithmetic. The characterisations are established via notations for propositional proofs obtained by translating Bounded Arithmetic proofs using the Paris-Wilkie-translation. For $\ell \leq k$, the $\Sigma^{\mathrm{b}}_{\ell+1}$-definable search problems of T^{k+1}_2 are exactly characterised by Π^{p}_k-PLS problems with Π^{p}_ℓ-goals. These Π^{p}_k-PLS problems can be defined in a weak base theory such as S^1_2, and proved to be total in T^{k+1}_2. Furthermore, the Π^{p}_k-PLS definitions can be Skolemised with simple polynomial time functions. The Skolemised Π^{p}_k-PLS definitions give rise to a new $\forall\Sigma^{\mathrm{b}}_1(\alpha)$ principle conjectured to separate $\mathrm{T}^k_2(\alpha)$ from $\mathrm{T}^{k+1}_2(\alpha)$.

1 Introduction

Bounded Arithmetic in the form introduced by the second author [Bus86] denotes a collection of theories of arithmetic which have a strong connection to computational complexity. An important goal in Bounded Arithmetic is to give good descriptions of the functions that are definable in a certain theory by a certain class of formulas. For the sake of simplicity of this introduction, we will concentrate only on the Bounded Arithmetic theories S^i_2. These theories are given as first order

*Supported in part by EPSRC grant EP/D03809X/1.
†Supported in part by NSF grant DMS-0700533.

theories of arithmetic in a language which suitably extends that of Peano Arithmetic, where induction is restricted in two ways. First, logarithmic induction is considered which only inducts over a logarithmic part of the universe of discourse.

$$\varphi(0) \wedge (\forall x)(\varphi(x) \rightarrow \varphi(x+1)) \rightarrow (\forall x)\varphi(|x|) \ .$$

Here, $|x|$ denotes the length of the binary representation of the natural number x, which defines a kind of logarithm on natural numbers. As in these theories exponentiation will not be a total function, this is a proper restriction. Second, the properties which can be inducted on, must be described by a suitably restricted ("bounded") formula. The class of formulas used here are the Σ^{b}_i-formulas which exactly characterise Σ^{p}_i, that is, properties of the i-th level of the polynomial time hierarchy of predicates. The main axioms of the theory S^i_2 are the instances of logarithmic induction for Σ^{b}_i formulas.

Let a (multi-)function f be called Σ^{b}_j-definable in S^i_2, if its graph can be expressed by a Σ^{b}_j-formula φ, such that the totality of f, which renders as $(\forall x)(\exists y)\varphi(x,y)$, is provable from the S^i_2-axioms in first-order logic. The main results characterising definable (multi-)functions in Bounded Arithmetic are the following.

- Buss [Bus86] characterised the Σ^{b}_i-definable functions of S^i_2 as $\mathrm{FP}^{\Sigma^{\mathrm{p}}_{i-1}}$, the i-th level of the polynomial time hierarchy of functions.

- Krajíček [Kra93] characterised the $\Sigma^{\mathrm{b}}_{i+1}$-definable multi-functions of S^i_2 as the class $\mathrm{FP}^{\Sigma^{\mathrm{p}}_i}[wit, O(\log n)]$ of multi-functions which can be computed in polynomial time using a witness oracle from Σ^{p}_i, where the number of oracle queries is restricted to $O(\log n)$ many (n being the length of the input).

- Buss and Krajíček [BK94] characterised the Σ^{b}_1-definable multi-functions of S^2_2 as projections of solutions to polynomial local search problems. This result extends to higher levels as well: the $\Sigma^{\mathrm{b}}_{i-1}$-definable multi-functions of S^i_2 are exactly the projections of solutions to problems from $\mathrm{PLS}^{\Sigma^{\mathrm{p}}_{i-2}}$, which is the class of polynomial local search problems relativised to $\Sigma^{\mathrm{p}}_{i-2}$-oracles.

- Pollett [Pol99] showed that the $\Sigma^{\mathrm{b}}_{j+1}$-definable multi-functions in S^i_2 for $j > i$ are exactly $\mathrm{FP}^{\Sigma^{\mathrm{p}}_j}[wit, O(1)]$.

The characterisation of the Σ^{b}_i-definable functions of S^{k+1}_2 for $0 < i < k$ turned out to be more difficult, but recently some advances have been made. Krajíček, Skelley, and Thapen [KST07] characterised the Σ^{b}_1-definable functions of S^3_2 in terms of coloured PLS problems, and the Σ^{b}_1-definable functions of S^4_2 in terms of a kind of reflection principle, and also in terms of a kind of recursion called *verifiable recursion.* Subsequently, Skelley and Thapen [ST07] characterised the Σ_1^b-definable functions of S_2^{k+1}, for all $k \geq 2$, in terms of a combinatorial principle for k-turn games. An earlier, more complex, game characterisation of the same functions was given by Pudlák [Pud06] using a combinatorial analysis of Herbrand disjunctions, which has been improved later by the same author [Pud07].

In this article we will provide characterisations for all pairs of bounded formula class $\Sigma_{\ell+1}^b$ and theory S_2^{k+2}, for $\ell < k$, in terms of generalisations of PLS problems which we call Π_k^b*-PLS problems with* Π_ℓ^b*-goals.* We will define the new complexity classes in Section 3. An instance of a Π_k^b-PLS problems with Π_ℓ^b-goals will consist, on input a, of a polynomially bounded set of feasible solutions of complexity Π_k^b and a goal set of complexity Π_ℓ^b, an initial value function computing a feasible solution, a cost function computing the cost of a feasible solution, and a neighbourhood function computing from a given feasible solution another feasible solution (its neighbour), such that either the computed neighbour is identical to the original solution, or the neighbour is of lower cost — these functions have to be polynomial time computable. The goal set has to satisfy that it consists of exactly those feasible solutions for which the neighbourhood function is the identity. An important requirement will be that these conditions are provable in a weak theory like S_2^1, as without such requirements we can easily construct for any total function f given by $(\forall x)\varphi(x, f(x))$ with φ polynomial time computable and polynomially bounded (that is, for any (x, y) with $\varphi(x, y)$, $|y|$ is polynomial in $|x|$), a Π_1^b-PLS problem with goal φ — some of the requirements will then depend on the totality of the function and can thus only be proved in a theory which already proves the totality of the function.

The new characterisations have been obtained during a research visit of the first author at the second author's institution in autumn 2007. Prior to this visit, these new characterisations had been partially guessed based on recent results about obtaining the above-mentioned known characterisations of definable functions via notations for propositional proofs and cut-reduction [AB09]. Then, during the research visit, two different proofs for the new characterisations have been obtained, one extending the idea of notations for propositional proofs, and the other based on witnessing arguments.

Witnessing arguments form the dominant method for characterising definable (multi-)functions in Bounded Arithmetic. For example, the above-mentioned known characterisation of definable (multi-)functions in Bounded Arithmetic all have been proven by specially tailored witnessing arguments. The new characterisation based on witnessing arguments will be reported in a different place [BB08].

In this article, we present the new characterisations based on proof notations, that is, via notations for propositional proofs which are obtained by translating first order proofs and applying cut-reduction. We will compare this approach with the above-mentioned witnessing argument at the end of this introduction after we have given an idea of how the new characterisations based on proof notations work. First, we briefly describe the general idea of proof notations as presented in [AB09], which will also be one half of the idea for the new characterisations. Suppose $(\forall x)(\exists y)\varphi(x,y)$, describing the totality of some multi-function, is provable in some Bounded Arithmetic theory. Fix a particularly nice formal proof P of this. Given $a \in \mathbb{N}$ we want to describe a procedure which finds some b such that $\varphi(\underline{a},\underline{b})$ is true ($\underline{a}$ is some canonical term in the language of Bounded Arithmetic with value a.) Invert the proof P of $(\forall x)(\exists y)\varphi(x,y)$ to a proof of $(\exists y)\varphi(x,y)$, where x is now a free variable of the proof, then substitute $\underline{a}$ for all occurrences of x. This yields a proof of $(\exists y)\varphi(\underline{a},y)$. Now we want to translate this proof to propositional logic. The propositional translation used here is well-known in proof-theoretic investigations; the translation has been described by Tait [Tai68], and later was independently discovered by Paris and Wilkie [PW85]. In the Bounded Arithmetic world it is known as the *Paris-Wilkie translation.* As these translations in general produce exponential size formulas and proofs, we cannot directly work with the resulting objects, but have to use notations for them. Applying cut-reduction appropriately to notations of propositional proofs, we obtain a proof with all cut-formulas of (at most) the same logical complexity as φ. It should be noted that a notation $h(a)$ for this proof can be computed in time polynomial in $|a|$ (cf. [AB09].)

The general local search problem which finds a witness for $(\exists y)\varphi(\underline{a},y)$ can now be characterised as follows. Its instance is given by a. The set of possible solutions are those notations of a suitable size which denote derivations of a suitable cut-rank (cut-rank is the maximal level of cut-formulas occurring in the derivation). Furthermore, they must satisfy that the formula which they derive is equivalent to $(\exists y)\varphi(\underline{a},y) \lor \psi_1 \lor \cdots \lor \psi_l$, where all ψ_i are of low complexity and false. An initial solution is given by $h(a)$. A neighbour to a solution h is a solution which denotes an immediate subderivation of the derivation denoted by h, if this exists, and h otherwise. The cost of a notation is the height of the proof-tree represented by the notation. The search task is to find a notation in the set of solutions which is a fixed point of the neighbourhood function. Obviously, a solution to the search task must exist. In fact, any solution of minimal cost has this property. Now consider any solution to the search problem. It must have the property that none of the immediate subderivations is in the solution space. This can only happen if the last inference derives $(\exists y)\varphi(\underline{a},y)$ from a true statement $\varphi(\underline{a},\underline{b})$ for some $b \in \mathbb{N}$. Thus b is a witness to $(\exists y)\varphi(\underline{a},y)$, and we can output b as a solution to our original witnessing problem.

This approach works fine if the difference between the complexity of induction and the level of definability we are interested in is not too big. For the known characterisations mentioned above, things can be arranged such that, depending on the complexity of logarithmic induction present in the Bounded Arithmetic theory we started with, and the level of definability, we obtain local search problems *defined by functions of some level of the polynomial time hierarchy,* and different bounds to the cost function [AB09]. For example, if we start with the Σ^{b}_i-definable functions of S^i_2, we obtain a local search problem defined by properties in $\mathrm{FP}^{\Sigma^{\mathrm{b}}_{i-1}}$, where the cost function is bounded by $|a|^{O(1)}$. Thus, by following the canonical path through the search problem which starts at the initial value and iterates the neighbourhood function until reaching a solution, we obtain a path of polynomial length, which describes a procedure in $\mathrm{FP}^{\Sigma^{\mathrm{b}}_{i-1}}$ to compute a witness.

For the new characterisations however, the complexity of induction is much bigger than the level of definability, Σ^{b}_j versus Σ^{b}_i with $j >> i$ say. The above-described strategy would deal with this difference by applying an appropriate number of cut-reductions $(j+1-i)$. But if $j+1-i$ is too big, too many cut-reductions would have to be applied, resulting in a search space which explodes: the search space would contain too many objects as well as objects of too big size (iterated exponentiation in input length.) In such a situation the solution will be to apply a maximal number of cut-reductions such that the search space does not explode, and then change the above-described local search problem so that a feasible solution still contains a notation for derivation as above, but now the complexity of ψ_j does not necessarily match that of φ but can be bigger. This is compensated by accompanying the notation with an auxiliary search problem for determining the truth of ψ_j. In other words, a feasible solution in the overall search problem contains a notation h and a position $\mathfrak{s}$ in an auxiliary search problem for a formula ψ which is related to h. A solution to the auxiliary search problem for ψ will determine the truth of ψ, and allow us to choose an appropriate immediate subderivation of h to continue the overall search problem. Overall, we end up with search problems where the set of feasible solutions has high computational complexity (due to the assertion that all ψ_j are false) but, e.g., the neighbourhood function is still of low computational complexity (due to the use of the auxiliary search problems.) For example, we obtain for the Σ^{b}_1-definable multi-functions of S^{k+2}_2 that the set of feasible solutions has complexity Π^{p}_k, but the neighbourhood function, cost function and initial value function are polynomial time computable — this defines an instance of the above-mentioned Π^{b}_k-PLS problems.

An important property of our characterisation is that the Π^{b}_k-PLS conditions that the functions and predicates have to satisfy, are provable in S^1_2. Furthermore, these conditions can be written in a prenex form which can be Skolemised with simple polynomial time computable functions, such that the resulting conditions are still provable in S^1_2. This has several consequences: First, we obtain a much stronger algorithmic description of the $\Sigma^{\mathrm{b}}_{\ell+1}$-definable functions, as Π^{b}_k-PLS problems with Π^{b}_ℓ-goals in Skolem form, in which all conditions are given as $\forall\Pi^{\mathrm{b}}_1$ conditions. Second, using the description in Skolem form we can define search principles classes based on some generic principle (involving second order symbols representing the functions and predicates that make up a Π^{b}_k-PLS problem with Π^{b}_ℓ-goal in Skolem form) which can be seen to characterise the $\forall\Sigma^{\mathrm{b}}_{\ell+1}$-consequences of T^{k+1}_2. This, third, leads to the conjecture that the generic principle separates relativised theories, i.e. $\mathrm{T}^{k+1}_2(\alpha)$ from $\mathrm{T}^k_2(\alpha)$.

It is worth mentioning at this point that there are connections between the approach described here and the approach considered in [ST07]. The main similarity is that [ST07] also makes use of a translation of T^{k+1}_2 proofs into exponential sized propositional proofs of some special purpose propositional proof systems which are described by polynomial time relations.

To come back to a comparison between the proof notation approach to the new characterisations presented here, and the witnessing arguments given in [BB08], the difference between them goes beyond obtaining the same results with two different methods. The layout of the witnessing argument is such that its inductive formulation has to incorporate the cut-reduction part of the proof notation argument. This in particular means that the witnessing argument directly deals with sequents of formulas as complex as induction formulas, where the notation argument directly deals with sequents of formulas of one level below that. So, on one hand the witnessing argument is direct but more involved, whereas for the notation argument it takes a while to set up the necessary machinery (mainly by repeating parts of [AB09]), but after that is pretty straightforward. Both approaches have in common that they use auxiliary search problems to determine the truth of formulas. The difference between the two approaches becomes even more visible when it comes to refinements of the results by Skolemising properties of the resulting search problems. While this is a technical but straightforward task for proof notations, it is more involved for the witnessing argument which needs to prove a stronger Skolemisation result due to its inductive layout and higher formula complexities.

The next section will briefly introduce Bounded Arithmetic in a way suitable for our proof-theoretic investigations. Section 3 defines the search problem classes of Π^{b}_k-polynomial local search, and their generalisations. Sections 4 and 5 review necessary definitions and results on notations and cut-reduction in general, and for Bounded Arithmetic in particular, from [AB09]. Section 6 then introduces the auxiliary search problems to determine the truth of formulas. This is followed by the section defining the search problems which come from proofs in Bounded

Arithmetic, and stating our main result concerning the characterisation of definable multi-functions in terms of Π^{b}_k-PLS. The next two sections deal with a strengthening of our main results by showing that the conditions for Π^{b}_k-PLS problems extracted from Bounded Arithmetic proofs can be Skolemised with simple, polynomial time computable Skolem-functions. In the final section we will use the Skolemised Π^{b}_k-PLS problems to define $\forall\Sigma^{\mathrm{b}}_1(\alpha)$-sentences which we conjecture to separate relativised Bounded Arithmetic theories $\mathrm{S}^{k+1}_2(\alpha)$ from $\mathrm{S}^{k+2}_2(\alpha)$.

2 Bounded Arithmetic

We introduce Bounded Arithmetic very briefly, and in a slightly nonstandard way which better suits our proof-theoretic investigations. The reader interested in the general theory of Bounded Arithmetic is kindly referred to the literature [Bus86].

The standard model for Bounded Arithmetic is $\mathbb{N}$, the set of natural numbers. For $a \in \mathbb{N}$ let $|a|$ denote the length of the binary representation of a.

Definition 2.1 (Language of Bounded Arithmetic)**.** We define the *language* $\mathcal{L}_{\mathrm{BA}}$ *of Bounded Arithmetic* as in [Bus86] with a few additional symbols for polynomial time computable functions:

$$\mathcal{L}_{\mathrm{BA}} = \{\mathrm{S}, +, \times, |\cdot|, \#, =, \leq\} \cup \{c_a \colon a{\in}\mathbb{N}\} \cup \{2^{|\cdot|}, \dot{-}, \min, \mathrm{pair}, (\cdot)_1, (\cdot)_2\}$$

To explain the meaning of these symbols we briefly indicate their interpretation in the standard model $\mathbb{N}$: $\{=, \leq\}$ denote the binary relations "equality" and "less than or equal". c_a for $a{\in}\mathbb{N}$ denotes a constant with standard interpretation $c_a^{\mathbb{N}} = a$. We will often write $\underline{a}$ instead of c_a, and 0 for c_0. S, $|\cdot|$ and $2^{|\cdot|}$ are unary function symbols whose standard interpretations are given by the successor function, $|\cdot|^{\mathbb{N}} \colon a \mapsto |a|$, and $2^{|\cdot|^{\mathbb{N}}} \colon a \mapsto 2^{|a|}$. $+$, $\times$, $\dot{-}$, min and $\#$ are binary function symbols whose standard interpretation are addition, multiplication, $\dot{-}^{\mathbb{N}} \colon a, b \mapsto \max(a{-}b, 0)$, minimisation, and $\#^{\mathbb{N}} \colon a, b \mapsto 2^{|a|\cdot|b|}$. pair, $(\cdot)_1$, $(\cdot)_2$ define some feasible pairing function like the Cantor pairing function with corresponding projections.

Atomic formulas are of the form $s = t$ or $s \leq t$ where s and t are terms. *Literals* are expressions of the form A or $\neg A$ where A is an atomic formula. Formulas are built up from literals by means of $\wedge$, $\vee$, $(\forall x)$, $(\exists x)$. The *negation* $\neg C$ *for a formula* C is defined via de Morgan's laws. Negation extends to sets of formulas in the usual way by applying it to their members individually. $A \rightarrow B$ is an abbreviation of $\neg A \vee B$.

Let $\mathrm{FV}(A)$ denote the free variables occurring in formula A. With $A_x(t)$ we denote the result of replacing all free occurrences of the variable x in A by t. Similar definitions are used for substitution into terms.

Definition 2.2 ($\overline{\mathrm{BASIC}}$)**.** With a *valid disjunction of literals* we mean a disjunction A of literals such that A is true in $\mathbb{N}$ under any assignment. Let $\overline{\mathrm{BASIC}}$ denote a set of valid disjunctions of literals which is sufficient to define the non-logical symbols in $\mathcal{L}_{\mathrm{BA}}$. More precisely, we consider the set $\overline{\mathrm{BASIC}}$ to be the natural reformulation of the axioms BASIC from [Bus86] into a set of disjunctions of literals, extended by suitable axioms defining the new symbols. We assume that the following axioms are included:

$$(\mathrm{pair}(a,b))_1 = a \qquad\qquad (\mathrm{pair}(a,b))_2 = b$$
$$(c)_1 \le c \qquad\qquad (c)_2 \le c$$
$$a, b \le t \;\rightarrow\; \mathrm{pair}(a,b) \le B(t) \text{ for some } \mathcal{L}_{\mathrm{BA}}\text{-term } B$$
$$\min(a,b) = a \;\vee\; \min(a,b) = b \qquad\qquad \min(a,b) = \min(b,a)$$
$$a \le b \;\rightarrow\; \min(a,b) = a \qquad\qquad \min(a,b) = a \;\rightarrow\; a \le b$$
$$a \dot{-} a = 0 \qquad\qquad (\mathrm{S}\, a) \dot{-} (\mathrm{S}\, b) = a \dot{-} b$$
$$a \le b \;\rightarrow\; a \dot{-} b = 0 \qquad\qquad a \dot{-} b = 0 \;\rightarrow\; a \le b$$

Definition 2.3 (Bounded Quantification)**.** Bounded quantifiers are introduced as follows: $(\forall x \le t)A$ denotes $(\forall x)A_x(\min(x,t))$, $(\exists x \le t)A$ denotes $(\exists x)A_x(\min(x,t))$, where x may not occur in t.

Our introduction of bounded quantifiers is a bit nonstandard. It has the advantage that already the usual cut-reduction procedure gives optimal results. The more standard abbreviation of bounded quantification, where e.g. $(\exists x \le t)A$ denotes $(\exists x)(x \le t \;\wedge\; A)$, would need a modification of cut-reduction to produce optimal bounds, as two logical connectives are to be removed for one bounded quantifier. Nevertheless, the two kind of abbreviations are equivalent over a weak base theory like Buss' BASIC (c.f. [Bus86]) assuming that this base theory includes some standard axiomatisation of $\min$ using $\le$ like $a \le b \;\rightarrow\; \min(x,y) = x$ and $\min(a,b) = \min(b,a)$. Also, either way makes use of a nonlogical symbol ("$\le$" versus "min").

Another approach to formalise bounded quantifiers is followed in [Bus86], where bounded quantifiers are treated as new logical symbols, not as abbreviations, and have their own, new kind of inference rules.

Definition 2.4 (Bounded Formulas)**.** The set Δ_0 of *bounded* $\mathcal{L}_{\mathrm{BA}}$*-formulas* is the set of $\mathcal{L}_{\mathrm{BA}}$-formulas consisting of literals and being closed under $\wedge$, $\vee$, $(\forall x \le t)$, $(\exists x \le t)$.

We now define a delineation of bounded formulas. The literature sometimes distinguishes between "strict" or "prenex" versions versus more liberal ones. We do not want to make such a distinction here to keep the focus on our proof-theoretic investigations, and define the classes only in their restricted form.

Definition 2.5. The set $\mathrm{s}\Sigma_i^{\mathrm{b}}$ is the smallest subset of bounded $\mathcal{L}_{\mathrm{BA}}$-formulas that is closed under taking subformulas and that contains all formulas of the form

$$(\exists x_1 \leq t_1)(\forall x_2 \leq t_2) \cdots (Qx_i \leq t_i)(\overline{Q}x_{i+1} \leq |t_{i+1}|)A(\vec{x}) \ ,$$

with Q and $\overline{Q}$ being of the corresponding alternating quantifier shape and A being quantifier free. A and the t_i's may involve variables not mentioned here.

Let $\mathrm{s}\Pi_i^{\mathrm{b}}$ be the set $\{\neg\varphi \colon \varphi \in \mathrm{s}\Sigma_i^{\mathrm{b}}\}$, and let $\mathrm{s}\Sigma_\infty^{\mathrm{b}}$ be $\bigcup_{d<\infty} \mathrm{s}\Sigma_d^{\mathrm{b}}$.

Definition 2.6 (Rank)**.** The *rank of a formula* φ, $\mathrm{rk}(\varphi)$, is defined as the minimal k such that $\varphi \in \mathrm{s}\Sigma_k^{\mathrm{b}} \cup \mathrm{s}\Pi_k^{\mathrm{b}}$, if such a k exists, and ∞ otherwise.

Definition 2.7. Let $\mathrm{Ind}(A, z, t)$ denote the expression

$$A_z(0) \ \wedge \ (\forall z \leq t)(A \ \rightarrow \ A_z(z+1)) \ \rightarrow \ A_z(t) \ .$$

We will base our definition of Bounded Arithmetic theories on a different normal form of induction than usually considered in the literature.

Definition 2.8. Let T_2^i denote the theory consisting of (universal closures of) formulas in $\overline{\mathrm{BASIC}}$ and of (universal closures of) formulas of the form $\mathrm{Ind}(A, z, 2^{|t|})$ with $A \in \mathrm{s}\Sigma_i^{\mathrm{b}}$, z a variable, and t an $\mathcal{L}_{\mathrm{BA}}$-term.

Let S_2^1 denote the theory consisting (of universal closures) of formulas in $\overline{\mathrm{BASIC}}$ and (of universal closures) of formulas of the form $\mathrm{Ind}(A, z, |t|)$ with $A \in \mathrm{s}\Sigma_1^{\mathrm{b}}$, z a variable and t an $\mathcal{L}_{\mathrm{BA}}$-term.

Our versions of T_2^i and S_2^1 are different from the standard versions as for example defined in [Bus86]. They are adapted to suit the proof-theoretic investigations we want to pursue. Nevertheless, they are equivalent in that the sets of consequences are the same. This follows from the fact that the restricted form of induction as defined in Definition 2.7 implies the usual form, because the following can be proven from $\overline{\mathrm{BASIC}}$ alone:

$$\mathrm{Ind}(A(\min(t, z)), z, 2^{|t|}) \ \rightarrow \ \mathrm{Ind}(A(z), z, t) \ .$$

Definition 2.9. Let Σ_i^{b} (Π_i^{b}) be the set of formulas φ such that there exist $\psi \in \mathrm{s}\Sigma_i^{\mathrm{b}}$ (resp. $\psi \in \mathrm{s}\Pi_i^{\mathrm{b}}$) with $\mathrm{S}_2^1 \vdash \varphi \ \leftrightarrow \ \psi$.

Let Δ_1^{b} be the set of formulas φ such that there exist formulas $\sigma \in \mathrm{s}\Sigma_1^{\mathrm{b}}$ and $\pi \in \mathrm{s}\Pi_1^{\mathrm{b}}$ with $\mathrm{S}_2^1 \vdash (\varphi \ \leftrightarrow \ \sigma) \ \wedge \ (\varphi \ \leftrightarrow \ \pi)$.

3 Π^{p}_{k}-Polynomial Local Search

A binary relation $R \subseteq \mathbb{N} \times \mathbb{N}$ is called *polynomially bounded* iff there is a polynomial p such that $(x, y) \in R$ implies $|y| \leq p(|x|)$. R is called *total* if for all x there exists a y with $(x, y) \in R$.

Definition 3.1 (Total and Definable Search Problems)**.** Let $R \subseteq \mathbb{N}\times\mathbb{N}$ be a polynomially bounded, total relation. The *(total) search problem* associated with R is this: Given input $x \in \mathbb{N}$, return a $y \in \mathbb{N}$ such that $(x, y) \in R$. R is called $\Sigma^{\mathrm{b}}_{\ell+1}$*-definable in* T^{k+1}_{2} if there exists a Π^{b}_{ℓ}-formula $\varphi(x, y)$ (Δ^{b}_{1} if $\ell = 0$) and an $\mathcal{L}_{\mathrm{BA}}$-term $t(x)$, both with all free variables shown, such that $(x, y) \in R$ iff $\mathbb{N} \vDash \varphi(x, y)$, and such that $\mathrm{T}^{k+1}_{2} \vdash (\forall x)(\exists y \leq t(x))\varphi(x, y)$.

Definition 3.2 (Π^{p}_{k}-PLS Problems with Π^{p}_{ℓ}-Goal)**.** A Π^{p}_{k}*-Polynomial Local Search (PLS) problem with* Π^{p}_{ℓ}*-goal*, for $k \geq \ell \geq 0$, is a tuple $L = (F, G, N, c, i)$ consisting of, for a given input x, a set $F(x)$ of *feasible solutions* with a polynomial bound d, a *goal set* $G(x)$, a *neighbourhood function* $N(x, s)$ mapping a configuration s to another configuration, a function $c(x, s)$ computing the *cost of a configuration* s, and a function $i(x)$ computing an *initial feasible solution*, such that the following properties are satisfied: the functions N, c and i are polynomial time computable, $F \in \Pi^{\mathrm{p}}_{k}$ and $G \in \Pi^{\mathrm{p}}_{\ell}$, and the following five conditions are satisfied:

$$(\forall x, s)(s \in F(x) \;\rightarrow\; |s| \leq d(|x|)) \tag{3.1}$$

$$(\forall x)(i(x) \in F(x)) \tag{3.2}$$

$$(\forall x, s)(s \in F(x) \;\rightarrow\; N(x, s) \in F(x)) \tag{3.3}$$

$$(\forall x, s)(N(x, s) = s \;\vee\; c(x, N(x, s)) < c(x, s)) \tag{3.4}$$

$$(\forall x, s)(s \in G(x) \;\leftrightarrow\; (N(x, s) = s \;\wedge\; s \in F(x))) \tag{3.5}$$

The search task is, for a given input x, to find some s with $s \in G(x)$.

Usually, the polynomial bound to F, d, is thought to be understood from the context and not explicitely mentioned. If we want to make it explicit we sometimes write $L = (d, F, G, N, c, i)$. We have introduced F and G as sets. When we focus on their complexity or their definability in Bounded Arithmetic, we treat "$s \in F(a)$" etc. as relations in s, a.

Without any further requirements, Π^{p}_{k}-PLS problems with Π^{p}_{ℓ}-goals do not say much about the complexity of the underlying search task. For example, let R be a polynomial time computable, total relation with polynomial bound p, defining a total search problem. Then we can define a Π^{p}_{1}-PLS problem with goal R as follows: Let $T(x)$ be $2^{p(|x|)}$. A feasible solution $s{\in}F(x)$ is given if $s{<}T(x) \;\wedge$

$R(x,s)$, or, in case $s{=}T(x){+}s'$, if $|s'|{\le}p(|x|) \ \wedge \ (\forall y{<}s')(x,y){\notin}R$; the initial value is $T(x)$; the neighbourhood function takes an s and outputs s if $s < T(x)$, or, in case $s{=}T(x){+}s'$, produces $T(x){+}s'{+}1$ if $|s'{+}1|{\le}p(|x|) \ \wedge \ (x,s'){\notin}R$, and s' otherwise; and the cost of an s is computed as $2T(x){\dot-}s$ for $s \ge T(x)$, and 0 otherwise. The problem with this definition is that its condition (3.3) cannot be proven only if one can already prove that R defines a *total* search problem.

To formulate a Π^{p}_k-PLS local search principle so as to guarantee the totality of a search problem without actually presupposing it, we have to ensure that the conditions (3.1)–(3.5) have "simple" proofs. We make this precise in the next definition by requiring that they are provable in S^1_2.

Definition 3.3 (Formalised Π^{p}_k-PLS Problems with Π^{p}_ℓ-Goals)**.** A Π^{p}_k-PLS problem with Π^{p}_ℓ-goal is *formalised in* S^1_2 provided the functions N, c, and i are Σ^{b}_1-definable in S^1_2, the predicate F is given by a Π^{b}_k-formula, the predicate G is given by a Π^{b}_ℓ-formula (Δ^{b}_1 if $\ell = 0$), and the defining conditions (3.1)–(3.5) are provable in S^1_2.

A Π^{p}_k-PLS problem with Π^{p}_ℓ-goal which is formalised in S^1_2 will be called a Π^{b}_k-*PLS problem with* Π^{b}_ℓ*-goal* (with superscript "b" instead of "p".)

The direction "$\leftarrow$" in condition (3.5) of a Π^{b}_k-PLS problem with Π^{b}_ℓ-goal is inessential, dropping it would result in an equivalent class of search problems. To make this more precise, let us denote with Π^{b}_k*-PLS' problems with* Π^{b}_ℓ*-goals* the class of search problems which are defined similar to Π^{b}_k-PLS problems with Π^{b}_ℓ-goals, with the only difference that in (3.5) equivalence "$\leftrightarrow$" is replaced by implication "$\rightarrow$". To see that Π^{b}_k-PLS' problems with Π^{b}_ℓ-goals are equivalent to Π^{b}_k-PLS problems with Π^{b}_ℓ-goals, first observe that any Π^{b}_k-PLS problem with Π^{b}_ℓ-goal is also a Π^{b}_k-PLS' problem with Π^{b}_ℓ-goal. Secondly, we can transform any Π^{b}_k-PLS' problem with Π^{b}_ℓ-goal $L' = (d', F', G', N', c', i')$ into a Π^{b}_k-PLS problem with Π^{b}_ℓ-goal $L = (d, F, G, N, c, i)$ which solves L', in the following way: Let $T(x)$ be $2^{d'(|x|)}$. We set d to $2d'$, and $i(x)$ as $T(x){+}i'(x)$. Let $s{\in}F(x)$ if either $s{<}T(x) \ \wedge \ s{\in}G(x)$, or, in case $s{=}T(x){+}s'$, if $s'{\in}F'(x)$. Set $N(x,s)$ to be s if $s{<}T(x)$, or, in case $s{=}T(x){+}s'$, to be $T(x){+}N'(x,s')$ if $N'(x,s'){\ne}s'$, and s' otherwise. Finally, define $c(x,s)$ to be 0 if $s{<}T(x)$, and $1{+}c'(x,s')$ in case $s{=}T(x){+}s'$.

Theorem 3.4. *Let* $k \ge \ell \ge 0$*. The* Π^{b}_k*-PLS problems with* Π^{b}_ℓ*-goals are* $\Sigma^{\mathrm{b}}_{\ell+1}$*-definable search problems in* T^{k+1}_2.

Proof. Let $L = (F, G, N, c, i)$ be a Π^{b}_k-PLS problem with Π^{b}_ℓ-goal. Let x be given. The set $A := \{c(x,s) \colon s \in F(x)\}$ is non-empty by (3.2), and can be expressed by a $\Sigma^{\mathrm{b}}_{k+1}$ formula. T^{k+1}_2 proves minimisation for $\Sigma^{\mathrm{b}}_{k+1}$-formulas, thus, arguing in

T^{k+1}_2, we can choose some minimal $c \in A$. Pick $s \in F(x)$ with $c(x,s) = c$, and let $s' := N(x,s)$. Then $s' \in F(x)$ by (3.3). By construction $c(x,s') \geq c(x,s)$, hence (3.4) shows $s' = N(x,s) = s$. Hence, (3.5) shows $s \in G(x)$.

That $\{(x,s)\colon s \in G(x)\}$ can be described by some Π^{b}_ℓ formula is clear by definition. □

The converse of the last theorem is also true and forms one of our main results in this article. It will be proven in Section 7.2.

Theorem 3.5. *Let $0 \leq \ell \leq k$. The $\Sigma^{\mathrm{b}}_{\ell+1}$-definable total search problems in T^{k+1}_2 can be characterised by Π^{b}_k-PLS problems with Π^{b}_ℓ-goals. This characterisation satisfies in addition that the goal formula is syntactically identical to the Π^{b}_ℓ-subformula of the original $\Sigma^{\mathrm{b}}_{\ell+1}$-formula.*

3.1 Search Problem Classes

Definition 3.2 gives rise to search principles expressed by one formula $\mathrm{PiPLS}(d,F,G,N,c,i)$ in second order parameters d,F,G,N,c,i, which is defined as

$$(3.1) \ \wedge \ (3.2) \ \wedge \ (3.3) \ \wedge \ (3.4) \ \wedge \ (3.5) \ \rightarrow \ (\forall x)(\exists y)G(x,y) \ .$$

By choosing appropriate substitutions for the parameters, this generic formula can be used to define syntactic search problem classes which characterise the $\forall\Sigma^{\mathrm{b}}_{\ell+1}$-consequences of T^{k+1}_2: Let $\mathrm{PiPLS}(k,\ell)$ be the set of all formulas obtained by replacing in $\mathrm{PiPLS}(d,F,G,N,c,i)$, d by some polynomial, N,c,i by polynomial time computable functions (represented by their Σ^{b}_1-definition in S^1_2), F by some formula in Π^{b}_k, and G by some formula in Π^{b}_ℓ. The proof of Theorem 3.4 shows that each formula in $\mathrm{PiPLS}(k,l)$ is provable in T^{k+1}_2. A converse is also true and can be shown using Theorem 3.5.

Corollary 3.6. *Over S^1_2, the theories $\mathrm{PiPLS}(k,l)$ and T^{k+1}_2 have the same $\forall\Sigma^{\mathrm{b}}_{\ell+1}$-consequences.*

Proof. We already argued for one inclusion. We still have to show that if T^{k+1}_2 proves $(\forall x)\varphi(x)$ with $\varphi \in \Sigma^{\mathrm{b}}_{\ell+1}$, then this formula also follows from a formula in $\mathrm{PiPLS}(k,l)$ over S^1_2.

Applying Theorem 3.5 we obtain a formalised Π^{p}_k-PLS problem with goal formula identical to φ. Consider the formula $\mathrm{PiPLS}(d,F,G,N,c,i)$ in $\mathrm{PiPLS}(k,l)$ coming from this characterisation. The conditions (3.1)–(3.5) are now provable in S^1_2, so over S^1_2 we immediately obtain $(\forall x)\varphi(x)$ from $\mathrm{PiPLS}(d,F,G,N,c,i)$. □

In Sections 8 and 9, we will see that a strengthening of Theorem 3.5 can also be proven, in which the conditions (3.1)–(3.5) will be transformed into some canonical Skolem form, see Corollary 9.8. This will reduce the complexity of the search principle class to match the complexity of the goal formulas. In particular we will obtain a set of $\forall\Sigma^{\mathrm{b}}_1$ formulas characterising the $\forall\Sigma^{\mathrm{b}}_1$-consequences of the theories T^{k+1}_2, for $k \geq 0$, see Corollary 10.2.

4 Notation Systems for Formulas and Proofs

In this section we review notation systems for propositional formulas and proofs, and cut-reduction for them from [AB09]. They provide the basic machinery for dealing with search problems based on proof notations.

4.1 Proof Systems

We begin with an abstract definition of proof systems, which will be at the heart of several derivation systems considered later.

Definition 4.1 (Notation System for Formulas)**.** A *notation system for formulas* is a triple $\langle \mathcal{F}, \approx, \mathrm{rk}\rangle$ where $\mathcal{F}$ is a set (of *formulas*), $\approx$ an equivalence relation on $\mathcal{F}$ (*identity between formulas*), and $\mathrm{rk}\colon \mathfrak{P}(\mathcal{F}) \times \mathcal{F} \to \mathbb{N}$ a function (*rank*). Here, $\mathfrak{P}(\mathcal{F})$ denotes the power set of $\mathcal{F}$.

We always write $\mathcal{C}$-$\mathrm{rk}(A)$ instead of $\mathrm{rk}(\mathcal{C}, A)$. With $\approx\Gamma$ we denote the set $\{G\colon (\exists F \in \Gamma)(G \approx F)\}$.

Definition 4.2. A *proof system* $\mathfrak{S}$ *over* $\langle \mathcal{F}, \approx, \mathrm{rk}\rangle$ is given by a set of formal expressions called *inference symbols* (syntactic variable $\mathcal{I}$), and for each inference symbol $\mathcal{I}$ an ordinal $|\mathcal{I}| \leq \omega$, a sequent $\Delta(\mathcal{I})$ and a family of sequents $(\Delta_\iota(\mathcal{I}))_{\iota<|\mathcal{I}|}$.

Proof systems may have inference symbols of the form Cut_C for $C \in \mathcal{F}$; these are called "cut inference symbols" and their use will (in Definition 4.4) be measured by the $\mathcal{C}$-cut-rank.

Notation 4.3. By writing $(\mathcal{I})\ \dfrac{\ldots\Delta_\iota\ldots(\iota < I)}{\Delta}$ we declare $\mathcal{I}$ as an inference symbol with $|\mathcal{I}| = I$ many hypotheses, with conclusion $\Delta(\mathcal{I}) = \Delta$, and ι-th hypothesis $\Delta_\iota(\mathcal{I}) = \Delta_\iota$ for $\iota < I$. If $|\mathcal{I}| = n$ we write $\dfrac{\Delta_0\ \Delta_1\ \ldots\ \Delta_{n-1}}{\Delta}$ instead of $\dfrac{\ldots\Delta_\iota\ldots(\iota < I)}{\Delta}$.

$\mathfrak{S}$-quasi derivations, to be defined next, are (infinite) terms built up from inference symbols. An $\mathfrak{S}$-quasi derivation will always have the form of an inference symbol $\mathcal{I}$, followed by "(", followed by a sequence of length $|\mathcal{I}|$ of $\mathfrak{S}$-quasi derivations, followed by ")". For example, the simplest $\mathfrak{S}$-quasi derivations are given as $\mathcal{I}()$ in case $\mathcal{I}$ is an inference symbol with $|\mathcal{I}| = 0$. We will write a sequence of the form $(d_0, \dots, d_{n-1})$ as $(d_\iota)_{\iota<n}$.

Definition 4.4 (Inductive definition of $\mathfrak{S}$-quasi derivations)**.** If $\mathcal{I}$ is an inference symbol of $\mathfrak{S}$, and $(d_\iota)_{\iota<|\mathcal{I}|}$ is a sequence of $\mathfrak{S}$-quasi derivations, then $d := \mathcal{I}(d_\iota)_{\iota<|\mathcal{I}|}$ is an $\mathfrak{S}$-*quasi derivation* with *end-sequent*

$$\Gamma(d) := \Delta(\mathcal{I}) \cup \bigcup_{\iota<|\mathcal{I}|} (\Gamma(d_\iota) \setminus \approx\Delta_\iota(\mathcal{I})) ,$$

last inference $\mathrm{last}(d) := \mathcal{I}$, *subderivations* $d(\iota) := d_\iota$ for $\iota < |\mathcal{I}|$, *height*

$$\mathrm{hgt}(d) := \sup \{\mathrm{hgt}(d_\iota) + 1 \colon \iota < |\mathcal{I}|\} ,$$

size (provided $\mathfrak{S}$ has inference symbols of finite arity only)

$$\mathrm{sz}(d) := (\sum_{\iota<|\mathcal{I}|} \mathrm{sz}(d_\iota)) + 1 ,$$

and *cut-rank*

$$\mathcal{C}\text{-crk}(d) := \sup(\{\mathcal{C}\text{-rk}(\mathcal{I})\} \cup \{\mathcal{C}\text{-crk}(d_\iota) \colon \iota < |\mathcal{I}|\}) .$$

Here we define $\mathcal{C}$-rk$(\mathcal{I})$, the *cut-rank of* $\mathcal{I}$, to be $\mathcal{C}$-rk$(C) + 1$ if $\mathcal{I}$ is of the form $\mathcal{I} = \mathrm{Cut}_C$ with $C \notin \mathcal{C}$, and to be 0 otherwise.

Definition 4.5. $d \vdash_{\approx} \Gamma$ is defined as $\Gamma(d) \subseteq \approx\Gamma$.

A translation of first order proofs into propositional ones, like the Paris-Wilkie translation, usually comes in two steps: First, first order formulas are translated into propositional ones; Second, first order proofs are translated into propositional proofs. In the next subsection, we introduce notation systems for propositional formulas of the type obtained by the Paris-Wilkie translation of first order formulas. The subsequent section defines our propositional proof system. Then, Subsection 4.4 describes polynomial-size notations for exponential-size propositional proofs that are obtained by the translation of first order proofs.

4.2 Notations for Propositional Formulas

Translating first order formulas into propositional ones via the Paris-Wilkie translation $^{\mathrm{PW}}$ transforms a bounded quantifier of the form $(\forall x \leq t(a))\varphi(x)$ into the propositional formula $\bigwedge_{i \leq t(a)^{\mathbb{N}}} \varphi(i)^{\mathrm{PW}}$. The length of this propositional formula is exponential in $|a|$, thus we need notation systems for such propositional formulas which allow us to deal with them in a feasible way. The next definition collects all necessary ingredients and properties of notation systems for propositional formulas.

Definition 4.6. We define $\neg$ as a function on the symbols $\{\top, \bot, \bigwedge, \bigvee\}$ in the following way: $\neg(\top) = \bot$, $\neg(\bot) = \top$, $\neg(\bigwedge) = \bigvee$, and $\neg(\bigvee) = \bigwedge$.

Definition 4.7. A *notation system* $\langle \mathcal{F}, \mathrm{tp}, \cdot[\cdot], \neg, \mathrm{rk}, \approx\rangle$ *for (infinitary) propositional formulas* is a notation system $\langle \mathcal{F}, \approx, \mathrm{rk}\rangle$ for formulas together with functions $\mathrm{tp}\colon \mathcal{F} \to \{\top, \bot, \bigwedge, \bigvee\}$, $\cdot[\cdot]\colon \mathcal{F} \times \mathbb{N} \to \mathcal{F}$, and $\neg\colon \mathcal{F} \to \mathcal{F}$, called *outermost connective*, *subformula*, and *negation*, respectively, such that $\mathrm{tp}(\neg(f)) = \neg(\mathrm{tp}(f))$, $\neg(f)[n] = \neg(f[n])$, $\mathcal{C}\text{-rk}(f) = \mathcal{C}\text{-rk}(\neg f)$, $\mathcal{C}\text{-rk}(f[n]) < \mathcal{C}\text{-rk}(f)$ for $f \notin \mathcal{C}$ and $n < |\,\mathrm{tp}(f)|$, and $f \approx g$ implies $\mathrm{tp}(f) = \mathrm{tp}(g)$, $f[n] \approx g[n]$, $\neg(f) \approx \neg(g)$ and $\mathcal{C}\text{-rk}(f) = \mathcal{C}\text{-rk}(g)$.

In the previous definition, the obvious idea behind $f[n]$ for $f \in \mathcal{F}$ and $n \in \mathbb{N}$ is that it denotes the n-th subformula of f. But observe that the situation we are describing is a bit more general. It does not exclude non-wellfounded notation systems, which may contain a notation f for which $0 < |\,\mathrm{tp}(f)|$ continues to hold for $f[0]$, $f[0][0]$, etc. ad infinitum. The cut-elimination results summarised in the following are still valid also in such a situation.

4.3 Propositional Proofs

The propositional proof system we are concerned with is directly based on notation systems for propositional formulas. There is of course a propositional proof system for (usual) propositional formulas in the background which is obtained by unfolding notations for propositional formulas into (usual) propositional formulas. This background proof system is not necessary for our technical developments, therefore we omit it. The interested reader will find a more detailed discussion in [AB09].

Definition 4.8. Let $\mathcal{F} = \langle \mathcal{F}, \mathrm{tp}, \cdot[\cdot], \neg, \mathrm{rk}, \approx\rangle$ be a notation system for propositional formulas. The *(propositional) proof system* $\mathfrak{S}_{\mathcal{F}}$ *over* $\mathcal{F}$ is the proof system over $\mathcal{F}$ which is given by the following set of inference symbols.

(Ax_A) $\quad \overline{\;A\;}$ for $A \in \mathcal{F}$ with $\mathrm{tp}(A) = \top$

$(\bigwedge_C)$ $\quad \dfrac{\dots \quad C[n] \quad \dots \quad (n \in \mathbb{N})}{C}$ for $C \in \mathcal{F}$ with $\mathrm{tp}(C) = \bigwedge$

$(\bigvee_C^i)$ $\quad \dfrac{C[i]}{C}$ for $C \in \mathcal{F}$ with $\mathrm{tp}(C) = \bigvee$ and $i \in \mathbb{N}$

(Cut_C) $\quad \dfrac{C \qquad \neg C}{\varnothing}$ for $C \in \mathcal{F}$ with $\mathrm{tp}(C) \in \{\top, \bigwedge\}$

(Rep) $\quad \dfrac{\varnothing}{\varnothing}$

Abbreviations
For $\mathrm{tp}(C) \in \{\bot, \bigvee\}$ let $(\mathrm{Cut}_C)\ \dfrac{C \qquad \neg C}{\varnothing}$ denote $(\mathrm{Cut}_{\neg C})\ \dfrac{\neg C \qquad C}{\varnothing}$

4.4 Notations for Propositional Proofs and Cut-Elimination

The translation of first order proofs in Bounded Arithmetic into the propositional proof system defined in Definition 4.8 may generate proofs of exponential size. E.g., an application of $(\forall)\ \dfrac{\varphi(\min(x, t(a)))}{(\forall x \le t(a))\varphi(x)}$ is translated into $(\bigwedge)\ \dfrac{\varphi(0)^{\mathrm{PW}} \quad \dots \quad \varphi(t(a)^{\mathbb{N}})^{\mathrm{PW}}}{(\forall x \le t(a))\varphi(x)^{\mathrm{PW}}}$ which may have exponentially in $|a|$ many premises. Thus, besides notations for propositional formulas, we also need notations for propositional proofs obtained by translation in order to be able to deal with them in a feasible way. The necessary ingredients for this are collected in the next definition.

Definition 4.9. Let $\mathcal{F}$ be a notation system for formulas, and $\mathfrak{S}_{\mathcal{F}}$ the propositional proof system over $\mathcal{F}$ from Definition 4.8.

A *notation system* $\mathcal{H} = \langle \mathcal{H}, \mathrm{tp}, \cdot[\cdot], \Gamma, \mathrm{crk}, \mathrm{o}, |\cdot| \rangle$ *for* $\mathfrak{S}_{\mathcal{F}}$ is a set $\mathcal{H}$ of *notations* and functions $\mathrm{tp}\colon \mathcal{H} \to \mathfrak{S}_{\mathcal{F}}$, $\cdot[\cdot]\colon \mathcal{H} \times \mathbb{N} \to \mathcal{H}$, $\Gamma\colon \mathcal{H} \to \mathfrak{P}_{\mathrm{fin}}(\mathcal{F})$, $\mathrm{crk}\colon \mathfrak{P}(\mathcal{F}) \times \mathcal{H} \to \mathbb{N}$, and $\mathrm{o}, |\cdot|\colon \mathcal{H} \to \mathbb{N} \setminus \{0\}$ called *denoted last inference*, *denoted subderivation*, *denoted end-sequent*, *denoted cut-rank*, *denoted height* and *size*, such that $\mathcal{C}$-crk$(h[n]) \le \mathcal{C}$-crk(h), $\mathrm{tp}(h) = \mathrm{Cut}_C$ implies $\mathcal{C}$-rk$(C) < \mathcal{C}$-crk(h) for $C \notin \mathcal{C}$, $\mathrm{o}(h[n]) < \mathrm{o}(h)$ for $n < |\mathrm{tp}(h)|$, and the following local faithfulness property holds for $h \in \mathcal{H}$:

$$\Delta(\mathrm{tp}(h)) \subseteq {\approx}\Gamma(h) \quad \text{and} \quad \forall \iota < |\mathrm{tp}(h)|\ h[\iota] \vdash_{\approx} \Gamma(h), \Delta_\iota(\mathrm{tp}(h)) \ .$$

We observe that the size function in the last definition is not denoted. The idea is that it measures the size of the notation, not of the denoted proof. The size function will be important later when we try to measure the effect which cut-elimination has on notations, to identify those cases where the effect is feasible, i.e. does not lead to an exponential blow-up typical for cut-elimination on (regular) proofs.

The next definition gives the canonical propositional translation of proof notations into propositional proofs. The observation following this definition states the connection between key structural functions for notations and for connected propositional derivations.

Definition 4.10. Let $\mathcal{H} = \langle \mathcal{H}, \mathrm{tp}, \cdot[\cdot], \Gamma, \mathrm{crk}, \mathrm{o}, |\cdot| \rangle$ be a notation system for $\mathfrak{S}_{\mathcal{F}}$. The *interpretation* $[\![h]\!]$ *of* $h \in \mathcal{H}$ is inductively defined as the following $\mathfrak{S}_{\mathcal{F}}$-derivation:

$$[\![h]\!] := \mathrm{tp}(h)([\![h[\iota]]\!])_{\iota < |\,\mathrm{tp}(h)|}$$

Observation 4.11. *We make use of the functions defined in Definition 4.4. For* $h \in \mathcal{H}$ *we have*

$$\begin{aligned} \mathrm{last}([\![h]\!]) &= \mathrm{tp}(h) \\ [\![h]\!](\iota) &= [\![h[\iota]]\!] \quad \textit{for } \iota < |\,\mathrm{tp}(h)| \\ \Gamma([\![h]\!]) &\subseteq {\approx}\Gamma(h) \end{aligned}$$

We explained in the introduction of this paper that our characterisation of definable search problems in Bounded Arithmetic will be based on translating Bounded Arithmetic proofs into propositional ones, and applying cut-reduction to the resulting propositional proofs. Thus, we also have to add to our notation system for propositional logic some notations for cut-reduction on propositional proofs. This can be done very uniformly, as presented in the next definition. Our approach following [AB09] is based on Mints' continuous cut-elimination procedure [Min78] in its technical smooth presentation by Buchholz [Buc91, Buc97] and utilises notations for certain operators of propositional proofs. Readers interested in a fuller account of this situation are kindly referred to [AB09]. The intuition behind the notations for operators for cut-reduction are as follows:

- The symbol I^k_C denotes an *inversion operator* which satisfies: If $h \vdash_{\approx} \Gamma, C$ and $\mathrm{tp}(C) = \bigwedge$ then $\mathsf{I}^k_C h \vdash_{\approx} \Gamma, C[k]$, $\mathcal{C}$-crk$(\mathsf{I}^k_C h) \leq \mathcal{C}$-crk$(h)$ and $\mathrm{o}(\mathsf{I}^k_C h) \leq \mathrm{o}(h)$.
- The symbol R_C denotes a *one-cut-reduction operator* which satisfies: If $h_0 \vdash_{\approx} \Gamma, C$, $h_1 \vdash_{\approx} \Gamma, \neg C$ and $\mathrm{tp}(C) \in \{\top, \bigwedge\}$, then $\mathsf{R}_C h_0 h_1 \vdash_{\approx} \Gamma$, $\mathcal{C}$-crk$(\mathsf{R}_C h_0 h_1) \leq \max\{\mathcal{C}\text{-crk}(h_0), \mathcal{C}\text{-crk}(h_1), \mathcal{C}\text{-rk}(C)\}$ and $\mathrm{o}(\mathsf{R}_C h_0 h_1) \leq \mathrm{o}(h_0) + \mathrm{o}(h_1)$.
- The symbol E denotes a *highest-cut-elimination operator* which satisfies: If $h \vdash_{\approx} \Gamma$ then $\mathsf{E}h \vdash_{\approx} \Gamma$ and $\mathcal{C}$-crk$(\mathsf{E}h) \leq \mathcal{C}$-crk$(h) \dot{-} 1$ and $\mathrm{o}(\mathsf{E}h) < 2^{\mathrm{o}(h)}$.

Definition 4.12. The *notation system* $\mathcal{CH}$ *for cut-elimination on* $\mathcal{H}$ is given by the set of terms $\mathcal{CH}$ which are inductively defined by

- $\mathcal{H} \subset \mathcal{CH}$,
- $h \in \mathcal{CH}$, $C \in \mathcal{F}$ with $\mathrm{tp}(C) = \bigwedge$, $k < \omega \quad \Rightarrow \quad \mathsf{I}^k_C h \in \mathcal{CH}$,
- $h_0, h_1 \in \mathcal{CH}$, $C \in \mathcal{F}$ with $\mathrm{tp}(C) \in \{\top, \bigwedge\} \quad \Rightarrow \quad \mathsf{R}_C h_0 h_1 \in \mathcal{CH}$,
- $h \in \mathcal{CH} \quad \Rightarrow \quad \mathsf{E}h \in \mathcal{CH}$,

where $\mathsf{I}, \mathsf{R}, \mathsf{E}$ are new symbols, and functions $\mathrm{tp}\colon \mathcal{CH} \to \mathfrak{S}_{\mathcal{F}}$, $\cdot[\cdot]\colon \mathcal{CH} \times \mathbb{N} \to \mathcal{CH}$, $\Gamma\colon \mathcal{CH} \to \mathfrak{P}_{\mathsf{fin}}(\mathcal{F})$, $\mathrm{crk}\colon \mathfrak{P}(\mathcal{F}) \times \mathcal{CH} \to \mathbb{N}$, $\mathrm{o}\colon \mathcal{CH} \to \mathbb{N} \setminus \{0\}$ and $|\cdot|\colon \mathcal{CH} \to \mathbb{N}$ defined by recursion on the complexity of $h \in \mathcal{CH}$:

- If $h \in \mathcal{H}$ then all functions are inherited from $\mathcal{H}$.
- $h = \mathsf{I}^k_C h_0$: Let $\Gamma(h) := \{C[k]\} \cup (\Gamma(h_0) \setminus \approx\{C\})$, $\mathcal{C}$-crk$(h) := \mathcal{C}$-crk(h_0), $\mathrm{o}(h) := \mathrm{o}(h_0)$, and $|h| := |h_0| + 1$.

  Case 1. $\mathrm{tp}(h_0) \in \{\bigwedge_D \colon D \approx C\}$. Then let $\mathrm{tp}(h) := \mathrm{Rep}$, and $h[0] := \mathsf{I}^k_C h_0[k]$.

  Case 2. Otherwise, let $\mathrm{tp}(h) := \mathrm{tp}(h_0)$, and $h[i] := \mathsf{I}^k_C h_0[i]$.
- $h = \mathsf{R}_C h_0 h_1$: Let $\mathcal{I} := \mathrm{tp}(h_1)$. We define $\Gamma(h) := (\Gamma(h_0) \setminus \approx\{C\}) \cup (\Gamma(h_1) \setminus \approx\{\neg C\})$, $\mathcal{C}$-crk$(h) := \max\{\mathcal{C}\text{-crk}(h_0), \mathcal{C}\text{-crk}(h_1), \mathcal{C}\text{-rk}(C)\}$, $\mathrm{o}(h) := \mathrm{o}(h_0) + \mathrm{o}(h_1)$, and $|h| := |h_0| + |h_1| + 1$. For $\mathrm{tp}(h)$ and $h[i]$ we consider the following two cases:

  Case 1. $\Delta(\mathcal{I}) \cap \approx\{\neg C\} = \varnothing$: Then let $\mathrm{tp}(h) := \mathcal{I}$, and $h[i] := \mathsf{R}_C h_0 h_1[i]$.

  Case 2. Otherwise, $\Delta(\mathcal{I}) \cap \approx\{\neg C\} \neq \varnothing$. Since $\mathrm{tp}(C) \in \{\top, \bigwedge\}$ and no inference symbol $\mathcal{I}'$ of $\mathfrak{S}_{\mathcal{F}}$ has $D \in \Delta(\mathcal{I}')$ with $\mathrm{tp}(D) = \bot$, we must have $\mathrm{tp}(C) = \bigwedge$. Thus $\mathcal{I} = \bigvee^k_D$ for some $k \in \mathbb{N}$ and $D \approx \neg C$. Then let $\mathrm{tp}(h) := \mathrm{Cut}_{C[k]}$ and $h[0] := \mathsf{I}^k_C h_0$, $h[1] := \mathsf{R}_C h_0 h_1[0]$.
- $h = \mathsf{E}h_0$: Let $\Gamma(h) := \Gamma(h_0)$, $\mathcal{C}$-crk$(h) := \mathcal{C}$-crk$(h_0) \dot{-} 1$, $\mathrm{o}(h) := 2^{\mathrm{o}(h_0)} - 1$, and $|h| := |h_0| + 1$.

  Case 1. $\mathrm{tp}(h_0) = \mathrm{Cut}_C$: Then let $\mathrm{tp}(h) := \mathrm{Rep}$ and
  let $h[0] := \mathsf{R}_C \mathsf{E}h_0[0]\mathsf{E}h_0[1]$ if $\mathrm{tp}(C) \in \{\top, \bigwedge\}$,
  let $h[0] := \mathsf{R}_{\neg C} \mathsf{E}h_0[1]\mathsf{E}h_0[0]$ if $\mathrm{tp}(C) \notin \{\top, \bigwedge\}$.

  Case 2. Otherwise, let $\mathrm{tp}(h) := \mathrm{tp}(h_0)$, and $h[i] := \mathsf{E}h_0[i]$.

It has been shown in [AB09] that the notation system for cut-elimination on $\mathcal{H}$ is a notation system in the sense of Definition 4.9.

4.5 Size Bounds of Notations for Cut-Elimination

Notation systems for propositional formulas and proofs will, as we will see later, be feasible in situations related to definable search problems of Bounded Arithmetic. We will now analyse the feasibility of notations for cut-reduction on propositional proofs, by studying the size of notations for cut-reduction. We will just state the necessary definitions and results, more details including full proofs can be found in [AB09].

Definition 4.13. $\mathcal{H}$ is called *bounded* if $|h[i]| \leq |h|$ for all $h \in \mathcal{H}$, $i < |\operatorname{tp}(h)|$.

Definition 4.14. We define a "size function" $\vartheta\colon \mathbb{N} \to \mathbb{N}$ by induction on the inductive definition of $\mathcal{CH}$ as follows.

- For $h \in \mathcal{H}$ we set $\vartheta(h) = |h|$.
- $\vartheta(\mathsf{I}^k_C h) = \vartheta(h) + 1$
- $\vartheta(\mathsf{R}_C h_0 h_1) = \max\{|h_0|+1+\vartheta(h_1)\ ,\ \vartheta(h_0)+1\}$
- $\vartheta(\mathsf{E}h) = o(h)(\vartheta(h) + 2)$

Proposition 4.15. *If $\mathcal{H}$ is bounded then for every $h \in \mathcal{CH}$ we have $|h| \leq \vartheta(h)$.*

Theorem 4.16. *If $\mathcal{H}$ is bounded, $h \in \mathcal{CH}$ and $i < |\operatorname{tp}(h)|$, then $\vartheta(h) \geq \vartheta(h[i])$.*

Definition 4.14, Proposition 4.15 and Theorem 4.16 together show that cut-reduction can behave feasibly on proof notations. E.g., assume that we have a proof notation $h(a)$ depending on some parameter a — such a notation may originate from a first order proof of a universal statement $(\forall x)\varphi(x)$, where we inverted the outermost universal quantifier and substituted the constant $\underline{a}$ for the new free variable x, thus considering a proof of $\varphi(\underline{a})$ for $a \in \mathbb{N}$ — such that $o(h(a))$ and $|h(a)|$ are polynomial in $|a|$. Applying cut-reduction once to $h(a)$ gives a propositional proof in which all subproofs can be denoted by a notation of size polynomial in $|a|$: Consider a subproof h' of $\mathsf{E}h(a)$ which is given by the path $i_1, \dots, i_k$, i.e. $h' = \mathsf{E}h(a)[i_1] \cdots [i_k]$. By Proposition 4.15, $|h'| \leq \vartheta(h')$, and by Theorem 4.16, $\vartheta(h') \leq \vartheta(\mathsf{E}h(a))$. By Definition 4.14, the latter can be computed to

$$\vartheta(\mathsf{E}h(a)) = o(h(a)) \cdot (\vartheta(h(a))+2) = o(h(a)) \cdot (|h(a)|+2)$$

which is polynomial in $|a|$.

In the next section we will define concrete notation systems for propositional formulas and proofs based on translating Bounded Arithmetic according to the Paris-Wilkie translation. Together with the results from this section they provide the concrete machinery for characterising definable search problems via proof notations.

5 Notations based on Bounded Arithmetic

We start by defining a notation system for propositional formulas obtained by translating the language of Bounded Arithmetic according to the Paris-Wilkie translation, as given in [AB09].

Let $\mathcal{F}_{\mathrm{BA}}$ be the set of closed formulas in Δ_0. We define the outermost connective function on $\mathcal{F}_{\mathrm{BA}}$ by

$$\mathrm{tp}(A) := \begin{cases} \top & A \text{ true literal} \\ \bot & A \text{ false literal} \\ \bigwedge & A \text{ is of the form } A_0 \ \wedge \ A_1 \text{ or } (\forall x)B \\ \bigvee & A \text{ is of the form } A_0 \ \vee \ A_1 \text{ or } (\exists x)B \ , \end{cases}$$

and the subformula function on $\mathcal{F}_{\mathrm{BA}} \times \mathbb{N}$ by

$$A[n] := \begin{cases} A & A \text{ literal} \\ A_{\min(n,1)} & A \text{ is of the form } A_0 \ \wedge \ A_1 \text{ or } A_0 \ \vee \ A_1 \\ B_x(\underline{n}) & A \text{ is of the form } (\forall x)B \text{ or } (\exists x)B \ . \end{cases}$$

To define a suitable rank function on $\mathcal{F}_{\mathrm{BA}}$, we first define an auxiliary rank function rk'. Let $\mathcal{C}$ be a subset of $\mathcal{F}_{\mathrm{BA}}$, and A in $\mathcal{F}_{\mathrm{BA}}$. We define $\mathcal{C}$-rk'(A) by induction on the complexity of A. If $A \in \mathcal{C} \cup \neg\mathcal{C}$, let $\mathcal{C}$-rk'$(A) := -1$. For $A \notin \mathcal{C} \cup \neg\mathcal{C}$, $\mathcal{C}$-rk'(A) is defined as follows:

- Let $\mathcal{C}$-rk'$(A) := 1 + \max\{\mathcal{C}\text{-rk'}(B), \mathcal{C}\text{-rk'}(C)\}$ in case $A = B \ \wedge \ C$ or $A = B \ \vee \ C$.
- If $A = (\forall x)B$ or $A = (\exists x)B$, let $\mathcal{C}$-rk'$(A) := 1 + \mathcal{C}$-rk'(B).

Using the auxiliary rank function rk', we define the *$\mathcal{C}$-rank of A*, denoted $\mathcal{C}$-rk(A), by $\mathcal{C}$-rk$(A) := \max\{0, \mathcal{C}\text{-rk'}(A)\}$. Observe that $\mathrm{s}\Sigma^{\mathrm{b}}_i$-rk$(A) \le \mathrm{s}\Sigma^{\mathrm{b}}_{i+1}$-rk$(A) + 1$. If $\mathcal{C}$ is the set of quantifier-free formulas, and $\varphi \in \mathrm{s}\Sigma^{\mathrm{b}}_\infty$, then the rank of φ as defined in Section 2 is the same as $\mathcal{C}$-rk(φ), i.e. $\mathcal{C}$-rk(φ) computes the minimal k such that $\varphi \in \mathrm{s}\Sigma^{\mathrm{b}}_k \cup \mathrm{s}\Pi^{\mathrm{b}}_k$.

The negation function for the notation system is the same as defined for $\mathcal{L}_{\mathrm{BA}}$. Intensional equality is defined in the following way: For t a closed term its numerical value $t^{\mathbb{N}} \in \mathbb{N}$ is defined in the obvious way. Let $\rightarrow^1_{\mathbb{N}}$ denote the rewriting relation on $\mathcal{L}_{\mathrm{BA}}$-terms and $\mathcal{L}_{\mathrm{BA}}$-formulas obtained from

$$\big\{(t, \underline{t^{\mathbb{N}}}) \colon t \text{ a closed term}\big\} \quad .$$

Let $\approx_{\mathbb{N}}$ denote the reflexive, symmetric and transitive closure of $\rightarrow^1_{\mathbb{N}}$. For example, $(\forall x)((\underline{3} + \underline{1}) \cdot x = \underline{1} + \underline{5}) \approx_{\mathbb{N}} (\forall x)(\underline{4} \cdot x = \underline{6})$.

Proposition 5.1. *The system* $\langle \mathcal{F}_{\mathrm{BA}}, \mathrm{tp}, \cdot[\cdot], \neg, \mathrm{rk}, \approx_{\mathbb{N}} \rangle$ *which we have just defined forms a notation system for formulas in the sense of Definition 4.7.*

Let ${\approx_{\mathbb{N}}}^k$ denote the restriction of $\approx_{\mathbb{N}}$ to expressions of depth $\leq k$. In a feasible Gödel numbering, like the one defined in [Bus86], the Gödel number for c_a has size proportional to $|a|$. Thus, for each k, the relation ${\approx_{\mathbb{N}}}^k$ is a polynomial time predicate. We will always assume that $\mathcal{F}_{\mathrm{BA}}$ implicitly contains such a constant k without explicitly mentioning it. All formulas and terms used in $\mathcal{F}_{\mathrm{BA}}$ are thus assumed to obey the abovementioned restriction on depth. We will come back to this restriction at relevant places. The next observation already makes use of this assumption.

Observation 5.2. *All relations and functions in* $\mathcal{F}_{\mathrm{BA}}$ *are polynomial time computable.*

Definition 5.3. Let BA^∞ denote the propositional proof system over $\mathcal{F}_{\mathrm{BA}}$ according to Definition 4.8.

Definition 5.4. The *finitary proof system* $\mathrm{BA}^\star$ is the proof system over $\langle \Delta_0, \approx_{\mathbb{N}}, \mathrm{rk} \rangle$ which is given by the following set of inference symbols.

(Ax_Δ) $\quad \dfrac{}{\Delta}$ $\quad$ if $\bigvee \Delta \in \overline{\mathrm{BASIC}}$

$(\bigwedge_{A_0 \wedge A_1})$ $\quad \dfrac{A_0 \qquad A_1}{A_0 \wedge A_1}$ $\qquad$ $(\bigvee^k_{A_0 \vee A_1})$ $\quad \dfrac{A_k}{A_0 \vee A_1}$

$(\bigwedge^y_{(\forall x)A})$ $\quad \dfrac{A_x(y)}{(\forall x)A}$ $\qquad$ $(\bigvee^t_{(\exists x)A})$ $\quad \dfrac{A_x(t)}{(\exists x)A}$

$(\mathrm{IND}^{y,t}_F)$ $\quad \dfrac{\neg F, F_y(y+1)}{\neg F_y(0), F_y(2^{|t|})}$ $\qquad$ $(\mathrm{IND}^{y,n,i}_F)$ $\quad \dfrac{\neg F, F_y(y+1)}{\neg F_y(\underline{n}), F_y(\underline{n+2^i})}$

(Cut_C) $\quad \dfrac{C \qquad \neg C}{\varnothing}$ for $C \in \Delta_0$ with C atomic or $\mathrm{tp}(C) = \bigwedge$

where in case $(\bigvee^k_{A_0 \vee A_1})$ we have that $k \in \{0, 1\}$, and in case $(\mathrm{IND}^{y,n,i}_F)$ that $n, i \in \mathbb{N}$.

According to Definition 4.4, $\mathrm{BA}^\star$-quasi derivations h are equipped with functions $\Gamma(h)$ denoting the endsequent of h, $\mathrm{hgt}(h)$ denoting the height of h, and $\mathrm{sz}(h)$ denoting the size of h.

In the following we will not need the cut-rank function which comes with $\mathrm{BA}^\star$-quasi derivations, but we will need a more general cut-rank function gcrk, which will also bound the rank of induction formulas.

Definition 5.5. Let h be a $\mathrm{BA}^\star$-quasi derivation, $h = \mathcal{I} h_0 \cdots h_{n-1}$. We define

$$\mathcal{C}\text{-gcrk}(h) := \sup(\{\mathcal{C}\text{-grk}(\mathcal{I})\} \cup \{\mathcal{C}\text{-gcrk}(h_i) \colon i < n\})$$

where $\mathcal{C}$-grk($\mathcal{I}$), the *generalised cut-rank of* $\mathcal{I}$, is $\mathcal{C}\text{-rk}(C) + 1$ if $\mathcal{I}$ is of the form Cut_C, $\mathrm{IND}_C^{y,t}$ or $\mathrm{IND}_C^{y,n,i}$ for $C \notin \mathcal{C}$, and 0 otherwise.

Observe that $\mathrm{s}\Sigma_i^\mathrm{b}\text{-gcrk}(h) \leq \mathrm{s}\Sigma_{i+1}^\mathrm{b}\text{-gcrk}(h) + 1$, which immediately follows from $\mathrm{s}\Sigma_i^\mathrm{b}\text{-gcrk}(\mathcal{I}) \leq \mathrm{s}\Sigma_{i+1}^\mathrm{b}\text{-gcrk}(\mathcal{I}) + 1$.

Definition 5.6 (Inductive definition of $\vec{x} \colon h$ and $\mathrm{BA}^\star$-derivations)**.** For $\vec{x}$ a finite list of disjoint variables and $h = \mathcal{I} h_0 \cdots h_{n-1}$ a $\mathrm{BA}^\star$-quasi-derivation we inductively define the relation $\vec{x} \colon h$ that h *is a* $\mathrm{BA}^\star$*-derivation with free variables among* $\vec{x}$ as follows.

- If $\vec{x}, y \colon h_0$ and $\mathcal{I} \in \{\bigwedge^y_{(\forall x)A}, \mathrm{IND}_F^{y,t}, \mathrm{IND}_F^{y,n,i}\}$ for some A, F, t, n, i, and $\mathrm{FV}(t) \cup \mathrm{FV}(\Gamma(\mathcal{I} h_0)) \subset \{\vec{x}\}$ then $\vec{x} \colon \mathcal{I} h_0$.
- If $\vec{x} \colon h_0$ and $\mathrm{FV}((\exists x)A), \mathrm{FV}(t) \subseteq \{\vec{x}\}$ then $\vec{x} \colon \bigvee^t_{(\exists x)A} h_0$.
- If $\vec{x} \colon h_0$, $\vec{x} \colon h_1$ and $\mathrm{FV}(C) \subseteq \{\vec{x}\}$ then $\vec{x} \colon \mathrm{Cut}_C h_0 h_1$.
- If $\mathrm{FV}(\Delta) \subseteq \{\vec{x}\}$ then $\vec{x} \colon \mathrm{Ax}_\Delta$,
- If $\vec{x} \colon h_0$, $\vec{x} \colon h_1$ and $\mathcal{I} = \bigwedge_{A_0 \wedge A_1}$ with $\mathrm{FV}(A_0 \wedge A_1) \subset \{\vec{x}\}$ then $\vec{x} \colon \mathcal{I} h_0 h_1$.
- If $\vec{x} \colon h_0$ and $\mathcal{I} = \bigvee^k_{A_0 \vee A_1}$ with $\mathrm{FV}(A_0 \vee A_1) \subset \{\vec{x}\}$ then $\vec{x} \colon \mathcal{I} h_0$.

We call a $\mathrm{BA}^\star$-derivation h *closed*, if $\varnothing \colon h$.

Definition 5.7. For h a $\mathrm{BA}^\star$-derivation, y a variable and t a closed term of Bounded Arithmetic we define the substitution $h(t/y)$ inductively by setting $(\mathcal{I} h_0 \cdots h_{n-1})(t/y)$ to be $\mathcal{I}(t/y) h_0(t/y) \cdots h_{n-1}(t/y)$ if $\mathcal{I}$ is not of the form $\bigwedge^y_{(\forall x)A}$, $\mathrm{IND}_F^{y,t}$, or $\mathrm{IND}_F^{y,n,i}$ with the same variable y, and $\mathcal{I} h_0 \cdots h_{n-1}$ otherwise.

Substitution for inference symbols is defined by setting

$$\begin{array}{llll}
\mathrm{Ax}_\Delta(t/y) = \mathrm{Ax}_{\Delta(t/y)} & & & \\
\bigwedge_{A_0 \wedge A_1}(t/y) = \bigwedge_{(A_0 \wedge A_1)(t/y)} & & \bigvee^k_{A_0 \wedge A_1}(t/y) = \bigvee^k_{(A_0 \wedge A_1)(t/y)} & \\
\bigwedge^z_{(\forall x)A}(t/y) = \bigwedge^z_{((\forall x)A)(t/y)} & & \bigvee^{t'}_{(\exists x)A}(t/y) = \bigvee^{t'(t/y)}_{((\exists x)A)(t/y)} & \\
\mathrm{IND}_F^{z,t'}(t/y) = \mathrm{IND}_{F(t/y)}^{z,t'(t/y)} & & \mathrm{IND}_F^{z,n,i}(t/y) = \mathrm{IND}_{F(t/y)}^{z,n,i} &
\end{array}$$

The next Lemma shows the substitution property for $\mathrm{BA}^\star$-derivations. The strange looking "⊆" instead of the expected equality comes from the fact that a substitution may make formulas equal which are not equal without the substitution.

Lemma 5.8. *Assume* $\vec{x}\colon h$ *and let* y *be a variable and* t *a closed term, then* $\vec{x}\setminus\{y\}\colon h(t/y)$ *and moreover* $\Gamma(h(t/y)) \subseteq (\Gamma(h))(t/y)$.

We will now define the ingredients for a notation system $\mathcal{H}_{\mathrm{BA}}$ for BA^∞ according to Definition 4.9. The interpretation $[\![h]\!]$ for $h \in \mathcal{H}_{\mathrm{BA}}$ according to Definition 4.10 formalises a translation of closed $\mathrm{BA}^\star$-derivations into BA^∞, which is called an *embedding*.

Let $\mathcal{H}_{\mathrm{BA}}$ be the set of closed $\mathrm{BA}^\star$-derivations. For each $h \in \mathcal{H}_{\mathrm{BA}}$ we define the denoted last inference $\mathrm{tp}(h)$ as follows: Let $h = \mathcal{I}h_0 \cdots h_{n-1}$,

$$\mathrm{tp}(h) := \begin{cases} \mathrm{Ax}_A & \text{if } \mathcal{I} = \mathrm{Ax}_\Delta,\ \text{where } A \text{ is the} \\ & \qquad\quad \text{“least” true literal in } \Delta \\ \bigwedge_{A_0 \,\wedge\, A_1} & \text{if } \mathcal{I} = \bigwedge_{A_0 \,\wedge\, A_1} \\ \bigvee^k_{A_0 \,\vee\, A_1} & \text{if } \mathcal{I} = \bigvee^k_{A_0 \,\vee\, A_1} \\ \bigwedge_{(\forall x)A} & \text{if } \mathcal{I} = \bigwedge^y_{(\forall x)A} \\ \bigvee^{t^{\mathbb{N}}}_{(\exists x)A} & \text{if } \mathcal{I} = \bigvee^t_{(\exists x)A} \\ \mathrm{Rep} & \text{if } \mathcal{I} = \mathrm{IND}^{y,t}_F \\ \mathrm{Rep} & \text{if } \mathcal{I} = \mathrm{IND}^{y,n,0}_F \\ \mathrm{Cut}_{F_y(\underline{n+2^i})} & \text{if } \mathcal{I} = \mathrm{IND}^{y,n,i+1}_F \\ \mathrm{Cut}_C & \text{if } \mathcal{I} = \mathrm{Cut}_C \end{cases}$$

For each $h \in \mathcal{H}_{\mathrm{BA}}$ and $j \in \mathbb{N}$ we define the denoted subderivation $h[j]$ as follows: Let $h = \mathcal{I}h_0 \cdots h_{n-1}$. If $j \geq |\,\mathrm{tp}(h)|$ let $h[j] := \mathrm{Ax}_{0=0}$. Otherwise, assume $j < |\,\mathrm{tp}(h)|$ and define

$$h[j] := \begin{cases} h_{\min(j,1)} & \text{if } \mathcal{I} = \bigwedge_{A_0 \,\wedge\, A_1} \\ h_0 & \text{if } \mathcal{I} = \bigvee^k_{A_0 \,\vee\, A_1} \\ h_0(\underline{j}/y) & \text{if } \mathcal{I} = \bigwedge^y_{(\forall x)A} \\ h_0 & \text{if } \mathcal{I} = \bigvee^t_{(\exists x)A} \\ \mathrm{IND}^{y,0,|t|^{\mathbb{N}}}_F h_0 & \text{if } \mathcal{I} = \mathrm{IND}^{y,t}_F \\ h_0(\underline{n}/y) & \text{if } \mathcal{I} = \mathrm{IND}^{y,n,0}_F \\ \mathrm{IND}^{y,n,i}_F h_0 & \text{if } \mathcal{I} = \mathrm{IND}^{y,n,i+1}_F \text{ and } j = 0 \\ \mathrm{IND}^{y,n+2^i,i}_F h_0 & \text{if } \mathcal{I} = \mathrm{IND}^{y,n,i+1}_F \text{ and } j = 1 \\ h_j & \text{if } \mathcal{I} = \mathrm{Cut}_C \end{cases}$$

The denoted end-sequent function on $\mathcal{H}_{\mathrm{BA}}$ is given by Γ. The size function $|\cdot|$ on $\mathcal{H}_{\mathrm{BA}}$ is given by $|h| := \mathrm{sz}(h)$. We define the denoted cut-rank function for $h \in \mathcal{H}_{\mathrm{BA}}$ to be $\mathcal{C}\text{-crk}(h) := \mathcal{C}\text{-gcrk}(h)$. We observe that $\mathcal{C}\text{-crk}(h[\iota]) \leq \mathcal{C}\text{-crk}(h)$ for $\iota < |\,\mathrm{tp}(h)|$, and that $\mathcal{C}\text{-rk}(C) < \mathcal{C}\text{-crk}(h)$ if $\mathrm{tp}(h) = \mathrm{Cut}_C$ and $C \notin \mathcal{C}$.

To define the denoted height function we need some analysis yielding an upper bound to the log of the lengths of inductions which may occur during the embedding (we take the log as this bounds the height of the derivation tree which embeds an application of induction). Let us first assume m is such an upper bound, and let us define the denoted height $\mathrm{o}_m(h)$ of h relative to m: For a $\mathrm{BA}^\star$-derivation $h = \mathcal{I}h_0 \cdots h_{n-1}$ we define

$$\mathrm{o}_m(h) := \begin{cases} \mathrm{o}_m(h_0) + i + 1 & \text{if } \mathcal{I} = \mathrm{IND}_F^{y,n,i} \\ \mathrm{o}_m(h_0) + m + 1 & \text{if } \mathcal{I} = \mathrm{IND}_F^{y,t} \\ 1 + \sup_{i<n} \mathrm{o}_m(h_i) & \text{otherwise} \end{cases}$$

Observe that $\mathrm{o}_m(h) > 0$ (in particular, $\mathrm{o}(\mathrm{Ax}_\Delta) = 1$).

To fill the gap of providing a suitable upper bound function of $\mathrm{BA}^\star$-derivations we first need to fix monotone bounding terms for any term in $\mathcal{L}_{\mathrm{BA}}$.

Bounding Terms for Language and Proofs

For a term t we define a term $\mathrm{bd}(t)$ which represents a monotone function with the following property: If $\mathrm{FV}(t) = \{\vec{x}\}$ then

$$(\forall \vec{n}) \qquad t_{\vec{x}}(\underline{\vec{n}})^{\mathbb{N}} \quad \leq \quad \mathrm{bd}(t)_{\vec{x}}(\underline{\vec{n}})^{\mathbb{N}}$$

The precise definition of $\mathrm{bd}(t)$ is not essential here, we can for example use the meta-function σ from [Bus86, p.77], or the explicit definition given in [AB09].

For $h \in \mathcal{H}_{\mathrm{BA}}$, the bounding term $\mathrm{bd}(h)$ is intended to bound any variable which occurs during the embedding of h. Then, the term $|\mathrm{bd}(h)|$ will bound the length of any induction which occurs during the embedding of h. This situation is related to the notion *proofs restricted by parameter variables* as defined in [Bus86, Section 4.5], where proofs are transformed in such a way that bounds to inductions and quantification only depend on the parameter variables of the proof — then the above mentioned bounding term $\mathrm{bd}(h)$ can simply be obtained by collecting all such bounds and taking their maximum. Let $h = \mathcal{I}h_0 \cdots h_{n-1}$ be in $\mathcal{H}_{\mathrm{BA}}$. Let $\max(n_1, \ldots, n_k)$ denote the maximal value amongst $\{n_1, \ldots, n_k\}$, where we set

$\max() = 0$. We define

$$\mathrm{bd}(h) := \begin{cases} \max(\mathrm{bd}(h_0(\underline{\mathrm{bd}(t)}/y)), \mathrm{bd}(t)) & \text{if } \mathcal{I} = \bigwedge^{y}_{(\forall x \le t)A} \\ \max(\mathrm{bd}(h_0), \mathrm{bd}(t)) & \text{if } \mathcal{I} = \bigvee^{t}_{(\exists x)A} \\ \max(\mathrm{bd}(h_0(\underline{2^{|\,\mathrm{bd}(t)|}}/y)), 2^{|\,\mathrm{bd}(t)|}) & \text{if } \mathcal{I} = \mathrm{IND}^{y,t}_{F} \\ \max(\mathrm{bd}(h_0(\underline{n+2^i}/y)), n+2^i) & \text{if } \mathcal{I} = \mathrm{IND}^{y,n,i}_{F} \\ \max(\mathrm{bd}(h_0), \dots, \mathrm{bd}(h_{n-1})) & \text{otherwise.} \end{cases}$$

Now we define for $h \in \mathcal{H}_{\mathrm{BA}}$ the denoted height function $\mathrm{o}(h)$ as $\mathrm{o}_{|\mathrm{bd}(h)|}(h)$.

Theorem 5.9. *The just defined system $\langle \mathcal{H}_{\mathrm{BA}}, \mathrm{tp}, \cdot[\cdot], \Gamma, \mathrm{crk}, \mathrm{o}(\cdot), |\cdot| \rangle$ forms a notation system for BA^{∞} in the sense of Definition 4.9. Furthermore, $\mathcal{H}_{\mathrm{BA}}$ is bounded in the sense of Definition 4.13.*

A proof of this Theorem can be found in [AB09]. The fact that $\mathcal{H}_{\mathrm{BA}}$ is bounded is easily observed by inspection.

Observation 5.10. *We assume that we have fixed a $k \in \mathbb{N}$ bounding depths of formulas and terms as explained in the remark on page 85, and some feasible Gödel numbering like the one in [Bus86]. Then, the following relations and functions are polynomial time computable (when interpreted as relations and functions on the corresponding Gödel numbers of syntactical objects): the finitary proof system $\mathrm{BA}^{\star}$, the set of $\mathrm{BA}^{\star}$-quasi derivations and the functions $h \mapsto \Gamma(h)$, $h \mapsto \mathrm{hgt}(h)$, and $h \mapsto \mathrm{sz}(h)$ denoting the endsequent, the height and the size for a $\mathrm{BA}^{\star}$-quasi derivation h; the bounding term $t \mapsto \mathrm{bd}(t)$ for terms t occurring in $\mathcal{F}_{\mathrm{BA}}$ and the relations $\mathrm{bd}(h) \le m$ on $\mathcal{H}_{\mathrm{BA}} \times \mathbb{N}$; the set $\mathcal{H}_{\mathrm{BA}}$ and the functions $h \mapsto \mathrm{tp}(h)$, $h, i \mapsto h[i]$, $h \mapsto \Gamma(h)$, $m, h \mapsto \mathrm{o}_m(h)$ and $h \mapsto |h|$.*

We now provide a connection between $\mathrm{BA}^{\star}/\mathcal{H}_{\mathrm{BA}}$ and the theories of Bounded Arithmetic as defined in Section 2. This step also includes some proof normalisation which is similar to known ones in the literature, for example free cut-elimination in [Bus86] or partial cut-elimination in [Bec03].

Theorem 5.11 (Partial Cut-elimination)**.** *Assume $\mathrm{T}^j_2 \vdash \varphi$ with $\varphi \in \Delta_0$ and $\mathrm{FV}(\varphi) \subseteq \{x\}$. Then, there is some $\mathrm{BA}^{\star}$-derivation h such that $\mathrm{FV}(h) \subseteq \{x\}$, $\Gamma(h) = \{\varphi\}$, $\mathrm{s}\Sigma^{\mathrm{b}}_j$-$\mathrm{gcrk}(h) = 0$ and $\mathrm{o}(h(\underline{a}/x)) = |a|^{O(1)}$.*

A proof of the last theorem can be found in [AB09].

5.1 Complexity Notions for $\mathrm{BA}^{\star}$

In order to describe local search problems based on proof notations we need some notions describing key complexity properties of $\mathrm{BA}^{\star}$ proof notations. Again, we will just state the necessary definitions and results, more details including full proofs can be found in [AB09].

Definition 5.12. We extend the definition of bounding terms $\mathrm{bd}(h)$ from $\mathcal{H}_{\mathrm{BA}}$ to $\mathcal{CH}_{\mathrm{BA}}$ by induction on $h \in \mathcal{CH}_{\mathrm{BA}}$ in the following way:

- If $h \in \mathcal{H}_{\mathrm{BA}}$ then the definition of $\mathrm{bd}(h)$ is inherited from the definition of $\mathrm{bd}(h)$ on $\mathcal{H}_{\mathrm{BA}}$.
- $\mathrm{bd}(\mathsf{I}^k_C h_0) := \mathrm{bd}(h_0)$.
- $\mathrm{bd}(\mathsf{R}_C h_0 h_1) := \max\{\mathrm{bd}(h_0), \mathrm{bd}(h_1)\}$.
- $\mathrm{bd}(\mathsf{E} h_0) := \mathrm{bd}(h_0)$.

Lemma 5.13. *Let* $h \in \mathcal{CH}_{\mathrm{BA}}$.

1. $\mathrm{bd}(h[j]) \le \mathrm{bd}(h)$ *for all* j.
2. *If* $\mathrm{tp}(h) = \bigvee^k_C$ *then* $k \le \mathrm{bd}(h)$.

Definition 5.14. For h a $\mathrm{BA}^\star$-derivation or $h \in \mathcal{CH}_{\mathrm{BA}}$, we define *the set of decorations of* h*,* $\mathrm{deco}(h)$, by induction on h. $\mathrm{deco}(h)$ will be a finite set of $\mathcal{L}_{\mathrm{BA}}$-terms and formulas in Δ_0. Let $h = \mathcal{I} h_0 \cdots h_{n-1}$, where $\mathcal{I}$ ranges over $\mathrm{BA}^\star \cup \{\mathsf{I}^k_C, \mathsf{R}_C, \mathsf{E}\}$. We define

$$\mathrm{deco}(h) := \mathrm{deco}(\mathcal{I}) \cup \bigcup_{i<n} \mathrm{deco}(h_i)$$

where

$$\begin{aligned}
\mathrm{deco}(\mathcal{I}) &:= \Delta(\mathcal{I}) \text{ for } \mathcal{I} = \mathrm{Ax}_\Delta, \textstyle\bigwedge_{A_0 \wedge A_1}, \bigvee^k_{A_0 \vee A_1} \\
\mathrm{deco}(\textstyle\bigwedge^y_{(\forall x)A}) &:= \{(\forall x)A, y\} \\
\mathrm{deco}(\textstyle\bigvee^t_{(\exists x)A}) &:= \{(\exists x)A, t\} \\
\mathrm{deco}(\mathrm{IND}^{y,t}_F) &:= \{F, \neg F_y(0), F_y(2^{|t|}), y, t\} \\
\mathrm{deco}(\mathrm{IND}^{y,n,i}_F) &:= \{F, \neg F_y(\underline{n}), F_y(\underline{n+2^i}), y, c_n\} \\
\mathrm{deco}(\mathrm{Cut}_C) &:= \{C\} \\
\mathrm{deco}(\mathsf{I}^k_C) &:= \{C, C[k], c_k\} \\
\mathrm{deco}(\mathsf{R}_C) &:= \{C\} \\
\mathrm{deco}(\mathsf{E}) &:= \varnothing \ .
\end{aligned}$$

Observation 5.15. *We have* $\Gamma(h) \subseteq \mathrm{deco}(h)$.

Definition 5.16. Let Φ be a set of $\mathcal{L}_{\mathrm{BA}}$-terms and formulas in Δ_0, and let $K \in \mathbb{N}$ be a size parameter. With Φ_K we denote the set obtained by enlarging Φ by the set $\{c_i : 0 \le i \le K\}$ and the set of formulas and terms which result from formulas and terms in Φ by substituting constants from $\{c_i : 0 \le i \le K\}$ for some (possibly none, possibly all) of the free variables.

Lemma 5.17. *Let Φ be a set of $\mathcal{L}_{\mathrm{BA}}$-terms and formulas in Δ_0, such that $\Phi \cap \Delta_0$ is closed under negation and taking subformulas. Let $j, K \in \mathbb{N}$ and y be a variable.*

1. *If $j \le K$ and $C \in \Phi \cap \Delta_0$, then $C[j] \in \Phi_K$.*
2. *If $h \in \mathrm{BA}^\star$ with $\mathrm{deco}(h) \subseteq \Phi$, and $j \le K$, then $\mathrm{deco}(h(\underline{j}/y)) \subseteq \Phi_K$.*
3. $\Delta(\mathrm{tp}(h)) \subseteq \mathrm{deco}(h)_{\mathrm{bd}(h)}$ *with the subscript understood in the sense of Definition 5.16.*
4. *If $h \in \mathcal{CH}_{\mathrm{BA}}$ with $\mathrm{deco}(h) \subseteq \Phi$ and $j \le K$, then $\mathrm{deco}(h[j]) \subseteq \Phi_{\max\{K, \mathrm{bd}(h)\}}$.*

Lemma 5.18. *For $h \in \mathcal{CH}_{\mathrm{BA}}$ we have that the cardinality of $\Gamma(h)$ is bounded above by $2 \cdot \mathrm{sz}(h)$.*

6 Searching for Truth

As explained in the introduction, the definition of search problems based on proof notations has to deal with properties whose computational complexity is too complicated to decide them directly. Therefore, instead of deciding them, we will replace them by some canonical search problem which determines their truth. This section will provide the definition and basic properties for such canonical search problems. In the next subsection we will present some general notation for tuples and sequences which will also be useful in later sections when we discuss the Skolemisation of prenex formulas that arise from search problems. The subsequent subsection then introduces canonical search problems for properties in $\mathrm{s}\Pi^{\mathrm{b}}_k$.

6.1 Notations for Tuples and Sequences

In order to have succinct notations for prenex formulas and for our discussion of Skolemisation, we introduce formal tuples, and in particular tuples of variables and quantifiers, and tuple quantification for tuples of variables. These tuples are formed and used on the meta level, they are not available in $\mathcal{L}_{\mathrm{BA}}$.

At the end of the section we will also introduce sequence coding which will be available within $\mathcal{L}_{\mathrm{BA}}$. Sequences will be used to define various functions and relations related to search problems.

Definition 6.1 (General Tuples)**.** A *tuple of length k* is an expression of the form $[t_1, \ldots, t_k]$ with t_i some formal expression. We will use the letter $\mathfrak{t}$ as a meta-variable for general tuples. We will use subscripts of the form $\mathfrak{t}_i$ only to denote the i-th element t_i of $\mathfrak{t}$. Let $[t_1, \ldots, t_k]\lceil_\ell$ denote $[t_1, \ldots, t_{\min(k,\ell)}]$.

Definition 6.2 (Tuples of Variables)**.** A *tuple of variables of length k* is an expression of the form $[z_1, \ldots, z_k]$ with z_i being a formal variable in $\mathcal{L}_{\mathrm{BA}}$. We will use the letter $\mathfrak{z}$ (possibly with superscripts) as a meta-variable for tuples of variables. $\mathfrak{z}_i$ and $\mathfrak{z}\lceil_\ell$ are defined as for general tuples.

Definition 6.3 (Tuples of Quantifiers)**.** A *tuple of quantifiers of length k* is an expression of the form $[Q_1, \ldots, Q_k]$ with $Q_i \in \{\exists, \forall\}$. We will use the letter $\mathfrak{Q}$ (possibly with super-scripts) as a meta-variable for tuples of quantifiers.

Let $\mathfrak{Q} = [Q_1, \ldots, Q_k]$ be a tuple of quantifiers of length k. The expression $\neg\mathfrak{Q}$ denotes the tuple $[\neg Q_1, \ldots, \neg Q_k]$ where $\neg\forall$ denotes $\exists$, and $\neg\exists$ denotes $\forall$. The expression $\forall^k$ denotes the tuple $[\forall, \ldots, \forall]$ of length k. The expression $\forall\!\!\!\exists^k$ denotes the tuple $[\forall, \exists, \forall, \exists, \ldots]$ of length k. The expression $\exists\!\!\!\forall^k$ denotes the tuple $\neg\forall\!\!\!\exists^k$.

Definition 6.4 (Tuple Quantification)**.** Let $\mathfrak{Q} = [Q_1, \ldots, Q_k]$ be a tuple of quantifiers of length k, and $\mathfrak{z} = [z_1, \ldots, z_k]$ a tuple of variables of length k. The expression $(\mathfrak{Q}\mathfrak{z})\beta$ denotes the formula

$$(Q_1 z_1)(Q_2 z_2) \cdots (Q_k z_k)\beta \enspace .$$

We now fix a coding of sequences of numbers of fixed length. As the length of sequences will always be fixed on the meta-level, we can choose a sequence coding based on a feasible pairing function. In principle we could define a concrete pairing function which does not use the #-function, but the mere existence will suffice for our investigations. This definition of sequence coding may however play a role in investigations of fragments of bounded arithmetic which do not include the #-function, but we do not pursue these here.

Let us remind that a feasible pairing function $a, b \mapsto \mathrm{pair}(a, b)$ with projection functions $c \mapsto (c)_1$ and $c \mapsto (c)_2$ are fixed in $\mathcal{L}_{\mathrm{BA}}$ which satisfy $(\mathrm{pair}(a, b))_1 = a$ and $(\mathrm{pair}(a, b))_2 = b$ and some natural bounding conditions like $(c)_i \leq c$ and $a, b \leq t \to \mathrm{pair}(a, b) \leq B(t)$ for some $\mathcal{L}_{\mathrm{BA}}$-term B.

Definition 6.5 (Sequence Coding)**.** We use pairing to define sequences of fixed length by letting $\langle\,\rangle = 0$, and $\langle a_1, \ldots, a_{k+1}\rangle = \mathrm{pair}(a_1, \langle a_2, \ldots, a_{k+1}\rangle)$ with corresponding projections p_i. The projection function p_i picks out the i-th element of a sequence; that is, $\mathrm{p}_i(\langle a_1, \dots, a_k \rangle) = a_i$.

We use $\mathfrak{s}$ (possibly with superscripts) as meta-variables to denote sequences. For sequences denoted by $\mathfrak{s}$, we often write $\mathfrak{s}_i$ to denote the i-th element, $\mathrm{p}_i(\mathfrak{s})$, of $\mathfrak{s}$. We also use well-known list notation for sequences. The empty sequence of length 0 is denoted by $\langle\,\rangle$. If $\mathfrak{s}$ is a sequence of length l, then $\langle a \,|\, \mathfrak{s} \rangle$ denotes the sequence of length $l+1$ given by $\langle a \,|\, \mathfrak{s} \rangle = \mathrm{pair}(a, \mathfrak{s})$. We also use expressions of the form $\langle a, b, c \,|\, \mathfrak{s} \rangle = \langle a \,|\, \langle b \,|\, \langle c \,|\, \mathfrak{s} \rangle\rangle\rangle$, and $\langle a, b, c \rangle = \langle a, b, c \,|\, \langle\,\rangle\rangle$, etc.

We also define the application of the projection function p_i to formal tuples $\mathfrak{t} = [t_1, \dots, t_k]$ to denote the application of p_i to each of the elements of $\mathfrak{t}$, that is, $\mathrm{p}_i(\mathfrak{t}) = [\mathrm{p}_i(t_1), \dots, \mathrm{p}_i(t_k)]$.

6.2 Canonical Search Problems for Properties in $\mathrm{s}\Pi^{\mathrm{b}}_k$

In this subsection we define a canonical search problem for each formula in $\mathrm{s}\Sigma^{\mathrm{b}}_\infty$. The canonical search problem will be used to determine the truth of the formula. To define the search space for a formula φ, we need an upper bound to all values which may occur as quantified values in the evaluation of φ. The next definition provides the necessary requirements which we will need for such upper bounds.

Definition 6.6 (Strict Upper Bounds)**.** Let φ be of the form $(\mathfrak{Q}\mathfrak{z})\beta$ for some quantifier-free β, $\mathfrak{Q} = [Q_1, \dots, Q_k]$ and $\mathfrak{z} = [z_1, \dots, z_k]$. An $\mathcal{L}_{\mathrm{BA}}$-term D is called a *strict upper bound (s.u.b.) for* φ if its free variables are amongst those of φ, and if it satisfies the following properties: Let $\mathfrak{Q}^i := [Q_{i+1}, \dots, Q_k]$ and $\mathfrak{z}^i := [z_{i+1}, \dots, z_k]$. For all $1 \le i \le k$ with $Q_i = \forall$,

$$\mathrm{S}^1_2 \vdash (\forall z_1) \cdots (\forall z_{i-1}) \Big((\forall z_i < D)(\mathfrak{Q}^i \mathfrak{z}^i)\beta \;\to\; (\forall z_i)(\mathfrak{Q}^i \mathfrak{z}^i)\beta \Big) \;\; ,$$

and for all i with $Q_i = \exists$,

$$\mathrm{S}^1_2 \vdash (\forall z_1) \cdots (\forall z_{i-1}) \Big((\exists z_i)(\mathfrak{Q}^i \mathfrak{z}^i)\beta \;\to\; (\exists z_i < D)(\mathfrak{Q}^i \mathfrak{z}^i)\beta \Big) \;\; .$$

Definition 6.7. For $\varphi \in \mathrm{s}\Sigma^{\mathrm{b}}_\infty$ we can define *the canonical s.u.b.* D_φ *for* φ inductively as follows:

- If φ is quantifier free, then let $D_\varphi := 0$.
- If φ is of the form $(\forall x \le t)\psi$ or $(\exists x \le t)\psi$, then let D_φ be the term $\max\{\mathrm{bd}(t) + 1, D_\psi(x/\mathrm{bd}(t))\}$.

We observe that D_φ represents a monotone function in its variables. Thus, D_φ is a s.u.b. in the sense of Definition 6.6, which can be shown immediately by induction on the complexity of φ.

Notation 6.8. Let 0^k denote the sequence of length k consisting only of zeros.

Let φ be a formula in $\mathrm{s}\Sigma^{\mathrm{b}}_\infty$ and $\vec{a}$ a list of variables such that $\mathrm{FV}(\varphi) \subseteq \{\vec{a}\}$. Let $D = D(\vec{a})$ be a s.u.b. for φ. We define the *canonical search problem* S^D_φ for φ whose aim is to determine the truth value for φ. S^D_φ is defined similar to a Π^{b}_k-PLS problem with Π^{b}_ℓ-goal from Definition 3.2, but instead of a goal set, S^D_φ has an *answer set* A^D_φ of low computational complexity which determines the truth of φ: For a solution $\mathfrak{s}$ to the search problem, φ is true iff $\mathfrak{s} \in A_\varphi$. The answer set will later be used to define the neighbourhood function for Π^{b}_k-PLS problems, which have to be of low complexity. The idea to determine the truth of φ of, say, the form $(\exists x < D)\psi(x)$ is to successively "search" for the truth of $\psi(0)$, $\psi(1)$,..., $\psi(D-1)$. If any of these intermediate searches are successful, the overall search will be successful and will yield a value d (usually the first such) for which $\psi(d)$ produces success; otherwise the overall search will yield a value D indicating that none of the intermediate searches were successful.

We start by defining the configuration space and cost function which only depend on rank of formulas and not on their actual form.

Definition 6.9 (Configuration Space)**.** Let $k \geq 0$ and $D \geq 1$. The *configuration space* $C^{k,D}$ is the set of all sequences of length k of elements $\leq D$, i.e. $\{\langle u_1, \ldots, u_k\rangle : u_1, \ldots, u_k \leq D\}$. The *cost function* c^D can be defined on all sequences as

$$c^D(\langle u_k, \ldots, u_1\rangle) := \sum_{i=1}^{k} (D \dot{-} u_i)(D+1)^{(i-1)}$$

It has the properties that $0 \leq c^D(\mathfrak{s}) < (D+1)^k$ for all $\mathfrak{s} \in C^{k,D}$, and that $c^D(\mathfrak{s}_1) > c^D(\mathfrak{s}_2)$ if $\mathfrak{s}_1$ is smaller than $\mathfrak{s}_2$ w.r.t. the lexicographical order on tuples on $C^{k,D}$.

Definition 6.10. The *canonical search problem* S^D_φ *of* φ, given by the system $(C^D_\varphi, F^D_\varphi, A^D_\varphi, N^D_\varphi, c^D_\varphi)$, consists of a *configuration space* C^D_φ, a *set of feasible solutions* F^D_φ which is a subset of the configuration space, an *answer set* A^D_φ which is a subset of the configuration space, a *neighbourhood function* N^D_φ which maps configurations to configurations, and a *cost function* c^D_φ defined for configurations. The goal of the search problem is to find some $\mathfrak{s} \in F^D_\varphi$ with $N^D_\varphi(\mathfrak{s}) = \mathfrak{s}$.

The defined sets and functions all implicitly depend on the parameters $\vec{a}$ of φ. We will usually not mention D as it is understood from the context.

The configuration space C^D_φ is $C^{\mathrm{rk}(\varphi),D}$ from the previous definition, and the cost function c^D_φ is the cost function c^D from the previous definition with domain restricted to C^D_φ.

The set of feasible solutions F_φ, the neighbourhood function N_φ and the answer set A_φ also implicitly include parameter variables $\vec{a}$. They are defined by induction on the complexity of φ.

If φ is in $\mathrm{s}\Sigma^{\mathrm{b}}_0 \cup \mathrm{s}\Pi^{\mathrm{b}}_0$ we define

$$\begin{aligned} F_\varphi &:= \{\langle\,\rangle\} \\ N_\varphi(\langle\,\rangle) &:= \langle\,\rangle \\ \langle\,\rangle \in A_\varphi &:\Leftrightarrow \varphi \end{aligned}$$

Let φ be in $\mathrm{s}\Sigma^{\mathrm{b}}_{k+1} \setminus \mathrm{s}\Pi^{\mathrm{b}}_{k+1}$ of the form $(\exists x)\psi$. ψ has (potentially) one free variable in addition to φ which is x. Thus, when defining F, N and A in the following, their first argument will denote the value for this additional parameter. We will make this dependency explicit by writing ψx in the index of F, N, A, resp. We define

$$F_\varphi := \{\langle d \,|\, \mathfrak{s}\rangle \in C_\varphi \colon \mathfrak{s} \in F_{\psi x}(d) \;\wedge\; (\forall x < d)\neg\psi(x)\}$$

$$N_\varphi(\langle d \,|\, \mathfrak{s}\rangle) := \begin{cases} \langle d \,|\, N_{\psi x}(d,\mathfrak{s})\rangle & \text{if } d < D \;\wedge\; N_{\psi x}(d,\mathfrak{s}) \neq \mathfrak{s} \\ \langle d \,|\, \mathfrak{s}\rangle & \text{if } d < D \;\wedge\; N_{\psi x}(d,\mathfrak{s}) = \mathfrak{s} \in A_{\psi x}(d) \\ \left\langle d+1 \,|\, 0^k\right\rangle & \text{if } d < D \;\wedge\; N_{\psi x}(d,\mathfrak{s}) = \mathfrak{s} \notin A_{\psi x}(d) \\ \langle d \,|\, \mathfrak{s}\rangle & \text{if } d = D \end{cases}$$

$$\langle d \,|\, \mathfrak{s}\rangle \in A_\varphi \Leftrightarrow d < D$$

For $\varphi \in \mathrm{s}\Pi^{\mathrm{b}}_{k+1} \setminus \mathrm{s}\Sigma^{\mathrm{b}}_{k+1}$ we define

$$\begin{aligned} F_\varphi &:= F_{\neg\varphi} \\ N_\varphi(\mathfrak{s}) &:= N_{\neg\varphi}(\mathfrak{s}) \\ A_\varphi &:= C_\varphi \setminus A_{\neg\varphi} \end{aligned}$$

The latter choices imply for φ of the form $(\forall x)\psi$ that

$$F_\varphi := \{\langle d \,|\, \mathfrak{s}\rangle \in C_\varphi \colon \mathfrak{s} \in F_{\psi x}(d) \;\wedge\; (\forall x < d)\psi(x)\}$$

$$N_\varphi(\langle d \,|\, \mathfrak{s}\rangle) := \begin{cases} \langle d \,|\, N_{\psi x}(d,\mathfrak{s})\rangle & \text{if } d < D \;\wedge\; N_{\psi x}(d,\mathfrak{s}) \neq \mathfrak{s} \\ \langle d \,|\, \mathfrak{s}\rangle & \text{if } d < D \;\wedge\; N_{\psi x}(d,\mathfrak{s}) = \mathfrak{s} \notin A_{\psi x}(d) \\ \left\langle d+1 \,|\, 0^k\right\rangle & \text{if } d < D \;\wedge\; N_{\psi x}(d,\mathfrak{s}) = \mathfrak{s} \in A_{\psi x}(d) \\ \langle d \,|\, \mathfrak{s}\rangle & \text{if } d = D \end{cases}$$

$$\langle d \,|\, \mathfrak{s}\rangle \in A_\varphi \Leftrightarrow d = D \qquad \text{assuming } \mathfrak{s} \in C_\psi$$

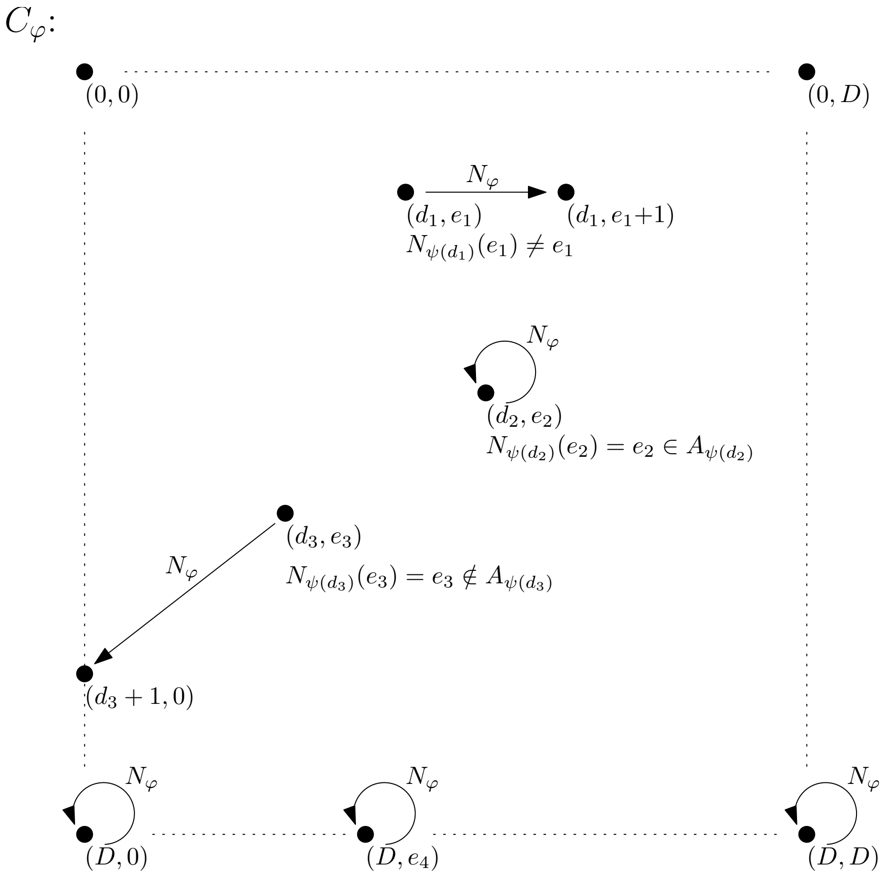

Figure 1: The canonical search problem $S_\varphi = (C_\varphi, F_\varphi, A_\varphi, N_\varphi, c_\varphi)$ for φ in $\mathrm{s}\Sigma^{\mathrm{b}}_2 \setminus \mathrm{s}\Pi^{\mathrm{b}}_2$ of the form $(\exists x)\psi(x)$, and D a strict upper bound for φ. The configuration space C_φ is a grid consisting of all points $\langle d,e\rangle$ with $0 \le d \le D$ and $0 \le e \le D$. N_φ is defined for all points on the grid. Its behaviour at $\langle d,e\rangle$ depends on the behaviour of the canonical search problem for $S_{\psi x}(d) = S_{\psi(d)}$, in particular on $N_{\psi x}(d,e) = N_{\psi(d)}(e)$ and $A_{\psi x}(d) = A_{\psi(d)}$.

Definition 6.11. Let $\varphi \in \mathrm{s}\Sigma^{\mathrm{b}}_{\infty}$, and let $S^D_\varphi = (C_\varphi, F_\varphi, A_\varphi, N_\varphi, c_\varphi)$ be the canonical search problem for φ. Let k be the rank of φ. We extend the definition of F_φ, A_φ and N_φ to sequences of length $\ell > k$ in the obvious way:

$$\langle u_1, \ldots, u_\ell \rangle \in F_\varphi \quad :\Longleftrightarrow \quad \langle u_1, \ldots, u_k \rangle \in F_\varphi$$
$$\langle u_1, \ldots, u_\ell \rangle \in A_\varphi \quad :\Longleftrightarrow \quad \langle u_1, \ldots, u_k \rangle \in A_\varphi$$

and if $N_\varphi(\langle u_1, \ldots, u_k \rangle) = \langle v_1, \ldots, v_k \rangle$ then

$$N_\varphi(\langle u_1, \ldots, u_\ell \rangle) \quad := \quad \langle v_1, \ldots, v_k, u_{k+1}, \ldots, u_\ell \rangle \ .$$

To explain the previous two definitions let us calculate F_φ for φ of rank $k > 0$: For $k = 1$ and $\varphi \equiv (\exists x)\beta(x)$ we have $\langle u \,|\, \mathfrak{s} \rangle \in F_\varphi \equiv (\forall x{<}u)\neg\beta(x)$. If $k = 2$ and $\varphi \equiv (\exists x)(\forall y)\beta(x, y)$ then $\langle u, v \,|\, \mathfrak{s} \rangle \in F_\varphi$ has the form

$$(\forall x{<}u)(\exists y)\neg\beta(x, y) \ \wedge \ (\forall y{<}v)\beta(u, y) \ .$$

If $k = 3$ and $\varphi \equiv (\exists x)(\forall y)(\exists z)\beta(x, y, z)$ we have that $\langle u, v, w \,|\, \mathfrak{s} \rangle \in F_\varphi$ is of the form

$$(\forall x{<}u)(\exists y)(\forall z)\neg\beta(x, y, z) \ \wedge \ (\forall y{<}v)(\exists z)\beta(u, y, z) \ \wedge \ (\forall z{<}w)\neg\beta(u, v, z) \ .$$

For the general case assume $\varphi \equiv (\exists x)(\forall y)\psi(x, y)$. Then $\langle u, v \,|\, \mathfrak{s} \rangle \in F_\varphi$ has the form

$$(\forall x{<}u)(\exists y)\neg\psi(x, y) \ \wedge \ (\forall y{<}v)\psi(u, y) \ \wedge \ \mathfrak{s} \in F_{\psi x y}(u, v)$$

Observation 6.12. *Let $\varphi \in \mathrm{s}\Sigma^{\mathrm{b}}_{\infty}$, and let k be the rank of φ according to Definition 2.6. Let $S_\varphi = (C_\varphi, F_\varphi, A_\varphi, N_\varphi, c_\varphi)$ be the canonical search problem for φ. Then, $C_\varphi, A_\varphi, N_\varphi$ and c_φ are polynomial time computable, and F_φ is in the level Π^{p}_k of the polynomial time hierarchy. More precisely, we observe that $\mathfrak{s} \in C_\varphi(a)$, $\mathfrak{s} \in A_\varphi(a)$ and $N_\varphi(a, \mathfrak{s}) = \mathfrak{s}'$ can be defined by $\mathrm{s}\Sigma^{\mathrm{b}}_0$-formulas, $c_\varphi(a, \mathfrak{s})$ can be defined by $\mathcal{L}_{\mathrm{BA}}$-terms, and $\mathfrak{s} \in F_\varphi(a)$ is equivalent to a $\mathrm{s}\Pi^{\mathrm{b}}_k$-formula in $\overline{\mathrm{BASIC}}$.*

Proposition 6.13. *Let $\varphi \in \mathrm{s}\Sigma^{\mathrm{b}}_{\infty} \setminus \mathrm{s}\Sigma^{\mathrm{b}}_1 \cup \mathrm{s}\Pi^{\mathrm{b}}_1$, D an s.u.b. for φ, and let $S^D_\varphi = (C^D_\varphi, F^D_\varphi, A^D_\varphi, N^D_\varphi, c^D_\varphi)$ be the canonical search problem for φ. The following is provable in $\overline{\mathrm{BASIC}}$. Assume $N^D_\varphi(\mathfrak{s}) = \mathfrak{s}$, then either $\mathfrak{s}_1 = D$, or $\mathfrak{s}_1 < D$ and $\mathfrak{s}_2 = D$.*

Proof. It is enough to consider φ of the form $(\exists x)(\forall y)\psi(x, y)$, as $N^D_{\neg\varphi}(\mathfrak{s}) = N^D_\varphi(\mathfrak{s})$. Let $\mathfrak{s} = \langle d, e \,|\, \mathfrak{s}' \rangle$ and assume $N^D_\varphi(\mathfrak{s}) = \mathfrak{s}$ and $d \neq D$, then we have to show $d < D$ and $e = D$. The definition of N^D_φ implies $d < D$ and $\langle e \,|\, \mathfrak{s}' \rangle \in A^D_{(\forall y)\psi(x,y)}(d)$. By definition of $A^D_{(\forall y)\psi(d,y)}$ the latter shows $e = D$. □

Corollary 6.14. *Let $S^D_\varphi = (C^D_\varphi, F^D_\varphi, A^D_\varphi, N^D_\varphi, c^D_\varphi)$ be the canonical search problem for a formula $\varphi \in \mathrm{s}\Sigma^{\mathrm{b}}_\infty$, and D an s.u.b. for φ. Then, the following are provable in $\overline{\mathrm{BASIC}}$:*

1. *If $\mathrm{rk}(\varphi) \geq 2$, $N_\varphi(\mathfrak{s}) = \mathfrak{s}$, and either $\mathrm{tp}(\varphi) = \bigvee$ and $\mathfrak{s} \in A_\varphi$, or $\mathrm{tp}(\varphi) = \bigwedge$ and $\mathfrak{s} \notin A_\varphi$, then $\mathfrak{s}_2 = D$.*
2. *If $\mathrm{rk}(\varphi) \geq 1$, $N_\varphi(\mathfrak{s}) = \mathfrak{s}$, and either $\mathrm{tp}(\varphi) = \bigvee$ and $\mathfrak{s} \notin A_\varphi$, or $\mathrm{tp}(\varphi) = \bigwedge$ and $\mathfrak{s} \in A_\varphi$, then $\mathfrak{s}_1 = D$.*

Proof. For both 1. and 2., it is enough to consider the case $\mathrm{tp}(\varphi) = \bigvee$ as the "either…or" cases are equivalent due to the definition of A_φ. For 1. we observe that the definition of $\mathfrak{s} \in A_\varphi$ implies $\mathfrak{s}_1 < D$. Thus, $\mathfrak{s}_2 = D$ by the previous Proposition. In case 2. the definition of $\mathfrak{s} \notin A_\varphi$ implies $\mathfrak{s}_1 \not< D$, hence $\mathfrak{s}_1 = D$. □

The next proposition validates that canonical search problems correctly determine truth.

Proposition 6.15. *Let $\varphi \in \mathrm{s}\Sigma^{\mathrm{b}}_\infty$, and let $S_\varphi = (C_\varphi, F_\varphi, A_\varphi, N_\varphi, c_\varphi)$ be the canonical search problem for φ. The following is provable in S^1_2:*

$$N_\varphi(\mathfrak{s}) = \mathfrak{s} \ \wedge\ \mathfrak{s} \in F_\varphi \quad \Rightarrow \quad (\varphi \quad \Leftrightarrow \quad \mathfrak{s} \in A_\varphi)$$

Proof. The proof is by induction on the rank of φ. It is enough to consider φ of the form $(\exists x)\psi(x)$, because the assertion is trivial for φ of rank 0, and for φ of the form $(\forall x)\psi(x)$ we can use $N_\varphi = N_{\neg\varphi}$, $F_\varphi = F_{\neg\varphi}$, and $A_\varphi = C_\varphi \setminus A_{\neg\varphi}$.

We argue in S^1_2. Let $\mathfrak{s} = \langle d \,|\, \mathfrak{s}' \rangle$ and assume $N_\varphi(\mathfrak{s}) = \mathfrak{s} \in F_\varphi$. Assume first $d < D$, hence $\mathfrak{s} \in A_\varphi$. We will show $\psi(d)$, which implies φ. By definition of N_φ we have $N_{\psi x}(d, \mathfrak{s}') = \mathfrak{s}'$ and $\mathfrak{s}' \in A_{\psi x}(d)$. The definition of F_φ shows $\mathfrak{s}' \in F_{\psi x}(d)$. If $\mathrm{rk}(\varphi) = 1$ we have $\mathfrak{s}' = \langle\,\rangle$. Thus, $\mathfrak{s}' \in A_{\psi x}(d)$ implies $\langle\,\rangle \in A_{\psi(d)}$, hence $\psi(d)$. For $\mathrm{rk}(\varphi) > 1$ we obtain by induction hypothesis $\psi(d)$ iff $\mathfrak{s}'_1 = D$. As $\mathfrak{s}_1 = d < D$, Proposition 6.13 shows $\mathfrak{s}'_1 = \mathfrak{s}_2 = D$. Hence $\psi(d)$.

Now assume $d = D$, hence $\mathfrak{s} \notin A_\varphi$. We have $(\forall x < D)\neg\psi(x)$ by definition of F_φ. As D is s.u.b. for φ, the latter shows $(\forall x)\neg\psi(x)$ (this is the only place where we need S^1_2). Hence $\neg\varphi$. □

The final proposition states that canonical search problems have the properties of search problems.

Proposition 6.16. *Let $S_\varphi = (C_\varphi, F_\varphi, A_\varphi, N_\varphi, c_\varphi)$ be the canonical search problem for a formula $\varphi \in \mathrm{s}\Sigma^{\mathrm{b}}_\infty$ of rank k. The following can be proven in S^1_2.*

1. *$0^k \in F_\varphi$.*

2. *If* $\mathfrak{s} \in F_\varphi$*, then* $N_\varphi(\mathfrak{s}) \in F_\varphi$

3. *If* $N_\varphi(\mathfrak{s}) = \mathfrak{s}'$ *and* $\mathfrak{s} \neq \mathfrak{s}'$*, then* $c_\varphi(\mathfrak{s}') < c_\varphi(\mathfrak{s})$*.*

Proof. The first and third assertion follow immediately from the definitions, and can be proven already in $\overline{\mathrm{BASIC}}$. The proof of the second assertion is by induction on the rank of φ. The non-trivial cases are that φ is of the form $(\exists x)\psi(x)$, and that $\mathfrak{s} = \langle d \,|\, \mathfrak{s}' \rangle$ and $N_\varphi(\mathfrak{s}) \neq \mathfrak{s}$. If $N_\varphi(\mathfrak{s}) = \langle d \,|\, N_{\psi x}(d, \mathfrak{s}') \rangle$ the assertion follows immediately from induction hypothesis. In case $N_\varphi(\mathfrak{s}) = \left\langle d+1 \,|\, 0^k \right\rangle$ the assertion follows using Proposition 6.15 to ensure $(\forall x{<}d{+}1)\psi(x)$. □

7 Search problems defined by proof notations

We are now ready to put things together: We first define a general local search problem based on proof notations which will be used in Subsection 7.2 to provide the characterisation of $\Sigma^{\mathrm{b}}_{\ell+1}$-definable search problems in T^{k+1}_2 in terms of Π^{b}_k-PLS problems with Σ^{b}_ℓ-goals.

7.1 Parameterised Local Search Problems based on Proof Notations

Let us start by describing the idea for computing witnesses using proof trees. Assume we have a T^{k+1}_2-proof of a formula $(\exists y)\varphi(y)$ in $\Sigma^{\mathrm{b}}_{\ell+1}$ and we want to compute an n such that $\varphi(n)$ is true — in case we are interested in definable search problems, such a situation is obtained from a proof of $(\forall x)(\exists y)\varphi(x, y)$ by inverting the universal quantifier to some $a \in \mathbb{N}$. Assume further, we have applied some proof theoretical transformations to obtain a BA^∞ derivation d_0 of $(\exists y)\varphi(y)$ with $\mathrm{s}\Sigma^{\mathrm{b}}_0\text{-crk}(d_0) \leq k$. Then we can define a path through d_0, represented by subderivations $d_1, d_2, d_3 \ldots$, such that

- $d_{j+1} = d_j(\iota)$ for some $\iota \in |\,\mathrm{last}(d_j)|$
- $\Gamma(d_j) = (\exists y)\varphi(y), \Gamma_j$ where all formulae $A \in \Gamma_j$ are false and in $\mathrm{s}\Sigma^{\mathrm{b}}_k \cup \mathrm{s}\Pi^{\mathrm{b}}_k$.

Such a path must be finite as $\mathrm{hgt}(d_j)$ is strictly decreasing. Say it ends with some d_ℓ. In this situation we must have that $\mathrm{last}(d_\ell) = \bigvee^k_{(\exists y)\varphi(y)}$ and that $\varphi(k)$ is true. Hence we found our witness.

The path which we have just described can be viewed as the canonical path through a related local search problem. Before explaining this, let us fix the notion of a local search problem.

Definition 7.1. *An instance of a local search problem* consists of a set F of possible solutions, a goal set G which is a subset of F, an initial value $d \in F$, a cost function $c\colon F \to \mathbb{N}$, and a neighbourhood function $N\colon F \to F$ which satisfy that $c(N(d)) < c(d)$ if $N(d) \neq d$, and that $d \in G$ iff $d \in F$ and $N(d) = d$. A solution to a local search problem, called a *local optimum*, is any $d \in G$.

Observe that the ingredients of a local search problem guarantee the existence of a local optimum, by starting with the initial value and iterating the neighbourhood function (this defines the *canonical path through the search problem*.)

Now we define a local search problem whose canonical path is the one described above. The set F of possible solutions is defined as the set of all BA^∞-derivations d which have the properties that $\mathrm{s}\Sigma^{\mathrm{b}}_0\text{-crk}(d_0) \leq k$, and that all formulae $A \in \Gamma(d) \setminus \{(\exists y)\varphi(y)\}$ are false and in $\mathrm{s}\Sigma^{\mathrm{b}}_k \cup \mathrm{s}\Pi^{\mathrm{b}}_k$. The cost of a possible solution $d \in F$ is given by the height $\mathrm{hgt}(d)$ of the proof tree d. We have already fixed some initial value $d_0 \in F$. The neighbourhood function $N\colon \mathrm{BA}^\infty \to \mathrm{BA}^\infty$ is defined by case distinction on the shape of $\mathrm{last}(d)$ for $d \in F$:

- $\mathrm{last}(d) = \mathrm{Ax}_A$ cannot occur as all atomic formulae in $\Gamma(d)$ are false by definition of F.
- $\mathrm{last}(d) = \bigwedge_{A_0 \,\wedge\, A_1}$, then $A_0 \;\wedge\; A_1$ must be false, hence some of A_0, A_1 must be false. Let $N(d) := d(0)$ if A_0 is false, and $d(1)$ otherwise.
- $\mathrm{last}(d) = \bigvee^k_{A_0 \,\vee\, A_1}$, then $A_0 \;\vee\; A_1$ must be false, hence both A_0, A_1 must be false. Let $N(d) := d(0)$.
- $\mathrm{last}(d) = \bigwedge_{(\forall x)A(x)}$. As $(\forall x)A(x)$ is false there is some i such that $A(i)$ is false. Let $N(d) := d(i)$.
- $\mathrm{last}(d) = \bigvee^k_{(\exists x)A(x)}$. If $(\exists x)A(x)$ is different from $(\exists y)\varphi(y)$ then $(\exists x)A(x)$ must be false; let $N(d) := d(0)$. Otherwise, if $\varphi(k)$ is false let $N(d) = d(0)$, and if it is true let $N(d) = d$. Observe that in the very last case we found our witness.
- $\mathrm{last}(d) = \mathrm{Cut}_C$. If C is false let $N(d) := d(0)$, otherwise let $N(d) := d(1)$.

Obviously, this defines a local search problem according to Definition 7.1. As remarked above, a local optimal solution to the search problem allows us to determine a witness.

The previous description covers the main idea for defining search problem via proof notations. It is not exactly the version we are looking for, as we want to have neighbourhood functions which are polynomial time computable, but the one that we describe above has to decide $\mathrm{s}\Sigma^{\mathrm{b}}_{k-1}$-formulas (in case of a cut) and maintain in some way the promise that the endsequent of elements in F consists of false formulas besides $(\exists y)\varphi(y)$. The adjustment we have to make is to incorporate the canonical search problems for deciding formulas from the previous section, instead of deciding them. We also have to store promised witnesses for false $\mathrm{s}\Pi^{\mathrm{b}}_{k}$ formulas in the endsequent of derivations, in order to obtain the optimal complexity for the set of feasible solutions, which is $\mathrm{s}\Pi^{\mathrm{b}}_{k}$. We do this by extending the set of possible solutions in the forthcoming Definition 7.3 to triples of the form $\langle d, f, \mathfrak{s}\rangle$, where d denotes a BA^{∞}-derivations as above, f stores witnesses of $\forall$ quantifiers, and $\mathfrak{s}$ is a position in a potential canonical search problems for deciding some formula related to the last inference of d.

In the next definition we fix some canonical choice function for the outermost quantifier of a sharply bounded formula. This is followed by the formal definition of parameterised local search problems, given as the adjustment of the local search problem described above.

Definition 7.2. Let ϵ denote the following choice function: For $\psi \in \mathrm{s}\Pi^{\mathrm{b}}_{0}$, let $\epsilon(\psi) = j$ for the smallest j such that $\psi[j]$ is false, and let $\epsilon(\psi) = 0$ if such a j cannot be found (including that $\psi \notin \mathrm{s}\Pi^{\mathrm{b}}_{0}$ and $\psi[j]$ is not defined etc).

Definition 7.3. We define a local search problem L which is parameterised by

- *complexity levels* ℓ, k with $0 \le \ell \le k$, denoting the formula classes $\mathrm{s}\Sigma^{\mathrm{b}}_{\ell}$ and $\mathrm{s}\Sigma^{\mathrm{b}}_{k}$,
- a $\mathrm{BA}^{\star}$-derivation $\bar{h}$ which is used to define an *initial value function* $h_{\bullet} \colon \mathbb{N} \to \mathcal{CH}_{\mathrm{BA}}$, mapping $a \mapsto h_a := \mathsf{E}\bar{h}(\underline{a}/x)$,
- a formula $(\exists y)\varphi(x, y) \in \mathrm{s}\Sigma^{\mathrm{b}}_{\ell+1}$,

such that S^1_2 proves, for $a \in \mathbb{N}$,

- $\Gamma(h_a) \subseteq \{(\exists y)\varphi(\underline{a}, y)\}$,
- $\mathrm{s}\Sigma^{\mathrm{b}}_{0}$-crk$(h_a) \le k$,
- $\mathrm{o}(h_a) = 2^{|a|^{O(1)}}$,
- $\vartheta(h_a) = |a|^{O(1)}$.

We denote such a parametrisation by $L = \langle \ell, k, \bar{h}, (\exists y)\varphi(x, y)\rangle$.

An instance of L is given by $a \in \mathbb{N}$ which defines the following functions and relations of a local search problem:

- Let Φ be $\mathrm{deco}(\bar{h})$ together with the closure of $\mathrm{deco}(\bar{h}) \cap \Delta_0$ under negation and taking subformulas.
- $D_a := \mathrm{bd}(h_a) + 1$ defines a strict upper bound for all formulas in $\Phi_{\max(a,\mathrm{bd}(h_a))}$ in the sense of Definition 6.6.
- The (finite) set of *potential configurations* $\widetilde{C}(a)$ consists of those pairs (h, f) of $h \in \mathcal{CH}_{\mathrm{BA}}$ and $f\colon A \to \{0, \ldots, D_a - 1\}$ for some finite subset A of $\mathcal{F}_{\mathrm{BA}}$, which satisfy:
  1. $\Gamma(h) \setminus \{(\exists y)\varphi(\underline{a}, y)\} \subset \mathrm{s}\Sigma^{\mathrm{b}}_k \cup \mathrm{s}\Pi^{\mathrm{b}}_k$,
  2. $\mathrm{dom}\, f$ consists of all $\psi \in \Gamma(h)$ with $\mathrm{tp}(\psi) = \bigwedge$ and $\psi \notin \mathrm{s}\Pi^{\mathrm{b}}_0$,
  3. $\mathrm{s}\Sigma^{\mathrm{b}}_0$-$\mathrm{crk}(h) \leq k$,
  4. $\mathrm{o}(h) \leq \mathrm{o}(h_a)$,
  5. $\mathrm{bd}(h) \leq \mathrm{bd}(h_a)$,
  6. $\vartheta(h) \leq \vartheta(h_a)$,
  7. $\mathrm{deco}(h) \subseteq \Phi_{\max(a,\mathrm{bd}(h_a))}$.
- The set of configurations is given by

$$C(a) \quad := \quad \{d \colon d < D_a\} \cup \left\{\langle h, f, \mathfrak{s}\rangle : (h, f) \in \widetilde{C}(a) \text{ and } \mathfrak{s} \in C^{k,D_a}\right\} .$$

- The *initial value function* is given by $i(a) := \left\langle h_a, \varnothing, 0^k \right\rangle$.
- The *cost function* is defined as

$$c(a, \langle h, f, \mathfrak{s}\rangle) := \mathrm{o}(h) \cdot (D_a + 1)^k + c(\mathfrak{s})$$

  and

$$c(a, d) := 0$$

  for $d < D_a$.
- The *neighbourhood function* is defined by case distinction as follows:
  for $d < D_a$ let $N(a, d) := d$;
  for $\mathrm{tp}(h) = \mathrm{Ax}_\psi$ let $N(a, \langle h, f, \mathfrak{s}\rangle) := \langle h, f, \mathfrak{s}\rangle$;
  for $\mathrm{tp}(h) = \mathrm{Rep}$ let $N(a, \langle h, f, \mathfrak{s}\rangle) := \left\langle h[0], f^r, 0^k \right\rangle$, where f^r denotes the restriction of f to $\Gamma(h[0])$ — similar in future cases;
  for $\mathrm{tp}(h) = \bigwedge_\psi$ let

$$N(a, \langle h, f, \mathfrak{s}\rangle) := \begin{cases} \left\langle h[f(\psi)], f^r, 0^k \right\rangle & \text{if } \psi \notin \mathrm{s}\Pi^{\mathrm{b}}_0, \\ \left\langle h[\epsilon(\psi)], f^r, 0^k \right\rangle & \text{if } \psi \in \mathrm{s}\Pi^{\mathrm{b}}_0 \ , \end{cases}$$

for $\mathrm{tp}(h) = \bigvee_{\psi}^{i}$ let $N(a, \langle h, f, \mathfrak{s}\rangle)$ be defined as

$$\begin{cases} \langle h[0], f^r, 0^k \rangle & \text{if } \psi \in \mathrm{s}\Sigma_0^{\mathrm{b}}, \\ \langle h, f, N_{\psi[i]}(\mathfrak{s}) \rangle & \text{if } \psi \notin \mathrm{s}\Sigma_0^{\mathrm{b}},\ N_{\psi[i]}(\mathfrak{s}) \neq \mathfrak{s}, \\ \langle h[0], f', 0^k \rangle & \text{if } \psi \notin \mathrm{s}\Sigma_0^{\mathrm{b}},\ N_{\psi[i]}(\mathfrak{s}) = \mathfrak{s},\ \mathfrak{s} \notin A_{\psi[i]} \\ & \quad \text{and } f' = (f \cup \{\psi[i] \mapsto \mathfrak{s}_1\})^r \text{ if } \psi \notin \mathrm{s}\Sigma_1^{\mathrm{b}} \\ & \quad \text{or } f' = f^r \text{ if } \psi \in \mathrm{s}\Sigma_1^{\mathrm{b}}, \\ \langle h, f, \mathfrak{s} \rangle & \text{if } \psi \notin \mathrm{s}\Sigma_0^{\mathrm{b}},\ \psi \neq (\exists y)\varphi(\underline{a}, y), \\ & \quad N_{\psi[i]}(\mathfrak{s}) = \mathfrak{s},\ \mathfrak{s} \in A_{\psi[i]}, \\ i & \text{if } \psi = (\exists y)\varphi(\underline{a}, y),\ N_{\varphi(\underline{a},\underline{i})}(\mathfrak{s}) = \mathfrak{s},\ \mathfrak{s} \in A_{\varphi(\underline{a},\underline{i})} \end{cases},$$

for $\mathrm{tp}(h) = \mathrm{Cut}_{\psi}$ let $N(a, \langle h, f, \mathfrak{s}\rangle)$ be defined as

$$\begin{cases} \langle h, f, N_{\psi}(\mathfrak{s}) \rangle & \text{if } N_{\psi}(\mathfrak{s}) \neq \mathfrak{s}, \\ \langle h[1], f^r, 0^k \rangle & \text{if } N_{\psi}(\mathfrak{s}) = \mathfrak{s},\ \mathfrak{s} \in A_{\psi}, \\ \langle h[0], f', 0^k \rangle & \text{if } N_{\psi}(\mathfrak{s}) = \mathfrak{s},\ \mathfrak{s} \notin A_{\psi} \\ & \quad \text{and } f' = (f \cup \{\psi \mapsto \mathfrak{s}_1\})^r \text{ if } \psi \notin \mathrm{s}\Pi_0^{\mathrm{b}} \\ & \quad \text{or } f' = f^r \text{ if } \psi \in \mathrm{s}\Pi_0^{\mathrm{b}} \end{cases}.$$

- The set of *feasible solutions* $F(a)$ is given by those $\langle h, f, \mathfrak{s}\rangle$ which satisfy
  - $\langle h, f, \mathfrak{s}\rangle \in C(a)$ and $\mathrm{tp}(h) \neq \mathrm{Ax}_{\psi}$;
  - for all $\psi \in \Pi := \mathrm{dom}(f)$ we have that $\psi[f(\psi)]$ is false;
  - for $\psi \in \Sigma := \Gamma(h) \setminus (\{(\exists y)\varphi(\underline{a}, y)\} \cup \Pi)$ we have that ψ is false;
  - $\mathrm{tp}(h) = \mathrm{Cut}_{\psi}$ implies $\mathfrak{s} \in F_{\psi}$;
  - $\mathrm{tp}(h) = \bigvee_{\psi}^{i}$ implies $\mathfrak{s} \in F_{\psi[i]}$;

  together with those $d < D_a$ such that $\varphi(a, d)$ holds.
- The *goal set* $G(a) := \{d < D_a \colon \varphi(a, d)\} \subset F(a)$.

We will now argue that the relations and functions defined above define a Π_k^{b}-PLS problem with Π_ℓ^{b}-goal according to Definition 3.2. One of the main considerations for this is to see that the computational complexity of the involved relations and functions fall into the right classes, in particular, that the set of configurations and the neighbourhood function are polynomial time computable. This is not difficult to see once we understood how notations for derivations are coded: any $h \in \mathcal{CH}_{\mathrm{BA}}$ is a term of inference symbols, and each inference symbol is given by its decoration consisting of formulas and terms and numbers — the formulas and terms have to come from Φ, and the numbers are bounded by $\max(a, \mathrm{bd}(h_a))$. Thus, a natural feasible Gödel numbering of such terms, as defined in [Bus86], will give us a suitable set of codes on which all necessary functions are easy to compute, as they all are either performing syntactic checks according to inference symbols and their decoration, or evaluating (in the case of feasible solutions) formulas in Φ (which is a *finite* set) under a numerical substitution.

Proposition 7.4. *The local search problem L from Definition 7.3, parameterised by $\langle \Phi, \ell, k, h, (\exists y)\varphi(x,y)\rangle$, provides a Π^{b}_k-PLS problem with Π^{b}_ℓ-goal according to Definition 3.2.*

Proof. As shown in [AB09] the functions $a \mapsto i(a) = h_a$, $a \mapsto \mathrm{bd}(h_a)$, $a \mapsto \mathrm{o}(h_a)$, $a \mapsto \vartheta(h_a)$, and $a \mapsto \mathrm{deco}(h_a)$ are polynomial time computable. Furthermore, the relations $\mathcal{CH}_{\mathrm{BA}}$, $\mathrm{s}\Sigma^{\mathrm{b}}_0$-crk$(h) \le k$, $\mathrm{bd}(h) \le m$ and $\mathrm{deco}(h) \subseteq \Phi_m$ are polynomial time computable, and once $\mathrm{bd}(h) \le m$ is established we also can compute $\mathrm{o}(h) \le m'$ and then $\mathrm{o}(h)$ in polynomial time. Hence $c \in \mathrm{FP}$. Also, the functions $\mathrm{tp}(h)$ and $h[i]$ are polynomial time computable on $\mathcal{CH}_{\mathrm{BA}}$. Using Observation 6.12, this shows that N is polynomial time computable, because the case distinction which defines N depends only on essentially finitely many N_ψ: Each such ψ is obtained from a formula in Φ (which is a finite set) by substituting constants for free variables.

To check that $F \in \Pi^{\mathrm{b}}_k$ we look at the critical cases — here we use, similar to the case above, that the definition of F depends essentially only on finitely many N_ψ. "$d \in F(a)$", for $d < D_a$, is a Π^{b}_l-property. The definition of "$(h, f, \mathfrak{s}) \in F(a)$" has three critical entries: Observe that $\Sigma \cup \Pi \subseteq \Gamma(h) \subseteq \mathrm{deco}(h) \subseteq \Phi_{\mathrm{bd}(h_a)}$, hence for $\psi \in \Pi$ the condition "$\psi[f(\psi)]$ is false" is a Π^{b}_{k-1}-property, and for $\psi \in \Sigma$ the condition "ψ is false" is a Π^{b}_k-property; the condition "$\mathfrak{s} \in F_\psi$" for ψ of rank $\le k$ is Π^{b}_k according to Observation 6.12.

That "$s \in G(a)$" is in Π^{b}_ℓ is obvious by definition.

So it remains to show that the properties (3.1)-(3.5) of Definition 3.2 do hold. For (3.1), $(\forall x, s)(s \in F(x) \;\rightarrow\; |s| \le d(|x|))$, we observe that if $(h, f) \in \widetilde{C}(a)$, then h is a term built up from inference symbols, the length of the term, i.e. the number of inference symbols, is $\vartheta(h) \le \vartheta(h_a) = |a|^{O(1)}$, and each occurring inference symbol is decorated with expressions from $\mathrm{deco}(h) \subseteq \Phi_{\max(a,\mathrm{bd}(h_a))}$ and $|\,\mathrm{bd}(h_a)| = |a|^{O(1)}$. Thus, the polynomial bound d can be found assuming a feasible Gödel numbering as in [Bus86]. Property (3.2), $(\forall x)(i(x) \in F(x))$, is obvious. The last one, (3.5),

$$(\forall x, s)(s \in G(x) \;\leftrightarrow\; (N(x,s) = s \;\wedge\; s \in F(x)))$$

also follows from the definition. For this, observe that for "←" the premise of the implication $N(x,s) = s \ \wedge \ s \in F(x)$ implies that s cannot be of the form $\langle h, f, \mathfrak{s}\rangle$: Assume it is, then either $\mathrm{tp}(h) = \mathrm{Ax}_\psi$ which would imply $\psi \in \Gamma(h)$ and ψ true, or $\mathrm{tp}(h) = \bigvee_\psi^i$, $\psi \neq (\exists y)\varphi(\underline{a}, y)$, $N_{\psi[i]}(\mathfrak{s}) = \mathfrak{s}$, $\mathfrak{s} \in A_{\psi[i]}$, and $s \in F(x)$ implies $\mathfrak{s} \in F_{\psi[i]}$, thus Proposition 6.15 shows $\psi[i]$, hence ψ, is true; both times we get a contradiction to the fact implied by $s \in F(a)$ that all formulas in $\Gamma(h) \setminus \{(\exists y)\varphi(\underline{a}, y)\}$ are false.

Property (3.3)

$$(\forall x, s)(s \in F(x) \ \rightarrow \ N(x,s) \in F(x))$$

follows by case distinction according to the definition $N(x,s)$, using the corresponding properties for canonical search problems as shown in Proposition 6.16. For example, consider the case that $s = \langle h, f, \mathfrak{s}\rangle \in F(x)$ with $\mathrm{tp}(h) = \bigwedge_\psi$ and $\psi \notin \mathrm{s}\Pi_0^\mathrm{b}$. Then $N(a,s) = \left\langle h[f(\psi)], f^r, 0^k \right\rangle$ and we have to show that $(h[f(\psi)], f^r) \in \widetilde{C}(a)$. Let $j = f(\psi)$, then $h[j] \vdash_\approx \Gamma(h), \psi[j]$ thus obviously $\Gamma(h[j]) \subset \mathrm{s}\Sigma_k^\mathrm{b} \cup \mathrm{s}\Pi_k^\mathrm{b}$. As $\mathrm{tp}(\psi) = \bigwedge$ and $\psi \in \mathrm{s}\Sigma_\infty^\mathrm{b}$, it follows that $\mathrm{tp}(\psi[j]) \neq \bigwedge$, thus $\mathrm{dom}\, f^r$ satisfies the property under 2. We compute $\mathrm{s}\Sigma_0^\mathrm{b}$-crk$(h[j]) \leq \mathrm{s}\Sigma_0^\mathrm{b}$-crk$(h) \leq k$, $\mathrm{o}(h[j]) < \mathrm{o}(h) \leq \mathrm{o}(h_a)$, $\mathrm{bd}(h[j]) \leq \mathrm{bd}(h) \leq \mathrm{bd}(h_a)$ by Lemma 5.13, 1., $\vartheta(h[j]) \leq \vartheta(h) \leq \vartheta(h_a)$ by Theorem 4.16, and that $\mathrm{deco}(h) \subseteq \Phi_{\max(a,\mathrm{bd}(h_a))}$, $j = f(\psi) \leq \mathrm{bd}(h_a)$ and $\mathrm{bd}(h) \leq \mathrm{bd}(h_a)$ imply $\mathrm{deco}(h[j]) \subseteq (\Phi_{\max(a,\mathrm{bd}(h_a))})_{\max(\mathrm{bd}(h_a),\mathrm{bd}(h))} = \Phi_{\max(a,\mathrm{bd}(h_a))}$ by Lemma 5.17, 4.

Other interesting cases occur when $s = \langle h, f, \mathfrak{s}\rangle \in F(x)$ with $\mathrm{tp}(h) = \bigvee_\psi^i$, $\psi \notin \mathrm{s}\Sigma_0^\mathrm{b}$ and $N_{\psi[i]}(\mathfrak{s}) = \mathfrak{s}$. If $\mathfrak{s} \notin A_{\psi[i]}$ and $\psi \notin \mathrm{s}\Sigma_1^\mathrm{b}$, then $N(a,s) = \left\langle h[0], f', 0^k \right\rangle$ and $f' = (f \cup \{\psi[i] \mapsto \mathfrak{s}_1\})^r$. The condition $\langle h[0], f'\rangle \in \widetilde{C}(a)$ can be shown as before. If $\psi[i] \in \mathrm{dom}(f')$ we also have to show that $\psi[i][\mathfrak{s}_1]$ is false. $\mathfrak{s}$ can be written as $\langle d \,|\, \mathfrak{s}'\rangle$ because $\mathrm{rk}(\psi) \geq 2$. As $s \in F(x)$ we have $\mathfrak{s} \in F_{\psi[i]}$ by definition of $F(x)$, which implies $\mathfrak{s}' \in F_{\psi[i][d]}$ by definition of $F_{\psi[i]}$. By assumptions we also have $N_{\psi[i]}(\mathfrak{s}) = \mathfrak{s}$ and $\mathfrak{s} \notin A_{\psi[i]}$. As $\mathrm{tp}(\psi[i]) = \bigwedge$, $\mathfrak{s} \notin A_{\psi[i]}$ shows $d < D_a$, hence both $N_{\psi[i]}(\mathfrak{s}) = \mathfrak{s}$ and $\mathfrak{s} \notin A_{\psi[i]}$ together with the definition of $N_{\psi[i]}$ show $N_{\psi[i][d]}(\mathfrak{s}') = \mathfrak{s}'$ and $\mathfrak{s}' \notin A_{\psi[i][d]}$. Now we can conclude using Proposition 6.15 that $\psi[i][d]$ is false.

If $\mathfrak{s} \in A_{\psi[i]}$ and $\psi \neq (\exists y)\varphi(\underline{a}, y)$, then $N(x,s) = s$ and there is nothing to show.

Finally, if $\mathfrak{s} \in A_{\psi[i]}$ and $\psi = (\exists y)\varphi(\underline{a}, y)$, then $N(x,s) = i$ and we have to show that $i < D_a$ and that $\varphi(\underline{a}, \underline{i})$ is true. Lemma 5.13, 2., shows that $i < \mathrm{bd}(h)$, thus $i < \mathrm{bd}(h_a) \leq D_a$. Again, $s \in F(x)$ implies $\mathfrak{s} \in F_{\psi[i]}$. Thus the assumptions $N_{\psi[i]}(\mathfrak{s}) = \mathfrak{s}$ and $\mathfrak{s} \in A_{\psi[i]}$ together with Proposition 6.15 show that $\psi[i]$ is true, that is $\varphi(\underline{a}, \underline{i})$ is true.

Finally, Property (3.4)

$$(\forall x, s)(N(x,s) = s \ \vee \ c(x, N(x,s)) < c(x,s))$$

also follows immediately from the definitions. Because, for $s = (h, f, \mathfrak{s})$ with $N(x,s) = (h', f', \mathfrak{s}') \neq s$, either $h' = h[j]$ for some j, and then $\mathrm{o}(h') < \mathrm{o}(h)$, or $h' = h$ and $\mathfrak{s}' = N_\psi(\mathfrak{s}) \neq \mathfrak{s}$ and then $c(\mathfrak{s}') < c(\mathfrak{s})$ using Proposition 6.16. □

7.2 $\Sigma^{\mathrm{b}}_{\ell+1}$-definable search problems in T^{k+1}_2 for $\ell \leq k$

Let $0 \leq \ell \leq k$ and assume that $\mathrm{T}^{k+1}_2 \vdash (\forall x)(\exists y)\varphi(x,y)$ with $(\exists y)\varphi(x,y) \in \mathrm{s}\Sigma^{\mathrm{b}}_{\ell+1}$, $\varphi \in \mathrm{s}\Pi^{\mathrm{b}}_{\ell}$. Inverting the $(\forall x)$ quantifier we also obtain $\mathrm{T}^{k+1}_2 \vdash (\exists y)\varphi(x,y)$. By partial cut-elimination, Theorem 5.11, we obtain some $\mathrm{BA}^\star$-derivation h such that $\mathrm{FV}(h) \subseteq \{x\}$, $\Gamma(h) = \{(\exists y)\varphi(x,y)\}$, $\mathrm{s}\Sigma^{\mathrm{b}}_{k+1}$-gcrk$(h) = 0$, and $\mathrm{o}(h(\underline{a}/x)) = |a|^{O(1)}$.

Let Φ be $\mathrm{deco}(h)$ together with the closure of $\mathrm{deco}(h) \cap \Delta_0$ under negation and taking subformulas. Then $L = \langle \Phi, \ell, k, h, (\exists y)\varphi(x,y)\rangle$ defines a local search problem according to Definition 7.3, because the following are provable in S^1_2:

- $\Gamma(h_a) = \Gamma(\mathsf{E}h(\underline{a}/x)) = \Gamma(h(\underline{a}/x)) \subseteq \Gamma(h)(\underline{a}/x) = \{(\exists y)\varphi(\underline{a}, y)\}$, where we used Lemma 5.8 for "$\subseteq$";
- $$\begin{aligned}\mathrm{s}\Sigma^{\mathrm{b}}_0\text{-crk}(h_a) &= \mathrm{s}\Sigma^{\mathrm{b}}_0\text{-crk}(\mathsf{E}h(\underline{a}/x)) = \mathrm{s}\Sigma^{\mathrm{b}}_0\text{-crk}(h(\underline{a}/x)) \dot{-} 1\\ &= \mathrm{s}\Sigma^{\mathrm{b}}_0\text{-gcrk}(h(\underline{a}/x)) \dot{-} 1 = \mathrm{s}\Sigma^{\mathrm{b}}_0\text{-gcrk}(h) \dot{-} 1\\ &\leq (\mathrm{s}\Sigma^{\mathrm{b}}_{k+1}\text{-gcrk}(h) + k + 1) \dot{-} 1 = k \ ,\end{aligned}$$
  using the properties mentioned directly after Definition 5.5 for "$\leq$";
- $\mathrm{o}(h_a) = \mathrm{o}(\mathsf{E}h(\underline{a}/x)) = 2^{\mathrm{o}(h(\underline{a}/x))} - 1 = 2^{|a|^{O(1)}}$;
- $$\begin{aligned}\vartheta(h_a) &= \vartheta(\mathsf{E}h(\underline{a}/x)) = \mathrm{o}(h(\underline{a}/x)) \cdot (\vartheta(h(\underline{a}/x)) + 2)\\ &= |a|^{O(1)} \cdot (|h(\underline{a}/x)| + 2) = |a|^{O(1)} \cdot (|h| + 2) = |a|^{O(1)} \ ;\end{aligned}$$
- $\mathrm{deco}(h_a) = \mathrm{deco}(\mathsf{E}h(\underline{a}/x)) = \mathrm{deco}(h(\underline{a}/x)) \subseteq \Phi_a$, where we have used Lemma 5.17, 2. for the last inclusion.

By Proposition 7.4, this defines a search problem in Π^{b}_k-PLS with Π^{b}_ℓ-goal. Thus we have proven Theorem 3.5, that the $\Sigma^{\mathrm{b}}_{\ell+1}$-definable total search problems in T^{k+1}_2 can be characterised by Π^{b}_k-PLS problems with Π^{b}_ℓ-goals. Together with Theorem 3.4 we obtain a full characterisation of the $\Sigma^{\mathrm{b}}_{\ell+1}$-definable total search problems in T^{k+1}_2:

Corollary 7.5. *Let $0 \leq \ell \leq k$. The $\Sigma^{\mathrm{b}}_{\ell+1}$-definable total search problems in T^{k+1}_2 are exactly characterised by Π^{b}_k-PLS problems with Π^{b}_ℓ-goals.*

8 Skolemising Search for Truth

In the remaining sections we will strengthen our results by showing that the properties (3.1)–(3.5) of the Π^b_k-PLS problems extracted from T^{k+1}_2-proofs according to Theorem 3.5 can be written in a prenex form which can be skolemised by simple polynomial time functions, provably in S^1_2.

Notation 8.1. We use α, β,... to range over formulas in Σ^b_0.

Definition 8.2 (Prenex forms)**.** ψ is called *a prenex form of* φ iff ψ has the shape $(\mathfrak{Q}\mathfrak{z})\beta$ for some $\beta \in \Sigma^b_0$, such that $\overline{\text{BASIC}} \vdash \varphi \leftrightarrow \psi$.

Definition 8.3 (Simple Skolemisation)**.** Let $(\mathfrak{Q}\mathfrak{z})\beta(x, \mathfrak{z})$ with $\beta \in \Sigma^b_0$ be a prenex form for $\varphi(x)$, where $\mathfrak{z} = [z_1, \dots, z_k]$ and $\mathfrak{Q} = [Q_1, \dots, Q_k]$. Let f be some function symbol. We say that

$$(\forall x)(\varphi(x) \rightarrow \varphi(f(x)))$$

admits simple Skolem functions iff there are polynomial time computable functions $f_1, \dots, f_k$ such that

$$(\forall x)(\forall^k \mathfrak{z})(\beta(x, t_1, \dots, t_k) \rightarrow \beta(f(x), t'_1, \dots, t'_k))$$

is provable in S^1_2, where

$$t_i := \begin{cases} z_i & \text{if } Q_i = \exists \\ f_i(x, z_1, \dots, z_i) & \text{otherwise} \end{cases}$$
$$t'_i := \begin{cases} f_i(x, z_1, \dots, z_i) & \text{if } Q_i = \exists \\ z_i & \text{otherwise} \end{cases}$$

The main result of this section will be to fix a suitable prenex form for $\mathfrak{s} \in F_\varphi$ in such a way that the *canonical* prenex form of

$$(\forall \mathfrak{s})(\mathfrak{s} \in F_\varphi \rightarrow N_\varphi(\mathfrak{s}) \in F_\varphi) \tag{8.1}$$

admits simple Skolem functions — we explain later what we mean by a canonical prenex form. In the next subsection we fix a suitable prenex form for $\mathfrak{s} \in F_\varphi$; that it enjoys the above mentioned property will be shown later in Theorem 8.5.

8.1 A suitable prenex form for $\mathfrak{s} \in F_\varphi$

Formulas have many prenex forms. We will now pick a suitable one for the formula $\mathfrak{s} \in F_\varphi$. Remember that we defined the application of the projection function p_i to formal tuples $\mathfrak{t} = [t_1, \ldots, t_k]$ as $\mathrm{p}_i(\mathfrak{t}) = [\mathrm{p}_i(t_1), \ldots, \mathrm{p}_i(t_k)]$.

Theorem 8.4. *Let φ be a strict formula of rank k, and D a s.u.b. for φ. Then there is a $\mathrm{s}\Sigma_0^{\mathrm{b}}$-formula γ_φ such that the following are provable in $\overline{\mathrm{BASIC}}$:*

1. $\mathfrak{s} \in F_\varphi \quad \Leftrightarrow \quad (\forall\exists^k \mathfrak{z})\gamma_\varphi(\mathfrak{s}, \mathfrak{z})$.
2. $(\forall \mathfrak{s})(\forall^k \mathfrak{z}^1)(\forall^k \mathfrak{z}^2)\Big(\bigwedge_{1 \le i,j \le k} \mathrm{p}_j(\mathfrak{z}_i^1) = \mathrm{p}_j(\mathfrak{z}_i^2) \ \wedge\ \gamma_\varphi(\mathfrak{s}, \mathfrak{z}^1) \ \to\ \gamma_\varphi(\mathfrak{s}, \mathfrak{z}^2)\Big)$.
3. $(\forall^k \mathfrak{z})\gamma_\varphi(0^k, \mathfrak{z})$.
4. *If $k \ge 1$ and $\varphi \equiv (\exists\forall^k \mathfrak{z})\beta(\mathfrak{z})$, then*
$$\gamma_\varphi(\mathfrak{s}, \mathfrak{z}) \ \to\ (\mathrm{p}_k(\mathfrak{z}_1) < \mathfrak{s}_1 \ \to\ \neg\beta(\mathrm{p}_k(\mathfrak{z})))$$
Here, $\mathrm{p}_k(\mathfrak{z})$ denotes $[\mathrm{p}_k(\mathfrak{z}_1), \ldots, \mathrm{p}_k(\mathfrak{z}_k)]$.
5. *If $k \ge 2$ and $\varphi \equiv (\exists\forall^k \mathfrak{z})\beta(\mathfrak{z})$, then*
$$\gamma_\varphi(\mathfrak{s}, \mathfrak{z}) \ \to\ (\mathrm{p}_{k-1}(\mathfrak{z}_1) < \mathfrak{s}_2 \ \to\ \beta(\mathfrak{s}_1, \mathrm{p}_{k-1}(\mathfrak{z}\lceil_{k-1})))$$
Observe that $\mathrm{p}_{k-1}(\mathfrak{z}\lceil_{k-1})$ denotes $[\mathrm{p}_{k-1}(\mathfrak{z}_1), \ldots, \mathrm{p}_{k-1}(\mathfrak{z}_{k-1})]$.

Proof. The definition and proof are by induction on k. If $k = 0$ let γ_φ be the formula $0 = 0$. All properties are obviously satisfied.

For $k > 0$ and $\varphi \equiv (\forall x)\beta(x)$ we define $\gamma_\varphi(\mathfrak{s}, \mathfrak{z})$ to be the same as $\gamma_{\neg\varphi}(\mathfrak{s}, \mathfrak{z})$.

For $k = 1$ and $\varphi \equiv (\exists x)\beta(x)$ we have $\langle u \rangle \in F_\varphi \equiv (\forall x{<}u)\neg\beta(x)$. Let $\gamma_\varphi(\langle u \rangle, x)$ be the formula

$$(\mathrm{p}_1(x){<}u \ \to\ \neg\beta(\mathrm{p}_1(x)))$$

Again it is easy to see that all properties are satisfied.

Although the general inductive case is for $k \ge 2$ already, we write out the cases for $k = 2$ and $k = 3$ explicitly, to make the definition of γ_φ more clear. The mentioning of "$\wedge\ 0 = 0$" in the following case is to suit the general inductive case. Let $k = 2$ and $\varphi \equiv (\exists x)(\forall y)\beta(x, y)$. Then $\langle u, v \rangle \in F_\varphi$ has the form

$$(\forall x{<}u)(\exists y)\neg\beta(x, y) \ \wedge\ (\forall y{<}v)\beta(u, y) \ .$$

Let $\gamma_\varphi(\langle u, v\rangle, x, y)$ be the formula

$$\begin{aligned} &(\mathrm{p}_2(x)<u \;\rightarrow\; \neg\beta(\mathrm{p}_2(x), \mathrm{p}_2(y))) \\ \wedge\; &(\mathrm{p}_1(x)<v \;\rightarrow\; \beta(u, \mathrm{p}_1(x))) \\ \wedge\; &0 = 0 \;\; . \end{aligned}$$

If $k = 3$ and $\varphi \equiv (\exists x)(\forall y)(\exists z)\beta(x, y, z)$ we have that $\langle u, v, w\rangle \in F_\varphi$ is of the form

$$(\forall x<u)(\exists y)(\forall z)\neg\beta(x, y, z) \;\wedge\; (\forall y<v)(\exists z)\beta(u, y, z) \;\wedge\; (\forall z<w)\neg\beta(u, v, z) \;\; .$$

Let $\gamma_\varphi(\langle u, v, w\rangle, x, y, z)$ be the formula

$$\begin{aligned} &(\mathrm{p}_3(x)<u \;\rightarrow\; \neg\beta(\mathrm{p}_3(x), \mathrm{p}_3(y), \mathrm{p}_3(z))) \\ \wedge\; &(\mathrm{p}_2(x)<v \;\rightarrow\; \beta(u, \mathrm{p}_2(x), \mathrm{p}_2(y))) \\ \wedge\; &(\mathrm{p}_1(z)<w \;\rightarrow\; \neg\beta(u, v, \mathrm{p}_1(z))) \;\; . \end{aligned}$$

For all cases considered so far it is easy to verify that the assertions 1.–6. are satisfied. We have explicitly written out case $k = 3$ to stress the dependency of quantifiers: It will be crucial for our later developments that the 3rd conjunct uses "z" and not "x" as a naive inductive continuation might suggest.

For the general inductive case we assume $\varphi \equiv (\exists x)(\forall y)\psi(x, y)$, $\psi \equiv (\exists\forall^k \mathfrak{z})\beta(x, y, \mathfrak{z})$ and $\mathrm{rk}(\psi) = k \geq 0$. Then $\langle u, v \,|\, \mathfrak{s}\rangle \in F_\varphi$ has the form

$$\begin{aligned} &(\forall x<u)(\exists y)\neg\psi(x, y) \;\wedge\; (\forall y<v)\psi(u, y) \;\wedge\; \mathfrak{s} \in F_{\psi xy}(u, v) \\ \Leftrightarrow\quad &\qquad (\forall x)(\exists y)(\forall\exists^k \mathfrak{z})(x<u \;\rightarrow\; \neg\beta(x, y, \mathfrak{z})) \\ &\quad\wedge\;\; (\forall y)(\exists\forall^k \mathfrak{z})(y<v \;\rightarrow\; \beta(u, y, \mathfrak{z})) \\ &\quad\wedge\;\; ((\forall\exists^k \mathfrak{z})\gamma_{\psi xy}(u, v, \mathfrak{s}, \mathfrak{z}) \\ \Leftrightarrow\quad &(\forall x)(\exists y)(\forall\exists^k \mathfrak{z})\gamma_\varphi(\langle u, v \,|\, \mathfrak{s}\rangle, x, y, \mathfrak{z}) \end{aligned}$$

where we define $\gamma_\varphi(\langle u, v \,|\, \mathfrak{s}\rangle, x, y, \mathfrak{z})$ to be the formula

$$\begin{aligned} &(\mathrm{p}_{k+2}(x)<u \;\rightarrow\; \neg\beta(\mathrm{p}_{k+2}(x), \mathrm{p}_{k+2}(y), \mathrm{p}_{k+2}(\mathfrak{z}))) \\ \wedge\; &(\mathrm{p}_{k+1}(x)<v \;\rightarrow\; \beta(u, \mathrm{p}_{k+1}(x), \mathrm{p}_{k+1}(y), \mathrm{p}_{k+1}(\mathfrak{z}\lceil_{k-1}))) \\ \wedge\; &\gamma_{\psi xy}(u, v, \mathfrak{s}, \mathfrak{z}) \;\; . \end{aligned}$$

This choice of γ_φ obviously satisfies all assertions. □

8.2 A simple Skolemisation for $(\forall \mathfrak{s})(\mathfrak{s} \in F_\varphi \rightarrow N_\varphi(\mathfrak{s}) \in F_\varphi)$

Now that we have fixed prenex forms for $\mathfrak{s} \in F_\varphi$, we choose a suitable prenex form of $(\forall \mathfrak{s})(\mathfrak{s} \in F_\varphi \rightarrow N_\varphi(\mathfrak{s}) \in F_\varphi)$ in a canonical way:

$$\begin{aligned} &(\forall \mathfrak{s})(\mathfrak{s} \in F_\varphi \rightarrow N_\varphi(\mathfrak{s}) \in F_\varphi) \\ \Leftrightarrow\quad &(\forall \mathfrak{s})\Big((\forall\exists^k \mathfrak{z})\gamma_\varphi(\mathfrak{s},\mathfrak{z}) \rightarrow (\forall\exists^k \bar{\mathfrak{z}})\gamma_\varphi(N_\varphi(\mathfrak{s}),\bar{\mathfrak{z}})\Big) \\ \Leftrightarrow\quad &(\forall \mathfrak{s})(\forall \bar{\mathfrak{z}}_1)(\exists \mathfrak{z}_1)(\forall \mathfrak{z}_2)(\exists \bar{\mathfrak{z}}_2)(\forall \bar{\mathfrak{z}}_3)(\exists \mathfrak{z}_3)\cdots\Big(\gamma_\varphi(\mathfrak{s},\mathfrak{z}) \rightarrow \gamma_\varphi(N_\varphi(\mathfrak{s}),\bar{\mathfrak{z}})\Big) \end{aligned}$$

The latter is the prenex form which we fix.

Theorem 8.5. *Let φ be a strict formula of rank k, and D a s.u.b for φ. The prenex form which we fixed for $(\forall \mathfrak{s})(\mathfrak{s} \in F_\varphi \rightarrow N_\varphi(\mathfrak{s}) \in F_\varphi)$ admits simple Skolem functions.*

Proof. We have to show that there are polynomial time computable functions

$$f_1(\mathfrak{s}, z_1),\ f_2(\mathfrak{s}, z_1, z_2),\ f_3(\mathfrak{s}, z_1, z_2, z_3),\ \ldots$$

such that

$$\begin{aligned} &(\forall \mathfrak{s}, z_1, z_2, z_3, \ldots) \\ &\quad\Big(\gamma_\varphi(\mathfrak{s}, f_1(\mathfrak{s}, z_1), z_2, f_3(\mathfrak{s}, z_1, z_2, z_3), z_4, \ldots) \\ &\qquad\quad \rightarrow \gamma_\varphi(N_\varphi(\mathfrak{s}), z_1, f_2(\mathfrak{s}, z_1, z_2), z_3, f_4(\ldots, z_4), \ldots)\Big)\ . \end{aligned} \tag{8.2}$$

In the following we suppress the argument $\mathfrak{s}$ from the Skolem functions. The Skolem functions may also depend on further parameters of φ which we also do not mention. We say that *the i-th slice of $f_1(z_1)$ ($f_2(z_1, z_2)$, $f_3(z_1, z_2, z_3)$, … respectively) is chosen canonically* if $\mathrm{p}_i(f_1(z_1)) = \mathrm{p}_i(z_1)$ ($\mathrm{p}_i(f_2(z_1, z_2)) = \mathrm{p}_i(z_2)$, $\mathrm{p}_i(f_3(z_1, z_2, z_3)) = \mathrm{p}_i(z_3)$, …respectively.) Choosing the i-th slice of $f_1, f_2, f_3, \ldots$ canonically implies that

$$\begin{aligned} &\mathrm{p}_i([f_1(z_1), z_2, f_3(z_1, z_2, z_3), z_4, \ldots]) \\ &\qquad = \mathrm{p}_i([z_1, z_2, z_3, z_4, \ldots]) \\ &\qquad = \mathrm{p}_i([z_1, f_2(z_1, z_2), z_3, f_4(z_1, z_2, z_3, z_4), \ldots]) \end{aligned}$$

We now define the Skolem functions and prove (8.2) by induction on k.

If $k = 0$ there is nothing to show. If $k = 1$ and $\varphi \equiv (\exists x)\beta(x)$ we choose the first slice of f_1 canonically. Then (8.2) is equivalent to

$$\begin{aligned}
&(\forall u, x)(\gamma_\varphi(\langle u\rangle, f_1(x)) \rightarrow \gamma_\varphi(N_\varphi(\langle u\rangle), x)) \\
&\quad\Leftrightarrow \quad (\forall u, \bar{u}, x)\Big(N_\varphi(\langle u\rangle) = \langle\bar{u}\rangle \ \wedge\ (\mathrm{p}_1(x) < u \rightarrow \neg\beta(\mathrm{p}_1(x))) \\
&\qquad\qquad \rightarrow (\mathrm{p}_1(x) < \bar{u} \rightarrow \neg\beta(\mathrm{p}_1(x)))\Big)
\end{aligned}$$

The non-trivial case is when $N_\varphi(\langle u\rangle) = \langle\bar{u}\rangle$, $\bar{u} = u{+}1$ and $\mathrm{p}_1(x) = u$. By definition of N_φ this implies $\neg\beta(u)$, hence (8.2) follows.

For the inductive case we consider $\varphi \equiv (\exists x)(\forall y)\psi(x, y)$ with $\psi(x, y) \equiv (\exists\forall^k \mathfrak{z})\beta(x, y, \mathfrak{z})$ and $k \geq 0$. Then (8.2) is equivalent to

$$\begin{aligned}
&(\forall u, v, \bar{u}, \bar{v}, \mathfrak{s}, \bar{\mathfrak{s}}, z_1, z_2, \dots) \\
&\Big(N_\varphi(\langle u, v \,|\, \mathfrak{s}\rangle) = \langle\bar{u}, \bar{v} \,|\, \bar{\mathfrak{s}}\rangle \\
&\qquad \wedge\ (\mathrm{p}_{k+2}(f_1(z_1)) {<} u \\
&\qquad\qquad \rightarrow \neg\beta(\mathrm{p}_{k+2}([f_1(z_1), z_2, f_3(z_1, z_2, z_3), z_4, \dots]))) \\
&\qquad \wedge\ (\mathrm{p}_{k+1}(f_1(z_1)) {<} v \rightarrow \beta(u, \mathrm{p}_{k+1}([f_1(z_1), z_2, f_3(\dots), \dots]))) \\
&\qquad \wedge\ \gamma_{\psi xy}(u, v, \mathfrak{s}, f_3(z_1, z_2, z_3), z_4, \dots) \\
&\quad \rightarrow (\mathrm{p}_{k+2}(z_1) {<} \bar{u} \rightarrow \neg\beta(\mathrm{p}_{k+2}([z_1, f_2(z_1, z_2), z_3, f_4(\dots), \dots]))) \\
&\qquad \wedge\ (\mathrm{p}_{k+1}(z_1) {<} \bar{v} \rightarrow \beta(\bar{u}, \mathrm{p}_{k+1}([z_1, f_2(z_1, z_2), z_3, \dots]))) \\
&\qquad \wedge\ \gamma_{\psi xy}(\bar{u}, \bar{v}, \bar{\mathfrak{s}}, z_3, f_4(z_1, z_2, z_3, z_4), \dots)\Big)
\end{aligned} \tag{8.3}$$

Let $u, v, \bar{u}, \bar{v}, \mathfrak{s}, \bar{\mathfrak{s}}, z_1, z_2, z_3, \dots$ be given with $N_\varphi(\langle u, v \,|\, \mathfrak{s}\rangle) = \langle\bar{u}, \bar{v} \,|\, \bar{\mathfrak{s}}\rangle$. The possible cases for N_φ are that $N_\varphi(\langle u, v \,|\, \mathfrak{s}\rangle) = \langle u, v \,|\, \mathfrak{s}\rangle$ which is trivial, or that $N_\varphi(\langle u, v \,|\, \mathfrak{s}\rangle) \neq \langle u, v \,|\, \mathfrak{s}\rangle$, in which case we distinguish the following three sub-cases according to the definition of N_φ:

1. $N_{\psi xy}(u, v, \mathfrak{s}) = \mathfrak{s}' \neq \mathfrak{s}$, thus

$$N_\varphi(\langle u, v \,|\, \mathfrak{s}\rangle) = \langle u, v \,|\, \mathfrak{s}'\rangle \quad .$$

2. $N_{\psi xy}(u, v, \mathfrak{s}) = \mathfrak{s}$ and $\mathfrak{s} \in A_{\psi xy}(u, v)$, thus

$$N_\varphi(\langle u, v \,|\, \mathfrak{s}\rangle) = \big\langle u, v + 1 \,|\, 0^k \big\rangle \quad .$$

3. $N_{\psi xy}(u, v, \mathfrak{s}) = \mathfrak{s}$ and $\mathfrak{s} \notin A_{\psi xy}(u, v)$, thus

$$N_\varphi(\langle u, v \,|\, \mathfrak{s}\rangle) = \big\langle u + 1, 0 \,|\, 0^k \big\rangle \quad .$$

As $N_{\psi xy}$ and $A_{\psi xy}$ are polynomial time computable, and $u, v, \bar{u}, \bar{v}, \mathfrak{s}, \bar{\mathfrak{s}}$ are parameters to all Skolem functions, we can define the Skolem functions by case distinction according to the above three cases.

Case 1. We have $\bar{u} = u$, $\bar{v} = v$, $\bar{\mathfrak{s}} = \mathfrak{s}'$. By induction hypothesis there are $f_3, f_4, \ldots$ such that

$$\begin{aligned} &\gamma_{\psi xy}(u, v, \mathfrak{s}, f_3(z_1, z_2, z_3), z_4, \ldots) \\ &\quad \to \gamma_{\psi xy}(u, v, \mathfrak{s}', z_3, f_4(z_1, z_2, z_3, z_4), \ldots) \end{aligned}$$

where the functions do not yet depend on z_1, z_2. By Definition 8.4, 2. this still holds if we modify slice $k+1$ and $k+2$ of $f_3, f_4, \ldots$. We choose slices $k+1$ and $k+2$ of $f_1, f_2, f_3, f_4, \ldots$ canonically. Then (8.3) turns into

$$\begin{aligned} &(\mathrm{p}_{k+2}(z_1)<u \ \to\ \neg\beta(\mathrm{p}_{k+2}([z_1, z_2, z_3, \ldots]))) \\ &\wedge\ (\mathrm{p}_{k+1}(z_1)<v \ \to\ \beta(u, \mathrm{p}_{k+1}([z_1, z_2, z_3, \ldots]))) \\ &\wedge\ \gamma_{\psi xy}(u, v, \mathfrak{s}, f_3(z_1, z_2, z_3), z_4, \ldots) \\ \to &(\mathrm{p}_{k+2}(z_1)<u \ \to\ \neg\beta(\mathrm{p}_{k+2}([z_1, z_2, z_3, \ldots]))) \\ &\wedge\ (\mathrm{p}_{k+1}(z_1)<v \ \to\ \beta(u, \mathrm{p}_{k+1}([z_1, z_2, z_3, \ldots]))) \\ &\wedge\ \gamma_{\psi xy}(u, v, \mathfrak{s}', z_3, f_4(z_1, z_2, z_3, z_4), \ldots) \end{aligned}$$

which is obviously satisfied using the induction hypothesis.

Case 2. We have $\bar{u} = u$, $\bar{v} = v{+}1$, and $\bar{\mathfrak{s}} = 0^k$. Observe that $\gamma_{\psi xy}(u, v{+}1, 0^k, \ldots)$ is always true by Theorem 8.4, 3. We choose slice $k{+}2$ of the Skolem functions canonically. Thus, (8.3) follows from

$$\begin{aligned} &(\mathrm{p}_{k+1}(f_1(z_1)) < v \ \to\ \beta(u, \mathrm{p}_{k+1}([f_1(z_1), z_2, f_3(\ldots), \ldots]))) \\ &\wedge\ \gamma_{\psi xy}(u, v, \mathfrak{s}, f_3(z_1, z_2, z_3), z_4, \ldots) \\ \to &(\mathrm{p}_{k+1}(z_1) < v+1 \ \to\ \beta(u, \mathrm{p}_{k+1}([z_1, f_2(z_1, z_2), z_3, \ldots]))) \end{aligned} \tag{8.4}$$

If $\mathrm{p}_{k+1}(z_1) \neq v$ choose all slices of Skolem functions canonically, then (8.4) is obviously satisfied.

Now assume $\mathrm{p}_{k+1}(z_1) = v$. If $k = 0$ we choose all slices of Skolem functions canonically. Then (8.4) is equivalent to $\beta(u, v)$ which is satisfied, because we have by construction of N_φ that $\mathfrak{s} \in A_{\beta(x,y)xy}(u, v)$, which implies that $\beta(u, v)$ is true.

If $k \geq 1$, we choose Skolem functions in the following way:

$$\begin{aligned} & & \mathrm{p}_{k+1}(f_2(z_1, z_2)) &= \mathfrak{s}_1 \\ \mathrm{p}_{k-1}(f_3(z_1, z_2, z_3)) &= \mathrm{p}_{k+1}(z_3) & \mathrm{p}_{k+1}(f_4(\ldots, z_4)) &= \mathrm{p}_{k-1}(z_4) \\ \mathrm{p}_{k-1}(f_5(\ldots, z_5)) &= \mathrm{p}_{k+1}(z_5) & \mathrm{p}_{k+1}(f_6(\ldots, z_6)) &= \mathrm{p}_{k-1}(z_6) \quad \ldots \end{aligned}$$

and all other slices canonically. Assuming the antecedent of (8.4) we have to show $\beta(u, v, \mathfrak{s}_1, \mathrm{p}_{k+1}(z_3), \mathrm{p}_{k-1}(z_4), \mathrm{p}_{k+1}(z_5), \dots)$.

If $k = 1$, the definition of N_φ shows $\mathfrak{s} \in A_{(\exists z)\beta(x,y,z)}(u, v)$, which by construction implies that $\beta(u, v, \mathfrak{s}_1)$ is true.

In case $k \geq 2$ we obtain from Theorem 8.4, 5 that

$$\gamma_{\psi xy}(u, v, \mathfrak{s}, \mathfrak{t}) \;\rightarrow\; (\mathrm{p}_{k-1}(\mathfrak{t}_1) < \mathfrak{s}_2 \;\rightarrow\; \beta(u, v, \mathfrak{s}_1, \mathrm{p}_{k-1}(\mathfrak{t}\lceil_{k-1})))$$

for $\mathfrak{t} = [f_3(z_1, z_2, z_3), z_4, f_5(\dots), \dots]$. Together with the antecedent of (8.4) this implies

$$\mathrm{p}_{k-1}(f_3(z_1, z_2, z_3)) < \mathfrak{s}_2 \;\rightarrow\; \beta(u, v, \mathfrak{s}_1, \mathrm{p}_{k-1}([f_3(z_1, z_2, z_3), z_4, \dots])) \;.$$

By our choice of Skolem functions the latter simplifies to

$$\mathrm{p}_{k+1}(z_3) < \mathfrak{s}_2 \;\rightarrow\; \beta(u, v, \mathfrak{s}_1, \mathrm{p}_{k+1}(z_3), \mathrm{p}_{k-1}(z_4), \mathrm{p}_{k+1}(z_5), \dots) \;.$$

As $\mathfrak{s} \in A_{\psi xy}(u, v)$ and $\mathrm{tp}(\psi) = \bigvee$ we have $\mathfrak{s}_2 = D$ by Corollary 6.14, thus $\beta(u, v, \mathfrak{s}_1, \mathrm{p}_{k+1}(z_3), \mathrm{p}_{k-1}(z_4), \mathrm{p}_{k+1}(z_5), \dots)$ is satisfied.

Case 3. We have $\bar{u} = u + 1$, $\bar{v} = 0$ and $\bar{\mathfrak{s}} = 0^k$. Observe that the formula $\gamma_{\psi xy}(u + 1, 0, 0^k, \dots)$ is always true by Theorem 8.4, 3. Thus, (8.3) follows from

$$\begin{aligned} &(\mathrm{p}_{k+2}(f_1(z_1)) < u \;\rightarrow\; \neg\beta(\mathrm{p}_{k+2}([f_1(z_1), z_2, f_3(\dots, z_3), \dots]))) \\ &\wedge \gamma_{\psi xy}(u, v, \mathfrak{s}, f_3(z_1, z_2, z_3), z_4, \dots) \\ &\rightarrow (\mathrm{p}_{k+2}(z_1) < u + 1 \;\rightarrow\; \neg\beta(\mathrm{p}_{k+2}([z_1, f_2(z_1, z_2), z_3, \dots]))) \end{aligned} \tag{8.5}$$

If $\mathrm{p}_{k+2}(z_1) \neq u$ choose all slices of Skolem functions canonically, then (8.5) is obviously satisfied.

Now assume $\mathrm{p}_{k+2}(z_1) = u$. We choose Skolem functions in the following way:

$$\begin{array}{rrr} & \mathrm{p}_{k+2}(f_2(z_1, z_2)) = v & \\ \mathrm{p}_k(f_3(z_1, z_2, z_3)) = \mathrm{p}_{k+2}(z_3) & \mathrm{p}_{k+2}(f_4(\dots, z_4)) = \mathrm{p}_k(z_4) & \\ \mathrm{p}_k(f_5(\dots, z_5)) = \mathrm{p}_{k+2}(z_5) & \mathrm{p}_{k+2}(f_6(\dots, z_6)) = \mathrm{p}_k(z_6) & \dots \end{array}$$

and all other slices canonically. Assuming the antecedent of (8.5) we thus have to show $\neg\beta(u, v, \mathrm{p}_{k+2}(z_3), \mathrm{p}_k(z_4), \mathrm{p}_{k+2}(z_5), \dots)$.

If $k = 0$, the definition of N_φ shows $\mathfrak{s} \notin A_{\beta(x,y)}(u, v)$ which by construction implies that $\neg\beta(u, v)$ is true.

In case $k \geq 1$ we obtain from Theorem 8.4, 4 that

$$\gamma_{\psi xy}(u, v, \mathfrak{s}, \mathfrak{t}) \;\rightarrow\; (\mathrm{p}_k(\mathfrak{t}_1) < \mathfrak{s}_1 \;\rightarrow\; \neg\beta(u, v, \mathrm{p}_k(\mathfrak{t})))$$

for $\mathfrak{t} = [f_3(z_1, z_2, z_3), z_4, f_5(\ldots), \ldots]$. Together with the antecedent of (8.5) this implies

$$\mathrm{p}_k(f_3(z_1, z_2, z_3)) < \mathfrak{s}_1 \ \rightarrow\ \neg\beta(u, v, \mathrm{p}_k([f_3(z_1, z_2, z_3), z_4, \ldots])) \ .$$

By our choice of Skolem functions the latter simplifies to

$$\mathrm{p}_{k+2}(z_3) < \mathfrak{s}_1 \ \rightarrow\ \neg\beta(u, v, \mathrm{p}_{k+2}(z_3), \mathrm{p}_k(z_4), \mathrm{p}_{k+2}(z_5), \ldots) \ .$$

As $\mathfrak{s} \notin A_{\psi xy}(u, v)$ and $\mathrm{tp}(\psi) = \bigvee$ we have $\mathfrak{s}_1 = D$ by Corollary 6.14, thus the latter implies $\neg\beta(u, v, \mathrm{p}_{k+2}(z_3), \mathrm{p}_k(z_4), \mathrm{p}_{k+2}(z_5), \ldots)$. □

Definition 8.6. Let φ be a strict formula of rank k, and D a s.u.b. for φ. We extend γ_φ from Definition 8.4 to sequences of length $\ell > k$ in the obvious way:

$$\gamma_\varphi(\langle u_1, \ldots, u_\ell\rangle, \mathfrak{z}) \quad :\Longleftrightarrow \quad \gamma_\varphi(\langle u_1, \ldots, u_k\rangle, \mathfrak{z}) \ .$$

9 Skolemising Π^{b}_k-PLS Conditions

We have seen in Proposition 7.4 that the local search problem L parameterised by $\langle \Phi, \ell, k, h, (\exists y)\varphi(x, y)\rangle$ defines a Π^{b}_k-PLS problem with Π^{b}_ℓ-goal. In this section, we are going to show that the Π^{b}_k-PLS conditions (3.1)-(3.5) for L can be skolemised by simple polynomial time Skolem functions. For the rest of this section, we assume the parametrisation for L is fixed.

Definition 9.1. For each strict formula we fix a notation of its syntactic form. Let $k = \mathrm{rk}(\psi)$ and choose $\beta_\psi(z_1, \ldots, z_k) \in \mathrm{s}\Sigma^{\mathrm{b}}_0 \cup \mathrm{s}\Pi^{\mathrm{b}}_0$ such that the following holds: If $\mathrm{tp}(\psi) = \bigvee$ then $\psi \equiv (\exists\forall^k \mathfrak{z})\beta_\psi(\mathfrak{z})$; if $\mathrm{tp}(\psi) = \bigwedge$ then $\psi \equiv (\forall\exists^k \mathfrak{z})\beta_\psi(\mathfrak{z})$. Further parameters to ψ may be denoted as convenient.

We are now going to fix a suitable prenex form of $s \in F(a)$, which will then be used to show that the Π^{b}_k-PLS conditions (3.1)-(3.5) admit simple Skolem functions.

First, let us bring the formula $s \in F(a)$ into a more readable form: $s \in F(a)$ is equivalent to

$$\begin{aligned}
&\Big[s <D_a \ \wedge\ \varphi(a, s)\Big]\\
&\vee\ \Big[\ s \geq D_a \ \wedge\ s = \langle h, f, \mathfrak{s}\rangle \ \wedge\ s \in C(a) \ \wedge\ \mathrm{tp}(h) \neq \mathrm{Ax}_\psi\\
&\quad \wedge\ (\forall\sigma)\Big(\big(\sigma = \langle 1, \psi, \nu\rangle \ \wedge\ \mathit{Cond}_1(s, \psi, \nu) \ \rightarrow\ \neg\psi[f(\psi)]\ \big)\\
&\qquad \wedge \big(\sigma = \langle 2, \psi, \nu\rangle \ \wedge\ \mathit{Cond}_2(s, \psi, \nu) \ \rightarrow\ \neg\psi\ \big)\\
&\qquad \wedge \big(\sigma = \langle 3, \psi, \nu\rangle \ \wedge\ \mathit{Cond}_3(s, \psi, \nu) \ \rightarrow\ \mathfrak{s} \in F_\psi\ \big)\Big)\ \Big]
\end{aligned}$$

using the following abbreviations:

- $Cond_1(\langle h, f, \mathfrak{s}\rangle, \psi, \nu)$ expresses

$$\nu = \mathrm{rk}(\psi) \ \wedge \ \psi \in \mathrm{dom}(f)$$

- $Cond_2(\langle h, f, \mathfrak{s}\rangle, \psi, \nu)$ expresses

$$\nu = \mathrm{rk}(\psi) \ \wedge \ \psi \in \Gamma(h) \setminus \big(\mathrm{dom}(f) \cup \{(\exists y)\varphi(\underline{a}, y)\}\big)$$

- $Cond_3(\langle h, f, \mathfrak{s}\rangle, \psi, \nu)$ expresses

$$\nu = \mathrm{rk}(\psi) \ \wedge \ \Big(\mathrm{tp}(h) = \mathrm{Cut}_\psi \ \vee \ \big(\mathrm{tp}(h) = \textstyle\bigvee_\chi^i \ \wedge \ \psi = \chi[i]\big)\Big)$$

To increase readability, we have used additional informal parameters as in "$s = \langle h, f, \mathfrak{s}\rangle$", which, when making everything formal, would have to be replaced by appropriate projections, e.g. "h" by "$\mathrm{p}_1(s)$" etc.

The occurrence of ν is currently superfluous but will play a role later. The conditions $s \in C(a)$ and $Cond_1$ to $Cond_3$ are obviously polynomial time computable and thus can be expressed by sharply bounded formulas. Thus, their exact shape is irrelevant for determining a suitable prenex form. The evaluation of formulas $\neg\psi[f(\psi)]$ and $\neg\psi$ can be expressed because each ψ has to be a numerical substitution of a formula from Φ which is a *finite* set.

We continue to determine a suitable prenex form of $s \in F(a)$. Using the suitable prenex form which we have fixed for $\mathfrak{s} \in F_\psi$ in Section 8, and the notation fixed in Definition 9.1, we transform $s \in F(a)$ equivalently into

$$\begin{aligned}
&\Big[s < D_a \ \wedge \ (\forall\exists^\ell \mathfrak{z})\beta_\varphi(a, s, \mathfrak{z})\Big] \\
&\vee \ \Big[s \geq D_a \ \wedge \ s = \langle h, f, \mathfrak{s}\rangle \ \wedge \ s \in C(a) \ \wedge \ \mathrm{tp}(h) \neq \mathrm{Ax}_\psi \ \wedge \ (\forall\sigma)\Big(\\
&\qquad \big(\sigma = \langle 1, \psi, \nu\rangle \ \wedge \ Cond_1(s, \psi, \nu) \ \rightarrow \ \neg\big((\forall\exists^\nu \mathfrak{z})\beta_\psi(\mathfrak{z})[f(\psi)]\big) \ \big) \\
&\quad \wedge \big(\sigma = \langle 2, \psi, \nu\rangle \ \wedge \ Cond_2(s, \psi, \nu) \ \rightarrow \ \neg(\exists\forall^\nu \mathfrak{z})\beta_\psi(\mathfrak{z}) \ \big) \\
&\quad \wedge \big(\sigma = \langle 3, \psi, \nu\rangle \ \wedge \ Cond_3(s, \psi, \nu) \ \rightarrow \ (\forall\exists^\nu \mathfrak{z})\gamma_\psi(\mathfrak{s}, \mathfrak{z})\big) \ \Big)\Big] .
\end{aligned}$$

This is equivalent to

$$(\forall\sigma)(\forall\exists^k \mathfrak{z})\Psi(a, s, \sigma, \mathfrak{z}) \tag{9.1}$$

for $\Psi(a,s,\sigma,\mathfrak{z})$ expressing

$$\begin{aligned}
&\Big[s < D_a \;\wedge\; \beta_\varphi(a,s,\mathrm{p}_\ell(\mathfrak{z}\lceil_\ell))\Big]\\
&\vee\; \Big[s \geq D_a \;\wedge\; s = \langle h,f,\mathfrak{s}\rangle \;\wedge\; s \in C(a) \;\wedge\; \mathrm{tp}(h) \neq \mathrm{Ax}_\psi\\
&\quad\wedge\; \big(\sigma = \langle 1,\psi,\nu\rangle \;\wedge\; \mathit{Cond}_1(s,\psi,\nu) \;\to\; \neg\beta_\psi(f(\psi),\mathrm{p}_{\nu-1}(\mathfrak{z}\lceil_{\nu-1}))\big)\\
&\quad\wedge\; \big(\sigma = \langle 2,\psi,\nu\rangle \;\wedge\; \mathit{Cond}_2(s,\psi,\nu) \;\to\; \neg\beta_\psi(\mathrm{p}_\nu(\mathfrak{z}\lceil_\nu))\big)\\
&\quad\wedge\; \big(\sigma = \langle 3,\psi,\nu\rangle \;\wedge\; \mathit{Cond}_3(s,\psi,\nu) \;\to\; \gamma_\psi(\mathfrak{s},\mathfrak{z}\lceil_\nu)\big)\Big] \;.
\end{aligned}$$

All these equivalences are provable in $\overline{\mathrm{BASIC}}$. The prenex form (9.1) is the one we fix for $s \in F(a)$.

We have implicitly used several independent quantifiers, i.e. we are reading $\mathfrak{z}$ as $[z_1,\ldots,z_k]$ where each variable z_i consists of k "slices" $\mathrm{p}_1(z_i)$, …, $\mathrm{p}_k(z_i)$. Slice i is used for formulas of rank i. As D_a is an s.u.b. for all formulas we have to consider, we may assume w.l.o.g. that the slices in each $\mathfrak{z}_i$ are strictly bounded by D_a, and that quantification and Skolem functions also respect this. We could enforce this by adding further conditions to Ψ, but we refrain from doing so as it only makes the exposition less clear.

Based on the above prenex form of $s \in F(a)$, we now consider the Π^{b}_k-PLS conditions (3.1)-(3.5) for the fixed parameterised local search problem L, and we show that they have prenex forms which admit simple Skolem functions, provable in $\overline{\mathrm{BASIC}}$. We start with the simplest case first.

9.1 Π^{b}_k-PLS condition (3.4)

Condition (3.4) of a Π^{b}_k-PLS problem in general has the form

$$(\forall a,s)(N(a,s) \neq s \;\to\; c(a,N(a,s)) < c(a,s)) \;.$$

As N and c are polynomial time functions, this condition is equivalent to a $\mathrm{s}\Pi^{\mathrm{b}}_1$-formula, so there is nothing to show.

9.2 Π^{b}_k-PLS condition (3.2)

This condition has the form

$$(\forall a)(i(a) \in F(a))$$

which, as we just showed, is equivalent to

$$(\forall a,\sigma)(\forall\exists^k \mathfrak{z})\Psi(a,i(a),\sigma,\mathfrak{z})$$

The latter obviously follows from the following stronger form:

$$(\forall a, \sigma)(\forall^k \mathfrak{z})\Psi(a, i(a), \sigma, \mathfrak{z}) \tag{9.2}$$

Theorem 9.2. (9.2) *is provable in* $\overline{\mathrm{BASIC}}$.

Proof. We argue in $\overline{\mathrm{BASIC}}$. Let $a, \sigma, \mathfrak{z}$ be given, and assume $\sigma = \langle j, \psi, \nu \rangle$. By definition, $i(a) = \langle h_a, \varnothing, 0^k \rangle$. The definition of L shows that $\langle h_a, \varnothing, 0^k \rangle \in C(a)$ and that $\mathrm{tp}(h_a) \neq \mathrm{Ax}_\psi$. We observe that $Cond_1(s, \psi, \nu)$ and $Cond_2(s, \psi, \nu)$ are false as $\Gamma(h_a) \subseteq \{(\exists y)\varphi(\underline{a}, y)\}$. For $j = 3$ we observe that $\mathfrak{s} = 0^k$ and $\gamma_\psi(0^k, \mathfrak{z}\lceil_{\mathrm{rk}(\psi)})$ is true by Theorem 8.4, 3 and Definition 8.6. Hence $\Psi(a, \langle h_a, \varnothing, 0^k \rangle, \sigma, \mathfrak{z})$ is true. □

9.3 Π_k^{b}-PLS condition (3.1)

This condition has the form

$$(\forall a, s)(s \in F(a) \ \rightarrow \ |s| \leq d(|a|))$$

which can be transformed equivalently over $\overline{\mathrm{BASIC}}$ in the following way:

$$\begin{aligned} &(\forall a, s)(s \in F(a) \ \rightarrow \ |s| \leq d(|a|)) \\ \Leftrightarrow \quad &(\forall a, s)\big[(\forall \sigma)(\forall\!\!\!\exists^k \mathfrak{z})\Psi(a, s, \sigma, \mathfrak{z}) \ \rightarrow \ |s| \leq d(|a|)\big] \\ \Leftrightarrow \quad &(\forall a, s)(\exists \sigma)(\exists\!\!\!\forall^k \mathfrak{z})\big[\Psi(a, s, \sigma, \mathfrak{z}) \ \rightarrow \ |s| \leq d(|a|)\big] \end{aligned}$$

The latter obviously follows from the following stronger form:

$$(\forall a, s, \sigma)(\forall^k \mathfrak{z})\big[\Psi(a, s, \sigma, \mathfrak{z}) \ \rightarrow \ |s| \leq d(|a|)\big] \tag{9.3}$$

Theorem 9.3. (9.3) *is provable in* $\overline{\mathrm{BASIC}}$.

Proof. We argue in $\overline{\mathrm{BASIC}}$. Let $a, s, \sigma, \mathfrak{z}$ be given with $\Psi(a, s, \sigma, \mathfrak{z})$. If $s < D_a$ then obviously $|s| \leq d(|a|)$ by definition of d. Otherwise, $s \geq D_a$, and we obtain $s \in C(a)$ by definition of Ψ. Again we obtain $|s| \leq d(|a|)$ by construction of d as indicated in the proof of Proposition 7.4. □

9.4 Π_k^{b}-PLS condition (3.3)

This condition has the form

$$(\forall a, s)(s \in F(a) \ \rightarrow \ N(a, s) \in F(a)) \ .$$

Using the prenex form fixed in (9.1), this formula can be transformed equivalently over $\overline{\text{BASIC}}$ in the following way:

$$\begin{aligned}&(\forall a,s)(s\in F(a)\ \to\ N(a,s)\in F(a))\\ &\quad\Leftrightarrow\quad (\forall a,s)\big[(\forall\sigma)(\forall\!\exists^k\mathfrak{z})\Psi(a,s,\sigma,\mathfrak{z})\ \to\ (\forall\bar\sigma)(\forall\!\exists^k\bar{\mathfrak{z}})\Psi(a,N(a,s),\bar\sigma,\bar{\mathfrak{z}})\big]\\ &\quad\Leftrightarrow\quad (\forall a,s,\bar\sigma,\bar z_1)(\exists\sigma,z_1)(\forall z_2)(\exists\bar z_2)(\forall\bar z_3)(\exists z_3)(\forall z_4)\cdots\\ &\qquad\qquad\big[\Psi(a,s,\sigma,z_1,z_2,\dots)\ \to\ \Psi(a,N(a,s),\bar\sigma,\bar z_1,\bar z_2,\dots)\big]\end{aligned}\tag{9.4}$$

Formula (9.4) is the prenex form which we fix for Condition (3.3).

Theorem 9.4. *The prenex formula* (9.4) *admits simple Skolem functions.*

Proof. We have to show that there are polynomial time functions

$$h_\sigma(a,s,\sigma,z_1),\ h_1(a,s,\sigma,z_1),\ h_2(a,s,\sigma,z_1,z_2),\ h_3(a,s,\sigma,z_1,z_2,z_3),\ \dots$$

such that S^1_2 proves

$$\begin{aligned}&(\forall a,s,\ \sigma,z_1,z_2,z_3,z_4,\dots)\\ &\quad\big[\Psi(a,s,h_\sigma(a,s,\sigma,z_1),h_1(a,s,\sigma,z_1),z_2,h_3(\dots,z_3),z_4,\dots)\\ &\quad\ \to\ \Psi(a,N(a,s),\sigma,z_1,h_2(\dots,z_2),z_3,h_4(\dots,z_4),\dots)\big]\end{aligned}\tag{9.5}$$

In the following we suppress the arguments a,s from the Skolem functions. We say that $h_\sigma(\sigma,z_1)$ (resp., $h_1(\sigma,z_1)$, $h_2(\sigma,z_1,z_2)$, …) is chosen *canonically* if $h_\sigma(\sigma,z_1)=\sigma$ (resp., $h_1(\sigma,z_1)=z_1$, $h_2(\sigma,z_1,z_2)=z_2$, ….)

Let $a,s,\sigma,z_1,z_2,z_3,\dots$ be given. We consider cases according to the definition of $N(a,s)$.

Let us start with some simple cases. Let $s=\langle h,f,0^k\rangle$, $\psi\notin \mathrm{s}\Pi^{\mathrm{b}}_0$ and $N(a,s)=\langle h[f(\psi)],f^r,0^k\rangle$ with $\mathrm{tp}(h)=\bigwedge_\psi$ and $0<\nu:=\mathrm{rk}(\psi)\le k$. If $\sigma\ne\langle 2,\psi[f(\psi)],\nu{-}1\rangle$ or $\mathit{Cond}_2(N(a,s),\psi[f(\psi)],\nu{-}1)$ is false, then choosing Skolem functions canonically obviously satisfies (9.5). So assume $\sigma=\langle 2,\psi[f(\psi)],\nu{-}1\rangle$ and $\mathit{Cond}_2(N(a,s),\psi[f(\psi)],\nu{-}1)$ is true. Choose $h_\sigma(\dots)=\langle 1,\psi,\nu\rangle$ and all other Skolem functions canonically. Then $\mathit{Cond}_1(s,\psi,\nu)$ is satisfied, and (9.5) is equivalent to

$$\neg\beta_\psi(f(\psi),\mathrm{p}_{\nu-1}([z_1,\dots,z_{\nu-1}]))\ \to\ \neg\beta_{\psi[f(\psi)]}(\mathrm{p}_{\nu-1}([z_1,\dots,z_{\nu-1}]))$$

which is obviously true.

Another simple case is if $s=\langle h,f,\mathfrak{s}\rangle$ and $N(a,s)=\langle h,f,N_\psi(\mathfrak{s})\rangle$ with $\mathrm{tp}(h)=\mathrm{Cut}_\psi$, $\nu:=\mathrm{rk}(\psi)$ and $N_\psi(\mathfrak{s})\ne\mathfrak{s}$. If $\sigma\ne\langle 3,\psi,\nu\rangle$ or $\mathit{Cond}_3(N(a,s),\psi,\nu)$ is false, then choosing Skolem functions canonically obviously satisfies (9.5). So assume $\sigma=\langle 3,\psi,\nu\rangle$ and $\mathit{Cond}_3(N(a,s),\psi,\nu)$ is true.

Choose h_σ canonically. As $Cond_3(s,\psi,\nu)$ is obviously satisfied, (9.5) is equivalent to

$$\gamma_\psi(\mathfrak{s}, h_1(\sigma,z_1), z_2, h_3(\dots),\dots) \rightarrow \gamma_\psi(N_\psi(\mathfrak{s}), z_1, h_2(\sigma,z_1,z_2), z_3,\dots) \ .$$

Choosing $h_1, h_2, \dots$ according to Theorem 8.5 will satisfy this implication.

We now list all non-trivial cases. In all other cases not mentioned here, choosing canonical Skolem functions immediately proves the assertion, as above. Let $s = \langle h, f, \mathfrak{s}\rangle$, then the following cases in the definition of $F(a,s)$ have to be considered:

1. $N(a,s) = \langle h[\epsilon(\psi)], f^r, 0^k\rangle$ with $\mathrm{tp}(h) = \bigwedge_\psi$ and $\psi \in \mathrm{s}\Pi^{\mathrm{b}}_0$.
2. $N(a,s) = \langle h[0], f', 0^k\rangle$ with $\mathrm{tp}(h) = \bigvee^i_\psi$, $\psi \notin \mathrm{s}\Sigma^{\mathrm{b}}_1$, $N_{\psi[i]}(\mathfrak{s}) = \mathfrak{s}$, $\mathfrak{s} \notin A_{\psi[i]}$ and $f' = (f \cup \{\psi[i] \mapsto \mathfrak{s}_1\})^r$.
3. $N(a,s) = i$ with $\mathrm{tp}(h) = \bigvee^i_{(\exists y)\varphi(\underline{a},y)}$, $N_{\varphi(\underline{a},\underline{i})}(\mathfrak{s}) = \mathfrak{s}$ and $\mathfrak{s} \in A_{\varphi(\underline{a},\underline{i})}$.
4. $N(a,s) = \langle h[1], f^r, 0^k\rangle$ with $\mathrm{tp}(h) = \mathrm{Cut}_\psi$, $\psi \notin \mathrm{s}\Pi^{\mathrm{b}}_0$, $N_\psi(\mathfrak{s}) = \mathfrak{s}$ and $\mathfrak{s} \in A_\psi$.
5. $N(a,s) = \langle h[0], f', 0^k\rangle$ with $\mathrm{tp}(h) = \mathrm{Cut}_\psi$, $\psi \notin \mathrm{s}\Pi^{\mathrm{b}}_0$, $N_\psi(\mathfrak{s}) = \mathfrak{s}$, $\mathfrak{s} \notin A_\psi$, and $f' = (f \cup \{\psi \mapsto \mathfrak{s}_1\})^r$.

We will now study these cases one by one, thereby considering only critical sub-cases; for all other sub-cases the canonical choices for Skolem functions will already satisfy (9.5).

Case 1. $N(a,s) = \langle h[\epsilon(\psi)], f^r, 0^k\rangle$ with $\mathrm{tp}(h) = \bigwedge_\psi$ and $\psi \in \mathrm{s}\Pi^{\mathrm{b}}_0$. If $\sigma = \langle 2, \psi[\epsilon(\psi)], 0\rangle$ such that $Cond_2(N(a,s), \psi[\epsilon(\psi)], 0)$ is true, we choose $h_\sigma(\sigma,\dots) = \langle 2,\psi,0\rangle$ and all other Skolem functions canonically. Then (9.5) is equivalent to $\neg\beta_\psi \rightarrow \neg\beta_{\psi[\epsilon(\psi)]}$ which is satisfied by definition of $\epsilon(\psi)$, cf. Definition 7.2.

Case 2. $N(a,s) = \langle h[0], f', 0^k\rangle$ with $\mathrm{tp}(h) = \bigvee^i_\psi$, $\psi \notin \mathrm{s}\Sigma^{\mathrm{b}}_1$, $N_{\psi[i]}(\mathfrak{s}) = \mathfrak{s}$, $\mathfrak{s} \notin A_{\psi[i]}$ and $f' = f^r \cup \{\psi[i] \mapsto \mathfrak{s}_1\}$. In this case we have that ψ is of the form $(\exists\forall^\nu \mathfrak{z})\beta_\psi(\mathfrak{z})$ with $\nu \geq 2$. Assume $\sigma = \langle 1, \psi[i], \nu{-}1\rangle$ and $Cond_1(N(a,s), \psi[i], \nu{-}1)$. Let $j := \mathfrak{s}_1$, then $f'(\psi[i]) = j$.

If $\nu = 2$ then $\mathfrak{s} \notin A_{\psi[i]}$ implies $\neg\psi[i][j]$, thus $\neg\beta_{\psi[i][j]}$. In this situation, the conclusion of (9.5) is of the form $\neg\beta_{\psi[i][j]}$ which is true. Hence, any choice of Skolem functions will satisfy (9.5).

Now assume $\nu > 2$. Choose $h_\sigma(\sigma,\dots) = \langle 3, \psi[i], \nu{-}1\rangle$ and all other Skolem functions canonically. $Cond_3(s, \psi[i], \nu{-}1)$ is obviously satisfied, thus (9.5) is equivalent to

$$\gamma_{\psi[i]}(\mathfrak{s}, z_1, z_2, \dots) \to \neg\beta_{\psi[i][j]}(\mathfrak{t})$$

with $\mathfrak{t} = \mathrm{p}_{\nu-2}([z_1, z_2, z_3, \dots])$. Assume $\gamma_{\psi[i]}(\mathfrak{s}, z_1, z_2, \dots)$. Theorem 8.4, 5, shows, as $\mathrm{rk}(\psi[i]) = \nu{-}1$ and $\mathrm{tp}(\psi[i]) = \bigwedge$, that $\mathrm{p}_{\nu-2}(z_1) < \mathfrak{s}_2 \to \neg\beta_{\psi[i]}(\mathfrak{s}_1, \mathfrak{t})$. As $N_{\psi[i]}(\mathfrak{s}) = \mathfrak{s}$, $\mathfrak{s} \notin A_{\psi[i]}$ and $\mathrm{tp}(\psi[i]) = \bigwedge$, we have $\mathfrak{s}_2 = D_a$ by Corollary 6.14, 1. Hence the latter implies $\neg\beta_{\psi[i]}(\mathfrak{s}_1, \mathfrak{t})$ which is the same as $\neg\beta_{\psi[i][j]}(\mathfrak{t})$.

Case 3. $N(a, s) = i$ with $\mathrm{tp}(h) = \bigvee^{i}_{(\exists y)\varphi(\underline{a}, y)}$, $N_{\varphi(\underline{a},\underline{i})}(\mathfrak{s}) = \mathfrak{s}$ and $\mathfrak{s} \in A_{\varphi(\underline{a},\underline{i})}$. We have that $\varphi(\underline{a}, \underline{i})$ is of the form $(\forall\exists^{\ell}\mathfrak{z})\beta_{\varphi(\underline{a},\underline{i})}(\mathfrak{z})$.

If $\ell = 0$ then $\mathfrak{s} \in A_{\varphi(\underline{a},\underline{i})}$ implies $\varphi(\underline{a}, \underline{i})$, which is the same as $\beta_{\varphi(\underline{a},\underline{i})}$. This implies the succedent of (9.5), which is of the form $\beta_\varphi(a, i)$.

If $\ell > 0$, choose $h_\sigma(\sigma, \dots) = \langle 3, \varphi(\underline{a}, \underline{i}), \ell\rangle$ and all other Skolem functions canonically. $Cond_3(s, \varphi(\underline{a}, \underline{i}), \ell)$ is obviously satisfied, thus (9.5) is equivalent to

$$\gamma_{\varphi(\underline{a},\underline{i})}(\mathfrak{s}, z_1, z_2, \dots) \to \beta_{\varphi(\underline{a},\underline{i})}(\mathfrak{t})$$

with $\mathfrak{t} = \mathrm{p}_\ell([z_1, z_2, z_3, \dots])$. Assume $\gamma_{\varphi(\underline{a},\underline{i})}(\mathfrak{s}, z_1, z_2, \dots)$. As $\mathrm{rk}(\varphi) = \ell$ and $\mathrm{tp}(\varphi) = \bigwedge$, Theorem 8.4, 4, shows $\mathrm{p}_\ell(z_1) < \mathfrak{s}_1 \to \beta_{\varphi(\underline{a},\underline{i})}(\mathfrak{t})$. As $N_{\varphi(\underline{a},\underline{i})}(\mathfrak{s}) = \mathfrak{s}$, $\mathfrak{s} \in A_{\varphi(\underline{a},\underline{i})}$ and $\mathrm{tp}(\varphi(\underline{a}, \underline{i})) = \bigwedge$, we have $\mathfrak{s}_1 = D_a$ by Corollary 6.14, 2. Hence the latter implies $\beta_{\varphi(\underline{a},\underline{i})}(\mathfrak{t})$.

Case 4. $N(a, s) = \langle h[1], f^r, 0^k\rangle$ with $\mathrm{tp}(h) = \mathrm{Cut}_\psi$, $N_\psi(\mathfrak{s}) = \mathfrak{s}$ and $\mathfrak{s} \in A_\psi$. We have that $\psi \equiv (\forall\exists^\nu\mathfrak{z})\beta_\psi(\mathfrak{z})$ for $\nu = \mathrm{rk}(\psi)$. Assume $\sigma = \langle 2, \neg\psi, \nu\rangle$ and $Cond_2(N(a, s), \neg\psi, \nu)$ is true.

If $\nu = 0$ choose Skolem functions arbitrarily. Then, the conclusion of (9.5) is equivalent to β_ψ, which is satisfied because $\mathfrak{s} \in A_\psi$ already implies ψ which is the same as β_ψ.

Now assume $\nu > 0$, and choose $h_\sigma(\sigma, \dots) = \langle 3, \psi, \nu\rangle$ and all other Skolem functions canonically. $Cond_3(s, \psi, \nu)$ is obviously satisfied. Then (9.5) is equivalent to

$$\gamma_\psi(\mathfrak{s}, z_1, z_2, \dots) \to \beta_\psi(\mathfrak{t})$$

with $\mathfrak{t} = \mathrm{p}_\nu([z_1, z_2, z_3, \dots])$. Assume $\gamma_\psi(\mathfrak{s}, z_1, z_2, \dots)$. As $\mathrm{rk}(\psi) = \nu$ and $\mathrm{tp}(\psi) = \bigwedge$, Theorem 8.4, 4, shows $\mathrm{p}_\nu(z_1) < \mathfrak{s}_1 \to \beta_\psi(\mathfrak{t})$. By assumption $N_\psi(\mathfrak{s}) = \mathfrak{s}$, $\mathfrak{s} \in A_\psi$ and $\mathrm{tp}(\psi) = \bigwedge$, so $\mathfrak{s}_1 = D_a$ by Corollary 6.14, 2. Hence $\beta_\psi(\mathfrak{t})$ follows.

Case 5. $N(a,s) = \langle h[0], f', 0^k \rangle$ with $\mathrm{tp}(h) = \mathrm{Cut}_\psi$, $\psi \notin \mathrm{s}\Pi^{\mathrm{b}}_0$, $N_\psi(\mathfrak{s}) = \mathfrak{s}$, $\mathfrak{s} \notin A_\psi$, and $f' = (f \cup \{\psi \mapsto \mathfrak{s}_1\})^r$. We have that $\psi \equiv (\forall\!\exists^\nu \mathfrak{z})\beta_\psi(\mathfrak{z})$ for $\nu = \mathrm{rk}(\psi)$, and $\nu > 0$. Let $j := \mathfrak{s}_1$.

If $\nu = 1$, the assumption $\mathfrak{s} \notin A_\psi$ implies $\neg\psi[\mathfrak{s}_1]$ which is $\neg\beta_{\psi[j]}$. Now the critical case is $\sigma = \langle 1, \psi, 1\rangle$, when the conclusion of (9.5) has the form $\neg\beta_{\psi[f(\psi)]}$ which is the same as $\neg\beta_{\psi[j]}$ and satisfied. Arbitrary choices for Skolem functions will satisfy (9.5).

Now assume $\nu > 1$. The critical case now is that $\sigma = \langle 1, \psi, \nu\rangle$ and that $Cond_1(N(a,s), \psi, \nu)$ is true, that is $\psi \in \mathrm{dom}(f')$, and $f'(\psi) = j$ by definition. Choose $h_\sigma(\sigma, \dots) = \langle 3, \psi, \nu\rangle$ and all other Skolem functions canonically. $Cond_3(s, \psi, \nu)$ is obviously satisfied. Then (9.5) is equivalent to

$$\gamma_\psi(\mathfrak{s}, z_1, z_2, \dots) \;\rightarrow\; \neg\beta_{\psi[j]}(\mathfrak{t})$$

with $\mathfrak{t} = \mathrm{p}_{\nu-1}([z_1, z_2, z_3, \dots])$, as $j = f(\psi)$. Assume $\gamma_\psi(\mathfrak{s}, z_1, z_2, \dots)$. As $\mathrm{tp}(\psi) = \bigwedge$ and $\mathrm{rk}(\psi) = \nu$, Theorem 8.4, 5, shows $\mathrm{p}_{\nu-1}(z_1) < \mathfrak{s}_2 \;\rightarrow\; \neg\beta_\psi(\mathfrak{s}_1, \mathfrak{t})$. As $N_\psi(\mathfrak{s}) = \mathfrak{s}$, $\mathfrak{s} \notin A_\psi$ and $\mathrm{tp}(\psi) = \bigwedge$, we have $\mathfrak{s}_2 = D_a$ by Corollary 6.14, 1. Hence $\neg\beta_\psi(\mathfrak{s}_1, \mathfrak{t})$ follows, which is the same as $\neg\beta_{\psi[j]}(\mathfrak{t})$. □

9.5 Π^{b}_k-PLS condition (3.5)

Condition (3.5) can be divided into two parts which we consider independently:

$$(\forall a, s)(s \in G(a) \;\rightarrow\; (N(a,s) = s \;\wedge\; s \in F(a))) \tag{9.6}$$

and

$$(\forall a, s)((N(a,s) = s \;\wedge\; s \in F(a)) \;\rightarrow\; s \in G(a)) \tag{9.7}$$

The goal set $G(a)$ is given as the set of all $s < D_a$ with $\varphi(a,s)$. Using the prenex form fixed for φ according to Definition 9.1, and the prenex form fixed for $s \in F(a)$ in (9.1), formula (9.6) can be transformed equivalently as follows, provably in $\overline{\mathrm{BASIC}}$:

$$\begin{aligned}
&(\forall a, s)\big(s \in G(a) \;\rightarrow\; (N(a,s) = s \;\wedge\; s \in F(a))\big)\\
&\Leftrightarrow\; (\forall a, s)\big(s < D_a \;\wedge\; (\forall\!\exists^\ell \mathfrak{z})\beta_\varphi(a, s, \mathrm{p}_\ell(\mathfrak{z}))\\
&\qquad\qquad \rightarrow\; N(a,s) = s \;\wedge\; (\forall\sigma)(\forall\!\exists^k \bar{\mathfrak{z}})\Psi(a, s, \sigma, \bar{\mathfrak{z}})\big)\\
&\Leftrightarrow\; (\forall a, s)(\forall\sigma)(\forall \bar{z}_1)(\exists z_1)(\forall z_2)(\exists \bar{z}_2)\cdots\\
&\qquad\quad \big(s < D_a \;\wedge\; \beta_\varphi(a, s, \mathrm{p}_\ell([z_1, z_2, \dots]))\\
&\qquad\qquad\quad \rightarrow\; N(a,s) = s \;\wedge\; \Psi(a, s, \sigma, \bar{z}_1, \bar{z}_2, \dots)\big)\;.
\end{aligned}$$

The latter assertion obviously follows from the following stronger one:

$$(\forall a, s, \sigma, z_1, z_2, z_3, \dots)\big(s < D_a \ \wedge\ \beta_\varphi(a, s, \mathrm{p}_\ell([z_1, z_2, \dots])) \\ \to\ N(a,s) = s \ \wedge\ \Psi(a, s, \sigma, z_1, z_2, \dots)\big)\ . \tag{9.8}$$

We show that (9.8) is provable in S^1_2.

Theorem 9.5. S^1_2 *proves* (9.8).

Proof. We argue in S^1_2. Let $a, s, \sigma, z_1, z_2, z_3, \dots$ be given, and assume $s < D_a$ and $\beta_\varphi(a, s, \mathrm{p}_\ell([z_1, z_2, \dots]))$. Hence, $N(a, s) = s$ by definition of N, and $\Psi(a, s, \sigma, z_1, z_2, \dots)$ by definition of Ψ. □

We now turn to condition (9.7). Instead of working directly with this condition we split it into two according to whether $s < D_a$ or not, and simplify the resulting conditions according to their meaning.

$$(\forall a, s)((N(a,s) = s \ \wedge\ s < D_a \ \wedge\ s \in F(a)) \ \to\ s \in G(a)) \tag{9.9}$$

and

$$(\forall a, s)((N(a,s) = s \ \wedge\ s \geq D_a \ \to\ s \notin F(a))) \tag{9.10}$$

We observe that (9.9) and (9.10) together imply (9.7) in $\overline{\mathrm{BASIC}}$.

We consider conditions (9.9) and (9.10) in turn. The former is straight forward to deal with. We transform (9.9) equivalently as follows, provable in $\overline{\mathrm{BASIC}}$:

$$\begin{aligned}
&(\forall a, s)\big((N(a,s) = s \ \wedge\ s < D_a \ \wedge\ s \in F(a)) \ \to\ s \in G(a)\big) \\
&\Leftrightarrow\ (\forall a, s)\big(N(a,s) = s \ \wedge\ s < D_a \ \wedge\ (\forall\sigma)(\forall\exists^k \mathfrak{z})\Psi(a, s, \sigma, \mathfrak{z}) \\
&\qquad\qquad \to\ (\forall\exists^\ell \bar{\mathfrak{z}})\beta_\varphi(a, s, \mathrm{p}_\ell(\bar{\mathfrak{z}}))\big) \\
&\Leftrightarrow\ (\forall a, s)(\forall \bar{z}_1)(\exists\sigma)(\exists z_1)\ (\forall z_2)(\exists \bar{z}_2)\ (\forall \bar{z}_3)(\exists z_3)\ (\forall z_4)(\exists \bar{z}_4)\ \dots \\
&\qquad\qquad \big(N(a,s) = s \ \wedge\ s < D_a \ \wedge\ \Psi(a, s, \sigma, z_1, z_2, z_3, \dots) \\
&\qquad\qquad\quad \to\ \beta_\varphi(a, s, \mathrm{p}_\ell([\bar{z}_1, \bar{z}_2, \bar{z}_3, \dots]))\big)\ .
\end{aligned}$$

The latter is the prenex form which we fix for (9.9). We now show that this prenex form admits simple Skolem functions.

Theorem 9.6. *The prenex form fixed for* (9.9) *admits simple Skolem functions.*

Proof. We have to show that there are polynomial time functions

$$h^\sigma(a,s,z_1),\ h_1(a,s,z_1),\ h_2(a,s,z_1,z_2),\ h_3(a,s,z_1,z_2,z_3),\ \dots$$

such that the following is provable in S^1_2:

$$\begin{aligned}&(\forall a,\ s,z_1,z_2,z_3,z_4,\dots)\big(N(a,s)=s\ \wedge\ s<D_a\\&\quad\wedge\ \Psi(a,s,h^\sigma(a,s,z_1),h_1(a,s,z_1),z_2,h_3(\dots,z_3),\dots)\\&\quad\to\ \beta_\varphi(a,s,\mathrm{p}_\ell([z_1,h_2(a,s,z_1,z_2),z_3,h_4(\dots,z_4),\dots]))\big)\ .\end{aligned} \tag{9.11}$$

We argue in S^1_2. Let $a,s,z_1,z_2,z_3,z_4,\dots$ be given with $N(a,s)=s$ and $s<D_a$. Choose $h^\sigma(\dots)=0$, and all other Skolem functions canonically. Assume $\Psi(a,s,0,z_1,z_2,z_3,z_4\dots)$, then $\beta_\varphi(a,s,\mathrm{p}_\ell([z_1,z_2,z_3,z_4,\dots]))$ follows by definition of $\Psi(a,s,0,0,z_1,z_2,z_3,z_4\dots)$ as $s<D_a$. □

We now turn to condition (9.10) to transform it into a suitable prenex form. This is not at all obvious because the canonical prenex form does not admit simple Skolem functions. The premise of the implication is of low complexity and can be ignored for the prenex form and later the Skolemisation. The only relevant part is the formula "$s\notin F(a)$". First, we double this part to the formula "$s\notin F(a)\vee s\notin F(a)$" to obtain two independent sets of quantifiers. This step is inessential and could have been incorporated already in the prenex form that we fixed for "$s\notin F(a)$". In the second step, we pull out quantifiers, but not in the canonical way (that is those of the same level at the same time, putting universal before existential ones.) Instead, we first pull out the first $(\exists,\forall)$ quantifier pair of the first "$s\notin F(a)$", followed by the first $(\exists,\forall)$ pair of the second "$s\notin F(a)$". Then comes the second $(\exists,\forall)$ pair of the first "$s\notin F(a)$", followed by the second $(\exists,\forall)$ pair of the second "$s\notin F(a)$", and so on. As "$s\notin F(a)$" is of rank k, we produce in this way a prenex formula of rank $2k$, where the canonical prenex form would be of rank k. Thus, we transform (9.10) equivalently as follows, provable in $\overline{\mathrm{BASIC}}$, where the very last equivalence just renames variables:

$$
\begin{aligned}
&(\forall a,s)\big((N(a,s)=s\ \wedge\ s\geq D_a\ \rightarrow\ s\notin F(a))\big)\\
&\Leftrightarrow\ (\forall a,s)\big((N(a,s)=s\ \wedge\ s\geq D_a\ \rightarrow\ s\notin F(a)\ \vee\ s\notin F(a))\big)\\
&\Leftrightarrow\ (\forall a,s)\big(N(a,s)=s\ \wedge\ s\geq D_a\ \rightarrow\ (\exists\sigma^1)(\exists\forall^k\mathfrak{z}^1)\neg\Psi(a,s,\sigma^1,\mathfrak{z}^1)\\
&\qquad\qquad \vee\ (\exists\sigma^2)(\exists\forall^k\mathfrak{z}^2)\neg\Psi(a,s,\sigma^2,\mathfrak{z}^2)\big)\\
&\Leftrightarrow\ (\forall a,s)(\exists\sigma^1,\sigma^2)(\exists z^1_1)\\
&\qquad (\forall z^1_2)(\exists z^2_1)\ (\forall z^2_2)(\exists z^1_3)\ (\forall z^1_4)(\exists z^2_3)\ \cdots\\
&\qquad\quad \big(N(a,s)=s\ \wedge\ s\geq D_a\ \rightarrow\ \neg\Psi(a,s,\sigma^1,z^1_1,z^1_2,z^1_3,\ldots)\\
&\qquad\qquad \vee\ \neg\Psi(a,s,\sigma^2,z^2_1,z^2_2,z^2_3,\ldots)\big)\\
&\Leftrightarrow\ (\forall a,s)(\exists\sigma^1,\sigma^2)(\exists z^1_1)\\
&\qquad (\forall z^1_2)(\exists z^2_2)\ (\forall z^2_3)(\exists z^1_3)\ (\forall z^1_4)(\exists z^2_4)\ \cdots\\
&\qquad\quad \big(N(a,s)=s\ \wedge\ s\geq D_a\ \rightarrow\ \neg\Psi(a,s,\sigma^1,z^1_1,z^1_2,z^1_3,\ldots)\\
&\qquad\qquad \vee\ \neg\Psi(a,s,\sigma^2,z^2_2,z^2_3,z^2_4,\ldots)\big)\ .
\end{aligned}
$$

The latter is the prenex form which we fix for (9.10). We now show that this prenex form admits simple Skolem functions.

Theorem 9.7. *The prenex form fixed for* (9.10) *admits simple Skolem functions.*

Proof. We have to show that there are polynomial time functions

$$
\begin{aligned}
&h^{\sigma^1}(a,s),\ h^{\sigma^2}(a,s),\\
&h_1(a,s),\ h_2(a,s,z_2),\ h_3(a,s,z_2,z_3),\ h_4(a,s,z_2,z_3,z_4),\ \ldots
\end{aligned}
$$

such that the following is provable in S^1_2:

$$
\begin{aligned}
&(\forall a,\ s,z_2,z_3,z_4,\ldots)\big(N(a,s)=s\ \wedge\ s\geq D_a\\
&\quad\rightarrow\ \neg\Psi(a,s,h^{\sigma^1}(a,s),h_1(a,s),z_2,h_3(a,s,z_2,z_3),\ldots)\\
&\qquad \vee\ \neg\Psi(a,s,h^{\sigma^2}(a,s),h_2(a,s,z_2),z_3,h_4(\ldots,z_4),\ldots)\big)\ .
\end{aligned}
\tag{9.12}
$$

We argue in S^1_2. Let $a,s,z_2,z_3,z_4,\ldots$ be given with $N(a,s)=s$ and $s\geq D_a$. Then $N(a,s)=s$ implies by definition of N that $s=\langle h,f,\mathfrak{s}\rangle$, $\mathrm{tp}(h)=\bigvee^i_\psi$, $\nu:=\mathrm{rk}(\psi[i])>0$, $N_{\psi[i]}(\mathfrak{s})=\mathfrak{s}$, $\mathfrak{s}\in A_{\psi[i]}$ and $\psi\not\equiv(\exists y)\varphi(\underline{a},y)$. Choose $h^{\sigma^1}(a,s)=\langle 2,\psi,\nu{+}1\rangle$, $h^{\sigma^2}(a,s)=\langle 3,\psi[i],\nu\rangle$, $\mathrm{p}_{\nu+1}(h_1(a,s))=i$,

$$
\mathrm{p}_\nu(h_j(\ldots,z_j))=\mathrm{p}_{\nu+1}(z_j)\qquad\qquad \mathrm{p}_{\nu+1}(h_j(\ldots,z_j))=\mathrm{p}_\nu(z_j)
$$

for $j = 2, \ldots, k$, and all remaining slices canonically. Let

$$\mathfrak{t} := [\mathrm{p}_{\nu+1}(z_2), \mathrm{p}_\nu(z_3), \mathrm{p}_{\nu+1}(z_4), \mathrm{p}_\nu(z_5), \ldots]$$

then we have

$$\mathrm{p}_{\nu+1}([h_1(a,s), z_2, h_3(a,s,z_2,z_3), \ldots]) = [i, \mathfrak{t}_1, \mathfrak{t}_2, \mathfrak{t}_3, \ldots] \tag{9.13}$$
$$\mathrm{p}_\nu([h_2(a,s,z_2), z_3, h_4(\ldots, z_4), \ldots]) = [\mathfrak{t}_1, \mathfrak{t}_2, \mathfrak{t}_3, \ldots] \tag{9.14}$$

Now, (9.12) is equivalent to

$$\begin{aligned}
&\neg\Psi(a,\ s, \langle 2, \psi, \nu{+}1\rangle\ , h_1(a,s), z_2, h_3(a,s,z_2,z_3), \ldots\,) \\
&\qquad \vee\ \neg\Psi(a, s, \langle 3, \psi[i], \nu\rangle\ , h_2(a,s,z_2), z_3, h_4(\ldots, z_4), \ldots\,) \\
&\Leftrightarrow\quad \beta_\psi(\mathrm{p}_{\nu+1}([h_1(a,s), z_2, h_3(a,s,z_2,z_3), \ldots])) \\
&\qquad\qquad \vee\ \neg\gamma_{\psi[i]}(\mathfrak{s}, h_2(a,s,z_2), z_3, h_4(\ldots, z_4), \ldots\,) \\
&\Leftrightarrow\quad \beta_\psi(i, \mathfrak{t}\lceil_\nu)\ \vee\ \neg\gamma_{\psi[i]}(\mathfrak{s}, h_2(a,s,z_2), z_3, h_4(\ldots, z_4), \ldots\,)
\end{aligned} \tag{9.15}$$

using (9.13) for the last equivalence. To show the last statement (9.15), assume $\gamma_{\psi[i]}(\mathfrak{s}, h_2(a,s,z_2), z_3, h_4(\ldots, z_4), \ldots\,)$. As $\mathrm{tp}(\psi[i]) = \bigwedge$ and $\mathrm{rk}(\psi[i]) = \nu$, Theorem 8.4, 4, shows

$$\mathrm{p}_\nu(h_2(\ldots, z_2)) < \mathfrak{s}_1\ \rightarrow\ \beta_{\psi[i]}(\mathfrak{t}\lceil_\nu)$$

using (9.14). Now, $N_{\psi[i]}(\mathfrak{s}) = \mathfrak{s}$, $\mathfrak{s} \in A_{\psi[i]}$ and $\mathrm{tp}(\psi[i]) = \bigwedge$ show $\mathfrak{s}_1 = D_a$ by Corollary 6.14, 2. Hence, the latter implies $\beta_{\psi[i]}(\mathfrak{t}\lceil_\nu)$ which is the same as $\beta_\psi(i, \mathfrak{t}\lceil_\nu)$. □

The next Corollary summarises the results obtained in this section.

Corollary 9.8. *Let* $0 \le \ell \le k$. *The* $\Sigma^{\mathrm{b}}_{\ell+1}$*-definable total search problems in* T^{k+1}_2 *can be characterised by some* Π^{b}_k*-PLS problems with* Π^{b}_ℓ*-goals, such that conditions* (3.1)–(3.5) *have prenex forms (over* $\overline{\mathrm{BASIC}}$*) which admit simple Skolem functions.*

10 A Proposed Hard Principle for T^k_2

The separation problem of Bounded Arithmetic, i.e. the question whether the hierarchy of Bounded Arithmetic theories is strict or not, is one of the central problems in this area, due to the connections of Bounded Arithmetic theories to complexity classes. There are several ways to approach the separation question. One path which is followed in current research, is by studying relativised theories. Relativised Bounded Arithmetic theories can be obtained by adding one unspecified set variable α to the language of Bounded Arithmetic, which counts as a new atomic formula and is allowed in $\mathrm{s}\Sigma^{\mathrm{b}}_k(\alpha)$-formulas and in induction formulas. Relativised separations have been obtained between all relativised Bounded Arithmetic theories [KPT91, Bus95, Zam96, Jeř09], the goal in current research is to improve the separations, ultimately to find $\forall\Sigma^{\mathrm{b}}_1(\alpha)$ principles which separate the theories, or even $\forall\Pi^{\mathrm{b}}_1(\alpha)$ principles — $\forall\Pi^{\mathrm{b}}_1$ is the complexity of consistency statements.

In this section we will derive, for each k, a generic $\forall\Sigma^{\mathrm{b}}_1(\alpha)$ principle from the results of the previous sections, and show that it gives rise to a class of $\forall\Sigma^{\mathrm{b}}_1$ formulas which characterise the $\forall\Sigma^{\mathrm{b}}_1$ consequences of T^{k+1}_2. The generic form of the principle is therefore conjectured to separate $\mathrm{T}^{k+1}_2(\alpha)$ from $\mathrm{T}^{k}_2(\alpha)$. Such generic principles are well-known in the literature. We will briefly discuss later the relation of the principle which we will define here to the game principles defined in [ST07].

Fix $k \geq 0$. The Skolemisation of the Π^{b}_k-PLS conditions from the previous section forms the basis for the generic $\forall\mathrm{s}\Sigma^{\mathrm{b}}_1(\alpha)$-principle which we will denote by $\mathcal{P}_k$. We replace the polynomial time functions and predicates in the Skolemised versions of (3.1)-(3.5) from the previous section by new function and predicate symbols in the following way: Let N, c, i be new function symbols which will be used for the neighbourhood function, the cost function, and the initial value function respectively. Let G, F' be new relation symbols, where G is binary and is used for the goal set, and F' is $k{+}2$-ary and represents $\Psi(a, s, \sigma, z_1, z_2, \dots, z_k)$ from the prenex form (9.1) fixed for $s \in F(a)$ in the previous section. Let b be a parameter, and let $a = \mathrm{p}_1(b)$, $a_1 = \mathrm{p}_1(\mathrm{p}_2(b))$ and $a_2 = \mathrm{p}_2(\mathrm{p}_2(b))$. We assume $D_a = a_1$, and that a_2 serves as a bound for all quantifiers. The Skolemised versions of (3.1)-(3.5) read as follows — strictly speaking, $(3.1)^{\mathrm{SK}}$ below is not the Skolemisation of (3.1), but a reformulation and adaptation to the current setting, as the original (3.1) is unsuitable. We take b as a parameter to these formulas, from which a, a_1 and a_2 can be computed.

$(3.1)^{\mathrm{SK}}$
$$i(a) < a_2 \ \wedge\ (\forall s < a_2)(N(a,s) < a_2)$$

$(3.2)^{\mathrm{SK}}$
$$(\forall \sigma, z_1, \dots, z_k < a_2) F'(a, i(a), \sigma, z_1, \dots, z_k)$$

$(3.3)^{\mathrm{SK}}$
$$\begin{aligned}&(\forall s, \sigma, z_1, \dots, z_k < a_2)\\ &\quad (F'(a, s, h_\sigma(a,s,\sigma,z_1), h_1(a,s,\sigma,z_1), z_2, h_3(a,s,\sigma,z_1,z_2,z_3), \dots)\\ &\quad \to\ F'(a, N(a,s), \sigma, z_1, h_2(a,s,\sigma,z_1,z_2), z_3, h_4(\dots, z_4), \dots))\end{aligned}$$

$(3.4)^{\mathrm{SK}}$
$$(\forall s < a_2)(N(a,s) = s \ \vee\ c(a, N(a,s)) < c(a,s))$$

$$(3.5\text{a})^{\text{SK}} \qquad (\forall s, z_1, z_2, \ldots, z_k < a_2)(N(a,s) = s \ \wedge \ s < a_1 \\ \wedge F'(a, s, 0, z_1, z_2, z_3, \ldots) \ \rightarrow \ G(a,s))$$

$$(3.5\text{b})^{\text{SK}} \qquad (\forall s, z_2, \ldots, z_k, z_{k+1} < a_2)(N(a,s) = s \ \wedge \ s \geq a_1 \\ \rightarrow \neg F'(a, s, g_{\sigma,1}(a,s), g_1(a,s), z_2, g_3(a, s, z_2, z_3), \ldots) \\ \vee \ \neg F'(a, s, g_{\sigma,2}(a,s), g_2(a, s, z_2), z_3, g_4(\ldots, z_4), \ldots))$$

where $h_\sigma, h_1, h_2, \ldots$ and $g_{\sigma,1}, g_{\sigma,2}, g_1, g_2, \ldots$ are further function symbols representing polynomial time Skolem functions. We have used only one direction of the equivalence in the Skolemisation of (3.5), as we only need this one to prove the principle $\mathcal{P}_k$. This direction comes in two parts, $(3.5\text{a})^{\text{SK}}$ and $(3.5\text{b})^{\text{SK}}$.

Let $\mathcal{X}$ be the list of new function and predicate symbols, that is

$$\mathcal{X} = G, F', N, c, i, h_\sigma, h_1, h_2, \ldots, g_{\sigma,1}, g_{\sigma,2}, g_1, g_2, \ldots$$

Observe that $(3.1)^{\text{SK}}$-$(3.5\text{b})^{\text{SK}}$ are all $\text{s}\Pi^{\text{b}}_1(\mathcal{X})$-formulas. Then the principle $\mathcal{P}_k(\mathcal{X})$ is given by the $\forall\text{s}\Sigma^{\text{b}}_1(\mathcal{X})$-formula obtained from

$$(\forall b)(a_1 < a_2 \ \wedge \ (3.1)^{\text{SK}} \ \wedge \ \cdots \ \wedge \ (3.5\text{b})^{\text{SK}} \ \rightarrow \ (\exists s < a) G(a,s)) \qquad (10.1)$$

by turning the independent bounded existential quantifiers into one using the pairing function and its bound $B(z)$. We observe that the shape of $\mathcal{P}_k$ depends on k.

Theorem 10.1. $\text{T}^{k+1}_2(\mathcal{X}) \vdash \mathcal{P}_k(\mathcal{X})$

Proof. The proof is similar to that of Theorem 3.4. We argue in $\text{T}^{k+1}_2(\mathcal{X})$. Let b be given, Let $a = \text{p}_1(b)$, $a_1 = \text{p}_1(\text{p}_2(b))$ and $a_2 = \text{p}_2(\text{p}_2(b))$. Assume that $a_1 < a_2$, and that $(3.1)^{\text{SK}}$-$(3.5\text{b})^{\text{SK}}$ are satisfied. Let $s \in F(b)$ denote the formula

$$s < a_2 \ \wedge \ (\forall \sigma{<}a_2)(\forall z_1{<}a_2)(\exists z_2{<}a_2) \cdots F'(a, s, \sigma, z_1, z_2, \ldots, z_k) \ .$$

Consider the set $X := \{c(a,s) \colon s < a_2 \ \wedge \ s \in F(b)\}$. This set can be described by some $\text{s}\Sigma^{\text{b}}_{k+1}(\mathcal{X})$-formula. By $(3.1)^{\text{SK}}$ and $(3.2)^{\text{SK}}$ we have $c(a, i(a)) \in X$. As $\text{T}^{k+1}_2(\mathcal{X})$ proves minimisation for $\text{s}\Sigma^{\text{b}}_{k+1}(\mathcal{X})$-properties, we can find some $c \in X$ which is minimal in X. Choose $s < a_2$ and $s \in F(b)$ with $c = c(a,s)$.

As $(3.3)^{\text{SK}}$ is derived from the Skolemisation of a prenex form for (3.3), we obtain $s \in F(b) \ \rightarrow \ N(a,s) \in F(b)$. Thus $N(a,s) \in F(b)$. Also $N(a,s) < a_2$ by $(3.1)^{\text{SK}}$, hence $c(a, N(a,s)) \in X$. As c is minimal in X we obtain $c(a,s) = c \leq c(a, N(a,s))$. Hence with $(3.4)^{\text{SK}}$

$$N(a,s) = s \ .$$

As $(3.5\mathrm{b})^{\mathrm{SK}}$ is derived from the Skolemisation of a prenex form for one part of (3.5), we obtain

$$N(a,s)=s \ \wedge\ s \geq a_1 \ \rightarrow\ s \notin F(b) \ .$$

As $N(a,s)=s$ and $s \in F(b)$ we thus have

$$s < a_1 \ .$$

Also, $(3.5\mathrm{a})^{\mathrm{SK}}$ is derived from the Skolemisation of a prenex form for another part of (3.5). Here we obtain

$$N(a,s)=s \ \wedge\ s < a_1 \ \wedge\ s \in F(b) \ \rightarrow\ G(a,s) \ .$$

Hence we have $G(a,s)$. Altogether this shows $s < a_1 \ \wedge\ G(a,s)$. □

By choosing appropriate substitutions for the parameters, this generic formula can be used to define syntactic search problem classes which characterise the $\forall\Sigma^{\mathrm{b}}_1$-consequences of T^{k+1}_2: Let $\mathrm{PiPLS}^{\mathrm{SK}}(k)$ be the set of all formulas obtained by replacing in $\mathcal{P}_k(\mathcal{X})$, the list of function and predicate symbols $\mathcal{X}$ by polynomial time computable functions and relations (i.e., their definitions in S^1_2.) Note that $\mathrm{PiPLS}^{\mathrm{SK}}(k)$ is a Skolemized version of the principle $\mathrm{PiPLS}(k,0)$ defined in Section 3. The last theorem shows that each formula in $\mathrm{PiPLS}^{\mathrm{SK}}(k)$ is provable in T^{k+1}_2. A converse is also true and can be shown using the results from Section 9. The next Corollary is a refinement of Corollary 3.6.

Corollary 10.2. *Over* S^1_2*, the theories* $\mathrm{PiPLS}^{\mathrm{SK}}(k)$ *and* T^{k+1}_2 *have the same* $\forall\Sigma^{\mathrm{b}}_1$*-consequences.*

Proof. We already argued for one inclusion. We still have to show that if T^{k+1}_2 proves $(\forall x)\varphi(x)$ with $\varphi \in \Sigma^{\mathrm{b}}_1$, then this formula also follows from a formula in $\mathrm{PiPLS}^{\mathrm{SK}}(k)$ over S^1_2.

By Theorem 3.5 and the strengthening in Section 9, we obtain a formalised Π^{p}_k-PLS problem with goal formula identical to φ, whose condition (3.1)–(3.5) have prenex forms which can be Skolemised as described in Section 9, and be proven in S^1_2. Let $\mathcal{X}$ be the list of polynomial time computable functions and predicates coming from this characterisation.

Let D_a be an s.u.b. for the search problem, and d the polynomial bound on the feasible solutions. W.l.o.g. we may assume that d also bounds all occurring σ, i.e., all triples $\langle i,\psi,\nu\rangle$ with $i \leq 3$, $\nu \leq k$, and ψ an instance of a formula in the set of decorations obtained from the original T^{k+1}_2-proof of $(\forall x)\varphi(x)$, by substituting free variables with constants for values $< D_a$. Let $E(a)$ be $2^{d(|a|)}$. Then let b be $\mathrm{pair}(a,\mathrm{pair}(a_1,a_2))$, for $a_1 = D_a$ and $a_2 = B(B(\ldots B(D_a + E(a))\ldots))$,

k iterations of B (here, B is the term giving a bound on the size of pairs: $x, y < z \to \mathrm{pair}(x,y) < B(z)$.) We define

$$N'(a,s) = \begin{cases} N(a,s) & \text{if } s{<}E(a) \ \wedge \ N(a,s){<}E(a) \\ E(a) & \text{otherwise} \end{cases}$$

$$c'(a,s) = \begin{cases} c(a,s)+2 & \text{if } s < E(a) \\ 1 & \text{if } s > E(a) \\ 0 & \text{if } s = E(a) \end{cases}$$

Let $\mathcal{X}'$ be $\mathcal{X}$ in which N, resp. c, has been replaced by N', resp. c'. Consider the formula $\mathcal{P}_k(\mathcal{X}')$ in $\mathrm{PiPLS}^{\mathrm{SK}}(k)$ defined by $\mathcal{X}'$. Given an input a, we choose an instance b for $\mathcal{P}_k(\mathcal{X}')$ as described above. Then it is easy to show that the strengthenings of the formalised Π^{p}_k-PLS problem proved in Section 9 imply

$$a_1 < a_2 \ \wedge \ (3.1)^{\mathrm{SK}} \ \wedge \ \cdots \ \wedge \ (3.5\mathrm{b})^{\mathrm{SK}}$$

in S^1_2, from which we immediately obtain $\varphi(a)$ over S^1_2 assuming $\mathcal{P}_k(\mathcal{X}')$.

We briefly discuss some cases for the above: $(3.1)^{\mathrm{SK}}$ follows immediately from the definitions. $(3.2)^{\mathrm{SK}}$ follows immediately from the related case in Section 9. Same for $(3.4)^{\mathrm{SK}}$.

To show $(3.3)^{\mathrm{SK}}$ let $s, \sigma, z_1, \ldots, z_k < a_2$ such that

$$F'(a, s, h_\sigma(a,s,\sigma,z_1), h_1(\ldots), z_2, \ldots) \ .$$

Thus $\Psi(a, s, h_\sigma(a,s,\sigma,z_1), h_1(\ldots), z_2, \ldots)$ which implies by Theorem 9.4 $\Psi(a, N(a,s), \sigma, z_1, h_2(a,s,\sigma,z_1,z_2), z_3, h_4(\ldots,z_4), \ldots)$. Theorem 9.3 shows that $s, N(a,s) < E(a)$. Thus, $N'(a,s) = N(a,s)$ and we obtain $F'(a, N'(a,s), \sigma, z_1, h_2(\ldots), \ldots)$. The cases $(3.5\mathrm{a})^{\mathrm{SK}}$ and $(3.5\mathrm{b})^{\mathrm{SK}}$ are similar. □

We observe that similar generic principles can be defined for the $\forall\Sigma^{\mathrm{b}}_{\ell+1}$-consequences of T^{k+1}_2.

As the principle $\mathrm{PiPLS}^{\mathrm{SK}}(k)$ characterises all $\forall\Sigma^{\mathrm{b}}_1$ consequences of T^{k+1}_2, we conjecture that its generic version $\mathcal{P}_k(\mathcal{X})$ defined in (10.1) will separate $\mathrm{T}^k_2(\mathcal{X})$ from $\mathrm{T}^{k+1}_2(\mathcal{X})$.

Conjecture 10.3. $\mathrm{T}^k_2(\mathcal{X}) \nvdash \mathcal{P}_k$

By applying standard techniques using bit-graphs of functions and coding different relations into one, the principle $\mathcal{P}_k$ can be transformed into a principle which depends on only one relation variable α. The resulting principle is still a $\forall \mathrm{s}\Sigma^{\mathrm{b}}_1(\alpha)$ sentence conjectured to separate $\mathrm{T}^k_2(\alpha)$ from $\mathrm{T}^{k+1}_2(\alpha)$.

The formula $\mathcal{P}_k$ can also be transformed into a propositional principle conjectured to provide exponential separations between constant-depth propositional proof systems, by using well-known connections between Bounded Arithmetic and constant-depth propositional proof systems via the Paris-Wilkie translation. There are different ways to view the resulting propositional principle. One way is to read it as a polynomial size set of clauses, where each clause is a logarithmic size set of literals.

We do not go into more depth on these constructions, as they are discussed in detail in the related paper [BB08]. The interested reader is kindly referred to that exposition.

We finish this section by comparing our approach to the characterisation of the $\forall\Sigma^{\mathrm{b}}_{\ell+1}$-consequences of T^{k+1}_2 to the results in [ST07]. The game principles GI_k from [ST07] and the principle $\mathrm{PiPLS}^{\mathrm{SK}}(k+1)$ defined here both characterise the $\forall\Sigma^{\mathrm{b}}_1$ consequences of T^{k+2}_2 over S^1_2. From this we immediately obtain that they are reducible to each other under the canonical reduction of total Σ^{b}_1 search problems as discussed e.g. in [ST07]: Let $A = (\forall x)(\exists y)\varphi(x, y)$ and $B = (\forall u)(\exists v)\psi(u, v)$ be two total Σ^{b}_1 search problems, then we call A reducible to B, in symbols $A \leq B$, if there are two polynomial time computable functions f and g, such that for any x, if v is a solution to B on input $f(x)$, i.e. $\psi(f(x), v)$, then $g(x, v)$ is a solution to A on input x, i.e. $\varphi(x, g(x, v))$. The results of [ST07] show that for any formula A in $\mathrm{PiPLS}^{\mathrm{SK}}(k+1)$, there is an instance B in GI_k with $A \leq B$, provable in S^1_2. In the other direction, using the results obtained here, we obtain that for any B in GI_k, there is a formula A in $\mathrm{PiPLS}^{\mathrm{SK}}(k+1)$ with $B \leq A$, provable in S^1_2. An inspection of the proof of Corollary 10.2 shows that in the latter case the reducing functions are given by the identity for f and a projection to the last component of the second argument (which codes, using the pairing function, the values of several existential quantifiers into one) for g. It is also possible to give a simple direct reduction from the GI_k principle to an instance of $\mathcal{P}_{k+1}$ by a construction that directly matches the combinatorial structure of GI_k. It is not clear whether there is a similarly simple direct reduction from $\mathcal{P}_{k+1}$ to GI_k.

References

[AB09] Klaus Aehlig and Arnold Beckmann. On the computational complexity of cut-reduction, 2009. Accepted for publication, DOI 10.1016/j.apal.2009.06.004.

[BB08] Arnold Beckmann and Samuel R. Buss. Polynomial local search in the polynomial hierarchy and witnessing in fragments of bounded arithmetic. Technical Report

CSR15-2008, Department of Computer Science, Swansea University, December 2008.

[Bec03] Arnold Beckmann. Dynamic ordinal analysis. *Arch. Math. Logic*, 42:303–334, 2003.

[BK94] Samuel R. Buss and Jan Krajíček. An application of Boolean complexity to separation problems in bounded arithmetic. *Proc. London Math. Soc. (3)*, 69(1):1–21, 1994.

[Buc91] Wilfried Buchholz. Notation systems for infinitary derivations. *Archive for Mathematical Logic*, 30:277–296, 1991.

[Buc97] Wilfried Buchholz. Explaining Gentzen's consistency proof within infinitary proof theory. In *Computational logic and proof theory (Vienna, 1997)*, volume 1289 of *Lecture Notes in Comput. Sci.*, pages 4–17. Springer, Berlin, 1997.

[Bus86] Samuel R. Buss. *Bounded arithmetic*, volume 3 of *Studies in Proof Theory. Lecture Notes*. Bibliopolis, Naples, 1986.

[Bus95] Samuel R. Buss. Relating the bounded arithmetic and the polynomial time hierarchies. *Annals of Pure and Applied Logic*, 75:67–77, 1995.

[Jeř09] Emil Jeřábek. Approximate counting by hashing in bounded arithmetic. *Journal of Symbolic Logic*, 74(3):829–860, 2009.

[KPT91] Jan Krajíček, Pavel Pudlák, and Gaisi Takeuti. Bounded arithmetic and the polynomial hierarchy. *Annals of Pure and Applied Logic*, 52:143–153, 1991.

[Kra93] Jan Krajíček. Fragments of bounded arithmetic and bounded query classes. *Trans. Amer. Math. Soc.*, 338(2):587–598, 1993.

[KST07] Jan Krajíček, Alan Skelley, and Neil Thapen. NP search problems in low fragments of bounded arithmetic. *J. Symbolic Logic*, 72(2):649–672, 2007.

[Min78] Grigori E. Mints. Finite investigations of transfinite derivations. *Journal of Soviet Mathematics*, 10:548–596, 1978. Translated from: Zap. Nauchn. Semin. LOMI 49 (1975). Cited after Grigori Mints. *Selected papers in Proof Theory*.Studies in Proof Theory. Bibliopolis, 1992.

[Pol99] Chris Pollett. Structure and definability in general bounded arithmetic theories. *Ann. Pure Appl. Logic*, 100(1-3):189–245, 1999.

[Pud06] Pavel Pudlák. Consistency and games—in search of new combinatorial principles. In *Logic Colloquium '03*, volume 24 of *Lect. Notes Log.*, pages 244–281. Assoc. Symbol. Logic, La Jolla, CA, 2006.

[Pud07] Pavel Pudlák. Fragments of bounded arithmetic and the lengths of proofs, 2007. Preprint.

[PW85] J. Paris and A. Wilkie. Counting problems in bounded arithmetic. In A. Dold and B. Eckmann, editors, *Methods in Mathematical Logic (Proceedings Caracas 1983)*, number 1130 in Lecture Notes in Mathematics, pages 317–340. Springer, 1985.

[ST07] Alan Skelley and Neil Thapen. The provable total search problems of bounded arithmetic, 2007. Preprint.

[Tai68] William W. Tait. Normal derivability in classical logic. In J. Barwise, editor, *The Syntax and Semantics of Infinitatry Languages*, number 72 in Lecture Notes in Mathematics, pages 204–236. Springer, 1968.

[Zam96] Domenico Zambella. Notes on polynomially bounded arithmetic. *Journal of Symbolic Logic*, 61:942–966, 1996.

On Topological Models of GLP

Lev Beklemishev*, Guram Bezhanishvili, and Thomas Icard

Abstract We develop topological semantics of a polymodal provability logic **GLP**. Our main result states that the bimodal fragment of **GLP**, although incomplete with respect to relational semantics, is topologically complete. The topological (in)completeness of **GLP** remains an interesting open problem.

1 Introduction

In this paper we initiate a study of topological models of an important polymodal provability logic **GLP** due to Japaridze [21, 22]. This system describes in the style of provability logic all the universally valid schemata for the reflection principles of restricted logical complexity in arithmetic. Thus, it is complete with respect to a very natural kind of arithmetical semantics.

The logic **GLP**, and its restricted bimodal version **GLB**, have been extensively studied in the early 1990s by Ignatiev [19, 20] and Boolos, who simplified and extended Japaridze's work. Boolos incorporated a very readable treatment of **GLB** into his popular book on provability logic [11]. More recently, interesting applications of **GLP** have been found in proof theory and ordinal analysis of arithmetic. In particular, **GLP** gives rise to a natural system of ordinal notation for the ordinal ϵ_0. Based on this system and the use of **GLP**, the first author of this paper gave a simple proof of consistency of Peano Arithmetic à la Gentzen and formulated a new independent combinatorial principle. This stimulated further interest towards **GLP** (see [3, 4] for a detailed survey).

The main difficulty in the modal-logical study of **GLP** comes from the fact that it is incomplete with respect to its relational semantics; that is, **GLP** is the logic of no class of *frames*. On the other hand, a suitable class of relational *models* for which **GLP** is sound and complete was developed in [5]. However, these models are sufficiently complicated to warrant a search for an alternative and simpler kind of semantics.

Many standard modal logics enjoy a natural topological interpretation. Topologically, propositions are interpreted as subsets of a topological space and boolean

*Supported by the Russian Foundation for Basic Research (RFBR), Dutch Organization for Scientific Research (NWO), Russian Presidential Council for Support of Leading Scientific Schools, and the ECO-NET cooperation project "Computational interpretations of modal logic."

connectives correspond to the standard set-theoretic operations. For logics containing the reflection axiom $\varphi \to \Diamond\varphi$, one usually interprets the modal $\Diamond$ as the closure operator of a topological space. However, provability logics fall outside this class due to the presence of Löb's axiom which contradicts reflection. For these logics one takes a different approach that reads $\Diamond$ as the derived set operator d mapping a set A to the set of limit points of A. The study of this interpretation was suggested in the Appendix of [25], and was developed by Esakia (see [13, 14] and [7] for a survey). In particular, Esakia noticed that a topological space satisfies Löb's axiom iff it is *scattered*. The concept of a scattered space goes all the way back to Cantor. Typical examples of scattered spaces are ordinals (in the interval topology). In fact, it was shown independently by Abashidze [2] and Blass [10] that the provability logic $\mathbf{GL}$ is complete with respect to any ordinal $\alpha \geq \omega^\omega$.

When generalizing topological interpretation to several modalities we deal with *polytopological spaces*; that is, sets equipped with several topologies τ_0, τ_1, ... The corresponding derived set operators d_0, d_1, ... then interpret the diamond modalities $\langle 0 \rangle$, $\langle 1 \rangle$, ... of our language in the usual way. The axioms of $\mathbf{GLP}$ impose restrictions on the relevant class of polytopological spaces, which leads to the concept of a $\mathbf{GLP}$-*space* (or, of a $\mathbf{GLB}$-*space* for the language with just two modalities).

It is well known that $\mathbf{GL}$ is complete with respect to its relational semantics; in fact, $\mathbf{GL}$ is the logic of finite irreflexive transitive trees (see, e.g., [11]). In contrast, $\mathbf{GLB}$ is incomplete with respect to its relational semantics. But the main result of this paper states that $\mathbf{GLB}$ is topologically complete. Thus, $\mathbf{GLB}$ appears to be the first naturally occurring example of a modal logic which is topologically complete but incomplete with respect to its relational semantics (artificial examples of this kind have already been known; see, e.g., [15, 16]). It is also worth pointing out that in [26] it was stated as an open problem whether there existed a topologically complete but relationally incomplete finitely axiomatizable modal logic. The question was stated for the case of modal logics with one modality and in this stronger form it still remains open. Nevertheless, since $\mathbf{GLB}$ is finitely axiomatizable, our results provide an answer to the bimodal version of the problem.

Our technique (which is based on the construction in [5]) does not obviously extend to the case with three or more modalities. Therefore, the topological completeness of $\mathbf{GLP}$ remains an interesting open problem. At the end of the paper we discuss some negative results indicating that the situation here could be significantly more complicated and the question of topological completeness of $\mathbf{GLP}$ might be independent of the axioms of Zermelo–Fraenkel set theory ZFC with the axiom of choice.

On the other hand, the third author of this paper established the topological completeness of the closed fragment of $\mathbf{GLP}$ (in the language with ω-many modalities) with respect to a natural polytopological space on the ordinal ϵ_0 (see [17, 18]). However, this space fails to be a $\mathbf{GLP}$-space.

The paper is organized as follows. In Section 2 we introduce $\mathbf{GLP}$ and its bimodal fragment $\mathbf{GLB}$, and discuss their relational, algebraic, and topological semantics. We also discuss Stone-like duality for $\mathbf{GLP}$-algebras and the resulting descriptive frames. In Section 3 we prove topological completeness of $\mathbf{GLB}$ with respect to the class of $\mathbf{GLB}$-spaces. We finish the paper with a discussion of some further results and remaining open questions.

2 Relational, algebraic, and topological semantics for GLP.

2.1 $\mathbf{GLP}$ and its relational semantics

$\mathbf{GLP}$ is a propositional modal logic formulated in a language with infinitely many modalities [0], [1], [2], ... As usual, $\langle n\rangle\varphi$ stands for $\neg[n]\neg\varphi$.

Definition 2.1. $\mathbf{GLP}$ is given by the following axiom schemata and rules.

Axioms:

(i) Boolean tautologies;

(ii) $[n](\varphi \to \psi) \to ([n]\varphi \to [n]\psi)$;

(iii) $[n]([n]\varphi \to \varphi) \to [n]\varphi$ (Löb's axiom);

(iv) $[m]\varphi \to [n]\varphi$, for $m < n$;

(v) $\langle m\rangle\varphi \to [n]\langle m\rangle\varphi$, for $m < n$.

Rules:

(i) $\vdash \varphi, \vdash \varphi \to \psi \;\Rightarrow \vdash \psi$ (modus ponens);

(ii) $\vdash \varphi \;\Rightarrow \vdash [n]\varphi$, for each $n \in \omega$ (necessitation).

In other words, for each modality we have the Gödel-Löb Logic $\mathbf{GL}$, and (iv) and (v) are the two axioms relating modalities to one another.

We denote by $\mathbf{GLB}$ the bimodal fragment of $\mathbf{GLP}$, restricted to the language with only [0] and [1], and by $\mathbf{GLP}_0$ the letterless fragment of $\mathbf{GLP}$, restricted to the language without variables (we assume propositional constants $\top$ and $\bot$ to be part of the language).

As usual, we would like to know what class of frames, if any, these logics define. Relational models of $\mathbf{GLP}_0$ have been studied extensively, first in [19] and [20], and later in [6]; see also [17, 18]. Unfortunately, for the fragments with variables, and already in the case of **GLB**, there is no single non-trivial frame for which we have soundness. To see this, we briefly recall relational semantics for **GL**.

A *(unimodal) frame* is a pair $\mathfrak{F} = \langle W, R\rangle$, where W is a nonempty set and R is a binary relation on W; $\mathfrak{F}$ is *transitive* if $wRvRu$ implies wRu for each $w, v, u \in W$ and *irreflexive* if wRw for no $w \in W$; a transitive frame $\mathfrak{F}$ is *dually well-founded* if for each nonempty subset U of W there exists $w \in U$ such that wRu for no $u \in U$. In such a case we call R a *dually well-founded relation*. It is well known that $\mathfrak{F} \models \mathbf{GL}$ iff $\mathfrak{F}$ is dually well-founded. Typical examples of dually well-founded frames are finite transitive irreflexive frames, and in fact, **GL** is the logic of these (see, e.g., [11]).

Next we recall that a *(polymodal) frame* is a tuple $\mathfrak{F} = \langle W, \{R_n\}_{n\in\omega}\rangle$, where W is a nonempty set and each R_n is a binary relation on W. For $A \subseteq W$ let $-A$ denote the complement of A in W. We recall that a *valuation* is a map $v : \mathrm{Var} \to 2^W$ from the set of propositional variables to the powerset of W and that v extends to all formulas as follows:

- $v(\varphi \vee \psi) = v(\varphi) \cup v(\psi)$, $v(\neg\varphi) = -v(\varphi)$, $v(\top) = W$, $v(\bot) = \varnothing$,
- $v(\langle n\rangle\varphi) = \{x \in W : \exists y\ (xR_ny \ \&\ y \in v(\varphi))\}$
- $v([n]\varphi) = \{x \in W : \forall y\ (xR_ny \Rightarrow y \in v(\varphi))\}$.

We will write $\mathfrak{F}, x \models_v \varphi$ for $x \in v(\varphi)$. If v is fixed, we abbreviate $\mathfrak{F}, x \models_v \varphi$ by $\mathfrak{F}, x \models \varphi$ or even $x \models \varphi$. A formula φ is *valid in* $\mathfrak{F}$, denoted $\mathfrak{F} \models \varphi$, if $v(\varphi) = W$ for all v.

In order for $\mathfrak{F}$ to be a **GLP**-frame, each R_n should be a dually well-founded relation and in addition $\mathfrak{F}$ should validate axioms (iv) and (v). The next lemma, which is well-known, gives necessary and sufficient conditions for this.

Lemma 2.2. *Let $m < n$. Then:*

1. $\mathfrak{F} \models [m]\varphi \to [n]\varphi$ *iff* wR_nv *implies* wR_mv.
2. $\mathfrak{F} \models \langle m\rangle\varphi \to [n]\langle m\rangle\varphi$ *iff* wR_mv *and* wR_nu *imply* uR_mv.

Proof. See, e.g., [11]. □

Remark 2.3. Let $\mathfrak{F}$ be a (polymodal) frame, R_n^{-1} denote the inverse of R_n, $R_n(U) := \{w \in W : \exists u \in U, uR_nw\}$, and $R_n^{-1}(U) := \{w \in W : \exists u \in U, wR_nu\}$. We call U an R_n-*upset* if it is upward closed with respect to R_n; that is, $u \in U$ and uR_nw imply $w \in U$. (Similarly, we call U an R_n-*downset* if $u \in U$ and wR_nu imply $w \in W$.) Then axiom (iv) states that $R_n \subseteq R_m$ and axiom (v) states that each set of the form $R_m^{-1}(U)$ is an R_n-upset.

We show that no non-trivial frame satisfies all of these requirements. Suppose for a contradiction that **GLB** is sound with respect to a frame $\mathfrak{F}$ with R_1 nonempty. Then there are $w, v \in W$ such that wR_1v. By Lemma 2.2(1), wR_0v, and by Lemma 2.2(2), vR_0v, which contradicts to R_0 being dually well-founded. Consequently, if $\mathfrak{F} \models$ **GLB** then $R_1 = \varnothing$, so $[1]\bot$ becomes valid. This obviously generalizes to **GLP**. Thus, we obtain:

Theorem 2.4. **GLP** *is incomplete with respect to its class of frames. In particular,* **GLP** *is not sound on any frame for which* $R_n \neq \varnothing$ *for* $n > 0$.

2.2 Algebraic semantics and descriptive frames

As we saw, **GLP** is incomplete with respect to relational semantics, and as we will see, topological completeness of **GLP** remains an open problem. Nevertheless, there is a semantics for which completeness of **GLP** is automatic, viz. algebraic semantics. Of course, algebraic semantics is not as informative as either relational or topological semantics, but completeness is straightforward through the well-known Lindenbaum construction. Moreover, Stone-like duality for **GLP**-algebras can be developed without much trouble.

We recall that a pair $\mathfrak{A} = \langle B, \delta \rangle$ is a **GL**-*algebra* (also known as a *diagonalizable algebra* or a *Magari algebra*) if B is a boolean algebra and $\delta : B \to B$ is a unary function on B such that $\delta 0 = 0$, $\delta(a \vee b) = \delta a \vee \delta b$, and $\delta a = \delta(a - \delta a)$. Given a **GL**-algebra $\mathfrak{A} = \langle B, \delta \rangle$, let $\tau a = -\delta(-a)$. It is well known that if we interpret formulas as elements of a **GL**-algebra $\mathfrak{A} = \langle B, \delta \rangle$, boolean connectives as boolean operations of B, and $\Diamond$ as δ (and hence $\Box$ as τ), then **GL**-algebras provide an adequate semantics for **GL**.

Definition 2.5. We call a tuple $\mathfrak{A} = \langle B, \{\delta_n\}_{n \in \omega} \rangle$ a **GLP**-*algebra* if

(i) $\langle B, \delta_n \rangle$ is a **GL**-algebra for each $n \in \omega$;

(ii) $\delta_n a \leq \delta_m a$ for each $m < n$ and $a \in B$;

(iii) $\delta_m a \leq \tau_n \delta_m a$ for each $m < n$ and $a \in B$.

In particular, a triple $\mathfrak{A} = \langle B, \delta_0, \delta_1 \rangle$ is a **GLB**-*algebra* if both $\langle B, \delta_0 \rangle$ and $\langle B, \delta_1 \rangle$ are **GL**-algebras, $\delta_1 a \leq \delta_0 a$, and $\delta_0 a \leq \tau_1 \delta_0 a$ for each $a \in B$.

A standard argument shows that **GLP**-algebras provide an adequate semantics for **GLP**, and **GLB**-algebras provide an adequate semantics for **GLB**. We give three types of examples of **GLP**-algebras.

Example 2.6 (free algebras)**.** Free n-generated $\mathbf{GLP}$-algebras, also known as Lindenbaum algebras, are obtained from the set of all formulas of $\mathbf{GLP}$ in the language with n propositional variables by identifying $\mathbf{GLP}$-equivalent formulas and defining the boolean algebra operations by logical connectives. The modal operators δ_n map the equivalence class of a formula φ to the equivalence class of the formula $\langle n\rangle\varphi$. In particular, the free 0-generated algebra is the Lindenbaum algebra of the letterless fragment $\mathbf{GLP}_0$.

Another kind of $\mathbf{GLP}$-algebras come from $\mathbf{GLP}$-spaces (see next section).

Example 2.7. Let $\mathcal{X}$ be a $\mathbf{GLP}$-space. The boolean algebra of all subsets of X enriched with the derived set operators d_n, for each $n \geq 0$, acting on 2^X is obviously a $\mathbf{GLP}$-algebra.

Perhaps the most intriguing examples of $\mathbf{GLP}$-algebras come from proof theory, where they have been introduced under the name of *graded provability algebras* [3].

Example 2.8 (provability algebras)**.** Let T be a first order arithmetical theory containing a sufficiently large fragment of Peano arithmetic PA. T is called *n-consistent* if the union of T and all true Π_n-sentences is consistent. If φ is an arithmetical sentence, let $\langle n\rangle_T\varphi$ denote a natural formalization of the statement that the theory $T+\varphi$ is n-consistent. (Such a formalization is equivalent to the so-called *uniform Σ_n-reflection principle* for $T+\varphi$.) This defines a function $\delta_n : \varphi \mapsto \langle n\rangle_T\varphi$, which is correctly defined on the equivalence classes of sentences modulo provable equivalence in T. The Lindenbaum algebra of T enriched with all the operators δ_n happens to be a $\mathbf{GLP}$-algebra. This example plays a fundamental role in the proof-theoretic analysis of PA based on provability logic (see [3,4]).

Of course, $\mathbf{GLP}$-algebras (respectively, $\mathbf{GLB}$-algebras) in general are rather abstract entities. Therefore, it is desirable to have a good representation for them. This is done through the well-known Stone construction.

Let X be a topological space. We recall that a subset A of X is *clopen* if A is both closed and open, and that X is *zero-dimensional* if clopen subsets form a basis for X. We also recall that X is a *Stone space* if it is compact, Hausdorff, and zero-dimensional.

It is a celebrated result of Stone that boolean algebras can be represented as the algebras of clopen subsets of Stone spaces. We recall that given a boolean algebra B, the dual Stone space X of B is constructed as the space of ultrafilters of B and that a topology on X is defined by declaring $\{\varphi(a) : a \in B\}$ to be a basis for the topology, where $\varphi(a) = \{x \in X : a \in x\}$. Let $\mathrm{Cp}(X)$ denote the set of clopen subsets of X. Then $\mathrm{Cp}(X)$ with set-theoretic operations $\cap, \cup, -$

is a boolean algebra, and $\varphi : B \to \mathrm{Cp}(X)$ is a boolean algebra isomorphism. This 1-1 correspondence between boolean algebras and Stone spaces extends to a categorical dual equivalence between the category of boolean algebras and boolean algebra homomorphisms and the category of Stone spaces and continuous maps.

This representation of boolean algebras was extended to a representation of $\mathbf{GL}$-algebras by Magari [24] and by Esakia and Abashidze [1] (see also [12] and [8]). Let X be a Stone space and R a transitive relation on X. For a clopen $A \subseteq X$ we call $x \in A$ a *strongly maximal point* of A if xRy for no $y \in A$. In particular, a strongly maximal point is irreflexive. Now we call a pair $\langle X, R\rangle$ a *descriptive* $\mathbf{GL}$*-frame* if X is a Stone space and R is a transitive binary relation on X such that $R(x)$ is closed for each $x \in X$, A clopen implies $R^{-1}(A)$ is clopen, and for each clopen A and $x \in A$, either x is strongly maximal or there exists a strongly maximal point $y \in A$ such that xRy.[1]

Let $\langle B, \delta\rangle$ be a $\mathbf{GL}$-algebra and let X be the Stone space of B. We define R on X by xRy iff $a \in y$ implies $\delta a \in x$ for each $a \in B$. Since in each $\mathbf{GL}$-algebra we have $\delta\delta a \leq \delta a$, it is easy to verify that R is transitive. It is also standard to show that $R(x)$ is closed for each $x \in X$, A clopen implies $R^{-1}(A)$ is clopen, and $\varphi(\delta a) = R^{-1}(\varphi(a))$. In fact, $\langle X, R\rangle$ is a descriptive $\mathbf{GL}$-frame. This follows from the following lemma proved in [1].

Lemma 2.9. *If $\langle B, \delta\rangle$ is a $\mathbf{GL}$-algebra and $\langle X, R\rangle$ is the dual of $\langle B, \delta\rangle$, then $\langle X, R\rangle$ is a descriptive $\mathbf{GL}$-frame.*

Proof. (Sketch) Let A be a clopen subset of X. It is sufficient to show that for each $x \in A$, either x is a strongly maximal point or there exists a strongly maximal point $y \in A$ such that xRy. If $x \notin R^{-1}(A)$, then x is a strongly maximal point. Suppose that $x \in R^{-1}(A)$. Since A is clopen, there exists $a \in B$ such that $A = \varphi(a)$. Therefore, $x \in R^{-1}(\varphi(a))$. As $R^{-1}(\varphi(a)) = \varphi(\delta a)$, we obtain $x \in \varphi(\delta a)$. But $\delta a = \delta(a - \delta a)$. Thus, $x \in \varphi(\delta(a - \delta a)) = R^{-1}(\varphi(a - \delta a))$. This implies that there exists $y \in \varphi(a - \delta a)$ such that xRy. Now as $y \in \varphi(a - \delta a) = \varphi(a) - R^{-1}(\varphi(a))$, y must be a strongly maximal point of $\varphi(a) = A$. □

It follows that if $\mathfrak{A} = \langle B, \delta\rangle$ is a $\mathbf{GL}$-algebra, then $\mathfrak{X} = \langle X, R\rangle$ is a descriptive $\mathbf{GL}$-frame and $\varphi : \langle B, \delta\rangle \to \langle \mathrm{Cp}(X), R^{-1}\rangle$ is an isomorphism of $\mathbf{GL}$-algebras. Thus, each $\mathbf{GL}$-algebra can be represented as the algebra of clopen subsets of the corresponding descriptive $\mathbf{GL}$-frame. In particular, if $\mathfrak{A}$ is countable, then $\mathfrak{X}$ is second-countable.

As in the case of boolean algebras and Stone spaces, this representation extends to a dual equivalence of the appropriate categories, however we will not address this here and refer the interested reader to [1, 8].

[1] Descriptive $\mathbf{GL}$-frames were called *strong transits* in [1].

This representation of $\mathbf{GL}$-algebras extends in an obvious way to $\mathbf{GLP}$-algebras and $\mathbf{GLB}$-algebras.

Definition 2.10. We call a tuple $\mathfrak{X} = \langle X, \{R_n\}_{n\in\omega}\rangle$ a *descriptive* $\mathbf{GLP}$*-frame* if

(i) $\langle X, R_n\rangle$ is a descriptive $\mathbf{GL}$-frame for each $n \in \omega$;

(ii) $R_n \subseteq R_m$ for each $m < n$;

(iii) xR_my and xR_nz imply zR_my for each $m < n$.

In particular, a triple $\mathfrak{X} = \langle X, R_0, R_1\rangle$ is a *descriptive* $\mathbf{GLB}$*-frame* if both $\langle X, R_0\rangle$ and $\langle X, R_1\rangle$ are descriptive $\mathbf{GL}$-frames, $R_1 \subseteq R_0$, and xR_0y and xR_1z imply zR_0y.

Let $\mathfrak{A} = \langle B, \{\delta_n\}_{n\in\omega}\rangle$ be a $\mathbf{GLP}$-algebra, X the Stone space of B, and xR_ny iff $a \in y$ implies $\delta_n a \in x$ for each $n \in \omega$ and $a \in B$.

Lemma 2.11. *If* $\mathfrak{A} = \langle B, \{\delta_n\}_{n\in\omega}\rangle$ *is a* $\mathbf{GLP}$*-algebra, then* $\mathfrak{X} = \langle X, \{R_n\}_{n\in\omega}\rangle$ *is a descriptive* $\mathbf{GLP}$*-frame. Moreover,* $\varphi : \langle B, \{\delta_n\}_{n\in\omega}\rangle \to \langle \mathrm{Cp}(X), \{R_n^{-1}\}_{n\in\omega}\rangle$ *is an isomorphism of* $\mathbf{GLP}$*-algebras.*

Proof. In view of the representation of $\mathbf{GL}$-algebras, all we have to verify is that $R_n \subseteq R_m$ and xR_my and xR_nz imply zR_my for each $m < n$. Let xR_ny and $a \in y$. Then $\delta_n a \in x$. Since $\delta_n a \le \delta_m a$, also $\delta_m a \in x$. Therefore, xR_my, and so $R_n \subseteq R_m$. Now let xR_my and xR_nz. Suppose that $a \in y$. Since xR_my, we have $\delta_m a \in x$. If $\delta_m a \notin z$, then $-\delta_m a \in z$. As xR_nz, we have $\delta_n(-\delta_m a) \in x$. But $\delta_m a \in x$ and $\delta_m a \le -\delta_n(-\delta_m a)$ imply $-\delta_n(-\delta_m a) \in x$, a contradiction. Thus, $\delta_m a \in z$, and so zR_my. □

In particular, Lemma 2.11 implies that if $\mathfrak{A} = \langle B, \delta_0, \delta_1\rangle$ is a $\mathbf{GLB}$-algebra, then $\mathfrak{X} = \langle X, R_0, R_1\rangle$ is a descriptive $\mathbf{GLB}$-frame, and $\varphi : \langle B, \delta_0, \delta_1\rangle \to \langle \mathrm{Cp}(X), R_0^{-1}, R_1^{-1}\rangle$ is an isomorphism of $\mathbf{GLB}$-algebras.

2.3 Topological semantics

Our main interest in this paper is in topological semantics. Ordinarily, when modal logics are interpreted topologically, modal diamond is read as topological closure. However, as we already pointed out in the introduction, this only works if the logic in question contains the reflection axiom, since each set is a subset of its closure. For logics that do not contain the reflection axiom, of which $\mathbf{GL}$, $\mathbf{GLB}$, and $\mathbf{GLP}$ are all examples, $\Diamond$ can instead be interpreted as the derived set operator.

Definition 2.12. Let X be a topological space and $A \subseteq X$. We recall that $x \in X$ is a *limit point* of A if for each neighborhood U of x we have $A \cap (U - \{x\}) \neq \varnothing$. Let $d(A)$ denote the set of limit points of A. As usual, we call $d(A)$ the *derived set* of A. Obviously, the topological closure of A can then be defined as $\mathrm{cl}(A) = A \cup d(A)$ and topological interior as $\mathrm{int}(A) = A \cap t(A)$, where $t(A) := -d(-A)$.

Interpreting $\Diamond$ as a derived set operator provides an adequate semantics for $\mathbf{GL}$. Let X be a topological space and let $v : \mathrm{Var} \to 2^X$ be a valuation. We extend v to the set of all formulas by setting

- $v(\varphi \vee \psi) = v(\varphi) \cup v(\psi)$, $v(\neg\varphi) = -v(\varphi)$, $v(\top) = X$, $v(\bot) = \varnothing$,
- $v(\Diamond\varphi) = d(v(\varphi))$, $v(\Box\varphi) = t(v(\varphi))$.

We will also write $X, x \overset{\text{top}}{\vDash}_v \varphi$ for $x \in v(\varphi)$. When the valuation v is clear from the context. this can also be written as $X, x \overset{\text{top}}{\vDash} \varphi$.

Definition 2.13. A formula φ is *valid in* X (denoted $X \overset{\text{top}}{\vDash} \varphi$) if $\forall v, v(\varphi) = X$. The *logic of* X is the set of all formulas valid in X. If $\mathcal{C}$ is a class of spaces, the *logic of* $\mathcal{C}$ is the set of formulas valid in all members $X \in \mathcal{C}$.

Given a topological space X, we recall that $x \in X$ is an *isolated point* of X if $\{x\}$ is an open subset of X. Note that the set of isolated points of a subspace Y of X coincides with $Y - d(Y)$. We call X a *scattered space* if each nonempty subspace of X has an isolated point.

Theorem 2.14 ([13]). *A topological space X is scattered iff $X \overset{top}{\vDash}$ $\mathbf{GL}$; moreover,* $\mathbf{GL}$ *is the logic of the class of all scattered spaces.*

Typical examples of scattered spaces are ordinals (in the interval topology). Theorem 2.14 can be improved by showing that $\mathbf{GL}$ is the logic of all ordinals. In fact, $\mathbf{GL}$ is the logic of any ordinal $\alpha \geq \omega^\omega$:

Theorem 2.15 ([2, 10]). $\mathbf{GL}$ *is the logic of the class of all ordinals. In fact,* $\mathbf{GL}$ *is the logic of any ordinal $\alpha \geq \omega^\omega$. In particular,* $\mathbf{GL}$ *is the logic of ω^ω.*[2]

For the case of the polymodal language of $\mathbf{GLP}$ we consider *polytopological spaces*; that is, sets X equipped with a family of topologies $\{\tau_n\}_{n\in\omega}$. As our immediate task, we would like to understand which polytopological spaces satisfy all the axioms of $\mathbf{GLP}$.

[2] For a simplified proof of this result we refer to [9].

Let $\mathcal{X} = \langle X, \{\tau_n\}_{n\in\omega}\rangle$ be a polytopological space. Let, for each $n \in \omega$, d_n denote the derived set operator and t_n its dual with respect to τ_n. Theorem 2.14 tells us that each τ_n should be a scattered topology. Now we give necessary and sufficient conditions for axioms (iv) and (v) to be valid in $\mathcal{X}$.

Proposition 2.16. *Let $\mathcal{X} = \langle X, \{\tau_n\}_{n\in\omega}\rangle$ be a polytopological space and let $m < n$.*

1. *For each $A \subseteq X$ we have $d_m(A)$ is τ_n-open iff $d_m(A) \subseteq t_n(d_m(A))$.*
2. *$\tau_m \subseteq \tau_n$ iff $d_n(A) \subseteq d_m(A)$ for each $A \subseteq X$.*

Proof. (1) We have:

$$\begin{aligned} d_m(A) \text{ is } \tau_n\text{-open iff } & d_m(A) = \mathrm{int}_n(d_m(A)) \\ \text{iff } & d_m(A) = d_m(A) \cap t_n(d_m(A)) \\ \text{iff } & d_m(A) \subseteq t_n(d_m(A)). \end{aligned}$$

(2) Let $\tau_m \subseteq \tau_n$. Suppose that $A \subseteq X$, $x \in d_n(A)$, and U is a τ_m-open neighborhood of x. Then U is also a τ_n-open neighborhood of x, and so $A \cap (U - \{x\}) \neq \varnothing$, which implies that $x \in d_m(A)$. Conversely, let $\tau_m \not\subseteq \tau_n$. Then there exists $U \in \tau_m$ such that $U \notin \tau_n$. Since $U \notin \tau_n$, there exists $x \in U$ such that for each τ_n-open neighborhood V of x we have $V \cap -U \neq \varnothing$. Therefore, $U \cap d_n(-U) \neq \varnothing$ and yet $U \cap d_m(-U) = \varnothing$. Thus, $d_n(-U) \not\subseteq d_m(-U)$. □

Theorem 2.14 and Proposition 2.16 suggest the following definition of a **GLP**-space.

Definition 2.17. Let $\mathcal{X} = \langle X, \{\tau_n\}_{n\in\omega}\rangle$ be a polytopological space. We call $\mathcal{X}$ a **GLP**-*space* if

(i) Each τ_n is a scattered topology;

(ii) $\tau_n \subseteq \tau_{n+1}$;

(iii) $d_n(A)$ is τ_{n+1}-open for each $A \subseteq X$.

In particular, a bitopological space $\langle X, \tau_0, \tau_1\rangle$ is a **GLB**-*space* if both τ_0 and τ_1 are scattered topologies, $\tau_0 \subseteq \tau_1$, and $d_0(A)$ is τ_1-open for each $A \subseteq X$.

Note that, because of condition (ii), condition (i) can be weakened to the requirement that only τ_0 be scattered. From Theorem 2.14 and Proposition 2.16 we directly obtain:

Theorem 2.18. *A polytopological space* $\mathcal{X} = \langle X, \{\tau_n\}_{n\in\omega}\rangle$ *is a* **GLP**-*space iff* $\mathcal{X} \overset{top}{\vDash}$ **GLP**, *and a bitopological space* $\langle X, \tau_0, \tau_1\rangle$ *is a* **GLB**-*space iff* $\mathcal{X} \overset{top}{\vDash}$ **GLB**.

An obvious question is whether **GLP** (resp. **GLB**) is complete with respect to this semantics. But first we should be able to give examples of **GLP**-spaces (resp. **GLB**-spaces). Note that conditions (i) and (ii) are natural topological conditions and are easy to satisfy. On the other hand, condition (iii) is rather strong and somewhat unusual. Nevertheless, we will see shortly how to satisfy it.

Of course, if $\langle X, \tau_0\rangle$ is a scattered space and τ_1 is a discrete topology on X, then $\langle X, \tau_0, \tau_1\rangle$ is trivially a **GLB**-space. The first example of a **GLB**-space with two non-discrete topologies was given by Leo Esakia (private communication).

Example 2.19 (Esakia space)**.** Let α be an ordinal. Let τ_0 consist of all $<$-downsets and let τ_1 be the interval topology. It is easy to verify that both τ_0 and τ_1 are scattered topologies and that $\tau_0 \subset \tau_1$. Let $A \subseteq \alpha$. To see that $d_0(A)$ is τ_1-open observe that $d_0(A) = \{x \in \alpha : x > \min(A)\}$, which is clearly τ_1-open. Thus, $\langle \alpha, \tau_0, \tau_1\rangle$ is a **GLB**-space.

On the other hand, the next lemma shows that in order to define a third non-discrete topology on α, the ordinal should be very large. Recall that a topological space X is *first-countable* if every point $x \in X$ has a countable basis of open neighborhoods.

Proposition 2.20. *For any* **GLB**-*space* $\langle X, \tau_0, \tau_1\rangle$, *if* τ_0 *is Hausdorff and first-countable, then* τ_1 *is discrete.*

Proof. It is easy to see that if $\langle X, \tau_0\rangle$ is first-countable and Hausdorff, then every point $a \in X$ is a (unique) limit of a countable sequence of points $A = \{a_n\}_{n\in\omega}$. Hence, there is a set $A \subseteq X$ such that $d_0(A) = \{a\}$. By condition (iii), this means that $\{a\}$ is τ_1-open. □

Going back to $\langle \alpha, \tau_0, \tau_1\rangle$, observe that $\langle \alpha, \tau_1\rangle$ is always Hausdorff, and that $\langle \alpha, \tau_1\rangle$ is first-countable iff $\alpha \leq \omega_1$. Therefore, in order for us to be able to define a non-discrete τ_2 on α, the ordinal should be at least $\omega_1 + 1$. This is, in fact, sufficient as the following example shows.

Example 2.21 (club topology)**.** Recall that *cofinality* $\mathrm{cf}(\alpha)$ of a limit ordinal α is the least order type of an unbounded subset of α. If α is not a limit ordinal, we set $\mathrm{cf}(\alpha) = 0$. A set $A \subseteq \alpha$ is called a *club in* α if it is τ_1-closed (in the interval topology on α) and unbounded in α.

Define a topology τ_2 on α as follows: a set U is τ_2-open if, for each $\beta \in U$, either $\mathrm{cf}(\beta) \leq \omega$ or there is a club C in β such that $C \subseteq U$.

If $\mathrm{cf}(\beta) > \omega$, the intersection of countably many clubs in β is again a club. Hence, it is easy to check that τ_2 is indeed a topology. The filter of neighborhoods of β in τ_2 (restricted to β) coincides with the so-called *club filter* on β — a well-known concept in set theory (see [23]). Therefore, we call this topology the *club topology*.

Proposition 2.22. *$\langle \alpha, \tau_1, \tau_2 \rangle$ is a $\mathbf{GLB}$-space. In fact, the club topology τ_2 is the coarsest topology τ such that $\langle \alpha, \tau_1, \tau \rangle$ is a $\mathbf{GLB}$-space.*

Proof. To verify condition (iii) notice that a set of the form $d_1(A) \cap \beta$ is a club in any $\beta \in d_1(A)$. Hence, $d_1(A)$ is τ_2-open. The other conditions are obvious.

On the other hand, assume $\langle \alpha, \tau_1, \tau \rangle$ is a $\mathbf{GLB}$-space. We show that every τ_2-open neighborhood of any $\beta \in \alpha$ contains a τ-open neighborhood. If $\mathrm{cf}(\beta) \leq \omega$ then either β is isolated already in τ_1 (in the case β is not a limit ordinal), or β is a unique limit of an increasing ω-sequence A of ordinals. Then $\{\beta\} = d_1(A)$ and hence β is isolated in τ. If $\mathrm{cf}(\beta) > \omega$ and C is a club in β, then $d_1(C) \subseteq C \cup \{\beta\}$ is a τ-open neighborhood of β. $\square$

We are mainly interested in topological completeness of $\mathbf{GLP}$ and $\mathbf{GLB}$. Note that no Esakia space can be an exact model of $\mathbf{GLB}$. Looking at $\langle \alpha, \tau_0, \tau_1 \rangle$, observe that τ_0 consists of the $<$-downsets of α. Since α is a linear order, the linearity axiom $[0]([0]^+ p \to q) \vee [0]([0]^+ q \to p)$ is valid in $\langle \alpha, \tau_0, \tau_1 \rangle$, where $[0]^+\varphi$ is an abbreviation of $\varphi \wedge [0]\varphi$.

As far as the $\mathbf{GLB}$-space $\langle \alpha, \tau_1, \tau_2 \rangle$ is concerned, the situation is more complicated. We know that it is consistent with ZFC that $\mathbf{GLB}$ is incomplete with respect to this space. This follows from a result of Blass [10] who analyzed the question of completeness of $\mathbf{GL}$ with respect to the club topology τ_2.[3] In particular, he has shown that it is consistent with ZFC that $\mathbf{GL}$ is incomplete with respect to τ_2 on any ordinal. He has also shown that, under the assumption $V = L$, $\mathbf{GL}$ *is* complete with respect to the space $\langle \aleph_\omega, \tau_2 \rangle$. We conjecture that this result can be extended to a completeness result for $\mathbf{GLB}$ with respect to $\langle \aleph_\omega, \tau_1, \tau_2 \rangle$.

In the next section we will be able to prove topological completeness of $\mathbf{GLB}$ while standing firmly on the basis of ZFC. However, the question of topological (in)completeness of any fragment of $\mathbf{GLP}$ with more than two modalities remains open. At the least, our method of proving completeness of $\mathbf{GLB}$ does not immediately generalize to three or more modalities.

While the full $\mathbf{GLP}$, so far, eludes completeness, we note that the letterless fragment $\mathbf{GLP}_0$ allows for a simple topological treatment. Namely, $\mathbf{GLP}_0$ is sound and

[3] Blass did not introduce the topology explicitly, but formulated an equivalent semantics in terms of the club filter.

complete with respect to a natural polytopological space defined on the ordinal ϵ_0. This space, however, is not a **GLP**-space (see [17, 18]).

3 Topological Completeness of GLB

In this section we work in the language with two modalities $[0]$ and $[1]$. Before we prove our main result, we need a few auxiliary notions.

3.1 The logic J

Our proof of topological completeness will make use of a subsystem of **GLB** introduced in [5] and denoted **J**. This logic is defined by weakening axiom (iv) of **GLB** to the following axioms (vi) and (vii) both of which are theorems of **GLB**:

(vi) $[0]\varphi \to [1][0]\varphi$;

(vii) $[0]\varphi \to [0][1]\varphi$.

J is the logic of a simple class of frames, which is established by standard methods ([5, Theorem 1]).

Lemma 3.1. **J** *is sound and complete with respect to the class of (finite) frames* $\langle W, R_1, R_2\rangle$ *such that, for all* $x, y, z \in W$,

1. R_0 *and* R_1 *are transitive and dually well-founded;*
2. *If* xR_1y*, then* xR_0z *iff* yR_0z*;*
3. xR_0y *and* yR_1z *imply* xR_0z.

If we let $\overline{R_1}$ denote the reflexive, symmetric, transitive closure of R_1, then we call each $\overline{R_1}$ equivalence class a *1-sheet*. By (2), all points in a 1-sheet are R_0 incomparable. But R_0 defines a natural ordering on 1-sheets in the following sense: if α and β are 1-sheets, then $\alpha R_0 \beta$, iff $\exists x \in \alpha, \exists y \in \beta$, xR_0y. By standard techniques, one can improve on Lemma 3.1 to show that **J** is complete for such frames, in which each 1-sheet is a tree under R_1, and if $\alpha R_0 \beta$ then xR_0y for all $x \in \alpha$, $y \in \beta$ (see [5, Theorem 2 and Corollary 3.3]). Thus, models of **J** can be seen as R_0-orders (and even tree-like orders), in which the nodes are 1-sheets that are themselves R_1-trees. We call such frames *tree-like J-frames*.

As shown in [5], **GLB** is reducible to **J** in the following sense. Let

$$M(\varphi) := \bigwedge_{i<s}([0]\varphi_i \to [1]\varphi_i),$$

where $[0]\varphi_i$, $i < s$, are all subformulas of φ of the form $[0]\psi$. Also, let

$$M^+(\varphi) := M(\varphi) \wedge [0]M(\varphi) \wedge [1]M(\varphi).$$

Proposition 3.2 ([5]). $\mathbf{GLB} \vdash \varphi$ *iff* $\mathbf{J} \vdash M^+(\varphi) \to \varphi$.

This proposition generalizes straightforwardly to the case of **GLP**. In fact, we obtain another proof of this proposition, for the case of $\mathbf{GLB}$, as a byproduct of the topological completeness proof below.[4]

3.2 Some notions related to partial orderings

Let $\langle X, \prec\rangle$ be a dually well-founded strict partial ordering. We consider bitopological spaces of the form $\langle X, \tau_0, \tau_1\rangle$, where τ_0 is the upset topology on $\langle X, \prec\rangle$ and τ_1 is generated by all semi-open intervals of the form

$$[a, b) := \{x \in X : a \preceq x \prec b\}$$

for $a \prec b$, and

$$[a, \infty) := \{x \in X : a \preceq x\}.$$

Notice that if $\langle X, \prec\rangle$ is a strict linear ordering, then τ_1 is the usual interval topology on X, and thus $\langle X, \tau_0, \tau_1\rangle$ is the Esakia space of the ordinal dual to $\langle X, \prec\rangle$.

Lemma 3.3. $\langle X, \tau_0, \tau_1\rangle$ *is a* $\mathbf{GLB}$*-space.*

Proof. Clearly $\tau_0 \subseteq \tau_1$. We show that sets of the form $d_0(A)$ are τ_1-open for any $A \subseteq X$. Let $\max(A)$ denote the set of maximal points of A. Since $\langle X, \prec\rangle$ is dually well-founded, $d_0(A)$ consists of all points below $\max(A)$; that is,

$$d_0(A) = \bigcup_{a \prec b \in \max(A)} [a, b).$$

Hence, $d_0(A)$ is a union of τ_1-open sets. □

We call such $\mathbf{GLB}$-spaces *general Esakia spaces*.

Next, we recall a few standard operations on strict partial orderings.[5] The *disjoint union* of the orderings X and Y is denoted $X \sqcup Y$. The *sum* of X and Y is denoted $X + Y$; that is, the ordering is obtained by putting Y on top of X. In particular,

[4] It is worth noting that Ignatiev's proof of *arithmetical* completeness of $\mathbf{GLP}$ establishes a similar reduction of $\mathbf{GLP}$ to a different frame complete subsystem of $\mathbf{GLP}$.

[5] The notations we use are dual to those given in [5], but they are more in line with the standard usage.

when X is a singleton $\{a\}$, $\{a\} + Y$ denotes the result of adding a new node at the bottom of Y.

A more general operation of *ordered sum* of a family $\{\mathcal{A}_i : i \in I\}$ of orderings $\mathcal{A}_i = \langle A_i, \prec_i \rangle$, where $\langle I, \prec \rangle$ is a strict partially ordered index set, is the ordering $\langle Y, \prec_Y \rangle$ such that $Y = \bigsqcup_{i \in I} A_i$. For $x, y \in Y$, we declare $x \prec_Y y$ iff either $x, y \in A_i$ and $x \prec_i y$ for some $i \in I$; or $x \in A_i$ and $y \in A_j$ for some $i \prec j$. We denote this ordering by $\sum_{i \in I} \mathcal{A}_i$. In particular, if $\langle I, \prec \rangle$ is the ordering $\langle \omega, > \rangle$ and all $\mathcal{A}_i$ are isomorphic to the same ordering $\mathcal{A}$, the ordering $\sum_{i \in I} \mathcal{A}$ consists of countably many copies of $\mathcal{A}$ ordered by ω^* and is denoted $\mathcal{A} \cdot \omega^*$.

3.3 Topological completeness theorem

Theorem 3.4 (Main Theorem)**. GLB** *is complete w.r.t. the class of general Esakia spaces.*

Proof. Assume $\mathbf{GLB} \nvdash \varphi$. Consider a finite tree-like J-model $\mathcal{A}$ such that $\mathcal{A} \nvDash M^+(\varphi) \to \varphi$. We denote by Greek letters $\alpha, \beta, \dots$ the elements of $\mathcal{A}$.

Following [5], we associate with $\mathcal{A}$ a strict partial ordering called the *topological blow-up of* $\mathcal{A}$. First, we associate with each 1-sheet $\mathcal{S}$ of $\mathcal{A}$ a strict partial ordering $\mathcal{S}^\omega$ by induction on the R_1-depth of $\mathcal{S}$. Second, we consider the set $\mathbf{S}(\mathcal{A})$ of all 1-sheets of $\mathcal{A}$ ordered by R_0 and take the ordered sum of orders $\mathcal{S}^\omega$ with respect to this index set. This idea is expressed by the following two formal definitions.

Definition 3.5.

- If $\mathcal{A}_\alpha = \langle A_\alpha, R_1 \rangle$ is a tree with the root α, define a strict partial ordering $\mathcal{A}_\alpha^\omega$ by induction on the depth of α:

$$\mathcal{A}_\alpha^\omega := \{\alpha\} + \Big(\bigsqcup_{i=1}^{n} \mathcal{A}_{\alpha_i}^\omega\Big) \cdot \omega^*,$$

  where α_i are all the R_1-children of α. $\mathcal{A}_\alpha^\omega := \{\alpha\}$ if $\mathcal{A}_\alpha$ is the singleton $\{\alpha\}$.

- $\mathfrak{B}_\omega(\mathcal{A}) := \sum_{\mathcal{S} \in \mathbf{S}(\mathcal{A})} \mathcal{S}^\omega$.

The ordering $\mathfrak{B}_\omega(\mathcal{A})$ is called the *topological blow-up of* $\mathcal{A}$ and will define the general Esakia space we seek. The order relation on $\mathfrak{B}_\omega(\mathcal{A})$ will be denoted $\prec$; τ_0 and τ_1 are the topologies of the associated general Esakia space; d_0 and d_1 are the corresponding derived set operators.

It is worth noting that the blow-up construction here is much simpler than the one in [5] for two main reasons. Firstly, we only deal with the case of two modalities which avoids the iterative process involved in [5] and the complicated limit construction. Secondly, the type of the resulting structure is simpler (it is just a strict partial order) and, in addition, it needs fewer new points. The latter seems to be a helpful feature of the topological semantics we consider compared to relational semantics.

Next, we make a couple observations about the defined structures. Firstly, there is a natural embedding of $\mathcal{A}^\omega_\beta$ as an upset into $\mathcal{A}^\omega_\alpha$ whenever $\alpha \prec \beta$. This is easy to verify by induction on R_1-depth of α. Secondly, a natural *projection map* $\pi_\alpha : \mathcal{A}^\omega_\alpha \to \mathcal{A}_\alpha$ is defined inductively as follows: if $x \in \mathcal{A}^\omega_{\alpha_i}$, then $\pi_\alpha(x) := \pi_{\alpha_i}(x)$; otherwise, $\pi_\alpha(x) := \alpha$. This extends to a map $\pi : \mathfrak{B}_\omega(\mathcal{A}) \to \mathcal{A}$ in the obvious way.

Lemma 3.6.

1. *Assume $x \in \mathcal{A}^\omega_\alpha$ and $\pi_\alpha(x) R_1 y$ in $\mathcal{A}_\alpha$. Then there is a sequence $(x_n)_{n\in\omega} \in \mathcal{A}^\omega_\alpha$ such that $x \in d_1(\{x_n : n \in \omega\})$ and $\pi_\alpha(x_n) = y$ for all $n \in \omega$.*

2. *For all $x, y \in \mathfrak{B}_\omega(\mathcal{A})$, if $\pi(x) R_1 y$, then $x \in d_1(\pi^{-1}(y))$.*

Proof. (1) We argue by induction on the R_1-depth of α. If α has depth 0, the claim is trivial (no such x, y exist). Otherwise, $\mathcal{A}^\omega_\alpha = \{\alpha\} + \left(\bigsqcup_{i=1}^n \mathcal{A}^\omega_{\alpha_i}\right) \cdot \omega^*$.

If x belongs to some copy of $\mathcal{A}^\omega_{\alpha_i}$, we can select a sequence x_n in (the same copy of) $\mathcal{A}^\omega_{\alpha_i}$ by the induction hypothesis. We obviously have that $\pi_\alpha(x_n) = y$ by the definition of π_α. Also, $x \in d_1(\{x_n : n \in \omega\})$ in $\mathcal{A}^\omega_{\alpha_i}$. Since $\langle \mathcal{A}^\omega_{\alpha_i}, \tau_1\rangle$ is a subspace of $\langle \mathcal{A}^\omega_\alpha, \tau_1\rangle$ (any interval in one space is an interval in the other), we also have $x \in d_1(\{x_n : n \in \omega\})$ in $\mathcal{A}^\omega_\alpha$.

If x is the root of $\mathcal{A}^\omega_\alpha$, then $\pi_\alpha(x) = \alpha$. Suppose y is an immediate successor of α. Then $\mathcal{A}^\omega_\alpha$ contains a sequence of copies of $\mathcal{A}^\omega_y$ the roots of which converge to x. Otherwise, let β be the son of α such that $\beta \prec y$. Select an element $z \in \mathcal{A}^\omega_\beta$ such that $\pi_\beta(z) = y$. Let z_n be the element corresponding to z within the n-th copy of $\mathcal{A}^\omega_\beta$ above x. Then z_n's converge to x in $\mathcal{A}^\omega_\alpha$.

(2) If $\pi(x) R_1 y$, then $\pi(x), y$ belong to the same 1-sheet $\mathcal{A}_\alpha$, $x \in \mathcal{A}^\omega_\alpha$ and $\pi = \pi_\alpha$ on $\mathcal{A}^\omega_\alpha$. Hence, one can apply (1) and obtain a sequence (x_n) in $\mathcal{A}^\omega_\alpha$ such that $x_n \in \pi^{-1}(y)$ and $x \in d_1(\{x_n : n \in \omega\})$. □

Lemma 3.7. *For all $x, y \in \mathfrak{B}_\omega(\mathcal{A})$, if $x \in d_1(Y)$, then $\pi(x) R_1 \pi(y)$ for infinitely many $y \in Y$.*

Proof. Let $\mathcal{A}_\alpha$ be the 1-sheet of $\pi(x)$. If $x \in d_1(Y)$, then Y is infinite because τ_1 is a T_1-topology; that is, each finite set is closed. Since $\mathcal{A}^\omega_\alpha$ is a semi-open interval in $\mathfrak{B}_\omega(\mathcal{A})$, there are infinitely many $y \in Y$ such that $y \in \mathcal{A}^\omega_\alpha$. Without loss of generality assume that this holds for all $y \in Y$ and that $x \notin Y$. We prove that $\pi(x) R_1 \pi(y)$ for infinitely many $y \in Y$ by induction on the R_1-depth of α.

If the depth of α is 0, then $\mathcal{A}^{\omega}_{\alpha} = \{\alpha\}$. Therefore, all $y \in Y$ must coincide with α, contradicting that Y is infinite. Otherwise, $\mathcal{A}^{\omega}_{\alpha} = \{\alpha\} + \left(\bigsqcup_{i=1}^{n} \mathcal{A}^{\omega}_{\alpha_i}\right) \cdot \omega^*$.

Suppose x belongs to some copy of $\mathcal{A}_{\alpha_i}$. Since this copy is a semi-open interval in $\mathcal{A}^{\omega}_{\alpha}$, infinitely many $y \in Y$ are in this interval. By induction hypothesis, $\pi(x) R_1 \pi(y)$ for infinitely many $y \in Y$.

If x is the root of $\mathcal{A}^{\omega}_{\alpha}$, then $\pi(x) = \pi_{\alpha}(x) = \alpha$. If $y \in Y$ then $y \neq x$ by assumption, and by the construction of $\mathcal{A}^{\omega}_{\alpha}$, $\pi(y) \neq \pi(x) = \alpha$. Since $\pi(y) \in \mathcal{A}_{\alpha}$ and α is the minimum of $\mathcal{A}_{\alpha}$, we have $\alpha R_1 \pi(y)$. □

We define a valuation $v : \mathrm{Var} \to 2^{\mathfrak{B}_{\omega}(\mathcal{A})}$ by

$$x \in v(p) \leftrightarrow \mathcal{A}, \pi(x) \vDash p.$$

Lemma 3.8. *For each subformula ψ of φ,*

$$\mathfrak{B}_{\omega}(\mathcal{A}), x \overset{top}{\vDash} \psi \leftrightarrow \mathcal{A}, \pi(x) \vDash \psi.$$

Proof. By induction on the build-up of ψ. We only treat the cases of modalities. Let $X := \mathfrak{B}_{\omega}(\mathcal{A})$ and $v(\psi) := \{x \in X : X, x \overset{\text{top}}{\vDash} \psi\}$.

1. Suppose $\mathcal{A}, \pi(x) \vDash \langle 1 \rangle \psi$. Then there is a y such that $\pi(x) R_1 y$ and $\mathcal{A}, y \vDash \psi$. Since $\pi(x) R_1 y$, we have $x \in d_1(\pi^{-1}(y))$. By inductive hypothesis, $\pi^{-1}(y) \subseteq v(\psi)$, hence $x \in d_1(v(\psi))$ and $X, x \overset{\text{top}}{\vDash} \langle 1 \rangle \psi$.

2. Suppose $X, x \overset{\text{top}}{\vDash} \langle 1 \rangle \psi$. Then $x \in d_1(v(\psi))$. Setting $Y := v(\psi)$, by Lemma 3.7, there is a $y \in Y$ such that $\pi(x) R_1 \pi(y)$. By inductive hypothesis, $\mathcal{A}, \pi(y) \vDash \psi$, hence $\mathcal{A}, \pi(x) \vDash \langle 1 \rangle \psi$.

3. If $\mathcal{A}, \pi(x) \vDash \langle 0 \rangle \psi$, then $\exists y \, (\pi(x) R_0 y \ \& \ \mathcal{A}, y \vDash \psi)$. Since π is a p-morphism, there is a $y' \succ x$ such that $\pi(y') = y$. This yields $X, y' \overset{\text{top}}{\vDash} \psi$ and $X, x \overset{\text{top}}{\vDash} \langle 0 \rangle \psi$.

4. If $X, x \overset{\text{top}}{\vDash} \langle 0 \rangle \psi$, then $\exists y \, (x \prec y \ \& \ X, y \overset{\text{top}}{\vDash} \psi)$. We have $\pi(x) R_0 \pi(y)$ or $\pi(x)$,$\pi(y)$ belong to the same 1-sheet. In the first case we are done. In the second case, let α be the R_1-maximal point such that $x \in \mathcal{A}^{\omega}_{\alpha}$, and let z be the $\prec$-minimal point of $\mathcal{A}^{\omega}_{\alpha}$. Obviously $\pi(z) = \alpha$.

Notice that $\mathcal{A}^{\omega}_{\beta}$ is an upwards closed submodel of $\mathcal{A}_{\alpha}$ whenever $\alpha R_1 \beta$. Then, since $x \in \mathcal{A}^{\omega}_{\alpha}$, we must also have $y \in \mathcal{A}^{\omega}_{\alpha}$; hence $z \prec y$. Since $\pi(z) = \alpha$ and $z \prec y$, we have $\pi(z) R_1 \pi(y)$. By inductive hypothesis, $\mathcal{A}, \pi(y) \vDash \psi$ and hence $\mathcal{A}, \pi(z) \vDash \langle 1 \rangle \psi$. By the monotonicity axioms in $\mathcal{A}$ this yields $\mathcal{A}, \pi(z) \vDash \langle 0 \rangle \psi$.

Since $\pi(x)$ belongs to the same 1-sheet as $\alpha = \pi(z)$, we also have $\mathcal{A}, \pi(x) \vDash \langle 0 \rangle \psi$. $\square$

Hence, we obtain $\mathfrak{B}_\omega(\mathcal{A}) \overset{\text{top}}{\nvDash} \varphi$, which proves the theorem. $\square$

Corollary 3.9. $\mathbf{GLB} \vdash \varphi$ *iff* $\mathbf{J} \vdash M^+(\varphi) \to \varphi$.

Proof. The non-trivial implication from left to right follows from the proof of topological completeness theorem. We have shown that if the conclusion is false, there must exist a $\mathbf{GLB}$-space falsifying φ, hence $\mathbf{GLB} \nvdash \varphi$. $\square$

4 Discussion

We have established topological completeness results for two fragments of $\mathbf{GLP}$: for the bimodal fragment $\mathbf{GLB}$, and for the the letterless fragment $\mathbf{GLP}_0$ (see [18]). There are some questions that remain open, which we summarize below.

1. Is $\mathbf{GLP}$ topologically complete?
2. Is $\mathbf{GLB}$ complete with respect to the $\mathbf{GLB}$-space $\langle \alpha, \tau_1, \tau_2 \rangle$, for some ordinal α, under the assumption $V = L$? (Here, τ_1 is the interval topology and τ_2 is the club topology on α.)
3. There is a natural notion of an *ordinal* $\mathbf{GLP}$*-space*. Consider a space of the form $\langle \alpha, \{\tau_n\}_{n \in \omega} \rangle$, where τ_1 is the interval topology on α, and τ_{n+1} is generated from τ_n and all sets of the form $d_n(A)$, for $A \subseteq \alpha$. Is $\mathbf{GLP}$ complete with respect to some ordinal $\mathbf{GLP}$-space?

From the results of Blass (see our discussion of Problem 2 at the end of Section 2.3) we know that a positive answer to Problem 3 would require some set-theoretic assumptions outside ZFC. Some partial results in this direction have already been obtained. In particular, we know that the assumption that the third topology τ_3 of an ordinal $\mathbf{GLP}$-space is nontrivial is equiconsistent with the existence of a weakly compact cardinal. In other words, non-discreteness of τ_3 (and similarly for further topologies τ_n) is a large cardinal assumption. We do not know the exact consistency strength of this assumption for $n > 3$. However, we know a reasonable sufficient condition for all τ_n to be non-discrete — the existence of the so-called Π^1_n-indescribable cardinals for each $n \in \omega$.[6] Therefore, it is hopeful to

[6]The first author thanks Philipp Schlicht for finding this condition and for further advice on set theory involved here.

obtain completeness of **GLP** with respect to an ordinal **GLP**-space if we simultaneously assume things like $V = L$ and the existence of Π^1_n-indescribable cardinals. These results, in fact, show that there are deeper connections between the theory of ordinal **GLP**-spaces and parts of set theory dealing with infinitary combinatorics and stationary reflection.

Acknowledgements. The first author initiated the work on this paper during his stay at the University of Münster in the summer of 2008 at the invitation of Wolfram Pohlers. He acknowledges support from Alexander von Humboldt Foundation and would like to thank Wolfram Pohlers, Ralf Schindler and the whole logic group in Münster for their hospitality and for the fruitful atmosphere they have created.

References

[1] M. Abashidze. Some properties of Magari algebras. In *Studies in logic and semantics*, pages 111–127. "Metsniereba", Tbilisi, 1981. In Russian.

[2] M. Abashidze. Ordinal completeness of the Gödel-Löb modal system. In *Intensional logics and the logical structure of theories*, pages 49–73. Metsniereba, Tbilisi, 1985. In Russian.

[3] L.D. Beklemishev. Provability algebras and proof-theoretic ordinals, I. *Annals of Pure and Applied Logic*, 128:103–123, 2004.

[4] L.D. Beklemishev. Reflection principles and provability algebras in formal arithmetic. *Uspekhi Matematicheskikh Nauk*, 60(2):3–78, 2005. In Russian. English translation in: *Russian Mathematical Surveys*, 60(2): 197–268, 2005.

[5] L.D. Beklemishev. Kripke semantics for provability logic GLP. *Annals of Pure and Applied Logic*, 161: 756–774, 2010. Preprint: Logic Group Preprint Series 260, University of Utrecht, November 2007. `http://preprints.phil.uu.nl/lgps/`.

[6] L.D. Beklemishev, J. Joosten, and M. Vervoort. A finitary treatment of the closed fragment of Japaridze's provability logic. *Journal of Logic and Computation*, 15(4):447–463, 2005.

[7] J. van Benthem and G. Bezhanishvili. Modal logics of space. In *Handbook of spatial logics*, pages 217–298. Springer, Dordrecht, 2007.

[8] C. Bernardi and P. d'Aquino. Topological duality for diagonalizable algebras. *Notre Dame Journal of Formal Logic*, 29(3):345–364, 1988.

[9] G. Bezhanishvili and P. J. Morandi. Scattered and hereditarily irresolvable spaces in modal logic. *Archive for Mathematical Logic*, 49(3):343–365, 2010.

[10] A. Blass. Infinitary combinatorics and modal logic. *Journal of Symbolic Logic*, 55(2):761–778, 1990.

[11] G. Boolos. *The Logic of Provability*. Cambridge University Press, Cambridge, 1993.

[12] W. Buszkowski and T. Prucnal. Topological representation of co-diagonalizable algebras. In G. Wechsung, editor, *Frege Conference' 1984*, pages 63–65, Berlin, 1984. Akademie-Verlag.

[13] L. Esakia. Diagonal constructions, Löb's formula and Cantor's scattered spaces. In *Studies in logic and semantics*, pages 128–143. Metsniereba, Tbilisi, 1981. In Russian.

[14] L. Esakia. Intuitionistic logic and modality via topology. *Annals of Pure and Applied Logic*, 127:155–170, 2003.

[15] M. Gerson. An extension of $S4$ complete for the neighbourhood semantics but incomplete for the relational semantics. *Studia Logica*, 34(4):333–342, 1975.

[16] M. Gerson. A neighbourhood frame for T with no equivalent relational frame. *Z. Math. Logik Grundlagen Math.*, 22(1):29–34, 1976.

[17] T.F. Icard, III. Models of the polymodal provability logic. M.Sc. Thesis, ILLC, University of Amsterdam, http://www.illc.uva.nl/Publications/ ResearchReports/MoL-2008-06.text.pdf, 2008.

[18] T.F. Icard, III. A topological study of the closed fragment of GLP. *Journal of Logic and Computation*, Advance Access published online on August 12, 2009, doi:10.1093/logcom/exp043.

[19] K.N. Ignatiev. The closed fragment of Dzhaparidze's polymodal logic and the logic of Σ_1-conservativity. ITLI Prepublication Series X–92–02, University of Amsterdam, 1992.

[20] K.N. Ignatiev. On strong provability predicates and the associated modal logics. *The Journal of Symbolic Logic*, 58:249–290, 1993.

[21] G.K. Japaridze. The modal logical means of investigation of provability. Thesis in Philosophy, in Russian, Moscow, 1986.

[22] G.K. Japaridze. The polymodal logic of provability. In *Intensional Logics and Logical Structure of Theories: Materials from the fourth Soviet–Finnish Symposium on Logic, Telavi, May 20–24, 1985*, pages 16–48. Metsniereba, Tbilisi, 1988. In Russian.

[23] T. Jech. *Set Theory. The Third Millenium Edition*. Springer, 2002.

[24] R. Magari. The diagonalizable algebras (the algebraization of the theories which express Theor.:II). *Bollettino della Unione Matematica Italiana,* Serie 4, 12, 1975. Suppl. fasc. 3, 117–125.

[25] J. C. C. McKinsey and A. Tarski. The algebra of topology. *Annals of Mathematics*, 45:141–191, 1944.

[26] V. Shehtman. On neighbourhood semantics thirty years later. In S. Artemov et al., editors, *We Will Show Them! Essays in Honour of Dov Gabbay, v.2*, pages 663–692. King's College Publications, 2005.

Program Extraction via Typed Realisability for Induction and Coinduction

Ulrich Berger and Monika Seisenberger

Swansea University, Swansea, SA2 8PP, Wales, UK
{u.berger,m.seisenberger}@swansea.ac.uk

Abstract We study a realisability interpretation for inductive and coinductive definitions and discuss its application to program extraction in constructive analysis. A speciality of this interpretation is that realisers are given by terms that correspond directly to programs in a lazy functional programming language such as Haskell.

1 Introduction

In this paper we give a realisability interpretation for a constructive theory of strictly positive inductive and coinductive definitions. The motivation is to provide a theoretical foundation for ongoing work on program extraction from proofs involving such definitions.

Our theory is an extension of intuitionistic first-order predicate logic with predicate variables and the definition of predicates as least and greatest fixed points of strictly positive operators. Since operators may depend strictly positively on other free predicate variables, these definitions may "interleave". An example of an interleaved inductive/coinductive definition is the predicate C_1, discussed in the conclusion of this paper, which characterises uniformly continuous functions on the real interval $[-1, 1]$. In the context of classical propositional modal logic a system allowing similar interleavings is known as the μ-calculus [BS07]. Möllerfeld [MÖ3] studied the first-order version of the μ-calculus, which is equivalent to the classical version of our system, and proved that it has the enormous proof-theoretic strength of Π^1_2-comprehension. Tupailo [Tup04] showed that the latter system can be embedded into its intuitionistic counterpart via a double-negation translation – hence preserving the proof-theoretic strength – however at the cost of introducing non-strictly positive inductive definitions. If one forbids interleavings, one obtains the proof-theoretically weaker system $\mathrm{ID}^{<\omega}$ of finitely iterated inductive definitions [BFPS81].

In the present paper we are concerned with the application of our theory to program extraction via realisability. The realisability interpretation we are going to study is related to interpretations given by Tatsuta [Tat98] and Miranda-Perea [MP05]. We try to point out the main similarities and differences. While Miranda-Perea extracts terms in a strongly normalising extension of the second-order polymorphic λ-calculus with "Mendler-style" (co)inductive types [Men91, Mat01, AMU05] (see also related work by Krivine and Parigot [KP90, Par92]), our realisers are taken from a λ-calculus with full recursion, ML-style polymorphic and recursive types and a call-by-name operational semantics. Hence our realisers can be directly understood as programs in a lazy functional programming language such as, for example, Haskell. Terms do not terminate in general, but those realising a formula do. Tatsuta's realisers can be viewed as an untyped version of ours. However, he works with realisability with truth whilst we omit the "truth" component.

From a practical point of view the most important difference to Tatsuta's interpretation is that we treat quantifiers uniformly in the realisability interpretation (as Miranda does): $M\,\mathbf{r}\,\forall x\,A(x)$ is defined as $\forall x\,(M\,\mathbf{r}\,A(x))$ (but not $\forall x\,(M\,x\,\mathbf{r}\,A(x))$ and $M\,\mathbf{r}\,\exists x\,A(x)$ is defined as $\exists x\,(M\,\mathbf{r}\,A(x))$ (but not $\pi_2(M)\,\mathbf{r}\,A(\pi_1(M))$). In general, a realiser never depends on variables of the object language in that language, i.e. the object language and the language of realisers are kept strictly separate. Realisers are extracted exclusively from the "propositional skeleton"of a proof ignoring the first-order part, the latter being important for the *correctness* of the realisers only. This widens the scope of applications considerably because it allows to deal with abstract structures that are not necessarily "constructively" given. Our uniform treatment of first-order quantifiers can also be seen as a special case of the interpretations studied by Schwichtenberg [Sch08], Hernest and Oliva [HO08] and Ratiu and Trifonov [RT10], which allow for a fine control of the amount of computational information extracted from proofs.

2 Inductive and coinductive definitions

We fix a first-order language $\mathcal{L}$. *Terms*, $r, s, t \ldots$, are built from constants, first-order variables and function symbols as usual. *Formulas*, $A, B, C \ldots$, are $s = t$, $\mathcal{P}(\vec{t})$ where $\mathcal{P}$ is a predicate (see below), $A \wedge B$, $A \vee B$, $A \to B$, $\forall x\,A$, $\exists x\,A$. A *predicate* is either a predicate constant P, or a predicate variable X, or a comprehension term $\{\vec{x} \mid A\}$, or an inductive predicate $\mu X.\mathcal{P}$, or a coinductive predicate $\nu X.\mathcal{P}$ where $\mathcal{P}$ is a predicate of the same arity as the predicate variable X and which is *strictly positive (s.p.)* in X, i.e. X does not occur free in any premise of a subformula of $\mathcal{P}$ which is an implication. The application, $\mathcal{P}(\vec{t})$, of a predicate $\mathcal{P}$ to a list of terms $\vec{t}$ is a primitive syntactic construct, except when $\mathcal{P}$ is a comprehension term, $\mathcal{P} = \{\vec{x} \mid A\}$, in which case $\mathcal{P}(\vec{t})$ stands for $A[\vec{t}/\vec{x}]$.

We will sometimes use the notation $\vec{x} \in \mathcal{P}$ for of $\mathcal{P}(\vec{x})$, $\mathcal{P} \subseteq \mathcal{Q}$ for $\forall \vec{x}\,(\mathcal{P}(\vec{x}) \to \mathcal{Q}(\vec{x}))$ and $\mathcal{P} \cap \mathcal{Q}$ for $\{\vec{x} \mid \mathcal{P}(\vec{x}) \ \wedge\ \mathcal{Q}(\vec{x})\}$ etc. We also write $\{t \mid A\}$ as an abbreviation for $\{x \mid \exists \vec{y}\,(x = t \ \wedge\ A)\}$ where x is a fresh variable and $\vec{y} = \mathrm{FV}(t) \cap \mathrm{FV}(A)$, as well as $f(\mathcal{P})$ for $\{f(x) \mid x \in \mathcal{P}\}$. Furthermore, we introduce *operators* $\Phi := \lambda \vec{X}.\mathcal{P}$, and write $\Phi(\vec{\mathcal{Q}})$ for the predicate $\mathcal{P}[\vec{\mathcal{Q}}/\vec{X}]$ where the latter is the usual substitution of the predicates $\vec{\mathcal{Q}}$ for the predicate variables $\vec{X}$. Φ is called a *s.p. operator* if $\mathcal{P}$ is s.p. in X. In this case we also write $\mu\Phi$ and $\nu\Phi$ for $\mu X.\mathcal{P}$ and $\nu X.\mathcal{P}$. A formula, predicate, or operator is called *non-computational*, if it contains neither free predicate variables nor the propositional connective $\vee$ nor the construct ν (formation of a greatest fixed point). Otherwise it is called *computational*.

The *proof rules* are the usual ones of intuitionistic predicate calculus with equality augmented by rules expressing that $\mu\Phi$ and $\nu\Phi$ are the least and greatest fixed points of the operator Φ. As is well-known, the fixed point property can be replaced by appropriate inclusions. Hence we stipulate the axioms

Closure	$\Phi(\mu\Phi) \subseteq \mu\Phi$	Induction	$\Phi(\mathcal{Q}) \subseteq \mathcal{Q} \to \mu\Phi \subseteq \mathcal{Q}$
Coclosure	$\nu\Phi \subseteq \Phi(\nu\Phi)$	Coinduction	$\mathcal{Q} \subseteq \Phi(\mathcal{Q}) \to \mathcal{Q} \subseteq \nu\Phi$

for all s.p. operators Φ and predicates $\mathcal{Q}$. In addition we allow any axioms expressible by non-computational formulas that hold in the intended model. We write $\Gamma \vdash A$ if A is derivable from assumptions in Γ in this system. If A is derivable without assumptions we write $\vdash A$, or even just A. Falsity can be defined as $\bot := \mu X.X$ where X is a propositional variable (i.e. a 0-ary predicate variable).

From the induction axiom for $\bot$ follows $\bot \to A$ for every formula A.

Lemma 2.1 (Instantiation)**.** *If* $\Gamma(X) \vdash A(X)$*, then* $\Gamma(\mathcal{P}) \vdash A(\mathcal{P})$.

Proof. Straightforward induction on derivations. □

Lemma 2.2 (Monotonicity)**.** *Let* Φ*,* Ψ *be s.p. operators,* $\mathcal{P}$*,* $\mathcal{Q}$ *predicates,* Γ *a context and* X *a predicate variable not free in* Γ.

(a) If $\Gamma \vdash \Phi(X) \subseteq \Psi(X)$*, then* $\Gamma \vdash \mu\Phi \subseteq \mu\Psi$ *and* $\Gamma \vdash \nu\Phi \subseteq \nu\Psi$.

(b) $\mathcal{P} \subseteq \mathcal{Q} \vdash \Phi(\mathcal{P}) \subseteq \Phi(\mathcal{Q})$.

Proof. (a) Assume Γ. We show $\mu\Phi \subseteq \mu\Psi$ using the Induction Axiom. Hence, we have to show $\Phi(\mu\Psi) \subseteq \mu\Psi$. By the hypothesis of the lemma, the Instantiation Lemma 2.1, and the closure axiom, we have $\Phi(\mu\Psi) \subseteq \Psi(\mu\Psi) \subseteq \mu\Psi$. The proof for ν is similar.

(b) Straightforward induction on the built-up of Φ, using (a) in the case of inductive and coinductive predicates. $\square$

Lemma 2.3 (Fixed Point)**.** *Let* Φ *be an operator.*

(a) $\Phi(\mu\Phi) = \mu\Phi$.

(b) $\Phi(\nu\Phi) = \nu\Phi$.

Proof. Because of the closure axiom, it suffices for (a) to show $\mu\Phi \subseteq \Phi(\mu\Phi)$. We use the induction rule. Thus it suffices to show $\Phi(\Phi(\mu\Phi)) \subseteq \Phi(\mu\Phi)$. But this follows from the closure axiom and the Monotonicity Lemma 2.2. The proof of (b) is similar. $\square$

As a running example we use the first-order language of the ordered real numbers. As axioms we adopt any non-computational formulas that are true in the structure of real numbers, e.g. the axioms of a real closed field where the linearity of the order is expressed non-computationally, e.g. by $\forall x, y\,(y \not< x \;\wedge\; x \not< y \to x = y)$. All sets we define in the following are subsets of the set of real numbers. We define the set $\mathbb{N}$ of natural numbers as usual inductively by

$$\mathbb{N} := \mu X.\{0\} \cup \{x + 1 \mid X(x)\}$$

Next we define coinductively a set which, as we will see later, is closely connected to the signed digit representation of real numbers. First we define the set of signed binary digits by $\mathrm{SD} := \{0, 1, -1\} = \{i \mid i = 0 \;\vee\; i = 1 \;\vee\; i = -1\}$. Now we define coinductively

$$\mathrm{C}_0 := \nu X.\{(i + x)/2 \mid \mathrm{SD}(i) \;\wedge\; X(x) \;\wedge\; |(i + x)/2| \le 1\}$$

It is easy to see that, classically, C_0 coincides with the closed interval $\mathbb{I} := [-1, 1]$. The point is that from a constructive proof of $\mathrm{C}_0(x)$ we can extract a program computing an infinite signed digit representation of x.

3 An idealised functional programming language

In this section we introduce an extended λ-calculus which we will use as the language of realisers in Sect. 4. For this calculus we define a denotational and an operational semantics and relate the two by an Adequacy Theorem. We also introduce types and define typable map operators, iterators and coiterators that will serve as realisers of monotonicity, induction and coinduction.

3.1 The untyped language

First, we introduce an untyped λ-calculus with constructors, pattern matching and recursion. Its terms are generated by the following formation rules.

Variables: $x, y, z, \ldots$.

Constructor terms: $C(M_1, \ldots, M_n)$ where C is taken from a set $\mathcal{C}$ of constructors each of which has a fixed arity and $M_1, \ldots, M_n$ are terms.

Case analysis: $\mathrm{case}\, M\, \mathrm{of}\{C_1(\vec{x_1}) \to R_1\,;\, \ldots;\, C_n(\vec{x_n}) \to R_n\}$ where M, $R_1, \ldots R_n$ are terms, the C_i are distinct constructors and each $\vec{x_i}$ is a vector of distinct variables.

λ-abstraction: $\lambda x.M$ where x is a variable and M is a term.

Application: $M\,N$ where M and N are terms.

Recursion: $\mathrm{rec}\, x \,.\, M$ where x is a variable and M is a term.

The free variables $\mathrm{FV}(M)$ of a term M is defined as expected, for example $\mathrm{FV}(\mathrm{case}\, M\, \mathrm{of}\{C_1(\vec{x_1}) \to R_1\,;\, \ldots\}) = \mathrm{FV}(M) \cup \bigcup_i(\mathrm{FV}(R_i) \setminus \vec{x_i})$, $\mathrm{FV}(\mathrm{rec}\, x \,.\, M) = \mathrm{FV}(M) \setminus x$. The usual conventions concerning bound variables apply. In particular, all definitions will be robust against bound renaming. $M[N/x]$ denotes the capture avoiding substitution of all free occurrences of x in M by N.

We axiomatise this calculus by the equations

$$\begin{aligned} \mathrm{case}\, C_i(\vec{K})\, \mathrm{of}\{C_1(\vec{x_1}) \to R_1\,;\, \ldots\} &= R_i[\vec{K}/\vec{x_i}] \\ (\lambda x.M)N &= M[N/x] \\ \mathrm{rec}\, x \,.\, M &= M[\mathrm{rec}\, x \,.\, M/x] \end{aligned}$$

We write $\vdash M = N$ if the equation $M = N$ can be derived from these axioms by the usual rules of equational logic.

Of particular interest are closed terms built exclusively from constructors. We call these terms *data* and denote them by $d, e, \ldots$.

3.2 Denotational semantics

In the following we mean by a *domain* a *Scott-domain*, i.e. an algebraic, countably based, bounded complete, dcpo [GHK$^+$03]. Every domain has a least element $\perp$ w.r.t. the domain ordering $\sqsubseteq$. Let $\mathcal{C}$ be a set of constructors and assume that every $C \in \mathcal{C}$ has a fixed arity. Let D be defined by the recursive domain equation

$$D = \sum_{C \in \mathcal{C}} D^{\mathrm{arity}(C)} + [D \to D]$$

where $+$ and the symbol $\sum$ denote the separated sum and $[\cdot \to \cdot]$ the continuous function space. Of course, this domain equation holds only "up to isomorphism", however, we will usually suppress the isomorphism notationally. Hence, every element of D is of exactly one of the following forms: $\bot$, $C(a_1, \ldots, a_n)$ where $C \in \mathcal{C}$, $n = \mathrm{arity}(C)$ and $a_i \in D$, or $\mathrm{abst}(f)$ where f is a continuous function from D to D. Moreover, the functions $C : D^{\mathrm{arity}(C)} \to D$ and $\mathrm{abst} : [D \to D] \to D$ are continuous injections with disjoint ranges covering all of $D \setminus \bot$. For $a, b \in D$ we define $a\,b := f(b)$ if $a = \mathrm{abst}(f)$ and $a\,b := \bot$, otherwise.

In the proof of the Adequacy Theorem we will exploit the algebraicity of the domain D. Let D_0 be the set of compact elements of D, i.e. those elements $a_0 \in D$ such that for every directed set $A \subseteq D$, if $a_0 \sqsubseteq \sqcup A$, then $a_0 \sqsubseteq a$ for some $a \in A$. That D is algebraic means that every element of D is the directed supremum of compact elements. Since compact elements are generated at some finite stage in the construction of D, there is a rank function $\mathbf{rk}(\cdot) : D_0 \to \mathbb{N}$ with the following properties:

(rk1) $C(a_1, \ldots, a_n)$ is compact iff all the a_i are, and in that case we have for all i, $\mathbf{rk}(C(a_1, \ldots, a_n)) > \mathbf{rk}(a_i)$.

(rk2) If $\mathrm{abst}(f)$ compact, then for every $a \in D$, $f(a)$ is compact with $\mathbf{rk}(f(a)) < \mathbf{rk}(\mathrm{abst}(f))$, and there exists a compact $a_0 \sqsubseteq a$ with $\mathbf{rk}(a_0) < \mathbf{rk}(\mathrm{abst}(f))$ and $f(a_0) = f(a)$.

The rank of a compact element can also be explained as the size of a suitable notation for the corresponding finite consistent set in the Information System representation of domains [Win93].

From standard facts in domain theory it follows that every program term M defines in a natural way a continuous function $[\![M]\!] : D^{\mathrm{Var}} \to D$ which for the purpose of this paper is most conveniently defined by the formula

$$[\![M]\!] := \bigsqcup_{n \in \mathbb{N}} [\![M]\!]^n$$

where the continuous functions $[\![M]\!]^n : D^{\mathrm{Var}} \to D$ are defined by recursion on $n \in \mathbb{N}$. We set $[\![M]\!]^0 \xi = \bot$. The definition of $[\![M]\!]^{n+1} \xi$ depends on the syntactic form of M. We use the notation $\vec{a} \mapsto \vec{b}$ to denote the partial map sending a_i to b_i

where $\vec{a} = a_1, \ldots, a_n$ with different a_i and $\vec{b} = b_1, \ldots, b_n$.

$$\begin{aligned}
[\![x]\!]^{n+1}\xi &= \xi(x) \\
[\![C(M_1, \ldots, M_k)]\!]^{n+1}\xi &= C([\![M_1]\!]^n\xi, \ldots, [\![M_k]\!]^n\xi) \\
[\![\mathrm{case}\, M \,\mathrm{of} \{C_1(\vec{x_1}) \to R_1\,;\, \ldots\}]\!]^{n+1}\xi &= [\![R_i]\!]^n\xi[\vec{x_i} \mapsto \vec{a}] \text{ if } [\![M]\!]^n\xi = C_i(\vec{a}) \\
&= \bot \text{ otherwise} \\
[\![\lambda x.M]\!]^{n+1}\xi &= \mathrm{abst}(f) \text{ where } f(a) := [\![M]\!]^n\xi[x \mapsto a] \\
[\![M\,N]\!]^{n+1}\xi &= ([\![M]\!]^n\xi)\,([\![N]\!]^n\xi) \\
[\![\mathrm{rec}\, x\,.\,M]\!]^{n+1}\xi &= [\![M]\!]^n\xi[x \mapsto [\![\mathrm{rec}\, x\,.\,M]\!]^n\xi]
\end{aligned}$$

Clearly, $[\![M]\!]^n$ increases in the domain ordering if n increases. Therefore, $[\![M]\!]$ is well-defined. If one removes the superscripts from the equations above one obtains valid equations for $[\![M]\!]$. By the definition of compactness we have the following:

Lemma 3.1. *If a is compact and $a \sqsubseteq [\![M]\!]\xi$, then $a \sqsubseteq [\![M]\!]^n\xi$ for some n.*

It is easy to see that this interpretation of terms turns D into a model of the axioms in Sect. 3:

Lemma 3.2 (Model)**.** *If $\vdash M = N$, then $[\![M]\!] = [\![N]\!]$.*

For closed terms M the value $[\![M]\!]\xi$ does not depend on ξ. Hence we sometimes write, somewhat ambiguously, $[\![M]\!]$ for $[\![M]\!]\xi$ where ξ is an arbitrary assignment, for example $\xi(x) = \bot$ for all variables x.

3.3 Operational semantics

A *closure* is a pair (M, η) where M is a term and η is an *environment*, i.e. a finite mapping from variables to closures, such that all free variables of M are in the domain of η. Note that this is an inductive definition on the meta-level. A *value* is a closure (M, η) where M is an *intro term*, i.e. a term of the form $C(M_1, \ldots, M_n)$, or $\lambda x.M_0$. We let range $c, c', \ldots$ over closures and $v, v', \ldots$ over values. We inductively define the relation $c \longrightarrow v$ (big-step reduction), where for partial maps f, g we write $f[g]$ to denote the partial map with domain $\mathrm{dom}(f) \cup \mathrm{dom}(g)$ and $f[g](a) = g(a)$ if $a \in \mathrm{dom}(g)$ and $= f(a)$ if $a \in \mathrm{dom}(f) \setminus \mathrm{dom}(g)$.

(i) $v \longrightarrow v$

(ii) $\dfrac{\eta(x) \longrightarrow v}{(x, \eta) \longrightarrow v}$

(iii) $$\frac{(M,\eta) \longrightarrow (C_i(\vec{K}),\eta') \qquad (R_i, \eta[\vec{x_i} \mapsto (\vec{K},\eta')]) \longrightarrow v}{(\text{case}\, M \,\text{of}\{C_1(\vec{x_1}) \to R_1\,;\,\ldots;\, C_n(\vec{x_n}) \to R_n\},\eta) \longrightarrow v}$$

(iv) $$\frac{(M,\eta) \longrightarrow (\lambda x.M_0,\eta') \qquad (M_0, \eta'[x \mapsto (N,\eta)]) \longrightarrow v}{(M\,N,\eta) \longrightarrow v}$$

(v) $$\frac{(M, \eta[x \mapsto (\text{rec}\, x\,.\, M,\eta)]) \longrightarrow v}{(\text{rec}\, x\,.\, M,\eta) \longrightarrow v}$$

Note that arguments of a constructor are not reduced: since $(C(\vec{M}),\eta)$ is a value, the only possible "reduction" is $(C(\vec{M}),\eta) \longrightarrow (C(\vec{M}),\eta)$. In order to reduce under a constructor we need a further 'print' relation $c \Longrightarrow d$ between closures c and data terms d.

$$\frac{c \longrightarrow (C(M_1,\ldots,M_n),\eta) \qquad (M_1,\eta) \Longrightarrow d_1 \quad \ldots \quad (M_n,\eta) \Longrightarrow d_n}{c \Longrightarrow C(d_1,\ldots,d_n)}$$

Clearly, the inductive definitions of $\longrightarrow$ and $\Longrightarrow$ give rise to an algorithm computing d from c whenever $c \Longrightarrow d$. Since this algorithm corresponds to a call-by-name evaluation of terms one can conclude that, for closed M, whenever $(M,\varnothing) \Longrightarrow d$, then in a call-by-name language such as Haskell the evaluation of the program corresponding to M will terminate with a result corresponding to d (provided M is typeable).

To every closure c we assign a term $\overline{c}$ by 'flattening', i.e. removing the structure provided by the nested environments:

$$\overline{(M,\eta)} = M[\overline{\eta(x)}/x \mid x \in \text{dom}(\eta)]$$

Note that this is a recursive definition on the meta-level.

Lemma 3.3 (Correctness)**.** *(a) If $c \longrightarrow v$, then $\vdash \overline{c} = \overline{v}$.*

(b) If $c \Longrightarrow d$, then $\vdash \overline{c} = d$.

Proof. (a) can be proven by straightforward induction on the definition of $c \longrightarrow v$. (b) Follows from (a) and induction on the definition of $c \Longrightarrow d$. □

3.4 Adequacy

Now we prove that denotational and operational semantics are equivalent w.r.t. data. By the Correctness Lemma 3.3 we know already that for a closed term M, if $(M,\varnothing) \Longrightarrow d$, then $\vdash M = d$ and therefore $[\![M]\!] = d$, by the Model Lemma 3.2. The Adequacy Theorem shows that the converse implication holds as well.

Theorem 3.4 (Adequacy)**.** *If* $[\![M]\!] = d$*, then* $(M, \varnothing) \Longrightarrow d$*.*

The rest of this section is devoted to the proof of this theorem. The proof we present can be viewed as a type free version of Plotkin's Adequacy Theorem for PCF [Plo77]. It is based on a variant of the reducibility or candidate method [Gir71, Tai75] where the role of types is taken over by compact domain elements (see also [Rey74, Win93, BC94, Pit94]). In [CS06] and [Ber05] a similar technique was applied to prove strong normalisation of typed λ-calculi with rewrite rules.

The properties of the rank function for compact domain elements discussed in Sect. 3.2 allow us to define for every compact a a set $\mathbf{Cl}(a)$ of closures, by recursion on $\mathbf{rk}(a)$:

$$\begin{aligned}
\mathbf{Cl}(\bot) &= \text{ the set of all closures} \\
\mathbf{Cl}(C(\vec{a})) &= \{c \mid \exists \vec{M}, \eta\, (c \longrightarrow (C(\vec{M}), \eta) \;\wedge\; \forall i\, (M_i, \eta) \in \mathbf{Cl}(a_i)\} \\
\mathbf{Cl}(\mathrm{abst}(f)) &= \{c \mid \exists x, M, \eta\, (c \longrightarrow (\lambda x.M, \eta) \wedge \forall a \in D_0\, (\mathbf{rk}(a) < \mathbf{rk}(\mathrm{abst}(f)) \\
&\qquad \to \forall c' \in \mathbf{Cl}(a)\, (M, \eta[x \mapsto c']) \in \mathbf{Cl}(f(a))))\}
\end{aligned}$$

Note that the sets $\mathbf{Cl}(a)$ are defined in analogy with the reducibility or computability predicates mentioned above.

Lemma 3.5. *If* a, b *are compact with* $a \sqsubseteq b$*, then* $\mathbf{Cl}(a) \supseteq \mathbf{Cl}(b)$*.*

Proof. Induction on the maximum of $\mathbf{rk}(a)$ and $\mathbf{rk}(b)$. The only interesting case is $\mathrm{abst}(f) \sqsubseteq \mathrm{abst}(g)$. Then $f \sqsubseteq g$ (pointwise). Let $c \in \mathbf{Cl}(\mathrm{abst}(g))$. Then $c \longrightarrow (\lambda x.M, \eta)$, and for all compact b with $\mathbf{rk}(b) < \mathbf{rk}(\mathrm{abst}(g))$ and all $c' \in \mathbf{Cl}(b)$ we have $(M, \eta[x \mapsto c']) \in \mathbf{Cl}(g(b))$. We show $c \in \mathbf{Cl}(\mathrm{abst}(f))$ using the same witness $(\lambda x.M, \eta)$. Let a be compact with $\mathbf{rk}(a) < \mathbf{rk}(\mathrm{abst}(f))$ and let $c' \in \mathbf{Cl}(a)$. By (rk2), there exists a compact $b \sqsubseteq a$ with $\mathbf{rk}(b) < \mathbf{rk}(\mathrm{abst}(g))$ and $g(b) = g(a)$. By induction hypothesis, $\mathbf{Cl}(b) \supseteq \mathbf{Cl}(a)$, hence $c' \in \mathbf{Cl}(b)$. It follows $(M, \eta[x \mapsto c']) \in \mathbf{Cl}(g(b)) = \mathbf{Cl}(g(a))$. □

Lemma 3.6. $c \in \mathbf{Cl}(a)$ *iff there exists a value* v *with* $c \longrightarrow v$ *and* $v \in \mathbf{Cl}(a)$*.*

Proof. This can be seen by a trivial induction in $\mathbf{rk}(a)$ using the fact that for values v, v' we have $v \longrightarrow v'$ iff $v = v'$. □

Lemma 3.7. *If* $c \in \mathbf{Cl}(d)$*, where* d *is a data, then* $c \Longrightarrow d$*.*

Proof. Straightforward induction on d. □

Lemma 3.8 (Coincidence)**.** *If* $(M, \eta) \in \mathbf{Cl}(a)$ *and* $\eta(x) = \eta'(x)$ *for all* $x \in \mathrm{FV}(M)$*, then* $(M, \eta') \in \mathbf{Cl}(a)$*.*

Proof. Straightforward induction on the $\mathbf{rk}(a)$. □

We call a total or partial assignment ξ compact if $\xi(x)$ is compact for all $x \in \mathrm{dom}(\xi)$, and write $\xi \sqsubseteq \xi'$ if $\mathrm{dom}(\xi) = \mathrm{dom}(\xi')$ and $\xi(x) \sqsubseteq \xi'(x)$ for all $x \in \mathrm{dom}(\xi)$. We write $\eta \in \mathbf{Cl}(\xi)$ if η is an environment with $\mathrm{dom}(\eta) \subseteq \mathrm{dom}(\xi)$, ξ is compact and $\eta(x) \in \mathbf{Cl}(\xi(x))$ for all $x \in \mathrm{dom}(\eta)$.

Lemma 3.9 (Approximation)**.** *If $\eta \in \mathbf{Cl}(\xi)$ and a is compact with $a \sqsubseteq [\![M]\!]\xi$, then $(M, \eta) \in \mathbf{Cl}(a)$.*

Proof. By Lemma 3.1 it is enough to show:

$$\text{If } \eta \in \mathbf{Cl}(\xi) \text{ and } a \text{ is compact with } a \sqsubseteq [\![M]\!]^n\xi, \text{ then } (M, \eta) \in \mathbf{Cl}(a).$$

We prove this by induction on $n \in \mathbb{N}$. The induction base, $n = 0$, is easy, since $[\![M]\!]^0\xi = \bot$ and therefore $a = \bot$, and $\mathbf{Cl}(\bot)$ is the set of all closures.

In the induction step, $n + 1$, we do a case analysis on the shape of M. We may assume $a \neq \bot$, since otherwise the assertion is trivial.

Case x. By assumption, $a \sqsubseteq [\![x]\!]^{n+1}\xi = \xi(x)$ and $\eta(x) \in \mathbf{Cl}(\xi(x))$. By Lemma 3.5, $\eta(x) \in \mathbf{Cl}(a)$. By Lemma 3.6, there exists a value v with $\eta(x) \longrightarrow v$ and $v \in \mathbf{Cl}(a)$. It follows $(x, \eta) \longrightarrow v$ and therefore $(x, \eta) \in \mathbf{Cl}(a)$, again by Lemma 3.6.

Case $C(\vec{M})$. By assumption we have $a \sqsubseteq [\![C(\vec{M})]\!]^{n+1}\xi = C([\![\vec{M}]\!]^n\xi)$. Hence $a = C(\vec{a})$ with $a_i \sqsubseteq [\![M_i]\!]^n\xi$. By induction hypothesis, $(M_i, \eta) \in \mathbf{Cl}(a_i)$. Since $(C(\vec{M}), \eta) \longrightarrow (C(\vec{M}), \eta)$ (recall that $(C(\vec{M}), \eta)$ is a value), it follows that $(C(\vec{M}), \eta) \in \mathbf{Cl}(C(\vec{a}))$.

Case $\lambda x.M$. By assumption, $a \sqsubseteq [\![\lambda x.M]\!]^{n+1}\xi = \mathrm{abst}(g)$ where $g(b) = [\![M]\!]^n\xi[x \mapsto b]$. Hence, $a = \mathrm{abst}(f)$ with $f \sqsubseteq g$. By induction hypothesis, $(M, \eta[x \mapsto c]) \in \mathbf{Cl}(f(b))$, for all compact b and all $c \in \mathbf{Cl}(b)$. Since $(\lambda x.M, \eta) \longrightarrow (\lambda x.M, \eta)$, it follows $(\lambda x.M, \eta) \in \mathbf{Cl}(\mathrm{abst}(f))$.

Case $\mathrm{case}\, M\, \mathrm{of}\{C(\vec{x_1}) \to R_1\, ; \ldots\}$. By assumption $a \sqsubseteq [\![\mathrm{case}\, M\, \mathrm{of}\{C(\vec{x_1}) \to R_1\, ; \ldots\}]\!]^{n+1}\xi$. Since $a \neq \bot$ we have, $[\![M]\!]^n\xi = C_i(\vec{b})$ for some i and $a \sqsubseteq [\![R_i]\!]^n\xi[\vec{x_i} \mapsto \vec{b}]$. Since a is compact and the function mapping $\vec{b}$ to $[\![R_i]\!]^n\xi[\vec{x_i} \mapsto \vec{b}]$ is continuous it follows that $a \sqsubseteq [\![R_i]\!]^n\xi[\vec{x_i} \mapsto \vec{b_0}]$ for some compact $\vec{b_0} \sqsubseteq \vec{b}$. By induction hypothesis, $(M, \eta) \in \mathbf{Cl}(C(\vec{b_0}))$. Hence, $(M, \eta) \longrightarrow (C(\vec{M_0}), \eta_0)$ with $(\vec{M_0}, \eta_0) \in \mathbf{Cl}(\vec{b_0})$. Again, by induction hypothesis, $(R_i, \eta[x \mapsto \eta_0\vec{M_0}]) \in \mathbf{Cl}(a)$. By Lemma 3.6, $(R_i, \eta[x \mapsto \eta_0\vec{M_0}]) \longrightarrow v)$ for some value $v \in \mathbf{Cl}(a)$. It follows $(\mathrm{case}\, M\, \mathrm{of}\{C(\vec{x_1}) \to R_1\, ; \ldots\}, \eta) \longrightarrow v$ and consequently $(\mathrm{case}\, M\, \mathrm{of}\{C(\vec{x_1}) \to R_1\, ; \ldots\}, \eta) \in \mathbf{Cl}(a)$, again by Lemma 3.6.

Case $M\, N$. By assumption, $a \sqsubseteq [\![M\, N]\!]^{n+1}\xi$. Since $a \neq \bot$ we have, $[\![M]\!]^n\xi = \mathrm{abst}(f)$ and $a \sqsubseteq f([\![N]\!]^n\xi)$. Since function application is continuous, there are a compact $f_0 \sqsubseteq f$ and a compact $b \sqsubseteq [\![N]\!]^n\xi$ with $a \sqsubseteq f_0(b)$. By (rk2), we may assume $\mathbf{rk}(b) < \mathbf{rk}(\mathrm{abst}(f_0))$. By induction hypothesis, $(M, \eta) \in \mathbf{Cl}(\mathrm{abst}(f_0))$ and $(N, \eta) \in \mathbf{Cl}(b)$. Therefore, $M \longrightarrow (\lambda x.M_0, \eta_0)$ such that $(M_0, \eta_0[x \mapsto (N, \eta)]) \in \mathbf{Cl}(f_0(b))$. By Lemma 3.6, $(M_0, \eta_0[x \mapsto (N, \eta)]) \longrightarrow v$ for some $v \in \mathbf{Cl}(f_0(b))$. It follows $(M\,N, \eta) \longrightarrow v$ and hence $(M\,N, \eta) \in \mathbf{Cl}(f_0(b)) \subseteq \mathbf{Cl}(a)$, by Lemma 3.6 and Lemma 3.5.

Case $\mathrm{rec}\,x\,.\,M$. By assumption, we have $a \sqsubseteq [\![\mathrm{rec}\,x\,.\,M]\!]^{n+1}\xi = [\![M]\!]^n\xi[x \mapsto [\![\mathrm{rec}\,x\,.\,M]\!]^n\xi]$. By a similar continuity argument as earlier in the proof, there exists a compact $b \sqsubseteq [\![\mathrm{rec}\,x\,.\,M]\!]^n\xi$ such that $a \sqsubseteq [\![M]\!]^n\xi[x \mapsto b]$. By induction hypothesis, $(\mathrm{rec}\,x\,.\,M, \eta) \in \mathbf{Cl}(b)$ and $(M, \eta[x \mapsto (\mathrm{rec}\,x\,.\,M, \eta)]) \in \mathbf{Cl}(a)$. By Lemma 3.6, $(M, \eta[x \mapsto (\mathrm{rec}\,x\,.\,M, \eta)]) \longrightarrow v$ for some value $v \in \mathbf{Cl}(a)$, therefore $(\mathrm{rec}\,x\,.\,M, \eta) \longrightarrow v$, and finally, $(\mathrm{rec}\,x\,.\,M, \eta) \in \mathbf{Cl}(a)$. $\square$

Proof of the Adequacy Theorem (Thm. 3.4). Assume $[\![M]\!] = d$ for some data d. Since d is compact, it follows, by the Approximation Lemma 3.9, $(M, \varnothing) \in \mathbf{Cl}(d)$. Hence $(M, \varnothing) \Longrightarrow d$, by Lemma 3.7.

3.5 Types, map, iteration and coiteration

The typing discipline we introduce now serves two purposes. First, types are used as indices for families of terms realising monotonicity, induction and coinduction. Second, we will show that all extracted programs are typeable and hence are valid programs in a typed functional programming language such as Haskell or ML.

Types are constructed from type variables $\alpha, \beta, \ldots \in \mathrm{TVar}$ according to the grammar

$$\mathrm{Type} \ni \rho, \sigma, \tau ::= \alpha \mid \mathbf{1} \mid \rho + \sigma \mid \rho \times \sigma \mid \rho \to \sigma \mid \mathrm{fix}\,\alpha.\rho$$

We consider the instance of our term language determined by the constructors Nil (nullary), Left, Right (unary), Pair (binary) and $\mathrm{In}_{\mathrm{fix}\,\alpha.\rho}$ (unary) for every fixed point type $\mathrm{fix}\,\alpha.\rho$, and define inductively the relation $\Gamma \vdash M : \rho$ (term M is of type ρ in typing context Γ).

(i) $\Gamma, x : \rho \vdash x : \rho$

(ii) $\Gamma \vdash \mathrm{Nil} : \mathbf{1}$

(iii) $\dfrac{\Gamma, x : \rho \vdash M : \sigma}{\Gamma \vdash \lambda x.M : \rho \to \sigma} \qquad \dfrac{\Gamma \vdash M : \rho \to \sigma \qquad \Gamma \vdash N : \rho}{\Gamma \vdash M\,N : \sigma}$

(iv) $$\frac{\Gamma \vdash M : \rho \qquad \Gamma \vdash N : \sigma}{\Gamma \vdash \mathrm{Pair}(M, N) : \rho \times \sigma}$$

$$\frac{\Gamma \vdash M : \rho \times \sigma \qquad \Gamma, x_1 : \rho, x_2 : \sigma \vdash R : \tau}{\Gamma \vdash \mathrm{case}\, M\, \mathrm{of}\{\mathrm{Pair}(x_1, x_2) \to R\} : \tau}$$

(v) $$\frac{\Gamma \vdash M : \rho}{\Gamma \vdash \mathrm{Left}(M) : \rho + \sigma} \qquad \frac{\Gamma \vdash M : \sigma}{\Gamma \vdash \mathrm{Right}(M) : \rho + \sigma}$$

$$\frac{\Gamma \vdash M : \rho + \sigma \qquad \Gamma, x_1 : \rho \vdash L : \tau \qquad \Gamma, x_2 : \sigma \vdash R : \tau}{\Gamma \vdash \mathrm{case}\, M\, \mathrm{of}\{\mathrm{Left}(x_1) \to L\, ; \mathrm{Right}(x_2) \to R\} : \tau}$$

(vi) Let $\rho = \rho(\vec{\alpha}) = \mathrm{fix}\, \alpha.\rho_0(\alpha, \vec{\alpha})$:

$$\frac{\Gamma \vdash M : \rho_0(\rho(\vec{\sigma}), \vec{\sigma})}{\Gamma \vdash \mathrm{In}_\rho(M) : \rho(\vec{\sigma})}$$

$$\frac{\Gamma \vdash M : \rho(\vec{\sigma}) \qquad \Gamma, x : \rho_0(\rho(\vec{\sigma}), \vec{\sigma}) \vdash R : \tau}{\Gamma \vdash \mathrm{case}\, M\, \mathrm{of}\{\mathrm{In}_\rho(x) \to R\} : \tau}$$

(vii) $$\frac{\Gamma, x : \tau \vdash M : \tau}{\Gamma \vdash \mathrm{rec}\, x\, .\, M : \tau}$$

The following definition refers to a fixed one-to-one assignment of variables f_α to type variables α. For every list of type variables $\vec{\alpha}$ and every type ρ which is s.p. in $\vec{\alpha}$ we define a program term $\mathbf{Map}_{\vec{\alpha};\rho}$ with $\mathrm{FV}(\mathbf{Map}_{\vec{\alpha};\rho}) = \{f_\alpha \mid \alpha \in \vec{\alpha} \cap \mathrm{FTV}(\rho)\}$ by induction on the structure of ρ.

$$\begin{aligned}
\mathbf{Map}_{\vec{\alpha};\alpha_i} &= f_i, \\
\mathbf{Map}_{\vec{\alpha};\rho} &= \lambda x.\, x, \text{if no } \alpha_i \text{ occurs in } \rho, \\
\mathbf{Map}_{\vec{\alpha};\rho+\sigma} &= \lambda x.\mathrm{case}\, x\, \mathrm{of}\{\mathrm{Left}(y) \to \mathrm{Left}(\mathbf{Map}_{\vec{\alpha};\rho} y)\, ; \\
&\qquad\qquad \mathrm{Right}(z) \to \mathrm{Right}(\mathbf{Map}_{\vec{\alpha};\sigma} z)\} \\
\mathbf{Map}_{\vec{\alpha};\rho\times\sigma} &= \lambda x.\, \mathrm{case}\, x\, \mathrm{of}\{\mathrm{Pair}(y, z) \to \mathrm{Pair}(\mathbf{Map}_{\vec{\alpha};\rho} y, \mathbf{Map}_{\vec{\alpha};\sigma} z)\} \\
\mathbf{Map}_{\vec{\alpha};\rho\to\sigma} &= \lambda x.\, \lambda y.\, \mathbf{Map}_{\vec{\alpha};\sigma}(x\, y) \\
\mathbf{Map}_{\vec{\alpha};\mathrm{fix}\,\alpha.\rho} &= \mathrm{rec}\, f_\alpha\, .\, \lambda x.\, \mathrm{case}\, x\, \mathrm{of}\{\mathrm{In}_{\mathrm{fix}\,\alpha.\rho}(y) \to \mathrm{In}_{\mathrm{fix}\,\alpha.\rho}(\mathbf{Map}_{\vec{\alpha},\alpha;\rho} y)\}
\end{aligned}$$

A type is called *regular* if in its construction the clause $\mathrm{fix}\,\alpha.\rho$ is applied only if ρ is s.p. in α. In the following all mentioned types are assumed to be regular.

Lemma 3.10. *Let $\rho = \rho(\vec{\alpha})$ be s.p. in $\vec{\alpha}$. Consider a context $\Gamma = f_{\alpha_1} : \sigma_1 \to \tau_1, \ldots, f_{\alpha_n} : \sigma_n \to \tau_n$. Then*

$$\Gamma \vdash \mathbf{Map}_{\vec{\alpha};\rho} : \rho(\vec{\sigma}) \to \rho(\vec{\tau})$$

Proof. We give a detailed derivation only for the case that $\rho(\vec{\alpha})$ is of the form $\mathrm{fix}\,\alpha.\rho_0(\alpha,\vec{\alpha})$. In the derivation below we set $\rho_{\vec{\sigma}} := \rho(\vec{\sigma}) = \mathrm{fix}\,\alpha.\rho_0(\alpha,\vec{\sigma})$.

$$\dfrac{\dfrac{\dfrac{x\colon \rho_{\vec\sigma} \vdash x\colon \rho_{\vec\sigma} \qquad \dfrac{\dfrac{\Gamma,\, f_\alpha\colon \rho_{\vec\sigma}\to\rho_{\vec\tau} \vdash \mathbf{Map}_{\vec\alpha,\alpha;\rho_0}\colon \rho_0(\rho_{\vec\sigma},\vec\sigma)\to\rho_0(\rho_{\vec\tau},\vec\tau) \qquad y\colon \rho_0(\rho_{\vec\sigma},\vec\sigma)\vdash y\colon \rho_0(\rho_{\vec\sigma},\vec\sigma)}{\Gamma,\, f_\alpha\colon \rho_{\vec\sigma}\to\rho_{\vec\tau},\, y\colon \rho_0(\rho_{\vec\sigma},\vec\sigma) \vdash \mathbf{Map}_{\vec\alpha,\alpha;\rho_0} y\colon \rho_0(\rho_{\vec\tau},\vec\tau)}}{\Gamma,\, f_\alpha\colon \rho_{\vec\sigma}\to\rho_{\vec\tau},\, y\colon \rho_0(\rho_{\vec\sigma},\vec\sigma) \vdash \mathrm{In}_\rho(\mathbf{Map}_{\vec\alpha,\alpha;\rho_0} y)\colon \rho_{\vec\tau}}}{\Gamma,\, x\colon \rho_{\vec\sigma},\, f_\alpha\colon \rho_{\vec\sigma}\to\rho_{\vec\tau} \vdash \mathrm{case}\,x\,\mathrm{of}\{\mathrm{In}_\rho(y)\to \mathrm{In}_\rho(\mathbf{Map}_{\vec\alpha,\alpha;\rho_0} y)\}\colon \rho_{\vec\tau}}}{\Gamma,\, f_\alpha\colon \rho_{\vec\sigma}\to\rho_{\vec\tau} \vdash \lambda x.\mathrm{case}\,x\,\mathrm{of}\{\mathrm{In}_\rho(y)\to \mathrm{In}_\rho(\mathbf{Map}_{\vec\alpha,\alpha;\rho_0} y)\}\colon \rho_{\vec\sigma}\to\rho_{\vec\tau}}}{\Gamma \vdash \mathrm{rec}\, f_\alpha\,.\,\lambda x.\mathrm{case}\,x\,\mathrm{of}\{\mathrm{In}_\rho(y)\to \mathrm{In}_\rho(\mathbf{Map}_{\vec\alpha,\alpha;\rho_0} y)\}\colon \rho_{\vec\sigma}\to\rho_{\vec\tau}}$$

$$\text{i.e. } \Gamma \vdash \mathbf{Map}_{\vec\alpha;\rho}\colon \rho_{\vec\sigma}\to\rho_{\vec\tau}$$

□

We introduce the abbreviations

$$\begin{aligned}
\mathbf{map}_{\vec\alpha;\rho} &:= \lambda f_{\alpha_1},\ldots,f_{\alpha_n}.\,\mathbf{Map}_{\vec\alpha;\rho},\\
\mathbf{in}_{\mathrm{fix}\,\alpha.\rho} &:= \lambda y.\,\mathrm{In}_{\mathrm{fix}\,\alpha.\rho}(y),\\
\mathbf{out}_{\mathrm{fix}\,\alpha.\rho} &:= \lambda x.\,\mathrm{case}\,x\,\mathrm{of}\{\mathrm{In}_{\mathrm{fix}\,\alpha.\rho}(y)\to y\}.
\end{aligned}$$

Lemma 3.11.

$$(\mathbf{map}_{\vec\alpha;\mathrm{fix}\,\alpha.\rho}\vec f)\circ \mathbf{in}_{\mathrm{fix}\,\alpha.\rho} = \mathbf{in}_{\mathrm{fix}\,\alpha.\rho}\circ(\mathbf{map}_{\vec\alpha,\alpha;\rho(\vec\alpha,\alpha)}\vec f(\mathbf{map}_{\vec\alpha;\mathrm{fix}\,\alpha.\rho}\vec f))$$

Proof. By the definition of $\mathbf{map}_{;}$ we have $\mathbf{map}_{\vec\alpha;\mathrm{fix}\,\alpha.\rho(\vec\alpha)}\vec f(\mathrm{In}_{\mathrm{fix}\,\alpha.\rho}(y)) =$ $(\mathrm{rec}\,f_\alpha\,.\,\lambda x.\,\mathrm{case}\,x\,\mathrm{of}\{\mathrm{In}_{\mathrm{fix}\,\alpha.\rho}(y)\to \mathrm{In}_{\mathrm{fix}\,\alpha.\rho}(\mathbf{map}_{\vec\alpha,\alpha;\rho(\vec\alpha,\alpha)}\vec f f_\alpha y)\})$ $(\mathrm{In}_{\mathrm{fix}\,\alpha.\rho}(y)) = \mathrm{In}_{\mathrm{fix}\,\alpha.\rho}(\mathbf{map}_{\vec\alpha,\alpha;\rho(\vec\alpha,\alpha)}\vec f(\mathbf{map}_{\vec\alpha;\mathrm{fix}\,\alpha.\rho(\vec\alpha)}\vec f)y)$. □

For every (regular) type $\mathrm{fix}\,\alpha.\rho$ we define the closed terms

$$\begin{aligned}
\mathbf{It}_{\mathrm{fix}\,\alpha.\rho} &:= \lambda s.\,\mathrm{rec}\,f\,.\,\lambda x.\,\mathrm{case}\,x\,\mathrm{of}\{\mathrm{In}_{\mathrm{fix}\,\alpha.\rho}(y)\to s(\mathbf{map}_{\alpha;\rho} f y)\}\\
\mathbf{Coit}_{\mathrm{fix}\,\alpha.\rho} &:= \lambda s.\mathrm{rec}\,f\,.\,\lambda x.\mathrm{In}_{\mathrm{fix}\,\alpha.\rho}(\mathbf{map}_{\alpha;\rho} f(s\,x))
\end{aligned}$$

which will later be used as realisers for induction and coinduction. The following is an immediate consequence of Lemma 3.10.

Lemma 3.12 (Typability of iterator and coiterator)**.** *For all types* $\sigma, \vec{\sigma}$

$$\vdash \mathbf{It}_{\mathrm{fix}\,\alpha.\rho} : (\rho(\sigma,\vec{\sigma}) \to \sigma) \to \mathrm{fix}\,\alpha.\rho(\vec{\sigma}) \to \sigma$$
$$\vdash \mathbf{Coit}_{\mathrm{fix}\,\alpha.\rho} : (\sigma \to \rho(\sigma,\vec{\sigma})) \to \sigma \to \mathrm{fix}\,\alpha.\rho(\vec{\sigma})$$

Lemma 3.13. *(a)* $\vdash \mathbf{It}_{\mathrm{fix}\,\alpha.\rho} s \circ \mathbf{in}_{\mathrm{fix}\,\alpha.\rho} = s \circ \mathbf{map}_{\vec{\alpha};\rho}(\mathbf{It}_{\mathrm{fix}\,\alpha.\rho} s)$

(b) $\vdash \mathbf{out}_{\mathrm{fix}\,\alpha.\rho} \circ \mathbf{Coit}_{\mathrm{fix}\,\alpha.\rho} s = \mathbf{map}_{\alpha;\rho}(\mathbf{Coit}_{\mathrm{fix}\,\alpha.\rho} s) \circ s$

Proof. Immediate by the definitions. □

4 Realisability

In this section we introduce a formalised realisability interpretation of the theory of inductive and coinductive definitions of Sect. 2. To this end we need a system that can talk about mathematical objects *and* realisers. Therefore we extend our first-order language $\mathcal{L}$ to a language $\mathbf{r}(\mathcal{L})$ by adding a new sort for program terms. All logical operations including inductive and coinductive definitions, as well as axioms and rules for $\mathcal{L}$ including closure, induction, coclosure and coinduction and the rules for equality, are extended mutatis mutandis for $\mathbf{r}(\mathcal{L})$. In addition, we have as extra axioms the equations given in Sect. 3.1.

4.1 Uniform realisability

We assign to every $\mathcal{L}$-formula A a unary $\mathbf{r}(\mathcal{L})$-predicate $\mathbf{r}(A)$ denoting a subset of D. Intuitively, $\mathbf{r}(A)(a)$, sometimes also written $a\,\mathbf{r}\,A$, states that a "realises" A. The predicate $\mathbf{r}(A)$ is defined relative to a fixed one-to-one mapping from $\mathcal{L}$-predicate variables X to $\mathbf{r}(\mathcal{L})$-predicate variables $\widetilde{X}$ with one extra argument place for domain elements. The definition of $\mathbf{r}(A)$ is such that if the formula A has the free predicate variables $X_1, \ldots, X_n$, then the predicate $\mathbf{r}(A)$ has the free predicate variables $\widetilde{X_1}, \ldots, \widetilde{X_n}$. Simultaneously with $\mathbf{r}(A)$ we define a predicate $\mathbf{r}(\mathcal{P})$ for every predicate $\mathcal{P}$, where $\mathbf{r}(\mathcal{P})$ has one extra argument place for domain elements. We also define regular types $\tau(A)$ and $\tau(\mathcal{P})$ relative to a fixed assignment of a type variable α_X to each predicate variable X.

If A is non-computational:

$$\mathbf{r}(A) = \{\mathrm{Nil} \mid A\} \qquad \tau(A) = \mathbf{1}$$

If A is non-computational but B is:

$$\mathbf{r}(A \wedge B) = \qquad \tau(A \wedge B) =$$
$$\mathbf{r}(B \wedge A) = \{x \mid A \wedge \mathbf{r}(B)(x)\} \qquad \tau(B \wedge A) = \tau(B)$$
$$\mathbf{r}(A \to B) = \{x \mid A \to \mathbf{r}(B)(x)\} \qquad \tau(A \to B) = \tau(B)$$

In all other cases:

$$\mathbf{r}(\mathcal{P}(\vec{t})) = \{x \mid \mathbf{r}(\mathcal{P})(x, \vec{t})\} \qquad \tau(\mathcal{P}(\vec{t})) = \tau(\mathcal{P})$$
$$\mathbf{r}(A \wedge B) = \mathrm{Pair}(\mathbf{r}(A), \mathbf{r}(B)) \qquad \tau(A \wedge B) = \tau(A) \times \tau(B)$$
$$\mathbf{r}(A \vee B) = \mathrm{Left}(\mathbf{r}(A)) \cup \mathrm{Right}(\mathbf{r}(B)) \qquad \tau(A \vee B) = \tau(A) + \tau(B)$$
$$\mathbf{r}(A \to B) = \{f \mid f(\mathbf{r}(A)) \subseteq \mathbf{r}(B)\} \qquad \tau(A \to B) = \tau(A) \to \tau(B)$$
$$\mathbf{r}(\forall y\, A) = \{x \mid \forall y\, (\mathbf{r}(A)(x))\} \qquad \tau(\forall y\, A) = \tau(A)$$
$$\mathbf{r}(\exists y\, A) = \{x \mid \exists y\, (\mathbf{r}(A)(x))\} \qquad \tau(\exists y\, A) = \tau(A)$$

If $\mathcal{P}$ is non-computational:

$$\mathbf{r}(\mathcal{P}) = \{(\mathrm{Nil}, \vec{x}) \mid \mathcal{P}(\vec{x})\} \qquad \tau(\mathcal{P}) = \mathbf{1}$$

Otherwise:

$$\mathbf{r}(\{\vec{x} \mid A\}) = \{(y, \vec{x}) \mid \mathbf{r}(A)(y)\} \qquad \tau(\{\vec{x} \mid A\}) = \tau(A)$$
$$\mathbf{r}(X) = \widetilde{X} \qquad \tau(X) = \alpha_X$$
$$\mathbf{r}(\mu X.\mathcal{P}) = \mu\widetilde{X}.\{(\mathrm{In}(y), x) \mid \mathbf{r}(\mathcal{P})(y, x)\} \qquad \tau(\mu X.\mathcal{P}) = \mathrm{fix}\, \alpha_X.\tau(\mathcal{P})$$
$$\mathbf{r}(\nu X.\mathcal{P}) = \nu\widetilde{X}.\{(\mathrm{In}(y), x) \mid \mathbf{r}(\mathcal{P})(y, x)\} \qquad \tau(\nu X.\mathcal{P}) = \mathrm{fix}\, \alpha_X.\tau(\mathcal{P})$$

where in the last two equations $\mathrm{In} := \mathrm{In}_{\mathrm{fix}\, \alpha_X.\tau(\mathcal{P})}$.

Let us see what we get when we apply realisability to our examples from the Introduction. The type associated with the inductively defined set $\mathbb{N}$ of natural numbers is $\tau(\mathbb{N}) = \mathrm{fix}\, \alpha.\mathbf{1} + \alpha$, the usual recursive definition of the data type of unary natural numbers. Its canonical inhabitants are the numerals $\underline{k} := \mathrm{inr}^k(\mathrm{inl}(\mathrm{Nil})))$ ($k \in \mathbb{N}$). Realisability for $\mathbb{N}$, $\mathbf{r}(\mathbb{N})$, is the least relation such that

$$\mathbf{r}(\mathbb{N}) = \{(\mathrm{inl}(\mathrm{Nil}), 0)\} \cup \{(\mathrm{inr}(n), x + 1) \mid \mathbf{r}(\mathbb{N})(n, x)\}$$

Hence, we have for a data d and $k \in \mathbb{R}$ that $d\, \mathbf{r}\, \mathbb{N}(k)$ holds iff k is a natural number and $d = \underline{k}$, i.e. d is a unary representation of k.

If in the second example we identify notationally the set SD with the type $\mathbf{1}+\mathbf{1}+\mathbf{1}$, then $\tau(\mathrm{C}_0) = \mathrm{fix}\,\alpha.\mathrm{SD}\times\alpha$, the type of infinite streams of signed digits. $\mathbf{r}(\mathrm{C}_0)$ is the largest predicate such that

$$\mathbf{r}(\mathrm{C}_0) = \{(\mathrm{Pair}(d_i, a), (i+x)/2)) \mid i \in \mathrm{SD} \ \wedge\ |(i+x)/2| \leq 1 \ \wedge\ \mathbf{r}(\mathrm{C}_0)(a, x)\}$$

It is easy to see that $\mathbf{r}(\mathrm{C}_0)(a, x)$ means that the signed digit stream $a = a_0, a_1, \ldots$ represents x i.e. $x = \Sigma_{i=0}^{\infty} 2^{-(i+1)} * a_i$.

4.2 Soundness

Now we prove that the realisability interpretation is sound in the sense that from every proof of a formula A one can extract a term M of type $\tau(A)$ and a proof that M realises A.

For every $\mathcal{L}$-operator $\Phi = \lambda X.\mathcal{P}$ we define a $\mathbf{r}(\mathcal{L})$-operator $\mathbf{r}(\Phi) := \lambda\widetilde{X}.\mathbf{r}(\mathcal{P})$.

Lemma 4.1 (Substitution)**.** $\mathbf{r}(\Phi)(\mathbf{r}(\mathcal{Q})) = \mathbf{r}(\Phi(\mathcal{Q}))$.

Proof. Straightforward induction on the (syntactic) size of Φ. □

In the next lemmas we consider predicates in the language $\mathbf{r}(\mathcal{L})$ whose first arguments range over predicate terms. The following definitions will be used:

$$\mathcal{P} \circ f := \{(x, \vec{y}) \mid (f\,x, \vec{y}) \in \mathcal{P}\}$$
$$f * \mathcal{P} := \{(f\,x, \vec{y}) \mid (x, \vec{y}) \in \mathcal{P}\}$$

Clearly, $(\mathcal{P} \circ f) \circ g = \mathcal{P} \circ (f \circ g)$ and $f * (g * \mathcal{P}) = (f \circ g) * \mathcal{P}$. The rationale for these definitions is that they allow us to neatly write certain sets of realisers:

$$\mathbf{r}(\mathcal{P} \subseteq \mathcal{Q}) = \{f \mid \mathbf{r}(\mathcal{P}) \subseteq \mathbf{r}(\mathcal{Q}) \circ f\} = \{f \mid f * \mathbf{r}(\mathcal{P}) \subseteq \mathbf{r}(\mathcal{Q})\}$$
$$\mathbf{r}(\mu X.\mathcal{P}) = \mu\widetilde{X}.\mathbf{in} * \mathbf{r}(\mathcal{P})$$
$$\mathbf{r}(\nu X.\mathcal{P}) = \nu\widetilde{X}.\mathbf{in} * \mathbf{r}(\mathcal{P})$$

where in the last two clauses $\mathbf{in} := \mathbf{in}_{\mathrm{fix}\,\alpha_X.\tau(\mathcal{P})}$.

The following easy lemma, which says that the operations $f \mapsto \mathcal{P} \circ f$ and $f \mapsto f * \mathcal{P}$ are adjoints, will allow for an analogous treatment of induction and coinduction.

Lemma 4.2 (Adjunction)**.** $\mathcal{Q} \subseteq \mathcal{P} \circ f \ \Leftrightarrow\ f * \mathcal{Q} \subseteq \mathcal{P}$

Setting $\mathcal{Q} := \mathcal{P} \circ f$ or $\mathcal{P} := f * \mathcal{Q}$ in the adjunction lemma, we immediately get $f * (\mathcal{P} \circ f) \subseteq \mathcal{P}$ and $\mathcal{Q} \subseteq (f * \mathcal{Q}) \circ f$.

Lemma 4.3 (Map). *Let $\Phi = \lambda X.\mathcal{P}'$ be a (strictly positive) operator in the language $\mathcal{L}$, $\alpha := \alpha_X$, and $\rho := \tau(\mathcal{P}')$. Then $\mathbf{map}_{\alpha;\rho}$ realises the monotonicity of Φ, that is*

$$\mathbf{map}_{\alpha;\rho}\,\mathbf{r}\,(\mathcal{P} \subseteq \mathcal{Q} \to \Phi(\mathcal{P}) \subseteq \Phi(\mathcal{Q}))$$

for all $\mathcal{L}$-predicates $\mathcal{P}$ and $\mathcal{Q}$. By the definition of realisability and the Adjunction Lemma 4.2 this is equivalent to each of the following two statements about arbitrary $\mathbf{r}(\mathcal{L})$-predicates $\mathcal{P}$ and $\mathcal{Q}$ of appropriate arity and all f:

(a) $\mathcal{P} \subseteq \mathcal{Q} \circ f \to \mathbf{r}(\Phi)(\mathcal{P}) \subseteq \mathbf{r}(\Phi)(\mathcal{Q}) \circ \mathbf{map}_{\alpha;\rho} f$

(b) $f * \mathcal{P} \subseteq \mathcal{Q} \to \mathbf{map}_{\alpha;\rho} f * \mathbf{r}(\Phi)(\mathcal{P}) \subseteq \mathbf{r}(\Phi)(\mathcal{Q})$

Furthermore, setting in (a) $\mathcal{P} := \mathcal{Q} \circ f$ and in (b) $\mathcal{Q} := f * \mathcal{P}$ one obtains

(c) $\mathbf{r}(\Phi)(\mathcal{Q} \circ f) \subseteq \mathbf{r}(\Phi)(\mathcal{Q}) \circ \mathbf{map}_{\alpha;\rho} f$

(d) $\mathbf{map}_{\alpha;\rho} f * \mathbf{r}(\Phi)(\mathcal{P}) \subseteq \mathbf{r}(\Phi)(f * \mathcal{P})$

Proof. We show a slight generalisation of (a). Let $\Phi = \lambda \vec{X}.\mathcal{P}'$ be an operator with n arguments, $\alpha_i = \alpha_{X_i}$ and $\rho = \rho(\vec{\alpha}) = \tau(\mathcal{P}')$. Then we have for all predicates $\vec{\mathcal{P}} = \mathcal{P}_1, \ldots, \mathcal{P}_n$, $\vec{\mathcal{Q}} = \mathcal{Q}_1, \ldots, \mathcal{Q}_n$ in the language $\mathbf{r}(\mathcal{L})$ and $\vec{f} = f_1, \ldots, f_n$

$$\mathcal{P}_1 \subseteq \mathcal{Q}_1 \circ f_1 \to \ldots \to \mathcal{P}_n \subseteq \mathcal{Q}_n \circ f_n \to \mathbf{r}(\Phi)(\vec{\mathcal{P}}) \subseteq \mathbf{r}(\Phi)(\vec{\mathcal{Q}}) \circ \mathbf{map}_{\alpha;\rho} \vec{f}$$

The proof is by induction on the structure of $\mathcal{P}'$. Recall that $\mathbf{r}(\Phi) = \lambda \vec{\tilde{X}}.\mathbf{r}(\mathcal{P}')$.

Case: No X_i occurs freely in $\mathcal{P}'$. Then $\mathbf{map}_{\vec{\alpha};\rho} \vec{f}$ is the identity. Furthermore, the operator $\mathbf{r}(\Phi)$ is constant. Therefore, the assertion clearly holds. In the following we assume that there is an X_i occurring freely in $\mathcal{P}'$.

We only look at the remaining interesting cases, namely those where $\mathcal{P}'$ is X_i for some i, $\mu Z.\mathcal{P}_0$ or $\nu Z.\mathcal{P}_0$.

Case $\mathcal{P}' = X_i$. Then $\mathbf{r}(\Phi)(\vec{\tilde{X}}) = \widetilde{X_i}$. Since $\mathbf{map}_{\vec{\alpha};\alpha_i} \vec{f} = f_i$, the assertion holds.

Case $\mathcal{P}' = \mu Z.\mathcal{P}_0$. Let $\Phi_0 := \lambda \vec{X}, Z.\mathcal{P}_0$. Then $\mathbf{r}(\Phi)(\vec{\tilde{X}}) = \mu \tilde{Z}.\mathbf{in}_\rho * \mathbf{r}(\Phi_0)(\vec{\tilde{X}}, \tilde{Z})$. Let $\mathcal{Q}_{n+1} := \mathbf{r}(\Phi)(\vec{\mathcal{Q}}) = \mu \tilde{Z}.\mathbf{in}_\rho * \mathbf{r}(\Phi_0)(\vec{\mathcal{Q}}, \tilde{Z})$. Assume $\mathcal{P}_i \subseteq \mathcal{Q}_i \circ f$ for all $i \leq n$. Then we need to show

$$\mu \tilde{Z}.\mathbf{in}_\rho * \mathbf{r}(\Phi_0)(\vec{\mathcal{P}}, \tilde{Z}) \subseteq \mathcal{Q}_{n+1} \circ \mathbf{map}_{\alpha;\rho} \vec{f}$$

We use induction on $\mu \tilde{Z}.\mathbf{in}_\rho * \mathbf{r}(\Phi_0)(\vec{\mathcal{P}}, \tilde{Z})$. Hence, it remains to show

$$\mathbf{in}_\rho * \mathbf{r}(\Phi_0)(\vec{\mathcal{P}}, \mathcal{Q}_{n+1} \circ \mathbf{map}_{\vec{\alpha};\rho} \vec{f}) \subseteq \mathcal{Q}_{n+1} \circ \mathbf{map}_{\vec{\alpha};\rho} \vec{f}$$

i.e., using the Adjunction Lemma 4.2,

$$\mathbf{r}(\Phi_0)(\vec{\mathcal{P}}, \mathcal{Q}_{n+1} \circ \mathbf{map}_{\vec{\alpha};\rho}\vec{f}) \subseteq \mathcal{Q}_{n+1} \circ \mathbf{map}_{\vec{\alpha};\rho}\vec{f} \circ \mathbf{in}_\rho$$

In the first step of the following we use the induction hypothesis with our assumption and $\mathcal{P}_{n+1} := \mathcal{Q}_{n+1} \circ \mathbf{map}_{\vec{\alpha};\rho}\vec{f}$.

$$\begin{aligned}
\mathbf{r}(\Phi_0)(\vec{\mathcal{P}}, \mathcal{P}_{n+1}) &\overset{\text{i.h.}}{\subseteq} \mathbf{r}(\Phi_0)(\vec{\mathcal{Q}}, \mathcal{Q}_{n+1}) \circ \mathbf{map}_{\vec{\alpha},\alpha;\rho(\vec{\alpha},\alpha)}\vec{f}(\mathbf{map}_{\vec{\alpha};\rho}\vec{f}) \\
&\overset{\text{Lemma 4.2}}{\subseteq} (\mathbf{in}_\rho * \mathbf{r}(\Phi_0)(\vec{\mathcal{Q}}, \mathcal{Q}_{n+1})) \circ \mathbf{in}_\rho \circ \mathbf{map}_{\vec{\alpha},\alpha;\rho(\vec{\alpha},\alpha)}\vec{f}(\mathbf{map}_{\vec{\alpha};\rho}\vec{f}) \\
&\overset{\text{Lemma 3.11}}{=} (\mathbf{in}_\rho * \mathbf{r}(\Phi_0)(\vec{\mathcal{Q}}, \mathcal{Q}_{n+1})) \circ ((\mathbf{map}_{\alpha;\rho}\vec{f}) \circ \mathbf{in}_\rho) \\
&= (\mathbf{in}_\rho * \mathbf{r}(\Phi_0)(\vec{\mathcal{Q}}, \mu\widetilde{Z}.\mathbf{in}_\rho * \mathbf{r}(\Phi_0)(\vec{\mathcal{Q}}, \widetilde{Z}))) \circ ((\mathbf{map}_{\alpha;\rho}\vec{f}) \circ \mathbf{in}_\rho) \\
&\overset{\text{fixed point}}{=} \mu\widetilde{Z}.\mathbf{in}_\rho * \mathbf{r}(\Phi_0)(\vec{\mathcal{Q}}, \widetilde{Z}) \circ ((\mathbf{map}_{\alpha;\rho}\vec{f}) \circ \mathbf{in}_\rho) \\
&= \mathcal{Q}_{n+1} \circ ((\mathbf{map}_{\alpha;\rho}\vec{f}) \circ \mathbf{in}_\rho)
\end{aligned}$$

Case $\mathcal{P}' = \nu Z.\mathcal{P}_0$. Let $\Phi_0 := \lambda\vec{X}, Z.\mathcal{P}_0$. Then $\mathbf{r}(\Phi)(\vec{\widetilde{X}}) = \nu\widetilde{Z}.\mathbf{in}_\rho * \mathbf{r}(\Phi_0)(\vec{\widetilde{X}}, \widetilde{Z})$. In this case it is more convenient to use and prove the formulation of (the generalisation of) (b). Assume $f_i * \mathcal{P}_i \subseteq \mathcal{Q}_i$ for all $i \leq n$. Setting $\mathcal{P}_{n+1} := \mathbf{r}(\Phi)(\vec{\mathcal{P}}) = \nu\widetilde{Z}.\mathbf{in}_\rho * \mathbf{r}(\Phi_0)(\vec{\mathcal{P}}, \widetilde{Z})$, we have to show

$$\mathbf{map}_{\vec{\alpha};\rho}\vec{f} * \mathcal{P}_{n+1} \subseteq \nu\widetilde{Z}.\mathbf{in}_\rho * \mathbf{r}(\Phi_0)(\vec{\mathcal{Q}}, \widetilde{Z})$$

We use coinduction on $\nu\widetilde{Z}.\mathbf{r}(\Phi_0)(\vec{\mathcal{Q}}, \widetilde{Z})$. This reduces the problem to showing

$$\mathbf{map}_{\vec{\alpha};\rho}\vec{f} * \mathcal{P}_{n+1} \subseteq \mathbf{in}_\rho * \mathbf{r}(\Phi_0)(\vec{\mathcal{Q}}, \mathbf{map}_{\vec{\alpha};\rho}\vec{f} * \mathcal{P}_{n+1})$$

$$\begin{aligned}
&\mathbf{in}_\rho * \mathbf{r}(\Phi_0)(\vec{\mathcal{Q}}, \mathbf{map}_{\vec{\alpha};\rho}\vec{f} * \mathcal{P}_{n+1}) \\
&\quad\overset{\text{i.h.}}{\supseteq} \mathbf{in}_\rho * ((\mathbf{map}_{\vec{\alpha},\alpha;\rho(\vec{\alpha},\alpha)}\vec{f}(\mathbf{map}_{\vec{\alpha};\rho}\vec{f})) * \mathbf{r}(\Phi_0)(\vec{\mathcal{P}}, \mathcal{P}_{n+1})) \\
&\quad= (\mathbf{in}_\rho \circ (\mathbf{map}_{\vec{\alpha},\alpha;\rho(\vec{\alpha},\alpha)}\vec{f}(\mathbf{map}_{\vec{\alpha};\rho}\vec{f}))) * \mathbf{r}(\Phi_0)(\vec{\mathcal{P}}, \mathcal{P}_{n+1}) \\
&\quad\overset{\text{Lemma 3.11}}{=} ((\mathbf{map}_{\vec{\alpha};\rho}\vec{f}) \circ \mathbf{in}_\rho) * \mathbf{r}(\Phi_0)(\vec{\mathcal{P}}, \mathcal{P}_{n+1}) \\
&\quad= (\mathbf{map}_{\vec{\alpha};\rho}\vec{f}) * (\mathbf{in}_\rho * \mathbf{r}(\Phi_0)(\vec{\mathcal{P}}, \mathcal{P}_{n+1})) \\
&\quad\overset{\text{fixed point}}{=} (\mathbf{map}_{\vec{\alpha};\rho}\vec{f}) * \mathcal{P}_{n+1}
\end{aligned}$$

□

Theorem 4.4 (Soundness)**.** *From a closed derivation of a formula A one can extract a program term M such that $\mathbf{r}(A)(M)$ and $M\colon \tau(A)$ are derivable.*

Proof. As usual, one shows by induction on derivations the following more general statement: From a derivation $B_1,\dots,B_n,\vec{C} \ \vdash \ A$ where $\vec{C}$ are non-computational assumptions one can extract a program term M with free variables among $x_1\colon \tau(B_1),\dots,x_n\colon \tau(B_n)$ such that

$$\mathbf{r}(B_1)(x_1),\dots,\mathbf{r}(B_n)(x_n),\vec{C} \vdash \mathbf{r}(A)(M)$$

and $x_1\colon \tau(B_1),\dots,x_n\colon \tau(B_n) \vdash M\colon \tau(A)$. In the following, we concentrate on the interesting cases: (Co)closure and (Co)induction. Let in the following $\alpha := \alpha_X$, $\rho := \rho(\alpha) := \tau(\Phi(X))$ and $\Phi = \lambda X.\mathcal{P}$.

Closure. We show that $M := \mathbf{in}_{\mathrm{fix}\,\alpha.\rho}$ realises closure, i.e.

$$\mathbf{r}(\Phi(\mu\Phi)) \subseteq \mathbf{r}(\mu\Phi) \circ \mathbf{in}_{\mathrm{fix}\,\alpha.\rho}$$

Using both, Adjunction Lemma 4.2 and Substitution Lemma 4.1, it suffices to show

$$\mathbf{in}_{\mathrm{fix}\,\alpha.\rho} * (\mathbf{r}(\Phi)(\mathbf{r}(\mu\Phi))) \subseteq \mathbf{r}(\mu\Phi)$$

i.e., since $*$ and substitution commute,

$$(\lambda\widetilde{X}.\mathbf{in}_{\mathrm{fix}\,\alpha.\rho} * \mathbf{r}(\mathcal{P}))(\mathbf{r}(\mu\Phi)) \subseteq \mathbf{r}(\mu\Phi)$$

But the latter is the closure axiom for $\mathbf{r}(\mu\Phi)$. Moreover, we have $x : \rho(\mathrm{fix}\,\alpha.\rho) \vdash \mathrm{In}_{\mathrm{fix}\,\alpha.\rho}(x) : \mathrm{fix}\,\alpha.\rho$, that is, $\vdash M : \rho(\mathrm{fix}\,\alpha.\rho) \to \mathrm{fix}\,\alpha.\rho\ (= \tau(\Phi(\mu\Phi) \subseteq \mu\Phi))$.

Coclosure. Similar, by setting $M := \mathbf{out}_{\mathrm{fix}\,\alpha.\rho}$.

$$\mathbf{out}_{\mathrm{fix}\,\alpha.\rho}\,\mathbf{r}\,\mu\Phi \subseteq \Phi(\mu\Phi)$$

can be derived from the coclosure axiom for $\mathbf{r}(\nu\varphi) = \nu\widetilde{X}.\mathbf{in}_{\mathrm{fix}\,\alpha.\rho} * \mathbf{r}(\mathcal{P})$.

Induction. By the Substitution Lemma 4.1, we have $\mathbf{r}(\Phi(\mathcal{Q}) \subseteq \mathcal{Q} \to \mu\Phi \subseteq \mathcal{Q}) = \{f \mid \forall s\,(\mathbf{r}(\Phi)(\mathbf{r}(\mathcal{Q})) \subseteq \mathbf{r}(\mathcal{Q}) \circ s \to \mathbf{r}(\mu\Phi) \subseteq \mathbf{r}(\mathcal{Q}) \circ fs)\}$. Hence, in order to show that $\mathbf{It}_{\mathrm{fix}\,\alpha.\rho}$ $(=: M)$ realises induction, we assume

$$\mathbf{r}(\Phi)(\mathbf{r}(\mathcal{Q})) \subseteq \mathbf{r}(\mathcal{Q}) \circ s \qquad (*)$$

and show $\mathbf{r}(\mu\Phi) \subseteq \mathbf{r}(\mathcal{Q}) \circ \mathbf{It}_{\mathrm{fix}\,\alpha.\rho}s$. We use induction on $\mathbf{r}(\mu\Phi)$ which reduces the problem to showing $(\lambda\widetilde{X}.\mathbf{in}_{\mathrm{fix}\,\alpha.\rho} * \mathbf{r}(\mathcal{P}))(\mathbf{r}(\mathcal{Q}) \circ \mathbf{It}_{\mathrm{fix}\,\alpha.\rho}s) \subseteq \mathbf{r}(\mathcal{Q}) \circ \mathbf{It}_{\mathrm{fix}\,\alpha.\rho}s$, i.e., using the Adjunction Lemma to

$$(\lambda\widetilde{X}.\mathbf{r}(\mathcal{P}))(\mathbf{r}(\mathcal{Q}) \circ \mathbf{It}_{\mathrm{fix}\,\alpha.\rho}s) \subseteq \mathbf{r}(\mathcal{Q}) \circ \mathbf{It}_{\mathrm{fix}\,\alpha.\rho}s \circ \mathbf{in}_{\mathrm{fix}\,\alpha.\rho}.$$

$$\begin{array}{rcl} \mathbf{r}(\Phi)(\mathbf{r}(\mathcal{Q}) \circ \mathbf{It}_{\mathrm{fix}\,\alpha.\rho} s) & \overset{\text{Lemma 4.3 (c)}}{\subseteq} & \mathbf{r}(\Phi)(\mathbf{r}(\mathcal{Q})) \circ \mathbf{map}_{\alpha;\rho}(\mathbf{It}_{\mathrm{fix}\,\alpha.\rho} s) \\ & \overset{(*)}{\subseteq} & \mathbf{r}(\mathcal{Q}) \circ s \circ \mathbf{map}_{\alpha;\rho}(\mathbf{It}_{\mathrm{fix}\,\alpha.\rho} s) \\ & \overset{\text{Lemma 3.13 (a)}}{=} & \mathbf{r}(\mathcal{Q}) \circ \mathbf{It}_{\mathrm{fix}\,\alpha.\rho} s \circ \mathbf{in}_{\mathrm{fix}\,\alpha.\rho} \end{array}$$

Moreover, Lemma 3.12 shows that M has the desired type.

Coinduction. Using the Substitution Lemma and the Adjunction Lemma we have

$$\begin{array}{l} \mathbf{r}(\mathcal{Q} \subseteq \Phi(\mathcal{Q}) \to \mathcal{Q} \subseteq \mu\Phi) \\ \qquad = \{f \mid \forall s\, (s * \mathbf{r}(\mathcal{Q}) \subseteq \mathbf{r}(\Phi)(\mathbf{r}(\mathcal{Q})) \to f s * \mathbf{r}(\mathcal{Q}) \subseteq \mathbf{r}(\nu\Phi))\} \end{array}$$

Similar to the induction case, we can now show that $M := \mathbf{Coit}_{\mathrm{fix}\,\alpha.\rho}$ is a realiser of the coinduction axiom by using the coinduction axiom for $\mathbf{r}(\nu\Phi)$ and Lemma 4.3 (d) as well as Lemma 3.13 (b).

□

4.3 Program extraction

We now combine the Soundness Theorem and the Adequacy Theorem to a theorem essentially saying that extracted programs compute the expected results. By "result" we can, from the user's perspective, only mean observable data, i.e. data as defined in Sect. 3.1, namely terms built from constructors only. Hence we restrict our attention to a class of formula where all realisers are data. We call an $\mathcal{L}$-formula a *data formula* if it contains neither free predicate variables nor coinductive definitions, and every subformula which is an implication is non-computational. Let Data be a formal representation of the set of all data, i.e. the $\mathbf{r}(\mathcal{L})$-predicate

$$\mathrm{Data} = \mu X. \bigcup_{C \text{ constructor}} \{C(\vec{x}) \mid \vec{x} \in X\}$$

Lemma 4.5 (Data formulas)**.** $\mathbf{r}(A) \subseteq \mathrm{Data}$ *for every data formula* A.

Proof. We show more generally: if A is an $\mathcal{L}$-formula such that every subformula which is an implication is non-computational (but A may contain free predicate variables), then

$$\mathbf{r}(A)^{\mathrm{Data}} \subseteq \mathrm{Data}$$

where $\mathbf{r}(A)^{\mathrm{Data}}$ is obtained from $\mathbf{r}(A)$ by replacing every $n+1$-ary $\mathbf{r}(\mathcal{L})$-predicate variable $\widetilde{X}$ by the $\mathbf{r}(\mathcal{L})$-predicate $\mathrm{Data}' := \{(x, \vec{y}) \mid \mathrm{Data}(x)\}$ of the same arity.

The proof is by induction on the structure of A. All other cases are straightforward, except $(\mu X.\mathcal{P})(\vec{t})$. In the latter case we have

$$\mathbf{r}((\mu X.\mathcal{P})(\vec{t})) = \{\mathbf{In}(y) \mid (\mu \widetilde{X}.\mathbf{r}(\mathcal{P}))(y,\vec{t})\}$$

Therefore, it suffices to show $\mu\widetilde{X}.\mathbf{r}(\mathcal{P})' \subseteq \mathrm{Data}'$. Where $\mathbf{r}(\mathcal{P})'$ is is obtained from $\mathbf{r}(\mathcal{P})$ by replacing every $\mathbf{r}(\mathcal{L})$-predicate variable $\widetilde{Y}$ by Data', except $\widetilde{X}$. We show this by induction. Hence we have to show $\mathbf{r}(\mathcal{P})'[\mathrm{Data}'/\widetilde{X}] \subseteq \mathrm{Data}'$, i.e. $\mathbf{r}(\mathcal{P})^{\mathrm{Data}} \subseteq \mathrm{Data}'$, i.e. $\forall \vec{x}\,(\mathbf{r}(\mathcal{P})(\vec{x})^{\mathrm{Data}} \subseteq \mathrm{Data})$. The latter follows from the (structural) induction hypothesis. □

Theorem 4.6 (Program Extraction)**.** *From a proof of a data formula A one can extract a program term M with the property that $(M,\varnothing) \Longrightarrow d$ for some data d provably realising A, i.e. $d\,\mathbf{r}\,A$ is provable.*

Proof. By the Soundness Theorem, we obtain from a proof of A a program term M and a proof of $M\,\mathbf{r}\,A$. By Lemma 4.5, $\mathrm{Data}(M)$ is provable and therefore true in D, i.e. $[\![M]\!] = d$ for some data d. By the Adequacy Theorem, $(M,\varnothing) \Longrightarrow d$, and by Lemma 3.3, $M = d$ is provable. It follows that $d\,\mathbf{r}\,A$ is provable. □

Let us continue our examples from Sect. 2 and Sect. 4.1. Suppose we have proved $\mathrm{C}_0(x)$ for some real number $x \in \mathbb{I}$. In order to obtain observable information about x, for example for a given natural number n a dyadic rational that approximates x with an error $\leq 2^{-10}$, we need to prove that there exist an integer $z < 2^{-n}$ such that $|x - z/(2^n)| \leq 2^{-n}$. From the proof we can then extract a representation of z and hence of the approximating rational $z/(2^{10})$. First, let us define inductively a predicate $\mathbb{Z}$ such that $\mathbb{Z}(z,n)$ means that n is a natural number and z is an integer $< 2^n$.

$$\mathbb{Z} = \mu X.\{(0,0)\} \cup \{(2^n i + z, n+1) \mid i \in \mathrm{SD} \ \wedge \ X(z,n)\}$$

It easy to see that a realiser of $\mathbb{Z}(z,n)$ is a signed binary representation of z (permitting leading zeros).

Lemma 4.7 (Printing digits)**.**

$$\forall n\,(\mathbb{N}(n) \to \forall x\,(\mathrm{C}_0(x) \to \exists z\,(\mathbb{Z}(z,n) \ \wedge \ |x - \frac{z}{2^n}| \leq \frac{1}{2^n})))$$

Proof. Induction on $\mathbb{N}(n)$. Set $\mathcal{P} := \{n \mid \forall x\,(\mathrm{C}_0(x) \to \exists z\,(\mathbb{Z}(z,n) \ \wedge \ |x - \frac{z}{2^n}| \leq \frac{1}{2^n}))\}$. We have to show (1) $\mathcal{P}(0)$, (2) $\forall n\,(\mathcal{P}(n) \to \mathcal{P}(n+1)$. For (1), we can take $z := 0$, since $\mathrm{C}_0(x)$ implies $|x| \leq 1$. For (2), assume $\mathcal{P}(n)$ (i.h.) and $\mathrm{C}_0(x)$. Let

$i \in \mathrm{SD}$ such that $x = (i + y)/2$ for some y with $\mathrm{C}_0(y)$. By i.h. there exists z such that $\mathbb{Z}(z, n)$ and $|y - \frac{z}{2^n}| \leq \frac{1}{2^n}$. It follows $\mathbb{Z}(2^n i + z, n + 1)$ and

$$|x - \frac{2^n i + z}{2^{n+1}}| = \frac{1}{2}|y - \frac{z}{2^n}| \leq \frac{1}{2^{n+1}}$$

□

The program extracted from this proof takes as inputs a (unary) natural number n and a signed digit stream a representing some real number in $\mathbb{I}$, and computes a signed binary representation of an integer $z < 2^n$ such that $|x - z/2^n| \leq 1/2^n$. In fact, the digits of that representation will be exactly the first n elements of the stream a. Hence, the extracted program is essentially Haskell's function `take` that computes the first n elements of a stream.

5 Conclusion and further work

In this paper we laid the programming-technological and proof-theoretic foundations for program extraction from proofs in a constructive theory of inductive and coinductive definitions. We showed that the realising programming language has an adequate denotational and operational semantics and the realisability interpretation is sound. Both results together imply that from proofs of formulas with associated observable types (data formulas) one can extract programs that compute data realising the formula.

In our opinion, one of the main advantages of program extraction over the traditional specify-implement-verify method is that it is possible to carry out proofs in a very simple formal system. Neither complicated data types (lists, streams, trees, function types, etc.) nor programming constructs (recursion, lambda-abstraction) need to be formalised by the user; these are all generated by the realisability interpretation automatically.

On the basis of the results of this paper one can now begin to formalise parts of constructive analysis and other branches of mathematics where inductive and coinductive definitions are used (or can be used), with the aim of extracting nontrivial certified programs. Currently, we are investigating a generalisation of the predicate $\mathrm{C}_0 \subseteq \mathbb{R}$ (one of our running examples) to predicates $\mathrm{C}_n \subseteq \mathbb{R}^{\mathbb{I}^n}$ characterising the (constructively) uniformly continuous function from $\mathbb{I}^n$ to $\mathbb{I}$ [Ber09]. For $n = 1$ the definition is

$$\mathrm{C}_1 :=$$
$$\nu F.\mu G.\{f \in \mathbb{I}^{\mathbb{I}} \mid \exists i \in \mathrm{SD}\, \exists f'\, (f = \mathrm{av}_i \circ f' \ \land\ F(f')) \ \lor\ \forall i \in \mathrm{SD}\, G(f \circ \mathrm{av}_i)\}$$

where F and G range over subsets of $\mathbb{R}^{\mathbb{I}}$ and $\mathrm{av}_i(x) := (i + x)/2$. To see the analogy with C_0 it is useful to rewrite the definition of the latter equivalently as

$$\mathrm{C}_0 := \nu X.\{x \in \mathbb{I} \mid \exists i \in \mathrm{SD}\, \exists x' \,(x = \mathrm{av}_i(x') \;\wedge\; X(x'))\}$$

The predicate C_0 characterises real numbers in $\mathbb{I}$ as objects perpetually emitting digits. A continuous function $f : \mathbb{I} \to \mathbb{I}$, which can be viewed as a real number in $\mathbb{I}$ that depends on an input in $\mathbb{I}$, perpetually emits digits as well, but before an emission can take place f may have to gain information about the input by absorbing finitely many digits from it in order to decide which digit to emit. The absorption part is formalised in C_1 by the inner "$\mu G \ldots G(f \circ \mathrm{av}_i)$". The data type associated with C_1 is

$$\tau(\mathrm{C}_1) = \nu\alpha.\mu\beta.\mathrm{SD} \times \alpha + \beta^3$$

which is the type of non-wellfounded trees with two kinds of nodes, one labelled by a signed digit and one child (emitting a digit), the other without label and three children (absorbing a digit). The fact that β is quantified by μ means that only those trees are legal members of $\tau(\mathrm{C}_1)$ that have on each path infinitely many emitting nodes. A similar type of trees has been studied independently in [GHP06], however, not in the context of analysis and realisability. The definition of C_1 is motivated by earlier works on the development and verification of exact real number algorithms based on the signed digit representation of real numbers [MRE07,GNSW07,EH02] some of which make use of coinductive methods [CDG06,Ber07,BH08,Niq08].

Based on the characterisation of uniformly continuous functions by the predicates C_n implementations of elementary arithmetic functions have been extracted [Ber09]. Further work in progress studies integration and analytic functions based on this approach. We are also extending this work to more general situations where the interval $\mathbb{I}$ and the maps av_i are replaced by an arbitrary bounded metric space with a system of contractions (see [Scr08] for related work), or even to non-metric situations.

Currently, we are adapting the existing implementation of program extraction in the Minlog proof system [BBS+98] to our setting.

References

[AMU05] A. Abel, R. Matthes, and T. Uustalu. Iteration and coiteration schemes for higher-order and nested datatypes. *Theor. Comput. Sci.*, 333:3–66, 2005.

[BBS+98] H. Benl, U. Berger, H. Schwichtenberg, M. Seisenberger, and W. Zuber. Proof theory at work: Program development in the Minlog system. In W. Bibel and P.H. Schmitt, editors, *Automated Deduction – A Basis for Applications*, volume II of *Applied Logic Series*, pages 41–71. Kluwer, Dordrecht, 1998.

[BC94] B. Blaaberg and C. Clausen. Adequacy for a lazy functional language with recursive and polymorphic types. *Theor. Comput. Sci.*, 136(1):243–275, 1994.

[Ber05] U. Berger. Strong normalization for applied lambda calculi. *Logical Methods in Comput. Sci.*, 1(2):1–14, 2005.

[Ber07] Y. Bertot. Affine functions and series with co-inductive real numbers. *Math. Struct. Comput. Sci.*, 17:37–63, 2007.

[Ber09] U. Berger. From coinductive proofs to exact real arithmetic. In E. Grädel and R. Kahle, editors, *Computer Science Logic*, volume 5771 of *LNCS*, pages 132–146. Springer, 2009.

[BFPS81] W. Buchholz, F. Feferman, W. Pohlers, and W. Sieg. *Iterated Inductive Definitions and Subsystems of Analysis: Recent Proof–Theoretical Studies*, volume 897 of *Lecture Notes in Mathematics*. Springer, Berlin, 1981.

[BH08] U. Berger and T. Hou. Coinduction for exact real number computation. *Theory of Computing Systems*, 43:394–409, 2008.

[BS07] J. Bradfield and C. Stirling. Modal mu-calculi. In P. Blackburn, J. van Benthem, and F. Wolter, editors, *Handbook of Modal Logic*, volume 3 of *Studies in Logic and Practical Reasoning*, pages 721–756. Elsevier, 2007.

[CDG06] A. Ciaffaglione and P. Di Gianantonio. A certified, corecursive implementation of exact real numbers. *Theor. Comput. Sci.*, 351:39–51, 2006.

[CS06] T. Coquand and A. Spiwack. A proof of strong normalisation using domain theory. In *Proceedings of the 21st Annual IEEE Symposium on Logic in Computer Science (LICS'06)*, pages 307–316. IEEE Computer Society Press, 2006.

[EH02] A. Edalat and R. Heckmann. Computing with real numbers: I. The LFT approach to real number computation; II. A domain framework for computational geometry. In G. Barthe, P. Dybjer, L. Pinto, and J. Saraiva, editors, *Applied Semantics - Lecture Notes from the International Summer School, Caminha, Portugal*, pages 193–267. Springer, 2002.

[GHK+03] G. Gierz, K.H. Hofmann, K. Keimel, J.D. Lawson, M. Mislove, and D.S. Scott. *Continuous Lattices and Domains*, volume 93 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, 2003.

[GHP06] N. Ghani, P. Hancock, and D. Pattinson. Continuous functions on final coalgebras. *Electr. Notes in Theoret. Comput. Sci.*, 164, 2006.

[Gir71] J-Y. Girard. Une extension de l'interprétation de Gödel à l'analyse, et son application à l'élimination des coupures dans l'analyse et la théorie des types. In J.E. Fenstad, editor, *Proceedings of the Second Scandinavian Logic Symposium*, pages 63–92. North–Holland, 1971.

[GNSW07] H. Geuvers, M. Niqui, B. Spitters, and F. Wiedijk. Constructive analysis, types and exact real numbers. *Math. Struct. Comput. Sci.*, 17(1):3–36, 2007.

[HO08] M. D. Hernest and P. Oliva. Hybrid functional interpretations. In A. Beckmann, C. Dimitracopoulos, and B. Löwe, editors, *CiE 2008: Logic and Theory of Algorithms*, volume 5028 of *LNCS*, pages 251–260. Springer, 2008.

[KP90] Jean-Louis Krivine and Michel Parigot. Programming with proofs. *Elektronische Informationsverarbeitung und Kybernetik*, 26(3):149–167, 1990.

[MÖ3] M. Möllerfeld. *Generalized inductive definitions*. PhD thesis, Westfälische Wilhelms-Universität Münster, 2003.

[Mat01] R. Matthes. Monotone inductive and coinductive constructors of rank 2. In L Fribourg, editor, *Computer Science Logic (Proceedings of the Fifteenth CSL Conference)*, number 2142 in LNCS, pages 600–615. Springer, 2001.

[Men91] N.P. Mendler. Inductive types and type constraints in the second-order lambda calculus. *Ann. Pure Appl. Logic*, 51:159–172, 1991.

[MP05] F. Miranda-Perea. Realizability for monotone clausular (co)inductive definitions. *Electr. Notes in Theoret. Comput. Sci.*, 123:179–193, 2005.

[MRE07] J. R. Marcial-Romero and M. H. Escardo. Semantics of a sequential language for exact real-number computation. *Theor. Comput. Sci.*, 379(1-2):120–141, 2007.

[Niq08] M. Niqui. Coinductive formal reasoning in exact real arithmetic. *Logical Methods in Computer Science*, 4(3:6):1–40, September 2008.

[Par92] M. Parigot. Recursive programming with proofs. *Theor. Comput. Sci.*, 94(2):335–356, 1992.

[Pit94] A.M. Pitts. Computational adequacy via ”mixed” inductive definitions. In *Proceedings of the 9th International Conference on Mathematical Foundations of Programming Semantics*, pages 72–82, London, UK, 1994. Springer-Verlag.

[Plo77] G.D. Plotkin. LCF considered as a programming language. *Theor. Comput. Sci.*, 5:223–255, 1977.

[Rey74] J.C. Reynolds. On the relation between direct and continuation semantics. In *Proceedings of the 2nd Colloquium on Automata, Languages and Programming*, pages 141–156, London, UK, 1974. Springer-Verlag.

[RT10] D. Ratiu and T. Trifonov. Exploring the computational content of the infinite pigeonhole principle. To appear in Journal of Logic and Computation, 2010.

[Sch08] H. Schwichtenberg. Realizability interpretation of proofs in constructive analysis. *Theory of Computing Systems*, 43(3-4):583–602, 2008.

[Scr08] A. Scriven. A functional algorithm for exact real integration with invariant measures. *Electron. Notes Theor. Comput. Sci.*, 218:337–353, 2008.

[Tai75] W.W. Tait. A realizability interpretation of the theory of species. In R. Parikh, editor, *Logic Colloquium Boston 1971/72*, volume 453 of *Lecture Notes in Mathematics*, pages 240–251. Springer, 1975.

[Tat98] M. Tatsuta. Realizability of monotone coinductive definitions and its application to program synthesis. In R. Parikh, editor, *Mathematics of Program Construction*, volume 1422 of *Lecture Notes in Mathematics*, pages 338–364. Springer, 1998.

[Tup04] S. Tupailo. On the intuitionistic strength of monotone inductive definitions. *Jour. Symb. Logic*, 69(3):790–798, 2004.

[Win93] G. Winskel. *The Formal Semantics of Programming Languages*. Foundations of Computing Series. The MIT Press, Cambridge, Massachusetts, 1993.

Another Reduction of Classical ID_ν to Constructive ID^i_ν

Wilfried Buchholz

Introduction. One of the major problems in reductive proof theory in the early 1970s was to give a proof-theoretic reduction of classical theories of iterated arithmetical inductive definitions to corresponding constructive systems. This problem was solved in [BFPS] in various ways which all where based on the method of cut-elimination (normalization, reps.) for infinitary Tait-style sequent calculi (infinitary systems of natural deduction, resp.). Only quite recently Avigad and Towsner [AT09] succeeded in giving a reduction of classical iterated ID theories to constructive ones by the method of functional interpretation. For a thorough exposition and discussion of all this cf. [Fef].

In the present paper we give yet another reduction of classical ID_ν to $\mathrm{ID}^i_\nu(\mathcal{W})$ based on cut-elimination arguments. $\mathcal{W}$ is a particularly simple accessibility ID; its corresponding operator form $\mathcal{W}(P,Q,y,x)$ (cf. [BFPS]) has the shape $A(x,y) \;\wedge\; \forall z(\widetilde{Q}(t(x),z) \;\rightarrow\; Pq(x,z))$ with primitive recursive A,t,q, and $\widetilde{Q}(u,z) :\equiv u \geq 1 \;\wedge\; (u \geq 2 \rightarrow Q(u \dot{-} 2, z))$. There are two reasons which, as we hope, justify a publication of this additional proof. First, it is considerably more direct then all the existing ones. Second, the method used here stems to a great extent from [Ge36] and therefore may be interesting for historical reasons too. Actually I have already used a variant of this method under the label "notations for infinitary derivations" in several papers (e.g. [Bu91], [Bu97], [Bu01]) without mentioning its close relationship to [Ge36]. When writing [Bu91] I was definitely not aware of this connection; but cf. [Bu95]. The method from [Ge36] can be roughly described as follows: By (primitive) recursion on the build-up of h, for each derivation h in a suitably designed finitary proof system Z of first order arithmetic a family $(h[i])_{i\in I_h}$ of Z-derivations is defined such that $\dfrac{\ldots\Gamma(d[i])\ldots(i\in I_h)}{\Gamma(h)}$ (where $\Gamma(h)$ denotes the endsequent of h) forms an inference in cutfree ω-arithmetic (with repetition-rule). Then the consistency of Z is obtained by quantifierfree transfinite induction over the relation $\prec := \{(h[i],h) : h \in Z \;\&\; i \in I_h\}$. In the present paper we proceed similarly. Let ID_ν be the finitary Tait-style system of ν-fold iterated inductive definitions as introduced in [Bu02]. We extend ID_ν by certain inferences E, D_σ, $\mathsf{S}^{\Pi}_{\mathcal{P},\mathcal{F}}$ (which do not alter the set of derivable sequents) to a finitary system ID^*_ν. This step corresponds very much to the passage from BI^-_1 to BI^*_1 in [Bu01]. Then by primitive recursion on the height of h, for each closed

ID^*_ν-derivation h we define a family $(h[\iota])_{\iota\in I_h}$ of closed ID^*_ν-derivations such that $\dfrac{\ldots\Gamma(h[\iota])\ldots(\iota\in I_h)}{\Gamma(h)}$ is an inference in the infinitary system ID^∞_ν. Formulated more technical, we assign to h an inference symbol $\mathsf{tp}(h)$ of ID^∞_ν, and for each $\iota\in|\mathsf{tp}(h)|$ a closed ID^*_ν-derivation $h[\iota]$ such that $\dfrac{\ldots\Gamma(h[\iota])\ldots(\iota\in|\mathsf{tp}(h)|)}{\Gamma(h)}$ is a $\mathsf{tp}(h)$-inference (0.10). On first sight the present system ID^∞_ν looks exactly like the system ID^∞_ν in [Bu02] (which itself is the Tait-style version of the natural deduction system ID^∞_ν from [Bu81]), but there is some subtle difference concerning the index sets $|\widetilde{\Omega}_P|$ of instances of the Ω-rule. In [Bu02], $|\widetilde{\Omega}_P|$ is a set of infinitary derivations while in the present paper $|\widetilde{\Omega}_P|$ is a set of finite derivations, namely $|\widetilde{\Omega}_P| = \mathbf{I}_\mu$ = set of all closed ID^*_ν-derivations h with $\deg(h) = 0$ and $\Gamma(h) \subseteq \mathrm{Pos}_\mu$, where $\mu := \mathrm{lev}(P)$. Now let $\mathcal{W}_\sigma$ be the accessible part of the relation $\{(h[\iota],h) : h\in\mathbf{I}_\sigma \;\&\; \iota\in|\mathsf{tp}(h)|_{\mathcal{W}}\}$, where $|\mathcal{I}|_{\mathcal{W}} := \mathcal{W}_\mu$ if $\mathcal{I} = \widetilde{\Omega}_P$ with $\mu := \mathrm{lev}(P) < \sigma$, and $|\mathcal{I}|_{\mathcal{W}} := |\mathcal{I}|$ otherwise. The proof-theoretic reduction of ID_ν to $\mathrm{ID}^i_\nu(\mathcal{W})$ will be established by a proof of transfinite induction over the relation $\{(h[i],h) : h\in\mathbf{I}_0 \;\&\; i\in|\mathsf{tp}(h)|\}$ which can be locally formalized in $\mathrm{ID}^i_\nu(\mathcal{W})$. The difficulty here is to come along without the uppermost set $\mathcal{W}_\nu$, which would be available in $\mathrm{ID}^i_{\nu+1}(\mathcal{W})$ but not in $\mathrm{ID}^i_\nu(\mathcal{W})$. We overcome this difficulty by using (a generalization of) Gentzen's technique (cf. [Ge43]) for proving transfinite induction up to ordinals $<\varepsilon_0$ within Z.

In order to avoid some annoying but inessential technicalities we restrict our treatment to $\nu<\omega$. So in the whole paper ν is a fixed natural number >0.

Preliminaries. For the reader's convenience we repeat some basic definitions and abbreviations from [Bu02] (with some minor deviations). Let $\mathcal{L}$ be an arbitrary first order language (i.e. set of function and predicate symbols). Atomic $\mathcal{L}$-formulas are $Rt_1...t_n$ where R is an n-ary predicate symbol (of $\mathcal{L}$), and $t_1,...,t_n$ are $\mathcal{L}$-terms. Expressions of the shape A or $\neg A$, where A is an atomic $\mathcal{L}$-formula, are called *literals*. $\mathcal{L}$-formulas are built up from literals by means of $\wedge$, $\vee$, $\forall x, \exists x$. $\mathrm{FV}(A)$ denotes the set of free variables of A. A formula or term A is called *closed* if $\mathrm{FV}(A)=\varnothing$. The *negation* $\neg A$ of a non-atomic formula A is defined via de Morgan's laws. The *rank* $\mathrm{rk}(A)$ of a formula A is defined by: $\mathrm{rk}(A):=0$ if A is a literal, $\mathrm{rk}(A\wedge B):=\mathrm{rk}(A\vee B):=\max\{\mathrm{rk}(A),\mathrm{rk}(B)\}+1$, $\mathrm{rk}(\forall xA):=\mathrm{rk}(\exists xA):=\mathrm{rk}(A)+1$. By $A(x/t)$ we denote the result of substituting t for (every free occurrence of) x in A (renaming bound variables if necessary). Expressions $\lambda x.F$ (where F is a formula) are called *predicates* and denoted by $\mathcal{F}$. For $\mathcal{F}=\lambda x.F$ we set $\mathcal{F}(t):=F(x/t)$. If $\mathcal{P}$ is a unary predicate symbol then $B(\mathcal{P}/\mathcal{F})$ denotes the result of substituting $\mathcal{F}$ for $\mathcal{P}$ in B, i.e. the formula resulting from B

be replacing every atom $\mathcal{P}t$ by $\mathcal{F}(t)$. Let X be a unary predicate symbol not in $\mathcal{L}$. A *positive operator form in* $\mathcal{L}$ is an $\mathcal{L} \cup \{X\}$-formula $\mathfrak{A}$ in which X occurs only positively (i.e. $\mathfrak{A}$ has no subformula $\neg Xt$) and which has at most one free variable x. We use the following abbreviations: $\mathfrak{A}(\mathcal{F},t) := \mathfrak{A}(x/t)(X/\mathcal{F})$, $\mathfrak{A}(\mathcal{F}) \subseteq \mathcal{F} := \forall x(\mathfrak{A}(\mathcal{F},x) \to \mathcal{F}(x))$. For each positive operator form $\mathfrak{A}$ we introduce a new unary predicate symbol $\mathcal{P}_{\mathfrak{A}}$. Finite sets of formulas are called *sequents*. They are denoted by Γ, Δ, Π. We mostly write $A_1, ..., A_n$ for $\{A_1, ..., A_n\}$, and A, Γ, Δ for $\{A\} \cup \Gamma \cup \Delta$, etc.

Definition 0.1 ($\mathcal{L}_\sigma$, Pos_σ, level)**.** Let $\mathcal{L}_0$ be a language consisting of the constant 0 (*zero*), the unary function symbol S (*successor*), and some predicate symbols R for primitive recursive relations, such that the set TRUE_0 of all true closed $\mathcal{L}_0$-literals is itself primitive recursive (under some canonical arithmetization of syntax). The only closed $\mathcal{L}_0$-terms are the *numerals* $0, S0, SS0, ...$ which we identify with the corresponding natural numbers (elements of $\mathbb{N}$). Arbitrary $\mathcal{L}_0$-terms will be denoted by $t, t_1, ...$, and (number) variables by x, y.

- $\mathcal{L}_{\sigma+1} := \mathcal{L}_0 \cup \{\mathcal{P}_{\mathfrak{A}} : \mathfrak{A} \text{ positive operator form in } \mathcal{L}_\sigma\} \quad (\sigma < \omega)$
- $\mathrm{Pos}_\sigma :=$ set of all $\mathcal{L}_{\sigma+1}$-formulas C such every $\mathcal{P}_{\mathfrak{A}}$ occurring negatively in C belongs to $\mathcal{L}_\sigma$.
- $\mathrm{lev}(\mathcal{P}_{\mathfrak{A}}) := \mathrm{lev}(\mathcal{P}_{\mathfrak{A}}t) := \min\{\sigma : \mathcal{P}_{\mathfrak{A}}t \in \mathrm{Pos}_\sigma\}$ (level)

Note that this "level" is not exactly the same as "level" in [Bu02].

Proposition 0.2.

(1) $\mathcal{L}_\sigma$*-formulas* $\subseteq \mathrm{Pos}_\sigma \subseteq \mathcal{L}_{\sigma+1}$*-formulas*

(2) $\mathcal{P}_{\mathfrak{A}}t \in \mathrm{Pos}_\sigma \;\Rightarrow\; \mathfrak{A}(\mathcal{P}_{\mathfrak{A}},t) \in \mathrm{Pos}_\sigma$.

Abbreviations.

- $\mathcal{L}_0$-lit := set of all $\mathcal{L}_0$-literals.
- $\bigwedge$-for := set of all formulas of the shape $A \wedge B$ or $\forall x A$.
- $C \in \bigwedge^+$-for $:\Leftrightarrow$ $C \in \bigwedge$-for or C has the shape $\mathcal{P}_{\mathfrak{A}}t$
- $C[k] := \begin{cases} C_k & \text{if } C = C_0 \stackrel{\wedge}{\vee} C_1 \text{ and } k \in \{0,1\} \\ A(x/k) & \text{if } C = \stackrel{\exists}{\forall} xA \text{ and } k \in \mathbb{N} \end{cases}$

Definition 0.3 (Inference symbols). An *inference symbol* is a formal expression $\mathcal{I}$ for which the following entities are given

- a set $|\mathcal{I}|$ (the *arity* of $\mathcal{I}$),
- a sequent $\Delta(\mathcal{I})$ *(principal formula(s))*,
- for each $\iota \in |\mathcal{I}|$ a sequent $\Delta_\iota(\mathcal{I})$ *(minor formula(s))*.

An inference symbol is called *(in)finitary* if its arity is (in)finite.

Notation. By writing

$$(\mathcal{I}) \quad \frac{\ldots\Delta_\iota\ldots(\iota{\in}I)}{\Delta}$$

we declare $\mathcal{I}$ as an inference symbol with $|\mathcal{I}| = I$, $\Delta(\mathcal{I}) = \Delta$, $\Delta_\iota(\mathcal{I}) = \Delta_\iota$.
If $I = \{0, ..., n{-}1\}$ we write

$$\frac{\Delta_0 \;\; \Delta_1 \;\; \ldots \;\; \Delta_{n-1}}{\Delta}, \quad \text{instead of} \; \frac{\ldots\Delta_\iota\ldots(\iota{\in}I)}{\Delta}.$$

Inference symbols $\mathcal{I}$ with $|\mathcal{I}| = \varnothing$ are called *axioms*.

Definition 0.4 (Proof systems). A *proof system* is given by a language $\mathcal{L}$ and a set of inference symbols in this language, where "$\mathcal{I}$ in $\mathcal{L}$" means that all elements of $\Delta(\mathcal{I}) \cup \bigcup_{\iota\in|\mathcal{I}|} \Delta_\iota(\mathcal{I})$ are $\mathcal{L}$-formulas. A proof system is called *finitary* if all its inference symbols are finitary; otherwise it is called *infinitary*.

From now on the letters A, B, C always denote $\mathcal{L}_\nu$-formulas, and $\mathcal{P}$ ranges over predicate symbols $\mathcal{P}_{\mathfrak{A}} \in \mathcal{L}_\nu$.

Definition 0.5 (The finitary proof systems ID_ν and ID^*_ν). The language of ID_ν is $\mathcal{L}_\nu$, and the inference symbols of ID_ν are

$(\mathsf{Ax}_\Gamma)\ \frac{}{\Gamma}$ if $\Gamma \in \mathsf{Ax}(\nu)$ where $\mathsf{Ax}(\nu)$ is a set of $\mathcal{L}_\nu$-sequents such that

(i) $\Gamma \in \mathsf{Ax}(\nu) \Longrightarrow \Gamma(\vec{x}/\vec{t}) \in \mathsf{Ax}(\nu)$

(ii) $\Gamma \in \mathsf{Ax}(\nu)\ \&\ \mathrm{FV}(\Gamma) = \varnothing \Longrightarrow \Gamma \cap \mathsf{TRUE}_0 \neq \varnothing$ or $\Gamma = \{\neg\mathcal{P}n, \mathcal{P}n\}$ or $\Gamma = \{n \neq n, \neg\mathcal{P}n, \mathcal{P}n\}$

(iii) $\{\neg A, A\} \in \mathsf{Ax}(\nu)$ for each atomic $\mathcal{L}_\nu$-formula A

$$(\textstyle\bigwedge_{A_0 \wedge A_1})\ \frac{A_0 \qquad A_1}{A_0 \wedge A_1}, \qquad (\textstyle\bigvee^k_{A_0 \vee A_1})\ \frac{A_k}{A_0 \vee A_1} \quad (k \in \{0,1\}),$$

$$(\textstyle\bigwedge^y_{\forall xA})\ \frac{A(x/y)}{\forall xA}, \qquad (\textstyle\bigvee^t_{\exists xA})\ \frac{A(x/t)}{\exists xA},$$

$$(\mathsf{Cut}_C)\ \frac{C \qquad \neg C}{\varnothing}\ (C \in \bigwedge^+\text{-for} \cup \mathcal{L}_0\text{-lit}),$$

$$(\mathsf{Ind}^t_{\mathcal{F}})\ \frac{}{\neg\mathcal{F}(0), \neg\forall x(\mathcal{F}(x) \to \mathcal{F}(Sx)), \mathcal{F}(t)}\,,$$

$$(\mathsf{Cl}_{\mathcal{P}_{\mathfrak{A}}t})\ \frac{\mathfrak{A}(\mathcal{P}_{\mathfrak{A}}, t)}{\mathcal{P}_{\mathfrak{A}}t}, \qquad (\mathsf{Ind}^{\mathcal{P}_{\mathfrak{A}}t}_{\mathcal{F}})\ \frac{}{\neg(\mathfrak{A}(\mathcal{F}){\subseteq}\mathcal{F}), \neg\mathcal{P}_{\mathfrak{A}}t, \mathcal{F}(t)}\,.$$

The inference symbols Ax_Γ, $\bigwedge_{A \wedge B}$, $\bigvee^k_{A \vee B}$, $\bigvee^t_{\exists xA}$, $\mathsf{Cl}_{\mathcal{P}_{\mathfrak{A}}t}$, and Cut_C with $C \in \mathcal{L}_0$-lit are called *simple*.

The proof system ID^*_ν is obtained from ID_ν by adding the following inference symbols

$$(\mathsf{J}^t_{\forall xA})\ \frac{\forall xA}{A(x/t)}\,, \qquad (\mathsf{J}^k_{A_0 \wedge A_1})\ \frac{A_0 \ \wedge\ A_1}{A_k}\,,$$

$$(\mathsf{S}^\Pi_{\mathcal{P},\mathcal{F}})\ \frac{\Pi}{\neg(\mathfrak{A}(\mathcal{F}) \subseteq \mathcal{F}), \Pi(\mathcal{P}/\mathcal{F})} \quad \text{with } \mathcal{P} = \mathcal{P}_{\mathfrak{A}} \text{ and } \Pi \subseteq \mathrm{Pos}_{\mathrm{lev}(\mathcal{P})}\,,$$

$$(\mathsf{E})\ \frac{\varnothing}{\varnothing}\,, \qquad (\mathsf{D}_\sigma)\ \frac{\varnothing}{\varnothing}\ (\sigma < \nu).$$

The role of E and D_σ will become clear in the definition of h^+ below.

Inductive Definition of ID^*_ν-derivations: If $\mathcal{I}$ is an inference symbol of ID^*_ν of arity l and $h_0, \dots, h_{l-1}$ are ID^*_ν-derivations such that for $\Gamma := \Delta(\mathcal{I}) \cup \bigcup_{i<l}(\Gamma(h_i) \setminus \Delta_i(\mathcal{I}))$ we have

- $\mathcal{I} = \bigwedge^y_{\forall xA} \Rightarrow y \notin \mathrm{FV}(\Gamma)$,
- $\mathcal{I} = \mathsf{Cut}_C \Rightarrow \mathrm{FV}(C) \subseteq \mathrm{FV}(\Gamma)$,
- $\mathcal{I} = \bigvee^t_C \Rightarrow \mathrm{FV}(t) \subseteq \mathrm{FV}(\Gamma)$,
- $\mathcal{I} = \mathsf{S}^\Pi_{\mathcal{P},\mathcal{F}} \Rightarrow \mathrm{FV}(\Pi) \subseteq \mathrm{FV}(\Gamma)$ and $h_0 = \mathsf{D}_\sigma h_{00}$ with $\sigma := \mathrm{lev}(\mathcal{P})$,
- $\mathcal{I} = \mathsf{D}_\sigma \Rightarrow \Gamma(h_0) \subseteq \mathrm{Pos}_\sigma\ \&\ \deg(h_0) = 0$,

then $h := \mathcal{I}h_0 \dots h_{l-1}$ is an ID^*_ν-derivation and $\Gamma(h) := \Gamma$ *(endsequent of h)*,

$$\deg(h) := \begin{cases} \deg(h_0) \dot{-} 1 & \text{if } \mathcal{I} = \mathsf{E} \\ \max\{\mathrm{rk}(C), \deg(h_0), \deg(h_1)\} & \text{if } \mathcal{I} = \mathsf{Cut}_C \\ \sup_{i<l} \deg(h_i) & \text{otherwise} \end{cases}$$

An ID_ν-derivation h is called *closed* if its endsequent $\Gamma(h)$ is closed, i.e. if $\mathrm{FV}(\Gamma(h)) = \varnothing$.

Abbreviations.

- ID^*_ν := set of all closed ID^*_ν-derivations.
- $h \vdash_m \Gamma \;:\Leftrightarrow\; h \in \mathrm{ID}^*_\nu$ with $\Gamma(h) \subseteq \Gamma$ and $\deg(h) \le m$.
- $h \vdash^\sigma_m \Gamma \;:\Leftrightarrow\; h \vdash_m \Gamma$ and $\Gamma \subseteq \mathrm{Pos}_\sigma$.
- $\mathbf{I}_\sigma := \{\mathsf{D}_\sigma h : h \vdash^\sigma_0 \Gamma(h)\}$ $(= \{\mathsf{D}_\sigma h : h \in \mathsf{ID}^*_\nu \;\&\; \deg(h) = 0 \;\&\; \Gamma(h) \subseteq \mathrm{Pos}_\sigma\})$ $(\sigma < \nu)$

Definition 0.6 (Substitution of numerals)**.** For $h = \mathcal{I}h_0 \dots h_{n-1}$ let

$$h(y/k) := \begin{cases} h & \text{if } \mathcal{I} = \bigwedge^y_{\forall x A} \\ \mathcal{I}(y/k)h_0(y/k)\dots h_{l-1}(y/k) & \text{otherwise} \end{cases}$$

where $\mathcal{I}(y/k)$ is defined as expected, i.e., in such a way that the following holds:
$h \vdash_m \Gamma \;\Rightarrow\; h(y/k) \vdash_m \Gamma(y/k)$.

Convention. From now on we use h as syntactic variable for closed ID^*_ν-derivations (i.e., elements of ID^*_ν).

Definition 0.7 (The infinitary proof system ID^∞_ν)**.** The language of ID^∞_ν consists of all *closed* $\mathcal{L}_\nu$-formulas.
We use P as syntactic variable for formulas of the form $\mathcal{P}_\mathfrak{A} n$ with $\mathcal{P}_\mathfrak{A} \in \mathcal{L}_\nu$.

The inference symbols of ID^∞_ν are

- All simple inference symbols of ID_ν (restricted to closed formulas) where $\Delta(\mathsf{Ax}_\Gamma)$ is slightly modified, namely

$$\Delta(\mathsf{Ax}_\Gamma) := \begin{cases} \Gamma \cap \mathsf{TRUE}_0 & \text{if } \Gamma \cap \mathsf{TRUE}_0 \ne \varnothing \\ \Gamma & \text{otherwise} \end{cases}$$

- $(\bigwedge_{\forall x A})$ $\dfrac{\dots A(x/i) \dots (i \in \mathbb{N})}{\forall x A}$, (Cut_C) $\dfrac{C \qquad \neg C}{\varnothing}$ $(C \in \bigwedge^+\text{-for})$, (Rep) $\dfrac{\varnothing}{\varnothing}$,
- $(\widetilde{\Omega}_P)$ $\dfrac{P \qquad \dots \Gamma(q) \setminus \{P\} \dots (q \in \mathbf{I}_\mu)}{\varnothing}$ with $\mu := \mathrm{lev}(P)$.

Definition 0.8. $(h^+,\ \mathtt{tp}(h),\ h[\iota])$. To each $h \in \mathsf{ID}^*_\nu$ we assign

- an inference symbol $\mathtt{tp}(h)$ of ID^∞_ν,

- for each $\iota \in |\mathsf{tp}(h)|$, a derivation $h[\iota] \in \mathsf{ID}^*_\nu$.

For the sake of conciseness we write

$$h^+ = \mathcal{I}(h_\iota)_{\iota \in I} \text{ for } \mathsf{tp}(h) = \mathcal{I} \ \& \ |\mathcal{I}| = I \ \& \ \forall \iota \in I(h[\iota] = h_\iota).$$

The definition proceeds by (primitive) recursion on the height of h. In clause 3. we make use of the following abbreviation:

$$\mathsf{Cut}^\circ_C(h_0, h_1) := \begin{cases} \mathsf{Cut}_C(h_0, h_1) & \text{if } C \in \bigwedge^+\text{-for} \cup \mathcal{L}_0\text{-lit} \\ \mathsf{Cut}_{\neg C}(h_1, h_0) & \text{otherwise} \end{cases}$$

Further we denote by $\mathbf{d}_A$ the canonical cutfree ID_ν-derivation of $\{\neg A, A\}$.

1.1. $(\mathcal{I}h_0 ... h_{l-1})^+ := \mathcal{I}(h_i)_{i<l}$ if $\mathcal{I}$ is simple.

1.2. $(\bigwedge^y_{\forall x A} \widetilde{h})^+ := \bigwedge_{\forall x A}\big(\widetilde{h}(y/i)\big)_{i \in \mathbb{N}}$

1.3. $(\mathsf{Ind}^{\mathcal{P}n}_{\mathcal{F}})^+ := \widetilde{\Omega}_{\mathcal{P}n}\mathsf{Ax}_{\{\neg\mathcal{P}n, \mathcal{P}n\}}\big(\mathsf{S}^{\{\mathcal{P}n\}}_{\mathcal{P},\mathcal{F}}\, q\big)_{q \in \mathbf{I}_\mu}$ with $\mu := \mathrm{lev}(\mathcal{P})$.

2. $(\mathsf{Ind}^n_{\mathcal{F}})^+ := \mathsf{Rep}(d_n)$ with $d_0 := \mathbf{d}_{\mathcal{F}(0)}$,

$$d_{i+1} := \bigvee^i_{\exists x(\mathcal{F}(x)\,\wedge\,\neg\mathcal{F}(Sx))} \bigwedge_{\mathcal{F}(i)\,\wedge\,\neg\mathcal{F}(Si)} d_i \mathbf{d}_{\mathcal{F}(Si)}$$

3. If $C \in \bigwedge^+$-for and $h_1^+ = \mathcal{I}(h_{1\iota})_{\iota \in I}$ then:

$$(\mathsf{Cut}_C h_0 h_1)^+ := \begin{cases} \mathcal{I}\big(\mathsf{Cut}_C h_0 h_{1\iota}\big)_{\iota \in I} & \text{if } \neg C \notin \Delta(\mathcal{I}) \\ \mathsf{Cut}^\circ_{C[k]}\big(\mathsf{J}^k_C h_0, \mathsf{Cut}_C h_0 h_{10}\big) & \text{if } \mathcal{I} = \bigvee^k_{\neg C} \\ \mathsf{Rep}(h_0) & \text{if } \neg C \in \Delta(\mathcal{I}) \text{ and } C = \mathcal{P}n \end{cases}$$

4. If $h^+ = \mathcal{I}(h_\iota)_{\iota \in I}$ then

$$(\mathsf{E}h)^+ := \begin{cases} \mathsf{Rep}\big(\mathsf{Cut}_C \mathsf{E}h_0 \mathsf{E}h_1\big) & \text{if } \mathcal{I} = \mathsf{Cut}_C \text{ with } C \in \bigwedge^+\text{-for} \\ \mathcal{I}\big(\mathsf{E}h_\iota\big)_{\iota \in I} & \text{otherwise} \end{cases}$$

5. If $C \in \bigwedge$-for and $h^+ = \mathcal{I}(h_\iota)_{\iota \in I}$ then

$$(\mathsf{J}^k_C h)^+ := \begin{cases} \mathsf{Rep}\big(\mathsf{J}^k_C h_k\big) & \text{if } \mathcal{I} = \bigwedge_C \\ \mathcal{I}\big(\mathsf{J}^k_C h_\iota\big)_{\iota \in I} & \text{otherwise} \end{cases}$$

6. If $\mathcal{P} = \mathcal{P}_{\mathfrak{A}}$, $\mu := \mathrm{lev}(\mathcal{P})$ $(< \nu)$, and $d \in \mathbf{I}_\mu$ with $d^+ = \mathcal{I}(d_\iota)_{\iota \in I}$ then

$$(\mathsf{S}^{\Pi}_{\mathcal{P},\mathcal{F}} d)^+ := \begin{cases} \bigvee^{n}_{\neg(\mathfrak{A}(\mathcal{F}) \subseteq \mathcal{F})} \big(\bigwedge_{\mathfrak{A}(\mathcal{F},n) \wedge \neg\mathcal{F}(n)} (\mathsf{S}^{\Pi \cup \Delta_0(\mathcal{I})}_{\mathcal{P},\mathcal{F}} d_0) \mathbf{d}_{\mathcal{F}(n)}\big) & \text{if } \mathcal{I} = \mathsf{Cl}_{\mathcal{P}n} \text{ with } \mathcal{P}n \in \Pi \\ \mathcal{I}^* \big(\mathsf{S}^{\Pi \cup \Delta_\iota(\mathcal{I})}_{\mathcal{P},\mathcal{F}} d_\iota\big)_{\iota \in I} & \text{if } \mathcal{I} = \bigwedge_A, \bigvee^k_A \text{ with } A \in \Pi \\ \mathcal{I}\big(\mathsf{S}^{\Pi}_{\mathcal{P},\mathcal{F}} d_\iota\big)_{\iota \in I} & \text{otherwise} \end{cases}$$

where $(\bigwedge_A)^* := \bigwedge_{A(\mathcal{P}/\mathcal{F})}$, $(\bigvee^k_A)^* := \bigvee^k_{A(\mathcal{P}/\mathcal{F})}$.

7. If $h^+ = \mathcal{I}\big(h_\iota\big)_{\iota \in I}$ then

$$(\mathsf{D}_\sigma h)^+ := \begin{cases} \mathsf{Rep}\big(\mathsf{D}_\sigma h_{\mathsf{D}_\mu h_0}\big) & \text{if } \mathcal{I} = \widetilde{\Omega}_P \text{ with } \mu := \mathrm{lev}(P) \geq \sigma \\ \mathcal{I}\big(\mathsf{D}_\sigma h_\iota\big)_{\iota \in I} & \text{otherwise} \end{cases}$$

Definition 0.9.

- $\mathrm{ID}^\infty_\nu \lceil \sigma := \mathrm{ID}^\infty_\nu \setminus \{\widetilde{\Omega}_P : \mathrm{lev}(P) \geq \sigma\}$
- $\deg(\mathcal{I}) := \begin{cases} \mathrm{rk}(C) + 1 & \text{if } \mathcal{I} = \mathsf{Cut}_C \text{ with } C \in \bigwedge^+\text{-for} \\ 0 & \text{otherwise} \end{cases}$

Lemma 0.10. *If* $h \vdash^\sigma_m \Gamma$ & $h^+ = \mathcal{I}(h_\iota)_{\iota \in I}$ *then*

$$\mathcal{I} \in \mathrm{ID}^\infty_\nu \lceil \sigma \;\;\&\;\; \Delta(\mathcal{I}) \subseteq \Gamma \;\;\&\;\; \deg(\mathcal{I}) \leq m \;\;\&\;\; \forall \iota \in I (h_\iota \vdash^\sigma_m \Gamma, \Delta_\iota(\mathcal{I})).$$

Proof. The proof of this lemma is routine and can be left to the reader (cf. Theorem 3 in [Bu97] and Theorem 5 in [Bu01]). □

Definition 0.11 (Iterated Inductive Definition of $\mathcal{W}_\sigma$ $(\sigma < \nu)$)**.**

1. If $h \in \mathbf{I}_\sigma$ with $|\mathsf{tp}(h)| \subseteq \mathbb{N}$ and $\forall i \in |\mathsf{tp}(h)| (h[i] \in \mathcal{W}_\sigma)$ then $h \in \mathcal{W}_\sigma$.
2. If $h \in \mathbf{I}_\sigma$ with $\mathsf{tp}(h) = \widetilde{\Omega}_P$, $\mathrm{lev}(P) < \sigma$ and $\forall \iota \in \mathcal{W}_{\mathrm{lev}(P)} (h[\iota] \in \mathcal{W}_\sigma)$ then $h \in \mathcal{W}_\sigma$.

Note that (according to Lemma 0.10) if $h \in \mathbf{I}_\sigma$ and $\mathsf{tp}(h) = \widetilde{\Omega}_P$ then $\mathrm{lev}(P) < \sigma$.

Note further that $\mathcal{W}_\sigma$ is by definition a subset of $\mathbf{I}_\sigma$.

Our goal is now to show that ID_ν is Π^0_2-conservative over $\mathrm{ID}^i_\nu(\mathcal{W})$ (where $\mathcal{W}$ denotes the operator form corresponding to the iterated inductive definition of $(\mathcal{W}_\sigma)_{\sigma<\nu}$). We will achieve this goal by giving an informal proof of

$$\text{“If } h \text{ is an } \mathrm{ID}_\nu\text{-derivation of a } \Pi^0_2\text{-sentence } A \text{ and if } h \text{ has height and degree} \le m \text{ then } A \text{ holds.”} \tag{1}$$

which for each fixed $m \in \mathbb{N}$ can be formalized in $\mathrm{ID}^i_\nu(\mathcal{W})$.

Abbreviations.

- $\mathcal{W}^* := \{h : \forall \sigma < \nu(h \vdash^\sigma_0 \Gamma(h) \Rightarrow \mathsf{D}_\sigma h \in \mathcal{W}_\sigma)\}$,
- $\mathsf{FALSE}_0 := \{\neg A : A \in \mathsf{TRUE}_0\}$,
- $\mathsf{E}^m h := \underbrace{\mathsf{E}\ldots\mathsf{E}}_{m\,\text{times}} h$.

Lemma 0.12. *Let R be a binary relation symbol of $\mathcal{L}_0$.*

(a) If $\widetilde{h}$ is an ID_ν-derivation of $\exists y R(x,y)$ with $\deg(\widetilde{h}) = m$, then for all n we have:

$$\mathsf{E}^m\widetilde{h}(x/n) \in \mathcal{W}^* \;\Rightarrow\; \mathcal{W}_0 \ni \mathsf{D}_0\mathsf{E}^m\widetilde{h}(x/n) \vdash \exists y R(n,y).$$

(b) $\mathcal{W}_0 \ni h \vdash \Gamma, \exists y R(n,y)$ with $\Gamma \subseteq \mathsf{FALSE}_0 \;\Rightarrow$ there exists k with $R(n,k)$.

Proof. (a) Obviously $\mathsf{E}^m h(x/n) \vdash^0_0 \exists y R(n,y)$ which yields the claim.

(b) Induction over $\mathcal{W}_0$: We have $h^+ = \mathcal{I}(h_i)_{i\in I}$ with $h_i \in \mathcal{W}_0$ for all $i \in I$. By Lemma 0.10 one of the following cases holds:

1. $\mathcal{I} = \mathsf{Rep}$ and $h_0 \vdash \Gamma, \exists y R(n,y)$.
2. $\mathcal{I} = \mathsf{Cut}_C$ with $C \in \mathsf{FALSE}_0$ and $h_0 \vdash \Gamma, C, \exists y R(n,y)$.
3. $\mathcal{I} = \mathsf{Cut}_C$ with $\neg C \in \mathsf{FALSE}_0$ and $h_1 \vdash \Gamma, \neg C, \exists y R(n,y)$.
4. $\mathcal{I} = \bigvee^k_{\exists y R(n,y)}$ with $R(n,k) \in \mathsf{FALSE}_0$ and $h_0 \vdash \Gamma, R(n,k), \exists y R(n,y)$.
5. $\mathcal{I} = \bigvee^k_{\exists y R(n,y)}$ and $R(n,k) \in \mathsf{TRUE}_0$.

In cases 1–4 the claim follows immediately from the IH (induction hypothesis). In case 5 we are done. $\square$

Now for establishing (1) it remains to prove:

$\mathsf{E}^m h \in \mathcal{W}^*$ holds for each closed ID_ν-derivation h and each $m \in \mathbb{N}$. (2)

Definition 0.13. For $\mathcal{I} \in \mathrm{ID}_\nu^\infty$ let

$$|\mathcal{I}|_{\mathcal{W}} := \begin{cases} \{0\} \cup \mathcal{W}_\mu & \text{if } \mathcal{I} = \widetilde{\Omega}_P \text{ and } \mu = \mathrm{lev}(P) \\ |\mathcal{I}| & \text{if } \mathcal{I} \text{ is not of the form } \widetilde{\Omega}_P \end{cases}$$

Note that $|\mathcal{I}|_{\mathcal{W}} \subseteq |\mathcal{I}|$ (since $\mathcal{W}_\mu \subseteq \mathbf{I}_\mu$).

$\Phi(\mathcal{X}) := \{h : \forall \iota \in |\mathsf{tp}(h)|_{\mathcal{W}}(h[\iota] \in \mathcal{X})\}$ and $\mathrm{Prog}(\mathcal{X}) \;:\Leftrightarrow\; \Phi(\mathcal{X}) \subseteq \mathcal{X}$, where $\mathcal{X}$ ranges over subsets of ID_ν^*.

Then $\mathcal{W}_\sigma$ (for $\sigma < \nu$) satisfies the following "axioms":

$(W_\sigma.1)$ $\mathbf{I}_\sigma \cap \Phi(\mathcal{W}_\sigma) \subseteq \mathcal{W}_\sigma$,

$(W_\sigma.2)$ $\mathbf{I}_\sigma \cap \Phi(\mathcal{X}) \subseteq \mathcal{X} \;\Rightarrow\; \mathcal{W}_\sigma \subseteq \mathcal{X}$.

Lemma 0.14. $\mathrm{Prog}(\mathcal{W}^*)$.

Proof. Let $\mathrm{H}_\sigma := \{h : \deg(h) = 0 \;\&\; \Gamma(h) \subseteq \mathrm{Pos}_\sigma\}$. Then $\mathcal{W}^* = \{h : \forall \sigma < \nu(h \in \mathrm{H}_\sigma \Rightarrow \mathsf{D}_\sigma h \in \mathcal{W}_\sigma)\}$.

Suppose $h \in \Phi(\mathcal{W}^*) \;\&\; \sigma < \nu \;\&\; h \in \mathrm{H}_\sigma$.

To prove: $\mathsf{D}_\sigma h \in \mathcal{W}_\sigma$. Trivially $\mathsf{D}_\sigma h \in \mathbf{I}_\sigma$.

1. $\mathsf{tp}(h) = \widetilde{\Omega}_P$ with $\sigma \leq \mu := \mathrm{lev}(P)$: From $h \in \mathrm{H}_\sigma$ by Lemma 0.10 we get $h[0] \in \mathrm{H}_\sigma \subseteq \mathrm{H}_\mu$. Together with $h \in \Phi(\mathcal{W}^*)$ this yields $q := \mathsf{D}_\mu h[0] \in \mathcal{W}_\mu$. From $q \in \mathcal{W}_\mu$ and $h \in H_\sigma \cap \Phi(\mathcal{W}^*)$ we conclude $h[q] \in H_\sigma \cap \mathcal{W}^*$. Hence $\mathsf{D}_\sigma h[q] \in \mathcal{W}_\sigma$ which yields $\mathsf{D}_\sigma h \in \mathcal{W}_\sigma$, since $(\mathsf{D}_\sigma h)^+ = \mathsf{Rep}(\mathsf{D}_\sigma h[q])$.

2. Otherwise: Then $\mathsf{tp}(\mathsf{D}_\sigma h) = \mathsf{tp}(h)$, $|\mathsf{tp}(h)|_{\mathcal{W}} \subseteq |\mathsf{tp}(h)|$ and $(\mathsf{D}_\sigma h)[\iota] = \mathsf{D}_\sigma h[\iota]$ for all $\iota \in |\mathsf{tp}(h)|$ $(*)$.

From $h \in \mathrm{H}_\sigma \cap \Phi(\mathcal{W}^*)$ by L.1 we get $\forall \iota \in |\mathsf{tp}(h)|_{\mathcal{W}}(h[\iota] \in \mathrm{H}_\sigma \cap \mathcal{W}^*)$, and then $\forall \iota \in |\mathsf{tp}(h)|_{\mathcal{W}}(\mathsf{D}_\sigma h[\iota] \in \mathcal{W}_\sigma)$. Together with $(*)$ this yields $\mathsf{D}_\sigma h \in \mathcal{W}_\sigma$. □

Remark 0.15. Now for establishing (2) it remains to prove

$$\mathrm{Prog}(\mathcal{X}) \Rightarrow h \in \mathcal{X}, \quad \text{for each closed } \mathrm{ID}_\nu\text{-derivation } h \text{ and each } \mathcal{X}, \tag{3}$$

and to find a *jump* operation $\mathcal{X} \mapsto \overline{\mathcal{X}}$ (á la [Ge43]) such that

$$h \in \overline{\mathcal{X}} \Rightarrow \mathsf{E}h \in \mathcal{X} \quad \text{and} \quad \mathrm{Prog}(\mathcal{X}) \Rightarrow \mathrm{Prog}(\overline{\mathcal{X}}). \tag{4}$$

Lemma 0.16. $\mathrm{Prog}(\mathcal{X}) \;\&\; \mathrm{lev}(\mathcal{P}) = \sigma < \nu \;\&\; d \in \mathcal{W}_\sigma \quad\Rightarrow\quad \mathsf{S}^{\Pi}_{\mathcal{P},\mathcal{F}} d \in \mathcal{X}$.

Proof. By induction on "$d \in \mathcal{W}_\sigma$": Assume $d \in \mathcal{W}_\sigma$ with $d^+ = \mathcal{I}(d_\iota)_{\iota\in|\mathcal{I}|}$. Then $d \in \mathbf{I}_\sigma$ and $\forall\iota \in |\mathcal{I}|_{\mathcal{W}}(d_\iota \in \mathcal{W}_\sigma)$. We have to prove: $h := \mathsf{S}^{\Pi}_{\mathcal{P},\mathcal{F}} d \in \mathcal{X}$.

1.1. $\mathcal{I} = \mathsf{Cl}_{\mathcal{P}n}$ with $\mathcal{P}n \in \Pi$: Then

$$h^+ = \bigvee\nolimits^{n}_{\neg(\mathfrak{A}(\mathcal{F})\subseteq\mathcal{F})}\big(\bigwedge\nolimits_{\mathfrak{A}(\mathcal{F},n)\,\wedge\,\neg\mathcal{F}(n)}(\mathsf{S}^{\Pi\cup\Delta_0(\mathcal{I})}_{\mathcal{P},\mathcal{F}} d_0)\mathbf{d}_{\mathcal{F}(n)}\big). \qquad (*)$$

By IH from $d_0 \in \mathcal{W}_\sigma$ we get $\mathsf{S}^{\Pi\cup\Delta_0(\mathcal{I})}_{\mathcal{P},\mathcal{F}} d_0 \in \mathcal{X}$. Further, the premise $\mathrm{Prog}(\mathcal{X})$ yields $\mathbf{d}_{\mathcal{F}(n)} \in \mathcal{X}$.

From $\mathsf{S}^{\Pi\cup\Delta_0(\mathcal{I})}_{\mathcal{P},\mathcal{F}} d_0 \in \mathcal{X}$ & $\mathbf{d}_{\mathcal{F}(n)} \in \mathcal{X}$ by $(*)$ and $\mathrm{Prog}(\mathcal{X})$ we get $h \in \mathcal{X}$.

1.2. $\mathcal{I} = \bigwedge_A, \bigvee^k_A$ with $A \in \Pi$: Then $h^+ = \mathcal{I}^*\big(\mathsf{S}^{\Pi\cup\Delta_i(\mathcal{I})}_{\mathcal{P},\mathcal{F}} d_i\big)_{i\in|\mathcal{I}|}$ $(*)$.

By IH we get $\forall i \in |\mathcal{I}|(\mathsf{S}^{\Pi\cup\Delta_i(\mathcal{I})}_{\mathcal{P},\mathcal{F}} d_i \in \mathcal{X})$, and then $h \in \mathcal{X}$ by $(*)$ and $\mathrm{Prog}(\mathcal{X})$.

1.3. otherwise: Then $h^+ = \mathcal{I}\big(\mathsf{S}^{\Pi}_{\mathcal{P},\mathcal{F}} d_\iota\big)_{i\in|\mathcal{I}|}$ $(*)$.

By IH we get $\forall\iota \in |\mathcal{I}|_{\mathcal{W}}(\mathsf{S}^{\Pi}_{\mathcal{P},\mathcal{F}} d_\iota \in \mathcal{X})$, and then $h \in \mathcal{X}$ by $(*)$ and $\mathrm{Prog}(\mathcal{X})$. □

Lemma 0.17. $\quad \mathrm{Prog}(\mathcal{X})$ & $C \in \bigwedge$*-for* $\quad\Rightarrow\quad \mathrm{Prog}(\{h_0 : \mathsf{J}^k_C h_0 \in \mathcal{X}\})$.

Proof. Left to the reader. □

Definition 0.18. $\mathcal{X}^{C,h_0} := \{h_1 : \mathsf{Cut}_C h_0 h_1 \in \mathcal{X}\}$

Lemma 0.19. *Assume* $\mathrm{Prog}(\mathcal{X})$.

(a) $C \in \bigwedge$*-for* & $\forall k(\mathsf{J}^k_C h_0 \in \mathcal{X}) \quad\Rightarrow\quad \mathrm{Prog}(\mathcal{X}^{C,h_0})$

(b) $h_0 \in \mathcal{X} \quad\Rightarrow\quad \mathrm{Prog}(\mathcal{X}^{P,h_0})$.

Proof. (a) Assume $C \in \bigwedge$-for & $\forall k(\mathsf{J}^k_C h_0 \in \mathcal{X})$ & $h_1 \in \Phi(\mathcal{X}^{C,h_0})$.

To prove: $h_1 \in \mathcal{X}^{C,h_0}$, i.e. $h := \mathsf{Cut}_C h_0 h_1 \in \mathcal{X}$.

Assume $h_1^+ = \mathcal{I}(h_{1\iota})_{\iota\in I}$.

Then $\forall\iota \in |\mathcal{I}|_{\mathcal{W}}(h_{1\iota} \in \mathcal{X}^{C,h_0})$ and thus $\forall\iota \in |\mathcal{I}|_{\mathcal{W}}(\mathsf{Cut}_C h_0 h_{1\iota} \in \mathcal{X})$.

1. $\neg C \notin \Delta(\mathcal{I})$: From $h^+ = \mathcal{I}(\mathsf{Cut}_C h_0 h_{1\iota})_{\iota\in I}$ and $\forall\iota \in |\mathcal{I}|_{\mathcal{W}}(\mathsf{Cut}_C h_0 h_{1\iota} \in \mathcal{X})$ we get $h \in \Phi(\mathcal{X}) \subseteq \mathcal{X}$.

2. $\mathcal{I} = \bigvee^k_{\neg C}$: Then $h^+ = \mathsf{Cut}^{\circ}_{C[k]}(\mathsf{J}^k_C h_0, \mathsf{Cut}_C h_0 h_{10})$ with $\mathsf{J}^k_C h_0 \in \mathcal{X}$ (by assumption) and $\mathsf{Cut}_C h_0 h_{10} \in \mathcal{X}$ as shown above. Hence $h \in \Phi(\mathcal{X}) \subseteq \mathcal{X}$.

(b) is proved in the same way as (a). □

Lemma 0.20. *For each closed* ID_ν*-derivation* h *and each* $\mathcal{X}$ *we have:*

$$\mathrm{Prog}(\mathcal{X}) \quad\Rightarrow\quad h \in \mathcal{X}.$$

Proof. By induction on the height of h: Assume $\mathrm{Prog}(X)$.

1. $h = \mathcal{I}h_0...h_{l-1}$ with simple $\mathcal{I}$: Then $h^+ = \mathcal{I}\big(h_i\big)_{i<l}$ and, by IH, $h_0,\ldots,h_{l-1} \in \mathcal{X}$. Hence $h \in \Phi(\mathcal{X}) \subseteq \mathcal{X}$.

2. $h = \bigwedge^y_{\forall x A}\widetilde{h}$: Then $h^+ = \bigwedge_{\forall x A}(\widetilde{h}(y/i))_{i\in\mathbb{N}}$ and, by IH, $\forall i \in \mathbb{N}(\widetilde{h}(y/i) \in \mathcal{X})$, i.e. $h \in \Phi(\mathcal{X}) \subseteq \mathcal{X}$.

3. $h = \mathsf{Cut}_C h_0 h_1$ with $C \in \bigwedge^+$-for: By IH we get $h_0 \in \mathcal{X}$.

3.1. $C \in \bigwedge$-for: By Lemma 0.17 we get $\forall k.\mathrm{Prog}(\{d : \mathsf{J}^k_C d \in \mathcal{X}\})$ and then, by IH, $\forall k(h_0 \in \{d : \mathsf{J}^k_C d \in \mathcal{X}\})$, i.e. $\forall k(\mathsf{J}^k_C h_0 \in \mathcal{X})$. From $\mathrm{Prog}(\mathcal{X})$ & $h_0 \in \mathcal{X}$ & $\forall k(\mathsf{J}^k_C h_0 \in \mathcal{X})$ by Lemma 0.19 a we conclude $\mathrm{Prog}(\mathcal{X}^{C,h_0})$ and then, by IH, $h_1 \in \mathcal{X}^{C,h_0}$, i.e. $h \in \mathcal{X}$.

3.2. $C = P$: From $\mathrm{Prog}(\mathcal{X})$ & $h_0 \in \mathcal{X}$ by Lemma 0.19 b we conclude $\mathrm{Prog}(\mathcal{X}^{P,h_0})$ and then, by IH, $h_1 \in \mathcal{X}^{P,h_0}$, i.e. $h \in \mathcal{X}$.

4. $h = \mathsf{Ind}^n_{\mathcal{F}}$: Then $h^+ = \mathsf{Rep}(d_n)$ with $d_0 := \mathbf{d}_{\mathcal{F}(0)}$,
$d_{i+1} := \bigvee^i_{\exists x(\mathcal{F}(x)\wedge\neg\mathcal{F}(Sx))}\bigwedge_{\mathcal{F}(i)\wedge\neg\mathcal{F}(Si)} d_i \mathbf{d}_{\mathcal{F}(Si)}$.
Using $\mathrm{Prog}(\mathcal{X})$ one easily shows $d_i \in \mathcal{X}$ by induction on i.

5. $h = \mathsf{Ind}^{\mathcal{P}n}_{\mathcal{F}}$: Then $h^+ = \widetilde{\Omega}_P \mathsf{Ax}_{\{\neg P,P\}}\big(\mathsf{S}^{\{P\}}_{\mathcal{P},\mathcal{F}} d\big)_{d\in\mathbf{I}_\sigma}$ with $\sigma := \mathrm{lev}(\mathcal{P})$ and $P := \mathcal{P}n$.

$\mathrm{Prog}(\mathcal{X})$ yields $\mathsf{Ax}_{\{\neg P,P\}} \in \mathcal{X}$, and by Lemma 0.16 we have $\forall d \in \mathcal{W}_\sigma(\mathsf{S}^{\{P\}}_{\mathcal{P},\mathcal{F}} d \in \mathcal{X})$. Hence $h \in \Phi(\mathcal{X}) \subseteq \mathcal{X}$. □

Now we come to the last part of our proof, which begins with the definition of the jump operation $\mathcal{X} \mapsto \overline{\mathcal{X}}$ mentioned in (4) above.

Remark. ID^*_ν-derivations have been introduced as terms in polish (prefix) notation build up from inference symbols each of which has a fixed finite arity. So every ID^*_ν-derivation is a finite sequence of inference symbols.

In the following we use $\mathfrak{a}$, $\mathfrak{a}'$ as syntactic variables for arbitrary finite sequences of inference symbols – including the empty sequence ε. Concatination is expressed by juxtaposition. Example: If $\mathfrak{a} = \mathsf{Cut}_C h_0 \mathsf{J}^k_D \mathsf{Cut}_B h_1$ then $\mathfrak{a}h_2$ is the derivation $\mathsf{Cut}_C h_0 h$ with $h := \mathsf{J}^k_D \mathsf{Cut}_B h_1 h_2$.

Definition 0.21 (Finitary Inductive Definition of $\mathbf{Q}(\mathcal{X})$)**.**

- (Q1) $\varepsilon \in \mathbf{Q}(\mathcal{X})$.
- (Q2) $\mathfrak{a} \in \mathbf{Q}(\mathcal{X})$ & $C \in \bigwedge$-for $\Rightarrow$ $\mathfrak{a}\,\mathsf{J}^k_C \in \mathbf{Q}(\mathcal{X})$.
- (Q3) $\mathfrak{a} \in \mathbf{Q}(\mathcal{X})$ & $C \in \bigwedge$-for & $\forall k(\mathfrak{a}\,\mathsf{J}^k_C h \in \mathcal{X})$ $\Rightarrow$ $\mathfrak{a}\,\mathsf{Cut}_C h \in \mathbf{Q}(\mathcal{X})$.
- (Q4) $\mathfrak{a} \in \mathbf{Q}(\mathcal{X})$ & $\mathfrak{a}h \in \mathcal{X}$ $\Rightarrow$ $\mathfrak{a}\,\mathsf{Cut}_P h \in \mathbf{Q}(\mathcal{X})$.

Note that $\mathbf{Q}(\mathcal{X})$ is arithmetical in $\mathcal{X}$.

Definition 0.22. $\overline{\mathcal{X}} := \{h : \forall \mathfrak{a} \in \mathbf{Q}(\mathcal{X})(\mathfrak{a}\mathsf{E}h \in \mathcal{X})\}$.

Remark 0.23.

(i) $h \in \overline{\mathcal{X}} \;\Rightarrow\; \mathsf{E}h \in \mathcal{X}$.

(ii) $h \in \overline{\mathcal{X}}\ \&\ \mathfrak{a} \in \mathbf{Q}(\mathcal{X})\ \&\ C \in \bigwedge^{+}\text{-for} \;\Rightarrow\; \mathfrak{a}\mathsf{Cut}_C\mathsf{E}h \in \mathbf{Q}(\mathcal{X})$.

Lemma 0.24. *Let* $\mathfrak{a} \in \mathbf{Q}(\mathcal{X})$.

(a) $h^+ = \mathsf{Cut}_A(h_0, h_1) \;\Rightarrow\; (\mathfrak{a}h)^+ = \mathsf{Cut}_A(\mathfrak{a}h_0, \mathfrak{a}h_1)$.

(b) $h^+ = \mathsf{Rep}(h_0) \;\Rightarrow\; (\mathfrak{a}h)^+ = \mathsf{Rep}(\mathfrak{a}h_0)$.

Proof. By induction on "$\mathfrak{a} \in \mathbf{Q}(\mathcal{X})$". □

Lemma 0.25. *If* $\mathrm{Prog}(\mathcal{X})$ *and* $\mathfrak{a} \in \mathbf{Q}(\mathcal{X})$ *then the following holds:*

$$h^+ = \mathcal{I}(h_\iota)_{\iota\in I}\ \&\ \forall\iota \in |\mathcal{I}|_{\mathcal{W}}(\mathfrak{a}h_\iota \in \mathcal{X}) \;\Rightarrow\; \mathfrak{a}h \in \mathcal{X}$$

Proof. By induction on "$\mathfrak{a} \in \mathbf{Q}(\mathcal{X})$":

1. $\mathfrak{a} = \varepsilon$: In this case the premises immediately yield $h \in \Phi(\mathcal{X}) \subseteq \mathcal{X}$.

2. $\mathfrak{a} = \mathfrak{a}'\mathsf{J}^k_C$ with $\mathfrak{a}' \in \mathbf{Q}(\mathcal{X})$:

2.1. $\mathcal{I} = \bigwedge_C$: Then $(\mathsf{J}^k_C h)^+ = \mathsf{Rep}(\mathsf{J}^k_C h_k)$ and $(\mathfrak{a}h)^+ = (\mathfrak{a}'\mathsf{J}^k_C h)^+ \overset{\text{L.8b}}{=} \mathsf{Rep}(\mathfrak{a}'\mathsf{J}^k_C h_k) = \mathsf{Rep}(\mathfrak{a}h_k)$.
From $\mathrm{Prog}(\mathcal{X})\ \&\ (\mathfrak{a}h)^+ = \mathsf{Rep}(\mathfrak{a}h_k)\ \&\ \mathfrak{a}h_k \in \mathcal{X}$ we get $\mathfrak{a}h \in \Phi(\mathcal{X}) \subseteq \mathcal{X}$.

2.2. otherwise: Then $(\mathsf{J}^k_C h)^+ = \mathcal{I}\big(\mathsf{J}^k_C h_\iota\big)_{\iota\in I}$ $(*)$.

$\mathrm{Prog}(\mathcal{X})\ \&\ \mathfrak{a}' \in \mathbf{Q}(\mathcal{X})\ \&\ (*)\ \&\ \forall\iota \in |\mathcal{I}|_{\mathcal{W}}(\mathfrak{a}'\mathsf{J}^k_C h_\iota = \mathfrak{a}h_\iota \in \mathcal{X}) \overset{\text{IH}}{\Rightarrow} \mathfrak{a}h = \mathfrak{a}'\mathsf{J}^k_C h \in \mathcal{X}$.

3. $\mathfrak{a} = \mathfrak{a}'\mathsf{Cut}_C h'$ with $\mathfrak{a}' \in \mathbf{Q}(\mathcal{X})\ \&\ C \in \bigwedge\text{-for}\ \&\ \forall k(\mathfrak{a}'\mathsf{J}^k_C h' \in \mathcal{X})$:

3.1. $\neg C \notin \Delta(\mathcal{I})$: Then $(\mathsf{Cut}_C h'h)^+ = \mathcal{I}\big(\mathsf{Cut}_C h'h_\iota\big)_{\iota\in I}$ $(*)$.

$\mathrm{Prog}(\mathcal{X})\ \&\ \mathfrak{a}' \in \mathbf{Q}(\mathcal{X})\ \&\ (*)\ \&\ \forall\iota \in |\mathcal{I}|_{\mathcal{W}}(\mathfrak{a}'\mathsf{Cut}_C h'h_\iota = \mathfrak{a}h_\iota \in \mathcal{X}) \overset{\text{IH}}{\Rightarrow} \mathfrak{a}h = \mathfrak{a}'\mathsf{Cut}_C h'h \in \mathcal{X}$.

3.2. $\mathcal{I} = \bigvee^k_{\neg C}$: Then $(\mathsf{Cut}_C h'h)^+ = \mathsf{Cut}^\circ_{C[k]}(\mathsf{J}^k_C h', \mathsf{Cut}_C h'h_0)$ and

$$(\mathfrak{a}h)^+ = (\mathfrak{a}'\mathsf{Cut}_C h'h)^+ \overset{\text{L.8a}}{=} \mathsf{Cut}^\circ_{C[k]}(\mathfrak{a}'\mathsf{J}^k_C h', \mathfrak{a}'\mathsf{Cut}_C h'h_0) = \mathsf{Cut}^\circ_{C[k]}(\mathfrak{a}'\mathsf{J}^k_C h', \mathfrak{a}h_0).$$

Further $\mathfrak{a}'\mathsf{J}^k_C h' \in \mathcal{X}$ and $\mathfrak{a}h_0 \in \mathcal{X}$. Hence $\mathfrak{a}h \in \Phi(\mathcal{X}) \subseteq \mathcal{X}$.

4. $\mathfrak{a} = \mathfrak{a}'\mathsf{Cut}_P h'$ with $\mathfrak{a}' \in \mathbf{Q}(\mathcal{X})$ & $\mathfrak{a}'h' \in \mathcal{X}$:
4.1. $\neg P \notin \Delta(\mathcal{I})$: As 3.1.
4.2. $\neg P \in \Delta(\mathcal{I})$: Then $(\mathsf{Cut}_P h'h)^+ = \mathsf{Rep}(h')$ and thus $(\mathfrak{a}h)^+ = (\mathfrak{a}'\mathsf{Cut}_C h'h)^+ \stackrel{\text{L.8b}}{=} \mathsf{Rep}(\mathfrak{a}'h')$.
Together with $\mathfrak{a}'h' \in \mathcal{X}$ this yields $\mathfrak{a}h \in \Phi(\mathcal{X}) \subseteq \mathcal{X}$. □

Lemma 0.26. $\mathrm{Prog}(\mathcal{X}) \Rightarrow \mathrm{Prog}(\overline{\mathcal{X}})$.

Proof. Assume $\mathrm{Prog}(\mathcal{X})$ & $h \in \Phi(\overline{\mathcal{X}})$ & $\mathfrak{a} \in \mathbf{Q}(\mathcal{X})$. To prove $\mathfrak{a}\mathsf{E}h \in \mathcal{X}$. For this it suffices to prove $\mathfrak{a}\mathsf{E}h \in \Phi(\mathcal{X})$.

Let $h^+ = \mathcal{I}(h_\iota)_{\iota \in I}$. Then $\forall \iota \in |\mathcal{I}|_{\mathcal{W}}(h_\iota \in \overline{\mathcal{X}})$ and thus $\forall \iota \in |\mathcal{I}|_{\mathcal{W}}(\mathfrak{a}\mathsf{E}h_\iota \in \mathcal{X})$.

1. $\mathcal{I} = \mathsf{Cut}_C$ with $C \in \bigwedge^+$-for: Then $(\mathsf{E}h)^+ = \mathsf{Rep}(\mathsf{Cut}_C \mathsf{E}h_0 \mathsf{E}h_1)$ and therefore, by Lemma 0.24 b,$(\mathfrak{a}\mathsf{E}h)^+ = \mathsf{Rep}(\mathfrak{a}\mathsf{Cut}_C \mathsf{E}h_0 \mathsf{E}h_1)$.

From $h_0, h_1 \in \overline{\mathcal{X}}$ & $\mathfrak{a} \in \mathbf{Q}(\mathcal{X})$ we get (by Remark (ii)) $\mathfrak{a}\mathsf{Cut}_C \mathsf{E}h_0 \in \mathbf{Q}(\mathcal{X})$ & $h_1 \in \overline{\mathcal{X}}$, and then $\mathfrak{a}\mathsf{Cut}_C \mathsf{E}h_0 \mathsf{E}h_1 \in \mathcal{X}$. Hence $\mathfrak{a}\mathsf{E}h \in \Phi(\mathcal{X}) \subseteq \mathcal{X}$.

2. otherwise: From $(\mathsf{E}h)^+ = \mathcal{I}(\mathsf{E}h_\iota)_{\iota \in I}$ & $\forall \iota \in |\mathcal{I}|_{\mathcal{W}}(\mathfrak{a}\mathsf{E}h_\iota \in \mathcal{X})$ we conclude $\mathfrak{a}\mathsf{E}h \in \mathcal{X}$ by Lemma 0.25. □

References

AT09 J. Avigad and H. Towsner, Functional interpretation and inductive definitions, JSL 74 (2009), pp. 1100-1120

BFPS W. Buchholz, S. Feferman, W. Pohlers and W. Sieg, *Iterated Inductive Definitions and Subsystems of Analysis: Recent proof-theoretical studies*, LNM 897, Springer (1981).

Bu81 W. Buchholz: The $\Omega_{\mu+1}$-rule, in Buchholz et al. (1981), 188-233.

Bu91 W. Buchholz: Notation systems for infinitary derivations, Arch. Math. Logic 30, pp. 277-296 (1991)

Bu95 W. Buchholz: On Gentzen's consistency proofs for arithmetic. Oberwolfach 1995.

Bu97 W. Buchholz, Explaining Gentzen's Consistency Proof within Infinitary Proof Theory, in G. Gottlob, A. Leitsch and D. Mundici (eds.) *Computational Logic and Proof Theory. KGC'97*, Lecture Notes in Computer Science 1289, pp. 4-17 (1997)

Bu01 W. Buchholz, Explaining the Gentzen-Takeuti reduction steps: a second order system, Arch. Math. Logic 40, pp. 255–272 (2001)

Bu02 W. Buchholz, Assigning ordinals to proofs in a perspicious way, in W. Sieg, R. Sommer and C. Talcott (eds.), *Reflections on the Foundations of Mathematics: Essays in honor of Solomon Feferman*, Lecture Notes in Logic 15, pp. 37-59 (2002)

Fef S. Feferman, The proof theory of classical and constructive inductive definitions. A 40 year saga, 1968-2008. This volume.

Ge36 G. Gentzen, Die Widerspruchsfreiheit der reinen Zahlentheorie, Math. Ann. 112 (1936), pp. 493-565

Ge43 G. Gentzen, Beweisbarkeit und Unbeweisbarkeit von Anfangsfällen der transfiniten Induktion in der reinen Zahlentheorie, Math. Ann. 119 (1943), pp. 149-161

Elementary Constructive Operational Set Theory

Andrea Cantini and Laura Crosilla*

[1] Dipartimento di Filosofia
Università degli studi di Firenze
via Bolognese, 52
50139 Firenze
Italy
Cantini@philos.unifi.it
[2] School of Mathematics
University of Leeds
LS2 9JT
UK
matmlc@leeds.ac.uk

Abstract We introduce an operational set theory in the style of [5] and [16]. The theory we develop here is a theory of *constructive* sets and operations. One motivation behind constructive operational set theory is to merge a constructive notion of set ([1], [2]) with some aspects which are typical of explicit mathematics [14]. In particular, one has non-extensional operations (or rules) alongside extensional constructive sets. Operations are in general partial and a limited form of self–application is permitted. The system we introduce here is a fully explicit, finitely axiomatised system of *constructive* sets and operations, which is shown to be as strong as $\mathbf{HA}$.

1 Introduction

This article is a follow-up of [9], where a constructive set theory with operations was introduced. Constructive operational set theory ($\mathbf{COST}$) is a constructive theory of sets and operations which has similarities with Feferman's (classical) Operational Set Theory ([16], [17], [20], [21], [22]) and Beeson's Intuitionistic set theory with rules [5]. In this article a fully explicit fragment, called $\mathbf{EST}$, of $\mathbf{COST}$ is singled out. This system is finitely axiomatized and is shown to be proof–theoretically as strong as Peano Arithmetic, $\mathbf{PA}$, (section 5).

*The research is part of a project supported by PRIN 2006 (Dimostrazioni, Operazioni, Insiemi). The second author gratefully acknowledges a grant by the John Templeton Foundation. She would also like to thank the School of Mathematics, University of Leeds, for the hospitality.

One motivation behind constructive operational set theory is to merge a constructive notion of set ([25], [1], [2]) with some aspects which are typical of explicit mathematics [14]. In particular, one has non-extensional operations (or rules) alongside extensional constructive sets. Operations are in general partial and a limited form of self–application is permitted.

The informal concept of rule plays a prominent role in constructive mathematics. Both Feferman and Beeson have repeatedly called attention to the distinction between rules and set–theoretic functions (see e.g. [15], [3]). There are several examples of intuitive rules which can not be represented by the set–theoretic concept of function. For example the operation of pair, which given two sets a and b enables us to form a new set, the set–theoretic pair of a and b. In operational set theory we have primitive operations corresponding to some set–theoretic rules, among which that of pair. In a sense, rules can be regarded as generalized algorithms or abstract rules. Without entering a detailed conceptual analysis of the notion of rule, we simply adopt the view that rules are represented by sets, and that it makes sense to *apply* a set c '*qua rule*' to another set b as input; and this possibly provides a result, whenever the algorithm encoded by c produces a computation converging to b. The application structure is specified by a ternary application relation, which satisfies very general closure conditions, in that it embodies at least pure combinatory logic with a number of primitive set–theoretic rules. As Beeson has emphasised e.g. in [5] this approach has the advantage of allowing for a natural computation system based on set theory. The idea is that while functions as graphs are hardly of any use in programming, a notion of operation can be utilised to obtain a polymorphic computation system based on set theory. Such a computation system is the main motivation for the theory of sets and rules, called **IZFR**, introduced in [5]. This is an operational version of intuitionistic Zermelo–Fraenkel set theory, **IZF** (see [3]); in particular, like that theory it is fully impredicative.

Quite different is Feferman's motivation in developing *operational set theory*. Feferman observes that analogues of 'small large cardinal notions' (those consistent with $V = L$) have emerged in different contexts, like admissible set theory, admissible recursion theory, explicit mathematics, recursive ordinal notations and constructive set and type theory. His aim in defining operational set theory is to develop a common language in which such notions can be expressed and can be interpreted both in their original classical form and in their analogue form in each of these special constructive and semi-constructive cases. Feferman's system **OST** is inherently classical, due to the presence of a choice operator (see section 3.4).

We see the present paper, though founded on [9], as a preliminary and rather experimental attempt in studying *constructive* operational set theory. It is hoped that the results here presented will contribute to both Feferman and Beeson's aims. We stress, however, our more parsimonious approach to the foundations for (constructive) mathematics: constructive operational set theory is based on intuitionistic logic and also complies with a notion of generalised predicativity.

The system **COST** of [9] had urelements at the base of the set–theoretic universe, representing the elements of an applicative structure with natural numbers. The main idea was to carefully endow the whole universe of sets with a natural extension of the base application relation. In [9] **COST** was shown to be of the same strength as **CZF** ([1], [2], [12]). Furthermore, a subtheory was singled out and shown to be of the same proof–theoretic strength as **PA**. The theory **COST** and its subsystems were introduced so to resemble as much as possible the constructive set theory **CZF** (and subsystems). In particular, **COST** had schemata of strong and subset collection, thus retaining all the mathematical expressivity of **CZF**. However, the presence of implicit principles of collection was not entirely satisfactory if one wished to have an explicit theory of sets and operations. In addition, as already noted in the introduction to [9], an inspection of the proofs in that paper (especially sections 3 - 5) shows that many of them can already be conducted in an explicit fragment of **COST**. For this reason we here single out such a fragment, **EST**, and show that it has the same strength as **PA**. Note further that in this article we work with pure sets, i.e. we do not introduce urelements.[3] One could also say that with **COST** and its subsystems we aimed at expressive theories, though of limited proof–theoretic strength. With **EST** we single out a more elegant, finitely axiomatized theory, though at the price of a more limited expressivity. We wish to note, however, that Friedman's system **B** ([18]) can be interpreted in the theory **EST** plus bounded (or limited) Dependent Choice (**LDC**) (section 4.3), so that we are persuaded we have a theory which is foundationally meaningful.

One contribution of the present paper is the use of the technique of partial cut elimination and asymmetric interpretation ([6]) to determine the strength of **EST**. We are not aware of other attempts to introduce this technique to systems of constructive set theory (see [21] for an application of this technique in the context of a proof–theoretic analysis of strong systems of classical operational set theory).

As to the contents of this paper, section 2 describes language and axioms of the theory **EST**. Section 3 collects elementary facts linking the set–theoretic and the applicative structures. In particular, we show that extensionality and totality of operations can not be assumed in general in the present context. In addition, we

[3] Urelements had a twofold motivation in [9]. On the one side, in the authors' opinion, including urelements at the ground of the set–theoretic universe appears as a constructively justified option. On the other side, urelements played a useful technical role, as they allowed for a separation between the principles of induction on the natural numbers and on sets. As a result we could define theories which had full induction on sets but bounded induction on the natural numbers. These theories had a considerable expressive power and a very limited proof–theoretic strength. However, in this paper we look for a more fundamental and simpler theory, and thus focus on a pure subsystem of **COST** with no set–induction.

study the relations between the notions of set–theoretic function and operation and also assess the status of some choice principles on the basis of $\mathbf{EST}$.

Section 4 is dedicated to clarifying the relation between $\mathbf{EST}$ and Beeson's $\mathbf{IZFR}$, Feferman's $\mathbf{OST}$ and Friedman's $\mathbf{B}$, respectively.

Finally, section 5 shows that $\mathbf{EST}$ has the same proof–theoretic strength as $\mathbf{PA}$. The lower bound is easily achieved. The upper bound is addressed by a series of steps. First an auxiliary constructive set theory, $\mathbf{ECST}^*$, is introduced. This is reduced to a classical axiomatic theory of abstract self–referential truth, $\mathbf{T_c}$, which is conservative over PA. The interpretation is obtained by an appropriate modification of [9]'s realisability interpretation. The reduction of $\mathbf{EST}$ to $\mathbf{ECST}^*$ is obtained by first introducing a Gentzen–style formulation of $\mathbf{EST}$ (in fact of a strengthening of it). A partial cut elimination theorem holds for such a system. Finally, we define an asymmetric interpretation of the operational set theory in $\mathbf{ECST}^*$, which allows us to obtain the desired upper bound.

2 The theory EST

2.1 Language and conventions

The language of $\mathbf{EST}$ is the following applicative extension, $\mathcal{L}^O$, of the usual first order language of Zermelo–Fraenkel set theory, $\mathcal{L}$.

The language includes the predicate symbols $\in$ and $=$. The logical symbols are all the intuitionistic operators: $\bot$, $\wedge$, $\vee$, $\rightarrow$, $\exists$, $\forall$. We have in addition:

- the combinators K and S;
- a ternary predicate symbol, App, for application; $App(x, y, z)$ is read as x applied to y yields z;
- el for the ground operation representing membership;
- pair, un, im, sep, for set operations;
- $\varnothing$, ω, set constants;
- IT for ω–iterator.[4]

For convenience we also use the bounded quantifiers $\exists x \in y$ and $\forall x \in y$, as abbreviations for $\exists x\,(x \in y \ \wedge \ \ldots)$ and $\forall x\,(x \in y \rightarrow \ldots)$.

[4] The idea of postulating an iteration principle as primitive is already present in Weyl's *Das Kontinuum* (chapter 1, section 7).

As customary, we define $\varphi \leftrightarrow \psi$ by $(\varphi \to \psi) \;\wedge\; (\psi \to \varphi)$ and $\neg\varphi$ by $\varphi \to \bot$. We also write $a \subseteq b$ for $\forall z\,(z \in a \to z \in b)$.

Terms and formulas. Terms and formulas are inductively defined as usual.

To increase perspicuity, we consider a definitional extension of $\mathcal{L}^O$ with application terms, defined inductively as follows.

(i) Each variable and constant is an application term.

(ii) If t, s are application terms then ts is an application term.

Application terms will be used in conjunction with the following abbreviations.

(i) $t \simeq x$ for $t = x$ when t is a variable or constant.

(ii) $ts \simeq x$ for $\exists y\,\exists z\,(t \simeq y \;\wedge\; s \simeq z \;\wedge\; App(y, z, x))$.

(iii) $t\downarrow$ for $\exists x\,(t \simeq x)$.

(iv) $t \simeq s$ for $\forall x\,(t \simeq x \leftrightarrow s \simeq x)$.

(v) $\varphi(t, \dots)$ for $\exists x\,(t \simeq x \;\wedge\; \varphi(x, \dots))$.

(vi) $t_1 t_2 \dots t_n$ for $(\dots(t_1 t_2)\dots)t_n$.

To ease readability we sometimes use the notation $t(x, y)$ for txy.

In the language $\mathcal{L}^O$, the notion of *bounded* formula needs to be appropriately modified.

Definition 2.1 (Bounded formulas)**.** A formula of $\mathcal{L}^O$ is *bounded*, or Δ_0, if and only if all quantifiers occurring in it, if any, are bounded *and in addition it does not contain application App.*

Classes are introduced as usual in set theory, as abbreviations for abstracts $\{x : \varphi(x)\}$ for any formula φ of the language $\mathcal{L}^O$. In particular, we let $\mathbf{V} := \{x : x\downarrow\}$. For A and B sets or classes, we write $f : A \to B$ for $\forall x \in A\,(fx \in B)$ and $f : \mathbf{V} \to B$ for $\forall x\,(fx \in B)$. By $f : A^2 \to B$ and $f : \mathbf{V}^2 \to B$ we indicate $\forall x \in A\,\forall y \in A\,(fxy \in B)$ and $\forall x\,\forall y\,(fxy \in B)$, respectively. This can be clearly extended to arbitrary exponents $n > 2$. Finally, for set a, $f : a \to \mathbf{V}$ means that f is everywhere defined on a.

Truth values. We may represent false and truth by the empty set and the singleton empty set, respectively; that is we let $\bot := \varnothing$ and $\top := \{\varnothing\}$.

Let Ω be the class $\mathcal{P}\top$, the powerset of $\top$. Then $x \in \Omega$ is an abbreviation for $\bot \subseteq x \subseteq \top$. The class Ω intuitively represents the class of truth values (or of propositions). Note that in the presence of exponentiation if Ω is taken to be a set then full powerset follows (see Aczel [1], Proposition 2.3).

Relations and set–theoretic functions. The notions of relation between two sets, of domain and range of a relation can be defined in the obvious way in **EST**. In the following we write $Dom(R)$ and $Ran(R)$ to denote the domain and the range of a relation, respectively. In remark 3.9 we shall see that in **EST** there is an operator **opair** internally representing the ordered pair of two sets. In addition, also the range and the domain of a relation correspond to internal operations, respectively.

We also have a standard notion of *set–theoretic function* which we can express by a formula, $Fun(F)$, stating that F is a set encoding a total binary relation which satisfies the obvious uniqueness condition. We shall use upper case letters $F, G, \ldots$ for set–theoretic functions and lower case letters $f, g, \ldots$ for operations (that is if they formally occur as operators in application terms or as first coordinates in *App*–contexts). Given a set–theoretic function F, we write $\langle x, y\rangle \in F$ or also $F(x) = y$ for $\mathbf{opair}\, xy \in F$. We shall investigate the relation between the notions of operation and set–theoretic function in section 3.3.

Finally, in defining the axiom of infinity we shall make use of the following successor operation.

Definition 2.2. Let $\mathsf{Suc} := \lambda x.\mathsf{un}\,(\mathsf{pair}\, x(\mathsf{pair}\, xx))$

2.2 Axioms of EST

Definition 2.3. **EST** is the $\mathcal{L}^O$ theory whose principles are all the axioms and rules of first order intuitionistic logic with equality, plus the following principles.

Extensionality

- $\forall x\,(x \in a \leftrightarrow x \in b) \to a = b$

General applicative axioms

- $App(x, y, z) \;\wedge\; App(x, y, w) \to z = w$
- $\mathsf{K}xy = x \;\wedge\; \mathsf{S}xy\!\downarrow \wedge\; \mathsf{S}xyz \simeq xz(yz)$

Membership operation

- $\mathsf{el} : \mathbf{V}^2 \to \Omega$ and $\mathsf{el}\, xy \simeq \top \leftrightarrow x \in y$

Set constructors

- $\forall x\,(x \notin \varnothing)$
- $\mathsf{pair}\,xy\downarrow \wedge\ \forall z\,(z \in \mathsf{pair}\,xy\ \leftrightarrow\ z = x\ \vee\ z = y)$
- $\mathsf{un}\,a\downarrow \wedge\ \forall z\,(z \in \mathsf{un}\,a\ \leftrightarrow\ \exists y \in a(z \in y))$
- $(f : a \to \Omega) \to \mathsf{sep}\,fa\downarrow \wedge\ \forall x\,(x \in \mathsf{sep}\,fa\ \leftrightarrow\ x \in a\ \wedge\ fx \simeq \top)$
- $(f : a \to V) \to \mathsf{im}\,fa\downarrow \wedge\ \forall x\,(x \in \mathsf{im}\,fa\ \leftrightarrow\ \exists y \in a(x \simeq fy))$

Strong infinity

- $(\omega 1)$ $\qquad \varnothing \in \omega\ \wedge\ \forall y \in \omega\,(\mathsf{Suc}\,y \in \omega)$
- $(\omega 2)$ $\qquad \forall x\,(\varnothing \in x\ \wedge\ \forall y(y \in x \to \mathsf{Suc}\,y \in x) \to \omega \subseteq x)$

ω–Iteration

- $$\begin{aligned} &\forall F[[Fun(F)\ \wedge\ \mathsf{dom}(F) = a\ \wedge\ Ran(F) \subseteq a] \\ &\quad \to \forall x \in a\,\exists z\,[IT(F, a, x) \simeq z\ \wedge\ Fun(z)\ \wedge\ Dom(z) = \omega \\ &\wedge\ Ran(z) \subseteq a\ \wedge\ z(\varnothing) = x\ \wedge\ \forall n \in \omega(z(\mathsf{Suc}\,n) = F(z(n)))]]. \end{aligned}$$

Remark 2.4. The principles ruling sep and im embody the explicit character of the separation and replacement schemata in the present operational context: sep provides – uniformly in any given $f : a \to \Omega$ – the set of all elements satisfying the "propositional function" defined by f; on the other hand, im yields – uniformly in any given operation f defined on a set a – the image of a under f.

Definition 2.5 (The theory $\mathbf{ESTE}$)**.** Let $\mathbf{ESTE}$ be obtained from $\mathbf{EST}$ by removing ω–iteration and by adding a new constant exp to the language together with the following explicit version of Myhill's exponentiation axiom [25]:

$$\mathsf{exp}\,ab\downarrow \wedge\ \forall x(x \in \mathsf{exp}\,ab\ \leftrightarrow\ (Fun(x)\ \ \wedge\ \ Dom(x) = a\ \ \wedge\ \ Ran(x) \subseteq b)).$$

3 Elementary properties of $\mathbf{EST}$

In this section we present some properties of $\mathbf{EST}$. In particular, we aim at clarifying the status of extensionality and intensionality in $\mathbf{EST}$. We also look at some aspects of the relationship between functions as operations and as graphs and the status of some choice principles. Finally, we show that the theory $\mathbf{ESTE}$ proves ω–iteration. Part of this section draws on [9], however adapting the arguments to the present context. For the reader's convenience we shall recall some of the arguments of [9]. First of all, as a consequence of the axioms for combinators, the universe of sets is closed under abstraction and recursion for operations (see e.g. [29]).

Lemma 3.1. *(i) For each term t, there exists a term $\lambda x.t$ with free variables those of t other than x and such that*

$$\lambda x.t \downarrow \wedge \ (\lambda x.t)y \simeq t[x := y].$$

*(ii) (Second recursion theorem) There exists a term **rec** with*

$$\textbf{\textit{rec}} f \downarrow \wedge \ (\textbf{\textit{rec}} f = e \to ex \simeq fex).$$

We now show that the logical operations generating bounded formulas are mirrored by internal operations.

Lemma 3.2. *There are application terms* eq, and, all, exists, imp, or *such that*

(i) $\mathsf{eq} : \mathbf{V}^2 \to \Omega$ *and* $\mathsf{eq}xy \simeq \top \leftrightarrow x = y$*;*

(ii) $x \in \Omega \ \wedge \ y \in \Omega \to \mathsf{and}xy \in \Omega \ \wedge \ (\mathsf{and}xy \simeq \top \leftrightarrow (x \simeq \top \ \wedge \ y \simeq \top))$*;*

(iii) $(f : a \to \Omega) \to \mathsf{all}fa \in \Omega \ \wedge \ (\mathsf{all}fa \simeq \top \leftrightarrow \forall x \in a\,(fx \simeq \top))$;

(iv) $(f : a \to \Omega) \to \mathsf{exists}fa \in \Omega \ \wedge \ (\mathsf{exists}fa \simeq \top \leftrightarrow \exists x \in a\,(fx \simeq \top))$*;*

(v) $x \in \Omega \ \wedge \ (x = \top \to y \in \Omega) \to \mathsf{imp}xy \in \Omega \ \wedge \ (\mathsf{imp}xy \simeq \top \leftrightarrow (x \simeq \top \to y \simeq \top))$;

(vi) $x \in \Omega \ \wedge \ y \in \Omega \to \mathsf{or}xy \in \Omega \ \wedge \ (\mathsf{or}xy \simeq \top \leftrightarrow (x \simeq \top \ \vee \ y \simeq \top))$.

Proof. See Lemma 3.2 of [9]. □

Proposition 3.3. *(i) For each Δ_0 formula φ with free variables contained in $\{x_1, \ldots, x_k\}$, there is an application term f_φ such that $f_\varphi \downarrow$, $f_\varphi : \mathbf{V}^k \to \Omega$ and*

$$f_\varphi\, x_1 \ldots x_k \simeq \top \leftrightarrow \varphi(x_1, \ldots, x_k).$$

(ii) To each Δ_0 formula $\varphi(x, y_1 \ldots y_k)$, we can associate an application term c_φ such that

$$\begin{aligned} &c_\varphi a y_1 \ldots y_k \downarrow \\ &\wedge \ \forall u(u \in c_\varphi a y_1 \ldots y_k \leftrightarrow u \in a \ \wedge \ \varphi(u, y_1, \ldots, y_k)). \end{aligned} \tag{3.1}$$

Proof. (i) A simple induction applies, making use of Lemma 3.2. (ii) follows from (i) and explicit separation. □

Remark 3.4.

(i) the schema (3.1) is naturally called *uniform bounded separation schema* (i.e. restricted to Δ_0-formulas, which *do not contain App*);

(ii) *uniform bounded separation* with application terms: we are allowed to use application terms as genuine terms insofar as they are defined. In the special case of separation, if t, s are application terms such that $t\downarrow$, $s\downarrow$ and $s : t \to \Omega$, then there exists an application term $r := \mathsf{sep}\, st$ such that

$$\forall u(u \in r \leftrightarrow u \in t \wedge su \simeq \top).$$

Instead of r, we write $\{u \in t : su \simeq \top\}$. Similarly, if φ is Δ_0 with free variables x, y, and t, s are application terms such that $t\downarrow$, $s\downarrow$, then there exists an application term $r_\varphi := c_\varphi ts$ such that

$$\forall u(u \in r_\varphi \leftrightarrow u \in t \wedge \varphi(u, s)).$$

Instead of r_φ, we again stick to the more familiar and perspicuous notation

$$\{u \in t : \varphi(u, s)\}.$$

The main tool in proving the results in the next subsection is the following Lemma. This is a consequence of proposition 3.3, and states that we can express an operator representing definition by cases on the universe for bounded predicates.

Lemma 3.5. *Let $\varphi(x, y)$ be Δ_0 (with the free variables shown). Then there exists an operation D_φ such that $D_\varphi uvab\downarrow$ and*

$$\varphi(u, v) \to D_\varphi uvab = a \tag{3.2}$$

$$\neg\varphi(u, v) \to D_\varphi uvab = b. \tag{3.3}$$

Proof. By uniform bounded separation (see proposition 3.3) and uniform union, there exists an operation D_φ such that

$$D_\varphi = \lambda u\lambda v\lambda a\lambda b.\{x \in a : \varphi(u, v)\} \cup \{x \in b : \neg\varphi(u, v)\}.$$

By λ-abstraction, $D_\varphi uvab\downarrow$. By extensionality, D_φ satisfies (3.2) – (3.3). □

Note that in the particular case in which a is $\top$ and b is $\bot$, even if $\varphi(u,v)$ is undecidable, then $D_\varphi uv\top\bot$ equals the proposition (the truth value) associated to $\varphi(u,v)$, i.e an element of Ω.

Indeed, as a special case we have the following.

Corollary 3.6. *There exists an operation* EQ *such that* $\mathsf{EQ}uv\downarrow$ *and*

$$u = v \to EQ(u,v) = \top$$
$$\neg u = v \to EQ(u,v) = \bot.$$

We stress again that $=$ is not decidable in general.

In the following we shall make use of the usual notation $\bigcup$ for the uniform operation of union, un, and write $\cup$ for the obvious definition of a uniform version of binary union.

3.1 Non–extensionality and partiality of operations

As observed in [9], the combination of operations and sets needs to be accomplished with care. The following argument shows that totality and extensionality can not be assumed in general.We also show that separation can not be extended to formulas with bounded quantifiers and App.

We say that two operations f and g are *extensional* if they satisfy the following:

$$\forall x\,(fx \simeq gx) \to f = g. \tag{3.4}$$

Proposition 3.7. **EST** *refutes extensionality for operations and totality of application:*

- $\neg[\forall x\,(fx \simeq gx) \to f = g]$;
- $\neg\forall x\,\forall y\,\exists z\,App(x,y,z)$.

Proof. The argument is standard. First of all, recall a (folklore) preliminary fact about *partial combinatory algebras* (pcas for short). By a pca we understand a non-empty set endowed with a partial binary function (i.e. application) and two special elements K and S satisfying the standard axioms for combinators (see definition 2.3). A pca is *extensional* if it satisfies extensionality for operations (3.4). Extensional pcas satisfy the fixed point property for total operations: if g is a total operation, then for some e, $ge = e$ (for the proof see [9] Lemma 3.11).

Now, assume extensionality, define $\varphi(u,v) \equiv (u = v)$ and let $NOTu = D_\varphi u\top\bot\top$. Then

$$u = \top \to NOTu = \bot$$
$$\neg u = \top \to NOTu = \top.$$

Note that NOT is total; hence by the previous remark, there exists a fixed point e such that $NOTe = e$ and

$$e = \top \to e = \bot \tag{3.5}$$
$$\neg e = \top \to e = \top. \tag{3.6}$$

The first implication implies $\neg e = \top$: if we assume $e = \top$, then by (3.5) $e = \bot$, which yields $\top = \bot$, i.e. $\varnothing \in \varnothing$, absurd. Hence by (3.6) we conclude $e = \top$: contradiction! On the other hand, if totality of application is assumed, the fixed point theorem of full lambda calculus holds and we can derive the inconsistency as well. □

Proposition 3.8. EST *with uniform separation for bounded conditions containing App*[5] *is inconsistent.*

Proof. By uniform separation including *App*-conditions, there would exist a total operation g such that

$$gfz = \{x \in \top : fz \simeq x\}.$$

By lemma 3.1 (second recursion theorem), there exists some e such that $gez \simeq ez$. Since g is total, e is total; hence $ee\downarrow$ and satisfies $ee = \{x \in \top : ee = x\}$. Were $x \in ee$, then $x = \varnothing \ \wedge \ x = ee$. Then $ee = \varnothing$ and hence $x \in \varnothing$: contradiction! □

3.2 $E_{\mathcal{P}}$–recursion

In [9] we noted that we can recast a form of set computability in a weak system of operational set theory. Already Beeson observed the link between his intuitionistic set theory with rules and a variant of set recursion (Beeson [5], see also [26]). In [27] Rathjen introduced a form of extended set recursion (inspired by [24]) named $E_{\mathcal{P}}$–computability. According to this form of set recursion, exponentiation is taken as one of the basic operations which are used to define set computability. Therefore, for a and b sets, the set ${}^{a}b$ of all set–theoretic functions from a to b, is computable. This notion of set recursion is used by Rathjen to develop an interpretation for **CZF** in itself which is a self validating semantics for that system of constructive

[5] This means the schema (3.1), where *App* is allowed to occur in the bounded formula φ; see also remark 3.4.

set theory. This interpretation is called the formulas–as–classes interpretation. We showed in [9] that we can naturally capture $E_{\mathcal{P}}$–computability in a subsystem of $\mathbf{COST}$. In particular, in operational set theory application is primitive and we can thus avoid the detour of [27] through coding and an inductive definition. In Proposition 4.3 of [9] we showed that the clauses defining $E_{\mathcal{P}}$-computability in Definition 4.1 of [27] can be carried out in a subsystem of $\mathbf{COST}$. Here we note that the proof of the proposition can be carried out in the theory $\mathbf{ESTE}$.[6]

For the reader's convenience we now briefly recall the content of Lemma 4.1 and that part of Proposition 4.3 of [9] which are needed in the following.

Remark 3.9.

(i) There are operations $\mathbf{int}$, $\mathbf{prod}$, $\mathbf{dom}$, $\mathbf{ran}$, $\mathbf{opair}$, $\mathbf{proj_i}$ ($i = 0, 1$), representing: binary intersection, cartesian product, domain and range of a set–theoretic function, ordered pair and projections, respectively. (See Lemma 4.1 of [9]).

(ii) There is a term $\overline{\mathbf{fa}}$ such that for any set–theoretic function F and for any $x \in Dom(F)$, $\overline{\mathbf{fa}}\, Fx \simeq F(x)$. In fact, we can take $\overline{\mathbf{fa}}$ to be : $\lambda F.\lambda x.\bigcup\{y \in Ran(F) : \langle x, y\rangle \in F\}$ (by uniform pair, union, separation). In addition, there is an operation $\overline{\mathbf{ab}}$ such that, for each f which is defined (or total) on a, $\overline{\mathbf{ab}}\, fa \simeq H$, with H a set–theoretic function with domain a and such that $\forall x \in a\,(H(x) \simeq fx)$. In fact, if $f : a \to \mathbf{V}$, then by im we can find b such that $\forall x \in a\, \exists y \in b\,(y \simeq fx)$. By (i) we have an operator $\mathbf{prod}$ which gives the cartesian product of a and b. Thus we can form $\{\langle x, y\rangle \in \mathbf{prod}\, ab : \mathsf{eq}\,(fx)y \simeq \top\}$ (see Remark 3.4) and obtain the desired operation. Note that both (i) and (ii) hold in $\mathbf{EST}$.

3.3 Operations and functions

In operational set theory we have set–theoretic functions and operations. We now wish to address the question of the relationship between them. Note that differences occur both with [9], where we had full replacement at our disposal, and with [17], where use is made of the choice operator.

According to Remark 3.9 (ii), in the theory $\mathbf{EST}$ to each set–theoretic function F there corresponds an operation which coincides with F on the common domain.

[6]Note, however, that due to the lack of set–induction, we can not prove in the present context Theorem 4.4 of [9] which showed that Rathjen's construction can be recast in $\mathbf{COST}$. Note also that the proof of the existence of dependent products in Proposition 4.3 of [9] needs exponentiation, and thus in the present context requires the theory $\mathbf{ESTE}$.

In addition, for every operation total on a set a there is a set–theoretic function representing it.

We can consistently (see section 5.3) achieve a sort of "harmony" between functions and operations by assuming Beeson's axiom $\mathbf{FO}$ (see [5]). $\mathbf{FO}$ asserts that every set–theoretic function *is* an operation, more precisely[7]:

$$(\mathbf{FO}) \quad \forall f\,(Fun(f) \to \forall x \forall y\,(\langle x, y\rangle \in f \;\leftrightarrow\; fx \simeq y)).$$

From Remark 3.9 (ii), when working in the theory $\mathbf{ESTE}$, the set $\mathsf{exp}\, ab$ contains a representative of each total operation $f : a \to b$. If we add $\mathbf{FO}$ to $\mathbf{ESTE}$ then every element of the set $\mathsf{exp}\, ab$ *is* an operation from a to b, that is

$$f \in \mathsf{exp}\, ab \to \forall x \in a \forall y \in b\,(\langle x, y\rangle \in f \leftrightarrow fx \simeq y).$$

One might now wonder if it is consistent to assume the existence of a set of *all operations* from a to b:

$$\mathsf{op} ab := \{\, f : \forall x \in a\, \exists y \in b\,(fx \simeq y)\}.$$

Pierluigi Minari has observed that if $\mathsf{op}ab$ is defined (and hence is a set), then one can reproduce the fixed point argument of Proposition 3.8.

Lemma 3.10. $\mathbf{EST} + \forall a \forall b\, \exists c(\mathsf{op}\, ab = c)$ *is inconsistent.*

The interaction between operations and functions is well exemplified in the section 3.5 on ω–Iteration in the theory $\mathbf{ESTE}$.

3.4 Choice principles

The full axiom of choice is validated in constructive type theory, where the Curry–Howard correspondence holds. However, the axiom of choice is not constructively acceptable in the context of set theory with extensionality and (bounded) separation, since it implies the (bounded) law of excluded middle by a well known argument (see [13] and [19]).

It is thus natural to ask what is the status of choice principles for operations.

In addition, as Feferman's theory $\mathbf{OST}$ is formulated with a choice operator ([16]), it is also worth exploring what is the status of such an operator on the basis of $\mathbf{EST}$.

[7] Unfortunately, in [9], section 5, the axiom $\mathbf{FO}$ appears to be stated incorrectly. However, the correct principle is used in the interpretation in Theorem 6.4.

First of all we consider two forms of choice *for operations*. Let $\mathbf{OAC}$ be the following principle:

$$\forall x \in a\, \exists y\, \varphi(x, y) \to \exists f\, \forall x \in a\, \varphi(x, fx). \tag{3.7}$$

Let $\mathbf{GAC}$ be its generalized class form:

$$\forall x\, (\varphi(x) \to \exists y\, \psi(x, y)) \to \exists f\, \forall x\, (\varphi(x) \to \psi(x, fx)). \tag{3.8}$$

Finally, let $\mathbf{GAC}\,!$ be $\mathbf{GAC}$ with the uniqueness restriction on the quantifier $\exists y$ in the antecedent of (3.8).

Lemma 3.11. *(i)* $\mathbf{EST} + \mathbf{OAC}$ *proves* $\varphi \lor \neg\varphi$ *for arbitrary bounded formulas.*

(ii) Moreover, $\mathbf{EST} + \mathbf{GAC}$ *and* $\mathbf{EST} + \mathbf{GAC}\,!$ *are inconsistent.*

Proof. (i) The standard argument, as presented for example by Goodman and Myhill [19], can be applied here, too. (ii) See Beeson [7, p. 228] or [9] Lemma 5.4. □

Let's consider Feferman's choice operator. Uniform choice is one of the principles of $\mathbf{OST}$ and is defined as follows (for a new constant $\mathcal{C}$):

$$(\mathcal{C}) \qquad \exists x\, (fx \simeq \top) \to (\mathcal{C}f \downarrow \land\ f(\mathcal{C}f) \simeq \top).$$

In [20], Theorem 6, Jäger shows that the theory $\mathbf{KP}_\omega$ + (AC) is a subsystem of $\mathbf{OST}$ (where $\mathbf{KP}_\omega$ is Kripke–Platek set theory with infinity axiom). An essential part of the proof consists in showing that $\mathbf{OST}$ proves bounded collection and that it proves the axiom of choice. The axiom of Choice is here taken in the form

$$(AC) \qquad \forall x \in a \exists y (y \in x) \to \exists F (Fun(F) \land\ Dom(F) = a\ \land\ \forall x \in a\, (F(x) \in x)).$$

It is not difficult to see that Jäger's proof that bounded Collection and (AC) hold in $\mathbf{OST}$ carries through to $\mathbf{EST}$ plus $(\mathcal{C})$.

Thus to conclude: $\mathbf{EST}$ *plus* $(\mathcal{C})$ *proves bounded Collection and* (AC). *Due to the latter fact, this theory is constructively unacceptable.*

3.5 The ω–iteration theorem in ESTE

We now show that in the theory **ESTE** we can prove the existence of an operation of ω-iteration.

First of all, note that strong infinity allows us to derive bounded induction on the natural numbers. In the following we also write 0 for $\varnothing$.

$$(\Delta_0 - IND_\omega) \qquad \varphi(0) \ \wedge \ \forall x \in \omega\,(\varphi(x) \to \varphi(\mathsf{Suc}\,x)) \to \forall x \in \omega\,(\varphi(x)),$$

where $\varphi(x)$ is Δ_0.

Lemma 3.12. *The principle ($\Delta_0 - IND_\omega$) holds in* **EST**.

Proof. This is proved by a simple application of proposition 3.3 and strong infinity. □

In the reminder of this section let F be a set–theoretic function with domain a and range $\subseteq a$, and $x \in a$. Let $Iter(H, F, a, x)$ be the *bounded* formula expressing the fact that $Fun(H)$, $Dom(H) = \omega$, $Ran(H) \subseteq a$ and H is defined by iterating F along ω with initial value x, i.e.

$$H(0) = x \ \wedge \ \forall n \in \omega(H(\mathsf{Suc}\,n) = F(H(n))).$$

By ($\Delta_0 - IND_\omega$) we easily verify the following.

Lemma 3.13. **EST** *without ω-iteration proves:*

$$Iter(H, F, a, x) \ \wedge \ Iter(G, F, a, x) \to H = G.$$

Thus the IT-operator chooses the unique such H uniformly in the data F, a, x.

Let's now consider the following bounded formula: $Iter^*(H, F, \mathsf{Suc}\,m, a, x)$ expressing the fact that $Fun(H)$ and $Dom(H) = \mathsf{Suc}\,m$ and $Ran(H) \subseteq a$ and H is defined by iterating F along $\mathsf{Suc}\,m$ with initial value x.

By bounded induction on the natural numbers we also have the following analogue of lemma 3.13.

Lemma 3.14. **EST** *without ω-iteration proves:*

$$Iter^*(H, F, \mathsf{Suc}\,m, a, x) \ \wedge \ Iter^*(G, F, \mathsf{Suc}\,j, a, x) \to (\forall n \in m \cap j)(H(n) = G(n)).$$

In addition, the following holds by uniform exponentiation exp and ($\Delta_0 - IND_\omega$).

Lemma 3.15. **ESTE** *proves:*

$$\forall m \in \omega(\exp(\mathsf{Suc}\, m)a)\downarrow \textit{(and hence is a set).}$$

We write $a^{\mathsf{Suc}\, m}$ for $\exp(\mathsf{Suc}\, m)a$.

Theorem 3.16. **ESTE** *proves ω-iteration.*

Proof. We first prove the following:

$$\forall m \in \omega\, \exists G \in a^{\mathsf{Suc}\, m}\, Iter^*(G, F, \mathsf{Suc}\, m, a, x). \tag{3.9}$$

Observe that we can apply $(\Delta_0 - IND_\omega)$ to verify (3.9) (here it is essential to have a set bound for G). The case $m = 0$ is obvious; at the successor step $m = \mathsf{Suc}\, j$, we simply expand any function G' such that $Iter^*(F, G', \mathsf{Suc}\, j, a, x)$ (which exists by IH) with the pair $\langle \mathsf{Suc}\, j, F(G'(j))\rangle$. The resulting set G satisfies $Iter^*(G, F, (\mathsf{Suc}\, m), a, x)$. For every $m \in \omega$ let

$$J(F, a, x, m) = \{G \in a^{\mathsf{Suc}\, m} : Iter^*(F, G, \mathsf{Suc}\, m, a, x)\}$$

is a set. By uniform bounded separation (see proposition 3.3), $J(F, a, x, m)$ can be regarded as an application term, as well as

$$H(F, a, x) = \bigcup\bigcup\{J(F, a, x, m) : m \in \omega\},$$

which is a set by explicit union, explicit replacement (im) and strong infinity. Now, $H(F, a, x)$ is a set (uniformly in F, a, x) and in fact a function with domain ω and range a, defined by iterating F along ω with initial value $x \in a$ (apply the uniqueness lemma above and (3.9)). Hence we can choose $IT = \lambda F \lambda a \lambda x. H(F, a, x)$. □

4 Relations with other theories

As already mentioned, the theory **EST** may be regarded as the pure and explicit fragment of **COST** ([9]). In particular, there are no urelements, no ∈–induction and no implicit principles, i.e. Strong Collection and Subset Collection.[8]

We now wish to explore the relations between **EST** and the operational theories **IZFR** of [5] and **OST** of [16]. We also clarify the relation of **EST** with Friedman's system **B** ([18]).

[8] As to the term 'implicit', we mean that strong collection and subset collection have no associated operation witnessing the sets asserted to exist, uniformly depending on the given data. For instance, if $\forall x \in a \exists y \varphi(x, y)$, by collection there exists some b, such that $\forall x \in a \exists y \in b\, \varphi(x, y, c)$; this schema is called implicit, since no operation coll_φ is assumed to exist, such that $\mathsf{coll}_\varphi(a, c)\downarrow$ and it yields d such that $\forall x \in a \exists y \in d\, \varphi(x, y, c)$.

4.1 Relation with Beeson's IZFR.

The theory **IZFR** is formulated on the basis of Beeson's logic of partial terms, LPT (see [4], [3]). We here consider a variant of **IZFR** with the application predicate App in place of LPT.

The theory has natural numbers as urelements, and is thus formulated in an extension of $\mathcal{L}^O$ with two predicates, S and N, for being a set and a natural number, respectively. In addition, there are constants 0, Suc_N, $\mathbf{d}$ for the natural number zero, successor and case distinction on the natural numbers, respectively. Finally, there are a new constant $\mathbf{P}$ for powerset and one c_φ for each primitive formula φ. A formula is *primitive* if it does not contain App or any constant c_ψ.

The theory **IZFR** is based on intuitionistic logic with equality and includes the following principles.

1. **Applicative axioms and extensionality:** as in **EST**.

2. **Basic set–theoretic axioms:** empty set, pair, union, image, all essentially as in **EST**. Note that in the presence of urelements the axiom of pair, for example, is written as follows:

$$S(\mathsf{pair}\,yz) \ \wedge\ \forall x\,(x \in \mathsf{pair}\,yz \ \leftrightarrow\ x = y \ \vee\ x = z).$$

   In addition:

   $\in$-induction axiom schema:

$$(\in -IND) \qquad \forall x\,(\forall y \in x\,\varphi(y) \rightarrow \varphi(x)) \rightarrow \forall x\,\varphi(x).$$

   The axiom of infinity, asserting the existence of a set of natural numbers as urelements.

3. **Ontological axiom and Natural numbers:** The following axiom:

$$z \in x \rightarrow S(x).$$

   In addition, principles expressing the desired properties of successor on the natural numbers and distinction by numerical cases and the schema of full induction on the natural numbers.

4. **Separation:**

$$(SEP) \qquad S(c_\varphi(a, y_1, \ldots, y_n)) \\ \wedge\ \forall x(x \in c_\varphi(a, y_1, \ldots, y_n) \ \leftrightarrow\ x \in a \ \wedge\ \varphi(x, y_1, \ldots, y_n)),$$

   where φ is primitive.

5. **Powerset**:

$$(POW) \qquad S(\mathbf{P}a) \wedge \forall x\,(x \in \mathbf{P}a \leftrightarrow S(x) \wedge \forall z \in x(z \in a)).$$

It is well–known that intuitionistic set theory with natural numbers as urelements can be interpreted in the corresponding "pure" (i.e. set only) theory. See e.g. Beeson [3], p. 166 (exercises 7 and 8). As a consequence, we can prove the following proposition.

Proposition 4.1. $\mathbf{IZFR}$ *is interpretable in* $\mathbf{EST}$+ *(SEP) + (POW) + (∈–Ind).*

Remark 4.2. The referee has asked about the converse direction of proposition 4.1. As far as we can see, there is no direct interpretation of the theory $\mathbf{EST}$+ (SEP) + (POW) + (∈–Ind) in $\mathbf{IZFR}$ because of the membership operation el and its corresponding axiom.

4.2 Relation with OST

Let $\mathbf{OST}$ be the theory defined in [16], see also [20]. Briefly, $\mathbf{OST}$ may be formulated in an extension of $\mathcal{L}^O$ with constants $\top$, $\bot$, $\mathbf{non}$, $\mathbf{dis}$, $\mathbf{all}$ and $\mathbf{C}$.[9] The theory $\mathbf{OST}$ is based on *classical logic* and includes the following principles.

1. **Applicative axioms and extensionality:** as in $\mathbf{EST}$.

2. **Basic set–theoretic axioms**: empty set, pair, union, infinity, ∈–induction (all formulated as in Zermelo–Fraenkel set theory).

3. **Logical operations axioms.** Let $\mathbb{B} := \{\top, \bot\}$ (which is a set by pair).
   (i) $\top \neq \bot$
   (ii) $(\mathsf{el} : \mathbf{V}^2 \to \mathbb{B}) \wedge \forall x \forall y\,(\mathsf{el}\,xy \simeq \top \leftrightarrow x \in y)$
   (iii) $(\mathbf{non} : \mathbb{B} \to \mathbb{B}) \wedge \forall x \in \mathbb{B}\,(\mathbf{non}(x) \simeq \top \leftrightarrow x \simeq \bot)$
   (iv) $(\mathbf{dis} : \mathbb{B}^2 \to \mathbb{B}) \wedge \forall x, y \in \mathbb{B}\,(\mathbf{dis}xy \simeq \top \leftrightarrow (x \simeq \top \vee y \simeq \top))$
   (v) $(f : a \to \mathbb{B}) \to (\mathbf{all}fa \in \mathbb{B} \wedge (\mathbf{all}fa \simeq \top \leftrightarrow \forall x \in a\,(fx \simeq \top)))$.

4. **Operational set–theoretic axioms**: uniform bounded separation and image (as in $\mathbf{EST}$, with $\mathbb{B}$ replacing Ω) and the uniform choice principle $(\mathcal{C})$ as defined in section 3.4.

[9]The constants pair, un, IT, $\varnothing, \omega$ of $\mathcal{L}^O$ are not needed for defining $\mathbf{OST}$. Note also that in [20] and subsequent papers, Jäger introduces a constant for the bounded existential quantifier, with corresponding axiom, instead of $\mathbf{all}$. In [21] Jäger investigates the proof–theoretic strength of extensions of $\mathbf{OST}$ by operators for Powerset and unbounded Existential quantifier.

Note first of all that $\in$–induction implies full induction on the natural numbers.

We now show that in the presence of the choice operator and of full induction on the natural numbers, we can derive the existence of an ω–iterator.

Lemma 4.3 ($\mathbf{OST}$). $\mathbf{OST}$ *proves ω-iteration.*

Proof. Similarly as in Theorem 3.16 we here show that for any set–theoretic function F with domain a and range $\subseteq a$, for $x \in a$

$$\forall m \in \omega \exists G\, [Iter^*(G, F, \mathsf{Suc}\, m, a, x)].$$

Note, however, that in the present case, where exponentiation is not available, the existential quantifier is unbounded. The claim is hence proved by unbounded induction on the natural numbers, which is available in $\mathbf{OST}$. We can now note that by proposition 3.3 there is a term, say t_{Iter^*}, representing the Δ_0 formula $Iter^*(G, F, \mathsf{Suc}\, m, a, x)$, that is

$$\forall m \in \omega \exists G[t_{Iter^*}(G, F, \mathsf{Suc}\, m, a, x) \simeq \top].$$

We can now apply uniform choice ($\mathcal{C}$) to obtain

$$\begin{aligned}&\forall m \in \omega\, [\mathcal{C}(\lambda y.t_{Iter^*}(y, F, \mathsf{Suc}\, m, a, x)) \downarrow\\ &\quad\wedge\; t_{Iter^*}(\mathcal{C}(\lambda y.t_{Iter^*}(y, F, \mathsf{Suc}\, m, a, x)), F, \mathsf{Suc}\, m, a, x) \simeq \top].\end{aligned}$$

Thus

$$\begin{aligned}&\forall m \in \omega\, [\mathcal{C}(\lambda y.t_{Iter^*}(y, F, \mathsf{Suc}\, m, a, x)) \downarrow\\ &\quad\wedge\; Iter^*(\mathcal{C}(\lambda y.t_{Iter^*}(y, F, \mathsf{Suc}\, m, a, x)), F, \mathsf{Suc}\, m, a, x)].\end{aligned}$$

We deduce that $\lambda m.\mathcal{C}(\lambda y.t_{Iter^*}(y, F, \mathsf{Suc}\, m, a, x)) : \omega \to \mathbf{V}$. We can now apply im and un to obtain the iterator:

$$\lambda F \lambda a \lambda x.\mathsf{un}\,(\mathsf{un}\,(\mathsf{im}\,(\lambda m.\mathcal{C}(\lambda y.t_{Iter^*}(y, F, \mathsf{Suc}\, m, a, x)))\omega)).$$

□

Let ($\mathbf{EM}$) denote the principle of Excluded Middle. Let $\mathbf{P}$ be a new constant for powerset and ($\mathbf{P}$) denote uniform powerset (that is the pure, i.e. set only, version of $\mathbf{IZFR}$'s (POW)):

$$(\mathbf{P}) \qquad \mathbf{P} : \mathbf{V} \to \mathbf{V} \;\wedge\; \forall a \forall x\, (x \in \mathbf{P}a \leftrightarrow \forall z \in x(z \in a)).$$

Proposition 4.4. *(i)* $\mathbf{EST} + (\mathcal{C}) + (\in -IND) + (\mathbf{EM}) = \mathbf{OST}$.

(ii) $\mathbf{ESTE} + (\mathcal{C}) + (\in -IND) + (\mathbf{EM}) = \mathbf{OST} + \mathbf{P}$.

Proof. (i): Note first of all that in the presence of $\mathbf{EM}$, $\Omega = \mathbb{B}$. The applicative axioms, extensionality and the operational axioms of membership, separation and image in $\mathbf{EST}$ and $\mathbf{OST}$ are thus equivalent. Showing first of all that $\mathbf{EST}$ is a subtheory of $\mathbf{OST}$, we note that one can show that in the latter theory there are terms representing operations of unordered pair and union (see [16], Corollary 2). The same corollary of Feferman shows that in $\mathbf{OST}$ we can define constants for the emptyset and for the first infinite ordinal. Thus one can easily derive $\mathbf{EST}$'s axioms of emptyset and infinity (where (ω 2) requires set–induction). Finally, by Lemma 4.3 we obtain ω–iteration.

In the opposite direction, showing that $\mathbf{OST}$ is contained in $\mathbf{EST} + (\mathcal{C}) + (\in -IND) + (\mathbf{EM})$, we note first of all that the implicit axioms of emptyset, pair and union are consequences of their explicit counterparts. Infinity follows from $(\omega 1)$. As to the logical operations axioms, we can interpret $\bot$ and $\top$ with $\varnothing$ and $\{\varnothing\}$ (i.e. $\mathsf{pair}\,\varnothing\varnothing$), respectively. Finally, by Lemma 3.2, we may let $\mathbf{non} = \lambda x.\mathsf{imp}\,x\varnothing$, $\mathbf{dis} = \mathsf{or}$ and $\mathbf{all} = \mathsf{all}$.

(ii) To see that $\mathbf{ESTE}$ is contained in $\mathbf{OST} + \mathbf{P}$, note that for sets a and b the following is a set

$$D \ := \ \{F \in \mathbf{P}(\mathbf{prod}\, ab) : Fun(F) \ \wedge \ Dom(F) = a \ \wedge \ Ran(F) \subseteq b\}.$$

By Proposition 3.3 the set D may be regarded as an application term, too, so that $\lambda a \lambda b.D$ uniformly represents exponentiation.

We now show that in the given extension of $\mathbf{ESTE}$ there is an application term representing the powerset operation. We note first of all that

$$\forall F \in \mathbb{B}^a \exists u\, (\forall x \in a(\langle x, \top\rangle \in F \ \leftrightarrow \ x \in u)).$$

Let's write $t(a, F, u)$ or simply t for the term representing the bounded formula $\forall x \in a(\langle x, \top\rangle \in F \ \leftrightarrow \ x \in u)$. We can thus apply $\mathbf{OST}$'s choice operator to obtain

$$\forall F \in \mathbb{B}^a [(\mathcal{C}\lambda y.t) \downarrow \wedge \ \forall x \in a\, (\langle x, \top\rangle \in F \ \leftrightarrow \ x \in (\mathcal{C}\lambda y.t))].$$

Thus we have an operation $\lambda F.(\mathcal{C}\lambda y.t(a, F, y)) : \mathbb{B}^a \to \mathbf{V}$. We can thus apply im to obtain $\lambda a.\mathsf{im}\,(\lambda F.(\mathcal{C}\lambda y.t(a, F, y)))\mathbb{B}^a$, which represents the powerset operation. □

4.3 Relation with Friedman's system B.

The theory $\mathbf{EST}$ has analogies with Friedman's constructive set theory $\mathbf{B}$ deprived of the principle of Δ_0–Dependent Choice (also called Limited Dependent Choice,

$\mathbf{LDC}$ in [18]. See also [3]). Let's call $\mathbf{B}^-$ the system obtained from $\mathbf{B}$ by omitting $\mathbf{LDC}$. It is easy to see that $\mathbf{B}^-$ can be interpreted in $\mathbf{EST}$.[10] Friedman's system includes a principle of abstraction which takes the place of $\mathbf{ZF}$'s replacement. This states:

$\forall x\, \exists z\, (z = \{\{u \in x : \varphi(\vec{y}, u)\} : \vec{y} \in x\})$, for $\varphi(\vec{y}, u)$ a Δ_0 formula.

Abstraction is clearly derivable in $\mathbf{EST}$ by bounded separation and image.

5 Proof theoretic reduction

In this section we show that the proof–theoretic strength of $\mathbf{EST}$ is the same as that of $\mathbf{PA}$.

Theorem 5.1 (The recursive content of $\mathbf{EST}$). *A number theoretic function f is of type $\omega \to \omega$ provably in $\mathbf{EST}$ iff f is provably recursive in $\mathbf{PA}$ (hence in $\mathbf{HA}$).*

The proof is given in two steps, the lower bound and the upper bound.

5.1 Lower bound

Theorem 5.2. $\mathbf{HA}$ *is interpretable in* $\mathbf{EST}$.

Proof. The domain of the interpretation is ω; the constant '0' is interpreted as the empty set, while the successor operation is the map $x \mapsto \mathsf{Suc}\, x$. The usual properties of 0 and successor are easily verified. Also $\mathbf{HA}$'s induction schema is given by $\Delta_0 - IND_\omega$ (Lemma 3.12). We now verify that we can define two ternary relations SUM and TIMES on ω, which exist as sets and encode the graphs of addition and multiplication on ω.

Existence of SUM

Let S be the set–theoretic function corresponding to Suc; this function exists in $\mathbf{EST}$ (by uniform union, pairing, $(\omega 1)$, explicit separation, image constructor and extensionality). Then by ω-iteration there exists an operation f such that, for $m \in \omega$,

$$fm = IT(S, \omega, m).$$

By explicit replacement there exists the set

$$H = \mathsf{im}\,(\lambda m.fm, \omega)$$

[10] The interpretation of $\mathbf{B}^-$ in $\mathbf{EST}$ can also be seen as another way of obtaining the lower bound for $\mathbf{EST}$'s proof–theoretic strength (see section 5.1).

of all set–theoretic functions defined by iterating S from m, when $m \in \omega$. Let $\omega^3 = \mathbf{prod}\,(\omega(\mathbf{prod}\,\omega\omega))$. Then by explicit bounded separation there is a set:

$$\begin{aligned}\mathsf{SUM} =\{u \in \omega^3 : (\exists F \in H)(\exists x, y, z \in \omega)[u = \langle x, y, z\rangle \\ \wedge\ Fun(F(x))\ \wedge\ Dom(F(x)) = \omega \\ \wedge\ Ran(F(x)) \subseteq \omega\ \wedge\ \langle y, z\rangle \in F(x)]\}.\end{aligned}$$

We claim that SUM is the graph of number theoretic addition.

First of all

$$\forall x \in \omega\, \forall y \in \omega\, \exists z \in \omega(\langle x, y, z\rangle \in \mathsf{SUM}).$$

Indeed, given x, $y \in \omega$, there exists a set–theoretic function $F(x) := IT(S, \omega, x)$, which is defined by ω-iteration with initial value x. Hence for every $y \in \omega$ we can find $z \in \omega$ such that $\langle y, z\rangle \in F(x)$. Then we can also verify uniqueness, for $x, y, z \in \omega$:

$$\langle x, y, z\rangle \in \mathsf{SUM}\ \wedge\ \langle x, y, w\rangle \in \mathsf{SUM} \to z = w.$$

Indeed, assume $\langle x, y, z\rangle \in \mathsf{SUM}$ and $\langle x, y, w\rangle \in \mathsf{SUM}$. Then there exist elements u_1,u_2,u_3, v_1,v_2,v_3 in ω , and $G, G' \in H$ such that

$$\begin{aligned}\langle x, y, z\rangle =&\langle u_1, u_2, u_3\rangle\ \wedge\ Fun(G(u_1))\ \wedge\ Dom(G(u_1)) = \omega \\ &\wedge\ Ran(G(u_1)) \subseteq \omega\ \wedge\ \langle u_2, u_3\rangle \in G(u_1) \\ \langle x, y, w\rangle =&\langle v_1, v_2, v_3\rangle\ \wedge\ Fun(G'(v_1))\ \wedge\ Dom(G'(v_1)) = \omega \\ &\wedge\ Ran(G'(v_1)) \subseteq \omega\ \wedge\ \langle v_2, v_3\rangle \in G'(v_1).\end{aligned}$$

By ordered pairing:

$$\begin{aligned}&Fun(G(x))\ \wedge\ Dom(G(x)) = \omega\ \wedge\ Ran(G(x)) \subseteq \omega\ \wedge\ \langle y, z\rangle \in G(x) \\ &Fun(G'(x))\ \wedge\ Dom(G'(x)) = \omega\ \wedge\ Ran(G'(x)) \subseteq \omega\ \wedge\ \langle y, w\rangle \in G'(x).\end{aligned}$$

Since G and G' are both defined by iterating S from the same initial value x, they coincide by lemma 3.13 and hence $z = w$.

Existence of TIMES.

Let F_m be the set–theoretic function:

$$\{c \in \omega^2 : (\exists u, v \in \omega)(c = \langle u, v\rangle\ \wedge\ \langle u, m, v\rangle \in \mathsf{SUM})\}$$

which exists by explicit separation. By ω-iteration, there exists an operation g such that for all $m \in \omega$:

$$gm = IT(F_m, \omega, 0).$$

Clearly $(gm)(n) = m \cdot n$. By explicit replacement there exists the set $G = \mathsf{im}\,(\lambda m.(gm))\omega$. Hence by explicit separation there exists a set:

$$\mathsf{TIMES} = \{u \in \omega^3 : (\exists H \in G)(\exists x, y, z \in \omega)(u = \langle x, y, z\rangle \\ \wedge\ Dom(H(x)) = \omega\ \wedge\ Ran(H(x)) \subseteq \omega\ \wedge\ \langle y, z\rangle \in H(x)\}.$$

Now, given x, $y \in \omega$, there exists a function $H(x) := IT(F_m, \omega, 0)$ defined by ω-iteration with initial value 0, and we can choose $\langle y, z\rangle \in H(x)$. Hence

$$(\forall x \in \omega)(\forall y \in \omega)(\exists z \in \omega)(\langle x, y, z\rangle \in \mathsf{TIMES}).$$

The verification of uniqueness, for x, y, z, w in ω

$$\langle x, y, z\rangle \in \mathsf{TIMES}\ \wedge\ \langle x, y, w\rangle \in \mathsf{TIMES} \to z = w$$

is similar to the case of addition and follows again by lemma 3.13. □

5.2 Upper bound

In this section we introduce two auxiliary theories, $\mathbf{ECST}^*$ and $\mathbf{T_c}$, and show that: (i) (a suitable extension of) $\mathbf{EST}$ can be interpreted in $\mathbf{ECST}^*$; (ii) $\mathbf{ECST}^*$ can be interpreted in $\mathbf{T_c}$ and thence has the same strength as $\mathbf{HA}$.

Elementary Constructive Set Theory

In [2] the authors introduce a subsystem of $\mathbf{CZF}$ called $\mathbf{ECST}$ (for Elementary Constructive Set Theory). They show that many standard set–theoretic constructions may be carried out already in this fragment of constructive set theory. We shall here be interested in a strengthening of $\mathbf{ECST}$ by addition of exponentiation.

The language of $\mathbf{ECST}$ is the same language as that of Zermelo–Fraenkel set theory. In this context, the notion of Δ_0 formula is the standard one, that is, a formula is Δ_0 or bounded if no unbounded quantifier occurs in it.

Definition 5.3. The theory $\mathbf{ECST}$ includes the principles of first order intuitionistic logic plus the following set–theoretic principles.

1. Extensionality;
2. Pair;
3. Union;
4. Δ_0-Separation;

5. Replacement;
6. Strong Infinity.

Here Strong Infinity is the following principle:

$$\exists a\, [Ind(a) \;\wedge\; \forall z\, (Ind(z) \to a \subseteq z)],$$

where we use the following abbreviations:

- $Empty(y)$ for $(\forall z \in y)\; \bot$,
- $Suc(x, y)$ for $\forall z\, [z \in y \;\leftrightarrow\; z \in x \;\vee\; z = x]$,
- $Ind(a)$ for $(\exists y \in a)Empty(y) \;\wedge\; (\forall x \in a)(\exists y \in a)Suc(x, y)$.

As usual, we write ω also for the set defined by strong infinity (which is unique by extensionality).

Note that $\mathbf{ECST}$ differs from $\mathbf{CZF}$ in that it only has Replacement in place of Strong Collection and it omits both Subset Collection and $\in$-Induction. Rathjen ([28]) has shown that $\mathbf{ECST}$ is very weak, as for example it does not prove the existence of the addition function on ω.

Let exponentiation be the axiom:

$$\forall a, b\, \exists c \forall F\, (F \in c \;\leftrightarrow\; (Fun(F) \;\wedge\; Dom(F) = a \;\wedge\; Ran(F) \subseteq b)),$$

where as usual Fun is a bounded formula expressing the fact that F is a set–theoretic function, $Dom(F)$ and $Ran(F)$ are the domain and range of F, respectively.

Definition 5.4. The theory $\mathbf{ECST}^*$ is obtained from $\mathbf{ECST}$ by adding the axiom of exponentiation.

To establish the upper bound we need to show that (a suitable extension of) $\mathbf{EST}$ can be interpreted in $\mathbf{ECST}^*$ and that in turn $\mathbf{ECST}^*$ can be reduced to $\mathbf{PA}$. We start from the latter problem.

Reducing $\mathbf{ECST}^*$ to $\mathbf{PA}$

We here modify the interpretation of [9] of a system of constructive set theory with urelements in a classical theory, $\mathbf{T_c}$, of abstract self–referential truth. The final result relies on the fact that $\mathbf{T_c}$ is conservative over $\mathbf{PA}$ ([7]). The main idea of the interpretation in [9] was to rephrase, in the new context, Aczel's interpretation of $\mathbf{CZF}$ in Constructive Type Theory and combine it with a suitable form of realizability.

First of all, let's recall the theory $\mathbf{T_c}$.

The theory $\mathbf{T_c}$

The basic first order language $\mathcal{L}_T$ of $\mathbf{T_c}$ comprises the predicate symbols $=$, $\mathcal{T}$, Nat , the binary function symbol ap (application), combinators K, S, successor, predecessor, definition by cases on numbers, pairing with projections. Terms are inductively generated from variables and individual constants via application. As usual $ts := ap(t, s)$; missing brackets are restored by associating to the left. Formulas are inductively generated from atoms of the form $t = s$, $\mathcal{T}(t)$, $Nat(t)$ by means of sentential operations and quantifiers. We adopt the following conventions:

(i) By $[\varphi]$ we denote a term representing the propositional function associated with φ and such that $\mathbf{FV}([\varphi]) = \mathbf{FV}(\varphi)$. We fix distinct closed *terms* $\hat{\forall}$, $\hat{\exists}$, $\hat{\neg}$ $\hat{\wedge}$, ..., naming the *logical constants*. In addition, $\hat{=}$, $\hat{N}$ name the equality and the number predicates, respectively. Then $[\varphi]$ is inductively defined by stipulating $[t = s] = (\hat{=} ts)$, $[Nat(s)] = \hat{Nat}s$, $[\mathcal{T}(s)] = s$ and closing under application of the "small hat" operations, noting that $[\forall x\varphi] = \hat{\forall}(\lambda x[\varphi])$, $[\exists x\varphi] = \hat{\exists}(\lambda x[\varphi])$.

(ii) Given a formula φ we define abstraction by letting $\{x : \varphi\} \ := \ \lambda x.[\varphi]$.

(iii) We define intensional membership, η , as follows:

$$x \,\eta\, a \ := \mathcal{T}(ax);$$
$$x \,\bar{\eta}\, a \ := \mathcal{T}(\hat{\neg}(ax)).$$

(iv) The notion of class (or classification) is so specified:

$$\mathbf{Cl}(a) \ := \ \forall x\, (x \,\eta\, a \ \vee \ x \,\bar{\eta}\, a).$$

(v) A formula φ is $\mathcal{T}$*–positive* iff φ is inductively generated from prime formulas of the form $\mathcal{T}(t)$, $t = s$, $\neg t = s$, $Nat(t)$, $\neg Nat(t)$ by means of $\vee$, $\wedge$, $\forall$, $\exists$.

(vi) A formula φ is $\mathcal{T}$*–positive operative in* v (in short, a *positive operator*) iff φ belongs to the smallest class of formulas inductively generated from prime formulas of the form $\mathcal{T}(t)$, $s \,\eta\, v$, $t = s$, $\neg t = s$, $Nat(t)$, $\neg Nat(t)$ by means of $\vee$, $\wedge$, $\forall y$, $\exists y$, where y is distinct from v and v does not occur in t, s.

(vii) For each formula φ, fixed points are defined by letting:

$$\mathbf{I}(\varphi) \ := \ \mathbf{Y}(\lambda v.\{x : \varphi(x, v)\})$$

where $\mathbf{Y}$ is Curry's fixed point combinator.

The system $\mathbf{T_c}$ comprises the following prinicples, besides *classical* predicate calculus with equality.

1. The base theory $\mathbf{TON}^-$ (see e. g. [23]), which formalises the notion of total extensional combinatory algebra expanded with natural numbers. This includes the obvious axioms on combinators, pairing, projections. In addition, closure axioms for the predicate Nat defining a copy of the natural numbers, together with number theoretic conditions on the basic operations of successor SUC, predecessor $PRED$, 0, definition by cases on the natural numbers.

2. A fixed point axiom (**Tr**) for abstract truth

$$\mathbf{Tr}(x, \mathcal{T}) \leftrightarrow \mathcal{T}(x).$$

Here $\mathbf{Tr}(x, \mathcal{T})$ is a formula encoding the closure properties:

$$\frac{a = b}{\mathcal{T}[a = b]} \qquad \frac{\neg(a = b)}{\mathcal{T}[\neg(a = b)]} \qquad \frac{Nat(a)}{\mathcal{T}[Nat(a)]} \qquad \frac{\neg Nat(a)}{\mathcal{T}[\neg Nat(a)]}$$

for the basic atomic formulas with $=$ and Nat. Further, the following additional clauses for the compound formulas:

$$\frac{\mathcal{T}(a)}{\mathcal{T}(\hat{\neg}\hat{\neg}a)} \qquad \frac{\mathcal{T}a \quad \mathcal{T}b}{\mathcal{T}(a \mathbin{\hat{\wedge}} b)} \qquad \frac{\mathcal{T}(\hat{\neg}a)\ [\ or\ \mathcal{T}\hat{\neg}b]}{\mathcal{T}(\hat{\neg}(a \mathbin{\hat{\wedge}} b))}$$

$$\frac{\forall x\, \mathcal{T}(ax)}{\mathcal{T}(\hat{\forall}a)} \qquad \frac{\exists x\, \mathcal{T}\hat{\neg}ax}{\mathcal{T}(\hat{\neg}\hat{\forall}a)}$$

3. Consistency axiom: $\neg(\mathcal{T}x \ \wedge\ \mathcal{T}\hat{\neg}x)$.

4. Induction on natural numbers Nat for *classes*:

$$\mathbf{Cl}(a) \ \wedge\ \mathbf{Clos}_{Nat}(a) \to \forall x(Nat(x) \to x\,\eta\,a)$$

with $\mathbf{Clos}_{Nat}(a) \ :=\ 0\eta a \ \wedge\ \forall x\,(x\eta a \to (SUCx)\eta a)$.

5. The principle **GID**, ensuring the minimality of the fixed points: if $\varphi(x, v)$ is a positive operator

$$\mathbf{Clos}_{\varphi}(\psi) \to \forall x\,(x\eta\mathbf{I}(\varphi) \to \psi(x))$$

with $\mathbf{Clos}_{\varphi}(\psi) \ :=\ \forall x\,(\varphi(x, \psi) \to \psi(x))$.[11]

[11] Here $\varphi(x, \psi)$ is the formula obtained by replacing each occurrence of the formula $t\,\eta\,v$ in $\varphi(x, v)$ by means of $\psi(t)$.

$\mathbf{T}^-$ is the theory $\mathbf{T_c}$ without number theoretic induction.

Let $\mathbf{CL}$ be $\{x : \mathbf{Cl}(x)\}$ (which is provably not a class). Then we can show that $\mathbf{CL}$ has natural closure conditions which are essential for the interpretation of $\mathbf{ECST}^*$. That is, $\mathbf{T}^-$ is closed under elementary comprehension, generalized disjoint union, generalized disjoint product. It satisfies a form of positive comprehension: if φ is $\mathcal{T}$–positive, then $\mathcal{T}[\varphi] \leftrightarrow \varphi$ and $\forall x\,(x\eta\{u : \varphi\} \leftrightarrow \varphi[u := x])$. Also a version of the second recursion theorem holds: if φ is positive $\forall x\,(x\eta\mathbf{I}(\varphi) \leftrightarrow \varphi(x, \mathbf{I}(\varphi)))$; for the proofs, see [8], II.9B, II.10A.

Theorem 5.5. $\mathbf{T_c}$ *is proof–theoretically equivalent to* $\mathbf{PA}$.

Proof. See [9], Theorem 7.3 or [7]. □

Reducing ECST^* to $\mathbf{T_c}$

In the following, unless otherwise stated, we work in the theory $\mathbf{T}^-$. We define a suitable counterpart of a universe $\mathcal{V}_N$ of sets, in a similar vein as in [9] (see also [10], [11]). A point of departure from [9] is however the treatment of infinity, as the subsystem of $\mathbf{COST}$ utilised there had urelements for natural numbers. For the present purpose it is instead crucial that the set of von Neumann natural numbers is interpreted in our weak theory, so to ensure that strong infinity holds under the given interpretation. For this purpose we add an initial condition to our version of Aczel's universe, adapting to our case a trick of Rathjen ([28]). In particular, in addition to the usual condition which defines sets as elements of the type of iterative sets, we also introduce a separate rule which defines the natural numbers as elements of the same type.

Let (x, y) denote the basic pairing operation which is built-in the axioms of $\mathbf{T}^-$; (x, y, z) stands for $(x, (y, z))$, and, if $u = (x, y, z)$, $u_0 = x$, $u_1 = y$ and $u_2 = z$. Let N be the class $\{x : Nat(x)\}$ and

$$N_k := \{m : m\,\eta\,N \ \wedge\ m <_N k\},$$

where $<_N$ represents the ordering relation on N. Henceforth, we simply write $<$ instead of $<_N$. Note that N_k is a class for every $k\,\eta\,N$. We also write $\sup(a, f)$ for $(1, a, f)$.

Choose by the fixed point theorem an operation ν such that

$$\nu x = sup(N_x, \nu). \tag{5.1}$$

Informally, the idea is that $sup(N_k, \nu)$ represents the von Neumann ordinal associated to the number k.

The universe of sets $\mathcal{V}_N$ is defined by means of two rules, one for initial finite segments of natural numbers and one for sets:

$$\frac{k\,\eta\,N}{sup(N_k,\nu)\,\eta\,\mathcal{V}_N}$$

and

$$\frac{\mathbf{Cl}(a) \qquad \forall u\,\eta\,a\,(fu\,\eta\,\mathcal{V}_N)}{sup(a,f)\,\eta\,\mathcal{V}_N}.$$

Lemma 5.6. *If $m\,\eta\,N$ and $k\,\eta\,N$ then $N_m = N_k \leftrightarrow m = k$.*

Proof. Obvious from right to left. Conversely, note that, if $N_m = N_k$ and $m \neq k$, we obtain a contradiction. □

Proposition 5.7. *There exists a closed term $\mathcal{V}_N$ such that*

(i)

$$\begin{aligned} a\,\eta\,\mathcal{V}_N \leftrightarrow & \exists n\,\eta\,N\,(a = \mathrm{sup}(N_n,\nu)) \\ & \vee\ (a = \mathrm{sup}(a_1,a_2)\ \wedge\ \mathbf{Cl}(a_1)\ \wedge\ \forall u\,\eta\,a_1\,((a_2u)\,\eta\,\mathcal{V}_N)); \end{aligned}$$

(ii) $\forall x(\mathcal{V}(x,\varphi) \to \varphi(x)) \to \forall x\,(x\eta\mathcal{V}_N \to \varphi(x))$,

where φ is an arbitrary formula and $\mathcal{V}(x,\varphi)$ is an abbreviation for $\exists n\,\eta\,N\,(x = \mathrm{sup}(N_n,\nu))\ \vee\ (x = \mathrm{sup}(x_1,x_2)\ \wedge\ \mathbf{Cl}(x_1)\ \wedge\ (\forall u\,\eta\,x_1)(\varphi(x_2u)))$.

Proof. See [9], Proposition 8.1. Observe that (ii) is an application of **GID**. □

Note that, as N_i is a class for each $i\,\eta\,N$, and $\nu i = \mathrm{sup}(N_i,\nu)$, we have

$$\mathrm{sup}(N_i,\nu)\,\eta\,\mathcal{V}_N \leftrightarrow \mathbf{Cl}(N_i)\ \wedge\ \forall k\,\eta\,N_i(\nu k\,\eta\,\mathcal{V}_N);$$

hence, by proposition 5.7 (i):

$$a\,\eta\,\mathcal{V}_N \leftrightarrow a = \mathrm{sup}(a_1,a_2)\ \wedge\ \mathbf{Cl}(a_1)\ \wedge\ \forall u\,\eta\,a_1\,((a_2u)\,\eta\,\mathcal{V}_N).$$

In the following, applications of proposition 5.7 (ii) will be simply referred to as *proofs by induction on $\mathcal{V}_N$*.

Proposition 5.8. *There are operations assigning $\bar{a}$ and $\widetilde{a}$ to each $a\,\eta\,\mathcal{V}_N$ and such that $\mathbf{Cl}(\bar{a})$ and $\widetilde{a} : \bar{a} \to \mathcal{V}_N$ (that is $\forall x\eta\bar{a}\,(\widetilde{a}x\eta\mathcal{V}_N)$).*

Proof. By induction on $\mathcal{V}_N$, using the recursion theorem. □

We next define recursively an equivalence relation, $\doteq$, on $\mathcal{V}_N$.
If $a \in \mathcal{V}_N$, let

$$Nat(a) := \exists k(k \,\eta\, N \ \wedge \ a = \sup(N_k, \nu)).$$

Lemma 5.9. *There exists a term* $\doteq$ *such that*

$$a \doteq b \leftrightarrow a \,\eta\, \mathcal{V}_N \ \wedge \ b \,\eta\, \mathcal{V}_N \ \wedge \ [\exists k(k \,\eta\, N \ \wedge \ N_k = \bar{a} = \bar{b} \ \wedge \ \widetilde{a} = \widetilde{b} = \nu) \ \vee$$
$$\vee \ (\neg(Nat(a) \ \wedge \ Nat(b)) \ \wedge \ \forall x \,\eta\, \bar{a} \, \exists y \,\eta\, \bar{b} \, (\widetilde{a}x \doteq \widetilde{b}y) \ \wedge \ \forall y \,\eta\, \bar{b} \, \exists x \,\eta\, \bar{a} \, (\widetilde{a}x \doteq \widetilde{b}y))].$$

Lemma 5.10. *For* $a, b, c \,\eta\, \mathcal{V}_N$ *the following holds*

1. $a \doteq a$
2. $a \doteq b \to b \doteq a$
3. $a \doteq b \ \wedge \ b \doteq c \to a \doteq c.$

Definition 5.11. Let $a, b \,\eta \mathcal{V}_N$:

$$a \dot{\in} b \ := \ \exists x \,\eta\, \bar{b} \, (a \doteq \widetilde{b}x).$$

The interpretation proceeds similarly as in [9], section 8. We here present only the most relevant steps of the interpretation.

Lemma 5.12 (Extensionality)**.** *Let* $a, b \,\eta\, \mathcal{V}_N$.

$$\forall x \,\eta\, \mathcal{V}_N \, (x \dot{\in} a \ \leftrightarrow \ x \dot{\in} b) \to a \doteq b.$$

Proof. **Case 1**: Assume $a = \sup(N_m, \nu)$, $b = \sup(N_k, \nu)$ and

$$\forall x(x \dot{\in} a \ \leftrightarrow \ x \dot{\in} b).$$

This easily implies

$$(\forall i < m)(\exists j < k)(\sup(N_i, \nu) \doteq \sup(N_j, \nu))$$
$$(\forall j < k)(\exists i < m)(\sup(N_j, \nu) \doteq \sup(N_i, \nu)).$$

By lemma 5.9

$$(\forall i < m)(i < k) \ \wedge \ (\forall i < k)(i < m),$$

which implies $m = k$, that is by definition $\sup(N_m, \nu) \doteq \sup(N_k, \nu)$.

Case 2: At least one between a, b is generated in $\mathcal{V}_N$ according to the second clause. Suppose $z \,\eta\, \bar{a}$. Then $\widetilde{a}z \,\eta\, \mathcal{V}_N$ and $\widetilde{a}z \dot{\in} a$, so that by hypothesis, also $\widetilde{a}z \dot{\in} b$. Then there exists a y such that $y \,\eta\, \bar{b}$ and $\widetilde{a}z \doteq \widetilde{b}y$. Similarly one proves the other conjunct in the definition of $a \doteq b$. □

Lemma 5.13. *For* $a, b\,\eta\,\mathcal{V}_N$,

$$\mathcal{T}[a \doteq b] \ \vee\ \mathcal{T}[\neg a \doteq b];$$
$$\mathcal{T}[a \dot{\in} b] \ \vee\ \mathcal{T}[\neg a \dot{\in} b].$$

Proof. See [9], Lemma 8.12. □

Proposition 5.14. *The structure* $\langle \mathcal{V}_N, \doteq, \dot{\in}\rangle$ *is a model of the theory* $\mathbf{ECST}^*$ *without replacement and exponentiation, provably in* $\mathbf{T_c}$.

Proof. See Proposition 8.1 of [9]. The main differences with that proposition concern extensionality, which is taken care of by Lemma 5.12, and strong infinity, which we address in the following.

Define $\hat{\omega} := sup(N, \mathbf{j})$ where, for $m\,\eta\,N$:

$$\mathbf{j}(m) = sup(N_m, \nu).$$

We need to show that:

1. $\hat{\omega}\,\eta\,\mathcal{V}_N$ and $\hat{\omega}$ is inductive (i.e. $\hat{\omega}$ contains the empty set and is closed under the set–theoretic successor, as defined within $\mathcal{V}_N$);

2. if $a\,\eta\,\mathcal{V}_N$ and a is inductive, then $\hat{\omega} \subseteq a$.

The first half of the first claim is obvious by construction. The second half requires class induction. As to the second claim, we assume that a is inductive and by class induction, using lemma 5.13, we show that

$$(\forall i\,\eta\,N)(\exists v\,\eta\,\bar{a})(\widetilde{a}v \doteq \mathbf{j}i = \mathrm{sup}(N_i, \nu)).$$

If $i = 0$, we are done by assumption on a. Let $i = SUCm$ and assume by IH that for some $v\,\eta\,\bar{a}$, $\widetilde{a}v \doteq \mathrm{sup}(N_m, \nu)$. For $c\,\eta\,\mathcal{V}_N$, let's write $(c \cup \{c\})$ also for the appropriate interpretation of the successor in $\mathcal{V}_N$ (obtained by interpreting pair and union as appropriate). Now $\widetilde{a}v \dot{\in} a$; by definition of inductive set, we also know that $(\widetilde{a}v \cup \{\widetilde{a}v\})\dot{\in} a$ and hence, for some $w\,\eta\,\bar{a}$, $\widetilde{a}w \dot{\in} a$ and $\widetilde{a}w \doteq (\widetilde{a}v \cup \{\widetilde{a}v\})$. Then also $(\mathbf{j}m \cup \{\mathbf{j}m\})\dot{\in} a$. Since we can easily verify that

$$(\mathbf{j}m \cup \{\mathbf{j}m\}) \doteq \mathbf{j}(SUCm)$$

we have the expected conclusion $\mathbf{j}(SUCm) \doteq \widetilde{a}i$. □

Finally, to give an interpretation of the theory $\mathbf{ECST}^*$ (including replacement and exponentiation) we can define a suitable notion of realisability in the theory

$\mathbf{T_c}$. First of all, if φ is a bounded formula of $\mathbf{ECST}^*$, we inductively define a map $\varphi \mapsto \|\varphi\|$, where (roughly) $\|\varphi\|$ collects the proof objects for φ, provided the parameters range over $\mathcal{V}_N$.

Let $\top$ denote the classification which only has the empty classification as element, while $a + b \ := \ \{u : u = (u_0, u_1) \ \wedge \ ((u_0 = 0 \ \wedge \ u_1 \,\eta\, a) \ \vee \ (u_0 = 1 \ \wedge \ u_1 \,\eta\, b))\}$ represents the direct sum of a, b.

Definition 5.15.

$$
\begin{aligned}
\|\bot\| &= \{e \,\eta\, \top : 0 = 1\};\\
\|a = b\| &= \{e : e = 0 \ \wedge \ \exists k (k \,\eta\, N \ \wedge \ N_k = \bar{a} = \bar{b} \ \wedge \ \widetilde{a} = \widetilde{b} = \nu)\}\\
&\quad + \{e :\, e = (e_0, e_1) \ \wedge \ \neg(Nat(a) \ \wedge \ Nat(b)) \ \wedge\\
&\quad \wedge \ \forall u \,\eta\, \bar{a}\, (e_0 u)_0 \,\eta\, \bar{b} \ \ \wedge \ (e_0 u)_1 \,\eta\, \|\widetilde{a}u = \widetilde{b}(e_0 u)_0\| \ \wedge\\
&\quad \wedge \ \forall v \,\eta\, \bar{b}\, (e_1 v)_0 \,\eta\, \bar{a} \ \ \wedge \ (e_1 v)_1 \,\eta\, \|\widetilde{a}(e_1 v)_0 = \widetilde{b}v\|\};\\
\|a \in b\| &= \{e : e = (e_0, e_1) \ \wedge \ e_0 \,\eta\, \bar{b} \ \wedge \ e_1 \,\eta\, \|a = \widetilde{b}e_0\|\};\\
\|\varphi \ \wedge \ \psi\| &= \{e : e = (e_0, e_1) \ \wedge \ e_0 \,\eta\, \|\varphi\| \ \wedge \ e_1 \,\eta\, \|\psi\|\};\\
\|\varphi \ \vee \ \psi\| &= \|\varphi\| + \|\psi\|;\\
\|\varphi \to \psi\| &= \{e : \forall q \,\eta\, \|\varphi\| (eq \,\eta\, \|\psi\|)\};\\
\|\exists x \in a\, \varphi(x)\| &= \{e : e = (e_0, e_1) \ \wedge \ e_0 \,\eta\, \bar{a} \ \ \wedge \ e_1 \,\eta\, \|\varphi(\widetilde{a}e_0)\|\};\\
\|\forall x \in a\, \varphi(x)\| &= \{e : \forall u \,\eta\, \bar{a}\, (eu \,\eta\, \|\varphi(\widetilde{a}u)\|)\}.
\end{aligned}
$$

Formally speaking, the definition of $\|\varphi\|$ above makes sense only after showing by a fixed point argument in $\mathbf{T}^-$ that there exists an operation $H(a, b)$ satisfying the equation for $\|a = b\|$ (hence the definition inductively extends H to arbitrary bounded conditions).

Definition 5.16. Let φ be an arbitrary formula of $\mathbf{ECST}^*$; we inductively define a formula $e \Vdash \varphi$ of $\mathbf{T_c}$ with the same free variables as φ and a fresh variable e:

1. if φ is a bounded formula of $\mathbf{ECST}^*$, then

$$e \Vdash \varphi \text{ iff } e \,\eta\, \|\varphi\|;$$

else:

2.

$$
\begin{aligned}
e \Vdash \varphi \to \psi &\text{ iff } \forall f (f \Vdash \varphi \to ef \Vdash \psi)\,;\\
e \Vdash \varphi \wedge \psi &\text{ iff } e = (e_0, e_1) \wedge e_0 \Vdash \varphi \wedge e_1 \Vdash \psi\,;\\
e \Vdash \varphi \vee \psi &\text{ iff } (e = (0, e_1) \wedge e_1 \Vdash \varphi) \vee (e = (1, e_1) \wedge e_1 \Vdash \psi)\,;\\
e \Vdash \forall x \in a\, \varphi(x) &\text{ iff } \forall x\, \eta\, \bar{a}\, (ex \Vdash \varphi(\widetilde{a}x))\,;\\
e \Vdash \exists x \in a\, \varphi(x) &\text{ iff } e = (e_0, e_1) \wedge e_0\, \eta\, \bar{a} \wedge e_1 \Vdash \varphi(\widetilde{a}e_0)\,;\\
e \Vdash \exists x\, \varphi &\text{ iff } e = (e_0, e_1) \wedge e_0\, \eta\, \mathcal{V}_N \wedge e_1 \Vdash \varphi(e_0)\,;\\
e \Vdash \forall x\, \varphi &\text{ iff } \forall x\, \eta\, \mathcal{V}_N\, (ex \Vdash \varphi(x))\,.
\end{aligned}
$$

Lemma 5.17. *Let φ be a bounded formula of $\mathbf{ECST}^*$. Then $\mathbf{T}^-$ proves*

$$
\begin{gathered}
\vec{x} \in \mathcal{V}_N \to Cl(\|\varphi(\vec{x})\|);\\
e \Vdash \varphi(\vec{x}) \text{ iff } e\, \eta\, \|\varphi(\vec{x})\|.
\end{gathered}
$$

Theorem 5.18. *Every theorem of $\mathbf{ECST}^*$ is realized in $\mathbf{T_c}$, i.e. if $\mathbf{ECST}^* \vdash \varphi(\vec{x})$, then there exists a closed term e such that, provably in $\mathbf{T_c}$, for $\vec{a} \in \mathcal{V}_N$*

$$e\vec{a} \Vdash \varphi(\vec{a}).$$

Proof. See Theorem 8.22 of [9]. □

5.3 Interpreting $\mathbf{\Gamma_{BEST}}$ in $\mathbf{ECST}^*$

Let $\mathbf{BEST}$ be $\mathbf{ESTE}$ + $\mathbf{FO}$. We shall prove that $\mathbf{BEST}$ is conservative over $\mathbf{ECST}^*$ for a suitable class of formulas in the common language. This is achieved through two steps. First we give a sequent style formulation of $\mathbf{BEST}$, called $\mathbf{\Gamma_{BEST}}$, so that the active formulas are positive in *App* and a partial cut elimination theorem holds. Then we give an asymmetric interpretation of $\mathbf{\Gamma_{BEST}}$ in $\mathbf{ECST}^*$, which yields the final result.

Step 1 We only give a sketch of the theory $\mathbf{\Gamma_{BEST}}$. As usual, capital Greek letters $\Gamma, \Lambda, \ldots$ denote finite sequences of formulas of $\mathbf{\Gamma_{BEST}}$. Sequents are of the form $\Gamma \Rightarrow \Lambda$. The system $\mathbf{\Gamma_{BEST}}$ is an extension of the intuitionistic Gentzen calculus ([30]). The logical rules consist of the usual rules for intuitionistic logic, including cut and $=$. In addition, there are the structural rules of weakening, exchange and contraction. In the following we first present the axioms and rules involving application; in particular, we include trivial independence conditions on constants for operations. Then we state the main rules for the set–theoretic constructors of $\mathbf{\Gamma_{BEST}}$.

In order to simplify the statements, we extend the language by adding new terms as follows:

(*) if t, s are terms, so are K_t, S_t, pair_t, im_t, sep_t, el_t, exp_t, S_{ts}.[12]

Finally, note that in the following, separation and explicit replacement are split into distinct rules to ease the asymmetric interpretation of section 5.4.

Gentzen-style presentation of non-logical axioms and rules. $\mathbf{\Gamma_{BEST}}$ includes (the closure under substitution of) the following sequents and rules:

1. Uniqueness:

$$\Gamma, ts \simeq p, ts \simeq q \Rightarrow p = q$$

2. let C be a constant among K, S, pair, im, sep, el, exp; then

$$\Gamma \Rightarrow \mathsf{C}t \simeq \mathsf{C}_t$$
$$\Gamma \Rightarrow \mathsf{S}_t s \simeq \mathsf{S}_{ts}$$

3. Combinatory completeness:

$$\Gamma \Rightarrow K_t s \simeq t$$

$$\frac{\Gamma \Rightarrow tr \simeq u \qquad \Gamma \Rightarrow sr \simeq v \qquad \Gamma \Rightarrow uv \simeq w}{\Gamma \Rightarrow \mathsf{S}_{ts} r \simeq w}$$

4. Independence:

- let $\mathsf{C}^1, \mathsf{C}^2 \in \{\mathsf{K}, \mathsf{S}, \mathsf{pair}, \mathsf{un}, \mathsf{im}, \mathsf{sep}, \mathsf{el}, \mathsf{exp}\}$; then

$$\Gamma, \mathsf{C}^1 = \mathsf{C}^2 \Rightarrow$$

- let $\mathsf{C}^1, \mathsf{C}^2 \in \{\mathsf{K}, \mathsf{S}, \mathsf{pair}, \mathsf{im}, \mathsf{sep}, \mathsf{el}, \mathsf{exp}\}$; then

$$\Gamma, \mathsf{C}^1_t = \mathsf{C}^2_s \Rightarrow t = s \ \wedge \ \mathsf{C}^1 = \mathsf{C}^2$$

[12] Formally, the special terms can be eliminated by means of a set–theoretically defined ordered pairing operation $\langle -, - \rangle$ and 8 distinct sets $c_1, \ldots, c_8$, e.g. to be identified with distinct elements of ω. For example, K_t, can be identified with $\langle c_1, t \rangle$.

- let $\mathsf{C}^1, \mathsf{C}^2 \in \{\mathsf{S}\}$; then

$$\mathsf{C}^1_{ts} = \mathsf{C}^2_{pq} \Rightarrow t = p \ \wedge \ s = q \ \wedge \ \mathsf{C}^1 = \mathsf{C}^2$$

5. Extensionality:

$$\Gamma, \forall x \, (x \in p \leftrightarrow x \in q) \Rightarrow p = q$$

6. Empty-set:

$$\Gamma \Rightarrow \forall x (x \notin \varnothing)$$

7. Representing elementhood:

$$\Gamma \Rightarrow \exists z [z \subseteq \top \ \wedge \ \mathsf{el}_a b \simeq z \ \wedge \ \forall u (u \in z \leftrightarrow u = \bot \ \wedge \ a \in b)]$$

8. Union:

$$\Gamma \Rightarrow \exists z [\mathsf{un} a \simeq z \ \wedge \ \forall u (u \in z \leftrightarrow \exists y \in a \, (u \in y))]$$

9. Pairing:

$$\Gamma \Rightarrow \exists z [\mathsf{pair}_a b \simeq z \ \wedge \ \forall u (u \in z \leftrightarrow u \in a \ \vee \ u \in b)]$$

10. Strong infinity:

$$\begin{aligned} &\Gamma \Rightarrow \varnothing \in \omega \\ &\Gamma, t \in \omega \Rightarrow \mathrm{S}\, t \in \omega \\ &\Gamma, \varnothing \in t \ \wedge \ \forall y (y \in t \to \mathsf{Suc}\, y \in t) \Rightarrow \omega \subseteq t \end{aligned}$$

11. Separation:

$$\frac{\Gamma \Rightarrow (\forall u \in a)(\exists y \subseteq \top)(fu \simeq y)}{\Gamma \Rightarrow \exists z [(\forall u \in z)(fu \simeq \top \ \wedge \ u \in a) \ \wedge \ (\forall u \in a)(\forall y (fu \simeq y \to y = \top) \to u \in z}$$

From the premisses

- $\Gamma \Rightarrow (\forall u \in a)(\exists y \subseteq \top)(fu \simeq y)$
- $\Gamma \Rightarrow (\forall u \in z)(fu \simeq \top \ \wedge \ u \in a)$
- $\Gamma \Rightarrow (\forall u \in a)(\forall y (fu \simeq y \to y = \top) \to u \in z)$

infer:

$$\Gamma \Rightarrow \mathsf{sep}_a f \simeq z$$

12. Explicit replacement:

$$\frac{\Gamma \Rightarrow (\forall x \in a)\exists y(fx \simeq y)}{\Gamma \Rightarrow \exists z[(\forall y \in z)(\exists x \in a)(fx \simeq y) \ \wedge \ (\forall x \in a)(\exists y \in z)(fx \simeq y)]}$$

From the premisses

- $\Gamma \Rightarrow (\forall u \in a)\exists y(fu \simeq y)$
- $\Gamma \Rightarrow (\forall y \in z)(\exists x \in a)(fx \simeq y)$
- $\Gamma \Rightarrow (\forall x \in a)(\exists y \in z)(fx \simeq y)$

infer:

$$\Gamma \Rightarrow \mathsf{im}_a f \simeq z$$

13. Exponentiation:

$$\Gamma \Rightarrow \exists z[\mathsf{exp}_a b \simeq z \ \wedge \ \forall F(F \in z \leftrightarrow (Fun(F) \ \wedge \ Dom(F) = a \ \wedge \ Ran(F) \subseteq b))]$$

14. Beeson's axiom **FO**: every function is an operation, i.e.

$$\Gamma, Fun(F), \langle x, y\rangle \in F \Rightarrow Fx \simeq y$$
$$\Gamma, Fun(F), Fx \simeq y \Rightarrow \langle x, y\rangle \in F.$$

We stress that *the active formulas of the inferences and axioms are positive in App.*

Theorem 5.19 (Quasi-normal form)**.** *A* $\mathbf{\Gamma_{BEST}}$*-derivation* $\mathcal{D}$ *can be effectively transformed into a* $\mathbf{\Gamma_{BEST}}$*-derivation* $\mathcal{D}^*$ *of the same sequent, such that every cut formula occurring in* $\mathcal{D}^*$ *is positive in* $\simeq$*.*

5.4 Step 2. The asymmetric interpretation

We now define an asymmetric interpretation of $\mathbf{\Gamma_{BEST}}$ into $\mathbf{ECST}^*$: the idea is to replace *App* by its finite stages App^n which, for each given *n*, can be explicitly defined and proved to exist in the pure set–theoretic language of $\mathbf{ECST}^*$. Thus the finite approximations of the rules can be justified in the *App*-free system $\mathbf{ECST}^*$. However, the interpretation is asymmetric in the sense that it depends on a pair of number parameters $m \leq n$; in particular the positive occurrences of *App* are separated from the negative ones (the former being replaced by App^n and the second by App^m).

Let $\mathcal{A}(x, y, z, P)$ be the *App-positive formula*, inductively generating the application predicate. The formula belongs to the language of $\mathbf{ECST}^*$, except (i) for the ternary predicate symbol P and (ii) for the terms of the form C_t, S_{ts} (C being a constant among K, S, im, sep, el, exp, pair). Since these special terms can be readily eliminated (in the sense that we can define a translation thereof in the pure set–theoretic language), we can assume that $\mathcal{A}(x, y, z, P)$ belongs to the language of $\mathbf{ECST}^*$, expanded with P.

Definition 5.20. Let $\bot$ also be an abbreviation for $\mathsf{K} = \mathsf{S}$ and define inductively:

$$\begin{aligned} App^0(x, y, z) &:= \bot \\ App^{k+1}(x, y, z) &:= \mathcal{A}(x, y, z, App^k). \end{aligned}$$

Here above $\mathcal{A}(x, y, z, App^k)$ is obtained from $\mathcal{A}(x, y, z, P)$ by replacing P everywhere with App^k.

Definition 5.21.

(i) We inductively define $A[m, n]$, where A is a formula of $\mathbf{\Gamma_{BEST}}$: uniformly in n, m.

$$\begin{aligned} A[m, n] &:= A \text{ provided } A \text{ has the form } t = s \text{ or } t \in s \\ App(t, s, r)[m, n] &:= App^n(t, s, r) \\ (A \to B)[m, n] &:= (A[n, m] \to B[m, n]); \end{aligned}$$

moreover $A \mapsto A[m, n]$ commutes with $\wedge$, $\vee$, $\forall$, $\exists$.

(ii) If $\Gamma := \{A_1, \ldots, A_p\}$, $\Gamma[m, n] := \{A_1[m, n], \ldots, A_p[m, n]\}$;

(iii) $(\Gamma \Rightarrow \Delta)[m, n] := \Gamma[n, m] \Rightarrow \Delta[m, n]$.

Lemma 5.22.

(i) For each $k \in \omega$, App^k is a formula of $\mathbf{ECST}^*$.

(ii) In addition we have, provably in $\mathbf{ECST}^*$,

$$k \leq m \Rightarrow App^k(x, y, z) \rightarrow App^m(x, y, z);$$

(iii) if A is App-positive (negative), then $A[m, n] := A^n$ ($A[m, n] := A^m$); if A is App-free, $A[m, n] := A$.

Lemma 5.23 (Persistence). *Let $m \leq p \leq q \leq n$. Then provably in* $\mathbf{ECST}^*$*:*

$$A[p, q] \rightarrow A[m, n];$$
$$A[n, m] \rightarrow A[q, p].$$

Below we also use the more suggestive notation $xy \simeq^m z$ instead of $App^m(x, y, z)$.

Lemma 5.24 (Uniqueness). *Provably in* $\mathbf{ECST}^*$*: If $Fun(F)$, $Dom(F) = a$, $Ran(F) \subseteq a$ and $x \in a$ then*

$$Iter(z, F, a, x) \ \wedge \ Iter(y, F, a, x) \rightarrow z = y. \tag{5.2}$$

Furthermore, for each given $m \in \omega$:

$$xy \simeq^m z \ \wedge \ xy \simeq^m w \ \rightarrow \ z = w. \tag{5.3}$$

Proof. As to (5.2), this is analogous to Lemma 3.13.
As to (5.3), we argue informally by outer induction on $m \in \omega$. If $m = 0$, the conclusion is trivial. As to the verification of the induction step $m = j + 1$, we first apply the independence axioms. This immediately yields uniqueness in all trivial cases where x is among un, pair, exp, K, S.

Assume $xy \simeq^{j+1} z$, $xy \simeq^{j+1} w$, i.e. $\mathcal{A}(x, y, z, App^j)$ and $\mathcal{A}(x, y, w, App^j)$. Then, for some a, b, c, d, we obtain $x = S_{ab}$ and $x = S_{cd}$. By independence, $a = c$, $b = d$ and hence $S_{ab}y \simeq^{j+1} z$, $S_{ab}y \simeq^{j+1} w$, which imply, for some p, q, r, s:

- $ay \simeq^j p$, $by \simeq^j q$, $pq \simeq^j z$
- $ay \simeq^j r$, $by \simeq^j s$, $rs \simeq^j w$.

By IH $p = r$, $q = s$ and hence $pq \simeq^j z$, $pq \simeq^j w$, which yields $z = w$ again by IH.

Consider the case where $\mathsf{im}_a f \simeq^{j+1} z$, $\mathsf{im}_a f \simeq^{j+1} w$ (we implicitly use independence conditions on terms of the form im_a). Then we have

- $(\forall u \in z)(\exists x \in a)(fx \simeq^j u) \ \wedge \ (\forall x \in a)(\exists u \in z)(fx \simeq^j u)$;
- $(\forall u \in w)(\exists x \in a)(fx \simeq^j u) \ \wedge \ (\forall x \in a)(\exists u \in w)(fx \simeq^j u)$.

We prove $z \subseteq w$. Let $u \in z$: then by the first condition above $fx \simeq^j u$, for some $x \in a$. Then by the second condition, $fx \simeq^j v$, for some $v \in w$. By IH $u = v$ and hence $u \in w$. We also easily verify that $w \subseteq z$ and hence $w = z$ by extensionality. □

Theorem 5.25. *Let $\mathcal{D}$ be a $\mathbf{\Gamma_{BEST}}$-derivation of $\Gamma \Rightarrow \Delta$. Then there exists a natural number $c \equiv c_{\mathcal{D}}$ such that, for every $m > 0$ and every n such that $n \geq c+m$,*

$$(\Gamma \Rightarrow \Delta)[m, n]$$

is derivable in $\mathbf{ECST}^*$.

Proof. By the preparation lemma we can assume that the given derivation of $\Gamma \Rightarrow \Delta$ is quasi-normal, i.e. cuts occur only on *App*-positive formulas. Furthermore, by the previous lemma 5.23 it is enough to check, for some constant c depending on the given quasi-normal derivation,

$$(\Gamma \Rightarrow \Delta)[m, c+m]. \tag{5.4}$$

Cut Assume that our derivation $\mathcal{D}$ ends with a cut on an *App*-positive formula C and that the immediate subderivations of $\mathcal{D}$ end with $\Gamma \Rightarrow C$ and $C, \Gamma \Rightarrow A$. By IH we have, for some c_0, c_1, for each $m > 0$:

$$\Gamma[c_0 + m, m] \Rightarrow C^{c_0+m}$$
$$C^m, \Gamma[c_1 + m, m] \Rightarrow A[m, c_1 + m].$$

Choose $m := c_0 + m$ in the second sequent. Then, for $c = c_0 + c_1$, we obtain:

$$C^{c_0+m}, \Gamma[c + m, c_0 + m] \Rightarrow A[c_0 + m, c + m].$$

Hence with a cut

$$\Gamma[c + m, c_0 + m], \Gamma[c_0 + m, m] \Rightarrow A[c_0 + m, c + m].$$

But $m \leq c_0 + m \leq c + m$ and hence by persistence:

$$\Gamma[c + m, m], \Gamma[c + m, m] \Rightarrow A[m, c + m].$$

The conclusion follows by contraction.

Explicit replacement By IH, for some c_0, for every $m > 0$, we have:

$$\cdot \quad \Gamma[c_0 + m, m] \Rightarrow (\forall x \in a)(\exists y)(fx \simeq^{c_0+m} y)$$

As y is unique, by replacement, there exists a function F (hence a set), depending on $c_0 + m$, such that

$$(\forall x \in a)(fx \simeq^{c_0+m} F(x)).$$

Hence we can choose a set $z = \{F(x) \mid x \in a\}$, depending on $c_0 + m$; z satisfies the asymmetric translation of the conclusion choosing $c := c_0$, i.e. we can derive in $\mathbf{ECST}^*$ the sequent whose antecedent is $\Gamma[c+m, m]$ and whose succedent is

$$(\forall y \in z)(\exists x \in a)(y \simeq^{c+m} fx) \ \wedge \ (\forall x \in a)(\exists y \in z)(fx \simeq^{c+m} y).$$

On the other hand, by IH we have

- $\Gamma[c_0 + m, m] \Rightarrow (\forall u \in a)(\exists y)(fu \simeq^{c_0+m} y)$
- $\Gamma[c_0 + m, m] \Rightarrow (\forall y \in z)(\exists x \in a)(fx \simeq^{c_0+m} y)$
- $\Gamma[c_0 + m, m] \Rightarrow (\forall x \in a)(\exists y \in z)(fx \simeq^{c_0+m} y)$.[13]

Hence by definition of the operator defining $\simeq$ we have, for $c = c_0 + 1$:

$$\Gamma[c + m, m] \Rightarrow \mathsf{im}_a f \simeq^{c+m} z.$$

Separation By IH, for some c_0, for every $m > 0$, we have:

$$\Gamma[c_0 + m, m] \Rightarrow (\forall x \in a)(\exists y \subseteq \top)(fx \simeq^{c_0+m} y).$$

By replacement, there exists a function F (hence a set), depending on $c_0 + m$, such that

$$(\forall x \in a)(F(x) \subseteq \top \ \wedge \ fx \simeq^{c_0+m} F(x)).$$

Hence

$$z = \{x \in a \mid \langle x, \top \rangle \in F\}$$

[13] Strictly speaking, each premiss will be assigned its own bounding constant c_i, where $i = 1, 2, 3$, but by persistence we can replace it by $c_0 = max\{c_1, c_2, c_3\}$.

is a set by bounded separation and it satisfies the asymmetric interpretation of the conclusion choosing $c = c_0$. As in the previous case, we can derive by definition of the operator defining $\simeq$, for $c = c_0 + 1$:

$$\Gamma[c+m,m] \Rightarrow \mathsf{sep}_a f \simeq^{c+m} z$$

provided z satisfies the asymmetric interpretation of the premisses of the second separation rule.

Exp, Union, Pairing, Elementhood by the appropriate corresponding axioms choosing $c = 0$.

□

Corollary 5.26. *Every* $\mathbf{\Gamma_{BEST}}$*-derivation of an App-free condition can be effectively transformed into a derivation in* $\mathbf{ECST}^*$.

References

[1] P. ACZEL, *The Type Theoretic Interpretation of Constructive Set Theory*, in: A. MacIntyre, L. Pacholski, J. Paris (eds.), ***Logic Colloquium '77*** (North–Holland, Amsterdam-New York, 1978).

[2] P. ACZEL, M. RATHJEN, *Notes on Constructive Set Theory*, Draft available at the address: `http://www.mittag-leffler.se/preprints/meta/AczelMon\_Sep\_24\_09\_16\_56.rdf.html`.

[3] M. BEESON, ***Foundations of Constructive Mathematics***, (Springer Verlag, Berlin, 1985).

[4] M. BEESON, *Proving programs and programming proofs*, in: R. BARCAN MARCUS ET AL., EDS., ***Logic, Methodology and Philosophy of Science VII***, Proceedings of the meeting in Salzburg, Austria, July 1983, (North–Holland, Amsterdam, 1986), 51-82.

[5] M. BEESON, *Towards a computation system based on set theory*, ***Theoretical Computer Science*** 60 (1988) pp. 297–340.

[6] A. CANTINI, *On the Relation Between Choice and Comprehension Principles in Second Order Arithmetic*, Journal of Symbolic Logic, 51 (1986) pp. 360–373.

[7] A. CANTINI, *Levels of implication and type free theories of partial classifications with approximation operator*, ***Zeitschrift für mathematische Logik und Grundlagen der Mathematik*** 38 (1992) pp. 107–141.

[8] A. CANTINI, ***Logical Frameworks for Truth and Abstraction***, (North Holland, Amsterdam, 1996).

[9] A. CANTINI, L. CROSILLA, *Constructive set theory with operations*, in A. Andretta, K. Kearnes, D. Zambella eds., *Logic Colloquium 2004*, Association of Symbolic Logic, Lecture notes in Logic, 29, 2008.

[10] L. CROSILLA, *Realizability Models for Constructive Set Theories with Restricted Induction Principles*, University of Leeds, Ph. D. Thesis, Department of Pure Mathematics, September 2000.

[11] L. CROSILLA, M. RATHJEN, *Inaccessible set axioms may have little consistency strength*, ***Annals of Pure and Applied Logic*** 115/1-3 (2001) pp. 33–70.

[12] L. CROSILLA, *Constructive and intuitionistic ZF*, in: Stanford Encyclopedia of Philosophy, February 2009, available at the address: http://plato.stanford.edu/entries/set-theory-constructive/ .

[13] R. DIACONESCU, *Axiom of choice and complementation* ***Proc. Amer. Math. Soc.*** 51 (1975) pp. 176–178.

[14] S. FEFERMAN, *A language and axioms for explicit mathematics* in: J. Crossley (ed.), *Algebra and Logic*, ***Lecture Notes in Mathematics***, vol 450, (Springer, Berlin 1975) pp. 87–139.

[15] S. FEFERMAN, *Constructive theories of functions and classes*, in M. Boffa, D. van Dalen, K. McAloon (eds.) ***Logic Colloquium '78***, (North Holland, Amsterdam, 1979) pp. 159–224.

[16] S. FEFERMAN, *Notes on Operational Set Theory I. Generalization of "small" large cardinals in classical and admissible set theory*, unpublished, Stanford University (2001) pp. 1–10.

[17] S. FEFERMAN, *Operational Set Theory and small large cardinals*, Information and Computation, Vol 207, issue 10, 2009, pp. 971-979 .

[18] H. FRIEDMAN, *Set-theoretic foundations for constructive analysis*, ***Annals of Mathematics*** 105 (1977) pp. 1–28.

[19] N.D. Goodman, J. Myhill: *Choice implies excluded middle*. Z. Math. Logik Grundlag. Math. 24 (1978) p. 461.

[20] G. JÄGER, *On Feferman's operational set theory OST*, ***Annals of Pure and Applied Logic***, 150 (2007) pp. 19–39 .

[21] G. JÄGER, *Full operational set theory with unbounded existential quantification and powerset*, ***Annals of Pure and Applied Logic***, 160(1), pp. 33-52.

[22] G. JÄGER, *Operations, sets and classes*, Logic, Methodology and Philosophy of Science: Proceedings of the thirteen International Congress, ed. by C. Glymour, W. Wei, E. Westerstahl, College Publications (2009)

[23] G. JÄGER, T. STRAHM, *Totality in applicative theories*, ***Annals of Pure and Applied Logic***, vol. 74 (1995) pp. 105–120.

[24] L. S. MOSS, *Power set Recursion*, ***Annals of Pure and Applied Logic*** 71 (1995) pp. 247–306.

[25] J. MYHILL, *Constructive Set Theory*, ***The Journal of Symbolic Logic*** 40 (1975) pp. 347–382.

[26] D. NORMANN, *Set Recursion*, ***Generalized Recursion Theory II*** (North Holland, Amsterdam, 1978) pp. 303–320.

[27] M. RATHJEN, *The formulae as classes interpretation of constructive set theory*, ***Proof technology and computation***, 279–322 (NATO Sci. Ser. III Comput. Sys. Sci., 2000, Amsterdam, 2006).

[28] M. RATHJEN, *The natural numbers in constructive set theory*, ***Mathematical Logic Quarterly*** 54 (2008) n.1, 83–97.

[29] A. S. TROELSTRA AND D. VAN DALEN, ***Constructivism in Mathematics: an Introduction***, volumes I and II (North–Holland, Amsterdam, 1988).

[30] A.S.TROELSTRA, H.SCHWICHTENBERG, *Basic Proof Theory*, Cambridge University Press, Cambridge 2000 (2nd ed.).

[31] H. WEYL, ***Das Kontinuum*** (Leipzig, 1918).

Functional Interpretations of Classical Systems

Justus Diller

Münster, Germany

Abstract In contrast to Gödel's Dialectica interpretation, the Diller-Nahm interpretation extends to systems of arithmetic in all finite types as well as to systems of set theory. We present a unified treatment of functional interpretations of Peano arthmetic and Kripke-Platek set theory, both of the standard classical theories as well as of their versions in all finite types. We also give axiomatic characterizations of the functional translations in question by weak axioms of choice.

1 Functional translations of classical systems, common features

Gödel's 1958 Dialectica interpretation D of Heyting arithmetic HA in his quantifier-free theory T of primitive recursive functionals of finite types [11] does not extend to Heyting arithmetic in all finite types HA^ω, as Howard's example shows. For the same reason, D as well as Shoenfield's interpretation S of Peano arithmetic PA [16] do not extend to Peano arithmetic in all finite types PA^ω. Also Kripke-Platek set theory (with axiom of infinity) $KP\omega$ cannot be D- or S-interpreted by constructive functionals, but only by use of a non-constructive choice functional (cf. [6]). Burr [4] gives a functional interpretation of $KP\omega$ by a hybrid $\vee$ of the $\wedge$- (cf. [9]) and the S-interpretation. Concerning the background of functional interpretations, see [1], [5], [7], and [18].

We present a unified approach by giving a $\wedge$-interpretation of PA^ω as well as of $KP\omega$ and its finite type extension $KP\omega^\omega$. For this purpose, it is adequate to work in the *negative fragment*, i.e. in the $\{\exists, \vee\}$free fragment of first order logic. In this fragment, stability is the one logical principle that extends intuitionistic to classical logic. It is an elementary theorem of intuitionistic logic (cf [19]) that in intuitionistic theories iTh stability of arbitrary *negative*, i.e. $\{\exists, \vee\}$free formulae is derivable from the stability of their atomic formulae:

1.1 Stability lemma. *For all negative formulae A in $L(iTh)$*

$$iTh + \{\neg\neg P \to P \mid P \text{ atomic } \} \vdash \neg\neg A \to A$$

The intuitionistic functional theories that we refer to are Gödel's theory T (cf. [11], [10], [17], [14]) and its extension $T_\wedge$ by a bounded universal quantifier $\forall x < t$ (cf.[9]) on the one hand and Burr's theory $T_\in$ of constructive set functionals (cf. [3], [8]) on the other.
These theories do not contain unbounded quantifiers. We choose formulations of T (as in [10] and [14]) and $T_\wedge$ in the negative fragment. The language of $T_\in$ (cf. [3]) is the closure of its Δ_0-language - which in turn is the closure of the type o equations of $L(T_\in)$ and $\perp$ under $\wedge, \vee, \rightarrow, \forall x \in t$ and $\exists x \in t$ - and its equations of higher type under $\wedge, \rightarrow$, and $\forall x \in t$.

1.2 Definition of classical functional theories $T^c, T^c_\wedge, T^c_\in$. Let

$$Stab(=) \equiv \{\neg\neg a = b \rightarrow a = b \mid a, b \text{ terms of the same type }\}$$

For FT one of the theories $T, T_\wedge$ or $T_\in$, the **classical version** FT^c of FT is

$$FT^c :\equiv FT + Stab(=)$$

1.3 Lemma. *For FT any of $T, T_\wedge, T_\in, FT^c$ satisfies stability in general:*

$$FT^c \vdash \neg\neg A \rightarrow A \quad \textit{for all formulae} \quad A \in L(FT^c)$$

Proof. For T and $T_\wedge$, the lemma is an immediate consequence of the stability lemma 1.1, because the atomic formulae of $L(T)$ and $L(T_\wedge)$ are, besides $\perp$, only equations. In $T_\in$, any Δ_0-formula A is equivalent to an equation $\{0 \mid A\} = 1$ by explicit Δ_0-separation (cf. [3],[8], and proposition 3.1 below). Therefore, any formula of $L(T_\in)$ is in $T_\in$ equivalent to a negative formula, and by the stability lemma, the lemma follows.
Since the type o fragments T_0 of T and $T_{\wedge 0}$ of $T_\wedge$ prove $Stab(=)$ within their respective language, T_0 and $T_{\wedge 0}$ are themselves already classical functional theories; we have $T^c_0 \equiv T_0$ and $T^c_{\wedge 0} \equiv T_{\wedge 0}$. This does not hold for $T_{\in 0}$.

1.4 Theories of classical arithmetic and set theory. We identify Peano arithmetic PA with the negative fragment of Heyting arithmetic HA. Similarly, Peano arithmetic in all finite types PA^ω is the *natural span* of PA and $T^c_\wedge$. It is defined as the negative fragment of HA^ω, extended by the schema $Stab(=)$, and with the bounded universal quantifier $\forall x < t$ restricted to $L(T_\wedge)$. Its type o fragment PA^ω_0, like $T_{\wedge 0}$, does not need $Stab(=)$ as an axiom.
Let $KP\omega$ (cf. [2]) be formulated in the negative fragment of the language of set theory. Its finite type version $KP\omega^\omega$ is the *natural span* of $KP\omega$ and $T^c_\in$, also formulated in the negative fragment: its language is the closure of the negative fragment of $L(T_\in)$ under $\wedge, \rightarrow$ and $\forall x^\tau$ for all types τ; its axioms and rules are the axioms of $KP\omega$ and of $T^c_\in$, the rule of transfinite induction extended to the full

language, and the rule of type-extensionality $T\text{-}EXT$ with side formulae restricted to $L(T_\in)$ (*weak extensionality*). (Because of the rule $T\text{-}EXT$, $KP\omega^\omega$ violates the deduction theorem.)

These classical theories are to be $\wedge$-interpreted and - if possible - Dialectica interpreted in the functional theories mentioned in 1.2. We attempt a simultaneous definition of functional translation and interpretation for the arithmetical as well as for the set-theoretic case.

1.5 Convention. For the remaining part of this section, let Th stand alternatively for PA^ω or $KP\omega^\omega$ or their subtheories, FT for a functional theory, $T_{\wedge\in}$ for $T_\wedge$ or $T_\in$, $\forall x <\in t$ for $\forall x < t$ or $\forall x \in t$, and I for D or $\wedge$.

1.6 Recursive definition of the I-translation on Th. To any formula $A \in L(Th)$, I assigns an expression $A^I \equiv \exists v \forall w A_I[v, w]$ with $A_I[v, w]$ a formula of $L(FT)$ and disjoint tuples of variables v, w not occurring free in A, as follows:

$$
\begin{array}{lll}
L(T)^D & A^D \equiv A \text{ for } A \in L(T) & \\
L(T_{\wedge\in})^I & A^I \equiv A \text{ for } A \in L(T_{\wedge\in}) & \text{otherwise}
\end{array}
$$

Let A^I be as above and $B^I \equiv \exists y \forall z B_I[y, z]$; then

$$
\begin{array}{lll}
(\wedge)^I & (A \wedge B)^I & \equiv \exists vy \forall wz (A_I[v, w] \wedge B_I[y, z]) \\
(\to)^D & (A \to B)^D & \equiv \exists WY \forall vz (A_D[v, Wvz] \to B_D[Yv, z]) \\
(\to)^\wedge & (A \to B)^\wedge & \equiv \exists XWY \forall vz (\forall x <\in Xvz\ A_\wedge[v, Wxvz] \to B_\wedge[Yv, z]) \\
 & & \quad \text{in case the tuple } w \text{ is not empty} \\
(\to)^\wedge_0 & (A \to B)^\wedge & \equiv \exists Y \forall vz (A_\wedge[v] \to B_\wedge[Yv, z]) \text{ for empty } w \\
(\forall)^I & (\forall u A[u])^I & \equiv \exists V \forall uw A_I[u, Vu, w]
\end{array}
$$

So, for D, the definition for the set theoretic case is identical with the one for the arithmetical case, with the (unavoidable) exception of the starting clause. For $\wedge$, the set theoretic case is generated from the arithmetical case by simply writing $\forall x \in t$ for $\forall x < t$ and thus substituting $T_\in$ for $T_\wedge$. If, for $\wedge$, closure of $L(Th)$ under $\forall x <\in t$ is preferred - which is not necessary, but occasionally useful - , a clause
$(\forall <\in)^\wedge \quad (\forall x <\in t\ A[x])^\wedge \equiv \exists V \forall w (\forall x <\in t\ A_\wedge[x, Vx, w])$
has to be added to the definition. - The Dialectica translation D is obtained from the $\wedge$-translation by writing D for $\wedge$ and by cancelling, in $(\to)^\wedge$, the bounded universal quantifier $\forall x <\in Xvz$ and the variables X and x.

1.7 Definition. The functional translation I is a **functional interpretation** of a theory Th in a functional theory FT, we write

$$Th \overset{I}{\hookrightarrow} FT,$$

if for $Th \vdash A$ and $A^I \equiv \exists v \forall w A_I[v,w]$, there is a tuple of terms b of $L(FT)$ (with variables among the variables free in A) such that

$$FT \vdash A_I[b,w]$$

In this case, A is called **I-interpretable in FT**, and the terms b are called (a tuple of) **I-interpreting terms** of A.

1.8 Negative version. Expressions $A^I \equiv \exists v \forall w A_I[v,w]$ are, for non-empty tuple v, not in the negative fragment of a language. The formula A^- is the *negative version* of A, if A^- is obtained from A by replacing any non-empty tuple of quantifiers $\exists v$ in A by $\neg\forall v \neg$ and any disjunction $B \vee C$ in A by $\neg(\neg B \wedge \neg C)$. Then A^{I-} is a formula of $L(Th)$, if A is.

1.9 Characterization problem. A class Γ of formulae of $L(Th)$ which does not refer to I is said to **characterize** I on the basis of Th, if

$$Th + \Gamma \equiv Th + \{A \leftrightarrow A^{I-} \mid A \in L(Th)\}$$

The characterization problem for I is the task to find a suitable class of additional axioms Γ characterizing I. The problem is independent of the I-interpretability of Th, and it may have different solutions for a theory Th in all finite types and its type o fragment.

We look at common features of characterization problems for $\wedge$ and D.
In intuitionistic theories, for quantifier free A, the I-translation of $\forall x \exists y A[x,y]$ is $\exists Y \forall x A[x, Yx]$, for $I = D$ as well as for $I = \wedge$. Thus,

$$B \to B^I \text{ with } B \equiv \forall x \exists y\ A[x,y]$$

is an axiom of choice with qf matrix A for both I in question. Combining these translations with the negative version complicates the situation.

1.10 Axioms of choice in classical context. For A in $L(T)$ or in $L(T_\in)$, respectively, and non-empty tuples x, y of variables of arbitrary type, let

$(qf - AC)$ $\qquad \forall x \neg \forall y \neg A[x,y] \to \neg \forall Y \neg \forall x\ A[x, Yx]$

which is $(AC)^-$ with quantifier free matrix A.
Up to a double negation which is irrelevant in Th due to the stability lemma, $(qf - AC)$ is of the form

$$B \to B^{D-} \text{ with } B \equiv (\forall x \exists y\ A[x,y])^-$$

Similarly, for A in $L(T_{\wedge\in})$ and tuples x, y as above, let

$(qf - ARC)\quad \forall x\neg\forall y\neg A[x,y] \to \neg\forall S, Y\neg\forall x\neg\forall s <\in Sx\,\neg A[x, Ysx]$

This **quantifier free axiom of restricting choice** is literally of the form

$$B \to B^{\wedge-} \text{ with } B \equiv (\forall x\exists y\ A[x,y])^-$$

We therefore put

$$(qf - AC_D) :\equiv (qf - AC) \text{ and } (qf - AC_\wedge) :\equiv (qf - ARC)$$

and let $(qf - AC_I)$ refer to either.

1.11 Lemma. *For $A, B, A[u]$ negative, u a - possibly empty - tuple of variables, Th proves:*
1. $A^{I-} \leftrightarrow A$ *for A in $L(FT)$*
2. $(A \wedge B)^{I-} \leftrightarrow A^{I-} \wedge B^{I-}$
3. $(A \to B)^{I-} \to A^{I-} \to B^{I-}$
4. $(\forall u\neg A[u])^{I-} \to \forall u\neg(A[u])^{I-}$; *furthermore*
5. $Th + (qf - AC_I) \vdash \forall u\neg(A[u])^{I-} \to (\forall u\neg A[u])^{I-}$

Proof. 1. to 3. are straightforward.
4.: For $A[u]^I \equiv \exists v\forall w\ A_I[u,v,w]$, we have

$$(\forall u\neg A[u])^{D-} \equiv (\exists W\forall uv\neg A_D[u,v,Wuv])^- \text{ and}$$

$$(\forall u\neg A[u])^{\wedge-} \equiv (\exists XW\forall uv\neg\forall x <\in Xuv\ A_\wedge[u,v,Wxuv])^-$$

Either formula implies
(1) $\quad (\forall uv\exists w\neg A_I[u,v,w])^-$
which up to two double negations is $\forall u\neg(A[u])^{I-}$.
5.: By $(qf - AC_I)$, (1) implies for $I = D$

$$(\exists W\forall uv\neg A_D[u,v,Wuv])^- \equiv (\forall u\neg A[u])^{D-}$$

and for $I = \wedge$

$$(\exists XW\forall uv\exists x <\in Xuv\neg A_\wedge[u,v,Wxuv])^-$$

which up to a double negation is $(\forall u\neg A[u])^{\wedge-}$.

1.12 Definition. A formula $B \in L(Th)$ is **prenex**, if

$$B \equiv \forall u_1\neg...\forall u_n\neg C[u_1,...,u_n]$$

with $n \geq 0$, - possibly empty - tuples $u_1,...,u_n$ of variables, and
$C[u_1,...,u_n] \in L(T_{\wedge\in})$.

For the Dialectica-translation D, the following result goes back to [12].

1.13 Relative characterization theorems

$$Th + (qf - AC_I) \;\equiv\; Th + \{B \leftrightarrow B^{I-} \mid B\ prenex\}$$

If Th is I-interpretable in FT, then

$$Th + (qf - AC_I) \;\equiv\; Th + \{A \leftrightarrow A^{I-}\}$$

Proof. Let $B \equiv \forall u_1 \neg ... \forall u_n \neg C[u_1, ..., u_n]$ be prenex. Then

$$Th + (qf - AC_I) \vdash B \leftrightarrow B^I$$

follows by 1. and n applications of 4. and 5. in Lemma 1.11.
Conversely, the schema $(qf - AC_I)$, as mentioned in 1.10, is a set of formulae $B \rightarrow B^{I-}$ with prenex B.
Let Th be I-interpretable, and for a given formula A of $L(Th)$, let B be a prenex normal form of A. Then
$Th \vdash A \leftrightarrow B$. Therefore, by I-interpretability of Th,
$Th \vdash (A \leftrightarrow B)^{I-}$, which by 2. and 3. in 1.11 implies
$Th \vdash A^{I-} \leftrightarrow B^{I-}$, and, as already shown,
$Th + (qf - AC_I) \vdash B \leftrightarrow B^{I-}$
Putting the first, the fourth, and the third of these equivalences together, we obtain

$$Th + (qf - AC_I) \vdash A \leftrightarrow A^{I-}.$$

2 Interpretations of classical arithmetical theories

Gödel's Dialectica interpretation of HA in T_0, $HA \overset{D}{\hookrightarrow} T_0$, which in fact is also a Dialectica interpretation of HA_0^ω in T_0, automatically yields Dialectica interpretations of Peano arithmetic PA and of PA_0^ω as subsystems of HA and of HA_0^ω respectively:

2.1 Dialectica interpretation theorems

$$PA \overset{D}{\hookrightarrow} T_0 \quad \text{and} \quad PA_0^\omega \overset{D}{\hookrightarrow} T_0$$

The Dialectica interpretation does not extend to PA^ω, as the following example shows.

2.2 Example, communicated by W. A. Howard

(1) $PA^\omega \vdash (\forall u^1 \exists y^o (y^o = 0 \leftrightarrow u^1 = 0^1))^-$,

a form of excluded middle for the equation $u^1 = 0^1$.
Modulo a double negation, the formula (1) is Dialectica translated into

$$\exists Y \forall u^1 \; (Y u^1 = 0 \leftrightarrow u^1 = 0^1)$$

However, there is no functional $Y : 1 \to o$ in T for which

(2) $T^c \vdash Y u^1 = 0 \leftrightarrow u^1 = 0^1$

This follows from the fact that the functionals $Y : 1 \to o$ in T are continuous, i.e. for each u^1, Yu^1 depends only on finitely many values of u^1, and solutions Y of (2) are not continuous (cf. [17]).
On the other hand, the formula (1) is $\wedge$-translated as

$$\exists XY \neg \forall x < Xu^1 \neg (Yu^1 x = 0 \leftrightarrow u^1 = 0^1),$$

and since

(3) $T_\wedge \vdash \neg \forall y < 2 \, \neg (y = 0 \leftrightarrow u^1 = 0^1)$,
the functionals $X = \lambda u^1.2$ and $Y = \lambda u^1 x.x$ are $\wedge$-interpreting terms for (1).

Since $Stab(=)$ is its own $\wedge$-translation and HA^ω, even $HA^\omega + \{A \leftrightarrow A^\wedge\}$ is $\wedge$-interpretable in $T_\wedge$ (cf. [9]), we have:

2.3 $\wedge$-interpretation theorems

$$PA^\omega \overset{\wedge}{\hookrightarrow} T^c_\wedge, \quad even \quad HA^\omega + \{A \leftrightarrow A^\wedge\} + Stab(=) \overset{\wedge}{\hookrightarrow} T^c_\wedge$$

The last statement does not transfer to PA^ω, because I-interpretability of A in $T_\wedge$ does not imply I-interpretability of A^- in $T^c_\wedge$. It is, however, easily seen by induction on deductions:

2.4 Lemma

$$HA^\omega + Stab(=) \vdash A \quad implies \quad PA^\omega \vdash A^-$$

The $\wedge$-interpretation theorem 2.3 and the relative characterization theorem 1.13 yield as an immediate corollary:

2.5 Characterization theorem for the $\wedge$-translation on PA^ω

$$PA^\omega + (qf - ARC) \equiv PA^\omega + \{A \leftrightarrow A^{\wedge -}\}$$

Any negative formula $\wedge$-interpretable in $T^c_\wedge$ is derivable in $PA^\omega + (qf - ARC)$.

Proof of the second statement. Let A be $\wedge$-interpretable in $T^c_\wedge$. Then $A^\wedge$ is derivable in $HA^\omega + Stab(=)$, and by lemma 2.4, $A^{\wedge -}$ is derivable in PA^ω. Thus, together with $A^{\wedge -} \leftrightarrow A$, A is derivable in $PA^\omega + (qf - ARC)$.

2.6 Extended $\wedge$-interpretation theorem

$$PA^\omega + (qf - ARC) \overset{\wedge}{\hookrightarrow} T^c_\wedge$$

Proof. In addition to the interpretation theorem 2.3, we only have to show that $(qf - ARC)$ is $\wedge$-interpretable in $T^c_\wedge$. However, $(qf - ARC)$ is an instance of $B \to B^{\wedge -}$, as pointed out in 1.10. So, it suffices to $\wedge$-interpret $B \to B^{\wedge -}$ in $T^c_\wedge$ for arbitrary negative B.
Let $B^\wedge \equiv \exists v \forall w\ B_\wedge[v, w]$. Then, after a change of bound variables,
$B^{\wedge -} \equiv \neg \forall y \neg \forall z\ B_\wedge[y, z]$,
$B^{\wedge - \wedge} \equiv \exists TY \forall SZ\, (\exists t < TSZ\, \forall s < S(YtSZ)\ B_\wedge[YtSZ, Zs(YtSZ)])^-$,
and finally

$$(B \to B^{\wedge -})^\wedge \equiv \exists XWTY \forall vSZ\, (\forall x < XvSZ\ B_\wedge[v, WxvSZ] \to \\ \exists t < TvSZ \forall s < S(YtvSZ)\ B_\wedge[YtvSZ, Zs(YtvSZ)])^-$$

$\wedge$-interpreting functionals X, W, T, Y are given by

$$YtvSZ = v, TvSZ = 1, WxvSZ = Zxv, XvSZ = Sv$$

The matrix $(B \to B^{\wedge -})_\wedge$ then reduces to

$$\forall x < Sv\ B_\wedge[v, Zxv] \to \forall s < Sv\ B_\wedge[v, Zsv]$$

which is a tautology in $T_\wedge$.

Corresponding results also hold for the Dialectica interpretation of PA^ω_0, because D may be viewed as a more efficient formulation of $\wedge$, if restricted to PA^ω_0:

2.7 Proposition. *For $A \in L(PA^\omega_0)$*

$$PA^\omega_0 \vdash A^{D-} \leftrightarrow A^{\wedge -}$$

On the basis of PA^ω_0, the schemata $(qf - AC)$ and $(qf - ARC)$ are equivalent.

Proof. The equivalence in the first statement is the negative version of the equivalence

$$HA^\omega_0 \vdash A^D \leftrightarrow A^\wedge$$

shown in [9]. Therefore, the statement follows by lemma 2.4, restricted to type o language, because, following [13], HA^ω is a conservative extension of HA^ω_0.
The second statement follows from the first, because, as remarked in 1.10, axioms

$(qf-AC)$ and $(qf-ARC)$ are of the form $B \to B^{D-}$ and $B \to B^{\wedge -}$ respectively, for the same B.

As corollaries to the characterization theorem 2.5 and the extended interpretation theorem 2.6, this proposition implies:

2.8 Characterization and extended interpretation theorem for the Dialectica translation
(1) *On the basis of* PA_0^ω, *the schemata* $(qf-AC)$ *and* $\{A \leftrightarrow A^{D-}\}$ *are equivalent:*

$$PA_0^\omega + (qf-AC) \equiv PA_0^\omega + \{A \leftrightarrow A^{D-}\}$$

(2) *Negative formulae* D*-interpretable in* T_0 *are derivable in* $PA_0^\omega + (qf-AC)$.
(3)

$$PA_0^\omega + (qf-AC) \overset{D}{\hookrightarrow} T_0$$

Proof. (1) follows from the characterization theorem 2.5 by restricting the language of PA^ω to its type *o* fragment and applying proposition 2.7. By the same argument, (2) follow from the second statement in 2.5.
To prove (3), only a D-interpretation of $(qf-AC)$ must be added to theorem 2.1. $(qf-AC)$ is of the form $B \to B^{D-}$, and that is D-interpreted as follows:
Let $B^D \equiv \exists v \forall w B_D[v,w]$ and $B^{D-} \equiv \neg\forall y \neg\forall z B_D[y,z]$. Then

$$B^{D-D} \equiv \exists Y \forall Z \neg\neg B_D[YZ, Z(YZ)]$$

and finally

$$(B \to B^{D-})^D \equiv \exists WY \forall vZ (B_D[v, WvZ] \to \neg\neg B_D[YvZ, Z(YvZ)])$$

D-interpreting functionals W, Y are now given by $YvZ = v$ and $WvZ = Zv$. These can also be extracted from the $\wedge$-interpretation of $(qf-ARC)$.

These results complete the discussion of the Dialectica interpretation of PA_0^ω and its characterization. Concerning the relation of the translations D and $\wedge$ on PA^ω proper, some details remain to be settled, some rest unsolved.

2.9 Proposition. *Let* $(qf-AC)_0$ *denote the schema* $(qf-AC)$, *restricted to formulae of type o.*
(1) $PA^\omega + (qf-AC) \vdash (qf-ARC)$
(2) $PA^\omega + (qf-ARC) \vdash (qf-AC)_0$
(3) $PA^\omega \nvdash (qf-AC)_0$
(4) $PA^\omega \nvdash (qf-ARC)$
(5) $PA^\omega + (qf-ARC) \vdash (qf-AC)$ *iff* $(qf-AC)$ *is* $\wedge$*-interpretable in* $T_\wedge^c$.

Proof. (1) Clearly, $(\exists Y' \forall x A[x, Y'x])^-$ implies $(\exists S, Y \forall x \exists s < SxA[x, Ysx])^-$,

simply by putting $Sx = 1$ and, given Y', $Ysx = Y'x$. Therefore, any axiom $(qf - AC)$ implies the corresponding axiom $(qf - ARC)$.
(2) is an immediate consequence of proposition 2.7.
(3) Let T denote Kleene's T-predicate. For any numeral e,

$$\forall x \exists y \, Texy \to \exists Y \forall x \, Tex(Yx)$$

is an instance of $(qf - AC)_0$. Now let e be an index of a total recursive function which is not provably recursive in PA^ω. In the model NF of primitive recursive functionals in normal form (cf. [17], there called $CTNF$), we have

$$NF \models \forall x \exists y Texy, \text{ but for no } Y : \; NF \models \forall x Tex(Yx)$$

Therefore $NF \not\models (qf - AC)_0$, and (3) follows, as NF is a model of PA^ω.
(4) is immediate from (2) and (3).
(5) is an application of the characterization theorem 2.5 (2) and the extended $\wedge$-interpretation theorem 2.6.

By (1) of this proposition, the sequence of theories

$$PA^\omega + (qf - AC) \qquad PA^\omega + (qf - ARC) \qquad PA^\omega$$

is of decreasing strength, and by (4), the second theory is properly stronger than the last. We conjecture that $PA^\omega + (qf - ARC) \nvdash (qf - AC)$.

Due to the lack of a D-interpretation theorem for PA^ω, it remains an open problem whether $PA^\omega + (qf - AC) \vdash A \leftrightarrow A^D$ for all $A \in L(PA^\omega)$.

3 Interpretation of Kripke-Platek set theories

The theories $T_\in$ and $T_\in^c$ allow quite flexible operations on sets. To a part, this is due to explicit Δ_0-separation in $T_\in$:

3.1 Proposition, explicit Δ_0-separation. *To any Δ_0-formula A and any term $t : o$ with $x : o$ not in t, there exists a separation term $\{x \in t \mid A[x]\}$ such that*

$$T_\in \vdash y \in \{x \in t \mid A[x]\} \leftrightarrow y \in t \wedge A[y]$$

Any Δ_0-formula A possesses a characteristic term $\{0 \mid A\} = \{x \in 1 \mid A\}$ with x not in A such that

$$T_\in \vdash 0 \in \{0 \mid A\} \leftrightarrow \{0 \mid A\} = 1 \leftrightarrow A$$

For a proof, see [3] or [8].

In analogy to the situation in $T_\wedge$, a principle of induction holds in $T_\in$ which is the essential technical tool for the $\wedge$-interpretation of transfinite induction $(T\ IND)$. As an alternative to [3], we give a proof of this principle closely related to the proof of the corresponding principle in $T_\wedge$ (Satz 1 in [9]).

3.2 Proposition, generalized transfinite induction. *Given a term X, a term tuple W with variables $a, u, x : o$ and a variable tuple z, all not in X, W, such that*

(1) $\quad T_\in \vdash \forall u \in a\, \forall x \in Xaz\ B[u, Wxaz] \to B[a, z]$

Then $\quad T_\in \vdash B[t, z]$ $\quad$ *for any term $t : o$.*

Proof. Let $TC\{t\}$ be the transitive closure of $\{t\}$ and $S := Finseq(TC\{t\})$ the set of finite sequences of elements from $TC\{t\}$ (cf. [2] and [3]). By simultaneous ω-recursion on S, we define a term tuple X_1, Z by

$$\begin{array}{ll} Zy\langle\rangle = z & Z\langle x, y\rangle\langle s, u\rangle = Wxa(Zys) \\ X_1\langle\rangle = 1 & X_1\langle s, u\rangle \qquad = \{\langle x, y\rangle \mid x \in Xa(Zys), y \in X_1 s\} \end{array}$$

Assuming $a \in TC\{t\}, s \in S$, we have

(2) $\forall u \in a(\forall y \in X_1\langle s, u\rangle\ B[u, Zy\langle s, u\rangle] \leftrightarrow$
$\qquad\qquad\qquad \forall y \in X_1 s\, \forall x \in Xa(Zys)\ B[u, Wxa(Zys)])$

(1), with Zys substituted for z, may be rewritten under this equivalence as

$$\forall u \in a \forall y \in X_1\langle s, u\rangle\ B[u, Zy\langle s, u\rangle] \to \forall y \in X_1 s\ B[a, Zys]$$

After distribution of $a \in TC\{t\}$ and $\forall s \in S$ over this implication, $(T\ IND)$ yields

$$T_\in \vdash t \in TC\{t\} \to \forall s \in S \forall y \in X_1 s\ B[t, Zys]$$

For $s = \langle\rangle$, this implies $T_\in \vdash B[t, z]$, as was to be shown.

Propositions 3.1 and 3.2 are derivability results within the constructive functional theory $T_\in$, independent of $Stab(=)$. We now turn to the problem of functional interpretability of $KP\omega$ and $KP\omega^\omega$ in $T^c_\in$.

3.3 Proposition. *Already the type o theory $KP\omega$ is not Dialectica interpretable in $T^c_\in$.*

Proof by example. $KP\omega$ (with a constant 0 for the empty set) proves

$$\forall x(\forall y\, \neg y \in x \to x = 0)$$

This has the D-translation

$$\exists Y \forall x(\neg Yx \in x \to x = 0)$$

Any Y satisfying this formula is necessarily a classical choice functional which is not a constructive set functional.

However, the $\wedge$-translation of this formula is

$$\exists ZY\forall x\ (\forall z \in Zx\ \neg Yz \in x \to x = 0),$$

and $\wedge$-interpreting functionals Z, Y are given by $Zx = x$ and $Yz = z$.
Here we made use of a simplification of the $\wedge$-translation which will be useful later, too:

3.4 Lemma. *A formula $\forall y\, A[y] \to B$ with A, B in $L(T_\in)$, $y : o$ may be $\wedge$-translated as*

$$\exists Y(\forall y \in Y\ A[y] \to B)$$

Proof. $(\forall y\, A[y] \to B)^\wedge$ is literally $\exists XY'(\forall x \in X\ A[Y'x] \to B)$. Y is obtained from X, Y' by $Y = \{Y'x \mid x \in X\}$, and X, Y' are obtained from Y by $X = Y$ and $Y'x = x$. Moreover, in the translation of longer formulae, the tuple of variables X, Y' and the variable Y are handled in exactly the same way.

3.5 $\wedge$-interpretation theorem for $KP\omega^\omega$
$KP\omega^\omega$ is $\wedge$-interpretable in $T^c_\in$:

$$KP\omega^\omega \overset{\wedge}{\hookrightarrow} T^c_\in$$

Proof by induction on deductions. $KP\omega^\omega$, including its Δ_0-sublanguage, is formulated in the negative fragment of first order logic. The $\wedge$-interpretation of the axioms and rules of this fragment, including identity, may be taken over from [9], replacing $<$ by $\in$. Results from $T_\wedge$ which are used there have to be transferred to $T_\in$. That, however, is easily done, in particular by exploiting explicit Δ_0-separation 3.1 (cf [3] and [8]).
$Stab(=)$, type extensionality $(T\text{-}EXT)$, with side formulae restricted to $L(T_\in)$, as well as the axiom of set-extensionality (ext), written as a Δ_0-formula, are all interpreted by the empty tuple.

(Δ_0-separation) $\quad \neg\forall b\neg\forall x(x \in b \leftrightarrow x \in t \wedge A[x])$
for all Δ_0-formulae $A[x]$ and terms $t : o$ with b not in $A[x]$ and b, x not in t.
The part of this axiom following $\forall b$ may be rewritten as a Δ_0-formula. Hence, by lemma 3.4, the axiom may be $\wedge$-translated as

$$\exists Y\neg\forall b \in Y\neg\forall x(x \in b \leftrightarrow x \in t \wedge A[x])$$

By explicit Δ_0-separation 3.1, the only relevant $b \in Y$ is the separation term $\{x \in t \mid A[x]\}$, and the singleton of this separation term is a $\wedge$-interpreting term Y.

Axioms $(Pair), (Union), (Infinity)$ are interpreted analogously, making use of the set functionals available in $T_\in$ (cf. [3]).

(Δ_0-collection) $\forall x \in a \neg\forall y \neg A[x, y] \to \neg\forall z \neg\forall x \in a \exists y \in z\, A[x, y]$
for Δ_0-formulae $A[x, y]$
The $\wedge$-translation of this formula is, using lemma 3.4,

$$\exists Z \forall Y (\forall x \in a\, \exists y \in Yx\, A[x, y] \to \exists z \in ZY \forall x \in a\, \exists y \in z\, A[x, y])^-$$

Given Y satisfying the antecedent, there is one canonical z satisfying the consequent, namely $z = \bigcup\{Yx \mid x \in a\}$. The implication ($\Delta_0$-collection) is therefore $\wedge$-interpreted by the term Z with the value $ZY = \{\bigcup\{Yx \mid x \in a\}\}$.

(T IND) $\forall u \in a\, F[u] \to F[a] \vdash F[t]$
By I.H., there are terms and term tuples X, W, Y_0 such that
(3) $T^c_\in \vdash \forall x \in Xavz\, \forall u \in a\, F_\wedge[u, vu, Wxavz] \to F_\wedge[a, Y_0va, z]$
We define terms Y by simultaneous transfinite recursion

$$Ya = Y_0(Y \upharpoonright a)a,$$

substitute $Y \upharpoonright a$ for v in (3), and obtain

$$T^c_\in \vdash \forall x \in Xa(Y \upharpoonright a)z\, \forall u \in a\; F_\wedge[u, Yu, Wxa(Y \upharpoonright a)z] \to F_\wedge[a, Ya, z]$$

Here, the terms $Xa(Y \upharpoonright a)z$, $Wxa(Y \upharpoonright a)z$ are terms $X'az, W'xaz$, and $F_\wedge[a, Ya, z]$ is a formula $B[a, z]$ satisfying (1) in proposition 3.2. So, by generalized transfinite induction, $F_\wedge[t, Yt, z]$ follows.
This completes the proof of the $\wedge$-interpretation theorem. For related proofs using, however, different translations, cf. [3], [4], [15]. An immediate consequence is:

3.6 Corollary, conservativity and relative consistency. *$KP\omega^\omega$ is a conservative extension of $T^c_\in$. The consistency of $T^c_\in$ implies the consistency of $KP\omega^\omega$.*

The characterization problem for the $\wedge$-translation on $KP\omega^\omega$ is solved by theorem 1.13:

3.7 Characterization theorem

$$KP\omega^\omega + (qf - ARC) \equiv KP\omega^\omega + \{A \leftrightarrow A^{\wedge-}\}$$

$KP\omega^\omega + (qf - ARC)$ proves any negative formula $\wedge$-interpretable in $T^c_\in$.

It may be conjectured that $KP\omega^\omega + (qf - ARC) \nvdash (qf - AC)$ and that therefore $(qf - AC)$ is not $\wedge$-interpretable in $T^c_\in$. On the other hand, a $\wedge$-interpretation of $(qf - ARC)$ in $T^c_\in$ is obtained by simply substituting $\in$ for $<$ in the proof of theorem 2.6:

3.8 Extended $\wedge$-interpretation theorem for $KP\omega^\omega + (qf - ARC)$

$$KP\omega^\omega + (qf - ARC) \overset{\wedge}{\hookrightarrow} T^c_\in$$

References

[1] Avigad, J., and S. Feferman: Gödel's functional ('Dialectica') interpretation, in: S. Buss (Ed.), Handbook of Proof Theory, Studies in Logic and the Foundations of Mathematics, Vol. 137, Elsevier, Amsterdam 1998, 337 - 406.

[2] Barwise, J.: Admissible Sets and Structures, Springer-Verlag, Berlin Heidelberg New York 1975.

[3] Burr, W.: Functional interpretation of Aczel's constructive set theory, APAL 104 (2000) 31 - 73.

[4] Burr, W.: A Diller-Nahm-style functional interpretation of $KP\omega$, Arch. Math. Logic 39 (2000) 599 - 604.

[5] Burr, W.: Concepts and aims of functional interpretations: Towards a functional interpretation of constructive set theory, Synthese 133 (2002) 257 - 274.

[6] Burr, W., and V. Hartung: A characterization of Σ_1-definable functions of $KP\omega + (uniformAC)$, Arch. Math. Logic 37 (1998) 199 - 214.

[7] Diller, J.: Logical problems of functional interpretations, APAL 114 (2002) 27 - 42.

[8] Diller, J.: Functional interpretations of constructive set theory in all finite types. Dialectica 62 (2008) 149 - 177.

[9] Diller, J., and W. Nahm: Eine Variante zur Dialectica-Interpretation der Heyting-Arithmetik endlicher Typen, Arch. Math. Logik Grundl. 16 (1974) 49 - 66.

[10] Diller, J., and K. Schütte: Simultane Rekursionen in der Theorie der Funktionale endlicher Typen, Arch. Math. Logik Grundl. 14 (1971) 69 - 74.

[11] Gödel, K.: Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes, Dialectica 12 (1958) 280 - 287.

[12] Kreisel, G.: Interpretation of analysis by means of constructive functionals of finite type, in: A. Heyting (Ed.), Constructivity in Mathematics, North Holland Publ. Co., Amsterdam 1959, 101 - 128.

[13] Rath, P.: Eine verallgemeinerte Funktionalinterpretation der Heyting-Arithmetik endlicher Typen, Ph.D. thesis, University of Münster, 1978.

[14] Schütte, K.: Proof Theory, Springer-Verlag, Heidelberg/New York 1977.

[15] Schulte, D.: Hybrids of the $\times$-translation for CZF^ω, J. Applied Logic 6 (2008) 443 - 458.

[16] Shoenfield, J.R.: Mathematical Logic, Addison-Wesley Publ. Comp., Reading, MA, 1967.

[17] Troelstra, A.S.: Metamathematical investigation of intuitionistic arithmetic and analysis, Lecture Notes in Mathematics 344, Springer, Heidelberg/New York 1973.

[18] Troelstra, A.S.: Introductory Note to 1958 and 1972, in: K. Gödel, Collected Works, vol. II, Publications 1938 - 1974, S. Feferman (Ed.), The Clarendon Press, Oxford University Press, New York 1990.

[19] Troelstra, A.S., and D. van Dalen: Constructivism in Mathematics, Vol. I, North Holland, Amsterdam 1988.

Towards a Formal Theory of Computability

Simon Huber, Basil A. Karádais, and Helmut Schwichtenberg

Abstract We sketch a constructive formal theory TCF^+ of computable functionals, based on the partial continuous functionals as their intended domain. Such a task had long ago been started by Dana Scott [12, 15], under the well-known abbreviation LCF (logic of computable functionals). The present approach differs from Scott's in two aspects.

(i) The intended semantical domains for the base types are non-flat free algebras, given by their constructors, where the latter are injective and have disjoint ranges; both properties do not hold in the flat case.

(ii) TCF^+ has the facility to argue not only about the functionals themselves, but also about their finite approximations.

In this setting we give an informal proof (based on Berger [2]) of Kreisel's density theorem [7], and an adaption of Plotkin's definability theorem [10, 11]. We then show that both proofs can be formalized in TCF^+.

The naive model of a finitely typed theory like TCF^+ is the full set theoretic hierarchy of functionals of finite types. However, this immediately leads to higher cardinalities, and does not lend itself well for a constructive theory of computability. A more appropriate semantics for typed languages has its roots in work of Kreisel [7] (where formal neighborhoods are used) and Kleene [6]. This line of research was developed in a mathematically more satisfactory way by Scott [13] and Ershov [3]. Today this theory is usually presented in the context of abstract domain theory (see [1, 16]); it is based on classical logic. The present work can be seen as an attempt to develop a constructive theory of formal neighborhoods for continuous functionals, in a direct and intuitive style. The task is to replace abstract domain theory by a more concrete, finitary theory of representations. As a framework we use Scott's information systems (see [8, 14, 16]). In this setup the basic notion is that of a "token", or unit of information. The elements or points of the domain appear as abstract or "ideal" entities: possibly infinite sets of tokens, which are "consistent" and "deductively closed".

The paper is organized as follows. Section 1 collects basic facts about information systems, and section 2 contains informal proofs of the density and definability theorems for the case of the non-flat natural numbers, in enough detail to guide the formalization. Section 3 develops the language and axioms of the theory TCF^+. The formalization of both theorems in TCF^+ is discussed in section 4.

1 Partial Continuous Functionals

1.1 Information systems

The basic idea of information systems is to provide an axiomatic setting to describe approximations of abstract objects (like functions or functionals) by concrete, finite ones. The axioms below are a minor modification of Scott's [14], due to Larsen and Winskel [8].

An *information system* is a structure $(A, \mathrm{Con}, \vdash)$ where A is a countable set (the *tokens*), Con is a nonempty set of finite subsets of A (the *consistent* sets) and $\vdash$ is a subset of $\mathrm{Con} \times A$ (the *entailment* relation), which satisfy

$$\begin{aligned}
&U \subseteq V \in \mathrm{Con} \to U \in \mathrm{Con}, \\
&\{a\} \in \mathrm{Con}, \\
&U \vdash a \to U \cup \{a\} \in \mathrm{Con}, \\
&a \in U \in \mathrm{Con} \to U \vdash a, \\
&U, V \in \mathrm{Con} \to \forall_{a \in V}(U \vdash a) \to V \vdash b \to U \vdash b.
\end{aligned}$$

The elements U of Con are called *formal neighborhoods*. We use U, V, W to denote *finite* sets, and write

$$\begin{aligned}
&U \vdash V \quad \text{for} \quad U \in \mathrm{Con} \ \wedge \ \forall_{a \in V}(U \vdash a), \\
&a \uparrow b \quad \text{for} \quad \{a, b\} \in \mathrm{Con} \qquad (a, b \text{ are } \textit{consistent}), \\
&U \uparrow V \quad \text{for} \quad \forall_{a \in U, b \in V}(a \uparrow b).
\end{aligned}$$

The *ideals* (also called *objects*) of an information system $\boldsymbol{A} = (A, \mathrm{Con}, \vdash)$ are defined to be those subsets x of A which satisfy

$$\begin{aligned}
&U \subseteq x \to U \in \mathrm{Con} \quad (x \text{ is } \textit{consistent}), \\
&x \supseteq U \vdash a \to a \in x \quad (x \text{ is } \textit{deductively closed}).
\end{aligned}$$

For example the *deductive closure* $\overline{U} := \{\, a \mid U \vdash a \,\}$ of U is an ideal. The set of all ideals of $\boldsymbol{A}$ is denoted by $|\boldsymbol{A}|$.

Examples. Every countable set A can be turned into a *flat* information system by letting the set of tokens be A, $\mathrm{Con} := \{\varnothing\} \cup \{\, \{a\} \mid a \in A \,\}$ and $U \vdash a$ mean $a \in U$. In this case the ideals are just the elements of Con.

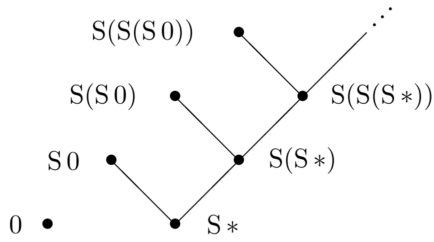

Figure 1: Tokens and entailment for $\mathbf{N}$

Consider the algebras $\mathbf{B}$ (booleans), $\mathbf{N}$ (natural numbers), $\mathbf{P}$ (positive numbers written binary), $\mathbf{D}$ (derivations) given by the constructors

$$\mathsf{tt}^{\mathbf{B}}, \mathsf{ff}^{\mathbf{B}} \quad \text{for } \mathbf{B},$$
$$0^{\mathbf{N}} \text{ and } \mathrm{S}^{\mathbf{N}\to\mathbf{N}} \text{ (successor) for } \mathbf{N},$$
$$\mathbf{1}^{\mathbf{P}}, \mathsf{s}_0^{\mathbf{P}\to\mathbf{P}} \text{ (append 0) and } \mathsf{s}_1^{\mathbf{P}\to\mathbf{P}} \text{ (append 1) for } \mathbf{P},$$
$$0^{\mathbf{D}} \text{ (axiom) and } \mathrm{C}^{\mathbf{D}\to\mathbf{D}\to\mathbf{D}} \text{ (rule) for } \mathbf{D}.$$

For each of them we define an information system $\boldsymbol{C}_\iota = (\mathrm{Tok}_\iota, \mathrm{Con}_\iota, \vdash_\iota)$:

(a) The *tokens* $a \in \mathrm{Tok}_\iota$ are the constructor expressions $\mathrm{C}a_1^* \dots a_n^*$ where a_i^* is an *extended token*, i.e., a token or the special symbol $*$ which carries no information.

(b) A finite set U of tokens in Tok_ι is *consistent* (i.e., $\in \mathrm{Con}_\iota$) if its elements start with the same n-ary constructor C, say $U = \{\mathrm{C}\vec{a_1^*}, \dots, \mathrm{C}\vec{a_m^*}\}$, and $U_i \in \mathrm{Con}_\iota$ where U_i consists of the (proper) tokens among $a_{1i}^*, \dots, a_{mi}^*$.

(c) $\{\mathrm{C}\vec{a_1^*}, \dots, \mathrm{C}\vec{a_m^*}\} \vdash_\iota \mathrm{C}'\vec{a^*}$ is defined to mean $\mathrm{C} = \mathrm{C}'$, $m \geq 1$ and $U_i \vdash a_i^*$, with U_i as in (b) above (and $U \vdash *$ defined to be true).

For example, the tokens for $\mathbf{N}$ are shown in Figure 1. For tokens a, b we have $\{a\} \vdash b$ if and only if there is a path from a (up) to b (down). In $\mathbf{D}$, the set $\{\mathrm{C}0*, \mathrm{C}{*}0\}$ is consistent, and $\{\mathrm{C}0*, \mathrm{C}{*}0\} \vdash \mathrm{C}00$.

A token is called *total* if it has the form $\mathrm{C}\vec{a}$ with a total token a_i at every argument position. For example, the total tokens for $\mathbf{N}$ are all $\mathrm{S}^n\, 0$, and for $\mathbf{D}$ all $*$-free constructor trees built from 0 and C.

By induction on the formation of tokens, one easily sees the following.

Lemma 1.1 (Comparability)**.** *If ι has at most unary constructors, then any two consistent tokens a, b are comparable, i.e., $\{a\} \vdash b$ or $\{b\} \vdash a$.*

1.2 Function spaces

Let $\boldsymbol{A} = (A, \mathrm{Con}_A, \vdash_A)$ and $\boldsymbol{B} = (B, \mathrm{Con}_B, \vdash_B)$ be information systems. Define the *function space* $\boldsymbol{A} \to \boldsymbol{B} = (C, \mathrm{Con}, \vdash)$ by

$$C := \mathrm{Con}_A \times B,$$
$$\{ (U_i, b_i) \mid i \in I \} \in \mathrm{Con} := \forall_{J \subseteq I} (\bigcup_{j \in J} U_j \in \mathrm{Con}_A \to \{ b_j \mid j \in J \} \in \mathrm{Con}_B).$$

For the definition of the entailment relation $\vdash$ it is helpful to first define the notion of an *application* of $W := \{ (U_i, b_i) \mid i \in I \} \in \mathrm{Con}$ to $U \in \mathrm{Con}_A$:

$$\{ (U_i, b_i) \mid i \in I \} U := \{ b_i \mid U \vdash_A U_i \}.$$

From the definition of Con we know that this set is in Con_B. Now define $W \vdash (U, b)$ by $WU \vdash_B b$. Clearly application is *monotone in the second argument*, in the sense that $U \vdash_A U'$ implies $WU' \subseteq WU$, hence $WU \vdash_B WU'$. Application is also *monotone in the first argument*, i.e.,

$$W \vdash W' \quad \text{implies} \quad WU \vdash_B W'U.$$

Using this one easily proves that $\boldsymbol{A} \to \boldsymbol{B}$ is an information system provided $\boldsymbol{A}$ and $\boldsymbol{B}$ are.

For any information system $\boldsymbol{A}$ the set of all $\mathcal{O}_U := \{ x \in |\boldsymbol{A}| \mid U \subseteq x \}$ with $U \in \mathrm{Con}$ forms the basis of a topology on $|\boldsymbol{A}|$, the *Scott topology*. The continuous functions (w.r.t. the Scott topology) from $|\boldsymbol{A}|$ to $|\boldsymbol{B}|$ are in a natural bijective correspondence with the ideals of $\boldsymbol{A} \to \boldsymbol{B}$:

(a) With any ideal $r \in |\boldsymbol{A} \to \boldsymbol{B}|$ we can associate a continuous function $|r| \colon |\boldsymbol{A}| \to |\boldsymbol{B}|$ by $|r|z := \{ b \in B \mid (U, b) \in r \text{ for some } U \subseteq z \}$. We call $|r|z$ the *application* of r to z.

(b) Conversely, with any continuous function $f \colon |\boldsymbol{A}| \to |\boldsymbol{B}|$ we can associate an ideal $\hat{f} \colon \boldsymbol{A} \to \boldsymbol{B}$ by $\hat{f} := \{ (U, b) \mid b \in f(\overline{U}) \}$.

These assignments are inverse to each other, i.e., $f = |\hat{f}|$ and $r = \widehat{|r|}$. We usually write rz for $|r|z$, and similarly $(U, b) \in f$ for $(U, b) \in \hat{f}$.

Lemma 1.2 (Approximable maps [14])**.** *Let $\boldsymbol{A} = (A, \mathrm{Con}_A, \vdash_A)$ and $\boldsymbol{B} = (B, \mathrm{Con}_B, \vdash_B)$ be information systems. The ideals of $\boldsymbol{A} \to \boldsymbol{B}$ are exactly the* approximable maps *from $\boldsymbol{A}$ to $\boldsymbol{B}$, i.e., the relations $r \subseteq \mathrm{Con}_A \times B$ with*

(a) *If $(U, b_1), \ldots, (U, b_n) \in r$, then $\{b_1, \ldots, b_n\} \in \mathrm{Con}_B$;*

(b) *If* $(U, b_1), \dots, (U, b_n) \in r$ *and* $\{b_1, \dots, b_n\} \vdash_B b$*, then* $(U, b) \in r$*;*

(c) *If* $(U', b) \in r$ *and* $U \vdash_A U'$*, then* $(U, b) \in r$.

Types are built from base types ι (the algebras above) by $\rho \to \sigma$. For every type ρ we define the information system $\boldsymbol{C}_\rho = (\mathrm{Tok}_\rho, \mathrm{Con}_\rho, \vdash_\rho)$ starting from the $\boldsymbol{C}_\iota$ by formation of function spaces $\boldsymbol{C}_{\rho\to\sigma} := \boldsymbol{C}_\rho \to \boldsymbol{C}_\sigma$. The set $|\boldsymbol{C}_\rho|$ of ideals in $\boldsymbol{C}_\rho$ is the set of *partial continuous functionals* of type ρ. A partial continuous functional $x \in |\boldsymbol{C}_\rho|$ is *computable* if it is recursively enumerable when viewed as a set of tokens. The information systems $\boldsymbol{C}_\rho$ enjoy the pleasant property of "coherence", which amounts to the possibility of locating inconsistencies in two-element sets of data objects. Generally, an information system $\boldsymbol{A} = (A, \mathrm{Con}, \vdash)$ is *coherent* if it satisfies: $U \subseteq A$ is consistent if and only if all of its two-element subsets are.

It is easy to see that every constructor C generates a continuous function $r_{\mathrm{C}} := \{\, (\vec{U}, \mathrm{C}\vec{a^*}\,) \mid \vec{U} \vdash \vec{a^*}\,\}$ in the function space (where $(\vec{U}, b)$ means $(U_1, \dots (U_n, b) \dots))$, and that

$$|r_{\mathrm{C}}|\vec{x} \subseteq |r_{\mathrm{C}}|\vec{y} \leftrightarrow \vec{x} \subseteq \vec{y}.$$

If $\mathrm{C}_1, \mathrm{C}_2$ are distinct constructors of ι, then $|r_{\mathrm{C}_1}|\vec{x} \neq |r_{\mathrm{C}_2}|\vec{y}$, since the two ideals are non-empty and disjoint. Hence constructors are injective and have disjoint ranges. Notice that neither property holds for flat information systems, since for them, by monotonicity, constructors need to be *strict* (i.e., if one argument is the empty ideal, then the value is as well). But then

$$|r_{\mathrm{C}}|\varnothing y = \varnothing = |r_{\mathrm{C}}|x\varnothing, \qquad |r_{\mathrm{C}_1}|\varnothing = \varnothing = |r_{\mathrm{C}_2}|\varnothing,$$

where C is a binary and C_1, C_2 are unary constructors.

2 Computable functionals

2.1 Terms and their denotational semantics

Terms are built from (typed) variables and (typed) constants (constructors C or defined constants D, see below) by application and abstraction:

$$M, N ::= x^\rho \mid \mathrm{C}^\rho \mid D^\rho \mid (\lambda_{x^\rho} M^\sigma)^{\rho\to\sigma} \mid (M^{\rho\to\sigma} N^\rho)^\sigma.$$

Every defined constant D comes with a system of *computation rules*, consisting of finitely many equations $D\vec{P}_i(\vec{y}_i) = M_i$ $(i = 1, \dots, n)$ with free variables of $\vec{P}_i(\vec{y}_i)$ and M_i among $\vec{y}_i$, where the $\vec{P}_i(\vec{y}_i)$ must be "constructor patterns", i.e., lists of applicative terms built from constructors and distinct variables, with each constructor C occurring in a context $\mathrm{C}\vec{P}$ (of base type). We assume that $\vec{P}_i$ and $\vec{P}_j$ for $i \neq j$ are non-unifiable. Examples are

(i) the predecessor function $\mathrm{P}\colon \mathbf{N} \to \mathbf{N}$ defined by the computation rules $\mathrm{P}0 = 0$, $\mathrm{P}(\mathrm{S}\,n) = n$,

(ii) Gödel's primitive recursion operators $\mathcal{R}^{\tau}_{\mathbf{N}}\colon \mathbf{N} \to \tau \to (\mathbf{N} \to \tau \to \tau) \to \tau$ with computation rules $\mathcal{R}0fg = f$, $\mathcal{R}(\mathrm{S}\,n)fg = gn(\mathcal{R}nfg)$, and

(iii) the least-fixed-point operators Y_ρ of type $(\rho \to \rho) \to \rho$ defined by the computation rule $Y_\rho f = f(Y_\rho f)$.

For every closed term $\lambda_{\vec{x}} M$ of type $\vec{\rho} \to \sigma$ we inductively define a set $[\![\lambda_{\vec{x}} M]\!]$ of tokens of type $\vec{\rho} \to \sigma$.

$$\frac{U_i \vdash b}{(\vec{U}, b) \in [\![\lambda_{\vec{x}} x_i]\!]}(V), \qquad \frac{(\vec{U}, V, c) \in [\![\lambda_{\vec{x}} M]\!] \qquad (\vec{U}, V) \subseteq [\![\lambda_{\vec{x}} N]\!]}{(\vec{U}, c) \in [\![\lambda_{\vec{x}} (MN)]\!]}(A).$$

For every constructor C and defined constant D we have

$$\frac{\vec{V} \vdash \vec{b^*}}{(\vec{U}, \vec{V}, \mathrm{C}\vec{b^*}) \in [\![\lambda_{\vec{x}} \mathrm{C}]\!]}(\mathrm{C}), \qquad \frac{(\vec{U}, \vec{V}, b) \in [\![\lambda_{\vec{x},\vec{y}} M]\!] \qquad \vec{W} \vdash \vec{P}(\vec{V})}{(\vec{U}, \vec{W}, b) \in [\![\lambda_{\vec{x}} D]\!]}(D),$$

with one such rule (D) for every computation rule $D\vec{P}(\vec{y}\,) = M$.

Here $(\vec{U}, V) \subseteq [\![\lambda_{\vec{x}} M]\!]$ means $(\vec{U}, b) \in [\![\lambda_{\vec{x}} M]\!]$ for all (finitely many) $b \in V$, and $(\vec{U}, b)$ denotes $(U_1, \dots (U_n, b) \dots)$. For a constructor pattern $\vec{P}(\vec{x}\,)$ and a list $\vec{V}$ of the same length and types as $\vec{x}$, $\vec{P}(\vec{V})$ is a list of formal neighborhoods of the same length and types as $\vec{P}(\vec{x}\,)$: $x(V)$ is V, and

$$(\mathrm{C}\vec{P})(\vec{V}) := \{\, \mathrm{C}\vec{b^*} \mid b^*_i \in P_i(\vec{V_i}) \text{ if } P_i(\vec{V_i}) \neq \varnothing, \text{ and } b^*_i = * \text{ otherwise} \,\}.$$

The *height* of a derivation of $(\vec{U}, b) \in [\![\lambda_{\vec{x}} M]\!]$ is defined as usual, by adding 1 at each rule. We define its *D-height* similarly, where only rules (D) count.

Theorem 2.1. (a) *For every term M, $[\![\lambda_{\vec{x}} M]\!]$ is an ideal.*

(b) *If a term M converts to M' by $\beta\eta$-conversion or application of a computation rule, then its value is preserved, i.e., $[\![M]\!] = [\![M']\!]$.*

For a term M with free variables among $\vec{x}$ and an assignment $\vec{x} \mapsto \vec{u}$ of ideals $\vec{u}$ to $\vec{x}$ let $[\![M]\!]^{\vec{u}}_{\vec{x}} := \bigcup_{\vec{U} \subseteq \vec{u}} [\![M]\!]^{\vec{U}}_{\vec{x}}$ with $[\![M]\!]^{\vec{U}}_{\vec{x}} := \{\, b \mid (\vec{U}, b) \in [\![\lambda_{\vec{x}} M]\!] \,\}$. Notice that a consequence of (A) is

$$c \in [\![MN]\!]^{\vec{u}}_{\vec{x}} \leftrightarrow \exists_{V \subseteq [\![N]\!]^{\vec{u}}_{\vec{x}}} ((V, c) \in [\![M]\!]^{\vec{u}}_{\vec{x}}) \qquad \text{(continuity of application).} \quad (2.1)$$

Proposition 2.2. *For every $n > 0$, there is a derivation of $(W, b) \in [\![Y]\!]$ with D-height n if and only if $W^n\varnothing \vdash b$.*

Proof. Every derivation of $(W, b) \in [\![Y]\!]$ must have the form

$$\dfrac{\dfrac{\dfrac{\hat{W} \vdash (V, b)}{(\hat{W}, V, b) \in [\![\lambda_f f]\!]} \qquad \dfrac{(\hat{W}, W_i, b_i) \in [\![\lambda_f Y]\!] \quad \dfrac{\hat{W} \vdash (V_{ij}, b_{ij})}{(\hat{W}, V_{ij}, b_{ij}) \in [\![\lambda_f f]\!]}}{(\hat{W}, b_i) \in [\![\lambda_f (Yf)]\!]}}{(\hat{W}, b) \in [\![\lambda_f (f(Yf))]\!]}}{(W, b) \in [\![Y]\!]}\,(D), \text{ assuming } W \vdash \hat{W}$$

with $V := \{\, b_i \mid i \in I \,\}$, $W_i := \{\, (V_{ij}, b_{ij}) \mid j \in I_i \,\}$.

"$\to$". By induction on the D-height. We have $(\hat{W}, W_i, b_i) \in [\![\lambda_f Y]\!]$, $\hat{W} \vdash W_i$ and $\hat{W} \vdash (V, b)$. By induction hypothesis $W_i^{n_i}\varnothing \vdash b_i$, and $\hat{W}^{n_i}\varnothing \vdash W_i^{n_i}\varnothing$ by monotonicity of application. Because of $\hat{W}^{n+1}\varnothing \vdash \hat{W}^n\varnothing$ (proved by induction on n, using monotonicity) we obtain $\hat{W}^n\varnothing \vdash b_i$ with $n := \max n_i$, i.e., $\hat{W}^n\varnothing \vdash V$. Recall that $\hat{W} \vdash (V, b)$ was defined to mean $\hat{W}V \vdash b$. Hence $\hat{W}(\hat{W}^n\varnothing) \vdash b$ and therefore $W^{n+1}\varnothing \vdash b$.

"$\leftarrow$". By induction on n. Let $W(W^n\varnothing) \vdash b$, i.e., $W \vdash (V, b)$ with $V := W^n\varnothing =: \{\, b_i \mid i \in I \,\}$. Then $W^n\varnothing \vdash b_i$, hence by induction hypothesis $(W, b_i) \in [\![Y]\!]$. Substituting W for $\hat{W}$ and all W_i in the derivation above gives the claim $(W, b) \in [\![Y]\!]$. □

Corollary 2.3. *The fixed point operator Y has the property*

$$b \in [\![Y]\!]w \leftrightarrow \exists_k (b \in w^{k+1}\varnothing). \tag{2.2}$$

Proof. Since $w^{k+1}\varnothing$ for fixed k is continuous in w, from $b \in w^{k+1}\varnothing$ we can infer $W^{k+1}\varnothing \vdash b$ for some $W \subseteq w$, and conversely. Moreover $b \in [\![Y]\!]w$ is equivalent to $(W, b) \in [\![Y]\!]$ for some $W \subseteq w$, by (A). Now apply the proposition. □

2.2 Total functionals

We now single out the total continuous functionals from the partial ones. Our main goal will be the density theorem, which says that every finite functional can be extended to a total one.

The *total* ideals x of type ρ (notation $x \in G_\rho$) and the equivalence relation $x_1 \approx x_2$ between them are defined inductively.

(a) For an algebra ι, the total ideals x are those of the form $\mathrm{C}\vec{z}$ with C a constructor of ι and $\vec{z}$ total (C denotes the continuous function $|r_{\mathrm{C}}|$). Two total ideals x_1, x_2 are equivalent (written $x_1 \approx_\iota x_2$) if both are of the form $\mathrm{C}\vec{z_i}$ with the same constructor C of ι, and $z_{1j} \approx_\iota z_{2j}$ for all j.

(b) An ideal r of type $\rho \to \sigma$ is total if and only if for all total z of type ρ, the result $|r|z$ of applying r to z is total. For $f, g \in G_{\rho\to\sigma}$ define $f \approx_{\rho\to\sigma} g$ by $\forall_{x\in G_\rho}(fx \approx_\sigma gx)$.

We show that $x \approx_\rho y$ implies $fx \approx_\sigma fy$, following Longo and Moggi [9].

Lemma 2.4 (Extension)**.** *If $f \in G_\rho$, $g \in |\boldsymbol{C}_\rho|$ and $f \subseteq g$, then $g \in G_\rho$.*

Proof. By induction on ρ. For base types ι use induction on the definition of $f \in G_\iota$. *Case* $\rho \to \sigma$. Assume $f \in G_{\rho\to\sigma}$ and $f \subseteq g$. We show $g \in G_{\rho\to\sigma}$. So let $x \in G_\rho$. We show $gx \in G_\sigma$. But $gx \supseteq fx \in G_\sigma$, so the claim follows by the induction hypothesis. □

Lemma 2.5. *$(f_1 \cap f_2)x = f_1x \cap f_2x$, for $f_1, f_2 \in |\boldsymbol{C}_{\rho\to\sigma}|$ and $x \in |\boldsymbol{C}_\rho|$.*

Proof. By the definition of $|r|$,

$$\begin{aligned}
&|f_1 \cap f_2|x \\
&= \{\, b \in \mathrm{Tok}_\sigma \mid \exists_{U\subseteq x}((U,b) \in f_1 \cap f_2) \,\} \\
&= \{\, b \in \mathrm{Tok}_\sigma \mid \exists_{U_1\subseteq x}((U_1,b) \in f_1) \,\} \cap \{\, b \in \mathrm{Tok}_\sigma \mid \exists_{U_2\subseteq x}((U_2,b) \in f_2) \,\} \\
&= |f_1|x \cap |f_2|x.
\end{aligned}$$

The part "$\subseteq$" of the middle equality is obvious. For "$\supseteq$", let $U_i \subseteq x$ with $(U_i, b) \in f_i$ be given. Choose $U = U_1 \cup U_2$. Then clearly $(U,b) \in f_i$ (as $\{(U_i,b)\} \vdash (U,b)$ and f_i is deductively closed). □

Lemma 2.6. *$f \approx_\rho g$ if and only if $f \cap g \in G_\rho$, for $f, g \in G_\rho$.*

Proof. By induction on ρ. For ι use induction on the definitions of $f \approx_\iota g$ and G_ι. *Case* $\rho \to \sigma$.

$$\begin{aligned}
f \approx_{\rho\to\sigma} g &\leftrightarrow \forall_{x\in G_\rho}(fx \approx_\sigma gx) \\
&\leftrightarrow \forall_{x\in G_\rho}(fx \cap gx \in G_\sigma) \quad \text{by induction hypothesis} \\
&\leftrightarrow \forall_{x\in G_\rho}((f \cap g)x \in G_\sigma) \quad \text{by the last lemma} \\
&\leftrightarrow f \cap g \in G_{\rho\to\sigma}.
\end{aligned}$$

□

Theorem 2.7. *$x \approx_\rho y$ implies $fx \approx_\sigma fy$, for $x, y \in G_\rho$ and $f \in G_{\rho\to\sigma}$.*

Proof. Since $x \approx_\rho y$ we have $x \cap y \in G_\rho$ by the previous lemma. Now $fx, fy \supseteq f(x \cap y)$ and hence $fx \cap fy \in G_\sigma$. But this implies $fx \approx_\sigma fy$ again by the previous lemma. □

We prove the density theorem, which says that every finitely generated functional (i.e., every $\overline{U}$ with $U \in \mathrm{Con}_\rho$) can be extended to a total one. A type ρ is called *dense* if

$$\forall_{U \in \mathrm{Con}_\rho} \exists_{x \in G_\rho} (U \subseteq x)$$

(i.e., $G_\rho \subseteq |\boldsymbol{C}_\rho|$ is dense w.r.t. the Scott topology), and *separating* if

$$\forall_{U,V \in \mathsf{Con}_\rho} (U \not\uparrow_\rho V \to \exists_{\vec{z} \in G} (U\vec{z} \not\uparrow_\iota V\vec{z})).$$

We prove that every type ρ is both dense and separating. Define the *height* $|a^*|$ of an extended token a^*, and $|U|$ of a formal neighborhood U, by

$$\begin{aligned}
&|\mathrm{C}a_1^* \dots a_n^*| := \max\{\, |a_i^*| \mid i = 1, \dots, n \,\} + 1, \qquad |*| := 0, \\
&|(U, b)| := \max\{|U|, |b|\} + 1, \\
&|\{\, a_i \mid i \in I \,\}| := \max\{\, |a_i| + 1 \mid i \in I \,\}.
\end{aligned}$$

Remark 2.8. Let $U \in \mathrm{Con}_\iota$ be non-empty. Then every token in U starts with the same constructor C. Let U_i consist of all tokens at the i-th argument position of some token in U. Then $\mathrm{C}\vec{U} \vdash U$ (and also $U \vdash \mathrm{C}\vec{U}$), and $|U_i| < |U|$ (where $\mathrm{C}\vec{U} := \{\, \mathrm{C}\vec{a^*} \mid a_i^* \in U_i$ if $U_i \neq \varnothing$, and $a_i^* = *$ otherwise $\}$).

We write $G_\iota a$ to mean that a is a total token (i.e., a constructor tree without $*$), and $G_\iota U$ to mean that U contains a total token. For $W = \{\, (U_i, a_i) \mid i < n \,\}$ we have $Wx := \{\, a_i \mid U_i \subseteq x \,\}$. Hence if x is decidable, then so is Wx.

Theorem 2.9 (Density)**.** *For every type $\rho = \rho_1 \to \dots \to \rho_p \to \iota$ we have decidable formulas TExt_ρ and Sep_ρ^i ($i = 1, \dots, p$) such that*

(a) $\forall_{U \in \mathrm{Con}_\rho} (U \subseteq \{\, a \mid \mathrm{TExt}_\rho(U, a) \,\} \in G_\rho)$ *and*

(b) $\forall_{U,V \in \mathrm{Con}_\rho} (U \not\uparrow_\rho V \to \vec{z}_{U,V} \in G \wedge U\vec{z}_{U,V} \not\uparrow_\iota V\vec{z}_{U,V})$, *where* $\vec{z}_{U,V} = z_{U,V,1}, \dots, z_{U,V,p}$ *and* $z_{U,V,i} = \{\, a \mid \mathrm{Sep}_\rho^i(U, V, a) \,\}$.

Proof. By induction on ρ.

Case ι, (a). Given $U \in \mathrm{Con}_\iota$ we define a token a_U by induction on the height $|U|$ such that $\{a_U\} \vdash U$ and $G_\iota a_U$. For $U = \varnothing$ let a_U be the nullary constructor of ι. If $U \neq \varnothing$, define U_i from U as in the remark above; then $\mathrm{C}\vec{U} \vdash U$ and $|U_i| < |U|$. Hence for $a_U := Ca_{U_1} \dots a_{U_n}$ we have $G_\iota a_U$ by induction hypothesis, and

$\{a_U\} \vdash C\vec{U} \vdash U$ by the definition of entailment. So we can put $\mathrm{TExt}_\iota(U, a) := (\{a_U\} \vdash a)$.

Case ι, (b). There is nothing to show.

Case $\rho \to \sigma$, (a). Fix $W = \{ (U_i, a_i) \mid i < n \} \in \mathrm{Con}_{\rho\to\sigma}$. Consider $i < j < n$ with $a_i \not\uparrow a_j$, thus $U_i \not\uparrow U_j$. By induction hypothesis (b) for ρ we have $\vec{z}_{ij} \in G$ such that $U_i\vec{z}_{ij} \not\uparrow_\iota U_j\vec{z}_{ij}$. Define for every $U \in \mathrm{Con}_\rho$ a set I_U of indices $k < n$ such that "U behaves as U_k with respect to the $\vec{z}_{ij}$":

$$I_U := \{ k < n \mid \forall_{i<k}(a_i \not\uparrow a_k \to U\vec{z}_{ik} \vdash_\iota U_k\vec{z}_{ik}) \ \land \\ \forall_{j>k}(a_k \not\uparrow a_j \to U\vec{z}_{kj} \vdash_\iota U_k\vec{z}_{kj}) \}.$$

Notice that $k \in I_{U_k}$. We first show

$$V_U := \{ a_k \mid k \in I_U \} \in \mathrm{Con}_\sigma.$$

It suffices to prove $a_i \uparrow a_j$ for $i, j \in I_U$ with $i < j$. Since $a_i \uparrow a_j$ is decidable we can argue indirectly. Assume $a_i \not\uparrow a_j$. Then $U\vec{z}_{ij} \vdash_\iota U_j\vec{z}_{ij}$ and $U\vec{z}_{ij} \vdash_\iota U_i\vec{z}_{ij}$, thus $U_i\vec{z}_{ij} \uparrow_\iota U_j\vec{z}_{ij}$. But $U_i\vec{z}_{ij} \not\uparrow_\iota U_j\vec{z}_{ij}$ by the choice of the $\vec{z}_{ij}$ for $U_i \not\uparrow U_j$.

By induction hypothesis (a) $V_U \subseteq y_{V_U} := \{ a \mid \mathrm{TExt}_\sigma(V_U, a) \} \in G_\sigma$. Let

$$r := \{ (U, a) \mid (a \in y_{V_U} \ \land \ \forall_{i,j<n}(a_i \not\uparrow a_j \to G_\iota(U\vec{z}_{ij}))) \ \lor \ V_U \vdash a \}, \quad (2.3)$$

We claim $W \subseteq r \in G_{\rho\to\sigma}$; then we can define $\mathrm{TExt}_{\rho\to\sigma}(W, (U, a))$ to be the defining formula of r. Since $k \in I_{U_k}$ we have $a_k \in V_{U_k}$, thus $(U_k, a_k) \in r$. For $r \in |\boldsymbol{C}_{\rho\to\sigma}|$ we verify the properties of approximable maps.

First we show that $(U, a) \in r$ and $(U, b) \in r$ imply $a \uparrow b$. But from the premises we obtain $a, b \in y_{V_U}$ and hence $a \uparrow b$.

Next we show that $(U, b_1), \dots, (U, b_n) \in r$ and $\{b_1, \dots, b_n\} \vdash b$ imply $(U, b) \in r$. We argue by cases. If the left hand side of the disjunction in (2.3) holds for one b_k, then $\{b_1, \dots, b_n\} \subseteq y_{V_U}$, hence $b \in y_{V_U}$ and thus $(U, b) \in r$. Otherwise $V_U \vdash \{b_1, \dots, b_n\} \vdash b$ and therefore $(U, b) \in r$ as well.

Finally we show that $(U, a) \in r$ and $U' \vdash U$ imply $(U', a) \in r$. We again argue by cases. If the left hand side of the disjunction in (2.3) holds, we have $a \in y_{V_U}$, and from $U' \vdash U$ we obtain $\forall_{i,j<n}(a_i \not\uparrow a_j \to G_\iota(U'\vec{z}_{ij}))$. We show $a \in y_{V_{U'}}$. But $U\vec{z}_{ij}$ and $U'\vec{z}_{ij}$ both contain a total token, for every i, j with $a_i \not\uparrow a_j$, which must be the same since $U' \vdash U$. Thus $I_U = I_{U'}$, hence $V_U = V_{U'}$. Now assume $V_U \vdash a$. But $U' \vdash U$ implies $I_U \subseteq I_{U'}$, hence $V_U \subseteq V_{U'}$, hence $V_{U'} \vdash a$ and therefore $(U', a) \in r$.

It remains to prove $r \in G_{\rho\to\sigma}$. Let $x \in G_\rho$. We show that $rx \in G_\sigma$, i.e.,

$$\{ a \in \mathrm{Tok}_\sigma \mid \exists_{U\subseteq x}((U, a) \in r) \} \in G_\sigma.$$

Recall $\vec{z}_{ij} \in G$ for all $i < j < n$ with $a_i \not\uparrow a_j$. Hence $x\vec{z}_{ij} \in G_\iota$ for all such i, j. Since every total ideal of base type contains a total token we have $U_{ij} \subseteq x$ with $G_\iota(U_{ij}\vec{z}_{ij})$. Let U be the union of all U_{ij}'s. Then $G_\iota(U\vec{z}_{ij})$. Hence $(U, a) \in r$ for all $a \in y_{V_U}$, i.e., $y_{V_U} \subseteq rx$ and therefore $rx \in G_\sigma$, by the Extension Lemma.

Case $\rho \to \sigma$, (b). Let $W_1, W_2 \in \mathrm{Con}_{\rho\to\sigma}$ with $W_1 \not\uparrow W_2$. Pick $(U_i, a_i) \in W_i$ such that $U_1 \uparrow U_2$ and $a_1 \not\uparrow a_2$. By induction hypothesis (a) for ρ

$$U_1 \cup U_2 \subseteq z_{U_1,U_2} := \{\, a \mid \mathrm{TExt}_\rho(U_1 \cup U_2, a)\,\} \in G_\rho.$$

Then $a_i \in W_i z_{U_1,U_2}$. From the induction hypothesis (b) for σ we obtain $\vec{z}_{a_1,a_2} \in G$ such that

$$\{a_1\}\vec{z}_{a_1,a_2} \not\uparrow_\iota \{a_2\}\vec{z}_{a_1,a_2},$$

where $\sigma = \sigma_1 \to \ldots \to \sigma_p \to \iota$ and $z_{a_1,a_2,i} := \{\, a \mid \mathrm{Sep}^i_\sigma(\{a_1\}, \{a_2\}, a)\,\}$ for $i = 1, \ldots, p$. Hence $W_1 z_{U_1,U_2}\vec{z}_{a_1,a_2} \not\uparrow_\iota W_2 z_{U_1,U_2}\vec{z}_{a_1,a_2}$. Therefore

$$\mathrm{Sep}^1_{\rho\to\sigma}(W_1, W_2, a) := \mathrm{TExt}_\rho(U_1 \cup U_2, a),$$
$$\mathrm{Sep}^{i+1}_{\rho\to\sigma}(W_1, W_2, a) := \mathrm{Sep}^i_\sigma(\{a_1\}, \{a_2\}, a). \qquad \square$$

2.3 Definability

There will be two kinds of (natural) numbers: (i) total tokens in the algebra $\mathbf{N}$, and (ii) total ideals of type $\mathbf{N}$. Recall that the total tokens in $\mathbf{N}$ are iterated applications of the successor constructor S to the zero constructor 0. We call them *index numbers* and write $n \in \mathbb{N}$ for the n-th such token. Then $\overline{n}$ is a total ideal of type $\mathbf{N}$.

In the statement of the definability theorem below we will need fixed enumerations $(e_n)_{n\in\mathbb{N}}$ of all tokens and $(E_n)_{n\in\mathbb{N}}$ of all formal neighborhoods, one for each type. We will also need some special computable functionals:

The parallel conditional $\mathrm{pcond}\colon \mathbf{B} \to \rho \to \rho \to \rho$

It is defined by the clauses

$$U \vdash \mathtt{tt} \to V \vdash a \to (U, V, W, a) \in \mathrm{pcond}, \tag{2.4}$$
$$U \vdash \mathtt{ff} \to W \vdash a \to (U, V, W, a) \in \mathrm{pcond}, \tag{2.5}$$
$$V \vdash a \to W \vdash a \to (U, V, W, a) \in \mathrm{pcond}. \tag{2.6}$$

We also need the least-fixed-point axiom, which says that any set of tokens (U, V, W, a) satisfying (2.4)–(2.6) is a superset of pcond. It is easy to see that pcond is an ideal.

Lemma 2.10 (Properties of pcond)**.**

$$\mathtt{tt} \in z \to \mathrm{pcond}(z, x, y) = x, \tag{2.7}$$

$$\mathtt{ff} \in z \to \mathrm{pcond}(z, x, y) = y, \tag{2.8}$$

$$a \in x \to a \in y \to a \in \mathrm{pcond}(z, x, y). \tag{2.9}$$

Proof. (2.7). Assume $\mathtt{tt} \in z$. "$\supseteq$". Let $a \in x$. We show $a \in \mathrm{pcond}(z, x, y)$. It suffices to find $U \subseteq z$, $V \subseteq x$ and $W \subseteq y$ such that $(U, V, W, a) \in \mathrm{pcond}$. Since $(\{\mathtt{tt}\}, \{a\}, \varnothing, a) \in \mathrm{pcond}$ by (2.4) we can take $\{\mathtt{tt}\}$ for U, $\{a\}$ for V and $\varnothing$ for W. "$\subseteq$". Let $a \in \mathrm{pcond}(z, x, y)$. We show $a \in x$. By continuity of application we have $U \subseteq z$, $V \subseteq x$ and $W \subseteq y$ such that $(U, V, W, a) \in \mathrm{pcond}$. It suffices to show $V \vdash a$. This will follow from the rules for pcond, since (because of $\mathtt{tt} \in z$) the token (U, V, W, a) must have entered pcond by clause (2.4) or (2.6). Formally we make use of the least-fixed-point axiom for pcond, and apply it to $C := \{ (U, V, W, a) \mid \{\mathtt{tt}\} \vdash U \to V \vdash a \}$. We show that C satisfies (2.4)–(2.6). For (2.5) we must show

$$U \vdash \mathtt{ff} \to W \vdash a \to \{\mathtt{tt}\} \vdash U \to V \vdash a.$$

This follows from ex-falso-quodlibet, since $\{\mathtt{tt}\} \vdash U$ and $U \vdash \mathtt{ff}$ implies $\{\mathtt{tt}\} \vdash \mathtt{ff}$, a contradiction. (2.4) and (2.6) have the desired conclusion $V \vdash a$ among their premises. But now the least-fixed-point axiom for pcond implies $(U, V, W, a) \in C$ (since $\mathtt{tt} \in z$ and $U \subseteq z$ imply $\{\mathtt{tt}\} \vdash U$) and hence $V \vdash a$.

(2.8) is proved similarly. (2.9). It suffices to have $V \subseteq x$ and $W \subseteq y$ such that $(\varnothing, V, W, a) \in \mathrm{pcond}$. Use (2.6) with $\{a\}$ for V and W. □

A continuous variant of the union for $\mathbf{N}$

For ideals in the algebra $\mathbf{N}$, the union (i.e., essentially the maximum) is not a continuous function. However, there is a continuous variant $\cup_{\#}$, which refers in its second argument to the fixed enumeration of the tokens of type $\mathbf{N}$. The type of $\cup_{\#}$ is $\mathbf{N} \to \mathbf{N} \to \mathbf{N}$, and its defining clauses are

$$U \vdash e_n \to V \vdash n \to U \vdash a \to (U, V, a) \in \cup_{\#}, \tag{2.10}$$

$$\{e_n\} \vdash a \to V \vdash n \to (U, V, a) \in \cup_{\#}, \tag{2.11}$$

and again we require the least-fixed-point axiom. It is easy to see that $\cup_{\#}$ is an ideal.

Lemma (Properties of $\cup_{\#}$)**.**

$$\forall_{a \in x}(a \uparrow e_n) \to x \cup_{\#} \overline{n} = x \cup \overline{\{e_n\}}, \tag{2.12}$$

$$e_n \in x \cup_{\#} \overline{n}. \tag{2.13}$$

Proof. (2.12). Assume $a \uparrow e_n$ for all $a \in x$.

"$\supseteq$". Let $a \in x \cup \overline{\{e_n\}}$. We show $a \in x \cup_{\#} \overline{n}$. It suffices to find $U \subseteq x$, $V \subseteq \overline{n}$ such that $(U, V, a) \in \cup_{\#}$. By the Comparability Lemma either $a \vdash \{e_n\}$ or $\{e_n\} \vdash a$. In the first case take $U = \{a\}$, and in the second $U = \varnothing$. Then $(U, \{n\}, a) \in \cup_{\#}$ by (2.10) or (2.11), respectively.

"$\subseteq$". Let $a \in x \cup_{\#} \overline{n}$. We show $a \in x \cup \overline{\{e_n\}}$. By continuity of application we have $U \subseteq x$ and $V \subseteq \overline{n}$ such that $(U, V, a) \in \cup_{\#}$. Let

$$C := (U, V, a) U \vdash a \ \vee\ \exists_{k \in \mathbb{N}}(\{e_k\} \vdash a \ \wedge\ V \vdash k).$$

C satisfies (2.10) and (2.11). Hence by the least-fixed-point axiom for $\cup_{\#}$ we have $(U, V, a) \in C$. If $U \vdash a$ the claim is immediate, since $U \subseteq x$. Otherwise we have $k \in \mathbb{N}$ such that $\{e_k\} \vdash a$ and $V \vdash k$. But $V \subseteq \overline{n}$ implies $k = n$. Hence $\{e_n\} \vdash a$ and therefore $a \in \overline{\{e_n\}}$.

(2.13). Assume $n \in \mathbb{N}$. It suffices to have $U \subseteq x$ and $V \subseteq \overline{n}$ such that $(U, V, e_n) \in \cup_{\#}$. Use (2.11) with e_n for a, $\varnothing$ for U and $\{n\}$ for V. □

A continuous variant of consistency

We define $\uparrow_{\#}$ of type $\rho \to \mathbf{N} \to \mathbf{B}$ by the clauses

$$U \vdash E_n \to V \vdash n \to (U, V, \mathtt{tt}) \in \uparrow_{\#}, \tag{2.14}$$

$$a \in U \to b \in E_n \to V \vdash n \to a \not\uparrow b \to (U, V, \mathtt{ff}) \in \uparrow_{\#}. \tag{2.15}$$

Again we require the least-fixed-point axiom; it is easy to see that $\uparrow_{\#}$ is an ideal.

Lemma 2.11 (Properties of $\uparrow_{\#}$).

$$\mathtt{tt} \in x \uparrow_{\#} \overline{n} \leftrightarrow x \supseteq E_n, \tag{2.16}$$

$$\mathtt{ff} \in x \uparrow_{\#} \overline{n} \leftrightarrow \exists_{a \in x, b \in E_n}(a \not\uparrow b). \tag{2.17}$$

Proof. (2.16). Let $n \in \mathbb{N}$. "$\to$". Assume $\mathtt{tt} \in x \uparrow_{\#} \overline{n}$. We show $x \supseteq E_n$. By continuity of application we have $U \subseteq x$ and $V \subseteq \overline{n}$ such that $(U, V, \mathtt{tt}) \in \uparrow_{\#}$. Let C be the predicate consisting of all (U, V, c) such that

$$(c = \mathtt{tt} \to \exists_{k \in \mathbb{N}}(U \vdash E_k \ \wedge\ V \vdash k)) \ \wedge$$
$$(c = \mathtt{ff} \to \exists_{a \in U, k \in \mathbb{N}, b \in E_k}(V \vdash k \ \wedge\ a \not\uparrow b)).$$

C satisfies (2.14) and (2.15). Hence by the least-fixed-point axiom for $\uparrow_{\#}$ we have $(U, V, \mathtt{tt}) \in C$, i.e., $k \in \mathbb{N}$ such that $U \vdash E_k$ and $V \vdash k$. Using $V \subseteq \overline{n}$ we obtain $k = n$. Now $U \subseteq x$ implies $x \supseteq E_n$.

"←". Assume $x \supseteq E_n$. We show $\mathrm{tt} \in x \uparrow_\# \overline{n}$. It suffices to find $U \subseteq x$ and $V \subseteq \overline{n}$ such that $(U, V, \mathrm{tt}) \in \uparrow_\#$. Take E_n for U and $\{n\}$ for V. Then $(U, V, \mathrm{tt}) \in \uparrow_\#$ by (2.14).

(2.17) is proved similarly. For "→" we can use the same C, and for "←" use (2.15) instead of (2.14). □

Let ι have at most unary constructors, i.e., be one of $\mathbf{N}$, $\mathbf{B}$ or $\mathbf{P}$. A partial continuous functional Φ of type $\rho_1 \to \cdots \to \rho_p \to \iota$ is *recursive in* pcond, $\cup_\#$ *and* $\uparrow_\#$ if it can be defined explicitly by a term involving the constructors for ι and $\mathbf{N}$, the constants predecessor, the fixed point operators Y_ρ, the parallel conditional pcond and the continuous variants of union and of consistency.

Theorem 2.12 (Definability). *A partial continuous functional is computable if and only if it is recursive in* pcond, $\cup_\#$ *and* $\uparrow_\#$.

Proof. The fact that the constants are defined by the rules above implies that the ideals they denote are recursively enumerable. Hence every functional recursive in pcond, $\cup_\#$ and $\uparrow_\#$ is computable. For the converse let Φ be computable of type $\rho_1 \to \cdots \to \rho_p \to \iota$. Then Φ is a primitive recursively enumerated set of tokens $(E_{f_1 n}, \ldots, E_{f_p n}, e_{gn})$ where $f_1, \ldots, f_p$ and g are fixed primitive recursive functions on index numbers. Let $\overline{f}$ denote a continuous extension of f to ideals, such that $\overline{fn} = \overline{f}\overline{n}$. Such an $\overline{f}$ is obtained by reading f's primitive recursion equations as computation rules in the sense of 2.1.

Let $\vec{\varphi} = \varphi_1, \ldots, \varphi_p$ be arbitrary continuous functionals of types $\rho_1, \ldots, \rho_p$, respectively. We show that Φ is definable by the equation $\Phi\vec{\varphi} = Y w_{\vec{\varphi}} \overline{0}$ with $w_{\vec{\varphi}}$ of type $(\mathbf{N} \to \iota) \to \mathbf{N} \to \iota$ given by

$$w_{\vec{\varphi}} \psi x := \mathrm{pcond}(\varphi_1 \uparrow_\# \overline{f_1} x \ \wedge \ \ldots \ \wedge \ \varphi_p \uparrow_\# \overline{f_p} x, \psi(x+1) \cup_\# \overline{g} x, \psi(x+1)).$$

Here $\wedge$ is the *parallel and* of type $\mathbf{B} \to \mathbf{B} \to \mathbf{B}$, defined by $\wedge\ (p, q) := \mathrm{pcond}(p, q, \{\mathrm{ff}\})$. To simplify notation we assume $p = 1$ in the argument to follow, and write w for w_φ. For later reference we split the rest of the argument into steps.

Step 1

We first prove that

$$\forall_n (a \in w^{k+1} \varnothing \overline{n} \to \exists_{n \le l \le n+k} (\varphi \supseteq E_{fl} \ \wedge \ \{e_{gl}\} \vdash a)). \tag{2.18}$$

The proof is by induction on k. For the base case assume $a \in w \varnothing \overline{n}$, i.e.,

$$a \in \mathrm{pcond}(\varphi \uparrow_\# \overline{fn}, \varnothing \cup_\# \overline{gn}, \varnothing).$$

Then clearly $\varphi \supseteq E_{fn}$ and $\{e_{gn}\} \vdash a$.

Step 2

For the step $k \mapsto k+1$ we have

$$a \in w^{k+2}\varnothing\overline{n} = w(w^{k+1}\varnothing)\overline{n} = \mathrm{pcond}(\varphi \uparrow_{\#} \overline{fn}, v \cup_{\#} \overline{gn}, v),$$

with $v := w^{k+1}\varnothing(\overline{n}+1)$. Then either $a \in v$ (and we are done by the induction hypothesis) or else $\varphi \supseteq E_{fn}$ and $\{e_{gn}\} \vdash a$.

Step 3

Now $\Phi\varphi \supseteq Yw\overline{0}$ follows easily. Assume $a \in Yw\overline{0}$. Then $a \in w^{k+1}\varnothing\overline{0}$ for some k, by (2.2). Therefore there is an l with $0 \le l \le k$ such that $\varphi \supseteq E_{fl}$ and $\{e_{gl}\} \vdash a$. But this implies $a \in \Phi\varphi$.

Step 4

For the converse assume $a \in \Phi\varphi$. Then for some $U \subseteq \varphi$ we have $(U, a) \in \Phi$. By our assumption on Φ this means that we have an n such that $U = E_{fn}$ and $a = e_{gn}$. We show

$$a \in w^{k+1}\varnothing(\overline{n-k}) \quad \text{for } k \le n.$$

The proof is by induction on k. For $k = 0$ because of $\varphi \supseteq E_{fn}$ we have $\mathtt{tt} \in \varphi \uparrow_{\#} \overline{fn}$ and hence $w\psi\overline{n} = \psi(\overline{n}+1) \cup_{\#} \overline{gn} \ni e_{gn} = a$, for any ψ.

Step 5

For the step $k \mapsto k+1$ by definition of w $(:= w_{\varphi})$

$$\begin{aligned} v' &:= w^{k+2}\varnothing(\overline{n-k-1}) \\ &= w(w^{k+1}\varnothing)(\overline{n-k-1}) \\ &= \mathrm{pcond}(\varphi \uparrow_{\#} \overline{f(n-k-1)}, v \cup_{\#} \overline{g(n-k-1)}, v) \end{aligned}$$

with $v := w^{k+1}\varnothing(\overline{n-k})$. By induction hypothesis $a \in v$; we show $a \in v'$. If a and $e_{g(n-k-1)}$ are inconsistent, $a \in \Phi\varphi$ and $(E_{f(n-k-1)}, e_{g(n-k-1)}) \in \Phi$ imply that $\varphi \cup E_{f(n-k-1)}$ is inconsistent, hence $\mathtt{ff} \in \varphi \uparrow_{\#} \overline{f(n-k-1)}$ and therefore $v' = v$.

Step 6

If a and $e_{g(n-k-1)}$ are consistent, a and $e_{g(n-k-1)}$ are comparable, since our underlying algebra ι has at most unary constructors.

Step 7

In case $\{e_{g(n-k-1)}\} \vdash a$ we have $v \cup_{\#} \overline{g(n-k-1)} \supseteq \{e_{g(n-k-1)}\} \vdash a$, and hence $a \in v'$ because of $a \in v$.

Step 8

In case $\{a\} \vdash e_{g(n-k-1)}$ we have $e_{g(n-k-1)} \in v$ because of $a \in v$, hence $v \cup_{\#} \overline{g(n-k-1)} = v$ and therefore again $a \in v'$.

Step 9

Now the converse inclusion $\Phi\varphi \subseteq Yw_\varphi\overline{0}$ can be seen easily. Since $a \in \Phi\varphi$, the claim just proved for $k := n$ gives $a \in w_\varphi^{n+1}\varnothing\overline{0}$, and this implies $a \in Yw_\varphi\overline{0}$. $\square$

3 The Theory TCF^+

We sketch a formal system TCF^+ intended to talk about computable functionals *plus* their finite approximations, i.e., tokens and formal neighborhoods. Since continuous functionals (i.e., ideals) are possibly infinite sets of tokens, TCF^+ contains for every type ρ set variables x^ρ. The only existence axiom for sets will be Σ-comprehension.

3.1 Types and token types

Recall that (object) types are built from base types ι (the algebras above) by $\rho \to \sigma$. Now in addition for every (object) type ρ we have *token types* Tok^*_ρ (extended tokens of type ρ), Tok_ρ (tokens of type ρ), LTok_ρ (lists of tokens of type ρ), LTok^*_ρ (lists of extended tokens of type ρ); let τ range over token types. The index ρ will be omitted if it is inessential or clear from the context.

We inductively define the extended tokens of an algebra ι. As a generic algebra we take the algebra $\mathbf{D}$ (of derivations), given by the constructors $0^{\mathbf{D}}$ (axiom) and $\mathrm{C}^{\mathbf{D}\to\mathbf{D}\to\mathbf{D}}$ (rule); for other algebras the definitions are similar. The clauses are

$$\mathrm{Tok}^*_{\mathbf{D}}(*), \quad \mathrm{Tok}^*_{\mathbf{D}}(0^{\mathbf{D}}), \quad \mathrm{Tok}^*_{\mathbf{D}}(a_1^*) \to \mathrm{Tok}^*_{\mathbf{D}}(a_2^*) \to \mathrm{Tok}^*_{\mathbf{D}}(\mathrm{C}^{\mathbf{D}\to\mathbf{D}\to\mathbf{D}}a_1^*a_2^*).$$

(Proper) tokens are defined similarly:

$$\mathrm{Tok}_{\mathbf{D}}(0^{\mathbf{D}}), \quad \mathrm{Tok}^*_{\mathbf{D}}(a_1^*) \to \mathrm{Tok}^*_{\mathbf{D}}(a_2^*) \to \mathrm{Tok}_{\mathbf{D}}(\mathrm{C}^{\mathbf{D}\to\mathbf{D}\to\mathbf{D}}a_1^*a_2^*).$$

Clearly every token can be viewed as an extended token.

It will be convenient to represent formal neighborhoods as lists of tokens. The algebra of lists of tokens of type $\mathbf{D}$ is defined by

$$\mathrm{LTok}_{\mathbf{D}}(\mathrm{nil}_{\mathbf{D}}), \quad \mathrm{Tok}_{\mathbf{D}}(a) \to \mathrm{LTok}_{\mathbf{D}}(U) \to \mathrm{LTok}_{\mathbf{D}}(a ::_{\mathbf{D}} U).$$

We use $\mathrm{nil}_{\mathbf{D}}$ to denote the empty list, and $a ::_{\mathbf{D}} U$ (or $\mathrm{cons}_{\mathbf{D}}(a, U)$) to denote the result of constructing a new list from a given one U by adding a in front. Similarly the algebra of lists of extended tokens is defined by

$$\mathrm{LTok}^*_{\mathbf{D}}(\mathrm{nil}_{\mathbf{D}}), \quad \mathrm{Tok}^*_{\mathbf{D}}(a) \to \mathrm{LTok}^*_{\mathbf{D}}(U) \to \mathrm{LTok}^*_{\mathbf{D}}(a ::_{\mathbf{D}} U).$$

We allow functions of *token-valued types* $\vec{\tau} \to \tau$, defined by primitive recursion. An easy example is $\dot{\in}_{\mathbf{D}} \colon \mathrm{Tok}^*_{\mathbf{D}} \to \mathrm{LTok}^*_{\mathbf{D}} \to \mathrm{Tok}_{\mathbf{B}}$; it is a boolean-valued function, i.e., with values in $\mathrm{Tok}_{\mathbf{B}}$. The recursion equations are

$$\begin{aligned} &(a^* \mathrel{\dot{\in}_{\mathbf{D}}} \mathrm{nil}) := \mathsf{ff}, \\ &(a^* \mathrel{\dot{\in}_{\mathbf{D}}} (b^* ::_{\mathbf{D}} U)) := (a^* =_{\mathbf{D}} b^*) \ \vee_{\mathbf{B}} a^* \mathrel{\dot{\in}} U, \end{aligned}$$

where equality $=_{\mathbf{D}} \colon \mathrm{Tok}^*_{\mathbf{D}} \to \mathrm{Tok}^*_{\mathbf{D}} \to \mathrm{Tok}_{\mathbf{B}}$ is defined by

$$\begin{aligned} &(* =_{\mathbf{D}} *) := (0 =_{\mathbf{D}} 0) := \mathsf{tt}, \\ &(* =_{\mathbf{D}} 0) := (* =_{\mathbf{D}} \mathrm{C}a_1^* a_2^*) := \mathsf{ff}, \\ &(0 =_{\mathbf{D}} *) := (0 =_{\mathbf{D}} \mathrm{C}a_1^* a_2^*) := \mathsf{ff}, \\ &(\mathrm{C}a_1^* a_2^* =_{\mathbf{D}} *) := (\mathrm{C}a_1^* a_2^* =_{\mathbf{D}} 0) := \mathsf{ff}, \\ &(\mathrm{C}a_1^* a_2^* =_{\mathbf{D}} \mathrm{C}b_1^* b_2^*) := (a_1^* =_{\mathbf{D}} b_1^*) \ \wedge_{\mathbf{B}} (a_2^* =_{\mathbf{D}} b_2^*), \end{aligned}$$

and $\vee_{\mathbf{B}} \colon \mathrm{Tok}_{\mathbf{B}} \to \mathrm{Tok}_{\mathbf{B}} \to \mathrm{Tok}_{\mathbf{B}}$ is the disjunction function on $\mathrm{Tok}_{\mathbf{B}}$, defined by $\mathsf{tt} \ \vee_{\mathbf{B}} b := \mathsf{tt}$ and $\mathsf{ff} \ \vee_{\mathbf{B}} b := b$.

From a list of extended tokens of $\mathbf{D}$ we obtain a list of (proper) tokens by removing the $*$'s. Define $\mathrm{clean} \colon \mathrm{LTok}^*_{\mathbf{D}} \to \mathrm{LTok}_{\mathbf{D}}$ by

$$\begin{aligned} &\mathrm{clean}(\mathrm{nil}) := \mathrm{nil}, && \mathrm{clean}(0 :: U) := 0 :: \mathrm{clean}(U), \\ &\mathrm{clean}(* :: U) := \mathrm{clean}(U), && \mathrm{clean}(\mathrm{C}a_1^* a_2^* :: U) := \mathrm{C}a_1^* a_2^* :: \mathrm{clean}(U). \end{aligned}$$

We define $\mathrm{args}_{\mathrm{C},i} \colon \mathrm{LTok}_{\mathbf{D}} \to \mathrm{LTok}^*_{\mathbf{D}}$ ($i = 1, 2$), which from a list of tokens of $\mathbf{D}$ constructs the list of the i-th arguments of C-tokens:

$$\begin{aligned} &\mathrm{args}_{\mathrm{C},i}(\mathrm{nil}) := \mathrm{nil}, \\ &\mathrm{args}_{\mathrm{C},i}(0 :: U) := \mathrm{args}_{\mathrm{C},i}(U), \\ &\mathrm{args}_{\mathrm{C},i}(\mathrm{C}a_1^* a_2^* :: U) := a_i^* :: \mathrm{args}_{\mathrm{C},i}(U). \end{aligned}$$

Now we can define entailment $\vdash$: $\mathrm{LTok}_{\mathbf{D}} \to \mathrm{Tok}^*_{\mathbf{D}} \to \mathrm{Tok}_{\mathbf{B}}$:

$$\begin{aligned} &U \vdash * := \mathsf{tt}, && 0 :: U \vdash \mathrm{C}b_1^* b_2^* := U \vdash \mathrm{C}b_1^* b_2^*, \\ &\mathrm{nil} \vdash 0 := \mathsf{ff}, && \mathrm{C}a_1^* a_2^* :: U \vdash 0 := U \vdash 0, \\ &\mathrm{nil} \vdash \mathrm{C}a_1^* a_2^* := \mathsf{ff}, && 0 :: U \vdash 0 := \mathsf{tt}, \end{aligned}$$

and

$$\begin{aligned} \mathrm{C}a_1^* a_2^* :: U \vdash \mathrm{C}b_1^* b_2^* := {} & \mathrm{clean}(a_1^* :: \mathrm{args}_{\mathrm{C},1}(U)) \vdash b_1^* \ \wedge_{\mathbf{B}} \\ & \mathrm{clean}(a_2^* :: \mathrm{args}_{\mathrm{C},2}(U)) \vdash b_2^*, \end{aligned}$$

where $\wedge_{\mathbf{B}}$: $\mathrm{Tok}_{\mathbf{B}} \to \mathrm{Tok}_{\mathbf{B}} \to \mathrm{Tok}_{\mathbf{B}}$ is the conjunction function on $\mathrm{Tok}_{\mathbf{B}}$, defined by $\mathsf{ff} \ \wedge_{\mathbf{B}} b := \mathsf{ff}$ and $\mathsf{tt} \ \wedge_{\mathbf{B}} b := b$.

To define consistency for lists of tokens we need an auxiliary function checking the outermost constructor only. Let $\mathrm{PreCon}\colon \mathrm{LTok}_{\mathbf{D}} \to \mathrm{Tok}_{\mathbf{B}}$ be defined by

$$\begin{aligned} &\mathrm{PreCon}(\mathrm{nil}) := \mathrm{PreCon}(a :: \mathrm{nil}) := \mathsf{tt}, \\ &\mathrm{PreCon}(0 :: \mathrm{C}a_1^* a_2^* :: U) := \mathrm{PreCon}(\mathrm{C}a_1^* a_2^* :: 0 :: U) := \mathsf{ff}, \\ &\mathrm{PreCon}(0 :: 0 :: U) := \mathrm{PreCon}(0 :: U), \\ &\mathrm{PreCon}(\mathrm{C}a_1^* a_2^* :: \mathrm{C}b_1^* b_2^* :: U) := \mathrm{PreCon}(\mathrm{C}b_1^* b_2^* :: U). \end{aligned}$$

Using PreCon we can define consistency $\mathrm{Con}\colon \mathrm{LTok}_{\mathbf{D}} \to \mathrm{Tok}_{\mathbf{B}}$ by

$$\begin{aligned} &\mathrm{Con}(\mathrm{nil}) := \mathrm{Con}(a :: \mathrm{nil}) := \mathsf{tt}, \\ &\mathrm{Con}(0 :: \mathrm{C}a_1^* a_2^* :: U) := \mathrm{Con}(\mathrm{C}a_1^* a_2^* :: 0 :: U) := \mathsf{ff}, \\ &\mathrm{Con}(0 :: 0 :: U) := \mathrm{Con}(0 :: U), \end{aligned}$$

and

$$\begin{aligned} \mathrm{Con}(\mathrm{C}a_1^* a_2^* :: \mathrm{C}b_1^* b_2^* :: U) := {} & \mathrm{PreCon}(\mathrm{C}b_1^* b_2^* :: U) \ \wedge_{\mathbf{B}} \\ & \mathrm{Con}(\mathrm{clean}(a_1^* :: b_1^* :: \mathrm{args}_{\mathrm{C},1}(U))) \ \wedge_{\mathbf{B}} \\ & \mathrm{Con}(\mathrm{clean}(a_2^* :: b_2^* :: \mathrm{args}_{\mathrm{C},2}(U))). \end{aligned}$$

We write $a^* \uparrow_\rho b^*$ for $\mathrm{Con}(a^* ::_\rho b^* ::_\rho \mathrm{nil})$.

We define $G_{\mathbf{D}}\colon \mathrm{Tok}^*_{\mathbf{D}} \to \mathrm{Tok}_{\mathbf{B}}$ expressing totality for extended tokens:

$$G_{\mathbf{D}}(*) := \mathsf{ff}, \qquad G_{\mathbf{D}}(0) := \mathsf{tt}, \qquad G_{\mathbf{D}}(\mathrm{C}a_1^* a_2^*) := G_{\mathbf{D}} a_1^* \ \wedge_{\mathbf{B}} G_{\mathbf{D}} a_2^*,$$

and also $G_{\mathrm{LTok}_{\mathbf{D}}}\colon \mathrm{LTok}_{\mathbf{D}} \to \mathrm{Tok}_{\mathbf{B}}$ doing the same for lists of tokens

$$G_{\mathrm{LTok}_{\mathbf{D}}}(\mathrm{nil}_{\mathbf{D}}) := \mathsf{ff}, \qquad G_{\mathrm{LTok}_{\mathbf{D}}}(a ::_{\mathbf{D}} U) := G_{\mathbf{D}} a \ \vee_{\mathbf{B}} G_{\mathrm{LTok}_{\mathbf{D}}} U.$$

Recall that total tokens of **N** are iterated applications of the successor constructor S to the zero constructor 0. They are called "index numbers", and written $n \in \mathbb{N}$. Since primitive recursion is available to define token-valued functions, we can construct standard auxiliary functions, like sequence coding. Thus every (index) number n can be written uniquely as $n = \langle a_0, a_1, \ldots, a_{k-1} \rangle$, and $k = \mathrm{lh}(n)$, $a_i = (n)_i$ for $i < k$.

Tokens of a function type $\rho \to \sigma$ are pairs (U, a) of lists of tokens of type ρ and tokens of type σ. Both projections are given by functions π_1, π_2. Consistency of lists of tokens, application WU and entailment $U \vdash a$ can be defined as described in 1.2.

3.2 Enumerations

We assume fixed enumerations $(e_n)_{n\in\mathbb{N}}$ of tokens and $(E_n)_{n\in\mathbb{N}}$ of lists of tokens, for each type. It seems easiest to define them explicitly. Fix for every constructor C of an algebra a unique "symbol number" $\mathrm{SN}(\mathrm{C})$. We also have a symbol number $\mathrm{SN}(\mathrm{Nhd})$ indicating the code of a formal neighborhood. We define a Gödel numbering $\ulcorner\cdot\urcorner \colon \mathrm{Tok}^*_{\mathbf{D}} \to \mathbb{N}$ by

$$\begin{aligned}
&\ulcorner * \urcorner := 0,\\
&\ulcorner 0 \urcorner := \langle \mathrm{SN}(0) \rangle,\\
&\ulcorner \mathrm{C} a_1^* a_2^* \urcorner := \langle \mathrm{SN}(\mathrm{C}), \ulcorner a_1^* \urcorner, \ulcorner a_2^* \urcorner \rangle.
\end{aligned}$$

Formal neighborhoods are gödelized by $\ulcorner\cdot\urcorner \colon \mathrm{LTok}_\rho \to \mathbb{N}$,

$$\ulcorner a_0 :: a_1 :: \ldots a_{k-1} :: \mathrm{nil} \urcorner := \langle \mathrm{SN}(\mathrm{Nhd}), \ulcorner \rho \urcorner, \ulcorner a_0 \urcorner, \ulcorner a_1 \urcorner, \ldots, \ulcorner a_{k-1} \urcorner \rangle,$$

where $\ulcorner \iota \urcorner := \langle \mathrm{SN}(\iota) \rangle$, $\ulcorner \rho \to \sigma \urcorner := \langle \mathrm{SN}(\to), \ulcorner \rho \urcorner, \ulcorner \sigma \urcorner \rangle$. It is clear that we can primitive recursively define the converse, mapping the Gödel number $\ulcorner a^* \urcorner$ of an extended token back to a^*, i.e., $e_{\ulcorner a^* \urcorner} = a^*$, and similarly for LTok_ρ.

3.3 Terms and formulas

We have variables a^* for Tok^*_ρ (extended tokens of type ρ), a for Tok_ρ (tokens of type ρ) and U for LTok_ρ (lists of tokens of type ρ). From these, the symbols for token-valued functions and constants for the constructors for tokens, extended tokens and lists of these we can build terms of token types. We identify terms of token type if they have the same normal form w.r.t. the defining primitive recursion equations for the token-valued functions involved.

Decidable (or Δ-) *prime formulas* are of the form $\mathrm{atom}(p)$, with p a term of token type $\mathrm{Tok}_{\mathbf{B}}$. They are decidable in the sense that for each such term p we can prove $p = \mathtt{tt} \lor p = \mathtt{ff}$; in fact, every closed term of type $\mathrm{Tok}_{\mathbf{B}}$ can be evaluated to either $\mathtt{tt}$ and $\mathtt{ff}$. Examples are $a \uparrow_\rho b$, $a \mathrel{\dot\in}_\rho U$, $U \vdash_\rho a$ (which are shorthand for $\mathrm{atom}(a \uparrow_\rho b)$, $\mathrm{atom}(a \mathrel{\dot\in}_\rho U)$, $\mathrm{atom}(U \vdash_\rho a)$). Δ*-formulas* are built from decidable prime formulas by $\to$, $\land$, $\lor$ and *bounded quantifiers*, i.e., $\forall_{a \dot\in U}$, $\exists_{a \dot\in U}$, with a a variable for tokens and U a term for a list of tokens.

In TCF^+ we also allow variables and constants of (object) type ρ, intended to denote sets of tokens. The constants are $[\![\lambda_{\vec{x}} M]\!]$ (with M a term as in 2.1) of type $\vec{\rho} \to \sigma$, and also pcond, $\cup_\#$, $\uparrow_\#$ of types $\mathbf{B} \to \rho \to \rho \to \rho$, $\rho \to \mathbf{N} \to \rho$ and $\rho \to \mathbf{N} \to \mathbf{B}$, respectively.

Prime Σ*-formulas* are either decidable prime formulas or else of the form $r \in_\rho x$, with r a term of token type Tok_ρ and x a variable or constant of type ρ. Σ*-formulas* are built as follows.

(a) Every prime Σ-formula is a Σ-formula.

(b) $A_0 \to B$ is a Σ-formula if A_0 is a Δ-formula and B a Σ-formula.

(c) Σ-formulas are closed under $\land$, $\lor$, bounded quantifiers and existential quantifiers over variables of a token type.

Prime formulas are either prime Σ-formulas or else of the form $G_\rho x$ (expressing totality of x) or $x \approx_\rho y$ (expressing equivalence of x and y); x, y are variables or constants of type ρ. *Formulas* are built from prime formulas by $\to$, $\land$, $\lor$, $\forall$, $\exists$, where the quantifiers are w.r.t all kinds of variables.

3.4 Axioms

TCF^+ is based on intuitionistic logic. In fact, minimal logic suffices, since falsity can be defined as $\mathrm{atom}(\mathtt{ff})$. Then $\mathrm{atom}(\mathtt{ff}) \to A$ ("ex-falso-quodlibet") can be proved provided one has it as an axiom for every prime formula (it can be proved for decidable prime formulas).

Therefore the axioms of TCF^+ are ex-falso-quodlibet for non-decidable prime formulas A, plus the usual ones of Heyting arithmetic, adapted to token types. In particular we have the ordinary induction schemes, for arbitrary formulas of the language. Examples are

$$A(\mathtt{tt}) \to A(\mathtt{ff}) \to A(a),$$
$$A(*) \to A(0) \to \forall_{a^*,b^*}(A(a^*) \to A(b^*) \to A(\mathrm{C}a^*b^*)) \to A(a^*).$$

Moreover $\mathrm{atom}(\mathtt{tt})$ is an axiom. For object types we have Σ*-comprehension*:

$$\exists_x \forall_a (a \in_\rho x \leftrightarrow A), \quad \text{for every } \Sigma\text{-formula } A.$$

A convenient notation for x is $\{\, a \mid A \,\}$. Further axioms are

(a) For every constant $[\![\lambda_{\vec{x}} M]\!]$ its defining clauses corresponding to the rules (V), (A), (C), (D) from 2.1, together with their least-fixed-point axioms.

(b) The defining clauses and corresponding least-fixed-point axioms, for pcond, $\cup_{\#}$ and $\uparrow_{\#}$, as listed in 2.3.

(c) The clauses from 2.2 defining the totality predicates G_ρ and the equivalence relations $x_1 \approx_\rho x_2$, together with their least-fixed-point axioms.

Notice that the latter imply $x_1 \approx_\rho x_2 \to Gx_1 \to Gx_2$.

3.5 First steps in TCF^+

We use the abbreviations

$$\begin{aligned}
&U \subseteq V \quad \text{for} \quad \forall_{a \dot{\in} U}(a \dot{\in} V),\\
&U \vdash V \quad \text{for} \quad \forall_{a \dot{\in} V}(U \vdash a),\\
&U \sim V \quad \text{for} \quad U \vdash V \ \wedge \ V \vdash U,\\
&a \sim b \quad \text{for} \quad \{a\} \vdash b \ \wedge \ \{b\} \vdash a,\\
&x \subseteq y \quad \text{for} \quad \forall_{a \in x}(a \in y),\\
&x = y \quad \text{for} \quad x \subseteq y \ \wedge \ y \subseteq x,\\
&U \subseteq x \quad \text{for} \quad \forall_{a \dot{\in} U}(a \in x).
\end{aligned}$$

Terms of (object) type are built from variables and constants by application ts and comprehension $\{\, a \mid A \,\}$. Then $r \in_\rho t$ for t a term of type ρ and r a term of token type Tok_ρ is defined by

$$\begin{aligned}
&(r \in_\rho \{\, a \mid A(a) \,\}) := A(r),\\
&(r \in_\rho ts) := \exists_{U \subseteq s}((U, r) \in t) \qquad \text{(continuity of application)}.
\end{aligned}$$

For a term M with free variables among $\vec{x}$ we write

$$a \in_\sigma [\![M]\!] \quad \text{for} \quad \exists_{\vec{U} \subseteq \vec{x}}((\vec{U}, a) \in_{\vec{\rho} \to \sigma} [\![\lambda_{\vec{x}} M]\!]).$$

We can prove Δ-comprehension for lists of tokens

$$\exists_U \forall_a (a \dot{\in} U \leftrightarrow a \dot{\in} V \ \wedge \ A), \quad \text{for every } \Delta\text{-formula } A,$$

by induction on V. A convenient notation for U is $[\, a \dot{\in} V \mid A \,]$.

We will need the *extension* $\overline{f}$ of a monotone token-valued function f to ideals. It suffices to do this for $f\colon \mathrm{Tok}^*_{\mathbf{N}} \to \mathrm{Tok}^*_{\mathbf{N}}$. Suppose f is *monotone*, i.e., $\{a^*\} \vdash b^*$ implies $\{fa^*\} \vdash fb^*$. Define $f[\cdot]\colon \mathrm{LTok}^*_{\mathbf{N}} \to \mathrm{LTok}^*_{\mathbf{N}}$ by

$$f[\mathrm{nil}] := \mathrm{nil}, \qquad f[a^* ::_{\mathbf{N}} U] := (fa^*) ::_{\mathbf{N}} f[U].$$

Then $\overline{f}\colon \mathbf{N} \to \mathbf{N}$ is defined by

$$\overline{f} = \{\, (U, a) \mid \mathrm{Con}(U) \ \wedge \ f[U] \vdash a \,\}.$$

Clearly $\overline{f}$ is a decidable ideal. If $f\colon \mathrm{Tok}_{\mathbf{N}} \to \mathrm{Tok}_{\mathbf{N}}$ is defined primitive recursively, then by reading f's primitive recursion equations as computation rules we obtain a defined constant $\overline{f}$ (in the sense of 2.1) such that $\overline{fn} = \overline{f}\overline{n}$.

Notice that $\forall_{i<n} A$ with i a variable and n a term of token type $\mathrm{Tok}_{\mathbf{N}}$ can be viewed as bounded quantification. Define $h\colon \mathrm{Tok}^*_{\mathbf{N}} \to \mathrm{LTok}^*_{\mathbf{N}}$ by

$$h(*) := h(0) := \mathrm{nil}, \qquad h(\mathrm{S}\, a^*) := h(a^*) * (a^* :: \mathrm{nil}),$$

where $*$ appends two lists from $\mathrm{LTok}^*_{\mathbf{N}}$. Then $h(\mathrm{S}^k\, 0) = [0, \mathrm{S}\,0, \dots, \mathrm{S}^{k-1}\,0]$ (i.e., $0 :: \mathrm{S}\,0 :: \dots \mathrm{S}^{k-1}\,0 :: \mathrm{nil}$), and we can read $\forall_{i<n} A$ as $\forall_{i \dot\in h(n)} A$.

Every W of token type $\mathrm{LTok}_{\rho\to\sigma}$ can be written as $\{\, (U_i, a_i) \mid i < n \,\}$. Here U_i, a_i are given as $f(W, i)$, $g(W, i)$ and n as the length $\mathrm{lh}(W)$ of W, with f, g and $\mathrm{lh}(\cdot)$ defined primitive recursively. Define

$$(a \mathrel{\dot\in} Wx) := \exists_{i<n} (U_i \subseteq x \ \wedge \ a = a_i).$$

Then $a \mathrel{\dot\in} Wx$ is a Δ-formula if x is given by $\{\, a \mid A \,\}$ with A a Δ-formula. Therefore by Δ-comprehension for list of tokens we obtain U consisting of all a_i's such that $a_i \mathrel{\dot\in} Wx$. Hence $Wx \vdash a$ can be seen as a Δ-formula as well.

4 Formalization

4.1 Density

The informal proof already was written in a form making its formalization in TCF^+ easy. We only discuss the more interesting issues.

The density theorem is parametrized by the type ρ, and its proof (by induction on ρ) is to be viewed as employing a "meta"-induction.

In the proof that $\rho \to \sigma$ is dense we fixed $W = \{\, (U_i, a_i) \mid i < n \,\} \in \mathrm{Con}_{\rho\to\sigma}$. Consider $i < j < n$ with $a_i \not\asymp a_j$, thus $U_i \not\asymp U_j$. The induction hypothesis (b) for ρ gives $\vec{z}_{ij} \in G$ such that $U_i \vec{z}_{ij} \not\asymp_\iota U_j \vec{z}_{ij}$. The definition of

$$V_U := [\, a_k \mid k \in I_U \,]$$

can be seen as an application of Δ-comprehension for lists of tokens, since $k \in I_U$ is a Δ-formula. Now the induction hypothesis that σ is dense yields $V_U \subseteq y_{V_U} := \{ a \mid \mathrm{TExt}_\sigma(V_U, a) \} \in G_\sigma$. The definition (2.3) of

$$r := \{ (U, a) \mid (a \in y_{V_U} \ \wedge \ \forall_{i,j<n}(a_i \not\uparrow a_j \to G_\iota(U\vec{z}_{ij}))) \ \vee \ V_U \vdash a \},$$

is by Σ-comprehension; in fact, the defining formula is a Δ-formula. The rest of the argument can be easily formalized.

The proof that $\rho \to \sigma$ is separating does not present any difficulties. We are given $W_1, W_2 \in \mathrm{Con}_{\rho\to\sigma}$ with $W_1 \not\uparrow W_2$, and pick $(U_i, a_i) \in W_i$ such that $U_1 \uparrow U_2$ and $a_1 \not\uparrow a_2$. Notice that the U_i, a_i can be defined primitive recursively from W_1, W_2, and hence are uniquely determined. By induction hypothesis (a) for ρ,

$$U_1 \cup U_2 \subseteq z_{U_1,U_2} := \{ a \mid \mathrm{TExt}_\rho(U_1 \cup U_2, a) \} \in G_\rho.$$

Then $a_i \mathrel{\dot\in} W_i z_{U_1,U_2}$. From the induction hypothesis (b) for σ we obtain $\vec{z}_{a_1,a_2} \in G$ such that (writing $\{a_i\}$ for $[a_i]$)

$$\{a_1\}\vec{z}_{a_1,a_2} \not\uparrow_\iota \{a_2\}\vec{z}_{a_1,a_2},$$

where $\sigma = \sigma_1 \to \ldots \to \sigma_p \to \iota$ and $z_{a_1,a_2,i} := \{ a \mid \mathrm{Sep}^i_\sigma(\{a_1\}, \{a_2\}, a) \}$ for $i = 1, \ldots, p$. Hence $W_1 z_{U_1,U_2}\vec{z}_{a_1,a_2} \not\uparrow_\iota W_2 z_{U_1,U_2}\vec{z}_{a_1,a_2}$.

4.2 Definability

We restrict ourselves to the more interesting direction and assume that Φ is given as a primitive recursively enumerated set of tokens (E_{fn}, e_{gn}) where f, g are fixed primitive recursive functions. We need to show that Φ is recursive in pcond, $\cup_\#$ and $\uparrow_\#$, i.e., that it can be defined explicitly by a term involving the constructors for ι and $\mathbf{N}$, the constants predecessor, the fixed point operators Y_ρ, the parallel conditional pcond and the continuous variants of union and of consistency. In doing so we follow the steps in the informal proof in 2.3. We show that Φ is definable by the equation $\Phi\varphi = Y w_\varphi \overline{0}$, with w_φ of type $(\mathbf{N} \to \iota) \to \mathbf{N} \to \iota$ given by

$$w_\varphi \psi x := \mathrm{pcond}(\varphi \uparrow_\# \overline{f}x, \psi(x+1) \cup_\# \overline{g}x, \psi(x+1)).$$

In Step 1 by continuity of application we obtain $U \subseteq \varphi \uparrow_\# \overline{fn}$ and $V \subseteq \varnothing \cup_\# \overline{gn}$ such that $(U, V, \varnothing, a) \in \mathrm{pcond}$. For $\varphi \supseteq E_{fn}$ it suffices by (2.16) to prove $\mathtt{tt} \in \varphi \uparrow_\# \overline{fn}$, which because of $U \subseteq \varphi \uparrow_\# \overline{fn}$ follows from $U \vdash \mathtt{tt}$. This will follow from the rules for pcond, because (since W is $\varnothing$) the token $(U, V, \varnothing, a)$ must have entered pcond by rule (2.4). Formally we make use of the least-fixed-point axiom for pcond, and apply it to $C := \{ (U, V, W, a) \mid W \subseteq \varnothing \to U \vdash \mathtt{tt} \}$. We show that C satisfies (2.4)–(2.6). For (2.5) we must show

$$U \vdash \mathtt{ff} \to W \vdash a \to (U, V, W, a) \in C, \quad \text{i.e.,}$$
$$U \vdash \mathtt{ff} \to W \vdash a \to W \subseteq \varnothing \to U \vdash \mathtt{tt}.$$

But this follows from ex-falso-quodlibet, since $W \vdash a$ and $W \subseteq \varnothing$ are contradictory. (2.6) is proved similarly, and (2.4) has the desired conclusion $U \vdash \mathtt{tt}$ among its premises. But now the least-fixed-point axiom for pcond implies $(U, V, \varnothing, a) \in C$ and hence $U \vdash \mathtt{tt}$. For $\{e_{gn}\} \vdash a$ we argue similarly, with $C := \{ (U, V, W, a) \mid W \subseteq \varnothing \to V \vdash a \}$, and obtain $V \vdash a$ and hence $a \in \varnothing \cup_{\#} \overline{gn}$. By (2.12) we conclude that $\{e_{gn}\} \vdash a$.

The next part of the informal proof was Step 2. Again by continuity of application we obtain $U \subseteq \varphi \uparrow_{\#} \overline{fn}$, $V \subseteq v \cup_{\#} \overline{gn}$ and $W \subseteq v$ such that $(U, V, W, a) \in \mathrm{pcond}$. We can prove $W \vdash a \vee (U \vdash \mathtt{tt} \wedge V \vdash a)$ as above from the rules for pcond. Hence either $a \in v$ (and we are done by the induction hypothesis), or else $\varphi \supseteq E_{fn}$ (which follows as above from $U \vdash \mathtt{tt}$) and $a \in v \cup_{\#} \overline{gn}$. From the latter by continuity of application we obtain $V \subseteq v$ and $W \subseteq \overline{gn}$ such that $(V, W, a) \in \cup_{\#}$. By a least-fixed-point argument (with $C := \{ (V, W, a) \mid \exists_m (m \mathrel{\dot\in} W \wedge \{e_m\} \vdash a) \vee V \vdash a \}$) we obtain either $V \vdash a$ (hence $a \in v$ and again we are done by the induction hypothesis), or else $\{e_m\} \vdash a$ for an $m \in G$ such that $m \mathrel{\dot\in} W$, hence $m = gn$, and therefore $\{e_{gn}\} \vdash a$. Now the induction used in the informal proof can be applied and we have proved (2.18) formally.

The informal proof proceeded by Step 3. Since corollary (2.2) referred to is available in TCF^+, we have proved the conclusion $a \in \Phi\varphi$ formally.

Let us now formalize the proof of the reverse direction. Step 4. In the formalization from $\varphi \supseteq E_{fn}$ we obtain $\mathtt{tt} \in \varphi \uparrow_{\#} \overline{fn}$ by (2.16). We show $a \in w\psi\overline{n}$ for an arbitrary ψ, i.e., $a \in \mathrm{pcond}(\varphi \uparrow_{\#} \overline{fn}, \psi(\overline{n}+1) \cup_{\#} \overline{gn}, \psi(\overline{n}+1))$. Because of $\mathtt{tt} \in \varphi \uparrow_{\#} \overline{fn}$ and (2.7) it is enough to show that $a \in \psi(\overline{n}+1) \cup_{\#} \overline{gn}$. But $e_{gn} \in \psi(\overline{n}+1) \cup_{\#} \overline{gn}$ by (2.13), and we have assumed $a = e_{gn}$.

Next we consider Step 5. Formally we can infer the existence of $b \in \varphi$ and $c \mathrel{\dot\in} E_{f(n-k-1)}$ such that $b \not\uparrow c$. Hence $\mathtt{ff} \in \varphi \uparrow_{\#} \overline{f(n-k-1)}$ by (2.17), and $v' = v$ by (2.8). Step 6 is immediate because of the Comparability Lemma. For Step 7: Here we can infer $a \in v \cup_{\#} \overline{g(n-k-1)}$ from (2.13). This and the induction hypothesis $a \in v$ yields the claim $a \in v'$ by (2.9). For Step 8: $v \cup_{\#} \overline{g(n-k-1)} = v$ follows from $e_{g(n-k-1)} \in v$ by (2.12). Again this and the induction hypothesis $a \in v$ yields the claim $a \in v'$ by (2.9). For Step 9: The final inference is justified by (2.2) (applied to $(\{0\}, a)$).

5 Future work

In this paper we attempted to have a first exploratory view on a constructive formal theory of computability TCF^{+}, where the functionals are studied together with their finite approximations. The attempt was guided by the semantics of non-flat Scott information systems; in particular, it was based on two case studies, namely, the density theorem and the definability theorem. Future work along these lines is to explain TCF^{+} in a rigorous and systematic way, as well as test it against further case studies, while an actual implementation on a theorem prover—which should be specially designed to allow for handling functionals and finite approximations alike—remains the ultimate goal of the whole enterprise.

References

[1] S. Abramsky and A. Jung. Domain theory. In S. Abramsky, D. Gabbay, and T. Maibaum, editors, *Handbook of Logic in Computer Science*, volume 3, pages 1–168. Clarendon Press, 1994.

[2] U. Berger. Total sets and objects in domain theory. *Annals of Pure and Applied Logic*, 60:91–117, 1993.

[3] Y. L. Ershov. Maximal and everywhere defined functionals. *Algebra i Logika*, 13(4):374–397, 1974.

[4] A. Heyting, editor. *Constructivity in Mathematics*. North–Holland, Amsterdam, 1959.

[5] S. Huber. On the computional content of choice axioms. Master's thesis, Mathematisches Institut der Universität München, 2010.

[6] S. C. Kleene. Countable functionals. In Heyting [4], pages 81–100.

[7] G. Kreisel. Interpretation of analysis by means of constructive functionals of finite types. In Heyting [4], pages 101–128.

[8] K. G. Larsen and G. Winskel. Using information systems to solve recursive domain equations. *Information and Computation*, 91:232–258, 1991.

[9] G. Longo and E. Moggi. The hereditary partial effective functionals and recursion theory in higher types. *The Journal of Symbolic Logic*, 49:1319–1332, 1984.

[10] G. D. Plotkin. LCF considered as a programming language. *Theoretical Computer Science*, 5:223–255, 1977.

[11] G. D. Plotkin. $\mathbf{T}^{\omega}$ as a universal domain. *Journal of Computer and System Sciences*, 17:209–236, 1978.

[12] D. Scott. A type-theoretical alternative to ISWIM, CUCH, OWHY. Manuscript, Oxford University. Published as [15], 1969.

[13] D. Scott. Outline of a mathematical theory of computation. Technical Monograph PRG–2, Oxford University Computing Laboratory, 1970.

[14] D. Scott. Domains for denotational semantics. In E. Nielsen and E. Schmidt, editors, *Automata, Languages and Programming*, volume 140 of *LNCS*, pages 577–613. Springer Verlag, Berlin, Heidelberg, New York, 1982.

[15] D. Scott. A type-theoretical alternative to ISWIM, CHUCH, OWHY. *Theoretical Computer Science*, 121:411–440, 1993.

[16] V. Stoltenberg-Hansen, E. Griffor, and I. Lindström. *Mathematical Theory of Domains*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 1994.

Σ^1_1 Choice in a Theory of Sets and Classes

Gerhard Jäger and Jürg Krähenbühl*

Institut für Informatik und angewandte Mathematik, Universität Bern
Neubrückstrasse 10, CH-3012 Bern, Switzerland
{jaeger,kraehenb}@iam.unibe.ch

Abstract Several decades ago Friedman showed that the subsystem Σ^1_1-AC of second order arithmetic is proof-theoretically equivalent – and thus equiconsistent – to $(\Pi^1_0\text{-CA})_{<\varepsilon_0}$. In this article we prove the analogous result for Σ^1_1 choice in the context of the von Neumann-Bernays-Gödel theory NBG of sets and classes.

Keywords: Proof theory, theories of sets and classes.

1 Introduction

Several decades ago Friedman showed that the subsystem Σ^1_1-AC of second order arithmetic is proof-theoretically equivalent – and thus equiconsistent – to $(\Pi^1_0\text{-CA})_{<\varepsilon_0}$ (cf. Friedman [7]). Later Feferman [2, 3], Tait [16], Feferman and Sieg [6] and Cantini [1] reproved and extended this result, always making use of different proof-theoretic techniques.

In this article we start off from the von Neumann-Bernays-Gödel theory NBG of sets and classes, extend it by the schema $(\mathcal{L}_2\text{-}\mathsf{I}_\in)$ of $\in$-induction for arbitrary formulas of the language $\mathcal{L}_2$ of NBG and study the effect of adding Σ^1_1 choice and Σ^1_1 collection,

$$\forall x \exists Y A[x, Y] \rightarrow \exists Z \forall x A[x, (Z)_x], \qquad (\Sigma^1_1\text{-AC})$$

$$\forall x \exists Y A[x, Y] \rightarrow \exists Z \forall x \exists y A[x, (Z)_y], \qquad (\Sigma^1_1\text{-Col})$$

where A is an elementary formula of $\mathcal{L}_2$, i.e. an $\mathcal{L}_2$ formula which does not contain bound class variables. We will show that the resulting theories are equiconsistent to the system $\mathsf{NBG}_{<E_0}$ which is obtained from $\mathsf{NBG} + (\mathcal{L}_2\text{-}\mathsf{I}_\in)$ by adding iterations of elementary comprehension along all initial segments of the notation system $(E_0, \lhd)$. E_0 is an elementarily definable class and $\lhd$ an elementary binary

*Research supported by the Swiss National Science Foundation

class relation on E_0 which, provably in NBG, well-orders all initial segments of E_0. The notation system $(E_0, \lhd)$ may be seen as the analogue of $(\varepsilon_0, <)$ with the ordinal ω replaced by the collection of all ordinals. In this sense, our result is the perfect analogue of Friedman's result mentioned above with natural numbers and sets of natural numbers replaced by sets and classes, respectively.

Our characterization of $\mathsf{NBG} + (\mathcal{L}_2\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{AC})$ is also interesting in connection with Feferman's operational set theory OST, introduced in Feferman [4,5]. As shown in Jäger [11], the extension $\mathsf{OST}(\mathbf{E}, \mathbb{P})$ of OST with unbounded existential quantification and power set is equiconsistent to $\mathsf{NBG}_{<E_0}$ and therefore, in view of the results of this paper, also to the more familiar system $\mathsf{NBG}+(\mathcal{L}_2\text{-}\mathsf{I}_\in)+(\Sigma^1_1\text{-}\mathsf{AC})$. The results of this paper are discussed from a broader perspective in Jäger [12].

The embedding of $\mathsf{NBG}_{<E_0}$ into $\mathsf{NBG} + (\mathcal{L}_2\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{AC})$ is straightforward. The difficult part of this paper is the reduction of $\mathsf{NBG} + (\mathcal{L}_2\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{AC})$ to $\mathsf{NBG}_{<E_0}$, and here an asymmetric interpretation plays a major rôle. Similar forms of asymmetric interpretations have been used, for example, in Cantini [1] to deal with subsystems of second order arithmetic and in Jäger [9–11] and Jäger and Strahm [13] in the context of theories of admissible sets, explicit mathematics and operational set theory.

First we observe that $(\Sigma^1_1\text{-}\mathsf{AC})$ can be replaced by $(\Sigma^1_1\text{-}\mathsf{Col})$. Then, in order to get rid of $(\mathcal{L}_2\text{-}\mathsf{I}_\in)$, we develop (within $\mathsf{NBG}_{<E_0}$) an infinitary sequent style reformulation G^∞ of $\mathsf{NBG} + (\Sigma^1_1\text{-}\mathsf{Col})$ in which constants for all sets are available. By making use of an infinitary rule for universal quantification over sets, we show

$$\mathsf{NBG} + (\mathcal{L}_2\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{Col}) \vdash A \quad \Longrightarrow \quad \mathsf{NBG}_{<E_0} \vdash \text{“}\mathsf{G}^\infty \text{ proves } A\text{”}.$$

A next step is to strengthen this assertion by a partial cut elimination argument for G^∞ to

$$\mathsf{NBG} + (\mathcal{L}_2\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{Col}) \vdash A \quad \Longrightarrow$$
$$\mathsf{NBG}_{<E_0} \vdash \text{“}\mathsf{G}^\infty \text{ proves } A \text{ with simple cuts”}.$$

Now the technical part begins: we have to go back from provability in G^∞ to provability in $\mathsf{NBG}_{<E_0}$. This is achieved in two further steps:

(i) Introduction of a sort of constructible hierarchy of classes and a truth definition based on this hierarchy which reflects all closed elementary formulas A,

$$\mathsf{NBG}_{<E_0} \vdash Tr[A] \leftrightarrow A.$$

(ii) An asymmetric interpretation of a suitable fragment of G^∞ with respect to this hierarchy such that, for all closed elementary formulas A of G^∞,

$$\mathsf{NBG}_{<E_0} \vdash (\text{“}\mathsf{G}^\infty \text{ proves } A \text{ with simple cuts”} \;\rightarrow\; Tr[A]).$$

Altogether, we thus have

$$\mathsf{NBG} + (\mathcal{L}_2\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{Col}) \vdash A \quad\Longrightarrow\quad \mathsf{NBG}_{<E_0} \vdash A$$

for all closed elementary formulas, and this is the required reduction. The definitions of our analogue of the constructible hierarchy and the associated notion of truth – although conceptually clear – require some care since everything has to be carried through within the restricted framework of $\mathsf{NBG}_{<E_0}$.

2 Von Neuman-Bernays-Gödel set theory

The von Neumann-Bernays-Gödel set theory NBG is a theory of sets and classes conservative over the system ZFC of Zermelo-Fraenkel set theory with the axiom of choice. NBG is known to be finitely axiomatizable although the version we are going to present below permits axiom schemas and as such is an infinite axiomatization.

Let $\mathcal{L}_1$ be a typical first order language of set theory with countably many set variables $a, b, c, f, g, u, v, w, x, y, z, \ldots$ and a single symbol for the element relation, but without function or individual constants.

$\mathcal{L}_2$, the language of NBG, augments $\mathcal{L}_1$ by a second sort of countably many variables $U, V, W, X, Y, Z, \ldots$ for classes; its *formulas* $(A, B, C, \ldots)$ are inductively generated as follows:

1. If a, b are set variables and if U is a class variable, then all expressions of the form $(a \in b)$ and $(a \in U)$ are (atomic) formulas of $\mathcal{L}_2$.

2. If A and B are formulas of $\mathcal{L}_2$, then so are are $\neg A$, $(A \vee B)$ and $(A \wedge B)$.

3. If A is a formula of $\mathcal{L}_2$, then $\exists x A$, $\forall x A$, $\exists X A$ and $\forall X A$ are formulas of $\mathcal{L}_2$.

The denotations for set variables, class variables and $\mathcal{L}_2$ formulas may be used with and without subscripts. Since we will be working within classical logic, the remaining logical connectives can be defined as usual.

We will often omit parentheses and brackets whenever there is no danger of confusion. Moreover, we frequently make use of the vector notation $\vec{a}$ as shorthand for a finite string $a_1, \ldots, a_n$ of set variables whose length is not important or evident from the context.

Equalities between sets/sets, sets/classes, classes/sets and classes/classes are not atomic formulas of $\mathcal{L}_2$ but defined as

$$(\mathit{Var}_1 = \mathit{Var}_2) \ := \ \forall x(x \in \mathit{Var}_1 \leftrightarrow x \in \mathit{Var}_2)$$

where Var_1 and Var_2 denote set or class variables. A formula of $\mathcal{L}_2$ is called *elementary* if it does not contain bound class variables; free class variables, however, are permitted. The Σ^1_1 formulas of $\mathcal{L}_2$ are those of the form $\exists X A$ with elementary A. Finally, an $\mathcal{L}_2$ formula A is called Σ^1 if all positive occurrences of class quantifiers are existential and all negative occurrences of class quantifiers are universal; it is called Π^1 if all positive occurrences of class quantifiers are universal and all negative occurrences of class quantifiers are existential. By a closed formula we mean one which does not contain free set or class variables.

The logic of NBG is classical two-sorted logic with equality for the first sort. The non-logical axioms of NBG are given in six groups. To increase readability, we freely use standard set-theoretic terminology.

I. Elementary comprehension For any elementary formula $A[u]$ of $\mathcal{L}_2$ and any class variable X not free in $A[u]$:

$$\exists X \forall y(y \in X \ \leftrightarrow \ A[y]). \tag{ECA}$$

Hence every elementary NBG formula $A[u]$ defines a class, which is typically written as $\{x : A[x]\}$. It may be (extensionally equal to) a set, but this is not necessarily the case.

II. Basic set existence

$$\forall x \forall y \exists z(z = \{x, y\}), \tag{Pair}$$

$$\forall x \exists y(y = \cup x), \tag{Union}$$

$$\forall x \exists y \forall z(z \in y \leftrightarrow z \subset x), \tag{Power set}$$

$$\exists x(\varnothing \in x \ \ \wedge \ \ (\forall y \in x)(y \cup \{y\} \in x)). \tag{Infinity}$$

In the following we write $\langle a, b\rangle$ for the ordered pair of the sets a and b à la Kuratowski. Class relations are classes which consist of ordered pairs only, and class functions are class relations which are right unique; i.e. for all U we set:

$$Rel[U] := (\forall x \in U)\exists y \exists z(x = \langle y, z\rangle),$$

$$Dom[U] := \{x : \exists y(\langle x, y\rangle \in U)\},$$

$$Fun[U] := Rel[U] \;\wedge\; \forall x \forall y \forall z(\langle x, y\rangle \in U \;\wedge\; \langle x, z\rangle \in U \;\rightarrow\; y = z).$$

If U is a function and x an element of $Dom[U]$, we write $U(x)$ for the unique y such that $\langle x, y\rangle \in U$. Replacement states that the range of a set under a function is a set.

III. Replacement For any class variable U:

$$Fun[U] \;\rightarrow\; \forall x \exists y(y = \{U(z) : z \in Dom[U] \cap x\}). \tag{REP}$$

Global choice is a very uniform principle of choice which claims the existence of a class function which picks an element of any non-empty set.

IV. Global choice

$$\exists X(Fun[X] \;\wedge\; Dom[X] = \{y : y \neq \varnothing\} \;\wedge\; \forall y(y \neq \varnothing \;\rightarrow\; X(y) \in y)). \tag{GC}$$

To complete the list of axioms of NBG, we add foundation. In NBG it is claimed that the element relation is well-founded with respect to classes.

V. Class foundation For any class variable U:

$$U \neq \varnothing \;\rightarrow\; (\exists x \in U)(\forall y \in x)(y \notin U). \tag{C-I$_\in$}$$

A set a is called an *ordinal* if a itself and all its elements are transitive, On stands for the class of all ordinals; i.e.

$$On \;:=\; \{x : Tran(x) \;\wedge\; (\forall y \in x)Tran(y)\}.$$

The axioms (Infinity) and (C-I$_\in$) imply that there exists a least infinite ordinal, which we denote by ω, as usual. The elements of ω are identified with the natural numbers in the sense that $0 := \varnothing,\ 1 := \{0\},\ 2 := 1 \cup \{1\}$ and so on. In the following small Greek letters are supposed to range over On.

One important property of NBG is the subset property: the intersection of a set a with a class is a subset of a. Its proof is standard.

There exist various alternative presentations of NBG. So it is an appealing feature of NBG that the schema of elementary comprehension can be replaced by finitely many axioms and thus a finite axiomatization of NBG is possible. Furthermore, according to a well-known result, see, e.g., Levy [14], NBG is a conservative extension of ZFC.

Theorem 2.1. *A sentence of the language $\mathcal{L}_1$ is provable in* NBG *if and only if it is provable in* ZFC.

In the following we will be mainly concerned with extensions of NBG. The first of those consists in adding to NBG the schema of $\in$-induction for arbitrary $\mathcal{L}_2$ formulas $A[u]$,

$$\forall x((\forall y \in x)A[y] \rightarrow A[x]) \rightarrow \forall x A[x]. \tag{$\mathcal{L}_2$-$\mathsf{I}_\in$}$$

Further interesting principles are the schemas of Σ^1_1 choice and Σ^1_1 collection which consists of all formulas

$$\forall x \exists Y A[x, Y] \rightarrow \exists Z \forall x A[x, (Z)_x], \tag{Σ^1_1-AC}$$

$$\forall x \exists Y A[x, Y] \rightarrow \exists Z \forall x \exists y A[x, (Z)_y] \tag{Σ^1_1-Col}$$

where $A[u, V]$ is an elementary $\mathcal{L}_2$ formula and $(Z)_a$ is the class given by

$$(Z)_a := \{x : \langle a, x\rangle \in Z\}.$$

Clearly, every instance of (Σ^1_1-Col) follows from (Σ^1_1-AC). However, in NBG also the converse is the case.

Theorem 2.2. *If $A[u, V]$ is an elementary $\mathcal{L}_2$ formula, then we have*

$$\mathsf{NBG} + (\Sigma^1_1\text{-Col}) \vdash \forall x \exists Y A[x, Y] \rightarrow \exists Z \forall x A[x, (Z)_x].$$

Proof. We work within NBG + (Σ^1_1-Col). Following the pattern of the usual proof of the well-ordering theorem in ZFC and exploiting the fact that we have global choice, it is easy to show that there exist a bijective class function W from On to the collection of all sets. We write W^{-1} for the inverse of W.

Suppose $\forall x \exists Y A[x, Y]$, where $A[u, V]$ is an elementary $\mathcal{L}_2$ formula. Then by (Σ^1_1-Col) there exists a class Z such that

$$\forall x \exists y A[x, (Z)_y]. \tag{$\star$}$$

Now the function W^{-1} comes into play in order to associate to any x a unique y for which $A[x,(Z)_y]$. Namely, by elementary comprehension and $(\star)$

$$Sel \;:=\; \{\langle x,y\rangle : A[x,(Z)_y] \;\wedge\; \forall z(A[x,(Z)_z] \to W^{-1}(y) \leq W^{-1}(z))\}$$

is a class function whose domain is the collection of all sets. Finally, if we write S for the class $\{\langle x,y\rangle : y \in (Z)_{Sel(x)}\}$, which exists by elementary comprehension, we have $(S)_x = (Z)_{Sel(x)}$ for all sets x. Hence $\forall x A[x,(S)_x]$. In other words, S is the required witness for $(\Sigma^1_1\text{-AC})$. □

Corollary 2.3. *The theories* $\mathsf{NBG} + (\Sigma^1_1\text{-AC})$ *and* $\mathsf{NBG} + (\Sigma^1_1\text{-Col})$ *prove the same formulas.*

In this paper we are interested in the consistency strength of the theories $\mathsf{NBG} + (\mathcal{L}_2\text{-I}_\in) + (\Sigma^1_1\text{-AC})$ and $\mathsf{NBG} + (\mathcal{L}_2\text{-I}_\in) + (\Sigma^1_1\text{-Col})$. The much simpler analysis of $\mathsf{NBG} + (\Sigma^1_1\text{-AC})$ and $\mathsf{NBG} + (\Sigma^1_1\text{-Col})$ will be presented elsewhere.

3 The notation system $(E_0, \lhd)$

In this section we work within $\mathsf{NBG} + (\mathcal{L}_2\text{-I}_\in)$ and set up the notation system $(E_0, \lhd)$. The underlying idea is very simple: $(E_0, \lhd)$ is designed to be the analogue of $(\varepsilon_0, <)$ with the set of the natural numbers, i.e. the ordinal ω, replaced by the class of all ordinals. All we have to do is to follow one of the standard introductions of the ordinal notation system up to ε_0 as, for example, in Schütte [15], taking care of the few additional complications arising by the fact that we now have all elements of On as basic entities.

Definition 3.1. By *finite sequences* we mean those functions whose domain is a finite ordinal; FS is defined to be the class of all finite sequences,

$$FS \;:=\; \{f : Fun[f] \;\wedge\; (\exists n < \omega)(Dom[f] = n)\}.$$

If we are given n sets $a_0,\ldots,a_{n-1}$ for some natural number n, we often write $(a_0,\ldots,a_{n-1})$ for that element f of FS which satisfies $Dom[f] = n$ as well as $(\forall i < n)(f(i) = a_i)$.

By elementary comprehension it can be easily shown in NBG that there exists a binary class relation $\lhd$ on FS satisfying the property (I) below. To simplify the formulation of this property, we abbreviate:

$$a \lhd b \;:=\; \langle a,b\rangle \in \lhd \qquad \text{and} \qquad a \unlhd b \;:=\; a \lhd b \;\vee\; a = b.$$

In addition, let $\lhd_{lex}$ be the lexicographic extension of $\lhd$; i.e. if a and b are finite sequences of sets, then $a \lhd_{lex} b$ is written for

$$(Dom[a] < Dom[b] \;\wedge\; (\forall i < Dom[a])(a(i) = b(i)) \;\vee$$
$$(\exists i < Dom[a])(i < Dom[b] \;\wedge\; a(i) \lhd b(i) \;\wedge\; (\forall j < i)(a(j) = b(j)).$$

(I) The binary relation $\lhd$ on FS. For all elements a and b of FS we have $a \lhd b$ if and only if $Dom[a]$ and $Dom[b]$ are at least 2 and one of the following cases holds:

(1) $a(0) = b(0) = 0 \;\wedge\; a(1) < b(1)$,

(2) $a(0) = 0 \;\wedge\; 0 < b(0)$,

(3) $a(0) = 1 \;\wedge\; 2 \leq b(0)$,

(4) $a(0) = b(0) = 2 \;\wedge\; a(1) \lhd b(1)$,

(5) $a(0) = 2 \;\wedge\; b(0) = 3 \;\wedge\; a \unlhd b(1)$,

(6) $a(0) = 3 \;\wedge\; b(0) = 2 \;\wedge\; a(1) \lhd b$,

(7) $a(0) = b(0) = 3 \;\wedge\; a \lhd_{lex} b$.

For the time being, this is a rather weird binary relation on finite sequences. Its real meaning will become transparent when restricted to the subclass E_0 of FS which is introduced in (III) and whose definition is based on $\lhd$.

For every ordinal α we let $\overline{\alpha}$ be the finite sequence $(0, \alpha)$. In addition, Ω is defined to be the finite sequence $(1, 0)$.

(II) The ω-exponentiation of elements of FS. There exists a class function $\widetilde{\omega}$ which is described by $Dom[\widetilde{\omega}] = FS$ and, for all elements a of FS,

$$\widetilde{\omega}(a) \;=\; \begin{cases} \overline{\omega^{\alpha}} & \text{if } \; a = \overline{\alpha} \; \text{ for some ordinal } \alpha, \\ \Omega & \text{if } \; a = \Omega, \\ (2, a) & \text{otherwise.} \end{cases}$$

In the following, the function $\widetilde{\omega}$ will be interesting four us only when restricted to those finite sequences which act as notations. They are collected in the class E_0 which can be defined be elementary comprehension and is characterized as follows.

(III) The class E_0 of notations. E_0 is defined to be the smallest subclass of FS which satisfies the following closure properties:

(1) For all ordinals α we have $\overline{\alpha} \in E_0$.

(2) $\Omega \in E_0$.

(3) If $a \in E_0$, then $\widetilde{\omega}(a) \in E_0$.

(4) If $a_0, \ldots, a_{n+1} \in E_0$ and $\Omega \trianglelefteq a_{n+1} \trianglelefteq \ldots \trianglelefteq a_1 \trianglelefteq a_0$, then

$$(3, \widetilde{\omega}(a_0), \widetilde{\omega}(a_1), \ldots, \widetilde{\omega}(a_{n+1})) \in E_0.$$

(5) If $a_0, \ldots, a_n \in E_0$ and $\Omega \trianglelefteq a_n \trianglelefteq \ldots \trianglelefteq a_1 \trianglelefteq a_0$ and $\alpha \neq 0$, then

$$(3, \widetilde{\omega}(a_0), \widetilde{\omega}(a_1), \ldots, \widetilde{\omega}(a_n), \overline{\alpha}) \in E_0.$$

The elements of E_0 of the form $(0, a)$ code the ordinals, the element $(1, 0) = \Omega$ is the least element greater than the codes of all ordinals, $(2, a)$ codes the ω-exponentiation of a and $(3, a_0, \ldots, a_{n-1})$ is for the sum of ω-powers and possibly the code of an ordinal, given in decreasing order. The proof of the following lemma is without any problems.

Lemma 3.2. NBG *proves that the relation* $\triangleleft$ *is a strict linear ordering on the class* E_0.

In the following we use the small Gothic type letters $\mathfrak{a}, \mathfrak{b}, \ldots$ (possibly with subscripts) for elements of E_0. Expressions like $\exists \mathfrak{a}(\ldots)$ and $\forall \mathfrak{a}(\ldots)$ are then to be read as $(\exists a \in E_0)(\ldots)$ and $(\forall a \in E_0)(\ldots)$, respectively. For simplicity of notation, we also write $\omega^{\mathfrak{a}}$ instead of $\widetilde{\omega}(\mathfrak{a})$.

Definition 3.3. For all positive natural numbers n and all $\mathfrak{a}_0, \ldots, \mathfrak{a}_{n-1} \in E_0$ we set

$$[\mathfrak{a}_0, \ldots, \mathfrak{a}_{n-1}] := \begin{cases} \mathfrak{a}_0 & \text{if } n = 1 \ \wedge \ (\mathfrak{a}_0 \trianglelefteq \Omega \ \vee \ \exists \mathfrak{b}(\mathfrak{a}_0 = \omega^{\mathfrak{b}})), \\ (3, \mathfrak{a}_0, \ldots, \mathfrak{a}_{n-1}) & \text{if } (3, \mathfrak{a}_0, \ldots, \mathfrak{a}_{n-1}) \in E_0. \end{cases}$$

In all other cases $[\mathfrak{a}_0, \ldots, \mathfrak{a}_{n-1}]$ may be taken to be undefined or to have the value $\varnothing$.

So every element $\mathfrak{a}$ of E_0 can be uniquely written as $[\mathfrak{a}_0, \ldots, \mathfrak{a}_{n-1}]$. This representation is useful for a compact description of the addition of ordinal terms. Once more, it can be introduced as a binary class function by elementary comprehension and is characterized by the following properties.

(IV) Addition of elements of E_0. For all $\mathfrak{a}$ and $\mathfrak{b}$ we have:

(1) If $\mathfrak{a} = \overline{0}$, then $\mathfrak{a} + \mathfrak{b} = \mathfrak{b}$, if $\mathfrak{b} = \overline{0}$, then $\mathfrak{a} + \mathfrak{b} = \mathfrak{a}$.

(2) If $\mathfrak{a} = [\mathfrak{a}_0, \dots, \mathfrak{a}_{m-1}, \overline{\alpha}]$ and $\mathfrak{b} = \overline{\beta}$ for some ordinals α and β greater than 0, then

$$\mathfrak{a} + \mathfrak{b} \; = \; [\mathfrak{a}_0, \dots, \mathfrak{a}_{m-1}, \overline{\alpha + \beta}].$$

(3) If $\mathfrak{a} = [\mathfrak{a}_0, \dots, \mathfrak{a}_{m-1}]$ such that $\Omega \trianglelefteq \mathfrak{a}_{m-1}$ and $\mathfrak{b} = \overline{\beta}$ for some ordinal β greater than 0, then

$$\mathfrak{a} + \mathfrak{b} \; = \; [\mathfrak{a}_0, \dots, \mathfrak{a}_{m-1}, \mathfrak{b}].$$

(4) If $\mathfrak{a} = [\mathfrak{a}_0, \dots, \mathfrak{a}_{m-1}]$ and $\mathfrak{b} = [\mathfrak{b}_0, \dots, \mathfrak{b}_{n-1}]$ such that $\Omega \trianglelefteq \mathfrak{b}_0$, then, if k is the largest natural number i for which $\mathfrak{b}_0 \trianglelefteq \mathfrak{a}_i$,

$$\mathfrak{a} + \mathfrak{b} \; = \; [\mathfrak{a}_0, \dots, \mathfrak{a}_k, \mathfrak{b}_0, \dots, \mathfrak{b}_{n-1}].$$

Before turning to the well-ordering of initial parts of E_0, a further class function, describing the finite addition of ω-powers of elements of E_0, has to be introduced.

(V) The function $\widehat{\omega}$ on elements of E_0 and finite numbers. There exists a class function $\widehat{\omega}$ which is described by $Dom[\widehat{\omega}] = E_0 \times \omega$ and, for all $\mathfrak{a}$ and all $n < \omega$,

$$\widehat{\omega}(\mathfrak{a}, n) \; = \; \begin{cases} 0 & \text{if } \; n = 0, \\ \widehat{\omega}(\mathfrak{a}, n-1) + \omega^{\mathfrak{a}} & \text{if } \; 0 < n < \omega. \end{cases}$$

We omit the proof of the following lemma since it is in complete analogy to the case of the notation system for $(\varepsilon_0, <)$.

Lemma 3.4. *The following assertions can be proved in* NBG*:*

1. $(\mathfrak{a} + \mathfrak{b}) + \mathfrak{c} \; = \; \mathfrak{a} + (\mathfrak{b} + \mathfrak{c})$.

2. $\mathfrak{a} \lhd \mathfrak{b} + \omega^{\mathfrak{c}} \;\; \wedge \;\; \overline{0} \lhd \mathfrak{c} \; \rightarrow \; (\exists \mathfrak{d} \lhd \mathfrak{c})(\exists n < \omega)(\mathfrak{a} \lhd \mathfrak{b} + \widehat{\omega}(\mathfrak{d}, n))$.

Starting with $\Omega + 1$ a sequence of terms which is cofinal in E_0 is obtained by simply iterating ω-exponentiation.

Definition 3.5. For all natural numbers n, the ordinal terms Ω_n are inductively defined by

$$\Omega_0 \; := \; \Omega + 1 \qquad \text{and} \qquad \Omega_{n+1} \; := \; \omega^{\Omega_n}.$$

The purpose of the next paragraphs is to show that $\mathsf{NBG} + (\mathcal{L}_2\text{-}\mathsf{I}_\in)$ proves the well-ordering of the relation $\lhd$ on E_0 up to each term Ω_k for k being any standard natural number. To do so, we need the following notations.

Definition 3.6. Let $A[u]$ be an arbitrary formula of the language $\mathcal{L}_2$ of NBG. Then we set:

$$\begin{aligned} \mathit{Prog}_\lhd[A] &:= \forall \mathfrak{u}((\forall \mathfrak{v} \lhd \mathfrak{u})A[\mathfrak{v}] \to A[\mathfrak{u}]), \\ \mathit{TI}_\lhd[\mathfrak{u}, A] &:= \mathit{Prog}_\lhd[A] \to (\forall \mathfrak{v} \lhd \mathfrak{u})A[\mathfrak{v}]. \\ A^*[\mathfrak{u}] &:= \forall \mathfrak{v}((\forall \mathfrak{w} \lhd \mathfrak{v})A[\mathfrak{w}] \to (\forall \mathfrak{w} \lhd \mathfrak{v} + \omega^{\mathfrak{u}})A[\mathfrak{w}]). \end{aligned}$$

The first two of these formulas express, as usual, the progressiveness of A with respect to $\lhd$ and transfinite induction for A along $\lhd$ up to $\mathfrak{u}$, respectively; A^* is the *jump* of A. The core of the well-ordering proofs up to Ω_k, for any standard natural number k, is provided by the following two properties of the jump-operation.

Lemma 3.7. *For any formula $A[u]$ of the language $\mathcal{L}_2$, we can prove in NBG:*

1. $\mathit{Prog}_\lhd[A] \to \mathit{Prog}_\lhd[A^*]$.
2. $\mathit{TI}_\lhd[\mathfrak{u}, A^*] \to \mathit{TI}_\lhd[\omega^{\mathfrak{u}}, A]$.

All our notations are chosen such that the proof of this lemma can be taken literally from the proof of the corresponding lemma for notations less than ε_0 in Schütte [15].

Theorem 3.8. *For any standard natural number k and for any formula $A[u]$ of the language $\mathcal{L}_2$ we have*

$$\mathsf{NBG} + (\mathcal{L}_2\text{-}\mathsf{I}_\in) \vdash \mathit{TI}_\lhd[\Omega_k, A].$$

Proof. We work informally in $\mathsf{NBG} + (\mathcal{L}_2\text{-}\mathsf{I}_\in)$ and prove this theorem by metainduction on k. Assume that $k = 0$. Then $\Omega_k = \Omega + 1$ and $\in$-induction on the ordinals yields, for arbitrary $\mathcal{L}_2$ formulas $A[u]$,

$$\mathit{Prog}_\lhd[A] \to (\forall \mathfrak{u} \lhd \Omega)A[\mathfrak{u}].$$

By the definition of progressiveness, this implies

$$\mathit{Prog}_\lhd[A] \to (\forall \mathfrak{u} \lhd \Omega + 1)A[\mathfrak{u}],$$

i.e. $\mathit{TI}_\lhd[\Omega_0, A]$. For $k > 0$ we have in view of the induction hypothesis for any $\mathcal{L}_2$ formulas $A[u]$ that $\mathit{TI}_\lhd[\Omega_{k-1}, A^*]$. Now we simply have to apply Lemma 3.7 in order to obtain $\mathit{TI}_\lhd[\Omega_k, A]$. □

In connection with the notation system $(E_0, \lhd)$ it only remains to introduce a few further notations which will be taken up again towards the end of Section 5.

Definition 3.9. The classes of limit notations and strong limit notations are defined by

$$Lim := \{x \in E_0 : x \neq \overline{0} \ \wedge \ (\forall y \in E_0)(x \neq y + \overline{1}\},$$

$$SLim := \{x \in Lim : (\forall y \in E_0)(x \neq y + \overline{\omega}\}.$$

In addition, we define $Lim_0 := \{\overline{0}\} \cup Lim$ and $SLim_0 := \{\overline{0}\} \cup SLim$ and, for any $U \subset E_0$ and $\mathfrak{a}, \mathfrak{b} \in E_0$,

$$\mathfrak{a} \in U \cap \mathfrak{b} \ := \ \mathfrak{a} \in U \ \wedge \ \mathfrak{a} \lhd \mathfrak{b}.$$

This means that the elements of Lim are the analogues of limit ordinals and the elements of $SLim$ correspond to those limit ordinals which cannot be obtained by adding ω. Clearly, any Ω_n belongs to $SLim$.

4 Elementary hierarchies

This section begins with introducing the theory $\mathsf{NBG}_{<E_0}$ which permits the iteration of elementary comprehension up to any Ω_k with k a standard natural number. It is easily verified afterwards that $\mathsf{NBG}_{<E_0}$ is contained in the system $\mathsf{NBG} + (\mathcal{L}_2\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{AC})$.

Definition 4.1. Let $A[U, V, u, v]$ be an elementary $\mathcal{L}_2$ formula with at most the variables U, V, u, v free. Then we write $Hier_A[\mathfrak{a}, U, V]$ for the elementary $\mathcal{L}_2$ formula

$$(\forall \mathfrak{b} \lhd \mathfrak{a})((V)_\mathfrak{b} = \{x : A[U, \Sigma(V, \mathfrak{b}), x, \mathfrak{b}]\}).$$

Here $\Sigma(V, \mathfrak{b})$ stands for the class $\{\langle x, \mathfrak{c}\rangle \in V : \mathfrak{c} \lhd \mathfrak{b}\}$ representing the disjoint union of the projections of V up to $\mathfrak{b}$.

$\mathsf{NBG}_{<E_0}$ is the theory of sets and classes which extends $\mathsf{NBG} + (\mathcal{L}_2\text{-}\mathsf{I}_\in)$ by claiming the existence of such hierarchies along each initial segment of E_0. Hence the axioms of $\mathsf{NBG}_{<E_0}$ comprise the axioms of NBG, the schema $(\mathcal{L}_2\text{-}\mathsf{I}_\in)$ plus

$$\forall X \exists Y \, Hier_A[\Omega_n, X, Y] \qquad \text{(It-ECA)}$$

for arbitrary elementary $\mathcal{L}_2$ formulas $A[U, V, u, v]$ with at most the variables U, V, u, v free and all standard natural numbers n.

Employing (Σ^1_1-AC), the following lemma is proved by transfinite induction along $\lhd$ up to Ω_n, which is available in $\mathsf{NBG} + (\mathcal{L}_2\text{-}\mathsf{I}_\in)$ according to Theorem 3.8. The argument is very similar to that of second order arithmetic, establishing that $\Pi^0_1\text{-}\mathsf{CA}_{<\varepsilon_0}$ is a subsystem of Σ^1_1-AC, and left to the reader.

Lemma 4.2. *Let $A[u,v,U,V]$ be an elementary $\mathcal{L}_2$ formula with at most the variables u,v,U,V free. For all standard natural numbers n and all class variables X, the theory $\mathsf{NBG} + (\mathcal{L}_2\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{AC})$ then proves*

$$(\forall \mathfrak{a} \lhd \Omega_n)\exists Y \, \mathit{Hier}_A[\mathfrak{a}, X, Y].$$

From this lemma we conclude that all axioms (It-ECA) are provable in the system $\mathsf{NBG} + (\mathcal{L}_2\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{AC})$. Therefore, the embedding of $\mathsf{NBG}_{<E_0}$ into $\mathsf{NBG} + (\mathcal{L}_2\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{AC})$ is an immediate consequence.

Theorem 4.3. *The theory $\mathsf{NBG}_{<E_0}$ is contained in $\mathsf{NBG} + (\mathcal{L}_2\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{AC})$; i.e. for all $\mathcal{L}_2$ formulas A we have*

$$\mathsf{NBG}_{<E_0} \vdash A \quad \Longrightarrow \quad \mathsf{NBG} + (\mathcal{L}_2\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{AC}) \vdash A.$$

For the reduction of $\mathsf{NBG} + (\mathcal{L}_2\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{Col})$ to $\mathsf{NBG}_{<E_0}$ it is convenient to have a global well-ordering of the set-theoretic universe at our disposal. Therefore, let $\mathcal{L}_\mathcal{W}$ be the extension of $\mathcal{L}_2$ by a fresh binary relation symbol $\mathcal{W}$ and include formulas $\mathcal{W}(u,v)$ into the list of atomic formulas. Then the global well-ordering axiom states

$$\forall x \exists! \alpha \mathcal{W}(x,\alpha) \ \land \ \forall x \forall y \forall \alpha (\mathcal{W}(x,\alpha) \ \land \ \mathcal{W}(y,\alpha) \to x = y). \qquad \text{(GWO)}$$

We write NBGW for the theory NBG – now all schemas formulated for $\mathcal{L}_\mathcal{W}$ formulas – in which the axiom of global choice (GC) has been replaced by the axiom global well-ordering (GWO). Accordingly, $\mathsf{NBGW}_{<E_0}$ is the theory $\mathsf{NBGW} + (\mathcal{L}_\mathcal{W}\text{-}\mathsf{I}_\in)$ extended by the iteration axiom (It-ECA), now formulated for all elementary $\mathcal{L}_\mathcal{W}$ formulas.

It goes without saying that NBG and $\mathsf{NBG}_{<E_0}$ are contained in NBGW and $\mathsf{NBGW}_{<E_0}$, respectively. Moreover, with little effort and by making use of standard techniques it can even be shown that we have the following theorem.

Theorem 4.4. NBGW *is a conservative extension of* NBG*, and* $\mathsf{NBGW}_{<E_0}$ *is a conservative extension of* $\mathsf{NBG}_{<E_0}$*, in both cases with respect to all $\mathcal{L}_2$ formulas.*

5 Reducing NBG + ($\mathcal{L}_2$-$\mathsf{I}_\in$) + (Σ^1_1-AC) to $\mathsf{NBG}_{<E_0}$

The eventual aim of this section is to show that $\mathsf{NBG} + (\mathcal{L}_2\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{AC})$ can be reduced to $\mathsf{NBG}_{<E_0}$. In order to achieve this it is sufficient – in view of what we have achieved so far – to reduce the theory $\mathsf{NBGW} + (\mathcal{L}_\mathcal{W}\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{Col})$ to $\mathsf{NBGW}_{<E_0}$, where in this context $(\Sigma^1_1\text{-}\mathsf{Col})$ is for $\mathcal{L}_\mathcal{W}$ formulas.

In the following we develop, within $\mathsf{NBGW}_{<E_0}$, an infinitary sequent calculus G^∞ for $\mathsf{NBGW} + (\mathcal{L}_\mathcal{W}\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{Col})$. For this purpose we code the set variables as pairs $\langle 0, n\rangle$ and the class variables as pairs $\langle 1, n\rangle$, n always a natural number. Moreover, to any set a we assign the set constant $\langle 2, a\rangle$. For natural numbers n and sets a we set

$$h_n := \langle 0, n\rangle, \qquad H_n := \langle 1, n\rangle, \qquad p_a := \langle 2, a\rangle.$$

We also fix several elementary class functions defined, for arbitrary sets a, b, c, by (some are written in infix or another mnemonically suitable notation):

$$\begin{array}{rclcrcl} (a \mathrel{\dot\in} b) & := & \langle 3, a, b\rangle, & & \dot{\mathcal{W}}(a, b) & := & \langle 4, a, b\rangle, \\ \dot\neg\, a & := & \langle 5, a\rangle, & & (a \mathrel{\dot\vee} b) & := & \langle 6, a, b\rangle, \\ (a \mathrel{\dot\wedge} b) & := & \langle 7, a, b\rangle, & & \dot\exists\, a\, b & := & \langle 8, a, b\rangle, \\ \dot\forall\, a\, b & := & \langle 9, a, b\rangle. & & & & \end{array}$$

We proceed with our development of G^∞ within $\mathsf{NBGW}_{<E_0}$ and present all formulas of G^∞ as sets, mimicking the build up of the formulas of $\mathcal{L}_\mathcal{W}$.

Definition 5.1. The class For^∞ is defined to be the smallest class which satisfies the following closure properties:

(1) For all natural numbers m, n and all sets a, b the class For^∞ contains

$$(h_m \mathrel{\dot\in} h_n), \quad (h_m \mathrel{\dot\in} p_a), \quad (p_a \mathrel{\dot\in} h_m), \quad (p_a \mathrel{\dot\in} p_b).$$

(2) For all natural numbers m, n and all sets a, the class For^∞ contains

$$(h_m \mathrel{\dot\in} H_n), \quad (p_a \mathrel{\dot\in} H_n).$$

(3) For all natural numbers m, n, all sets a, b, the class For^∞ contains

$$\dot{\mathcal{W}}(h_m, h_n), \quad \dot{\mathcal{W}}(h_m, p_a), \quad \dot{\mathcal{W}}(p_a, h_m), \quad \dot{\mathcal{W}}(p_a, p_b).$$

(4) For all $x, y \in \mathit{For}^\infty$, the class For^∞ also contains

$$\dot{\neg}\, x, \quad (x \mathbin{\dot{\vee}} y), \quad (x \mathbin{\dot{\wedge}} y).$$

(5) For all $x \in \mathit{For}^\infty$ and all natural numbers n, the class For^∞ also contains

$$\dot{\exists}\, h_n\, x, \quad \dot{\forall}\, h_n\, x, \quad \dot{\exists}\, H_n\, x, \quad \dot{\forall}\, H_n\, x.$$

This definition could be reformulated as an explicit elementary formula, for the prize of being less perspicuous. We are not going to work out the details, only formulate the corresponding assertion.

Lemma 5.2. *For^∞ is an elementarily definable class of* $\mathsf{NBGW}_{<E_0}$.

Clearly, for any sets a and b, $(a \mathbin{\dot{\rightarrow}} b)$ stands for $(\dot{\neg}\, a \mathbin{\dot{\vee}} b)$ and $(a \mathbin{\dot{\leftrightarrow}} b)$ for $((a \mathbin{\dot{\rightarrow}} b) \mathbin{\dot{\wedge}} (b \mathbin{\dot{\rightarrow}} a))$; other abbreviations of this sort are used as expected.

It is also elementarily decidable whether a set or class variable occurs freely (in the usual sense) within an element of For^∞. Moreover, there is an elementary class function Sub taking care of all sorts of simultaneous substitutions of free occurrences of set and class variables within an element of For^∞ by constants and variables of the appropriate sort. For instance, given a $\varphi \in \mathit{For}^\infty$, a set a and $i_1, i_2, j, m, n < \omega$,

$$\mathit{Sub}(\langle p_a, h_m, H_n\rangle, \langle h_{i_1}, h_{i_2}, H_j\rangle, \varphi)$$

is the element of For^∞ obtained from φ by simultaneously replacing all free occurrences of h_{i_1}, h_{i_2} and H_j by p_a, h_m and H_n, respectively. Also, if φ is given in the form $\psi[h_{i_1}, h_{i_2}, H_j]$, we often simply write $\psi[p_a, h_m, H_n]$ instead of $\mathit{Sub}(\langle p_a, h_m, H_n\rangle, \langle h_{i_1}, h_{i_2}, H_j\rangle, \varphi)$.

The previous definition is so that Gödel numbers, all belonging to For^∞, can be canonically assigned to the formulas of $\mathcal{L}_\mathcal{W}$. For this purpose we begin with fixing an mapping $\natural$ which assigns natural numbers to all set and class variables, making sure that different variables are mapped onto different natural numbers.

If u, v are set variables and if U is a class variable of $\mathcal{L}_\mathcal{W}$, we define

$$\ulcorner(u \in v)\urcorner := (h_{\natural(u)} \mathbin{\dot{\in}} h_{\natural(v)}), \qquad \ulcorner(u \in U)\urcorner := (h_{\natural(u)} \mathbin{\dot{\in}} H_{\natural(U)}),$$

$$\ulcorner\mathcal{W}(u, v)\urcorner := \dot{\mathcal{W}}(h_{\natural(u)}, h_{\natural(v)}).$$

The Gödel numbers of the non-atomic formulas of $\mathcal{L}_\mathcal{W}$ are inductively calculated in compliance with the equations

$$\ulcorner \neg A \urcorner := \dot\neg \ulcorner A \urcorner,$$

$$\ulcorner (A \lor B) \urcorner := (\ulcorner A \urcorner \ \dot\lor \ \ulcorner B \urcorner),$$

$$\ulcorner (A \land B) \urcorner := (\ulcorner A \urcorner \ \dot\land \ \ulcorner B \urcorner),$$

$$\ulcorner \exists x A \urcorner := \dot\exists\, h_{\natural(x)} \ulcorner A \urcorner,$$

$$\ulcorner \forall x A \urcorner := \dot\forall\, h_{\natural(x)} \ulcorner A \urcorner,$$

$$\ulcorner \exists X A \urcorner := \dot\exists\, H_{\natural(X)} \ulcorner A \urcorner,$$

$$\ulcorner \forall X A \urcorner := \dot\forall\, H_{\natural(X)} \ulcorner A \urcorner.$$

The elements of For^∞ are called $\mathcal{L}_\mathcal{W}^\infty$ formulas and will be denoted by the small Greek letters θ, φ, χ and ψ (possibly with subscripts). To increase the readability we often omit the dots when it is clear from the context that we speak about elements of For^∞.

The *set-closed* formulas are those $\mathcal{L}_\mathcal{W}^\infty$ formulas which do not contain free set variables (but they may contain free class variables and set constants); the closed formulas of $\mathcal{L}_\mathcal{W}^\infty$ are those $\mathcal{L}_\mathcal{W}^\infty$ formulas which contain neither free set variables nor free class variables. We collect the set-closed formulas in the class SC^∞ and the closed formulas of $\mathcal{L}_\mathcal{W}^\infty$ in the class $CFor^\infty$; both classes are elementarily definable.

The capital Greek letters $\Theta, \Phi, \Psi, \dots$ (possibly with subscripts) denote finite sequences of set-closed formulas. If Φ is the sequence of set-closed formulas $\varphi_1, \dots, \varphi_m$ and Ψ the sequence of set-closed formulas $\psi_1, \dots, \psi_n$, then

$$\langle 12, m, n, \varphi_1, \dots, \varphi_m, \psi_1, \dots, \psi_n \rangle$$

is the sequent with antecedent Φ and succedent Ψ; typically, it will be written as $(\Phi \supset \Psi)$ or simply as $\Phi \supset \Psi$.

The elementary, Σ_1^1, Σ^1 and Π^1 formulas of $\mathcal{L}_\mathcal{W}^\infty$ are defined analogously to the corresponding classes of $\mathcal{L}_\mathcal{W}$ formulas; set constants are now, of course, permitted as parameters.

Looking at the basic set existence and replacement axioms and at the global well-ordering axiom (GWO) of NBGW, we can convince ourselves that the corresponding axioms, formulated within the language $\mathcal{L}_\mathcal{W}^\infty$, are elementary $\mathcal{L}_\mathcal{W}^\infty$ formulas. We collect the resulting set-closed formulas in the class AX^∞.

Definition 5.3. The *degree* $dg(\varphi)$ of a set-closed formula φ is inductively defined as follows:

1. If φ is a set-closed elementary or Σ^1_1 formula of $\mathcal{L}^\infty_\mathcal{W}$, then $dg(\varphi) := 0$.
2. For all set-closed formulas which are neither elementary nor Σ^1_1 we set

$$
\begin{aligned}
dg(\neg\psi) &:= dg(\psi) + 1,\\
dg(\psi_1 \vee \psi_2) &:= \max(dg(\psi_1), dg(\psi_2)) + 1,\\
dg(\psi_1 \wedge \psi_2) &:= \max(dg(\psi_1), dg(\psi_2)) + 1,\\
dg(\exists h_n \psi[h_n]) &:= dg(\psi[p_\varnothing]) + 1,\\
dg(\forall h_n \psi[h_n]) &:= dg(\psi[p_\varnothing]) + 1,\\
dg(\exists H_n \psi[H_n]) &:= dg(\psi[H_n]) + 1,\\
dg(\forall H_n \psi[H_n]) &:= dg(\psi[H_n]) + 1.
\end{aligned}
$$

G^∞ is an extension of the classical Gentzen sequent calculus LK (cf., e.g., Girard [8] or Takeuti [17]) by additional axioms and rules of inference which take care of the non-logical axioms of NBGW. Universal set quantification in the succedent and the corresponding existential set quantification in the antecedent are infinitary rules branching over the collection of all sets. The axioms and rules of G^∞ can be grouped as follows.

I. Axioms. For all set-closed elementary formulas φ, all elements ψ of AX^∞, all sets a, b, all set-closed elementary formulas $\theta[p_\varnothing]$ and all H_m, h_n so that no variable conflicts arise:

(A1) $\varphi \supset \varphi$,

(A2) $\supset \psi$,

(A3) $\supset (p_a \in p_b)$ if $a \in b$,

(A4) $\supset (p_a \notin p_b)$ if $a \notin b$,

(A5) $\supset \exists H_m \forall h_n (h_n \in H_m \leftrightarrow \theta[h_n])$.

II. Structural rules. The structural rules of G^∞ consist of the usual weakening, exchange and contraction rules.

III. Propositional rules. The propositional rules of G^∞ consist of the usual rules for introducing the propositional connectives on the left and right hand sides of sequents.

IV. Quantifier rules for sets. Formulated only for succedents; there are also corresponding rules for the antecedents. For all set variables h_n, all set constants p_a and all set-closed formulas $\varphi[p_\varnothing]$:

$$\frac{\Phi \ \supset \ \Psi, \varphi[p_a]}{\Phi \ \supset \ \Psi, \exists h_n \varphi[h_n]}\ , \qquad \frac{\Phi \ \supset \ \Psi, \varphi[p_b] \quad \text{for all sets b}}{\Phi \ \supset \ \Psi, \forall h_n \varphi[h_n]}\ .$$

V. Quantifier rules for classes. Formulated only for succedents; there are also corresponding rules for the antecedents. By $(\star)$ we mark those rules where the designated free class variables are not to occur in the conclusion. For all set-closed formulas $\varphi[H_0]$and all class variables H_m, H_n so that no variable conflicts arise:

$$\frac{\Phi \ \supset \ \Psi, \varphi[H_m]}{\Phi \ \supset \ \Psi, \exists H_n \varphi[H_n]}\ , \qquad \frac{\Phi \ \supset \ \Psi, \varphi[H_m]}{\Phi \ \supset \ \Psi, \forall H_n \varphi[H_n]}\ (\star).$$

VI. Σ^1_1 collection rules. For all set-closed elementary formulas $\varphi[p_\varnothing, H_0]$ and all variables h_m, H_n, H_k so that no variable conflicts arise:

$$\frac{\Phi \ \supset \ \Psi, \forall h_m \exists H_n \varphi[h_m, H_n]}{\Phi \ \supset \ \Psi, \exists H_i \forall h_m \exists h_n \varphi[h_m, (H_i)_{h_n}]}\ .$$

VII. Cuts. For all set-closed formulas φ:

$$\frac{\Phi \ \supset \ \Psi, \varphi \qquad \Phi, \varphi \ \supset \ \Psi}{\Phi \ \supset \ \Psi}\ .$$

The formula φ is called the cut formula of this cut; the degree of a cut is the degree of its cut formula.

Since G^∞ has inference rules which branch over all sets, namely the rules for introducing universal quantification over sets in the succedents and existential quantification over sets in the antecedents, infinite proof trees may occur. We confine ourselves to those whose depths are bounded by initial segments of E_0.

Definition 5.4. Let k be an arbitrary standard natural number. For any notation $\mathfrak{a} \lhd \Omega_k$, any $n < \omega$ and any sequent $\Phi \supset \Psi$, we define $\mathsf{G}^\infty_k \overset{\mathfrak{a}}{\underset{n}{\vdash}} \Phi \supset \Psi$ by induction on $\mathfrak{a}$.

1. If $\Phi \supset \Psi$ is an axiom of G^∞, then we have $\mathsf{G}^\infty_k \overset{\mathfrak{a}}{\underset{n}{\vdash}} \Phi \supset \Psi$ for all $n < \omega$.

2. If $\mathsf{G}^\infty_k \overset{\mathfrak{a}_x}{\underset{n}{\vdash}} \Phi_x \supset \Psi_x$ and $\mathfrak{a}_x \lhd \mathfrak{a}$ for every premise of a rule which is not a cut, then we have $\mathsf{G}^\infty_k \overset{\mathfrak{a}}{\underset{n}{\vdash}} \Phi \supset \Psi$ for the conclusion $\Phi \supset \Psi$ of this rule.

3. If $\mathsf{G}^\infty_k \overset{\mathfrak{a}_i}{\underset{n}{\vdash}} \Phi_i \supset \Psi_i$ and $\mathfrak{a}_i \lhd \mathfrak{a}$ for the two premises $\Phi_i \supset \Psi_i$ of a cut $(i = 1, 2)$ whose degree is less than n, then we have $\mathsf{G}^\infty_k \overset{\mathfrak{a}}{\underset{n}{\vdash}} \Phi \supset \Psi$ for the conclusion $\Phi \supset \Psi$ of this cut.

To be precise, given a standard natural number k, we employ axiom (It-ECA) to introduce a class U such that, for any $\mathfrak{a} \lhd \Omega_k$, the projection $(U)_\mathfrak{a}$ consists of all pairs $(\Phi \supset \Psi, n)$ for which we have $\mathsf{G}^\infty_k \overset{\mathfrak{a}}{\underset{n}{\vdash}} \Phi \supset \Psi$.

$\mathsf{G}^\infty_k \overset{\mathfrak{a}}{\underset{0}{\vdash}} \Phi \supset \Psi$ says that there exists a cut-free proof in G^∞ whose depth is bounded by the notation $\mathfrak{a}$ and $\mathfrak{a} \lhd \Omega_k$. If we have $\mathsf{G}^\infty_k \overset{\mathfrak{a}}{\underset{1}{\vdash}} \Phi \supset \Psi$, then only set-closed formulas which are elementary or Σ^1_1 are permitted as cut formulas.

Since the main formulas of all axioms and the main formulas of the conclusions of all Σ^1_1 collection rules are elementary or Σ^1_1 formulas of $\mathcal{L}^\infty_\mathcal{W}$, partial cut elimination – eliminating all cuts whose cut formulas are neither elementary nor Σ^1_1 formulas – can be proved following standard patterns; see, for example, Schütte [15].

Theorem 5.5 (Partial cut elimination)**.** *Let k be a standard natural number. Then* $\mathsf{NBGW}_{<E_0}$ *proves for all $n < \omega$, all $\mathfrak{a} \in E_0$ such that $\omega^\mathfrak{a} \lhd \Omega_k$ and all sequents $\Phi \supset \Psi$ that*

$$\mathsf{G}^\infty_k \overset{\mathfrak{a}}{\underset{n+2}{\vdash}} \Phi \supset \Psi \quad \rightarrow \quad \mathsf{G}^\infty_k \overset{\omega^\mathfrak{a}}{\underset{n+1}{\vdash}} \Phi \supset \Psi.$$

The axioms and rules of G^∞ are so that apart from $\in$-induction, all axioms of $\mathsf{NBGW} + (\Sigma^1_1\text{-Col})$ are directly verified within G^∞. For proving the instances of $(\mathcal{L}_\mathcal{W}\text{-I}_\in)$ infinite derivations are required in general.

Lemma 5.6. *Let k be a standard natural number. Then* $\mathsf{NBGW}_{<E_0}$ *proves for all set-closed formulas $\varphi[p_\varnothing]$:*

1. *For all ordinals α, all sets a of set-theoretic rank α and all ordinals β such that $\beta = \omega^\alpha + \omega + 2$,*

$$\mathsf{G}^\infty_k \overset{\overline{\beta}}{\underset{0}{\vdash}} \forall h_m((\forall h_n \in h_m)\varphi[h_n] \rightarrow \varphi[h_m]) \supset \varphi[p_a].$$

2. $\mathsf{G}^\infty_k \overset{\Omega}{\underset{0}{\vdash}} \forall h_m((\forall h_n \in h_m)\varphi[h_n] \rightarrow \varphi[h_m]) \supset \forall h_m \varphi[h_m].$

Proof. We let ψ be the formula $\forall h_m((\forall h_n \in h_m)\varphi[h_n] \to \varphi[h_m])$ and show the first assertion by induction on α. Given a set a of rank α, the induction hypothesis implies for all $b \in a$

$$\mathsf{G}_k^\infty \left|\frac{\overline{\gamma}}{0}\right. \psi \supset \varphi[p_b] \tag{5.1}$$

where $\gamma := \omega^\alpha$. If $b \notin a$, then according to (A4) and weakening

$$\mathsf{G}_k^\infty \left|\frac{\overline{1}}{0}\right. \psi \supset p_b \notin p_a. \tag{5.2}$$

From (5.1) and (5.2) we conclude, for any set b,

$$\mathsf{G}_k^\infty \left|\frac{\overline{\gamma+1}}{0}\right. \psi \supset p_b \notin p_a \;\vee\; \varphi[p_b].$$

By universal set quantification we thus have

$$\mathsf{G}_k^\infty \left|\frac{\overline{\gamma+2}}{0}\right. \psi \supset (\forall h_n \in p_a)\varphi[h_n],$$

and from this, simple manipulations within G^∞ also lead to

$$\mathsf{G}_k^\infty \left|\frac{\overline{\gamma+\omega}}{0}\right. \psi,\, (\forall h_n \in p_a)\varphi[h_n] \to \varphi[p_a] \supset \varphi[p_a].$$

Universal set quantification and contraction within the antecedent therefore finish the proof of our first assertion. The second assertion follows from the first by a universal set quantification in the succedent. □

It is now routine to verify by induction on the lengths of the proofs in the system $\mathsf{NBGW}+(\mathcal{L}_\mathcal{W}\text{-}\mathsf{I}_\in)+(\Sigma^1_1\text{-}\mathsf{Col})$ that every theorem of $\mathsf{NBGW}+(\mathcal{L}_\mathcal{W}\text{-}\mathsf{I}_\in)+(\Sigma^1_1\text{-}\mathsf{Col})$ is derivable in G^∞.

Theorem 5.7. *Let k be a standard natural number greater 0 and A a formula of $\mathcal{L}_\mathcal{W}$ without free set variables. If A is derivable in $\mathsf{NBGW}+(\mathcal{L}_\mathcal{W}\text{-}\mathsf{I}_\in)+(\Sigma^1_1\text{-}\mathsf{Col})$, then there exist standard natural numbers m and n such that $\mathsf{NBGW}_{<E_0}$ proves*

$$\mathsf{G}_k^\infty \left|\frac{\Omega+\overline{m}}{n}\right. \supset \ulcorner A \urcorner.$$

Applying Theorem 5.5 finitely often we can strengthen this theorem to an interpretation of $\mathsf{NBGW}+(\mathcal{L}_\mathcal{W}\text{-}\mathsf{I}_\in)+(\Sigma^1_1\text{-}\mathsf{Col})$ in G^∞ with proofs whose cut formulas are either elementary or Σ^1_1 formulas and whose depths are bounded by Ω_k for suitable standard natural numbers k.

Corollary 5.8. *Let A be a formula of $\mathcal{L}_{\mathcal{W}}$ without free set variables. If A is derivable in* $\mathsf{NBGW} + (\mathcal{L}_{\mathcal{W}}\text{-}\mathsf{I}_{\in}) + (\Sigma^1_1\text{-}\mathsf{Col})$, *then there exists a standard natural number k such that* $\mathsf{NBGW}_{<E_0}$ *proves that there is a notation $\mathfrak{a} \lhd \Omega_k$ such that*

$$\mathsf{G}^{\infty}_{k} \overset{\mathfrak{a}}{\underset{1}{\vdash}} \supset \ulcorner A \urcorner.$$

The next step is to introduce a truth definition for the set-closed formulas. This truth definition will always depend on a class U such that the class parameters are interpreted as projections $(U)_a$ (a any set) of U and the class quantifiers range over all projections of U; the set quantifiers range over the universe of all sets.

In the following we let Lh be the elementary class function which assigns to any element φ of For^{∞} the number $Lh(\varphi) < \omega$ of occurrences of logical connectives in φ. Also, F_ω is defined to be the class of all functions with domain ω; i.e. we set

$$F_\omega \ := \ \{f : Fun[f] \ \wedge \ Dom[f] = \omega\}.$$

For an $f \in F_\omega$, a set a and an $n < \omega$, we write $f_{(a|n)}$ for the element of F_ω which maps n to a and otherwise agrees with f.

Definition 5.9.

1. $Sat[U, V, u, v]$ is defined to be the elementary $\mathcal{L}_{\mathcal{W}}$ formula

   $$(\exists \varphi \in SC^{\infty})(\exists f \in F_\omega)(u = \langle \varphi, f\rangle \ \wedge \ Lh(\varphi) = v \ \wedge \ A[U, V, f, \varphi]),$$

   where $A[U, V, f, \varphi]$ is the auxiliary formula taken to be the disjunction of the following clauses:

   (1) $\exists x \exists y(\varphi = (p_x \mathrel{\dot{\in}} p_y) \ \wedge \ x \in y)$,

   (2) $\exists x(\exists n < \omega)(\varphi = (p_x \mathrel{\dot{\in}} H_n) \ \wedge \ x \in (U)_{f(n)})$,

   (3) $\exists x \exists y(\varphi = \dot{\mathcal{W}}(p_x, p_y) \ \wedge \ \mathcal{W}(x, y))$,

   (4) $\exists x(\varphi = \dot{\neg}\, x \ \wedge \ \langle x, f\rangle \notin V)$,

   (5) $\exists x \exists y(\varphi = (x \ \dot{\vee} \ y) \ \wedge \ (\langle x, f\rangle \in V \ \vee \ \langle y, f\rangle \in V))$,

   (6) $\exists x \exists y(\varphi = (x \ \dot{\wedge} \ y) \ \wedge \ \langle x, f\rangle \in V \ \wedge \ \langle y, f\rangle \in V)$,

   (7) $\exists x(\exists n < \omega)(\varphi = \dot{\exists}\, h_n\, x \ \wedge \ \exists y(\langle Sub(p_y, h_n, x), f\rangle \in V))$,

   (8) $\exists x(\exists n < \omega)(\varphi = \dot{\forall}\, h_n\, x \ \wedge \ \forall y(\langle Sub(p_y, h_n, x), f\rangle \in V))$,

   (9) $\exists x(\exists n < \omega)(\varphi = \dot{\exists}\, H_n\, x \ \wedge \ \exists y(\langle x, f_{(y|n)}\rangle \in V))$,

   (10) $\exists x(\exists n < \omega)(\varphi = \dot{\forall}\, H_n\, x \ \wedge \ \forall y(\langle x, f_{(y|n)}\rangle \in V))$.

2. A class V is called a *satisfaction hierarchy* with respect to U if it satisfies iterating this formula Sat along the natural numbers; i.e.

$$SH[U,V] \;:=\; (\forall n < \omega)((V)_n = \{x : Sat[U, \bigcup\{(V)_i : i < n\}, x, n]\}).$$

In this definition, the parameter U codes a universe of classes; the class V collects those pairs $\langle \varphi, f \rangle \in SC^\infty \times F_\omega$ such that φ is satisfied with respect to U if its class parameters are interpreted according to f. This leads directly to the definition of the truth of set-closed formulas with respect to a class U and an $f \in F_\omega$.

Definition 5.10. For all classes U and sets f, φ we set

$$Tr[U,f,\varphi] \;:=\; \varphi \in SC^\infty \;\wedge\; f \in F_\omega \;\wedge\; \exists X(SH[U,X] \;\wedge\; \langle \varphi, f \rangle \in (X)_{Lh(\varphi)}).$$

Note that the principle $(\mathsf{It\text{-}ECA})$ makes sure that, provable in $\mathsf{NBGW}_{<E_0}$, for every class U there exists a satisfaction hierarchy with respect to U which is essentially unique: if $SH[U,V_1]$ and $SH[U,V_2]$, then $(V_1)_n = (V_2)_n$ for all $n < \omega$. It is now an easy exercise to verify that this definition of truth has the expected closure properties

Lemma 5.11. *The theory* $\mathsf{NBGW}_{<E_0}$ *proves, for all classes* U*, all* $f \in F_\omega$*, all set-closed formulas* φ, ψ*, all sets* x, y *and all* $n < \omega$*, that*

$$\begin{aligned}
Tr[U,f,(p_x \mathrel{\dot\in} p_y)] &\leftrightarrow x \in y,\\
Tr[U,f,(p_x \mathrel{\dot\in} H_n)] &\leftrightarrow x \in (U)_{f(n)},\\
Tr[U,f,\dot{\mathcal{W}}(p_x,p_y)] &\leftrightarrow \mathcal{W}(x,y),\\
Tr[U,f,\dot\neg\,\varphi] &\leftrightarrow \neg\, Tr[U,f,\varphi],\\
Tr[U,f,(\varphi \mathrel{\dot\vee} \psi)] &\leftrightarrow (Tr[U,f,\varphi] \vee Tr[U,f,\psi]),\\
Tr[U,f,(\varphi \mathrel{\dot\wedge} \psi)] &\leftrightarrow (Tr[U,f,\varphi] \wedge Tr[U,f,\psi]),\\
Tr[U,f,\dot\exists\, h_n\varphi] &\leftrightarrow \exists x\, Tr[U,f,Sub(p_x,h_n,\varphi)],\\
Tr[U,f,\dot\forall\, h_n\varphi] &\leftrightarrow \forall x\, Tr[U,f,Sub(p_x,h_n,\varphi)],\\
Tr[U,f,\dot\exists\, H_n\varphi] &\leftrightarrow \exists x\, Tr[U,f_{(x|n)},\varphi],\\
Tr[U,f,\dot\forall\, H_n\varphi] &\leftrightarrow \forall x\, Tr[U,f_{(x|n)},\varphi].
\end{aligned}$$

A further expected property of this truth definition is that the truth of an set-closed elementary formula only depends on the interpretation of its class parameters. The following is obvious from, for example, the previous lemma.

Lemma 5.12. *In* $\mathsf{NBGW}_{<E_0}$ *we have, for all classes* U, V*, all* $f, g \in F_\omega$ *and all set-closed elementary formulas* φ*, that*

$$(\forall n < \omega)((U)_{f(n)} = (V)_{g(n)}) \;\rightarrow\; (Tr[U, f, \varphi] \leftrightarrow Tr[V, g, \varphi]).$$

This definition of truth reflects $\mathcal{L}_\mathcal{W}$ formulas without bound class variables in the appropriate way. To simplify the formulation of the following lemma, we state it only for formulas without class parameters.

Lemma 5.13 (Truth reflection)**.** *Let* A *be a closed elementary formula of* $\mathcal{L}_\mathcal{W}$ *and* B *a closed* Π^1 *formula of* $\mathcal{L}_\mathcal{W}$*. Then the theory* $\mathsf{NBGW}_{<E_0}$ *proves, for any* U *and* $f \in F_\omega$*:*

1. $A \;\leftrightarrow\; Tr[U, f, \ulcorner A \urcorner]$.
2. $B \;\rightarrow\; Tr[U, f, \ulcorner B \urcorner]$.

In the following Elm stands for the class of all elementary $\mathcal{L}^\infty_\mathcal{W}$ formulas which contain h_0 as the only free set variable; additional free occurrences of class variables are permitted. Then we write

$$Def[U, V, u] \;:=\; Def_1[U, u] \;\vee\; Def_2[U, V, u],$$

where

$$Def_1[U, u] := \exists v(u = \langle 0, v\rangle \;\wedge\; v \in U),$$

$$Def_2[U, V, u] := \begin{cases} \exists z(\exists \varphi \in Elm)(\exists f \in F_\omega)(u = \langle\langle \varphi, f\rangle, z\rangle \\ \qquad\qquad \wedge\; Sat[U, V, \langle Sub(p_z, h_0, \varphi), f\rangle, Lh(\varphi)]). \end{cases}$$

For carrying through an asymmetric interpretation of the (quasi cut-free) derivations of the systems G^∞_k in Theorem 5.21 below, we need hierarchies of classes with sufficiently strong closure properties. One possible approach to provide such hierarchies is to turn to an analogue of the constructible hierarchy.

Definition 5.14. Let k be a standard natural number. Then a class W is said to be a *k-constructible hierarchy* if, for all $\mathfrak{a} \in SLim_0 \cap \Omega_k$, $\mathfrak{b} \in Lim_0 \cap \Omega_k$ and $n < \omega$, we have:

$$(W)_\mathfrak{a} = \{\langle\langle x, y\rangle, z\rangle : x \in Lim_0 \cap \mathfrak{a} \;\wedge\; \langle y, z\rangle \in (W)_x\},$$

$$(W)_{\mathfrak{b}+\overline{(n+1)}} = \{x : Sat[(W)_{\mathfrak{b}}, \bigcup\{(W)_{\mathfrak{b}+\overline{y}} : 0 < y < (n+1)\}, x, n]\},$$

$$(W)_{\mathfrak{b}+\overline{\omega}} = \{x : Def[(W)_{\mathfrak{b}}, \bigcup\{(W)_y : \mathfrak{b} \lhd y \lhd \mathfrak{b}+\overline{\omega}\}, x]\}.$$

The following lemma follows more or less directly, by coding two formulas into one, from the hierarchy axiom of $\mathsf{NBGW}_{<E_0}$; its proof can therefore be omitted.

Lemma 5.15. *Let k be a standard natural number. Then $\mathsf{NBGW}_{<E_0}$ proves the existence of a k-constructible hierarchy.*

Now assume that W is a k-constructible hierarchy. For any $\mathfrak{a} \in Lim_0$, the class $(W)_{\mathfrak{a}}$ may be considered as a code of the collection of all classes $((W)_{\mathfrak{a}})_u$, where u is an arbitrary set. The idea of this hierarchy then is as follows:

(i) $(W)_0$ codes the empty collection of classes.

(ii) For any $\mathfrak{b} \in Lim_0$, the successor stages $\mathfrak{b} + \overline{(n+1)}$ are used to collect all set-closed formulas of length n together with $f \in F_\omega$ which are true if their class parameters are interpreted by projections of $(W)_{\mathfrak{b}}$ via f and their class quantifiers range over the projections of $(W)_{\mathfrak{b}}$.

(iii) At limit stages of the form $\mathfrak{b} + \overline{\omega}$ the class $(W)_{\mathfrak{b}+\omega}$ collects $(W)_{\mathfrak{b}}$ and all classes which are definable by elementary formulas and interpretations of class parameters as projections of $(W)_{\mathfrak{b}}$.

(iv) At strong limits simply all projections of the previous limit stages are coded together.

Lemma 5.16. *Let k be a standard natural number. Then $\mathsf{NBGW}_{<E_0}$ proves for all k-constructible hierarchies W, all $f \in F_\omega$, all $\mathfrak{a} \in Lim_0 \cap \Omega_k$ and all set-closed formulas φ with $Lh(\varphi) = n$ and all $\psi \in Elm$:*

1. $Tr[(W)_{\mathfrak{a}}, f, \varphi] \leftrightarrow \langle \varphi, f \rangle \in (W)_{\mathfrak{a}+\overline{(n+1)}}$.
2. $((W)_{\mathfrak{a}+\overline{\omega}})_{\langle \psi, f\rangle} = \{x : Tr[(W)_{\mathfrak{a}}, f, Sub(p_x, h_0, \psi)]\}$.

The proof of this lemma is by carefully carrying out the informal considerations above; its details can be left out. Some further useful properties of hierarchies of this sort are listed in the following lemma. For its formulation and for later use we introduce the abbreviations

$$U \mathrel{\dot\in} V := \exists x(U = (V)_x),$$

$$U \mathrel{\dot\subset} V := \forall x((U)_x \mathrel{\dot\in} V),$$

$$U \mathrel{\dot\subset_\omega} V := (\forall n < \omega)((U)_n \mathrel{\dot\in} V).$$

Lemma 5.17. *Let k be a standard natural number. Then* $\mathsf{NBGW}_{<E_0}$ *proves for all k-constructible hierarchies W, all $\mathfrak{a} \in Lim_0 \cap \Omega_k$ and all $\mathfrak{b} \in Lim_0 \cap \mathfrak{a}$:*

1. $(W)_\mathfrak{a} \mathrel{\dot\in} (W)_{\mathfrak{a}+\overline{\omega}}$ *and* $(W)_\mathfrak{a} \mathrel{\dot\subset} (W)_{\mathfrak{a}+\overline{\omega}}$.

2. $(W)_\mathfrak{b} \mathrel{\dot\subset} (W)_\mathfrak{a}$ *and* $(W)_\mathfrak{b} \mathrel{\dot\in} (W)_\mathfrak{a}$.

Proof. Assume that W, $\mathfrak{a}$ and $\mathfrak{b}$ satisfy the assumptions of this lemma. Then $(W)_\mathfrak{a} \mathrel{\dot\in} (W)_{\mathfrak{a}+\overline{\omega}}$ follows from $(W)_\mathfrak{a} = ((W)_{\mathfrak{a}+\omega})_0$. In order to show $(W)_\mathfrak{a} \mathrel{\dot\subset} (W)_{\mathfrak{a}+\overline{\omega}}$, pick any set x and an $f \in F_\omega$ such that $f(0) = x$. If φ is the elementary $\mathcal{L}^\infty_\mathcal{W}$ formula $(h_0 \in H_0)$, then $((W)_\mathfrak{a})_x = ((W_{\mathfrak{a}+\overline{\omega}})_{\langle \varphi, f\rangle}$. This establishes the first assertion.

If $\mathfrak{a}$ is an element of $SLim_0$ and $\mathfrak{b} \in Lim_0 \cap \mathfrak{a}$, then $(W)_\mathfrak{b} \mathrel{\dot\subset} (W)_\mathfrak{a}$ directly follows from the definition of $(W)_\mathfrak{a}$.

From $\mathfrak{a} \in SLim_0$ and $\mathfrak{b} \in Lim_0 \cap \mathfrak{a}$ it also follows that $\mathfrak{b} + \overline{\omega} \in Lim_0 \cap \mathfrak{a}$, hence $(W)_{\mathfrak{b}+\overline{\omega}} \mathrel{\dot\subset} (W)_\mathfrak{a}$. In view of the first assertion this implies $(W)_\mathfrak{b} \mathrel{\dot\in} (W)_\mathfrak{a}$. A simple transfinite induction on $\mathfrak{a}$, combined with the first assertion, finishes the proof of the second. □

The formula $Tr[U, f, \varphi]$ interprets the class parameters of φ by projections of U which are provided by the element f of F_ω. Sometimes it is more practical to have them coded into a class V.

Definition 5.18. For classes U, V and set-closed formulas φ we set

$$TR[U, V, \varphi] \;:=\; (\exists f \in F_\omega)((\forall n < \omega)((V)_n = (U)_{f(n)}) \;\wedge\; Tr[U, f, \varphi]).$$

For classes V, X, Y and an $n < \omega$ we write $Y = V(X|n)$ to express that $(Y)_n = X$ and $(Y)_m = (V)_m$ for any $m < \omega$ which is different from n. Then

$$TR[U, V, \varphi(X/H_n)] \;:=\; X \mathrel{\dot\in} U \;\wedge\; \exists Y(Y = V(X|n) \;\wedge\; TR[U, Y, \varphi]).$$

Hence in $TR[U, V, \varphi(X/H_n)]$ all free occurrences of the class variable H_n within φ are interpreted by X and all others according to V. Naturally, the predicate $TR[U, V, \varphi]$ inherits the closure properties stated in Lemma 5.11 from $Tr[U, f, \varphi]$. We collect them for later reference.

Lemma 5.19. *The theory* $\mathsf{NBGW}_{<E_0}$ *proves, for all classes U, V, all set-closed formulas φ, ψ, all sets x, y and all $n < \omega$, that*

$$
\begin{aligned}
TR[U,V,(p_x \mathrel{\dot\in} p_y)] &\leftrightarrow x \in y,\\
TR[U,V,(p_x \mathrel{\dot\in} H_n)] &\leftrightarrow x \in (V)_n,\\
TR[U,V,\dot{\mathcal{W}}(p_x,p_y)] &\leftrightarrow \mathcal{W}(x,y),\\
TR[U,V,\dot\neg\,\varphi] &\leftrightarrow \neg\, TR[U,V,\varphi],\\
TR[U,V,(\varphi \mathrel{\dot\vee} \psi)] &\leftrightarrow (TR[U,V,\varphi] \;\vee\; TR[U,V,\psi]),\\
TR[U,V,(\varphi \mathrel{\dot\wedge} \psi)] &\leftrightarrow (TR[U,V,\varphi] \;\wedge\; TR[U,V,\psi]),\\
TR[U,V,\dot\exists\, h_n\varphi] &\leftrightarrow \exists x\, TR[U,V,\mathit{Sub}(p_x,h_n,\varphi)],\\
TR[U,V,\dot\forall\, h_n\varphi] &\leftrightarrow \forall x\, TR[U,V,\mathit{Sub}(p_x,h_n,\varphi)],\\
TR[U,V,\dot\exists\, H_n\varphi] &\leftrightarrow (\exists X \mathrel{\dot\in} U)\, TR[U,V,\varphi(X/H_n)],\\
TR[U,V,\dot\forall\, H_n\varphi] &\leftrightarrow (\forall X \mathrel{\dot\in} U)\, TR[U,V,\varphi(X/H_n)].
\end{aligned}
$$

Utilizing these properties, it is routine to show (by simultaneous induction on the length of φ and ψ) that set-closed Σ^1 formulas are upward persistent and set-closed Π^1 formulas downward persistent.

Lemma 5.20. *Let k be a standard natural number. Then $\mathsf{NBGW}_{<E_0}$ proves for all k-constructible hierarchies W, all classes U, all set-closed Σ^1 formulas φ, all set-closed Π^1 formulas ψ and all $\mathfrak{a}, \mathfrak{b} \in \mathit{Lim}_0 \cap \Omega_k$:*

1. $\mathfrak{a} \lhd \mathfrak{b} \;\wedge\; TR[(W)_\mathfrak{a}, U, \varphi] \;\rightarrow\; TR[(W)_\mathfrak{b}, U, \varphi]$.
2. $\mathfrak{a} \lhd \mathfrak{b} \;\wedge\; U \mathrel{\dot\subset_\omega} (W)_\mathfrak{a} \;\wedge\; TR[(W)_\mathfrak{b}, U, \psi] \;\rightarrow\; TR[(W)_\mathfrak{a}, U, \psi]$.

If Φ and Ψ are finite sequences of set-closed formulas, $(\Phi \supset \Psi)^\bullet$ denotes (the Gödel number of) the disjunction whose disjuncts are the negated formulas of Φ and the formulas of Ψ.

Theorem 5.21. *Let k be a standard natural number. In $\mathsf{NBGW}_{<E_0}$ we can prove that, for all k-constructible hierarchies W, all classes U, all finite sequences Φ of set-closed Π^1 formulas, all finite sequences Ψ of set-closed Σ^1 formulas, all $\mathfrak{a} \lhd \Omega_k$ and all $\mathfrak{b}, \mathfrak{c} \in \mathit{Lim}_0 \cap \Omega_k$, we have the implication*

$$
\mathsf{G}_k^\infty \,\Big|\!\frac{\mathfrak{a}}{1}\; \Phi \supset \Psi \;\wedge\; \mathfrak{b} + \omega^{\mathfrak{a}+\bar{1}} \trianglelefteq \mathfrak{c} \;\wedge\; U \mathrel{\dot\subset_\omega} (W)_\mathfrak{b} \;\rightarrow\; TR[(W)_\mathfrak{c}, U, (\Phi \supset \Psi)^\bullet].
$$

Proof. We show this theorem by induction on $\mathfrak{a}$, which is justified by Theorem 3.8, and distinguish the following cases:

1. $\Phi \supset \Psi$ is an axiom (A1)–(A4) or a conclusion of a structural rule, a propositional rule, a quantifier rule for set or a quantifier rule for classes. Then the assertion is trivially satisfied, is a consequence of Lemma 5.13 and Lemma 5.19 or follows from the induction hypothesis.

2. $\Phi \supset \Psi$ is an axiom (A5). Then Φ is empty and Ψ consists of a single formula $\exists H_m \forall h_n (h_n \in H_m \leftrightarrow \varphi[h_n])$, where $\varphi[p_\varnothing]$ is a set-closed elementary formula. In this case, the assertion is a consequence of Lemma 5.16, Lemma 5.19, and Lemma 5.20.

3. $\Phi \supset \Psi$ is a conclusion of a Σ^1_1 collection rule. Then the sequence Ψ is of the form $\Psi_0, \exists H_i \forall h_m \exists h_n \theta[h_m, (H_i)_{h_n}]$ for some set-closed elementary formula $\theta[p_\varnothing, H_0]$, and there exists an $\mathfrak{a}_0 \lhd \mathfrak{a}$ such that

$$\mathsf{G}^\infty_k \overset{\mathfrak{a}_0}{\underset{1}{\vdash}} \Phi \supset \Psi_0, \forall h_m \exists H_n \theta[h_m, H_n].$$

For $\mathfrak{c}_0 := \mathfrak{b} + \omega^{\mathfrak{a}_0 + \overline{1}}$ the induction hypothesis gives us

$$TR[(W)_{\mathfrak{c}_0}, U, (\Phi \supset \Psi_0, \forall h_m \exists H_n \theta[h_m, H_n])^\bullet].$$

Clearly, $\mathfrak{c}_0 \lhd \mathfrak{c}$, and therefore Lemma 5.17 implies

$$(W)_{\mathfrak{c}_0} \mathrel{\dot{\subset}} (W)_{\mathfrak{c}} \qquad \text{and} \qquad (W)_{\mathfrak{c}_0} \mathrel{\dot{\in}} (W)_{\mathfrak{c}}. \tag{5.3}$$

Now we set $\theta_1[h_m] := \theta[h_m, H_n]$ and $\theta_2[h_m, h_n] := \theta[h_m, (H_i)_{h_n}]$. Then by Lemma 5.19

$$TR[(W)_{\mathfrak{c}_0}, U, (\Phi \supset \Psi_0)^\bullet] \;\vee\; \forall x \exists y\, TR[(W)_{\mathfrak{c}_0}, U, \theta_1[p_x]((((W)_{\mathfrak{c}_0})_y / H_n)],$$

and a simple persistency argument, see Lemma 5.20, together with (5.3) yields

$$TR[(W)_{\mathfrak{c}}, U, (\Phi \supset \Psi_0)^\bullet] \;\vee\; \forall x \exists y\, TR[(W)_{\mathfrak{c}}, U, \theta_1[p_x]((((W)_{\mathfrak{c}_0})_y / H_n)].$$

This can also be written as

$$TR[(W)_{\mathfrak{c}}, U, (\Phi \supset \Psi_0)^\bullet] \;\vee\; \forall x \exists y\, TR[(W)_{\mathfrak{c}}, U, \theta_2[p_x, p_y]((W)_{\mathfrak{c}_0} / H_i)].$$

In view of $(W)_{\mathfrak{c}_0} \mathrel{\dot{\in}} (W)_{\mathfrak{c}}$, see (5.3), we continue with

$$TR[(W)_{\mathfrak{c}}, U, (\Phi \supset \Psi_0)^\bullet] \vee (\exists Z \mathrel{\dot{\in}} (W)_{\mathfrak{c}}) \forall x \exists y\, TR[(W)_{\mathfrak{c}}, U, \theta_2[p_x, p_y](Z / H_i)].$$

By Lemma 5.19 this tells us

$$TR[(W)_{\mathfrak{c}}, U, (\Phi \supset \Psi_0, \exists H_i \forall h_m \exists h_n \theta[h_m, (H_i)_{h_n}])^\bullet],$$

completing the treatment of this case.

4. $\Phi \supset \Psi$ is a conclusion of a cut. By assumption, its cut formula has to be a set-closed elementary formula or a set-closed formula of the form $\exists H_n \theta$, where θ is set-closed elementary. In the remainder we concentrate on the second and more complicated case. Then there exists $\mathfrak{a}_1, \mathfrak{a}_2 \lhd \mathfrak{a}$ such that

$$\mathsf{G}_k^\infty \left|\frac{\mathfrak{a}_1}{1}\right. \Phi \supset \Psi, \exists H_n \theta, \tag{5.4}$$

$$\mathsf{G}_k^\infty \left|\frac{\mathfrak{a}_2}{1}\right. \Phi, \exists H_n \theta \supset \Psi. \tag{5.5}$$

Set $\mathfrak{c}_1 := \mathfrak{b} + \omega^{\mathfrak{a}_1 + \bar{1}}$ and apply the induction hypothesis to (5.4). Then we obtain

$$TR[(W)_{\mathfrak{c}_1}, U, (\Phi \supset \Psi, \exists H_n \theta)^\bullet]$$

and from that, because of Lemma 5.19,

$$TR[(W)_{\mathfrak{c}_1}, U, (\Phi \supset \Psi)^\bullet] \;\vee\; (\exists X \mathrel{\dot\in} (W)_{\mathfrak{c}_1})\, TR[(W)_{\mathfrak{c}_1}, U, \theta(X/H_n)]. \tag{5.6}$$

Furthermore, by an inversion argument (we did not formulate it explicitly but it can be proved in a straightforward way), assertion (5.5) gives

$$\mathsf{G}_k^\infty \left|\frac{\mathfrak{a}_2}{1}\right. \Phi, \mathit{Sub}(\langle H_m \rangle, \langle H_n \rangle, \theta) \supset \Psi, \tag{5.7}$$

where H_m is a fresh class variable which does not occur in $\Phi \supset \Psi$ and $\exists H_n \theta$. For $\mathfrak{c}_2 := \mathfrak{c}_1 + \omega^{\mathfrak{a}_2 + \bar{1}}$ and all $V \mathrel{\dot\subset_\omega} (W)_{\mathfrak{c}_1}$ the induction hypothesis applied to (5.7) – with $\mathfrak{a}$, $\mathfrak{b}$ and $\mathfrak{c}$ replaced by $\mathfrak{a}_2$, $\mathfrak{c}_1$ and $\mathfrak{c}_2$, respectively – yields

$$TR[(W)_{\mathfrak{c}_2}, V, (\Phi, \mathit{Sub}(\langle H_m \rangle, \langle H_n \rangle, \theta) \supset \Psi)^\bullet].$$

In particular, this is the case for any $V \mathrel{\dot\subset_\omega} (W)_{\mathfrak{c}_1}$ satisfying $(V)_m \mathrel{\dot\in} (W)_{\mathfrak{c}_1}$ as well as $(V)_i = (U)_i$ if $i < \omega$ and $i \neq m$. Once more we apply Lemma 5.19 and deduce

$$\begin{aligned} &TR[(W)_{\mathfrak{c}_2}, U, (\Phi \supset \Psi)^\bullet] \;\vee \\ &\qquad (\forall X \mathrel{\dot\in} (W)_{\mathfrak{c}_1}) \neg TR[(W)_{\mathfrak{c}_2}, U, \mathit{Sub}(\langle H_m \rangle, \langle H_n \rangle, \theta)(X/H_m)]. \end{aligned}$$

In view of the persistency properties formulated in Lemma 5.20 and an obvious exchange of variables, $TR[(W)_{\mathfrak{c}_2}, U, \mathit{Sub}(\langle H_m \rangle, \langle H_n \rangle, \theta)(X/H_m)]$ is equivalent, for $X \mathrel{\dot\in} (W)_{\mathfrak{c}_1}$, to $TR[(W)_{\mathfrak{c}_1}, U, \theta(X/H_n)]$, and it follows that

$$TR[(W)_{\mathfrak{c}_2}, U, (\Phi \supset \Psi)^\bullet] \;\vee\; (\forall X \mathrel{\dot\in} (W)_{\mathfrak{c}_1}) \neg TR[(W)_{\mathfrak{c}_1}, U, \theta(X/H_n)].$$

Together with (5.6) this implies

$$TR[(W)_{\mathfrak{c}_1}, U, (\Phi \supset \Psi)^\bullet] \quad \vee \quad TR[(W)_{\mathfrak{c}_2}, U, (\Phi \supset \Psi)^\bullet].$$

Since $\mathfrak{c}_2 = \mathfrak{c}_1 + \omega^{\mathfrak{a}_2+\bar{1}} = \mathfrak{b} + \omega^{\mathfrak{a}_1+\bar{1}} + \omega^{\mathfrak{a}_2+\bar{1}} \lhd \mathfrak{b} + \omega^{\mathfrak{a}+\bar{1}} \unlhd \mathfrak{c}$, Lemma 5.20 proves $TR[(W)_{\mathfrak{c}}, U, (\Phi \supset \Psi)^\bullet]$, as desired.

Therefore all possible cases for deriving the sequent $\Phi \supset \Psi$ within G^∞_k have been considered, proving our theorem. □

Corollary 5.22. *Let k be a standard natural number and A a closed elementary $\mathcal{L}_\mathcal{W}$ formula. Then the theory $\mathsf{NBGW}_{<E_0}$ proves, for all $\mathfrak{a} \lhd \Omega_k$, that*

$$\mathsf{G}^\infty_k \left|\frac{\mathfrak{a}}{1}\right. \supset \ulcorner A \urcorner \quad \rightarrow \quad A.$$

Proof. First of all, Lemma 5.15 implies that there exists a k-constructible hierarchy W. Then, assuming $\mathsf{G}^\infty_k \left|\frac{\mathfrak{a}}{1}\right. \supset \ulcorner A \urcorner$ and setting $\mathfrak{c} := \omega^{\mathfrak{a}+\bar{1}}$, the previous theorem implies $TR[(W)_{\mathfrak{c}}, \varnothing, \ulcorner A \urcorner]$. Because of truth reflection, c.f. Lemma 5.13, we therefore also have A. □

Theorem 5.23 (Reduction)**.** *The theory* $\mathsf{NBGW} + (\mathcal{L}_\mathcal{W}\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{Col})$ *can be reduced to the theory* $\mathsf{NBGW}_{<E_0}$ *with respect to all closed elementary $\mathcal{L}_\mathcal{W}$ formulas; i.e. for all closed elementary $\mathcal{L}_\mathcal{W}$ formulas A we have*

$$\mathsf{NBGW} + (\mathcal{L}_\mathcal{W}\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{Col}) \vdash A \quad \Longrightarrow \quad \mathsf{NBGW}_{<E_0} \vdash A.$$

Proof. Let A be a closed elementary $\mathcal{L}_\mathcal{W}$ formula provable in the theory $\mathsf{NBGW} + (\mathcal{L}_\mathcal{W}\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{Col})$. According to Corollary 5.8 we thus have

$$\mathsf{NBGW}_{<E_0} \vdash (\exists \mathfrak{a} \lhd \Omega_k)(\mathsf{G}^\infty_k \left|\frac{\mathfrak{a}}{1}\right. \supset \ulcorner A \urcorner)$$

for a suitable standard natural number k. Hence the previous corollary yields $\mathsf{NBGW}_{<E_0} \vdash A$. □

Corollary 5.24 (Final result)**.** *The four theories* $\mathsf{NBG} + (\mathcal{L}_2\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{AC})$, $\mathsf{NBGW} + (\mathcal{L}_\mathcal{W}\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{Col})$, $\mathsf{NBGW}_{<E_0}$ *and* $\mathsf{NBG}_{<E_0}$ *are equiconsistent.*

To prove this summary, we simply recall what we have shown before: In view of Theorem 4.3, $\mathsf{NBG}_{<E_0}$ is contained in $\mathsf{NBG}+(\mathcal{L}_2\text{-}\mathsf{I}_\in)+(\Sigma^1_1\text{-}\mathsf{AC})$, which, according to Corollary 2.3, is equivalent to $\mathsf{NBG} + (\mathcal{L}_2\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{Col})$. However, this system is obviously contained in $\mathsf{NBGW} + (\mathcal{L}_\mathcal{W}\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{Col})$. The above reduction theorem provides the reduction of $\mathsf{NBGW} + (\mathcal{L}_\mathcal{W}\text{-}\mathsf{I}_\in) + (\Sigma^1_1\text{-}\mathsf{Col})$ to $\mathsf{NBGW}_{<E_0}$, a conservative extension of $\mathsf{NBG}_{<E_0}$ by Theorem 4.4. Thus the circle is closed.

References

[1] A. Cantini, *On the relation between choice and comprehension principles in second order arithmetic*, Journal of Symbolic Logic **51** (1986), 360–373.

[2] S. Feferman, *Ordinals and functionals in proof theory*, Actes du Congrès International des Mathématiciens (Nice), Gauthier-Villars, 1971, pp. 229–233.

[3] ______, *Theories of finite type related to mathematical practice*, Handbook of Mathematical Logic (J. Barwise, ed.), North-Holland, 1977, pp. 913–971.

[4] ______, *Notes on operational set theory, I. Generalization of "small" large cardinals in classical and admissible set theory*, `http://math.stanford.edu/~feferman/papers/OperationalST-I.pdf`, 2001.

[5] ______, *Operational set theory and small large cardinals*, `http://math.stanford.edu/~feferman/papers/ostcards.pdf`, 2006.

[6] S. Feferman and W. Sieg, *Proof-theoretic equivalences between classical and constructive theories for analysis*, Iterated Inductive Definitions and Subsystems of Analysis: Recent Proof-Theoretical Studies (W. Buchholz, S. Feferman, W. Pohlers, and W. Sieg), Lecture Notes in Mathematics, vol. 897, Springer, 1981, pp. 78–142.

[7] H. Friedman, *Iterated inductive definitions and* Σ^1_2-AC, Intuitionism and Proof Theory, Proceedings of the Summer Conference at Buffalo, New York, 1968 (A. Kino, J. Myhill, and R.E. Vesley, eds.), Studies in Logic and the Foundations of Mathematics, North-Holland, 1970, pp. 435–442.

[8] J.-Y. Girard, *Proof Theory and Logical Complexitiy*, Studies in Proof Theory, Bibliopolis, 1987.

[9] G. Jäger, *Beweistheorie von* KPN, Archiv für mathematische Logik und Grundlagenforschung **20** (1980), 53–64.

[10] ______, *On Feferman's operational set theory* OST, Annals of Pure and Applied Logic **150** (2007), 19–39.

[11] ______, *Full operational set theory with unbounded existential quantification and power set*, Annals of Pure and Applied Logic **160** (2009), 33–52.

[12] ______, *Operations, sets and classes*, Logic, Methodology, and Philosophy of Science; Proceedings of the Thirteenth International Congress (C. Glymour, W. Wei, and D. Westerståhl, eds.), College Publications, King's College London, 2009, pp. 74–96.

[13] G. Jäger and T. Strahm, *Upper bounds for metapredicative Mahlo in explicit mathematics and admissible set theory*, Journal of Symbolic Logic **66** (2001), 935–958.

[14] A. Levy, *The role of classes in set theory*, Sets and Classes. On the Work by Paul Bernays (G.-H. Müller, ed.), Studies in Logic and the Foundations of Mathematics, North-Holland, 1976, pp. 277–323.

[15] K. Schütte, *Proof Theory*, Springer, 1977.

[16] W.W. Tait, *Normal derivability in classical logic*, The Syntax and Semantics of Infinitary Languages (J. Barwise, ed.), Lecture Notes in Mathematics, vol. 72, Springer, 1968, pp. 204–236.

[17] G. Takeuti, *Proof Theory*, Studies in Logic and the Foundations of Mathematics, North-Holland, 1987.

An Extended Predicative Definition of the Mahlo Universe

Reinhard Kahle and Anton Setzer *

CENTRIA and Departamento de Matemática
Universidade Nova de Lisboa
P–2829-516 Caparica, Portugal
kahle@mat.uc.pt and Department of Computer Science
University of Wales Swansea
Singleton Park
Swansea SA2 8PP, UK
a.g.setzer@swan.ac.uk

Abstract In this article we develop a Mahlo universe in Explicit Mathematics using extended predicative methods. Our approach differs from the usual construction in type theory, where the Mahlo universe has a constructor that refers to all total functions from families of sets in the Mahlo universe into itself; such a construction is, in the absence of a further analysis, impredicative. By extended predicative methods we mean that universes are *constructed from below*, even if they have impredicative characteristics.

1 Predicativity[1]

After the discovery of set theoretic paradoxes at the end of the 19th and beginning of the 20th century, especially Burali-Forti's [BF97] and Russell's (1901, [Rus02]), RUSSELL [Rus06] introduced in 1906 the notion of *predicativity*. POINCARÉ (1906, [Poi06]) made this notion more precise and proposed a foundation of mathematics, which is entirely based on *predicative* constructions. A concept is called predicative, if its definition only refers to concepts introduced before and therefore does not presuppose its own existence. Many mathematical notions are introduced impredicatively. The most prominent example is the set of real numbers defined as Dedekind cuts.

*Both authors were supported by the Volkswagen-Stiftung (RiP program at Oberwolfach). The first author was partially supported by the ESF research project *Dialogical Foundations of Semantics* within the ESF Eurocores program *LogICCC* (funded by the Portuguese Science Foundation, FCT LogICCC/0001/2007). The second author was supported by EPSRC grant EP/G033374/1, *Theory and applications of induction-recursion*.

[1] This historic introduction is partially based on [Fef05].

HERMANN WEYL (1918, [Wey18]) was the first to carry out a systematic development of predicative mathematics. But it soon turned out that significant parts of established mathematics could not be developed using predicative methods. KREISEL [Kre60] proposed in 1958 that ramified analysis RA^*, autonomously iterated, should be considered as the limit of predicative analysis. Using proof theoretic methods KURT SCHÜTTE [Sch65b, Sch65a] and SOLOMON FEFERMAN [Fef64] determined (independently, in 1964-5) Γ_0 as the autonomous ordinal of RA^*. (See also SCHÜTTE's book [Sch77, p. 220] for an excellent presentation and discussion of this result.) Therefore, in proof theory Γ_0 is usually considered as the *limit of predicativity*. Because of this result, predicative analysis is rather weak compared to other, more commonly used mathematical theories (e.g., Zermelo-Fraenkel set theory or full analysis). Already the first substantially impredicative theory ID_1 has a proof theoretic ordinal which is substantially stronger than Γ_0.

Before moving beyond Γ_0, one should note that the results of reverse mathematics show that a substantial portion of ordinary "mathematical theorems" can be proven in the theory ATR_0, Arithmetical Transfinite Recursion, a theory of strength Γ_0, i.e., a theory which is predicative in the proof theoretic sense (see e.g. [Sim99]). However, some mathematical theorems require an extension of ATR_0, called $(\Pi^1_1\text{-}\mathsf{CA})_0$, which (from a proof theoretic perspective) is substantially impredicative (it has the strength of finitely iterated inductive definitions $\mathsf{ID}_{<\omega}$).

For theories whose proof theoretic ordinal is greater than Γ_0, but which can nonetheless be *analysed* using predicative methods (especially without the use of *collapsing functions*), GERHARD JÄGER introduced the notion of *metapredicativity*. The first metapredicative treatment is [Jäg80], the first published metapredicative treatments are [JKSS99] and [Str99].

One should note that there are different understandings of what can be considered as predicative. For instance, in Martin-Löf type theory, inductive and inductive-recursive definitions (the latter allows to define strictly positive universes) are in general considered as predicative, referring to an intuitive understanding of what is meant by a least set closed under certain monotone operators. With inductive-recursive definitions one reaches the strength of KPM ([DS03], Theorem 6.4.2 and Corollary 6.4.3). A Mahlo Universe has been proposed by the second author in [Set00] as a predicatively justified extension of Martin-Löf Type Theory that goes beyond even KPM. In this article we explain how a Mahlo universe can in fact be considered as a predicative construction.

The other extreme position regarding predicativity is the observation that the natural numbers as defined in Peano Arithmetic can be considered as impredicative: they are defined as the least set closed under zero and successor, where "least" is characterized by the induction principle, which refers to the totality of the natural numbers. So the natural numbers are defined by referring to the totality of natural numbers. See EDWARD NELSON [Nel86], DANIEL LEIVANT [Lei94, Lei95], and CHARLES PARSONS [Par92], where PARSONS refers this to an observation by MICHAEL DUMMETT (no citation given).

In this article we introduce an extended predicative version of the *Mahlo universe* in the context of *Explicit Mathematics*. The corresponding theory is impredicative using the proof theoretic understanding (i.e., it goes beyond Γ_0; we expect it to even exceed slightly the strength of KPM). A Mahlo universe M is usually defined as, roughly speaking, a collection of sets such that for every function $f : \mathsf{M} \to \mathsf{M}$ there exist a subuniverse $\mathsf{sub}\, f$ of the Mahlo universe closed under f which is an element of the Mahlo universe. Closure under f means that $f : \mathsf{sub}\, f \to \mathsf{sub}\, f$. This definition of M is impredicative, since it refers to the set of total functions from M into itself, which refers to the totality of M.

Our goal is to introduce the Mahlo universe "from below" so that the *definition* has an *extended predicative character*. For this we will refer to the collection of *arbitrary, (possibly) partial functions* (which is unproblematic from a predicative point of view). This collection is not directly available in *Martin-Löf Type Theory* but in *Explicit Mathematics*, a framework developed by SOLOMON FEFERMAN and further explored by the group of GERHARD JÄGER. Therefore we develop the extended predicative Mahlo universe within the framework of Explicit Mathematics.

2 Mahloness

Mahlo cardinals were introduced 1911 by PAUL MAHLO [Mah11, Mah12]. Mahlo cardinals are the first substantial step in the development of large cardinals beyond inaccessible cardinals (weakly inaccessible cardinals were introduced 1908 by FELIX HAUSDORFF [Hau08]). A (weakly) Mahlo cardinal is a cardinal κ which is (weakly) inaccessible and such that the set of (weakly) inaccessible cardinals less than κ is stationary in κ, i.e., every closed unbounded set in κ contains a (weakly) inaccessible cardinal.

For the proof-theoretic analysis of subsystems of analysis, proof theory makes extensive use of the *recursive analogues* of large cardinals ([Poh96, Poh98]). The recursive analogue of a regular cardinal is an admissible or recursively regular ordinal κ, which is an ordinal closed under all κ-partial recursive functions. (See [Hin78], Def. VIII.2.1). Recursively inaccessible ordinals are recursively regular ordinals κ that are the κ^{th} recursively regular ordinal ([Hin78], Def. VIII.6.1). The recursive analogue of a Mahlo cardinal is a recursively Mahlo ordinal. An admissible ordinal κ is a *recursively Mahlo ordinal* ([Hin78], Def. VIII.6.7) if for all $f : M \to M$, which are M-recursive with parameters in M, there exists an admissible $\kappa < M$ such that $\forall\alpha < \kappa.f(\alpha) < \kappa$. (If one replaces "admissible" by "recursively inaccessible" in this definition, one obtains an equivalent definition.) *Recursively Mahlo sets* are sets of the form L_M for recursively Mahlo ordinals M.

The theory of recursively regular ordinals is often developed in the context of Kripke-Platek set theory KP. KP was introduced by RICHARD PLATEK 1966 in his PhD thesis [Pla66] with a variant introduced independently 1964 by SAUL KRIPKE [Kri64]. The book of JON BARWISE [Bar75] contains an excellent exposition of KP, with the historical background described in Notes I.2.7. In the context of KP, an admissible set ([Bar75], Def. II.1.1) is a transitive set a which is a model of KP, where, apart from closure under pair, union and Δ_0-separation, the main property is closure under Δ_0-collection: If

$$b \in a \ \wedge \ \forall x \in b.\exists y \in a.\varphi(x, y)$$

then there exists $c \in a$ such that

$$\forall x \in b.\exists y \in c.\varphi(x, y)$$

for any Δ_0-formula φ with parameters in a. Recursively inaccessible sets ([Bar75], Def. V.6.7) are admissible sets closed under the operation of stepping to the next admissible set. Recursively Mahlo sets ([Bar75], Exercise. V.7.25) are admissible sets $\mathrm{ad}_{\mathrm{Mahlo}}$ such that for all Δ_0 formulas $\varphi(x, y, \vec{z})$ and variables $\vec{z}$ such that

$$\vec{z} \in \mathrm{ad}_{\mathrm{Mahlo}} \ \wedge \ \forall x \in \mathrm{ad}_{\mathrm{Mahlo}}.\exists y \in \mathrm{ad}_{\mathrm{Mahlo}}.\varphi(x, y, \vec{z})$$

there exists an admissible $b \in \mathrm{ad}_{\mathrm{Mahlo}}$ such that

$$\vec{z} \in b \ \wedge \ \forall x \in b.\exists y \in b.\varphi(x, y, \vec{z})$$

holds. Admissible, recursively inaccessible and recursively Mahlo ordinals are the supremum of the ordinals in an admissible, recursively inaccessible and recursively Mahlo set, respectively. Alternatively they are the ordinals α such that L_α is admissible, recursively inaccessible or recursively Mahlo, respectively.

The step towards an analysis of recursively Mahlo ordinals was an important step in the development of impredicative proof theory. The first step in impredicative proof theory was the analysis of one inductive definition by WILLIAM ALVIN HOWARD ([How72]) based on the Bachmann Ordinal (introduced by HEINZ BACHMANN, [Bac50]). Today, this line of research is continued by two schools in proof theory, one founded by KURT SCHÜTTE (see [Sch77]) and one founded by GAISI TAKEUTI (see [Tak87]). The latter one is based on ordinal diagrams which are closer to Gentzen's original paper [Gen36]. The most productive researcher following this approach is TOSHIYASU ARAI who pushed it beyond $(\Pi^1_2\text{-CA}) + (\text{BI})$ [Ara96a, Ara96b, Ara97a, Ara97b, Ara00a, Ara00b, Ara03, Ara04].

In the other school, iterated inductive definitions were analysed, culminating in a complete analysis in the seminal monograph [BFPS81] by BUCHHOLZ, POHLERS, FEFERMAN and SIEG. With GERHARD JÄGER's dissertation [Jäg79] the focus shifted from the analysis of subsystems of analysis to the analysis of extensions of $\mathsf{KP}\omega$ which allowed a much more fine grained development of intermediate theories. Here $\mathsf{KP}\omega$ is KP plus the existence of the set of natural numbers. This turned out to be very successful with the analysis by WOLFRAM POHLERS and GERHARD JÄGER in 1982 of the equivalent theories KPI, $(\Delta^1_2 - \mathrm{CA}) + (\mathrm{BI})$, and T_0 in [JP82]. Here KPI is $\mathsf{KP}\omega$ plus axioms stating the inaccessibility of the set theoretic universe, $(\Delta^1_2 - \mathrm{CA}) + (\mathrm{BI})$ is the subsystem of analysis with comprehension (CA) restricted to Δ^1_2-formulas which is extended by bar induction BI, and T_0 is a system of explicit mathematics discussed in Sect. 3. The article [JP82] concentrates on the upper bound; the lower bound is based on the embedding of T_0 into $(\Delta^1_2 - \mathrm{CA}) + (\mathrm{BI})$ by FEFERMAN [Fef79] and a well-ordering proof for T_0 by JÄGER [Jäg83]. A more direct well-ordering proof can be found in [BS83] and [BS88] by WILFRIED BUCHHOLZ and KURT SCHÜTTE. The state-of-the-art technique for determining upper bounds is based on the simplified version of local predicativity by BUCHHOLZ [Buc92]. A constructive underpinning was obtained by the second author, by carrying out a proof theoretic analysis of Martin-Löf type theory [Set98], showing that it is slightly stronger than KPI (see as well independent work by MICHAEL RATHJEN and E. GRIFFOR [GR94].)

The first significant step beyond inaccessibles which are in some sense two level inductive definitions, was taken by MICHAEL RATHJEN ([Rat90, Rat91, Rat94a]) with his analysis of KPM, i.e. $\mathsf{KP}\omega$ with the Mahloness of its universe, and a corresponding subsystem of analysis [Rat96]. The second author of this article introduced in [Set00] a Mahlo universe in Martin-Löf type theory and showed that its strength goes slightly beyond that of KPM. This provided a first constructive underpinning of this proof-theoretic development. Later GERHARD JÄGER (e.g. [Jäg05]) introduced a Mahlo universe in Explicit Mathematics $(\mathsf{T}_0(\mathsf{M}))$, which we will revisit in Sect. 4.

The analysis of KPM was the main stepping stone for RATHJEN to jump to an analysis of $\mathsf{KP}\omega$ with Π_3-reflection ([Rat92,Rat94b]) and later of $(\Pi^1_2\text{-CA}) + (\mathrm{BI})$ ([Rat95, Rat05a, Rat05b]).

We look now at the rules and axioms for formulating Mahlo in Explicit Mathematics. There are two versions, *internal Mahlo* $(\mathsf{T}_0(\mathsf{M})^+)$, corresponding to having a universe in Explicit Mathematics having the Mahlo property, and *external Mahlo* $(\mathsf{T}_0(\mathsf{M}))$, corresponding to the fact that the overall collection of sets has the Mahlo property. We first focus on the internal Mahlo universe, and then indicate how to modify this in order to obtain the external Mahlo universe.

The first part is that a recursively Mahlo set is a recursively inaccessible set (remember that we could replace admissibles by recursively inaccessible sets). Recursively inaccessible sets correspond to universes closed under inductive generation, so in $\mathsf{T}_0(\mathsf{M})^+$ we demand for some constant M corresponding to the recursively Mahlo set $\mathrm{ad}_{\mathrm{Mahlo}}$ that it is a universe which is closed under inductive generation (which would correspond in type theory to closure under the W-type, in subsystems of analysis to the formation of inductively defined sets, and in KP to the formation of the next admissible above a given set). We note here that the metapredicative versions are obtained by omitting inductive generation—which is an impredicative concept in the proof theoretic sense. Thus, for metapredicative Mahlo, closure under inductive generation is omitted.

The assumption for the main closure property of $\mathrm{ad}_{\mathrm{Mahlo}}$ is $\vec{z} \in \mathrm{ad}_{\mathrm{Mahlo}}$ and $\forall x \in \mathrm{ad}_{\mathrm{Mahlo}}.\exists y \in \mathrm{ad}_{\mathrm{Mahlo}}.\varphi(x, y, \vec{z})$. We can collect the elements $\vec{z}$ together into one set a and replace the closure under φ by a function $f \in (\mathsf{M} \to \mathsf{M})$.

The reader with a background in Martin-Löf Type Theory might wonder why this is sufficient, since in type theory this assumption is translated as having a function $f \in (\mathsf{Fam}(\mathsf{M}) \to \mathsf{Fam}(\mathsf{M}))$, where[2]

$$\mathsf{Fam}(u) := \{(a, b) \mid a \,\dot{\in}\, u \;\wedge\; b \in (a \to u)\} \ .$$

The reason why this can be avoided is that for any universe u we can write encoding functions $\mathsf{pair} \in (\mathsf{Fam}(u) \to u)$ and decoding functions $\mathsf{proj}_0 \in (u \to u)$ and $\mathsf{proj}_1 \in ((x \,\dot{\in}\, u) \to \mathsf{proj}_0\, x \to u)$ for families of sets such that for $a \,\dot{\in}\, u$ and $b \in (a \to u)$ we have $\mathsf{proj}_0\ (\mathsf{pair}\ (a, b)) \doteq a$ and $\mathsf{proj}_1\ (\mathsf{pair}\ (a, b)) \doteq b$. We use here notations inherited from dependent type theory, $\mathsf{proj}_1 \in ((x \,\dot{\in}\, u) \to \mathsf{proj}_0\, x \to u)$ means that proj_1 is a defined constant such that

$$\forall x \,\dot{\in}\, u.\forall y \,\dot{\in}\, \mathsf{proj}_0\, x.\mathsf{proj}_1\, x\, y \,\dot{\in}\, u \ .$$

For this one defines (using join and arithmetic comprehension)

$$\begin{aligned} \mathsf{pair}\ (a, b) &:= \{(0, x) \mid x \,\dot{\in}\, a\} \cup \{(1, (x, y)) \mid x \,\dot{\in}\, a \;\wedge\; y \,\dot{\in}\, b\, x\} \ , \\ \mathsf{proj}_0\ a &:= \{x \mid (0, x) \,\dot{\in}\, a\} \ , \\ \mathsf{proj}_1\ a\ x &:= \{y \mid (1, (x, y)) \,\dot{\in}\, a\} \ . \end{aligned}$$

Now a function $f \in (\mathsf{Fam}(u) \to \mathsf{Fam}(u))$ can be encoded as a function $g \in (u \to u)$ s.t. $g\ x = \mathsf{pair}\ (f\ (\mathsf{proj}_0\ x, \mathsf{proj}_1\ x))$, and a universe u is closed under f if and only if it is closed under g (modulo $\doteq$). In the same way we can replace

[2] The notations $\dot{\in}$, $\doteq$, $\dot{\subset}$, $\Re$ and related notions are introduced in Section 3, which introduces as well the theory T_0.

$\vec{z}$ occurring above, which would be translated into an element of $\mathsf{Fam}(u)$, by one single element of u.

Assuming the closure of $\mathrm{ad}_{\mathrm{Mahlo}}$ under $\vec{z}$ and φ the recursively Mahlo property gave us the existence of a recursively inaccessible b containing $\vec{z}$ and closed under φ. The existence of b translates into the existence of a subuniverse $\mathsf{m}\,(a, f)$. So we have $\mathsf{m}\,(a, f)$ is a universe, $\mathsf{m}\,(a, f) \subseteq \mathsf{M}$. (Note that in type theory an explicit embedding from $\mathsf{m}\,(a, f)$ into M needs to be defined, which we can avoid in Explicit Mathematics because there universes are à la Russell rather than à la Tarski). $\vec{z} \in b$ translates into $a \mathrel{\dot{\in}} \mathsf{m}\,(a, f)$ and that $\mathrm{ad}_{\mathrm{Mahlo}}$ is closed under φ is translated into $f \in (\mathsf{m}\,(a, f) \to \mathsf{m}\,(a, f))$. (In type theory it was necessary to introduce a constructor reflecting f in $\mathsf{m}\,(a, f)$, which is implicit in Explicit Mathematics. Furthermore, in the formulation of the Mahlo universe in [Set00] the parameter a doesn't occur. This is because closure under a can be avoided by replacing closure under f by closure under g such that $g\,x$ is the union of $f\,x$ and a.)

Universes in Explicit mathematics are usually not closed under inductive generation, and we follow this convention. We observe that M needs in addition to being a universe to be closed under inductive generation. However, $\mathsf{m}\,(a, f)$ does not need to be closed under inductive generation: We can use again the trick of encoding of families of sets into sets and define for every $f \in (\mathsf{M} \to \mathsf{M})$ a function $g \in (\mathsf{M} \to \mathsf{M})$ such that u is closed under g if u is closed under f and inductive generation (modulo $\doteq$). So we obtain that, even if $\mathsf{m}\,(a, f)$ is not necessarily closed under inductive generation, there still exists for every $f \in (\mathsf{M} \to \mathsf{M})$ a subuniverse of M closed under f and inductive generation (modulo $\doteq$).

Up to now, the strength of the rules does not exceed possessing T_0 plus the existence of one universe, since we could easily model $\mathsf{m}\,(a, f) := \mathsf{M}$. What is still missing is to model that the admissible is an element of M, which is modelled by

$$\mathsf{m}\,(a, f) \mathrel{\dot{\in}} \mathsf{M}$$

Note that this means that M has a constructor that depends negatively on M, namely

$$\mathsf{m} \in ((\mathsf{M}, (\mathsf{M} \to \mathsf{M})) \to \mathsf{M})$$

This completes the internal version of the Mahlo universe, which can be summarized as follows (notations such as $\mathcal{U}(t)$ will be explained in the next section):

$$\begin{array}{l} \mathcal{U}(\mathsf{M}) \ \wedge\ \mathsf{i} \in (\mathsf{M}^2 \to \mathsf{M}) \\ \quad a \mathrel{\dot{\in}} \mathsf{M} \wedge f \in (\mathsf{M} \to \mathsf{M}) \to \mathsf{m}\,(a, f) \mathrel{\dot{\subset}} \mathsf{M} \ \wedge\ \mathcal{U}(\mathsf{m}\,(a, f)) \wedge a \mathrel{\dot{\in}} \mathsf{m}\,(a, f) \\ \quad a \mathrel{\dot{\in}} \mathsf{M} \wedge f \in (\mathsf{M} \to \mathsf{M}) \to f \in (\mathsf{m}\,(a, f) \to \mathsf{m}\,(a, f)) \ \wedge\ \mathsf{m}\,(a, f) \mathrel{\dot{\in}} \mathsf{M} \end{array}$$

An external Mahlo universe is obtained by giving the collection $\Re$ of names for sets in Explicit Mathematics the rôle of M. So we obtain as conditions the axioms developed by JÄGER (in addition to T_0 which contains closure of $\Re$ under i):

$$\Re(a) \wedge f \in (\Re \to \Re) \to \mathcal{U}(\mathsf{m}\,(a,f)) \wedge a \mathrel{\dot\in} \mathsf{m}\,(a,f),$$
$$\Re(a) \wedge f \in (\Re \to \Re) \to f \in (\mathsf{m}\,(a,f) \to \mathsf{m}\,(a,f)).$$

3 Explicit Mathematics

We work in the framework of Feferman's *Explicit Mathematics*, [Fef75, Fef79]. It was introduced in the 1970s to formalize BISHOP-style constructive mathematics.

Explicit Mathematics is based on a two-sorted language, comprising *individuals* (combinatory logic plus additional constants) and *types* (i.e., collections of individuals). As a general convention, individual constants are given as lower case letters (or letter combinations) in sans serif font, individual variables as roman lower case letters, such as x, y, individual terms as roman lower case letters such as r, s, t, and type variables in roman upper case letters such as U, V, X, Y (we do not use type constants). Types are *named* by individuals, which are formally expressed by a *naming relation* $\Re(x, U)$, and one has an axiom expressing that every type has a name:

$$\forall U.\exists x.\Re(x, U).$$

Based on the primitive element relation $t \in X$, it is convenient to introduce the following abbreviations:

$$\begin{aligned}
\Re(s) &:= \exists X.\Re(s, X),\\
s \mathrel{\dot\in} t &:= \exists X.\Re(t, X) \wedge s \in X,\\
\exists x \mathrel{\dot\in} s.\varphi(x) &:= \exists x.x \mathrel{\dot\in} s \wedge \varphi(x),\\
\forall x \mathrel{\dot\in} s.\varphi(x) &:= \forall x.x \mathrel{\dot\in} s \to \varphi(x),\\
s \mathrel{\dot\subset} t &:= \forall x \mathrel{\dot\in} s.x \mathrel{\dot\in} t,\\
s \doteq t &:= s \mathrel{\dot\subset} t \wedge t \mathrel{\dot\subset} s,\\
\Re_{\Re}(s) &:= \Re(s) \wedge \forall x \mathrel{\dot\in} s.\Re(x),\\
f \in (\Re \to \Re) &:= \forall x.\Re(x) \to \Re(f\,x),\\
f \in (s \to s) &:= \forall x.x \mathrel{\dot\in} s \to f\,x \mathrel{\dot\in} s,\\
f \in (s^2 \to s) &:= \forall x, y.x \mathrel{\dot\in} s \wedge y \mathrel{\dot\in} s \to f\,(x, y) \mathrel{\dot\in} s.
\end{aligned}$$

The usual starting point of Explicit Mathematics is the theory EETJ of *explicit elementary types with join*, cf. [FJ96]. It is based on Beeson's classical *logic of partial terms* (see [Bee85] or [TvD88]) for individuals and classical logic for types.

The first order part is given by *applicative theories* which formalize partial combinatory algebra, pairing and projection, and axiomatically introduced natural numbers, cf. [JKS99]. EETJ adds types on the second order level, and axiomatize *elementary comprehension* and *join* as type construction operations. We dispense here with a detailed description of EETJ which can be found in many papers on Explicit Mathematics (e.g., [JKS01], [JS02] or [Kah07]). Let us just briefly address the finite axiomatization of elementary comprehension and join. For these, we have the following individual constants in the language: nat (natural numbers), id (identity), co (complement), int (intersection), dom (domain), inv (inverse image), and j (join). These constants together make up a set of *generators*, to which also belong—depending on the particular theory under consideration—other constants used to introduce names, such as i (inductive generation) in T_0 or m in the approaches to Mahlo; for the extended predicative version we have also the additional generators M, pre and sub. From the axiomatization we just give as an example the one for *intersections*:

$$\Re(a) \wedge \Re(b) \rightarrow \Re(\mathsf{int}\,(a,b)) \wedge \forall x. x \mathrel{\dot{\in}} \mathsf{int}\,(a,b) \leftrightarrow x \mathrel{\dot{\in}} a \wedge x \mathrel{\dot{\in}} b.$$

The generators for elementary comprehension and join will appear again below when we define the notion of universe in Explicit Mathematics as a type which is closed under elementary comprehension and join.

3.1 Inductive Generation

Let us shortly address the most famous theory of Explicit Mathematics, T_0 [Fef75], which is obtained from EETJ by adding *inductive generation* and the standard induction scheme on natural numbers for arbitrary formulae of the language. Using the abbreviation

$$\mathsf{Closed}(a,b,S) := \forall x \mathrel{\dot{\in}} a.(\forall y \mathrel{\dot{\in}} a.(y,x) \mathrel{\dot{\in}} b \rightarrow y \in S) \rightarrow x \in S$$

inductive generation is given by the following two axioms, expressing that $\mathsf{i}\,(a,b)$ is the least fixed point of the operator $X \mapsto \mathsf{Closed}(a,b,X)$, or the accessible part of the relation b restricted to a:[3]

(IG.1) $\Re(a) \wedge \Re(b) \rightarrow \exists X. \Re(\mathsf{i}\,(a,b), X) \wedge \mathsf{Closed}(a,b,X)$,

(IG.2) $\Re(a) \wedge \Re(b) \wedge \mathsf{Closed}(a,b,\varphi) \rightarrow \forall x \mathrel{\dot{\in}} \mathsf{i}\,(a,b).\varphi(x)$.

[3] Formulas such as $\mathsf{Closed}(a,b,\varphi)$ are to be understood in the obvious way (replace in $\mathsf{Closed}(a,b,S)$ formulas $t \in S$ by $\varphi(t)$). This convention will apply later even to formulas where a name for a set a is replaced by φ – then $s \mathrel{\dot{\in}} a$ is to be replaced by $\varphi(s)$.

As mentioned before, the theory T_0 played an important role in the proof-theoretic analysis of the proof theoretically equivalent theories $(\Delta^1_2 - \mathrm{CA}) + (\mathrm{BI})$ and KPI (see [Fef79, Jäg83]); since T_0 has the same strength as KPI, one can say that inductive generation is a way of formalizing *inaccessibility* in Explicit Mathematics, and formalizing it "from below".

3.2 Universes

We now turn to the notion of *universes* as discussed, for instance, in [JKS01]. In the context of Mahloness, universes are considered by JÄGER, STRAHM, and STUDER [JS01, JS02, Str02, Jäg05, JS05].

The concept of *universes* can be introduced as a defined notion: A universe is a type W such that:

1. all elements of W are names and
2. W is closed under elementary comprehension and join.

For the formal definition we introduce the auxiliary notation of the closure condition $\mathcal{C}(W, a)$ as the disjunction of the following formulas:

(1) $a = \mathsf{nat} \vee a = \mathsf{id}$,

(2) $\exists x.a = \mathsf{co}\, x \wedge x \in W$,

(3) $\exists x.\exists y.a = \mathsf{int}\,(x, y) \wedge x \in W \wedge y \in W$,

(4) $\exists x.a = \mathsf{dom}\, x \wedge x \in W$,

(5) $\exists f.\exists x.a = \mathsf{inv}\,(f, x) \wedge x \in W$,

(6) $\exists x.\exists f.a = \mathsf{j}\,(x, f) \wedge x \in W \wedge \forall y \,\dot{\in}\, x.f\, y \in W$.

The formula $\forall x.\mathcal{C}(W, x) \to x \in W$ expresses that W is a type closed under the type constructions of EETJ, i.e., elementary comprehension and join. Now, we define a universe as a collection of names which satisfies this closure condition, and we write $\mathsf{U}(W)$ to express that W is a *universe*:

$$\mathsf{U}(W) := (\forall x \in W.\Re(x)) \wedge \forall x.\mathcal{C}(W, x) \to x \in W \quad .$$

We write $\mathcal{U}(t)$ to express that t is a *name of a universe*:

$$\mathcal{U}(t) := \exists X.\Re(t, X) \wedge \mathsf{U}(X).$$

A detailed discussion of the concept of universes in Explicit Mathematics can be found in [JKS01], including *least universes* and *name induction*. Universes can be considered as a formalization of *admissibility*. However, since, if one adds induction axioms expressing *least universes* or *name induction*, one reaches *inacessibilty*, they can serve as alternatives to inductive generation in T_0.

4 Axiomatic Mahlo

The first formulation of Mahlo in Explicit Mathematics was given in a metapredicative setting by JÄGER and STRAHM [JS01]. Its proof theoretic strength was determined in [Str02] (with the upper bound given in [JS01]) as $\varphi\, \varepsilon_0\, 0\, 0$ (with induction restricted to types the strength is $\varphi\, \omega\, 0\, 0$). The non-metapredicative version, which is obtained by adding inductive generation, was studied by JÄGER and STUDER [JS02]. The resulting theory $\mathsf{T}_0(\mathsf{M})$ (Explicit Mathematics with Mahlo) is defined as the extension of T_0 by the following two axioms:

(M1) $\qquad \Re(a) \wedge f \in (\Re \to \Re) \to \mathcal{U}(\mathsf{m}\,(a,f)) \wedge a \,\dot{\in}\, \mathsf{m}\,(a,f),$

(M2) $\qquad \Re(a) \wedge f \in (\Re \to \Re) \to f \in (\mathsf{m}\,(a,f) \to \mathsf{m}\,(a,f)).$

The axioms state that for every function from names to names there is a universe which is closed under f. This universe is defined uniformly in f by use of the universe constructor m.

An overview over what is known about $\mathsf{T}_0(\mathsf{M})$ can be found in JÄGER's article [Jäg05]. Together with THOMAS STUDER [JS02] he determined an upper bound for the proof theoretic strength of Explicit Mathematics with impredicative Mahlo, using specific nonmonotone inductive definitions introduced by RICHTER [Ric71], see also [Jäg01]. A lower bound can be combined according to JÄGER [Jäg05] by using the realization of CZF with Mahloness into Explicit Mathematics with the Mahlo universe (SERGEI TUPAILO [Tup03]) together with a not-worked out adaption of the well-ordering proof by MICHAEL RATHJEN [Rat94a] for KPM:[4]

Theorem 4.1. $\mathsf{T}_0(\mathsf{M}) \equiv \mathsf{KPM}$ *and the proof-theoretic ordinal is* $\mathbf{\Psi}_\Omega(\varepsilon_{M_0+1})$.

The axiomatization of the universe $\mathsf{m}\,(a,f)$ for a given function f (and given name a) is *impredicative* in the following sense: f is assumed to be a total function from names to names but this totality has to hold, of course, also with respect to

[4]The second author regards the latter as a good hint why this theorem is true, but details in well-ordering proofs can be quite tricky and more details need to be worked out before we can regard this result as a full theorem.

the name of the "newly introduced" universe $\mathsf{m}\,(a, f)$. In other words, in order to verify the premise $f \in (\Re \to \Re)$ one already needs to "know" $\mathsf{m}\,(a, f)$.

We call this approach to Mahlo universes *axiomatic*.

JÄGER and STUDER, in [JS02], also consider a variant of $\mathsf{T}_0(\mathsf{M})$ which is based on partial functions, partial with respect to the definedness predicate of the underlying applicative theory. It is easy to see from the model construction that this does not change the proof-theoretic strength. Note that, when we speak about partiality of functions in the following, we have something else in mind, namely that there are no "a priori" conditions given on the behaviour of a function outside of the subuniverse under consideration.

In the given form, $\mathsf{T}_0(\mathsf{M})$ axiomatizes an "external" Mahlo universe, in the sense that the "universe" of all names—the extension of $\Re$—has the Mahlo property. However, the collection of all names is not a universe in the defined sense of the theory.

TUPAILO [Tup03, p. 172, IX] also considers an extension of T_0, which he called $\mathsf{T}_0 + \mathsf{M}^+$, which formalizes an "internal" Mahlo universe, i.e., there is a universe—in the sense defined within the theory—, named by M which has the Mahlo property. We formalise a variant $\mathsf{T}_0(\mathsf{M})^+$ which consists of the axioms of T_0 plus the following axioms:

$$
\begin{array}{ll}
(\mathsf{M}^+1) & \mathcal{U}(\mathsf{M}) \;\wedge\; \mathsf{i} \in (\mathsf{M}^2 \to \mathsf{M}), \\
(\mathsf{M}^+2) & a \mathrel{\dot\in} \mathsf{M} \wedge f \in (\mathsf{M} \to \mathsf{M}) \to \mathsf{m}\,(a, f) \mathrel{\dot\subset} \mathsf{M}, \\
(\mathsf{M}^+3) & a \mathrel{\dot\in} \mathsf{M} \wedge f \in (\mathsf{M} \to \mathsf{M}) \to \mathcal{U}(\mathsf{m}\,(a, f)) \wedge a \mathrel{\dot\in} \mathsf{m}\,(a, f) \\
 & \qquad\qquad\qquad\qquad\qquad \wedge\; f \in (\mathsf{m}\,(a, f) \to \mathsf{m}\,(a, f)), \\
(\mathsf{M}^+4) & a \mathrel{\dot\in} \mathsf{M} \wedge f \in (\mathsf{M} \to \mathsf{M}) \to \mathsf{m}\,(a, f) \mathrel{\dot\in} \mathsf{M}
\end{array}
$$

We note some differences to the axioms of $\mathsf{T}_0 + \mathsf{M}^+$ given by TUPAILO:

- In $\mathsf{T}_0 + \mathsf{M}^+$, one has the *limit* operator u which gives (the name of) the next universe above a given name (see [Kah97]). Now, M is also closed under this operator: $\mathsf{u} : \mathsf{M} \to \mathsf{M}$. This is not necessary, since using m we can define easily for every universe a universe on top of it $\mathsf{m}(a, \lambda x.x)$ (see also [JS02, Sect. 6]).

- Also, $\Re$ is closed under the *limit* operator u. Since universes are not closed under inductive generation, adding u most likely doesn't add any strength to it. This is at least the case without the Mahlo universe: At the end of Sect. 4 in [JS02] as a consequence of a sophisticated model construction an outline of the argument is given, why adding closure under u to T_0 doesn't increase its proof-theoretic strength.

- $\mathsf{T}_0 + \mathsf{M}^+$ has no parameter a of m, so m only depends on $f \dot{\in} (\mathsf{M} \to \mathsf{M})$. This doesn't make any difference, since we can define for every $a \dot{\in} \mathsf{M}$ and $f \in (\mathsf{M} \to \mathsf{M})$ a $g : \mathsf{M} \to \mathsf{M}$ such that a universe is closed under g if and only if it is closed under f and a. (In Sect. 2 we showed how to encode a family of sets into a set such that a universe contains the code for the family if it contains the index and the elements of the family. We can do the same trick and encode two sets into one. Now let $g\,x$ be the code for the two sets $f\,x$ and a, and use the fact that universes are non-empty.)

- $\mathsf{T}_0 + \mathsf{M}^+$ doesn't demand $\mathsf{m}\,(a, f) \dot{\subset} \mathsf{M}$. In this respect, $\mathsf{T}_0(\mathsf{M})^+$ seems to be slightly stronger. However, any standard model used for determining an upper bound will fulfil this condition, and the well-ordering proof shouldn't make use of it, therefore this condition should not add any proof theoretic strength to the theory. However, we believe that having this axiom is more aesthetically appealing, since $\mathsf{m}\,(a, f)$ should be a subuniverse of M.

For the extended predicative version of Mahlo, we formalize an *internal* Mahlo universe corresponding to $\mathsf{T}_0(\mathsf{M})^+$.

5 Extended Predicative Mahlo

We aim to introduce new universes "from below": given a "potential Mahlo universe", i.e., a universe which should have the Mahlo property, we will enlarge this universe "carefully" by stages such that we get the desired property. The key difference between this approach compared to the axiomatic approach above is that we will not assume that f is a total function from names to names, but we will assume that it is total on the *subuniverse* which should be closed under f.

5.1 Relative f-Pre-Universe

For a given universe v—which is to be extended to a Mahlo universe—a name a and a given (arbitrary, possibly *partial*) function f we first define what it means that u is (the name of) a *pre-universe*, containing a, closed under f relative to v.

$$\mathsf{RPU}(a, f, u, v) := (\forall x. \mathcal{C}(u, x) \wedge x \dot{\in} v \to x \dot{\in} u) \wedge \tag{5.1}$$

$$(a \dot{\in} v \to a \dot{\in} u) \wedge \tag{5.2}$$

$$(\forall x \in u. f\,x \in v \to f\,x \in u) \tag{5.3}$$

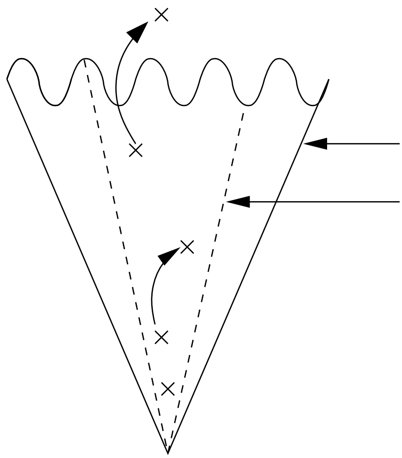

Figure 1: A pre-universe

Thus, for given a, f, and v, a pre-universe u has the following properties:

- u is closed under the generators of EETJ, as long as the generated names are in v (5.1);
- if a is an element of v, it is an element of u (5.2);
- if f maps an element x of u to an element of v, then $f\,x$ is in u; i.e., $f\,x$ cannot be in v but outside u (5.3).

Figure 1 illustrates a pre-universe. We see that a and $f\,b$ are included in u, since they are in v. $f\,c$ is not (yet) in v, so it is not included in u.

From a foundational point of view, this is a well-understood predicative inductive definition and we can introduce a straightforward *induction principle* to obtain *least f-pre-universes*. Using the new generator pre to name a pre-universe u, a least f-pre-universe $\mathsf{pre}\,(a, f, v)$ is characterized by the following axioms:

I. Least f-pre-universes

(EPM.1) $\Re_{\Re}(v) \to \mathsf{RPU}(a, f, \mathsf{pre}\,(a, f, v), v)$.

(EPM.2) $\Re_{\Re}(v) \wedge \mathsf{RPU}(a, f, \varphi, v) \to \forall x \mathrel{\dot\in} \mathsf{pre}\,(a, f, v).\varphi(x)$.

With (EPM.2) one gets immediately: $\Re_{\Re}(v) \to \mathsf{pre}\,(a, f, v) \mathrel{\dot\subset} v$.

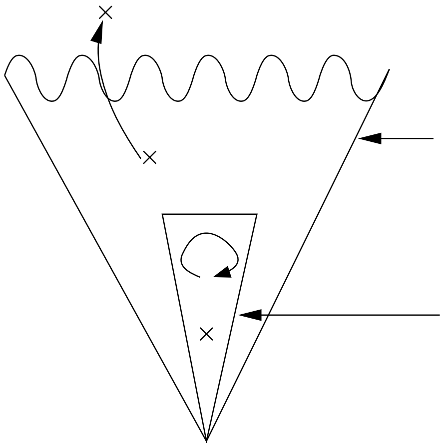

Figure 2: $\mathsf{Indep}(a, f, u, v)$

5.2 Independence

The f-pre-universes are defined relative to v; what we want is, of course, universes that no longer depend on v. Formally, we can express this *independence* by a formula $\mathsf{Indep}(a, f, u, v)$ which expresses that the "relativization to v" in the closure condition of $\mathsf{RPU}(a, f, u, v)$ is already fulfilled:

$$\begin{aligned}\mathsf{Indep}(a, f, u, v) := {} & (\forall x.\mathcal{C}(u, x) \to x \in v) \\ & \wedge a \mathrel{\dot\in} v \\ & \wedge (\forall x \mathrel{\dot\in} u. f\, x \mathrel{\dot\in} v)\end{aligned}$$

Figure 2 illustrates what it means for u to be independent of v, in case $u = \mathsf{pre}\,(a, f, v)$: $a \mathrel{\dot\in} v$ and therefore $a \mathrel{\dot\in} u$; for $a \mathrel{\dot\in} u$ we have $f\, a \mathrel{\dot\in} v$ and therefore $f\, a \mathrel{\dot\in} u$, so u is closed under f. How f operates outside u does not really matter: it is possible that for some $b \mathrel{\dot\in} v$ we have $f\, b \mathrel{\dot\notin} v$.

The following lemma follows now directly from the definitions:

Lemma 5.1.

$$\begin{aligned}&\Re(u) \;\wedge\; \Re_{\Re}(v) \;\wedge\; \mathsf{RPU}(a, f, u, v) \wedge \mathsf{Indep}(a, f, u, v) \\ &\qquad\qquad \to \mathcal{U}(u) \wedge a \mathrel{\dot\in} u \wedge f \in (u \to u).\end{aligned}$$

Thus, under the condition $\mathsf{Indep}(a, f, \mathsf{pre}\,(a, f, v), v)$, the least f-pre-universes $\mathsf{pre}\,(a, f, v)$ are actually universes. But the main property is that they are now independent of v in the sense that an enlargement of v will not change the extension of $\mathsf{pre}\,(a, f, v)$. This gives them, in fact, their *predicative character*. Formally this property is expressed in the following *extended predicativity lemma*.

Lemma 5.2 (Extended Predicativity)**.** *In* $\mathsf{EETJ} + (\mathsf{EPM.1}) + (\mathsf{EPM.2})$ *we can prove:*

$$\Re(v) \;\wedge\; \Re_{\Re}(w) \wedge \mathsf{Indep}(a, f, \mathsf{pre}\,(a, f, v), v) \wedge v \mathrel{\dot\subset} w \\ \rightarrow \mathsf{pre}\,(a, f, v) \doteq \mathsf{pre}\,(a, f, w).$$

As a corollary we get that an enlargement of v does not influence the independence property considered with respect to the bigger universe.

Corollary 5.3. *In* $\mathsf{EETJ} + (\mathsf{EPM.1}) + (\mathsf{EPM.2})$ *we can prove:*

$$\Re(v) \;\wedge\; \Re_{\Re}(w) \wedge \mathsf{Indep}(a, f, \mathsf{pre}\,(a, f, v), v) \wedge v \mathrel{\dot\subset} w \\ \rightarrow \mathsf{Indep}(a, f, \mathsf{pre}\,(a, f, w), w).$$

5.3 The Mahlo Universe

Intuitively, the idea to build the Mahlo universe is now to enlarge a potential Mahlo universe u and $\mathsf{pre}\,(a, f, u)$ in parallel up to the stage that $\mathsf{pre}\,(a, f, u)$ is independent of u (and, of course, doing this for all a and f). When the preuniverse is complete, it will not depend on any future additions to u.

Thus, axiomatically expressed, the Mahlo universe, named by M, has to be a universe, it has to be closed under inductive generation, and it has to collect, for every f, provided $\mathsf{pre}\,(a, f, \mathsf{M})$ is complete, an element representing $\mathsf{pre}\,(a, f, \mathsf{M})$ to it. Since in this case $\mathsf{pre}\,(a, f, \mathsf{M})$ is independent of M, we introduce a new name $\mathsf{sub}\,(a, f)$ which names the same type as $\mathsf{pre}\,(a, f, \mathsf{M})$, and add this element to M.

Figure 3 illustrates the construction of M: If $\mathsf{pre}\,(a, f, \mathsf{M})$ is independent of M, it contains a and is closed under f; then the name $\mathsf{sub}\,(a, f)$ is added to M (and the addition of $\mathsf{sub}\,(a, f)$ to M doesn't affect the reason for originally adding it to M). Note again, that how f operates outside $\mathsf{pre}\,(a, f, \mathsf{M})$ does not really matter: it is possible that for some $b \mathrel{\dot\in} \mathsf{M}$ we have $f\,b \mathrel{\dot\notin} \mathsf{M}$, i.e., that M is not closed under f.

II. Mahlo universe

$(\mathsf{EPM.3})\qquad \mathcal{U}(\mathsf{M}) \wedge \mathsf{i} \in (\mathsf{M}^2 \rightarrow \mathsf{M}).$

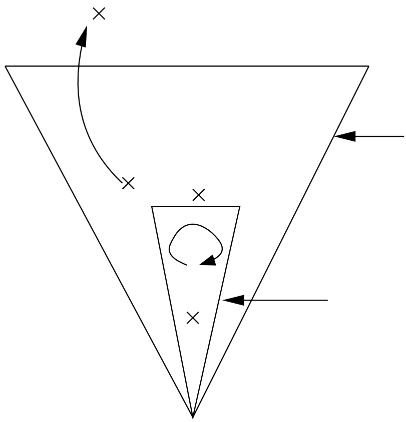

Figure 3: The extended predicative Mahlo universe

(EPM.4) $\quad \mathsf{Indep}(a, f, \mathsf{pre}\,(a, f, \mathsf{M}), \mathsf{M}) \quad \rightarrow \quad \mathsf{sub}\,(a, f) \,\dot{\in}\, \mathsf{M} \quad \wedge \quad \mathsf{sub}\,(a, f) \doteq \mathsf{pre}\,(a, f, \mathsf{M})$.

From (EPM.4) the theory will get its strength: Whenever we have a pre-universe $\mathsf{pre}\,(a, f, \mathsf{M})$, which is independent of M, we will have a name $\mathsf{sub}\,(a, f)$ of this universe in M. Note that by (EPM.1) $\mathsf{pre}\,(a, f, \mathsf{M})$ is already a pre-universe relative to M. Therefore, by Lemma 5.1 the premise of (EPM.4) implies that $\mathsf{pre}\,(a, f, \mathsf{M})$ is in fact a universe which is closed under a and f.

By Lemma 5.2 and Corollary 5.3 we know that independent universes do not depend on the universe used in the last parameter. Using the additional generator sub we can get rid of this redundant dependence in the name of the sub-universe which is actually added to M. More concretely, under the assumption $\mathsf{Indep}(a, f, \mathsf{pre}\,(a, f, \mathsf{M}), \mathsf{M})$ the addition of $\mathsf{sub}\,(a, f)$ to M does not affect the universe named by $\mathsf{pre}\,(a, f, \mathsf{M})$ (or $\mathsf{sub}\,(a, f)$). “Philosophically spoken”, it does not affect the *reason for its addition*.

5.4 M is a Mahlo Universe

To show that M is indeed a Mahlo universe, we interpret $\mathsf{T}_0(\mathsf{M})^+$ into $\mathsf{T}_0 +$ (EPM.1–4). This can be done translating $\mathsf{m}\,(a, f)$ by $\mathsf{sub}\,(a, f)$ and using the following lemma and theorem.

Lemma 5.4.

$$\Re(u) \ \wedge \ \mathcal{U}(v) \ \wedge \ a \,\dot{\in}\, v \ \wedge \ f \in (v \to v) \ \wedge \ u \,\dot{\subset}\, v \ \wedge \ \mathsf{RPU}(a, f, u, v) \\ \to \mathsf{Indep}(a, f, u, v) \ \wedge \ \mathcal{U}(u) \ \wedge \ a \,\dot{\in}\, u \ \wedge \ f \in (u \to u)$$

Theorem 5.5.

$$a \,\dot{\in}\, \mathsf{M} \ \wedge \ f \in (\mathsf{M} \to \mathsf{M}) \\ \to \mathsf{sub}\,(a, f) \,\dot{\in}\, \mathsf{M} \ \wedge \ \mathsf{sub}\,(a, f) \,\dot{\subset}\, \mathsf{M} \\ \wedge \ \mathcal{U}(\mathsf{sub}\,(a, f)) \ \wedge \ a \,\dot{\in}\, \mathsf{sub}\,(a, f) \ \wedge \ f \in (\mathsf{sub}\,(a, f) \to \mathsf{sub}\,(a, f))$$

It is a straightforward exercise to formalise variants of (EPM.1–4) to capture an *extended predicative external Mahlo universe* corresponding to $\mathsf{T}_0(\mathsf{M})$. These axioms might seem no more convincing than the axioms of axiomatic Mahlo, which just express that for every name a and function from names to names we can find a type closed under it. But these axioms are impredicative, since the collection of names has to have those closure princples. An extended predicative version of external Mahlo doesn't have these problems, because the premise for introducing $\mathsf{sub}\,(a, f)$ doesn't require $f \in (\Re \to \Re)$ which would refer to $\mathsf{sub}\,(a, f)$.

Dag Normann has in [Nor99] developed a domain theoretic construction of a Mahlo universe and shown that the closure ordinal is the first recursively Mahlo ordinal. It can be regarded as a domain theoretic construction of an extended predicative external Mahlo universe.

5.5 The Least Mahlo Universe

The addition of (EPM.1–4) to T_0 yields already a theory of Mahloness with an appropriate proof-theoretic strength. However, the specific feature of the given approach is the possibility to axiomatize a *least Mahlo universe*.

For this we observe that, working in a set theoretical model of explicit mathematics, the extended predicative Mahlo universe can be defined as the least fixed point of the following operator

$$\Gamma(X) := \{x \mid \mathcal{C}(X, x)\} \cup \{\mathsf{i}\,(a, b) \mid a, b \in X\} \\ \cup \{\mathsf{sub}\,(a, f) \mid \mathsf{Indep}(a, f, \mathsf{pre}(a, f, X), X)\}$$

where Corollary 5.3 (adapted to the set theoretical setting) shows that Γ is monotone. The corresponding induction principle in set theory would be

$$\Gamma(A) \subseteq A \to \mathsf{M} \subseteq A$$

which means

$$\begin{aligned}&(\mathsf{U}(A)\ \wedge\ \mathsf{i}\in(A^2\to A)\\&\ \wedge\ (\forall a,f.\mathsf{Indep}(a,f,\mathsf{pre}\,(a,f,A),A)\to\mathsf{sub}\,(a,f)\in A))\\&\to\mathsf{M}\subseteq A\end{aligned}$$

It doesn't make sense to define $\mathsf{pre}\,(a, f, \varphi)$ for arbitrary formulas φ in Explicit Mathematics, and therefore we have to restrict the induction on M to "small sets", i.e., elements of $\Re$. We obtain the following

III. Induction for M

$$(\mathsf{EPM}.5)\qquad \begin{aligned}&\mathcal{U}(u)\wedge\mathsf{i}\in(\mathsf{u}^2\to\mathsf{u})\\&\wedge(\forall f.\forall a.\mathsf{Indep}(a,f,\mathsf{pre}\,(a,f,u),u)\to\mathsf{sub}\,(a,f)\,\dot{\in}\,u)\\&\to\mathsf{M}\,\dot{\subset}\,u\end{aligned}$$

Now, the theory EPM of *extended predicative Mahlo* can be defined as the extension of T_0 by the axioms $(\mathsf{EPM}.1)$ – $(\mathsf{EPM}.5)$.

Note that such an induction principles as $(\mathsf{EPM}.5)$ cannot be formulated in the axiomatic approach, as the quantifier in the "induction step" has to range over arbitrary functions, not only those which are total from names to names. For the approach to Mahlo in Martin-Löf type theory, which is also based on total functions, the addition of an induction principle leads to a contradiction (see [Pal98, Theorem 6.1]), and this is probably also the case for axiomatic Mahlo in Explicit Mathematics. As, so far, there is no account for partial functions in Martin-Löf type theory which allows to refer to the collection of all terms, there is yet no possibility to define an extended predicative version of Mahlo. We note however that we don't expect that the induction principles expressing minimality of M strengthen the theory. We expect the situation in this case to be similar to that in Martin-Löf type theory, where the second author has shown [Set97] that if one has a universe with certain closure conditions, one can define a set corresponding to the least universe having the same closure conditions—therefore having a least universe doesn't add any strength.

6 Remarks on the Analysis of EPM

A proof-theoretic analysis of EPM will be given by the authors elsewhere. As we formalize an internal Mahlo universe, the strength of EPM is slightly above the one of KPM. One needs one extra recursively inaccessible above KPM, i.e., a model of EPM has to be given in KPMI, $\mathsf{KP}\omega$ *plus the existence of one recursively Mahlo ordinal* M *plus* $\forall x\exists y.Ad(y)\wedge x\in y$. For the lower bound one can use an embedding of the theory $\mathsf{T}_0(\mathsf{M})^+$ and then follow arguments of Tupailo [Tup03] to get a realization of an appropriate extension of CZF into $\mathsf{T}_0(\mathsf{M})^+$. It seems to be feasible to get a lower bound by a well-ordering proof for that extension of CZF. The argument above would show as well that the theory $\mathsf{T}_0(\mathsf{M})^+$ has the same strength as EPM and KPMI.

However, there are still a couple of questions concerning modifications of the theory. For instance, in [JKS01], a concept of *name strictness* is introduced. It expresses that generators only generate names for appropriate arguments (e.g., $\Re(\mathsf{co}\, x) \to \Re(x)$).[5] In this context, also *name induction* is considered, which serves as an alternative to inductive generation or least universes to get a theory of the strength of T_0. The addition of name strictness and/or name induction may allow to simplify the definitions of relative f-pre-universe; however, there seems to be a subtle problem with formulating name strictness for generators of subuniverses of the Mahlo universe.

Also, one may investigate the potential of the induction axioms, for both the subuniverses and the Mahlo universe itself, in concrete applications. As noted above, it is the specific feature of the extended predicative approach that it allows to formulate such induction axioms.

Finally, the formulation of an extended predicative Mahlo universe in a metapredicative setting (both with an external and an internal Mahlo universe) is still lacking. It should result, in principle, from the omission of inductive generation (and therefore $(\mathsf{EPM}.3)$) and the induction axioms $(\mathsf{EPM}.2)$ and $(\mathsf{EPM}.5)$, and one probably needs to add $\Re_{\Re}(v) \to \mathsf{pre}\,(a, f, v) \mathrel{\dot{\subset}} v$, which is no longer provable without $(\mathsf{EPM}.2)$. These axioms allow an embedding of the metapredicative axiomatic external Mahlo universe (Theorem 5.5 holds with this modifications), which gives a lower bound for its proof theoretic strength. However one needs to carefully check whether any other adaptations of the axioms are needed, in order to avoid obtaining a theory which is stronger than the metapredicative axiomatic external Mahlo universe.

References

[Ara96a] Toshiyasu Arai. Proof theory of theories of ordinals I: Reflecting ordinals. Draft, 1996.

[Ara96b] Toshiyasu Arai. Systems of ordinal diagrams. Draft, 1996.

[Ara97a] Toshiyasu Arai. Proof theory of theories of ordinals II: Σ_1 stability. Draft, 1997.

[5] This concept is analogous to—and motivated by—the usual strictness for definedness or strictness for the predicate N for natural numbers in applicative theories, cf. [Kah00].

[Ara97b] Toshiyasu Arai. Proof theory of theories of ordinals III: Π_1 collection. Draft, 1997.

[Ara00a] Toshiyasu Arai. Ordinal diagrams for Π_3-reflection. *J. Symbolic Logic*, 65(3):1375 – 1394, 2000.

[Ara00b] Toshiyasu Arai. Ordinal diagrams for recursively Mahlo universes. *Arch. Math. Logic*, 39(5):353 – 391, 2000.

[Ara03] Toshiyasu Arai. Proof theory for theories of ordinals. I. Recursively Mahlo ordinals. *Ann. Pure Appl. Logic*, 122(1 – 3):1 – 85, 2003.

[Ara04] Toshiyasu Arai. Proof theory for theories of ordinals. II. Π_3-reflection. *Ann. Pure Appl. Logic*, 129(1 – 3):39 – 92, 2004.

[Bac50] Heinz Bachmann. Die Normalfunktionen und das Problem der ausgezeichneten Folgen von Ordnungszahlen. *Vierteljahresschrift der Naturforschenden Gesellschaft in Zürich*, XCV:5–37, 1950.

[Bar75] Jon Barwise. *Admissible Sets and Structures. An Approach to Definability Theory*. Ω-series. Springer, 1975.

[Bee85] Michael Beeson. *Foundations of Constructive Mathematics*. Ergebnisse der Mathematik und ihrer Grenzgebiete; 3.Folge, Band 6. Springer, 1985.

[BF97] C. Burali-Forti. Una questione sui numeri transfiniti. *Rendiconti del Circolo Matematico di Palermo (1884 - 1940)*, 11(1):154 – 164, December 1897. Translation in [Hei67], pp. 104 – 112.

[BFPS81] Wilfried Buchholz, Solomon Feferman, Wolfram Pohlers, and Wilfried Sieg. *Iterated Inductive Definitions and Subsystems of Analysis: Recent Proof-Theoretical Studies*, volume 897 of *Lecture Notes in Mathematics*. Springer, 1981.

[BS83] Wilfried Buchholz and Kurt Schütte. Ein Ordinazahlbezeichnungssystem für die beweistheoretische Abgrenzung der Π^1_2-Separation und Bar-Induktion. *Sitzungsberichte der Bayerischen Akademie der Wissenschaften, Math.-Nat. Klasse*, pages 99 – 132, 1983.

[BS88] Wilfried Buchholz and Kurt Schütte. *Proof Theory of Impredicative Subsystems of Analysis*. Bibliopolis, Naples, 1988.

[Buc92] Wilfried Buchholz. A simplified version of local predicativity. In P. Aczel, H. Simmons, and S. S. Wainer, editors, *Proof Theory. A selection of papers from the Leeds Proof Theory Programme 1990*, pages 115 – 147, Cambridge, 1992. Cambridge.

[DS03] Peter Dybjer and Anton Setzer. Induction-recursion and initial algebras. *Annals of Pure and Applied Logic*, 124:1 – 47, 2003.

[Fef64] Solomon Feferman. Systems of predicative analysis. *Journal of Symbolic Logic*, 29:1 – 30, 1964.

[Fef75] Solomon Feferman. A language and axioms for explicit mathematics. In J. Crossley, editor, *Algebra and Logic*, volume 450 of *Lecture Notes in Mathematics*, pages 87–139. Springer, 1975.

[Fef79] Solomon Feferman. Constructive theories of functions and classes. In M. Boffa, D. van Dalen, and K. McAloon, editors, *Logic Colloquium 78*, pages 159–224. North–Holland, Amsterdam, 1979.

[Fef05] Solomon Feferman. Predicativity. In S. Shapiro, editor, *The Oxford Handbook of Philosophy of Mathematics and Logic*, pages 590 – 624, Oxford, 2005. Oxford University Press.

[FJ96] Solomon Feferman and Gerhard Jäger. Systems of explicit mathematics with non-constructive μ-operator. Part II. *Annals of Pure and Applied Logic*, 79(1):37–52, 1996.

[Gen36] Gerhard Gentzen. Die Widerspruchsfreiheit der reinen Zahlentheorie. *Mathematische Annalen*, 112:493 – 565, 1936.

[GR94] Edward Griffor and Michael Rathjen. The strength of some Martin-Löf type theories. *Archive for Mathematical Logic*, 33(5):347, 1994.

[Hau08] Feliz Hausdorff. Grundzüge einer Theorie der geordneten Mengen. *Mathematische Annalen*, 65(4):435 – 505, 1908.

[Hei67] Jean van Heijenoort. *From Frege to Gödel*. Harward University Press, 1967.

[Hei86] G. Heinzmann. *Poincaré, Russell, Zermelo et Peano. Textes de la discussion (1906 – 1912) sur les fondements des mathématiques: des antinomies á la preédicativité*. Albert Blanchard, Paris, 1986.

[Hin78] P.G. Hinman. *Recursion-Theoretic Hierarchies*. Springer, 1978.

[How72] W. A. Howard. A system of abstract constructive ordinals. *J. Symbolic Logic*, 37:355 – 374, 1972.

[Jäg79] Gerhard Jäger. *Die konstruktible Hierarchie als Hilfsmittel zur beweistheoretischen Untersuchung von Teilsystemen der Analysis*. PhD thesis, Universität München, 1979. Dissertation.

[Jäg80] Gerhard Jäger. Theories for iterated jumps. Handwritten Notes, 1980.

[Jäg83] Gerhard Jäger. A well-ordering proof for Feferman's theory T_0. *Archiv f. Math. Logik und Grundlagenforschung*, 23:65 – 77, 1983.

[Jäg01] Gerhard Jäger. First order theories for nonmonotonic inductive definitions: Recursively inaccessible and Mahlo. *Journal of Symbolic Logic*, 66(3):1073–1089, September 2001.

[Jäg05] Gerhard Jäger. Metapredicative and explicit Mahlo: a proof-theoretic perspective. In René Cori, Alexander Razborov, Stevo Todorčević, and Carol Wood, editors, *Logic Colloquium 2000*, Lecture Notes in Logic, pages 272 – 293. A K Peters. Association for Symbolic Logic, 2005.

[JKS99] Gerhard Jäger, Reinhard Kahle, and Thomas Strahm. On applicative theories. In A. Cantini, E. Casari, and P. Minari, editors, *Logic and Foundation of Mathematics*, pages 83–92. Kluwer, 1999.

[JKS01] Gerhard Jäger, Reinhard Kahle, and Thomas Studer. Universes in explicit mathematics. *Annals of Pure and Applied Logic*, 109(3):141–162, 2001.

[JKSS99] Gerhard Jäger, Reinhard Kahle, Anton Setzer, and Thomas Strahm. The proof-theoretic analysis of transfinitely iterated fixed point theories. *Journal of Symbolic Logic*, 64(1):53 – 67, 1999.

[JP82] Gerhard Jäger and Wolfram Pohlers. Eine beweistheoretische Untersuchung von $(\Delta^1_2 - CA) + BI$ und verwandter Systeme. *Sitzungsberichte der Bayer. Akad. d. Wiss., Math.-Nat. Kl.*, pages 1 – 28, 1982.

[JS01] Gerhard Jäger and Thomas Strahm. Upper bounds for metapredicative Mahlo in explicit mathematics and admissible set theory. *Journal of Symbolic Logic*, 66(2):935–958, 2001.

[JS02] Gerhard Jäger and Thomas Studer. Extending the system T_0 of explicit mathematics: the limit and Mahlo axioms. *Annals of Pure and Applied Logic*, 114(1–3):79–101, 2002.

[JS05] Gerhard Jäger and Thomas Strahm. Reflections on reflections in explicit mathematics. *Annals of Pure and Applied Logic*, 136(1–2):116–133, 2005.

[Kah97] Reinhard Kahle. Uniform limit in explicit mathematics with universes. Technical Report IAM-97-002, IAM, Universität Bern, 1997.

[Kah00] Reinhard Kahle. N-strictness in applicative theories. *Archive for Mathematical Logic*, 39(2):125 – 144, February 2000.

[Kah07] Reinhard Kahle. *The applicative realm*, volume 40 of *Textos de Matemática*. Departamento de Matemática, Universidade de Coimbra, 2007. Habilitationsschrift, Fakultät für Informations- und Kommunikationswissenschaften, Universität Tübingen.

[Kre60] Georg Kreisel. Ordinal logics and the characterization of informal concepts of proof. In John A. Todd, editor, *Proceedings of the International Congress of Mathematicians, 14 – 21 Aug 1958*, pages 289 – 299, Cambridge, 1960. Cambridge University Press.

[Kri64] S. Kripke. Transfinite recursion on admissible ordinals, I, II (abstracts). *J. Symbolic Logic*, 29:161 – 162, 1964.

[Lei94] Daniel Leivant. Ramified recurrence and computational complexity i: Word recurrence and poly-time. In Peter Clote and Jeffrey Remmel, editors, *Feasible Mathematics II*, pages 320 – 343, New York, 1994. Birkhäuser-Boston.

[Lei95] Daniel Leivant. Intrinsic theories and computational complexity. In Daniel Leivant, editor, *Logic and Computational Complexity*, pages 177 – 194. Springer Lecture Notes in Computer Science 960, 1995.

[Mah11] Paul Mahlo. Über linear transfinite Mengen. *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig. Mathematisch-Physische Klasse*, 63:187 – 225, 1911.

[Mah12] Paul Mahlo. Zur Theorie und Anwendung der ϱ_0-Zahlen. *Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig. Mathematisch-Physische Klasse*, 64:108 – 112, 1912.

[Nel86] E. Nelson. *Predicative Arithmetic*. Princeton University Press, Princeton, 1986.

[Nor99] Dag Normann. A Mahlo-universe of effective domains with totality. In S. Barry Cooper and J. Truss, editors, *Models and Computability,* volume 259 of *London Mathematical Society Lecture Note Series,* pages 293 – 312. Cambridge University Press, 1999

[Pal98] E. Palmgren. On universes in type theory. In G. Sambin and J. Smith, editors, *Twenty five years of constructive type theory*, pages 191 – 204, Oxford, 1998. Oxford University Press.

[Par92] Charles Parsons. Impredicativity of induction. In Michael Detlefsen, editor, *Proof, Logic and formalization*, pages 139 – 161. Routledge, 1992.

[Pla66] R. Platek. *Foundations of recursion theory.* PhD thesis, Stanford University, Standford, California, 1966. Contains Supplement.

[Poh96] Wolfram Pohlers. Pure proof theory. *Bulletin of Symbolic Logic*, 2(2):159–188, 1996.

[Poh98] Wolfram Pohlers. Subsystems of set theory and second order number theory. In S. Buss, editor, *Handbook of Proof Theory*, chapter IV, pages 209–335. North-Holland, 1998.

[Poi06] H. Poincaré. Les mathématiques et la logique. *Revue de métaphysique et de morale*, 14:294 – 317, 1906. Reprinted in [Hei86], 11 – 53.

[Rat90] Michael Rathjen. Ordinal notations based on a weakly Mahlo cardinal. *Arch. Math. Log.*, 29:249 – 263, 1990.

[Rat91] Michael Rathjen. Proof-theoretical analysis of KPM. *Arch. Math. Log.*, 30:377 – 403, 1991.

[Rat92] Michael Rathjen. Eine Ordinalzahlanalyse der Π_3–Reflexion. Habilitationsschrift, Westfälische Wilhelms–Universität Münster, 1992.

[Rat94a] Michael Rathjen. Collapsing functions based on recursively large cardinals: A well-ordering proof for KPM. *Archive for Mathematical Logic*, 33:35–55, 1994.

[Rat94b] Michael Rathjen. Proof theory of reflection. *Ann. Pure Appl. Logic*, 68:181 – 224, 1994.

[Rat95] Michael Rathjen. Recent advances in ordinal analysis: Π^1_2-CA and related systems. *Bulletin of Symbolic Logic*, 1:468 – 485, 1995.

[Rat96] Michael Rathjen. The recursively Mahlo property in second order arithmetic. *Mathematical Logic Quarterly*, 42:59–66, 1996.

[Rat05a] Michael Rathjen. An ordinal analysis of parameter free Π^1_2-comprehension. *Arch. Math. Log.*, 44(3):263 – 362, April 2005.

[Rat05b] Michael Rathjen. An ordinal analysis of stability. *Arch. Math. Logic*, 44(1):1 – 62, January 2005.

[Ric71] Wayne Richter. Recursively Mahlo ordinals and inductive definitions. In R. O. Gandy and C. E. M. Yates, editors, *Logic Colloquium '69*, pages 273–288. North-Holland, 1971.

[Rus02] Bertrand Russell. Letter to Frege. Published in [Hei67], pp. 124 – 125, 1902.

[Rus06] Bertrand Russell. On some difficulties in the theory of transfinite numbers and order types. *Proc. London Mathematical Society*, 4:29 –53, 1906. Reprinted in [Rus73], pp. 135 – 164.

[Rus73] Bertrand Russell. *Essays in Analysis*. George Braziller, New York, 1973. Edited by D. Lackey.

[Sch65a] Kurt Schütte. Einge Grenze für die Beweisbarkeit der Transfiniten Induktion in der verzweigten Typenlogik. *Archiv für Mathematische Logik und Grundlagenforschung*, 7:45 – 60, 1965.

[Sch65b] Kurt Schütte. Predicative well-orderings. In J. Crossley and M. Dummett, editors, *Formal Systems and recursive functions. Proc. of the 8th Logic Colloquium Oxford 1963*, pages 280 – 303, Amsterdam, 1965. North Holland.

[Sch77] Kurt Schütte. *Proof Theory*. Springer, 1977.

[Set97] Anton Setzer. Defining the least universe in Martin-Löf's type theory. *Bulletin of Symbolic Logic*, 3:276 – 277, 1997.

[Set98] Anton Setzer. Well-ordering proofs for Martin-Löf type theory. *Annals of Pure and Applied Logic*, 92:113 – 159, 1998.

[Set00] Anton Setzer. Extending Martin-Löf type theory by one Mahlo-universe. *Archive for Mathematical Logic*, 39:155–181, 2000.

[Sim99] Stephen G. Simpson. *Subsystems of second-order arithmetic*. Springer, 1999.

[Str99] Thomas Strahm. First steps into metapredicativity in explicit mathematics. In S. B. Cooper and J. Truss, editors, *Sets and Proofs*, pages 383 – 402, Cambridge, 1999. Cambridge University Press.

[Str02] Thomas Strahm. Wellordering proofs for metapredicative Mahlo. *Journal of Symbolic Logic*, 67(1):260–278, 2002.

[Tak87] G. Takeuti. *Proof Theory*. North–Holland Publishing Company, Amsterdam, second edition, 1987.

[Tup03] Sergei Tupailo. Realization of constructive set theory into explicit mathematics: a lower bound for impredicative Mahlo universe. *Annals of Pure and Applied Logic*, 120(1-3):165 – 196, 2003.

[TvD88] Anne Troelstra and Dirk van Dalen. *Constructivism in Mathematics, vol II*. North Holland, 1988.

[Wey18] Hermann Weyl. *Das Kontinuum. Kritische Untersuchungen über die Grundlagen der Analysis*. Veit, Leipzig, 1918.

ITTMs with Feedback

Robert S. Lubarsky

Florida Atlantic University
Dept. of Mathematical Sciences
777 Glades Rd., Boca Raton, FL 33431, USA
Robert.Lubarsky@alum.mit.edu

Abstract Infinite time Turing machines are extended in several ways to allow for iterated oracle calls. The expressive power of these machines is discussed and in some cases determined.

1 Introduction

Infinite time Turing machines, or ITTMs, introduced in [2], are regular Turing machines that are allowed to run for transfinitely many steps. The only changes to the standard definition of a Turing machine that need making are what to do at limit stages: the head goes to the front of the tape(s), the state entered is a dedicated state for limits, and the value of each cell is the lim sup of the previous values.

That introductory paper also discussed various kinds of oracles computations and corresponding jump operators. One such jump operator encodes the information "does the ITTM with index e on input r converge?" If e is allowed to call an oracle A, this is called the **strong jump** $A^{\blacktriangledown}$ of A: $\{(e, x) \mid \{e\}^A(x)\downarrow\}$. The jump can of course be used as an oracle itself, and the process iterated: you can, for instance, ask whether $\{e\}(r)$ converges, where $\{e\}$ can itself ask oracle questions of simple (non-oracle) ITTMs.

We would like to investigate ultimate iterations of this jump, for several reasons. Iterations of a procedure can lead to new phenomena. A well-known example of that in a context similar to the current one is transfinite iterations of the regular Turing jump. If you iterate the Turing jump along any well-order that appears along the way, you get the least admissible set containing your starting set, admissible computability theory being a quantum leap beyond ordinary computability theory [1]. Arguably the next example right after this one would be iterations of inductive definitions. Admissible set theory is exactly what you need to develop a theory of positive inductive definitions, the least fixed point of such being Σ_1 definable over any admissible set containing the definition in question (e.g. its parameters) [1]. If the language of least fixed points of positive inductive definitions is closed in a straightforward manner, you end up with the μ-calculus. Determining the sets definable in the μ-calculus is however anything but a straightforward extension of admissibility, needing a generalization of the notion of reflection, gap reflection [3–5]. Something similar happens with ITTMs, as some of the extensions are quite different from the base case, as we will see.

A potential application of this work is in proof theory. The strongest fragment of second-order arithmetic for which an ordinal analysis has been done to date is Π^1_2 Comprehension [6]. Regular (i.e. non-iterated) ITTMs are already more powerful than that. Perhaps having descriptions of stronger subsystems of analysis other than the straightforward hierarchy of Π^1_n Comprehension principles will help the proof theorists make progress.

The goal of this line of inquiry is to examine what kind of iterations of ITTMs make sense, and to quantify how powerful those iterations are by characterizing the reals, or what amounts to the same thing ordinals or sets, that can be so written. This situation is different from that for regular Turing machines, because an ITTM computation can halt after infinitely many steps, and so ITTMs have the power to write reals. Hamkins and Lewis insightfully classified the reals that come up in this context as **writable** if they appear as the output of a halting computation, **eventually writable** if they are eventually the unchanging content of the output tape for a divergent computation, and **accidentally writable** if they appear anywhere on any tape during any ITTM computation, even if they are overwritten later. The same concepts apply to ordinals, where an ordinal is writable (resp. eventually, accidentally) if some real coding that order-type is writable (resp. eventually, accidentally). This distinction among these kinds of reals and ordinals turned out to be crucial to their characterization, as announced in [7] and detailed in [8], with improved proofs and other results in [9]. Let λ, ζ, and Σ respectively be the suprema of the writable, eventually writable, and accidentally writable ordinals.

Theorem 1.1. *(Welch) ζ is the least ordinal α such that L_α has a Σ_2 elementary extension, L_λ is the smallest Σ_1 substructure of L_ζ, and L_Σ is the unique Σ_2 extension of L_ζ.*

The relativization of this theorem to a real parameter holds straightforwardly.

In the next section, we give some notions of the syntax and semantics of these iterations fundamental to what follows. The three after that each gives a different kind of extension of ITTMs, and about as much as is currently know about them. Some are characterized pretty fully, others only to the point where it's clear that there's something very different going on. The final section offers a generalization of the semantics.

2 Feedback ITTMs and the Tree of Subcomputations

A **feedback ITTM** (**FITTM**) is an ITTM with two additional tapes, and an additional state, which is the oracle query "does the feedback ITTM with program the content of the first additional tape on input the content of the second converge?" Clearly, the additional tapes are merely an expository convenience, as they could be coded as dedicated parts of the original tape.

The semantics of feedback ITTMs is defined via the tree of subcomputations. The idea is that the tree keeps track of oracle calls by having each one be a child of the calling computation. This tree is in general not itself ITTM computable. Rather, it is defined within ZF, even if a fragment of ZF would suffice, inductively on the ordinals. At every ordinal stage, each extant node is labeled with some computation, and control is with one node.

At stage 0, control is with the root, which we think of as at the top of the downward growing subcomputation tree. The root is labeled with the index (and input, if any) of the main computation.

At a successor stage, if the node currently in control is in any state other than the oracle call, action is as with a regular Turing machine. If taking that action places that machine in a halting state, then, if there is a parent, the parent gets the answer "convergent" to its oracle call, and control passes to the parent. If there is no parent, then the current node is the root, and the computation halts. If the additional step does not place the machine in a halting state, then control stays with the current node. If the current node makes an oracle call, a new child is formed, after (to the right of) all of its siblings, labeled with the calling index and parameter; a new machine is established at that node, with program the given index and with the parameter written on the input tape; and control passes to that node.

At a limit stage, there are three possibilities. One is that on some final segment of the stages there were no oracle calls, and so control was always at one node. Then the rules for limit stages of regular ITTMs apply, and the snapshot of the computation at the node in question is determined (where the snapshot includes all of the current information about the computation – the state, the tape contents, and so on). If that snapshot repeats an earlier one, then that computation is divergent. (Here we are using the standard convention, first articulated in [2], that a snapshot qualifies as repeating only if it guarantees an infinite loop. In point of fact, a snapshot might be identical to an earlier one, which guarantees that it will recur ω-many times, but it is possible that at the limit of those snapshots, we escape the loop. So by convention, a repeating snapshot is taken to be one that guarantees that you're in a loop.) At that point, if there is a parent, then the parent gets the answer "divergent" to its oracle call, and control is passed to the parent. If there is no parent, then the node in question is the root, and the entire computation is divergent.

A second possibility is that cofinally many oracle calls were made, and there is a node ρ such that cofinally many of those calls were ρ's children. Note that such a node must be unique. Then ρ was active cofinally often, and again the rules for regular ITTMs at limit stages apply. If ρ is seen at that stage to be repeating, then control passes to ρ's parent, if any, which also gets the answer that ρ is divergent; if ρ is the root, then the main computation is divergent. If ρ is not seen to be repeating at this stage, then ρ retains control and the computation continues.

The final possibility is that, among the cofinally many oracle calls made, there is an infinite descending sequence, which is the right-most branch of the tree. This is bad. It is troublesome, at best, to define what to do at the next step. Various ways to avoid this last situation are the subject of the next sections.

3 Pre-Qualified Iterations

The problem cited above is that the subcomputation tree has an infinite descending sequence. The most obvious way around that is to ensure that that does not happen, that the tree is well-founded. That can be enforced by attaching an ordinal to each node of the tree and requiring that children of a node have smaller ordinals.

That is in essence what is done with the strong jump $\varnothing^{\blacktriangledown}$ of [2]. $\varnothing^{\blacktriangledown}$ is $\{(e, x) \mid e(x) \downarrow\}$, which is the same thing as labeling the root of the subcomputation tree with 1, so none of its children, the oracle calls, can themselves make oracle calls. In unpublished work, Phil Welch has show that $\zeta^{\varnothing^{\blacktriangledown}}$ is the smallest Σ_2-extendible limit of Σ_2-extendibles, and that $\lambda^{\varnothing^{\blacktriangledown}}$ and $\Sigma^{\varnothing^{\blacktriangledown}}$ are such that $L_{\lambda^{\varnothing^{\blacktriangledown}}}$ is the least Σ_1 substructure of $L_{\zeta^{\varnothing^{\blacktriangledown}}}$, which is itself a Σ_2 substructure of $L_{\Sigma^{\varnothing^{\blacktriangledown}}}$.

We would like to generalize this to ordinals as large as possible, certainly to ordinals greater than 1. An **ordinal oracle ITTM** is an FITTM with not two but three additional tapes. On the third tape is written a real coding an ordinal α. The oracle calls allowed are about other ordinal oracle ITTMs, and on the third tape must be written some ordinal $\beta < \alpha$. Since one of the other tapes is for parameter passing, it is unimportant just how the ordinals are written on the latest tape. With this restriction, the third outcome above can never happen, and all computations are well-defined (as either convergent or divergent).

An **iterated ITTM**, or **IITTM**, is an FITTM that may make an oracle call about any ordinal oracle ITTM writing on the third tape any ordinal at all. So an IITTM is like an ordinal oracle ITTM only the length of the ordinal iteration is not fixed in advance. Rather, it is limited only by what the machine figures out to write down.

Definition 3.1. λ^{it}, ζ^{it}, and Σ^{it} are the respective suprema of the ordinals writable, eventually writable, and accidentally writable by IITTMs.

Definition 3.2. An ordinal α is

- 0-extendible if it is Σ_2 extendible,
- $\beta + 1$-extendible if it is a Σ_2 extendible limit of β-extendibles, and
- γ-extendible (γ a limit) if it is Σ_2 extendible and a limit of β-extendibles for each $\beta < \gamma$.

As pointed out by the referee, the limit clause actually works perfectly well for all three clauses.

The definition above relativizes to any parameter x. The corresponding notation is for α to be $\beta[x]$-extendible. Notice that, in the limit case, when $\gamma < \alpha$, α is also the limit of ordinals which are themselves limits of β-extendibles for each $\beta < \gamma$.

Theorem 3.3. *For ordinal oracle ITTMs with ordinal α coded by the input real x_α and parameter y, the supremum ζ of the eventually writable ordinals is the least $\alpha[x_\alpha, y]$-extendible. Moreover, the supremum Σ of the accidentally writable ordinals is such that $L_\Sigma[x_\alpha, y]$ is the (unique) Σ_2 extension of $L_\zeta[x_\alpha, y]$, and the supremum λ of the writable ordinals is such that $L_\lambda[x_\alpha, y]$ is the smallest Σ_1 substructure of $L_\zeta[x_\alpha, y]$. Finally, the writable (resp. eventually, accidentally) reals are those in the corresponding segment of $L[x_\alpha, y]$.*

Proof. By induction on α.

$\alpha = 0$: This is the relativized version of Welch's theorem cited above.

$\alpha = \beta+1$: Let γ be any ordinal less than ζ. Run some machine which eventually writes γ. Dovetail that computation with the following. Simulate running all ordinal oracle ITTMs with input β and as parameters the output of the first machine, which is eventually γ, and y. This is essentially running a universal machine: clear infinitely many cells on the scratch tape, split them up into countably many infinite sequences, and on the i^{th} sequence run a copy of the i^{th} machine. For each of those simulations, keep asking whether the current output will ever change. (That is, ask whether the computation that continues that simulation until the output tape changes, at which point it halts, is convergent.) This is a legitimate question for the oracle, as $\beta < \alpha$. Whenever you get the answer "no," indicate as much on a dedicated part of the output tape. Eventually you will get all and only the indices of the eventually stable computations. So the least $\beta[x_\alpha, y]$-extendible ordinal is less than ζ, and so ζ is the limit of such.

Because of this closure under β-extendibility, $L_\zeta[x_\alpha, y]$ can run correctly the computation of the ordinal oracle ITTMs with input β. So the rest of the proof – that the computations of eventually writable reals stabilize by ζ, and that the eventually writable reals form a Σ_2 substructure of the accidentally writables and a Σ_1 extension of the writables – follows by the same arguments pioneered in [8] and improved upon in [9]. In order to keep this paper self-contained, and to verify that the new context here really makes no difference, we present these arguments here.

Suppose, toward a contradiction, that $L_\zeta[x_\alpha, y]$ satisfies some Π_2 sentence φ, but $L_\Sigma[x_\alpha, y]$ does not. By the nature of Π_2 sentences, the set of ordinals $\xi \leq \Sigma$ such that $L_\xi[x_\alpha, y] \models \varphi$ is closed, and so contains its maximum. By hypothesis, that maximum is strictly less than Σ. Take some machine that accidentally writes each of the ordinals less than Σ. A universal machine will do, for instance, so we will call this machine u. We also need a machine, say p, which eventually writes the φ's parameter. It is safe to assume that there is only one parameter, as finitely many can be combined into one set by pairing. If no parameter is necessary, then $\varnothing$ as a dummy parameter can be used. Our final machine, call it e, runs p and u simultaneously. It takes the output of u and uses it to generate the various $L_\xi[x_\alpha, y]$s. When it finds such a set modeling φ, with parameter the current output of p, it compares ξ to the current content of the output tape. If the current content is an ordinal greater than or equal to ξ, nothing is written and the computation continues. Else ξ is written on the output. Eventually the output of p settles down. Once that happens, when the largest such ξ ever appears, it will be so written, after which point it will never be overwritten, making ξ eventually writable. This is a contradiction.

Regarding λ, suppose $L_\zeta[x_\alpha, y]$ satisfies some Σ_1 formula ψ with parameters from $L_\lambda[x_\alpha, y]$. Consider the computation which first computes the parameters using a halting computation, then runs a machine which eventually writes a witness to ψ and halts when it finds one. This is a halting computation for such a witness.

By the foregoing, ζ is $\alpha[x_\alpha, y]$-extendible. That it is the least such is ultimately because the assertion that any particular cell in a computation stabilizes is Σ_2. In detail, let ζ_α be the least $\alpha[x_\alpha, y]$-extendible ordinal and Σ_α its Σ_2 extension. Since stabilization is a Σ_2 assertion, any computation has the same eventually stable cells at ζ_α as at Σ_α. Moreover, if δ is a stage beyond which a certain cell is stable in ζ_α, the assertion that that cell beyond δ is stable is Π_1, so that same δ is also a stabilization point in Σ_α. So the snapshot of a computation at ζ_α is that same at Σ_α, and all looping has occurred by then.

α a limit: Since ordinal oracle ITTMs with input α subsume those with input $\beta < \alpha$, ζ is $\beta[x_\alpha, y]$-extendible for each $\beta < \alpha$, and hence, considering successor βs, a limit of $\beta[x_\alpha, y]$-extendibles. The rest follows as above. $\square$

Theorem 3.4. *ζ^{it} is the least κ which is κ-extendible, λ^{it} its smallest Σ_1 substructure, and Σ^{it} its (unique) Σ_2 extension.*

Proof. For every $\alpha < \zeta^{it}$, the ordinal oracle ITTMs with input α are also IITTMs. Hence the least α-extendible is $\leq \zeta^{it}$, and ζ^{it} is a limit of α-extendibles. The rest, again, follows as above. □

4 Freezing Computations

Another way to deal with the possible ill-foundedness of the subcomputation tree is not to worry about it. That is, while no steps are taken to rule out such computations, there will be some with perfectly well-founded subcomputation trees, even if only by accident. We remain positive, and focus our attention on those, where we have a well-defined semantics, including whether a computation converges or diverges. So we can define the reals writable, eventually writable, and accidentally writable by FITTMs.

Proposition 4.1. *Every feedback eventually writable real is feedback writable.*

Proof. Let e be a computation which writes a feedback eventually writable real. Consider an alternative computation which runs e on a dedicated part of the tapes. Every time e's output tape changes, the main computation asks the oracle: "Consider the computation which begins at the current snapshot of e, and continues e's computation until the output tape changes once more, and then halts. Does that converge or diverge?" Since e's tree of subcomputations is well-founded, so is that of the oracle call, and the oracle call will return a definite answer. If that answer is "converge," then the construction continues; if "diverge", then the construction halts. By hypothesis, this computation eventually halts, at which point e's output is written on the output tape. □

Even worse:

Proposition 4.2. *Every feedback accidentally writable real is feedback writable.*

Proof. Suppose e is a divergent computation. As in [2], e then has to loop, and does so already at some countable stage. The sledgehammer way to see that is that there are only set-many possible snapshots, so if a computation never halts then it has to repeat itself. As to why that would happen at some countable stage, that follows from Levy absoluteness. More concretely, the argument in [2] for regular ITTMs applies unchanged in the current setting. There are only countably many cells. So only countably many stop changing beneath $\aleph_1$. Moreover, there is some countable bound α by which those have all stopped changing. List the remaining cells in an ω-sequence $c_0, c_1, ...$ Let α_0 be the least stage beyond α at which c_0 changes. Inductively, let α_n be the least stage beyond α_{n-1} by which all of $c_0, c_1, ..., c_n$

have changed since stage α_{n-1}. The configuration at stage $\alpha_\omega = \lim_n \alpha_n$ repeats unboundedly beneath $\aleph_1$, and so is a looping stage.

Let α be such that e has already started to loop by α many steps. Suppose we could write (a real coding) α via a halting computation. Then any real written at any time during e's computation would be writable, via the program "write α, then compute e for the number of steps given by the integer n in the coding of α, then output whatever's on e's tapes then" (with the desired choice of n, of course). So it suffices to write the looping time of a computation.

First we determine the first looping snapshot of the machine. At every stage of the computation in a simulation of e, the oracle is asked: "Consider the computation that begins with the current snapshot of e, saves it on a dedicated part of the tape, and continues with a simulation of e on a different part of the tape, halting whenever the original snapshot is reached again; does this computation halt?" If the answer is "no," the simulation continues. Eventually the answer will be "yes." That is the first looping snapshot. (Actually, as pointed out in [2], that's not quite right. A snapshot can repeat itself, which would then force it to repeat ω-many times, but the limit could be unequal to that repeating snapshot, and so this loop could be escaped. The constructions here could be modified easily enough to avoid this problem.)

The next thing to do would be to write the ordinal number of steps it took to get to that looping snapshot, and the ordinal number of steps it would take to make one loop, and then to add them. Since those ordinals are constructed the same way, we will describe only how to do the second.

During the construction, we will assign integers to ordinals in such a way that the $<$-relation will be immediate. The construction will take ω-many stages, during each of which we will use up countably (or finitely) many integers, so beforehand assign to each $n \in \omega$ countably many integers disjointly to be available at stage n. Furthermore, each integer has its own infinite part of the tape for its scratchwork.

Let C_i ($i \in \omega$) be the (simulated) i^{th} cell of the tape on which we're running (the simulation of) e. We will need to know which cells change value cofinally in the stage of interest (the return of the looping stage) and which don't. So simulate the run of e from the looping stage until its reappearance. Every time C_i changes value, toggle the i^{th} cell on another dedicated tape from 0 to 1 to 0. At the end of the computation, the i^{th} cell on the dedicated tape will be 0 iff C_i changed value boundedly often; so it will be 1 iff C_i changed value cofinally often.

Stage 0 starts in the looping snapshot, and is itself split into ω-many steps. Those steps interleave consideration of the cells that changed boundedly often and those that change cofinally. At step $2i$ continue the computation until the i^{th} cell with bounded change stops changing. That can be determined by asking the oracle whether the cell in question changes before the looping snapshot reappears. While this is not a converges-or-diverges question on the face of it, since the computation converges in any case (either when the cell changes or when the looping snapshot is reached, whichever happens first), one of those outcomes can be changed to a trivial loop, so that the question is a standard oracle call. If the answer is "yes," then continue the computation until the answer becomes "no," which is guaranteed to happen. At that point, use an available integer to mark that ordinal stage, which integer is then larger in the ordinal ordering than all other integers used so far. Also write the current snapshot in that integer's scratchwork part of the tape. Then proceed to the next step, $2i + 1$.

At step $2i + 1$, we will consider not just the i^{th} cofinally changing cell, but also the j^{th} such for all $j \leq i$, for purposes of dovetailing. Sequentially for each j from 0 to i, go to the next stage at which the j^{th} cofinally changing cell changes value again. After doing so for i, use an available integer to mark that ordinal stage, which integer is then larger in the ordinal ordering than all other integers used so far. Also write the current snapshot in that integer's scratchwork part of the tape. Then proceed to step $2(i + 1)$.

Because stage 0 consists of ω-many steps, each of which picks out only one integer in an increasing sequence, it picks out a strictly increasing ω-sequence of ordinals. The limit of that ordinal sequence is the ordinal in the computation at which its looping snapshot reappears. That's because by then we're beyond the ordinal at which any cell with boundedly many changes will change again, thanks to the even steps, and those cells with cofinal changing change cofinally in that ordinal, thanks to the dovetailing in the odd steps.

To summarize, we have produced an ω-sequence cofinal in the ordinal at which the looping snapshot reappears. Inductively, suppose at stage $i > 0$ we have an integer assignment, with $<$, to a subset of e's ordinal stage, as well as a picture of the snapshot of the computation each at such stage of the computation. Then for each integer which is a successor in this partial assignment, replicate the construction above with the starting snapshot being the snapshot of e at the predecessor and the ending snapshot being the snapshot of e at the integer under consideration. By the well-foundedness of the ordinals, this process ends after ω-many stages.

□

It is easy to see that the feedback writable reals are those contained in the initial segment of L given by the feedback writable ordinals, which are also the FITTM clockable ordinals. We call the set of these ordinals Λ.

This result removes the basis of the analysis used in weaker forms of ITTM computation. It comes about because the divergence of a computation in this paradigm can be determined convergently by a computation of the same type. Why doesn't this run afoul of some kind of diagonalization result? The answer is that there's no universal machine! That is, the computations and oracle calls used in the proofs above were sometimes convergent and sometimes divergent, but conveniently they were in any case all well-defined: the tree of subcomputations was well-founded. If it is not, we have no semantical notion of how the computation should continue or what the outcome should be. This notion is captured in the following.

Definition 4.3. A computation is **freezing** if its tree of subcomputations is ill-founded.

Proposition 4.4. *There is no FITTM computation which decides on an input e whether the e^{th} FITTM is freezing.*

Proof. If there were, you could diagonalize against the non-freezing computations, for a contradiction. □

We expect that as with most models of computation, the key to understanding what's computable will be an analysis of the uncomputable. While the freezing computations do not have an output or even a divergent computation, they are perfectly well-defined up until the point when an oracle call is made about a freezing subcomputation. For that matter, on the tree of subcomputations, that freezing subcomputation generates a good tree underneath it, until it calls its own freezing subcomputation. More generally, even for a freezing computation, its subcomputation tree, albeit ill-founded, is well-defined. Hence the following definition makes sense.

Definition 4.5. A real is **freezingly writable** if it appears anywhere on a tape during a freezing computation or any of its subcomputations.

We expect that the role that the eventually and accidentally writable reals played in the understanding of the writable reals for basic ITTMs will be played here by the freezingly writable reals. In any case, it should be of interest to understand better the freezing computations. Centrally, what does the subcomputation tree of a freezing computation look like? Since the computation cannot continue once an infinite path through the tree develops, that infinite path is unique, and is the right-most path. So each of the ω-many levels on the tree has width some successor ordinal. For each freezing computation e, let λ_n^e be the width of level n of e's subcomputation tree. For a fixed e, there are three possibilities for the λ_n^es:

a) λ_n^e is bounded beneath Λ.

b) λ_n^e is cofinal in Λ.

c) Some λ_n^e is greater than Λ.

Option a) is simply unavoidable: it is a simple task to write a machine which immediately makes an oracle call about itself, producing a subcomputation tree of order-type ω^* (ω backwards).

Options b) and c), as it turns out, are incompatible with each other. To see this, first note that if c) holds for some computation, then n can be chosen to be 1 (level 0 consisting of the root alone). After all, if this is not the case for some given e, let e_1 be some computation that halts at a stage larger than $\max_{m<n} \lambda^e_m$. Use e_1 to write e_1's run-time (using methods like those in the main proposition above). Use that ordinal to run e substituting for the oracle calls an explicit computation until the right-most node on level $n-1$ (of e's original subcomputation tree) becomes active. That is the node which has more than Λ-many children, and which is now the root node of the tree of this modified computation.

Now assume we have indices e_b and e_c of types b and c respectively (and $\lambda_1^{e_c} > \Lambda$). Simulate e_c. Whenever an oracle call is made, write the new length of the top level in the subcomputation tree (using techniques as above). Use that ordinal to simulate the computation of e_b substituting explicit computation for oracle calls and building explicitly the subcomputation tree. Whenever the run of e_b demands an ordinal greater than that provided by e_c yet, break off the former computation and return to the latter. By hypothesis, at a certain point you will be able see that e_b's subcomputation tree is ill-founded. Then write $\sup_{n\in\omega} \lambda_n^{e_b}$, and halt. This would then be a halting computation of Λ, contradiction.

Unfortunately, we do not know which of b) or c) is excluded. For that matter, there could be no examples of either! Possibly all freezing computations are of type a), where those bounds over all freezing es are cofinal in Λ.

5 Parallel Oracle Calls

With sequential computation, as defined above, once an ill-founded oracle call is made, the entire computation is freezing. Parallel computation provides an alternative. In its essence, this is the same as with finite computation. In that setting, what should be the semantics of “A or B”? That both converge and one is true, or that one is true regardless of whether the other even converges? Similarly here, a machine could make a parametrized oracle call. This is perhaps most easily modeled by having another tape as part of the oracle call. The called computation asks for the convergence of a computation with index given on the first tape and inputs the second and third tapes. When making a call, the third tape is blank, but in generating the answer, the oracle substitutes all possible finite strings (equivalently: all integers) on the blank tape. If any return a convergent computation, the oracle answers “yes.” If none of them freeze and all return a divergent computation, the oracle answers “no.” If at least one of the parallel calls freezes and all those that do not diverge, then the oracle gives no answer and the current computation freezes.

Notice that the roles of convergence and divergence could be interchanged here, as convergent and divergent computations can be interchanged with each other:

given e, ask the oracle whether e converges; if yes, diverge, if no, halt. Of course, if e freezes, so does this.

Arguments similar to those above show that the parallel writable, parallel eventually writable, and parallel accidentally writable reals are all the same.

Although it seems likely, we do not have a proof that the parallel writable reals include strictly more than the feedback writables do.

6 Extending Convergence and Divergence Consistently

For both (sequential) feedback and parallel computation above, the semantics was given conservatively. That is, the convergence/divergence answers to oracle calls were forced on us. Evidence for such was an explicit computation in which some tree was well-founded, as so is absolute. Once well-foundedness is brought into the picture, induction cannot be too far behind. In fact, the process can be described via an inductive definition.

Let $\downarrow$ and $\uparrow$ be a disjoint pair of sets of computation calls, where a computation call is a pair consisting of (an index for) a program and a parameter. Given $(\downarrow, \uparrow)$, computations can be defined as convergent or divergent relative to that pair. For the sake of concreteness we will restrict attention to feedback computation; analogous considerations apply to parallel computation. When making oracle calls, the given pair $(\downarrow, \uparrow)$ is used as the oracle. This is deterministic, as $\downarrow$ and $\uparrow$ are disjoint. It is also monotonic: any computation that asks only oracle calls already in $\downarrow$ or $\uparrow$ will be unaffected by increasing either or both of those; all other computations are freezing, and so can only thaw by increasing those. As a monotonic operator, it has a least fixed point. This is the semantics given for feedback computation, that is the sense in which the semantics was conservative. This description of the matter does allow for considering other fixed points as possible semantics for these computational languages.

References

[1] Jon Barwise, **Admissible Sets and Structures**, Perspectives in Mathematical Logic, Springer, 1975

[2] Joel Hamkins and Andy Lewis, “Infinite Time Turing Machines,” **The Journal of Symbolic Logic**, v. 65 (2000), p. 567-604

[3] Robert Lubarsky, “Building Models of the μ-calculus,” unpublished

[4] Robert Lubarsky, “μ-definable Sets of Integers,” **Journal of Symbolic Logic**, v. 58 (1993), p. 291-313

[5] Michael Möllerfeld, "Generalized Inductive Definitions," Ph.D. thesis, Universität zu Münster, 2002

[6] Michael Rathjen, "Recent Advances in Ordinal Analysis," **The Bulletin of Symbolic Logic**, v. 1 (1995), p. 468-485

[7] Philip Welch, "The Length of Infinite Time Turing Machine Computations," **The Bulletin of the London Mathematical Society**, v. 32 (2000), p. 129-136

[8] Philip Welch, "Eventually Infinite Time Turing Machine Degrees: Infinite Time Decidable Reals," **The Journal of Symbolic Logic**, v. 65 (2000), p. 1193-1203

[9] Philip Welch, "Characteristics of Discrete Transfinite Turing Machine Models: Halting Times, Stabilization Times, and Normal Form Theorems," **Theoretical Computer Science**, v. 410 (2009), p. 426-442

Logspace without Bounds

Isabel Oitavem*

Dept. Matemática, Faculdade de Ciências e Tecnologia
Universidade Nova de Lisboa
P-2829-516 Caparica, Portugal
oitavem@fct.unl.pt

Abstract This paper provides a recursion-theoretic characterization of the functions computable in logarithmic space, without explicit bounds in the recursion schemes. It can be seen as a variation of the Clote and Takeuti characterization of logspace functions [7], which results from the implementation of an intrinsic growth-control within an input-sorted context.

Keywords. Implicit complexity, logarithmic space, normal/safe.

1 Introduction

A lot of work has been done in order to provide recursion-theoretic characterizations of relevant classes of computational complexity. Some of these machine independent characterizations address explicitly the resource constraints of the complexity classes by imposing bounds in the recursion schemes. This is, for instance, the case of the Cobham characterization of polytime functions [6], the Thompson characterization of polyspace functions [18], and the Clote and Takeuti characterization of NC and logspace functions [7]. Different techniques to implement an intrinsic growth-control have been developed and with them characterizations without explicit bounds in the recursion schemes have been achieved. Besides others, we mention: for the class of polytime functions [3], [10], [14]; for the class of polyspace functions [15], [17]; for NC [11], [5], [16], [4]. The aim of the present paper is to provide such a characterization for the logspace functions. Working in an input-sorted context, as in [3], and based on [7], we establish an implicit characterization of the class of functions computable in logarithmic space — *Logspace*.

There exists an implicit characterization of the small-output logspace functions due to Bellantoni, see [1]. By "small-output" we mean that the length of the output is logarithmically dominated by the length of the input. The functions considered

*This work was carried out at CMAF-UL, and was supported by FCT (project LogICCC/0001/2007 and PTDC/MAT/104716/2008).

by Bellantoni have numeric inputs and a sort of unary outputs. This non standard aspect is avoided in the characterization given here. Moreover, we characterize all logspace functions.

For other implicit characterizations of logspace (decidable sets) we refer to Jones [9], Møller Neergaard [12] and [13], and work of Hofmann [8]. As Hofmann wrote, separating logspace from P or NP seems more accessible than separating P from NP or such like, but surprisingly little work exists concerning implicit characterizations of logspace. In particular, little work exists concerning recursion-theoretic implicit characterizations of this class.

1.1 Notation

We work over the set $\{0,1\}^*$ of all finite binary sequences (i.e., leading zeroes are allowed), and we consider the standard notation related with it: $|x|$ is the length of the sequence x, ϵ is the sequence of length zero, $x\hat{\ }y$ is the concatenation of x with the sequence y, the string product $x \times y = x\hat{\ }\cdots\hat{\ }x$ is the concatenation of x with itself $|y|$ times (similar in growth to Buss' smash function). We consider $\{0,1\}^*$ ordered according to length and, within the same length, lexicographically and we denote this order by $<$. Thus, $\epsilon < 0 < 1 < 00 < 01 < 10 < 11 < 000 < \cdots$. x' denotes the successor sequence of x with respect to the order $<$, and $\min\{x, y\}$ denotes x if $x < y$ and y otherwise.

2 Inductive characterization of Logspace

Clote and Takeuti established, in [7], an inductive characterization of *Logspace* with explicit bounds in one of the recursion schemes. The bounded inductive characterization of *Logspace* we describe in this section results basically from rewriting over $\{0,1\}^*$ the characterization of Clote and Takeuti, which was originally given in numeric notation. Therefore, since there is no essential difference between them, we consider *Logspace* as being the smallest class of functions which contains the initial functions 1-12 and that is closed under the composition, bounded* recursion on notation[1] and concatenation recursion on notation schemes. For further reference, we denote this description of *Logspace* by *Logspace*$_{CT}$.

1) $E^n(x_1, \ldots, x_n) = \epsilon$ (zeroes)
2) $\Pi^n_j(x_1, \ldots, x_n) = x_j, 1 \leq j \leq n$ (projections)
3) $S_i(x) = xi,\ i \in \{0,1\}$ (string successors)

[1]For technical reasons, this scheme will be replaced by another bounded scheme, therefore at this stage we use the designation *bounded**.

4)	$P(\epsilon) = \epsilon \quad P(xi) = x,\ i \in \{0,1\}$	(string predecessor)
5)	$IP(\epsilon, x) = x \quad IP(yi, x) = P(IP(y,x)),$ $i \in \{0,1\}$	(iterated string predecessor)
6)	$s(x) = x'$	(numeric successor)
7)	$p(\epsilon) = \epsilon \quad p(x') = x$	(numeric predecessor)
8)	$Ip(\epsilon, x) = x \quad Ip(y', x) = p(Ip(y,x))$	(iterated numeric predecessor)
9)	$lh(\epsilon) = \epsilon \quad lh(xi) = s(lh(x)),\ i \in \{0,1\}$	(length)
10)	$U(\epsilon) = 0 \quad U(xi) = i,\ i \in \{0,1\}$	(last digit)
11)	$c(\epsilon, y, z) = y, c(x', y, z) = z$	(conditional)
12)	$\times(x, y) = x \times y$	(string product)

Composition: $f(\bar{x}) = g(\bar{r}(\bar{x}))$

Bounded* recursion on notation:

$$f(\epsilon, \bar{x}) = g(\bar{x})$$
$$f(yi, \bar{x}) = h_i(y, \bar{x}, f(y, \bar{x})), i \in \{0,1\}$$

provided that $f(y, \bar{x}) < lh(b(y, \bar{x}))$ holds for all $y, \bar{x}$, where b is a function already in the class.

Concatenation recursion on notation[2]:

$$f(\epsilon, \bar{x}) = g(\bar{x})$$
$$f(yi, \bar{x}) = S_{U(h_i(y,\bar{x}))}(f(y, \bar{x})), i \in \{0,1\}$$

In this context the bounded* recursion on notation scheme can be reformulated as follows, without affecting the defined class.

Bounded recursion on notation:

$$f(\epsilon, \bar{x}) = g(\bar{x})$$
$$f(yi, \bar{x}) = h_i(y, \bar{x}, \min\{f(y, \bar{x}), lh(b(y, \bar{x}))\}), i \in \{0,1\}$$

where, again, b is a function already in the class.

Let us denote by *Logspace*$_{\min}$ the class which results from replacing in *Logspace*$_{CT}$ the bounded* recursion on notation scheme by the bounded recursion on notation scheme above. It is obvious that the bounded recursion on notation scheme is not more restrictive than the bounded* recursion on notation scheme, and so *Logspace*$_{CT}$ $\subseteq$ *Logspace*$_{\min}$. If one notices that $\min$ can be defined in *Logspace*$_{CT}$ by $\min(x, y) = c(Ip(y, x), x, y)$, then the inclusion *Logspace*$_{\min}$ $\subseteq$ *Logspace*$_{CT}$ is also straightforward. Therefore, *Logspace*$_{CT}$ $=$ *Logspace*$_{\min}$ and one may define *Logspace* as follows:

[2]In the following $S_{U(z)}(w)$ abbreviates the function $c(p(U(z)), S_0(w), S_1(w))$.

Definition 2.1. *Logspace* is the smallest class of functions containing the initial functions 1-12 that is closed under the composition, bounded recursion on notation and concatenation recursion on notation schemes.

Remark 2.2. Given symbols α, β ($\alpha \neq \beta$), the natural numbers are bijectively encoded by strings in the alphabet $\{\alpha, \beta\}$. This is particularly well-known for the alphabet $\{1, 2\}$, where the natural number assigned to a string $s_l\, s_{l-1} \cdots s_1$ is $\sum_{i=1}^{l} s_i \cdot 2^{i-1}$. In general, the bijection $\pi : \mathbb{N} \to \{\alpha, \beta\}^*$ goes as follows:

$$\begin{array}{ccccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & \cdots \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \\ \epsilon & \alpha & \beta & \alpha\alpha & \alpha\beta & \beta\alpha & \beta\beta & \alpha\alpha\alpha & \cdots \end{array}$$

The usual addition function over $\mathbb{N}$ induces, by π, a 2-ary function over $\{\alpha, \beta\}^*$ that we represent by "$+$". We mean that, for all $m, n \in \mathbb{N}$, $\pi(m) + \pi(n) = \pi(m + n)$. It is routine to check that the function $+$ can be computed bit-by-bit (only an extra bit is required to carry information). Therefore, the function $+$ is computable in logarithmic space. In this paper, as in [10] and [15], we are interested in the case $\alpha = 0$ and $\beta = 1$. Therefore, the basic equations here are $x + \epsilon = x$, $\epsilon + x = x$, $0 + 0 = 1$, $0 + 1 = 1 + 0 = 00$, and $1 + 1 = 01$.

3 Implicit characterization of Logspace

Following ideas of Bellantoni and Cook, [3] or [2], we consider functions with two sorts of input positions — normal and safe positions. They are written by this order and they are separated by a semicolon as follows: $f(\bar{x}; y)$. Notice that here we only consider functions that have at most one variable in safe position. However, this restriction is not mandatory. As a matter of fact we could formalize everything with an arbitrary number of variables in safe positions. The point is that for our purposes it is enough to have one safe position.

Let us introduce the class of functions *Logs*.

Definition 3.1. *Logs* is the smallest class of functions containing the initial functions 1-12, described below, that is closed under the normal composition, safe recursion on notation, safe concatenation recursion on notation and safe log-transition recursion on notation schemes.

1) $E^n(x_1, \ldots, x_n;) = \epsilon$ (zeroes)
2) $\Pi^n_j(x_1, \ldots, x_n;) = x_j, 1 \leq j \leq n$ (projections)
3) $S_i(x;) = xi, i \in \{0, 1\}$ (string successors)
4) $P(\epsilon;) = \epsilon \quad P(xi;) = x, i \in \{0, 1\}$ (string predecessor)
5) $IP(\epsilon, x;) = x \quad IP(yi, x;) = P(IP(y, x;);)$, (iterated string

	$i \in \{0,1\}$	predecessor)
6)	$s(x;) = x'$	(successor)
7)	$p(\epsilon;) = \epsilon \quad p(x';) = x$	(numeric predecessor)
8)	$Ip(\epsilon, x;) = x \quad Ip(y', x;) = p(Ip(y, x;);)$	(iterated numeric predecessor)
9)	$lh(\epsilon;) = \epsilon \quad lh(xi;) = s(lh(x;);), i \in \{0,1\}$	(length)
10)	$U(\epsilon;) = 0 \quad U(xi;) = i, i \in \{0,1\}$	(last digit)
11)	$c(\epsilon, y, z;) = y, c(x', y, z;) = z$	(conditional)
12)	$\times(x, y;) = x \times y$	(string product)

Normal composition: $f(\bar{x}; y) = g(\bar{r}(\bar{x};); y)$

Safe recursion on notation:

$$f(\epsilon, \bar{x};) = g(\bar{x};)$$
$$f(yi, \bar{x};) = h_i(y, \bar{x}; f(y, \bar{x};)), i \in \{0,1\}$$

Safe concatenation recursion on notation[3]:

$$f(\epsilon, \bar{x};) = g(\bar{x};)$$
$$f(yi, \bar{x};) = S_{U(h_i(y,\bar{x};);)}(f(y, \bar{x};);), i \in \{0,1\}$$

Safe log-transition recursion on notation

$$f(xi, \bar{y}, z; \epsilon) = f(x, \bar{y}, z; \epsilon)$$
$$f(xi, \bar{y}, z; w') = f(x, \bar{y}, z'; w), i \in \{0,1\}$$
$$f(\epsilon, \bar{y}, z; w) = h(\bar{y}, z;)$$

It is obvious that the strength of this characterization is concentrated on the normal positions: all initial functions involve only variables in normal positions, the same happens with the safe concatenation recursion on notation scheme and we only have normal composition. However, in the safe recursion on notation scheme the recursive value $f(\bar{x}, y;)$ is placed in the safe position of h. Thus, the unique way to use the power of the safe recursion on notation scheme is via safe log-transition recursion on notation which, for x in normal position and w in safe position such that $w \leq lh(x)$, enables us to use w as if it was in a normal position. Informally, if f is defined by log-transition based on h, then $f(x, \bar{y}, z; w)$ leads to $h(\bar{y}, z + \min\{lh(x), w\};)$.

In other words, the safe composition scheme imposes a complete separation between normal and safe input positions. This separation is respected by all recursion schemes, except by the safe log-transition recursion scheme. The goal of this scheme is to allow $f(x, \bar{y}; w) = h(\bar{y}, w;)$ whenever $w \leq lh(x)$. Therefore, we call it "log-transition".

[3]In the following $S_{U(z;)}(w;)$ abbreviates the function $c(p(U(z;);), S_0(w;), S_1(w;);)$.

In fact, the safe log-transition recursion scheme can be replaced by the following scheme: $f(x, \bar{y}; w) = h(\bar{y}, \min(lh(x;), w;);)$, where $\min(z, w;)$ abbreviates $c(Ip(w, z;), z, w;)$. This, in particular, means that the log-transition is conceptually a composition scheme that we formulate recursively.

In order to prove that *Logspace* $=$ *Logs* we establish two lemmas:

Lemma 3.2. *For all* $F \in$ *Logspace there exists* $f \in$ *Logs such that*

$$\forall \bar{x}\ F(\bar{x}) = f(\bar{x};).$$

Proof. The proof is by induction on the complexity of F. The only relevant case is when F is defined by bounded recursion on notation:

$F(\epsilon, \bar{x}) = G(\bar{x})$

$F(yi, \bar{x}) = H_i(y, \bar{x}, \min\{F(y, \bar{x}), lh(B(y, \bar{x}))\}), i \in \{0, 1\}$.

By induction assumption, there exist $g, h_0, h_1, b \in$ *Logs* such that $\forall \bar{x}\ G(\bar{x}) = g(\bar{x};), \forall y, \bar{x}, z\ H_i(y, \bar{x}, z) = h_i(y, \bar{x}, z;)$ and $\forall y, \bar{x}\ B(y, \bar{x}) = b(y, \bar{x};)$. Therefore, we just have to define f, by safe recursion on notation, as follows:

$f(\epsilon, \bar{x};) = g(\bar{x};)$

$f(yi, \bar{x};) = h_i^*(y, \bar{x}; f(y, \bar{x};)), i \in \{0, 1\}$,

where h_i^* for any $i \in \{0, 1\}$ is defined, by normal composition, according to the expression $h_i^*(y, \bar{x}; z) = h_i^{**}(b(y, \bar{x};), y, \bar{x}, \epsilon; z)$ and h_i^{**} is defined by log-transition based on h_i, i.e.

$h_i^{**}(wj, y, \bar{x}, u; \epsilon) = h_i^{**}(w, y, \bar{x}, u; \epsilon)$

$h_i^{**}(wj, y, \bar{x}, u; w') = h_i^{**}(w, y, \bar{x}, u'; z),\ j \in \{0, 1\}$

$h_i^{**}(\epsilon, y, \bar{x}, u; z) = h_i(y, \bar{x}, u;)$. □

Lemma 3.3. *For all* $f \in$ *Logs there exist* $F, b \in$ *Logs such that* $\forall \bar{x}, y\ f(\bar{x}; y) = F(\bar{x}; \min\{y, lh(b(\bar{x};))\})$. *Moreover, the function* f *is not used in the definition of* b.

Proof. The proof is by induction on the complexity of f. The only non trivial case is when f is defined by log-transition. In this case we have

$f(xi, \bar{y}, z; \epsilon) = f(x, \bar{y}, z; \epsilon)$

$f(xi, \bar{y}, z; w') = f(x, \bar{y}, z'; w), i \in \{0, 1\}$

$f(\epsilon, \bar{y}, z; w) = h(\bar{y}, z;)$.

Setting $b(x, \bar{y}, z;) = x$ and considering $F(x, \bar{y}, z; \min\{w, lh(x)\})$ defined by log-transition based on h we achieve the desired result. □

Theorem 3.4. *Logspace* $=$ *Logs.*

Proof. It is immediate, by lemma 3.2, that *Logs* contains *Logspace*. Let us check the other inclusion. We have to show that for all $f \in$ *Logs* there exists $F \in$ *Logspace* such that $\forall \bar{x}, y\ f(\bar{x}; y) = F(\bar{x}, y)$. Let us proceed by induction on the complexity of f. For all initial functions of *Logs*, and for functions obtained by the normal composition or safe concatenation recursion on notation schemes the result is obvious. Lemma 3.3 turns the result into an obvious statement for functions defined by the safe recursion on notation scheme. Thus, we just have to prove that any function in *Logs* defined by the log-transition scheme is also definable in *Logspace*. Let us consider f defined by log-transition as follows:

$f(xi, \bar{y}, z; \epsilon) = f(x, \bar{y}, z; \epsilon)$
$f(xi, \bar{y}, z; w') = f(x, \bar{y}, z'; w)$
$f(\epsilon, \bar{y}, z; w) = h(\bar{y}, z;)$.

By induction hypothesis, there exist $H \in$ *Logspace* such that $\forall \bar{y}, z\ h(\bar{y}, z;) = H(\bar{y}, z)$. Then, since $+$ is a logspace function as claimed in remark 2.2, we achieve the desired result defining $F(x, \bar{y}, z, w) = H(\bar{y}, z + \min\{w, lh(x)\})$, where $\min\{w, lh(x)\}$ is $c(Ip(lh(x), w), w, lh(x))$. □

References

[1] Bellantoni S.: Predicative Recursion and Computational Complexity. Ph. D. Dissertation, University of Toronto (1993).

[2] Bellantoni S.: Predicative recursion and the polytime hierarchy. Feasible Mathematics II, ed. P.Clote and J.B.Remmel, pp.15-29 (1995).

[3] Bellantoni S., Cook S.: A new recursion-theoretic characterization of polytime functions. Computational Complexity 2, pp.97-110 (1992).

[4] Bonfante G., Kahle R., Marion J.-Y.: Recursion schemeta for NC^k. LNCS, Springer-Verlag, pp.49-63 (2008).

[5] Bellantoni S., Oitavem I.: Separating NC along the δ axis. ICC Special issue of Theoretical Computer Science 318, pp.57-78 (2004).

[6] Cobham A.: The intrinsic computational difficulty of functions. Proc. of the 1964 International Congress for Logic, Methodology, and the Philosophy of Science, ed. Y. Bar-Hillel, North Holland, Amsterdam, pp.24-30 (1965).

[7] Clote P., Takeuti G.: First order bounded arithmetic and small boolean circuit complexity classes. Feasible Mathematics II, ed. P.Clote and J.B.Remmel, pp.154-218 (1995).

[8] Hofmann M.: Programming languages for logarithmic space. Talk given at Geocal'06, 13th Feb.2006, http://www-lipn.univ-paris13.fr/~baillot/GEOCAL06/SLIDES/Hofmann1302.pdf

[9] Jones N.: Logspace and Ptime characterized by programming languages. Theoretical Computer Science 228, pp.151-174 (1999).

[10] Leivant D.: Ramified recurrence and computational complexity I: word recurrence and polytime. Feasible Mathematics II, ed. P.Clote and J.B.Remmel, pp.320-343 (1995).

[11] Leivant D.: A characterization of NC by tree recurrence. Proceedings of the 39th Annual Symposium on Foundations of Computer Science, FOCS 1998, IEEE Computer Society, pp.716-724 (1998).

[12] Møller Neergaard P.: A functional language for logarithmic space. Prog. Lang. and Systems: 2nd Assian Symp. (APLAS 2004), LNCS 3302, pp.311-326.

[13] Møller Neergaard P.: BC_{ϵ}^{-}: A recursion-theoretic characterization of Logspace. Technical report, Brandeis University, March 2004.

[14] Niggl K.-H.: The μ-Measure as a tool for classifying computational complexity. Archive for Mathematical Logic 39(7), pp. 509-514 (2000).

[15] Oitavem I.: New recursive characterizations of the elementary functions and the functions computable in polynomial space. Revista Matematica de la Universidad Complutense de Madrid 10(1), pp.109-125 (1997).

[16] Oitavem I.: Characterizing NC with tier 0 pointers. Mathematical Logic Quarterly 50(1), pp.9-17 (2004).

[17] Oitavem I.: Characterizing Pspace with pointers. Mathematical Logic Quarterly 54(3), pp.317-323 (2008).

[18] Thompson D.: Subrecursiveness: machine-independent notions of computability in restricted time and storage. Mathematical Systems Theory 6(1), pp.3-15 (1971).

Investigations of Subsystems of Second Order Arithmetic and Set Theory in Strength between Π^1_1-CA and Δ^1_2-CA + BI: Part I

Michael Rathjen

Abstract This paper is the first of a series of two. It contains proof–theoretic investigations on subtheories of second order arithmetic and set theory. Among the principles on which these theories are based one finds autonomously iterated positive and monotone inductive definitions, Π^1_1 transfinite recursion, Δ^1_2 transfinite recursion, transfinitely iterated Π^1_1 dependent choices, extended Bar rules for provably definable well-orderings as well as their set-theoretic counterparts which are based on extensions of Kripke-Platek set theory. This first part introduces all the principles and theories. It provides lower bounds for their strength measured in terms of the amount of transfinite induction they achieve to prove. In other words, it determines lower bounds for their proof-theoretic ordinals which are expressed by means of ordinal representation systems. The second part of the paper will be concerned with ordinal analysis. It will show that the lower bounds established in the present paper are indeed sharp, thereby providing the proof-theoretic ordinals. All the results were obtained more then 20 years ago (in German) in the author's PhD thesis [43] but have never been published before, though the thesis received a review (MR 91m#03062). I think it is high time it got published.

1 Introduction

To set the stage for the following, a very brief history of ordinal-theoretic proof theory from the time after Gentzen's death until the early 1980s reads as follows: In the 1950's proof theory flourished in the hands of Schütte. In [57] he introduced an infinitary system for first order number theory with the so-called ω-rule, which had already been proposed by Hilbert [23]. Ordinals were assigned as lengths to derivations and via cut-elimination he re-obtained Gentzen's ordinal analysis for number theory in a particularly transparent way. Further, Schütte extended his approach to systems of ramified analysis and brought this technique to perfection in his monograph "Beweistheorie" [58]. Independently, in 1964 Feferman [13] and Schütte [59], [60] determined the ordinal bound Γ_0 for theories of autonomous ramified progressions.

A major breakthrough was made by Takeuti in 1967, who for the first time obtained an ordinal analysis of a strong fragment of second order arithmetic. In [67] he gave an ordinal analysis of Π^1_1 comprehension, extended in 1973 to Δ^1_2 comprehension in [68]. For this Takeuti returned to Gentzen's method of assigning ordinals (ordinal diagrams, to be precise) to purported derivations of the empty sequent (inconsistency).

The next wave of results, which concerned theories of iterated inductive definitions, were obtained by Buchholz, Pohlers, and Sieg in the late 1970's (see [10]). Takeuti's methods of reducing derivations of the empty sequent ("the inconsistency") were extremely difficult to follow, and therefore a more perspicuous treatment was to be hoped for. Since the use of the infinitary ω-rule had greatly facilitated the ordinal analysis of number theory, new infinitary rules were sought. In 1977 (see [5]) Buchholz introduced such rules, dubbed Ω-rules to stress the analogy. They led to a proof-theoretic treatment of a wide variety of systems, as exemplified in the monograph [11] by Buchholz and Schütte. Yet simpler infinitary rules were put forward a few years later by Pohlers, leading to the *method of local predicativity*, which proved to be a very versatile tool (see [40–42]). With the work of Jäger and Pohlers (see [28, 29, 33]) the forum of ordinal analysis then switched from the realm of second-order arithmetic to set theory, shaping what is now called *admissible proof theory*, after the models of *Kripke-Platek set theory*, $\mathbf{KP}$. Their work culminated in the analysis of the system with Δ^1_2 comprehension plus Bar induction, (BI), [33]. In essence, admissible proof theory is a gathering of cut-elimination techniques for infinitary calculi of ramified set theory with Σ and/or Π_2 reflection rules[1] that lend itself to ordinal analyses of theories of the form $\mathbf{KP}+$ *"there are x many admissibles"* or $\mathbf{KP}+$ *"there are many admissibles"*. By way of illustration, the subsystem of analysis with Δ^1_2 comprehension and Bar induction can be couched in such terms, for it is naturally interpretable in the set theory $\mathbf{KPi} := \mathbf{KP} + \forall y \exists z (y \in z \wedge z \textit{ is admissible})$ (cf. [33]).

The investigations of this paper focus, as far as subsystems of second order arithmetic are concerned, on theories whose strength strictly lies in between that of Δ^1_2-$\mathbf{CA}$ and Δ^1_2-$\mathbf{CA} + (\mathrm{BI})$. Δ^1_2-$\mathbf{CA}$ is actually not much stronger than Π^1_1-$\mathbf{CA}$, the difference being that the latter theory allows one to carry out iterated hyperjumps of length $< \omega$ while the former allows one to carry out iterated hyperjumps of length $< \varepsilon_0$. The jump from Δ^1_2-$\mathbf{CA}$ to Δ^1_2-$\mathbf{CA} + (\mathrm{BI})$ is indeed enormous. By comparison, even the ascent from Π^1_1-$\mathbf{CA}$ to Δ^1_2-$\mathbf{CA} + \mathrm{BR}$ (with BR referring to the Bar rule) is rather benign. To get an appreciation for the difference one might also point out that all hitherto investigated subsystems of second order arithmetic in

[1] Recall that the salient feature of admissible sets is that they are models of Δ_0 collection and that Δ_0 collection is equivalent to Σ reflection on the basis of the other axioms of $\mathbf{KP}$ (see [3]). Furthermore, admissible sets of the form L_α also satisfy Π_2 reflection.

the range from Π^1_1-$\mathbf{CA}_0$ to Δ^1_2-$\mathbf{CA}+\mathrm{BR}$ can be reduced (as far as strength is concerned) to first order theories of iterated inductive definitions. The theories investigated here are beyond that level. Among the principles on which these theories are based one finds autonomously iterated positive and monotone inductive definitions, Π^1_1 transfinite recursion, Δ^1_2 transfinite recursion, transfinitetely iterated Π^1_1 dependent choices, extended Bar rules for provably definable well-orderings as well as their set-theoretic counterparts which are based on extensions of Kripke-Platek set theory. This first part introduces all the principles and theories. It provides lower bounds for their strength measured by the amount of provable transfinite induction. In other words, it determines lower bounds for their proof-theoretic ordinals which are expressed by means of ordinal representation systems. The second part of the paper will be concerned with ordinal analysis. It will show that the lower bounds established in the present paper are indeed sharp, thereby providing the proof-theoretic ordinals. All the results were obtained more then 20 years ago (in German) in the author's PhD thesis [43] but have never been published before, though the thesis received a review (MR 91m#03062). I always thought that the results in my thesis were worth publishing but in the past I never seemed to have enough time to sit down for six weeks and type the entire PhD thesis again. The thesis was produced by the now obsolete word processing system "Signum" and it was also written in German. Over the past 20 years or so academic life has changed in that time, e.g. for doing research, has become a luxury good. I would like to thank Andreas Weiermann for nudging me again and again to publish it.

Zueignung

Den kreatürlichen Freunden Bobby, Honky Tonk, Schnuffi und Marlene gewidmet.

Outline of the paper

In the following I give a brief outline of the contents of this paper. It is roughly divided into two chapters. The first chapter, entitled "THEORIES", introduces the background and presents all the principles and theories to be considered. It also establishes interrelationships between various theories. The second chapter, entitled "WELL-ORDERING PROOFS", introduces an ordinal representation system and establishes lower bounds for the proof-theoretic ordinals of most of the theories considered.

Section 2 carefully defines the basic theory of arithmetical comprehension, $\mathbf{ACA}_0$, which forms the basis for all subsystems of second order arithmetic, and also the basic set theory $\mathbf{BT}$ which forms the basis of all set theories. While such attention to detail will not matter that much for the present paper it will certainly be of importance to its sequel which features proof analyses of infinitary calculi. Section 3 introduces second order theories of iterated inductive definitions. Systems investigated in the literature before used to be first order theories with the inductively defined sets being captured via additional predicates and iterations restricted to arithmetical well-orderings. Going to second order theories allows one to formalize iterations along arbitrary well-orderings and also to address the more general scenario of monotone inductive definitions. Section 4 compares the theories of the foregoing section with theories of transfinite Π^1_1 comprehension. In section 5 it is shown that theories of iterated inductive definitions can be canonically translated into set theories of iterated admissibility. This translation exploits the structure theory of Σ_+-inductive definitions on admissible sets originating in Gandy's Theorem (cf. [3, VI]). Section 6 features iterations based on stronger operations such as Δ^1_2 comprehension and Σ^1_2 dependent choices. Section 7 deals with their set-theoretic counterparts which are to be found in certain forms of Σ recursion.

In order to approach the strength of Δ^1_2-$\mathbf{CA}$ + (BI) it is natural to restrict the schema (BI) to specific syntactic complexity classes of formulae, ($\mathcal{F}$-BI). An alternative consists in directing the attention to the well-ordering over which transfinite induction is allowed in that one requires them to be provably well-ordered or parameter-free. This will be the topic of section 8. Particular rules and schemata considered include the rule $\mathrm{BR}(\text{impl-}\Sigma^1_2)$ and the schema $\mathrm{BI}(\text{impl-}\Sigma^1_2)$:

$$(\mathrm{BR}(\text{impl-}\Sigma^1_2)) \quad \frac{\exists! X\,(\mathrm{WO}(X) \wedge G[X])}{\forall X\,(\mathrm{WO}(X) \wedge G[X] \to \mathrm{TI}(X,H))}$$

where $G[U]$ is a Σ^1_2 formula (without additional parameters), $H(a)$ is an arbitrary $\mathcal{L}_2$ formula, $\mathrm{WO}(X)$ expresses that X is a well-ordering, and $\mathrm{TI}(X,H)$ expresses the instance of tranfinite induction along X with the formula $H(a)$.

$$(\mathrm{BI}(\text{impl-}\Sigma^1_2)) \quad \exists! X\,(\mathrm{WO}(X) \wedge G[X]) \to \forall X\,(\mathrm{WO}(X) \wedge G[X] \to \mathrm{TI}(X,H))$$

where $G[U]$ is a Σ^1_2 formula (without additional parameters) and $H(a)$ is an arbitrary $\mathcal{L}_2$ formula.

The rule $\mathrm{BR}(\text{impl-}\Sigma^1_2)$ is, on the basis of Δ^1_2-$\mathbf{CA}$, much stronger than the rule BR whereas $\mathrm{BR}(\text{impl-}\Sigma^1_2)$ is still much weaker than (BI). The difference in strength between (BI) and $\mathrm{BR}(\text{impl-}\Sigma^1_2)$ is of course owed to the fact that the first is a rule while the second is a schema. But one can say something more illuminative about it. As it turns out, $\mathrm{BR}(\text{impl-}\Sigma^1_2)$ and $\mathrm{BI}(\text{impl-}\Sigma^1_2)$ are of the same strength (on the basis of Δ^1_2-$\mathbf{CA}$), in actuality the theories Δ^1_2-$\mathbf{CA}$+$\mathrm{BR}(\text{impl-}\Sigma^1_2)$ and Δ^1_2-$\mathbf{CA}$ + $\mathrm{BI}(\text{impl-}\Sigma^1_2)$ prove the same Π^1_1 statements. Thus the main difference between $\mathrm{BR}(\text{impl-}\Sigma^1_2)$ and (BI) is to be found in the premiss of $\mathrm{BI}(\text{impl-}\Sigma^1_2)$ requiring the well-ordering to be describable via a Σ^1_2 formula without parameters.

Section 8 also considers set-theoretic versions of $(\mathrm{BR}(\text{impl-}\Sigma^1_2)$ and $(\mathrm{BI}(\text{impl-}\Sigma^1_2))$ which can be viewed as formal counterparts of the notion of a good Σ_1 definition of an ordinal/set known from the theory of admissible sets (cf. [3, II.5.13]).

With the next section we enter the second chapter of this paper. Sections 9 and 10 develop an ordinal representation system $\mathrm{OT}(\Phi)$ which will be sufficient unto the task of expressing the proof-theoretic ordinals of all the foregoing theories.

Section 11 introduces the technical basis for well-ordering proofs. By a well-ordering proof in a given theory T we mean a proof formalizable in T which shows that a certain ordinal representation system (or a subset of it) is well-ordered. The notion of a *distinguished set* (of ordinals) (in German: ausgezeichnete Menge) will be central to carrying out well ordering proofs in the various subtheories of second order arithmetic introduced in earlier sections. A theory of distinguished sets developed for this purpose emerged in the works of Buchholz and Pohlers [4, 6, 7].

The remaining sections 12-15 are concerned with well-ordering proofs for most of the theories featuring in this paper. The lower bounds for the proof-theoretic ordinals of theories established in this article turn out to be sharp. Proofs of upper bounds, though, will be dealt with in the second part of this paper which will be devoted to ordinal analysis. The final section of this paper provides a list of all theories and their proof-theoretic ordinals.

I. THEORIES

2 The formal set-up

This section introduces the languages of second order arithmetic and set theory with the natural numbers as urelements. Moreover, a collection of theories, comprehension and induction principles formalized in these languages will be introduced. Our presentation of second order arithmetic is equivalent to those found in the standard literature (e.g. [10, 63]). The same applies to set theory with urelements, where we follow the standard reference [3]. Slight deviations are of a purely technical nature, one pecularity being that we define formulae in such a way that negations occur only in front of prime formulae, another being that function symbols will be avoided. Instead, we axiomatize number theory by means of relation symbols representing their graphs.

2.1 The language $\mathcal{L}_2$

The vocabulary of $\mathcal{L}_2$ consists of free number variables $a_0, a_1, a_2, \ldots$, bound number variables $x_0, x_1, x_2, \ldots$, free set variables $U_0, U_1, U_2, \ldots$, bound set variables $X_0, X_1, X_2, \ldots$, the logical constants $\neg, \wedge, \vee, \forall, \exists$, the constants (numerals) $\bar{n}$ for each $n \in \mathbb{N}$, a 1-place relation symbol P, three 2-place relation symbols $\in, \equiv$, SUC and two 3-place relation symbols ADD, MULT. In addition, $\mathcal{L}_2$ has auxiliary symbols such as parentheses and commas. The intended interpretation of these symbols is the following:

1. Number variables range over natural numbers while set variables range over sets of natural numbers.
2. The constant $\bar{n}$ denotes the nth natural number.
3. P stands for an arbitrary set of natural numbers.
4. $\in$ denotes the elementhood relation between natural numbers and sets of natural numbers.
5. $\equiv$ denotes the identity relation between natural numbers.
6. SUC, ADD, and MULT denote the graphs of the numerical functions $n \mapsto n+1$, $(n,m) \mapsto n+m$, and $(n,m) \mapsto n \cdot m$, respectively.

The *terms* of $\mathcal{L}_2$ are the free number variables and the constants $\bar{n}$. As syntactical we also use a, b, c, d, e for free number variables, R, S, U, V for free set variables, u, v, w, x, y, z, i, j for bound number variables, W, X, Y, Z for bound set variables, r, s, t for terms, and A, B, C, D, F, G, H for formulae of $\mathcal{L}_2$. If E is an expression, $\tau_1, \ldots, \tau_n$ are distinct primitive symbols and $\sigma_1, \ldots, \sigma_n$ are arbitrary expressions, then by $E(\tau_1, \ldots, \tau_n \mid \sigma_1, \ldots, \sigma_n)$ we mean the expression obtained from A by writing σ_i in place of τ_i at each occurrence of τ_i. If A is a formula of the form $B(\tau_1, \ldots, \tau_n \mid \sigma_1, \ldots, \sigma_n)$ then this fact will also be expressed (less accurately) by writing B as $B(\tau_1, \ldots, \tau_n)$ and A as $B(\sigma_1, \ldots, \sigma_n)$.

Definition 2.1. *The* atomic formulae *of $\mathcal{L}_2$ are of the form $(s \equiv t)$, $(s \in U)$,* $\mathrm{SUC}(s,t)$, $\mathrm{P}(s)$, $\mathrm{ADD}(s,t,r)$, *and* $\mathrm{MULT}(s,t,r)$.

*The $\mathcal{L}_2$-*formulae *are defined inductively as follows: If A is an atomic formula then A and $\neg A$ are $\mathcal{L}_2$-formulae. If A and B are $\mathcal{L}_2$-formulae then so are $(A \wedge B)$ and $(A \vee B)$. If $F(a)$ is an $\mathcal{L}_2$-formula in which the bound number variable x does not occur then $\forall x\, F(x)$ and $\exists x\, F(x)$ are $\mathcal{L}_2$-formulae. If $G(U)$ is an $\mathcal{L}_2$-formula in which the bound set variable X does not occur then $\forall X\, G(X)$ and $\exists X\, G(X)$ are $\mathcal{L}_2$-formulae.*

The negation*, $\neg A$, of a non–atomic formula A is defined to be the formula obtained from A by*
(i) putting $\neg$ in front any atomic subformula,
(ii) replacing $\wedge, \vee, \forall x, \exists x, \forall X, \exists X$ by $\vee, \wedge, \exists x, \forall x, \exists X, \forall X$, respectively, and
(iii) dropping double negations.

As usual, $(A \to B)$ abbreviates $(\neg A \vee B)$ and $(A \leftrightarrow B)$ stands for $((A \to B) \wedge (B \to A))$. Outer most parentheses will usually be dropped. We write $s \neq t$ for $\neg(s \equiv t)$ and $s \notin U$ for $\neg(s \in U)$. To avoid parenthesis we also adopt the conventions that $\neg$ binds more strongly than the other connectives and that $\wedge, \vee$ bind more strongly than $\to$ and $\leftrightarrow$.

We also use the following abbreviations with $Q \in \{\forall, \exists\}$:
$Qx_1 \ldots x_n\, F(x_1, \ldots, x_n) \;:=\; Q\,x_1 \ldots Q\,x_n\, F(x_1, \ldots, x_n)$,
$Q\,X_1 \ldots, X_n\, F(X_1, \ldots, X_n) \;:=\; QX_1 \ldots QX_n\, F(X_1, \ldots, X_n)$,
and $\forall x \exists! y\, H(x, y) \;:=\; \forall x \exists y\, H(x, y) \;\wedge\; \forall xyz\, (H(x, y) \;\wedge\; H(x, z) \to y \equiv z)$.

Definition 2.2. *The formula class Π^1_0 (as well as Σ^1_0) consists of all arithmetical $\mathcal{L}_2$-formulae, i.e., all formulae which do not contain set quantifiers.*

If $F(U)$ is a Σ^1_n-formula (Π^1_n-formula) then $\forall X\, F(X)$ ($\exists X\, F(X)$) is a Π^1_{n+1} (Σ^1_{n+1}) formula.

2.2 The theory $\mathbf{ACA_0}$

As a base for all theories in the language $\mathcal{L}_2$ we use the theory $\mathbf{ACA}_0$ which in addition to the usual number-theoretic axioms has the axiom schema of arithmetical comprehension and an induction axiom for sets. As we will subject these theories to proof-theoretic treatment we shall present the axiomatization of $\mathbf{ACA}_0$ in more detail than would otherwise be necessary.

Definition 2.3. *The mathematical axioms of $\mathbf{ACA}_0$ are the following:*

(i) Equality axioms

(G1) $\forall x\, (x \equiv x)$.

(G2) $\forall xy\, (x \equiv y \to [F(x) \leftrightarrow F(y)])$ *for $F(a)$ in Π^1_0.*

(G3) $\bar{n} \equiv \bar{n}$.

(G4) $\bar{n} \not\equiv \bar{m}$ *if n, m are different natural numbers.*

(i) Axioms for $\mathrm{SUC}, \mathrm{ADD}, \mathrm{MULT}$.

(SUC1) $\forall x\, \exists! y\, \mathrm{SUC}(x, y)$.

(SUC2) $\forall y\,[y \equiv \bar{0} \;\vee\; \exists x\,\mathrm{SUC}(x,y)]$.

(SUC3) $\forall xyz\,(\mathrm{SUC}(x,z) \;\wedge\; \mathrm{SUC}(y,z) \to x \equiv y)$.

(SUC4) $\mathrm{SUC}(\bar{n}, \overline{n+1})$.

(SUC5) $\neg\mathrm{SUC}(\bar{n}, \bar{m})$ *if* $n+1 \neq m$.

(ADD1) $\forall xy\,\exists! z\,\mathrm{ADD}(x,y,z)$.

(ADD2) $\forall x\,\mathrm{ADD}(x,\bar{0},x)$.

(ADD3) $\forall uvwxy\,[\mathrm{ADD}(u,v,w) \wedge \mathrm{SUC}(v,x) \wedge \mathrm{SUC}(w,y) \;\to\; \mathrm{ADD}(u,x,y)]$.

(ADD4) $\mathrm{ADD}(\bar{n}, \bar{m}, \overline{n+m})$.

(ADD5) $\neg\mathrm{ADD}(\bar{n}, \bar{m}, \bar{k})$ *if* $n+m \neq k$.

(MULT1) $\forall xy\,\exists! z\,\mathrm{MULT}(x,y,z)$.

(MULT2) $\forall x\,\mathrm{MULT}(x,\bar{0},\bar{0})$.

(MULT3) $\forall uvwxy\,[\mathrm{MULT}(u,v,w) \;\wedge\; \mathrm{SUC}(v,x) \;\wedge\; \mathrm{ADD}(w,u,y) \;\to\; \mathrm{MULT}(u,x,y)]$.

(MULT4) $\mathrm{MULT}(\bar{n}, \bar{m}, \overline{n \cdot m})$.

(MULT5) $\neg\mathrm{MULT}(\bar{n}, \bar{m}, \bar{k})$ *if* $n \cdot m \neq k$.

(iii) Induction Axiom

$$\text{(Ind)} \qquad \forall X\,[\bar{0} \in X \wedge \forall xy\,[\mathrm{SUC}(y,x) \wedge y \in X \to x \in X] \;\to\; \forall x\,(x \in X)].$$

(iv) Arithmetical Comprehension

$$\text{(ACA)} \qquad \exists X\,\forall y\,[y \in X \leftrightarrow F(y)]$$

where $F(a)$ is Π^1_0 and X does not occur in $F(a)$.

As logical rules and axioms for every theory formulated in the language of $\mathcal{L}_2$ we choose the following:

(L1) All formulae of $\mathcal{L}_2$ that are valid in propositional logic.

(L2) The number quantifier axioms $\forall x\,F(x) \;\to\; F(t)$ and $F(t) \;\to\; \exists x\,F(x)$ for every $\mathcal{L}_2$-formula $F(a)$ in which x does not occur and every term t .

(L3) The set quantifier axioms $\forall X\, H(X) \to H(U)$ *and* $H(U) \to \exists X\, F(X)$ *for every* $\mathcal{L}_2$*-formula* $H(V)$ *in which* X *does not occur and set variable* U*.*

(L4) Modus ponens*: From* A *and* $A \to B$ *deduce* B*.*

(L5) From $A \to F(a)$ *deduce* $A \to \forall x\, F(x)$ *and from* $F(a) \to A$ *deduce* $\exists x\, F(x) \to A$ *providing the free number variable* a *does not occur in the conclusion and* x *does not occur in* $F(a)$*.*

(L6) From $A \to H(U)$ *deduce* $A \to \forall X\, H(X)$ *and from* $H(U) \to A$ *deduce* $\exists X\, F(X) \to A$ *providing the free set variable* U *does not occur in the conclusion and* X *does not occur in* $F(U)$*.*

We write $\mathbf{T} \vdash A$ *when* T *is a theory in the language of* $\mathcal{L}_2$ *and* A *can be deduced from* T *using the axioms of* T *and any combination of the preceding axioms and rules of* $\mathbf{ACA}_0$*.*

By $\mathbf{ACA}$ *we denote the theory* $\mathbf{ACA}_0$ *augmented by the scheme of induction for all* $\mathcal{L}_2$*-formulae:*

$$\text{(IND)} \qquad F(\bar{0}) \wedge \forall xy\, [\mathrm{SUC}(y,x) \wedge F(y) \to F(x)] \to \forall x\, F(x)$$

where $F(a)$ *is an arbitrary formula of* $\mathcal{L}_2$*.*

The sublanguage of $\mathcal{L}_2$ *without set variables will be denoted by* $\mathcal{L}_1$*.*

2.3 The languages $\mathcal{L}^*$ and $\mathcal{L}^*(\mathrm{M})$

$\mathcal{L}^*$ will be the language of set theory with the natural numbers as urelements. $\mathcal{L}^*$ comprises $\mathcal{L}_1$ and in addition has a constant N for the set of natural numbers, a 1-place predicate symbol Set for the class of sets, and a 1-place predicate symbol Ad for the class of admissible sets. The intended interpretations of $\mathcal{L}_1$ and $\mathcal{L}^*$ diverge with respect to the scopes of the quantifiers $\forall x$ and $\exists x$ which in the case of $\mathcal{L}^*$ are viewed as ranging over all sets and urelements. Moreover, $\mathcal{L}^*$ has also bounded quantifiers $(\forall x \in t)$ and $(\exists x \in t)$ which will be treated as quantifiers in their own right.

We will also have use for an extended language $\mathcal{L}^*(\mathrm{M})$ which has a constant M, intended to denote the smallest admissible set.

The terms of $\mathcal{L}^*$ ($\mathcal{L}^*(\mathrm{M})$) consist of the free variables and the constants $\bar{n}$ and N (and M).

The atomic formulae of $\mathcal{L}^*$ ($\mathcal{L}^*(\mathrm{M})$) consists of all strings of symbols of the forms $(s \equiv t)$, $(s \in t)$, $\mathrm{P}(t)$, $\mathrm{SUC}(s,t)$, $\mathrm{ADD}(s,t,r)$, $\mathrm{MULT}(s,t,r)$, $\mathrm{Ad}(s)$, and $\mathrm{Set}(s)$, where s, t, r are arbitrary terms of $\mathcal{L}^*$ ($\mathcal{L}^*(\mathrm{M})$).

Definition 2.4. *$\mathcal{L}^*$-formulae are inductively defined as follows:*

1. *A and $\neg A$ are $\mathcal{L}^*$-formulae whenever A is an atomic $\mathcal{L}^*$-formula.*
2. *If A and B are $\mathcal{L}^*$-formulae so are $(A \wedge B)$ and $(A \vee B)$.*
3. *If $F(a)$ is an $\mathcal{L}^*$-formula in which x does not appear and t is an $\mathcal{L}^*$ term then $\forall x\, F(x)$, $\exists x\, F(x)$, $(\forall x \in t)\, F(x)$, and $(\exists x \in t)\, F(x)$ are $\mathcal{L}^*$-formulae.*

$\mathcal{L}^*(\mathrm{M})$-formulae are defined in a similar vein. The negation, *$\neg A$, of a formula A is defined as in Definition 2.1, but extended by the clauses $\neg(\forall x \in t)F(x) := (\exists x \in t)\neg F(x)$ and $\neg(\exists x \in t)F(x) := (\forall x \in t)\neg F(x)$ for the bounded quantifiers.*

Definition 2.5 (Translating $\mathcal{L}_2$ into $\mathcal{L}^*$)**.** *Let $U_i^* := a_{2\cdot i}$, $X_i^* := x_{2\cdot i+2}$, $a_i^* := a_{2\cdot i+1}$, $x_i^* := x_{2\cdot i+1}$, and $\bar{n}^* := \bar{n}$.*

To every $\mathcal{L}_2$-formula A we assign an $\mathcal{L}^*$-formula A^* as follows: Replace every free variable $\mathcal{X}$ in A by $\mathcal{X}^*$. Replace number quantifiers $\forall x \,...x...$ and $\exists x \,...x...$ by $(\forall x^* \in \mathrm{N})\,...x^*...$ and $(\exists x^* \in \mathrm{N})\,...x^*...$, respectively. Replace set quantifiers $\forall X \,...X...$ and $\exists X \,...X...$ by $\forall X^*[\mathrm{Set}(X^*) \wedge X^* \subseteq \mathrm{N} \rightarrow ...X^*...]$ and $\exists X^*[\mathrm{Set}(X^*) \wedge X^* \subseteq \mathrm{N} \wedge ...X^*...]$, respectively, where $X^* \subseteq \mathrm{N}$ stands for $\forall u\,[u \in X^* \rightarrow u \in \mathrm{N}]$. The translation $A \mapsto A^*$ provides an embedding of $\mathcal{L}_2$ into $\mathcal{L}^*$, preserving the intended interpretations. In what follows we view $\mathcal{L}_2$ as sublanguage of $\mathcal{L}^*$, formally fixed by the natural translation $\ ^*$.

2.4 Syntactic classifications

Definition 2.6. *The Δ_0* formulae *are the smallest collection of $\mathcal{L}^*$ formulae containing all quantifier-free formulae closed under $\neg, \wedge, \vee$ and bounded quantification. Spelled out in detail the last closure clause means that if $F(a)$ is Δ_0, t is a term and x is a bound variable not occurring in $F(a)$ then $(\exists x \in t)F(x)$ and $(\forall x \in t)F(x)$ are Δ_0.*

$\mathcal{L}^*$ formulae which are Δ_0 or of the form $\exists x\, F(x)$ with $F(a)$ Δ_0 are said to be Σ_1. Dually, a formula is Π_1 if it is the negation of a Σ_1 formula.

A formula is a Σ formula *(Π* formula*) if it belongs to the smallest collection of formulae containing the Δ_0 formulae which is closed under $\wedge, \vee$, bounded quantification, and existential (universal) quantification.*

Definition 2.7. *The collection of P-positive Δ_0 formulae of $\mathcal{L}^*$, $\Delta(\mathrm{P}^+)$, is inductively generated from the Δ_0 formulae in which P does not occur and all formulae $\mathrm{P}(t)$ by closing off under $\vee, \wedge, (\forall x \in s)$, and $(\exists x \in s)$.*

The collection of P-positive arithmetical formulae, $\Pi^1_0(\mathrm{P}^+)$, is the collection of $\mathcal{L}_2$

formulae generated from the arithmetical formulae in which P *does not occur and the atomic formulae* $\mathrm{P}(t)$ *by closing off under* $\wedge, \vee$*, and numerical quantification.*

Remark 2.8. If A is $\Pi^1_0(\mathrm{P}^+)$ then A^* is $\Delta_0(\mathrm{P}^+)$.

Definition 2.9. *For* $\mathcal{L}^*$ *formulae* A *and terms* s *the relativization of* A *to* s*,* A^s*, arises from* A *by restricting all unbounded quantifiers* $\forall x \ldots$ *and* $\exists x \ldots$ *to* s*, i.e., by replacing them with* $(\forall x \in s) \ldots$ *and* $(\exists x \in s) \ldots$*, respectively.*

Note that A^s *is always a* Δ_0 *formula.*

Many mathematical and set-theoretic predicates have Δ_0 formalizations. For those that occur most frequently we introduce abbreviations:
$\mathrm{Tran}(s) := (\forall x \in s)(\forall y \in x)\,(y \in s)$.
$\mathrm{Ord}(s) := \mathrm{Tran}(s) \wedge (\forall x \in s)\mathrm{Tran}(x)$.
$\mathrm{Lim}(s) := \mathrm{Ord}(s) \wedge (\exists x \in s)(x \in s) \wedge (\forall x \in s)(\exists y \in s)(x \in y)$.
$s \subseteq t := (\forall x \in s) s \in t$.
$(s = t) := (\mathrm{Set}(s) \wedge \mathrm{Set}(t) \wedge s \subseteq t \wedge t \subseteq s) \vee (s \in \mathrm{N} \wedge t \in \mathrm{N} \wedge s \equiv t)$.
$(s = \{t, r\}) := t \in s \wedge r \in s \wedge (\forall x \in s)(x = t \vee x = r)$.
$(s = \langle t, r \rangle) := s = \{\{t, r\}, \{r\}\}$.
$\mathrm{Fun}(f) := f$ is a function.
$(\mathrm{dom}(f) = s) := f$ is a function with domain s.
$(\mathrm{rng}(f) = s) := f$ is a function with range s.
$(f(r) = t) := \mathrm{Fun}(f) \wedge \langle r, t \rangle \in f$.
$(s = \bigcup r) := (\forall x \in r)(x \subseteq s) \wedge (\forall y \in s)(\exists x \in r) y \in x$.

To save us from writing too many symbols we shall adopt the following conventions. Frequently parentheses around bounded quantifiers will be dropped. In writing $A[\vec{a}]$ we intend to convey that all free variables in the formula occur in the list of variables $\vec{a}$. Boldface versions of variables are meant to stand for tuples of variables. If $\vec{x} = (x_1, \ldots, x_r)$ we write $\forall \vec{x}$, $\forall \vec{x} \in s$, $\exists \vec{x}$, and $\exists \vec{x} \in s$ for $\forall x_1 \ldots \forall x_r$, $(\forall x_1 \in s) \ldots (\forall x_r \in s)$, $\exists x_1 \ldots \exists x_r$, and $(\exists x_1 \in s) \ldots (\exists x_r \in s)$, respectively.

We also use class notations $\{x \mid F(x)\}$ as abbreviations with the usual meaning: $s \in \{x \mid F(x)\} := F(s)$, $t = \{x \mid F(x)\} := \forall z[z \in t \leftrightarrow F(z)]$, $\{x \in t \mid F(x)\} := \{x \mid x \in t \wedge F(x)\}$, etc.

Lower case Greek variables $\alpha, \beta, \gamma, \ldots$ range over ordinals. The letters f, g, h will be reserved for functions, i.e., $\forall f \ldots$ and $\exists f \ldots$ stand for $\forall f(\mathrm{Fun}(f) \to \ldots)$ and $\exists f(\mathrm{Fun}(f) \wedge \ldots)$, respectively. We write $\alpha < \beta$ instead of $\alpha \in \beta$.

2.5 A base system for set theory

We fix a formal theory $\mathbf{BT}$ to serve as a base system for all our set theories. The language of $\mathbf{BT}$ is $\mathcal{L}^*$.

Definition 2.10. *The axioms of* $\mathbf{BT}$ *come in four groups.*

Logical Axioms

1. *Every propositional tautology is an axiom.*
2. $\forall x F(x) \rightarrow F(s)$.
3. $F(s) \rightarrow \exists x F(x)$.
4. $(\forall x \in t)F(x) \rightarrow (s \in t \rightarrow F(s))$.
5. $(s \in t \wedge F(s)) \rightarrow (\exists x \in t)F(x)$.

Ontological Axioms

(O1) $s = t \rightarrow [F(s) \leftrightarrow F(t)]$ *for* $F(a)$ *in* Δ_0.

(O2) $\mathrm{Set}(t) \rightarrow t \notin \mathrm{N}$.

(O3) $\bar{n} \in \mathrm{N}$ *for all* $n \in \mathbb{N}$.

(O4) $s \in t \rightarrow \mathrm{Set}(t)$.

(O5) $\mathrm{R}(s_1, \ldots, s_k) \rightarrow s_1 \in \mathrm{N} \wedge \ldots \wedge s_k \in \mathrm{N}$ *for* $\mathrm{R} \in \{\equiv, \mathrm{SUC}, \mathrm{ADD}, \mathrm{MULT}, \mathrm{P}\}$ *and k being the arity of* R.

(O6) $\mathrm{Ad}(s) \rightarrow \mathrm{N} \in s \wedge \mathrm{Tran}(s)$.

(O7) $\mathrm{Ad}(s) \wedge \mathrm{Ad}(t) \rightarrow s \in t \vee s = t \vee t \in s$.

(O8) $\mathrm{Ad}(s) \rightarrow \forall x \in s \forall y \in s \exists z \in s\, (x \in z \vee y \in z)$.

(O9) $\mathrm{Ad}(s) \rightarrow \forall x \in s \exists y \in s\, (y = \bigcup x)$.

(O10) $\mathrm{Ad}(s) \rightarrow \forall \vec{u} \in s \forall x \in s\, \exists y \in s\, [\mathrm{Set}(y) \wedge \forall z \in s(z \in y \leftrightarrow z \in x \wedge A[z, x, \vec{u}])]$ *for all* Δ_0 *formulae* $A[a, b, \vec{c}]$.

(O11) $\mathrm{Ad}(s) \rightarrow \forall \vec{u} \in s \forall x \in s\, (\forall y \in x\, \exists z \in s\, B[y, z, x, \vec{u}] \rightarrow \exists w \in s\, \forall y \in x\, \exists z \in w\, B[y, z, x, \vec{u}])$ *for all* Δ_0 *formulae* $B[a, b, c, \vec{d}]$.

The axioms (O8)–(O11) assert that every admissible set is a model of pairing, union, Δ_0 *separation and* Δ_0 *collection, respectively. (O7) asserts that admissible sets are linearly ordered with respect to* $\in$.

Arithmetical Axioms.

All **-translations of the equality axioms and the axioms pertaining to* SUC, ADD, *and* MULT *of Definition 2.3 (i) and (ii).*

Set Existence Axioms.

(Pairing) $\exists z\, s \in z \,\wedge\, t \in z)$.

(Union) $\exists z\,(z = \bigcup s)$.

(Δ_0 Separation) $\exists z[\mathrm{Set}(z) \wedge \forall x \in z(x \in s \wedge A(x)) \wedge \forall x \in s(A(x) \to x \in z)]$ *for* $A(a)$ *in* Δ_0.

Induction Axioms.

(Δ_0-FOUND) $\mathrm{Tran}(s) \,\wedge\, \forall x \in s(\forall y \in x\, A(y) \to A(x)) \to \forall x \in s\, A(x)$ *whenever* $A(a)$ *is* Δ_0.

(*Ind*)* $\bar{0} \in s \,\wedge\, \forall xy \in \mathrm{N}[\mathrm{SUC}(y, x) \,\wedge\, y \in s \to x \in s] \to \forall x \in \mathrm{N}\, x \in s$.

As logical rules *of* $\mathbf{BT}$ *we choose* Modus Ponens *and the following quantifier rules:*

$$A \to F(a) \vdash A \to \forall x F(x)$$

$$F(a) \to A \vdash \exists F(x) \to A$$

$$A \to (a \in s \to F(a)) \vdash A \to \forall x \in sF(x)$$

$$(F(a) \,\wedge\, a \in s) \to A \vdash \exists x \in sF(x) \to A$$

with the proviso that a *does not occur in the conclusion.*

If $\mathbf{T}$ *is a theory in the language* $\mathcal{L}^*(\mathrm{M})$ *which comprises* $\mathbf{BT}$ *then* $\mathbf{T} \vdash A$ *is meant to convey that* A *is deducible from the axioms of* $\mathbf{T}$ *via the above rules of inference.*

Remark 2.11. All non-logical axioms of $\mathbf{BT}$ are of the form $G[\vec{s}]$ where $G[\vec{a}]$ is a Σ_1 formula.

Also note that none of the axioms of $\mathbf{BT}$ asserts that any admissible sets exist.

Lemma 2.12. $\mathbf{ACA}_0 \subseteq \mathbf{BT}$.

Proof. The $\subseteq$ symbol is meant to convey that every theorem of $\mathbf{ACA}_0$ is a theorem of $\mathbf{BT}$ via the *-translation. This is proved by induction on the length of derivations in $\mathbf{ACA}_0$. The only interesting cases to inspect are the arithmetical comprehension axioms. The *-translation turns them into instances of Δ_0 separation. □

At this point, having introduced a great deal of the formal background for the paper, we can rejoice. Perhaps a few words about $\mathbf{BT}$ are in order. We assume that the reader is acquainted with the theory of admissible sets. The standard reference for admissible sets and an excellent presentation at that is [3]. The axioms (O6), (O8)–(O11) and Δ_0-FOUND ensure that, provably in $\mathbf{BT}$, every set $\mathcal{A}$ which satisfies $\mathrm{Ad}(\mathcal{A})$ is a model of the theory $\mathbf{KPU}^+$ of [3] with the set of natural numbers as urelements.

Definition 2.13. *The theory* $\mathbf{KPu}$ *comprises* $\mathbf{BT}$ *and has the following additional axioms.*

(Δ_0-Collection) $\forall x \in s\, \exists y\, A(x,y) \to \exists z \forall x \in s \exists y \in z\, A(x,y)$ *where* $A(a,b) \in \Delta_0$*;*

(IND^*) $\forall xy \in \mathrm{N}\, (\mathrm{SUC}(y,x)\, \wedge\, F(y) \to F(x)) \to \forall x \in \mathrm{N}\, F(x)$*;*

(FOUND) $\forall x(\forall y \in x F(y) \to F(x)) \to \forall x F(x)$*,*

where in the last two schemata $F(a)$ *may be any formula of* $\mathcal{L}^*(\mathrm{M})$*.*

In naming this theory $\mathbf{KPu}$ *we follow the usage of [27].*

Remark 2.14. *One cannot prove the existence of an admissible set in* $\mathbf{KPu}$*. As a result, the axioms (O6)–(O11) are immaterial as far as the proof strength of* $\mathbf{KPu}$ *is concerned. In more detail, letting* $\mathbf{KPu}^-$ *denote* $\mathbf{KPu}$ *restricted to the language* $\mathcal{L}^*$ *without the predicate symbol* Ad*, we have* $\mathbf{KPu} \vdash A \;\Rightarrow\; \mathbf{KPu}^- \vdash A^-$ *for every formula* A *of* $\mathcal{L}^*$*, where* A^- *results from* A *by replacing any occurrence* $\mathrm{Ad}(s)$ *in* A *by* $\bar{0} \neq \bar{0}$ *and any occurrence of* M *by* N*. This shows that* $\mathbf{KPu}$ *is a conservative extension of* $\mathbf{KPu}^-$*.*

As regards the predicate Ad, it plays a role in extensions of $\mathbf{BT}$ which prove $\exists x \mathrm{Ad}(x)$. Examples of such systems are the theories $\mathbf{KPl}$ and $\mathbf{KPi}$ introduced in [27]. $\mathbf{KPl}$ axiomatizes a set universe which is a limit of admissible sets while $\mathbf{KPi}$ also demands that the universe itself be an admissible set (or class).

Definition 2.15. $\mathbf{KPl}$ *is the theory* $\mathbf{BT} + (\mathrm{IND}^*) + (\mathrm{FOUND}) + (\mathrm{Lim}) + (\mathrm{M})$*, where* (Lim) *is the axiom schema* $\exists y(\mathrm{Ad}(y) \,\wedge\, s \in y)$ *and* (M) *is the axiom* $\mathrm{Ad}(\mathrm{M}) \,\wedge\, \forall x \in \mathrm{M}\, \neg\mathrm{Ad}(x)$*.*

$\mathbf{KPi}$ *is the theory* $\mathbf{KPu} + \mathbf{KPl}$*.*

In what follows, theories having only restricted versions of (FOUND) or (IND^*) as axioms will be of great importance. Such theories are interesting because of the following observations. In mathematics one mostly uses only limited amounts of induction. From proof theory we know that restricting the amount of induction tends to give rise to theories of much weaker proof-theoretic strength.

Definition 2.16. *If* $\mathbf{T}$ *is a theory whose axiom schemata comprise* (IND*) *and* (FOUND) *we denote by* $\mathbf{T}^w$ *the theory without* (FOUND) *and by* $\mathbf{T}^r$ *the theory without* (FOUND) *and* (IND*).

Using this convention, $\mathbf{KPl}^w$ *is* $\mathbf{BT}$ + (IND*) + (Lim) + (M) *and* $\mathbf{KPl}^r$ *is* $\mathbf{BT}$ + (Lim) + (M).

Remark 2.17. *The combination of* (Lim), *(O7) and* (Δ_0-FOUND) *implies* $\exists! y(\mathrm{Ad}(y) \wedge \forall z \in y \neg \mathrm{Ad}(z))$. *Therefore* $\mathbf{KPl}^r$ *is a definitional extension of* $\mathbf{KPl}^r$ *without* (M).

2.6 Some derivable consequences

We show show some basic principles that can be deduced in theories introduced so far. For future reference some will be labeled with traditional names.

Lemma 2.18. $\mathbf{KPu} \vdash F[\vec{a}] \;\Rightarrow\; \mathbf{KPl}^r \vdash \mathrm{Ad}(s) \to \forall \vec{x} \in s\, F^s[\vec{x}]$.

Proof. Proceed by induction on the length of the derivation in $\mathbf{KPu}$. □

The main use we shall make of the foregoing lemma is that every statement that can be proved in $\mathbf{KPu}$ about the universe of sets can be transferred to $\mathbf{KPl}^r$ by relativizing it to any admissible set.

Proofs for the following four results can be found in [3], I.4.2-4.4.5.

Lemma 2.19. (Σ Persistence) *For every* Σ *formula* A *we have:*

(i) $\mathbf{BT} \vdash A^s \wedge s \subseteq t \to A^t$*;*

(ii) $\mathbf{BT} \vdash A^S \to A$.

Lemma 2.20. (Σ Reflection) *For* $A \in \Sigma$ *we have* $\mathbf{KPu}^r \vdash A \to \exists x\, A^x$.

Lemma 2.21. (Σ Collection) *For every* Σ *formula* $F(a,b)$,

$$\mathbf{KPu}^r \vdash \forall x \in s\, \exists y\, F(x,y) \to \exists z[\forall x \in s\, \exists y \in z\, F(x,y) \wedge \forall y \in z\, \exists x \in s\, F(x,y)].$$

Lemma 2.22. (Δ Separation) *If* $A(a)$ *is in* Σ *and* $B(a)$ *is in* Π *then*

$$\mathbf{KPu}^r \vdash \forall x\, [A(x) \leftrightarrow B(x)] \to \exists z(\mathrm{Set}(z) \wedge \forall x[x \in z \leftrightarrow A(x)]).$$

Lemma 2.23. (Σ Replacement) *For every* Σ *formula* $F(a,b)$,

$$\mathbf{KPu}^r \vdash \forall x \in s\, \exists! y\, F(x,y) \to \exists f[\mathrm{Fun}(f) \wedge \mathrm{dom}(f) = s \wedge \forall x \in s\, F(x, f(x))].$$

An powerful tool in set theory is definition by transfinite recursion. The most important applications are definitions of functions by Σ recursion. The axioms of $\mathbf{KPu}$ are sufficient for this task. A closer inspection of the well known proof (see [3], I.6.1)) reveals that a restricted form of foundation, dubbed (Σ-FOUND), suffices.

(Σ-FOUND) is the schema

$$\forall x[\forall y \in x\, F(y) \;\rightarrow\; F(x)] \;\rightarrow\; \forall x\, F(x)$$

for $F(a) \in \Sigma$.

Lemma 2.24. (Σ Recursion) $\mathbf{KPu}^r$ + (Σ-FOUND) *proves all instances of* Σ *recursion,* (Σ-REC),

$$\forall\alpha\forall x\exists! y\, G(\alpha, x, y) \;\rightarrow\; \forall\alpha\exists f[\mathrm{Fun}(f) \wedge \mathrm{dom}(f) = \alpha \wedge (\forall\beta < \alpha)G(\beta, f\restriction\beta, f(\beta))]$$

where $G(a, b, c) \in \Sigma$ *and* $f\restriction\beta = \{\langle\delta, z\rangle \mid \delta < \beta \;\wedge\; \langle\delta, z\rangle \in f\}$.

Remark 2.25. *The formalization of systems of set theory with a predicate earmarked for the class of admissible sets was introduced in [27] for proof-theoretic purposes. Singling out* $\mathbf{KPl}$ *as a system worthy of attention owes much to the observation (see [3], V.6.12) that* L_α *is a model of axiom Beta (see [3, I.9.4]) if* α *is a limit of admissible ordinals.*

3 Theories of iterated inductive definitions

In this section we introduce second theories of iterated inductive definitions formalized in the language of second order arithmetic. Till now such theories were formulated as first order theories with quantifiers just ranging over the natural numbers but with the aid of predicate symbols to represent inductively defined sets (see [10]). In this restricted language one can only talk about well orderings defined by arithmetical formulae. Moving to $\mathcal{L}_2$ enables one to reformulate these theories via set existence axioms and to talk about iterated inductive definitions where the iteration is carried out along arbitrary well orderings.

Instead of pursuing a proof-theoretic analysis of such theories directly via specific infinitary proof systems taylored to accommodate iterated inductive definitions (as e.g. in [10]), we analyze them by first embedding them into germane set theories and subsequently use the general machinery for ordinal analysis of set theories. In this way we obtain uniform and simultaneous ordinal analyses of almost all theories of inductive definitions. An example which illustrates the uniformity of the method is the theory $\mathbf{ID}_{\prec^*}$ introduced in [15]. An analysis of $\mathbf{ID}_{\prec^*}$ was carried out in [5] by means of the Ω_ν-rules. An sketch of an ordinal analysis of this theory by Pohlers' so-called method of local predicativity was adumbrated in [42]. But a full analysis using this technique didn't materialize before [71] (169 pages), and turned out to be quite a formidable task. Therefore I consider it in order to add a further proof, in particular as it is more or less a by-product of the investigation of yet stronger theories of iterated inductive definitions.

As per usual, to begin with we need to set up some terminological conventions.

Definition 3.1. *Let $Q(a,b)$ be a formula of $\mathcal{L}_2$ (which may contain additional parameters, i.e. free variables). When we view Q as a binary relation we shall write sQt for $Q(s,t)$. Let $F(a)$ be an arbitrary formula of $\mathcal{L}_2$. We will use the following abbreviations:*

$$\begin{aligned}
\mathrm{Fld}(s) &:= \exists x(sQx \vee xQs) \quad \textit{(s is in the field of Q)};\\
\mathrm{LO}(Q) &:= \forall x \neg(xQx) \wedge \forall xy[\mathrm{Fld}(x) \wedge \mathrm{Fld}(y) \to xQy \vee x \equiv y \vee yQx]\\
&\qquad \wedge \forall xyz[xQy \wedge yQz \to xQz] \quad \textit{(Q is a linear order)};\\
\mathrm{PROG}(Q,F) &:= \forall x[\forall y(yQx \to F(y)] \to F(x)] \quad \textit{(F is Q progressive)};\\
\mathrm{TI}(Q,F) &:= \mathrm{PROG}(Q,F) \to \forall x\, F(x) \quad \textit{(Q induction for F)};\\
\mathrm{WF}(Q) &:= \forall X\, \mathrm{TI}(Q,X) \quad \textit{(Q is well founded)};\\
\mathrm{WO}(Q) &:= \mathrm{LO}(Q) \wedge \mathrm{WF}(Q) \quad \textit{(Q is a well-ordering)}.
\end{aligned}$$

We shall use the primitive recursive pairing function $\langle m,n\rangle := \frac{1}{2}(m+n)(m+n+1)+m$. If U is a set we denote by U_s the set $\{x \mid \langle s,x\rangle \in U\}$.

We shall use the notation $F(\mathrm{P}^+)$ to convey that $F(\mathrm{P}) \in \Pi^1_0(\mathrm{P}^+)$ (see Definition 2.7). Such formulae are said to be P-*positive arithmetical formulae*. If $H(a)$ is any formula then $F(H)$ denotes the formula obtained by replacing all occurrences of the form $\mathrm{P}(t)$ and $\mathrm{P}(x)$ by $H(t)$ and $H(x)$, respectively, with the usual proviso that we may have to rename some bound variables to avoid any unintended capture of variables.

Definition 3.2 ($\mathbf{ID}_{\prec^*}$)**.** *Let $\prec$ be a linear ordering on $\mathbb{N}$, definable via an arithmetical formula $\mathrm{Q}[a,b]$ such that $\mathbf{ACA}_0 \vdash \mathrm{LO}(\prec)$. The vocabulary of $\mathbf{ID}_{\prec^*}$, $\mathcal{L}(\mathbf{ID}_{\prec^*})$, comprises that of $\mathcal{L}_1$ but in addition has a unary predicate symbol Q^* to denote the accessible part of $\prec$ and moreover for every P-positive arithmetical formula $F[\mathrm{P}^+,U,a,b]$ it has a two-place predicate symbol I^F.*

The axioms of $\mathbf{ID}_{\prec^*}$ comprise those of Definition 2.3(i),(ii) and the induction schema (IND) *for all formulae of $\mathcal{L}(\mathbf{ID}_{\prec^*})$. Further axioms of $\mathbf{ID}_{\prec^*}$ are the following:*

$$(\mathrm{Q}^*1)\ \ \mathrm{PROG}(\prec,\mathrm{Q}^*)$$

$$
\begin{array}{ll}
(\mathrm{Q}^*2) & \mathrm{PROG}(\prec, H) \to \forall x\,[\mathrm{Q}^*(x) \to H(x)] \\
(\mathrm{I}^{\mathrm{Q}^*}1) & \mathrm{Q}^*(t) \to \forall x\,(\,F[\mathrm{I}^F_t, \mathrm{I}^F_{\prec t}, t, x] \to x \in \mathrm{I}^F_t] \\
(\mathrm{I}^{\mathrm{Q}^*}2) & \mathrm{Q}^*(t) \,\wedge\, \forall x\,(F[H, \mathrm{I}^F_{\prec t}, t, x] \to H(x)) \to \forall x(x \in \mathrm{I}^F_t \to H(x))
\end{array}
$$

for every arithmetical formula $F[\mathrm{P}^+, U, a, b]$ and every $\mathcal{L}(\mathbf{ID}_{\prec^*})$ formula $H(a)$, where we used the notations $s \in \mathrm{I}^F_t := \mathrm{I}^F(t, s)$ and $\mathrm{I}^F_{\prec t} := \{z \mid \exists y(y \prec t \,\wedge\, z \in \mathrm{I}^F_y\}$.

All arithmetical well-orderings have an order-type less than the first recursively inaccessible ordinal ω_1^{CK}. Since the accessible part of a primitive recursive ordering can have order-type ω_1^{CK} (see [21]) it seems that $\mathbf{ID}_{\prec^*}$ may be able to axiomatize iterated inductive definitions along non-arithmetical well orderings in contrast to what we said at the beginning of this section about previous investigations of such theories in proof theory. This point will be clarified later once we have embedded $\mathbf{ID}_{\prec^*}$ into a second order system.

Definition 3.3. *For formulae $F(\mathrm{P}, U, a, b)$ and $Q(a, b)$ we use the abbreviations*

$$
\begin{aligned}
\mathrm{Cl}^F(V, U, s) &:= \forall x[F(V, U, s, x) \to x \in V] \\
\mathrm{IT}^F(Q, U) &:= \forall i\,[\mathrm{Cl}^F(U_i, U_{Qi}, i) \,\wedge\, \forall Y(\mathrm{Cl}^F(Y, U_{Qi}, i) \to U_i \subseteq Y)],
\end{aligned}
$$

where $U_i := \{x \mid \langle i, x\rangle \in U\}$ and $U_{Qi} := \{x \mid \exists j(jQi \,\wedge\, \langle j, x\rangle \in U\}$.

Definition 3.4. *(i) Let $\mathbf{ID}^*$ be the theory $\mathbf{ACA}$ augmented by the schema*

$$(\mathrm{IT}^*1) \qquad \forall x[\mathrm{WO}(\prec_x) \to \exists Z\,\mathrm{IT}^F(\prec_x, Z)]$$

for every formula $F[P^+, U, a, b]$ (having no further parameters) and every family of relations $(\prec_n)_{n\in\mathbb{N}}$, which is definable by some arithmetical formula $Q[a, b, c]$ via $s \prec_r t := Q[s, t, r]$.

(ii) $\mathbf{ID}^*_2$ arises from $\mathbf{ID}^*$ by adding the schema

$$(\mathrm{IT}^*2) \qquad \forall x\forall i\forall Z[\mathrm{WO}(\prec_x) \,\wedge\, \mathrm{IT}^F(\prec_x, Z) \,\wedge\, \mathrm{Cl}^F(H, Z_{\prec_x i}, i) \to Z_i \subseteq H]$$

with the same conditions on F as above and every formula $H(a)$.
$\mathbf{ID}^*_2$ makes greater demands than $\mathbf{ID}^*$ in that if $\mathrm{WO}(\prec_s)$ holds and U satisfies $\mathrm{IT}^F(\prec_s, U)$, then every section U_i of U will also be contained in all $\mathcal{L}_2$-definable classes closed under the operator $X \mapsto \{z \mid F[X, U_{\prec_s i}, z]\}$.

It is obvious that $\mathbf{ID}^*_2$ merely axiomatizes iterations of length $< \omega_1^{CK}$. However, the strength of $\mathbf{ID}^*_2$ is owed to the fact that the arithmetical well-orderings may depend on numerical parameters which can be quantified over. E.g. if $\prec_e$ is defined *by* $n \prec_e m := \exists y\, \mathrm{T}_2(e,n,m,y)$ *where* T_2 *stands for the well-known primitive recursive predicate from Kleene's normal form theorem, then via the statement* $\forall x[\mathrm{WO}(\prec_x) \to \exists Z\, \mathrm{IT}^F(\prec_x, Z_x)]$ *we quantify over all iterations below* ω_1^{CK}.

On the other hand, in general it is not possible to deduce $\exists Z \forall x[\mathrm{WO}(\prec_x) \to \mathrm{IT}^F(\prec_x, Z_x)]$ *within* $\mathbf{ID}_2^*$ *which would amount to an iteration of length* ω_1^{CK}.

If one lifts the restriction to arithmetical well-orderings and allows arbitrary parameters in the operator forms in the schemata (IT^*1) and (IT^*2) one arrives at autonomous theories of arithmetical inductive definitions which provide the natural limit of all such theories. Furthermore, we also consider the wider class of monotone inductive definitions.

Definition 3.5. *(i)* $\mathbf{AUT}\text{-}\mathbf{ID}^{pos}$ *is the theory* $\mathbf{ACA}$ *augmented by the schema*

$$(\mathrm{IT}^{pos}1) \qquad \forall X[\mathrm{WO}(X) \to \exists Z\, \mathrm{IT}^F(X,Z)]$$

for every arithmetical formula $F(P^+, U, a, b)$.

(ii) $\mathbf{AUT}\text{-}\mathbf{ID}^{mon}$ *is the theory* $\mathbf{ACA}$ *augmented by the schema*

$$(\mathrm{IT}^{mon}1) \qquad \mathrm{MON}(F) \to \forall X[\mathrm{WO}(X) \to \exists Z\, \mathrm{IT}^F(X,Z)]$$

for every arithmetical formula $F(P, U, a, b)$*, where*

$$\mathrm{MON}(F) := \forall i \forall x \forall X \forall Y \forall Z[X \subseteq Y \;\wedge\; F(X,Z,i,x) \to F(Y,Z,i,x)].$$

(iii) By $\mathbf{AUT}\text{-}\mathbf{ID}_2^{pos}$ *we denote the theory* $\mathbf{AUT}\text{-}\mathbf{ID}^{pos}$ *plus the scheme*

$$(\mathrm{IT}^{pos}2) \qquad \mathrm{WO}(R) \wedge \mathrm{IT}^F(R,U) \;\wedge\; \mathrm{Cl}^F(H, U_{Rs}, s) \to U_s \subseteq H$$

for every arithmetical formula $F(P^+, U, a, b)$ *and arbitrary* $\mathcal{L}_2$*-formula* $H(b)$*. In the same vein one defines* $(\mathrm{IT}^{mon}2)$ *and the theory* $\mathbf{AUT}\text{-}\mathbf{ID}_2^{mon}$.

(iv) If $\mathbf{T}$ *is an* $\mathcal{L}_2$ *theory defined by adding axioms to* $\mathbf{ACA}$*, then* $\mathbf{T}_0$ *denotes the theory which is obtained by adding the same axioms to* $\mathbf{ACA}_0$.

Remark 3.6. *For* P*-positive formulae* $F(\mathrm{P}^+, U, a, b)$*,* $\mathrm{MON}(F)$ *is provable in pure logic. Thus axioms for monotone inductive definitions imply the corresponding axioms for positive ones.*

In Definition 3.1 we defined well-foundedness of a relation in a somewhat unusual way. One can prove in $\mathbf{ACA}_0$ that our definition is equivalent to the usual one.

Lemma 3.7. $\mathbf{ACA}_0 \vdash \mathrm{WF}(R) \leftrightarrow \forall Z[Z \neq \varnothing \rightarrow \exists x \in Z\, \forall y(yRx \rightarrow y \notin Z)]$.

Proof. Exercise or see [17], 6.1.5. □

Lemma 3.8. $\mathbf{ACA}_0 \vdash \mathrm{WO}(R) \wedge \mathrm{IT}^F(R,U) \wedge \mathrm{IT}^F(R,V) \rightarrow \forall i\,(U_i = V_i)$.

Proof. Assume $\mathrm{WO}(R)$, $\mathrm{IT}^F(R,U)$, $\mathrm{IT}^F(R,V)$ but $U_i \neq V_i$ for some i. By Lemma 3.7 we can pick an R-minimal i_0 with $U_{i_0} \neq V_{i_0}$. By minimality, $U_{Ri_0} = V_{Ri_0}$, and thus $\mathrm{Cl}(U_{i_0}, V_{Ri_0}, i_0)$ as well as $\mathrm{Cl}(V_{i_0}, U_{Ri_0}, i_0)$. As a result, $V_{i_0} \subseteq U_{i_0}$ and $U_{i_0} \subseteq V_{i_0}$, yielding the contradiction $U_{i_0} = V_{i_0}$. □

Definition 3.9. *We define an interpretation* ${}^{\curlywedge} : \mathbf{ID}_{\prec^*} \longrightarrow \mathbf{ID}_2^*$*, where* $\prec$ *is given by a formula* $Q[a,b]$*. Let* $s \prec_r t := s \prec t \wedge (t \equiv r \vee t \prec r)$*,*

$$\mathrm{Acc}(\prec, U) := \mathrm{PROG}(\prec, U) \wedge \forall Z[\mathrm{PROG}(\prec, Z) \rightarrow U \subseteq Z],$$
$$\mathrm{P}^F(r,s) := \exists Z[\mathrm{IT}^F(\prec_r, Z) \wedge s \in Z_r]$$

for every arithmetical formula $F[\mathrm{P}^+, U, a, b]$*.*

If A *is a formula of* $\mathbf{ID}_{\prec^*}$ *then* $A^{\curlywedge}$ *arises from* A *by replacing all subformulas* $\mathrm{Q}^*(s)$ *and* $\mathrm{I}^F(r,s)$ *in* A *by* $\exists X[s \in X \wedge \mathrm{Acc}(\prec, X)]$ *and* $\mathrm{P}^F(r,s)$*, respectively.*

Theorem 3.10. *If* $\mathbf{ID}_{\prec^*} \vdash A$ *then* $\mathbf{ID}_2^* \vdash A^{\curlywedge}$*.*

Proof. It suffices to prove the translations of axioms arising from the schemata (Q^*1), (Q^*2), $(\mathrm{I}^{\mathrm{Q}^*}1)$ and $(\mathrm{I}^{\mathrm{Q}^*}2)$ in $\mathbf{ID}_2^*$. We reason in the target theory $\mathbf{ID}_2^*$. Let

$$G[\mathrm{P}^+, U, a, b] \;:=\; \forall y[y \prec b \rightarrow \mathrm{P}(y)].$$

There exists a set V such $\mathrm{IT}^G(\varnothing, V)$. For $S := V_{\bar{0}}$ we have $\mathrm{Acc}(\prec, S)$. As in Lemma 3.8 one shows that thereby S is uniquely determined, i.e. $\forall X[\mathrm{Acc}(\prec, X) \rightarrow X = S]$. Therefore we get $\forall x[(\mathrm{Q}^*(x))^{\curlywedge} \leftrightarrow x \in S)$ and thus $(\mathrm{PROG}(\prec, \mathrm{Q}^*))^{\curlywedge}$. As a result, provability of the translation of (Q^*2) follows with the help of (IT^*2).

To prove the translations of $(\mathrm{I}^{\mathrm{Q}^*}1)$ and $(\mathrm{I}^{\mathrm{Q}^*}2)$ suppose that $(\mathrm{Q}^*(r))^{\curlywedge}$ holds, i.e. $r \in S$. If $\neg\mathrm{WO}(\prec_r)$ then by Lemma 3.7 there exists a set U such that $r \in U$ and $\forall x \in U\, \exists y \in U\, (y \prec_r x)$. Letting $U^c := \{i \mid i \notin U\}$ we have $\mathrm{PROG}(\prec, U^c)$ and thus $S \subseteq U^c$ by choice of S. But this is incompatible with $r \in U$. Hence $\prec_r$ must be a well-ordering.

Now let $F[\mathrm{P}^+, U, a, b]$ be an arithmetical P-positive formula. From (IT^*1) we obtain the existence of a set V satisfying $\mathrm{IT}^F(\prec_r, V)$. By Lemma 3.8 we conclude that

$$\forall i[i \prec r \vee i \equiv r \rightarrow \forall x((\mathrm{I}^F(i,x))^{\curlywedge} \leftrightarrow x \in V_i)],$$

and hence $(\mathrm{I}^F_{\prec r})^{\curlywedge} = V_{\prec r}$ and $(\mathrm{I}^F_r)^{\curlywedge} = V_r$. From the foregoing and (IT^*2) we obtain the derivability of the ${}^{\curlywedge}$-translations of $(\mathrm{I}^{\mathrm{Q}^*}1)$ and $(\mathrm{I}^{\mathrm{Q}^*}2)$. □

Definition 3.11. Bar induction*, abbreviated* (BI)*, is the schema*

$$\forall X\,[\mathrm{WO}(X) \to \mathrm{TI}(X, H)]$$

for every $\mathcal{L}_2$-formula $H(a)$.

Lemma 3.12. (BI) *is a consequence of* $\mathbf{AUT\text{-}ID}_2^{pos}$.

Proof. Assume $\mathrm{WO}(R)$ and $\mathrm{PROG}(R, H)$. We aim at showing $\forall x\, H(x)$. Let $F(\mathrm{P}^+, U, a, b) := \forall z[zRb \to \mathrm{P}(z)]$. Owing to $(\mathrm{IT}^{pos}1)$ there exists a set V such that $\mathrm{IT}^F(\varnothing, V)$. Letting $S := V_{\bar{0}}$ and employing $(\mathrm{IT}^{pos}2)$ we obtain

$$\mathrm{PROG}(R, S) \quad \text{and} \quad \mathrm{PROG}(R, H) \to S \subseteq H.$$

Hence $\forall x\,(x \in S) \wedge S \subseteq H$, thus $\forall x\, H(x)$. □

Theorem 3.13. *(i)* $\mathbf{ID}_2^* \subseteq \mathbf{ID}^* + (\mathrm{BI})$.

(ii) $\mathbf{AUT\text{-}ID}_2^{pos} = \mathbf{AUT\text{-}ID}^{pos} + (\mathrm{BI})$.

(iii) $\mathbf{AUT\text{-}ID}_2^{mon} = \mathbf{AUT\text{-}ID}^{mon} + (\mathrm{BI})$.

Proof. (ii) In view of Lemma 3.12 it suffices to show that the instances of $(\mathrm{IT}^{pos}2)$ can be derived in $\mathbf{AUT\text{-}ID}^{pos}$ with the help of (BI). Assume $\mathrm{WO}(R)$, $\mathrm{IT}^F(R, V)$ and $\mathrm{Cl}^F(H, V_{Rs}, s)$. Letting $G(U) := \mathrm{Cl}^F(U, V_{Rs}, s) \to V_s \subseteq U$, we have

$$\forall Z\, G(Z). \tag{3.1}$$

Moreover,

$$\mathbf{ACA}_0 + (\mathrm{BI}) \vdash \forall Z\, A(Z) \to A(H) \tag{3.2}$$

holds for every arithmetical formula $A(U)$ and arbitrary $\mathcal{L}_2$-formula $H(a)$ (see [15], Lemma 1.6.3). As $G(U)$ is arithmetical we get $V_s \subseteq H$ from (3.1) and (3.2). This shows $(\mathrm{IT}^{pos}2)$.

(i) and (iii) are proved similarly, crucially using (3.2) and also Lemma 3.12 for the "$\supseteq$" entailments. □

Remark 3.14. *Up to the year 1981, the monograph [10] gives a comprehensive account of the proof theory of iterated inductive definitions. The preface written by Feferman provides a detailed history of the subject.*

4 Theories of iterated Π^1_1-comprehension

It is a classical result (see [34, 66]) that every Π^1_1-set of the structure $\mathfrak{N} = (\mathbb{N}, 0, +, \cdot)$ can be obtained as a section of a fixed point of a positive arithmetical inductive definition. As an extension of this result there is a close connection between iterated inductive definitions and iterated Π^1_1-comprehensions. The next definition provides a precise definition of the latter sort of theory.

Definition 4.1. *For every* Π^1_1*-formula* $B(U, a, b)$ *let*

$$\mathrm{HJ}^B(R, U) := \forall x \forall i [x \in U_i \leftrightarrow B(U_{Ri}, i, x)],$$

where $U_{Ri} = \{y \mid \exists j\, (jRi \wedge y \in U_i)\}$.

(i) $(\Pi^1_1\text{-TR})$ *is the schema*

$$\forall X\, [\mathrm{WO}(X) \rightarrow \exists Y\, \mathrm{HJ}^B(X, Y)]$$

where $B(U, a, b)$ *is* Π^1_1.

(ii) The theory of Π^1_1 transfinite recursion, Π^1_1-$\mathbf{TR}$, *is* $\mathbf{ACA}$ *augmented by the schema* $(\Pi^1_1\text{-TR})$.

Lemma 4.2 ($\mathbf{ACA}_0$). *If* $\mathrm{WO}(R)$*,* $\mathrm{HJ}^B(R, U)$ *and* $\mathrm{HJ}^B(R, V)$*, then* $U_i = V_i$ *holds for all* i.

Proof. Analogous to Lemma 3.8. □

Lemma 4.3. $\mathbf{AUT}$-$\mathbf{ID}_0^{mon} \subseteq \Pi^1_1$-$\mathbf{TR}_0$.

Proof. We have to show that $(\mathrm{IT}^{mon}1)$ is deducible in Π^1_1-$\mathbf{TR}_0$. Let $F(\mathrm{P}, U, a, b)$ be arithmetical and set $B(U, a, b) := \forall Z[\mathrm{Cl}^F(Z, U, a) \rightarrow b \in Z]$. Now assume $\mathrm{WO}(R)$ and $\mathrm{MON}(F)$. Invoking $(\Pi^1_1\text{-TR})$, there exists a set S such that $\mathrm{HJ}^B(R, S)$. For all i we then have

$$S_i = \{x \mid \forall Z[\mathrm{Cl}^F(Z, S_{Ri}, i) \rightarrow x \in Z]\}$$

and hence

$$\forall Z[\mathrm{Cl}^F(Z, S_{Ri}, i) \rightarrow S_i \subseteq Z].$$

Thus if $F(S_i, S_{Ri}, i, s)$ and $\mathrm{Cl}^F(U, S_{Ri}, i)$ hold, then from (4.1) we obtain $S_i \subseteq U$ and, since $\mathrm{MON}(F)$, we have $F(U, S_{Ri}, i, s)$ and consequently $s \in U$. So in view of (4.1) the preceding argument shows that $F(S_i, S_{Ri}, i, s) \rightarrow s \in S_i$ holds. Thence

$$\mathrm{Cl}(S_i, S_{Ri}, i).$$

(4.1) and (4.1) yield $\mathrm{IT}^F(R, S)$. □

Corollary 4.4. *For every Π^1_1-formula $F(\mathrm{P}, U, a, b)$,*

$$\Pi^1_1\text{-}\mathbf{TR}_0 \vdash \mathrm{MON}(F) \to \forall X[\mathrm{WO}(X) \to \exists Z\, \mathrm{IT}^F(X, Z)].$$

Proof. Since (Σ^1_1-AC) is deducible in Π^1_1-$\mathbf{CA}_0$ (cf. 5.11(i)), the formula $\mathrm{Cl}^F(V, U, a, b)$ is provably equivalent to a Σ^1_1-formula in Π^1_1-$\mathbf{TR}_0$, thus $B(U, a, b)$ is equivalent to a Π^1_1-formula. □

To prove the inclusion Π^1_1-$\mathbf{TR}_0 \subseteq \mathbf{AUT}$-$\mathbf{ID}^{pos}_0$, we shall need the inductive characterization of Π^1_1 classes mentioned at the beginning of this section. Unfortunately, the usual proofs in the literature (cf. [3], VI.1.11 and [25], III.3.2) cannot be directly carried out in $\mathbf{AUT}$-$\mathbf{ID}^{pos}_0$ since they utilize ordinals or use Π^1_1-comprehension. Albeit Π^1_1-comprehension is a consequence of $\mathbf{AUT}$-$\mathbf{ID}^{pos}_0$, we cannot use it at this point since this is part of what we want to prove. $\mathbf{ACA}_0$ suffices for the desired characterization of Π^1_1 classes.

Lemma 4.5. *For every Π^1_1-formula $B(U, a, b)$ one can construct an arithmetical formula*
$F(\mathrm{P}^+, U, a, b)$ such that

$$\mathbf{ACA}_0 \vdash \forall X \forall i \forall z\, [B(Y, i, z) \leftrightarrow \forall Z\, [\mathrm{Cl}^F(Z, Y, i) \to \langle z, \bar{1}\rangle \in Z]].$$

Proof. By the Π^1_1 normal form theorem (cf. [25, IV.1.]) one finds an arithmetical formula $Q(U, a, b, c, d)$ such that with $c \prec^b d := Q(U, a, b, c, d)$ one has

$$\mathbf{ACA}_0 \vdash \forall z\, [B(U, a, z) \leftrightarrow \mathrm{WO}(\prec^z)];$$
$$\mathbf{ACA}_0 \vdash \forall z \forall x\, [\mathrm{Fld}(x, \prec^z) \to (x \prec^z \bar{1} \lor x \equiv \bar{1})].$$

((4.1) follows from the fact that $\prec^s$ is a relation on codes of sequences of natural numbers and $\bar{1}$ encodes the empty sequence.) Immediately from (4.1) we have

$$\forall x\, [x \prec^s \bar{1} \to x \in V) \to \mathrm{TI}(\prec^s, V).$$

Define

$$F(Z^+, U, a, \langle b, c\rangle) := \forall x\, (x \prec^b c \to y \in V),$$
$$A(V, b, c) := \mathrm{TI}(\prec^b, V) \lor \forall y (y \prec^b c \to y \in V),$$
$$C(b, c) := \forall Z\, [\mathrm{Cl}^F(Z, U, a) \to \langle b, c\rangle \in Z).$$

We aim at showing

$$\forall Y\, A(Y, s, t) \leftrightarrow C(s, t).$$

"⇒": $\mathrm{Cl}^F(R, U, a)$ implies $\mathrm{PROG}(\prec^s, R_s)$, thus $\forall y(y \in R_s) \vee (t \in R_s)$ since $A(R_s, s, t)$ holds. Thus $\langle s, t\rangle \in R$.

"⇐": Given a set V, let $V^* := \{\langle z, u\rangle \mid A(V, z, u)\}$. Suppose that $F(V^*, U, a, \langle b, c\rangle)$. Then $\forall x[x \prec^b c \rightarrow A(V, b, x)]$, thus $\mathrm{TI}(\prec^b, V) \vee \forall xy\,[x \prec^b c \wedge y \prec^b x \rightarrow y \in V]$. Consequently, we have $\mathrm{TI}(\prec^b, V) \vee \neg\mathrm{PROG}(\prec^b, V) \vee \forall x\,[x \prec^b c \rightarrow x \in V]$, hence $A(V, b, c)$, thus $\langle b, c\rangle \in V^*$. As a result we have shown $\mathrm{Cl}^F(V^*, U, a)$. Thence the assumption $C(s, t)$ yields $\langle s, t\rangle \in V^*$, so that $A(V, s, t)$ holds.

We have thus shown (4.1). From (4.1) we can conclude

$$A(V, s, \bar{1}) \leftrightarrow \mathrm{TI}(\prec^s, V).$$

Combining (4.1), (4.1), and (4.1) we arrive at $\forall z\,[B(U, a, z) \leftrightarrow C(z, \bar{1})]$, as desired. □

Theorem 4.6. *(i)* Π^1_1-$\mathbf{TR}_0 = \mathbf{AUT\text{-}ID}_0^{pos} = \mathbf{AUT\text{-}ID}_0^{mon}$.

(ii) Π^1_1-$\mathbf{TR} = \mathbf{AUT\text{-}ID}^{pos} = \mathbf{AUT\text{-}ID}^{mon}$.

(iii) Π^1_1-$\mathbf{TR} + (\mathrm{BI}) = \mathbf{AUT\text{-}ID}_2^{pos} = \mathbf{AUT\text{-}ID}_2^{mon}$.

Proof. (i) implies (ii) and by Theorem 3.13 also (iii). For (i), in view of Lemma 4.3, it suffices to show Π^1_1-$\mathbf{TR}_0 \subseteq \mathbf{AUT\text{-}ID}_0^{pos}$. Let $B(U, a, b)$ be Π^1_1 and choose $F(\mathrm{P}^+, U, a, b)$ as in Lemma 4.5. Let

$$G(\mathrm{P}^+, U, a, b) := F(\mathrm{P}, \{x \mid \langle x, \bar{1}\rangle \in U\}, a, b).$$

On account of $(\mathrm{IT}^{pos}1)$, for every well ordering R there exists a set V such that

$$\mathrm{IT}^G(R, V).$$

(4.1) implies

$$\forall Z\,[\mathrm{Cl}^F(Z, \{x \mid \langle x, \bar{1}\rangle \in V_{Ra}\}, a) \rightarrow V_a \subseteq Z]$$

and $\mathrm{Cl}^F(V_a, \{x \mid \langle x, \bar{1}\rangle \in V_{Ra}\}, a)$, which by choice of $F(\mathrm{P}, U, a, b)$ entails

$$\forall z\,[B(\{x \mid \langle x, \bar{1}\rangle \in V_{Ra}\}, a, z) \rightarrow \langle z, \bar{1}\rangle \in V_a].$$

With $V^* := \{\langle i, z\rangle \mid \langle i, \langle z, \bar{1}\rangle\rangle \in V\}$, (4.1) implies $\forall iz\,[B(V^*_{Ri}, i, z) \leftrightarrow z \in V^*_i]$, and hence $\mathrm{HJ}^B(R, V^*)$. □

Corollary 4.7. *Owing to Corollary 4.4, the identities of Theorem 4.6 also hold for iterated Π^1_1 inductive definitions instead of iterated arithmetical inductive definitions.*

Remark 4.8. *(i) The main purpose of this section was preparatory work for the interpretation of* Π^1_1-$\mathbf{TR}_0$ *into systems of set theory. The equivalences of Theorem 4.6 lend itself to rather transparent interpretations into theories of iterated admissibility.*

(ii) Historically, reductions of subsystems of second order arithmetic played an important role in proof theory (cf. [10, 15, 19]). Theorem 4.6 can be viewed as a tribute to those times.

(iii) The idea of proof of Lemma 4.3 using the characterization of a fixed point I_Γ *of an operator* Γ *by means of* $I_\Gamma = \bigcap\{X \mid \Gamma(X) \subseteq X\}$ *is of course the standard one.*

5 Set theories of iterated admissibility

Theories of iterated inductive definitions have a canonical interpretation in set theories of iterated admissibility. The structure theory of Σ_+-inductive definitions on admissible sets (cf. [3, VI.2]) will be useful here.

Definition 5.1. *(i) Let* $\mathfrak{D}[\alpha, f]$ *denote the conjunction of the formulae* $\mathrm{Ord}(\alpha)$, $\mathrm{Fun}(f)$, $\mathrm{dom}(f) = \alpha$, *and*

$$(\forall\beta < \alpha)[\mathrm{Ad}(f(\beta)) \wedge (\forall\eta < \beta)\,(f(\eta) \in f(\beta)) \wedge (\forall x \in f(\beta))[\mathrm{Ad}(x) \to (\exists\eta < \beta)\,(x = f($$

(ii) $(\text{AUT-Ad}) := \forall\alpha\exists f\,\mathfrak{D}[\alpha, f]$.

(iii) $\mathbf{AUT}$-$\mathbf{KPl} := \mathbf{KPl} + (\text{AUT-Ad})$.

(iv) $(\mathrm{Ad}^*) := (\forall\alpha \in \mathrm{M})\,\exists f\,\mathfrak{D}[\alpha, f]$.

(v) $\mathbf{KPl}^* := \mathbf{KPl} + (\mathrm{Ad}^*)$.

$\mathbf{AUT}$-$\mathbf{KPl}$ axiomatizes a set universe that has as many admissible sets as ordinals and in which the admissible sets are linearly ordered, whereas $\mathbf{KPl}^*$ only asserts that there are at least as many admissible sets as there are ordinals below ω_1^{CK}, i.e. ordinals in the least admissible set above the urelement structure of the natural numbers.

Lemma 5.2. *(i)* $\mathbf{KPl}^r \vdash \mathfrak{D}[\alpha, f] \wedge \mathfrak{D}[\alpha, g] \to f = g$.

(ii) $\mathbf{KPl}^* \vdash (\forall\alpha \in \mathrm{M})\,\exists! f\,\mathfrak{D}[\alpha, f]$.

Proof. (i): Suppose that $\mathfrak{D}[\alpha, f]$ and $\mathfrak{D}[\alpha, g]$. We show $\forall\beta < \alpha\,(f(\beta) = g(\beta)$ by induction on β, a principle justified by (Δ_0-FOUND). Let $\beta < \alpha$ and assume by induction hypothesis that $f \restriction \beta = g \restriction \beta$ (where $h \restriction \beta$ is the restriction of the function h to the domain β). Since $\mathrm{Ad}(f(\beta))$ and $\mathrm{Ad}(g(\beta))$ hold it follows from axiom (O7) that $f(\beta) \in g(\beta) \,\vee\, f(\beta) = g(\beta) \,\vee\, g(\beta) \in f(\beta)$. As $\mathrm{Ad}(f(\beta))$ and $\mathfrak{D}[\alpha, g]$ hold, $f(\beta) \in g(\beta)$ would entail the existence of an ordinal $\eta < \beta$ such that $f(\beta) = g(\eta) = f(\eta)$, contradicting $\mathfrak{D}[\alpha, f]$. Likewise one can rule out that $g(\beta) \in f(\beta)$. Hence $f(\beta) = g(\beta)$.

Finally, $f \restriction \alpha = g \restriction \alpha$ yields $f = g$.

(ii) is an immediate consequence of (i). □

Lemma 5.3. *Let* $\mathfrak{D}_0[a, \alpha, f]$ *be the conjunction of the following formulas:* $\mathrm{Ord}(\alpha)$, $\mathrm{Fun}(f)$, $\mathrm{dom}(f) = \alpha \cup \{0\}$, $\mathrm{Ad}(a)$, $f(0) = a$, *and*

$$(\forall\beta < \alpha)[\mathrm{Ad}(f(\beta)) \wedge (\forall\eta < \beta)\,(f(\eta) \in f(\beta)) \wedge (\forall x \in f(\beta))[\mathrm{Ad}(x) \wedge a \in x \rightarrow (\exists\eta <$$

We then have $\mathbf{AUT\text{-}KPl}^r \vdash \mathrm{Ad}(a) \rightarrow \forall\alpha\,\exists! f\,\mathfrak{D}_0[a, \alpha, f]$.

Proof. Uniqueness of f can be proved as in Lemma 5.2. To prove existence suppose $\mathrm{Ad}(a)$. Invoking the axiom (Lim), there are admissible sets b and c such that $a, \alpha \in b$ and $b \in c$. Δ_0 separation relativized to c ensures the existence of $\rho := \{\eta \in b \mid \mathrm{Ord}(\eta)\}$ with $\rho \in c$. We also have $\rho \notin b$. Moreover, by (AUT-Ad) there exists a function g such that $\mathfrak{D}[\rho + \rho, g]$. The existence of the ordinal $\rho + \rho \in c$ can be established in the usual way since c is admissible. Using (Δ_0-FOUND) one easily shows that $(\forall\eta < \rho + \rho)\,\eta \in g(\eta + 1)$. Since $\rho \in g(\rho + 1)$ and $\rho \notin a$, the axiom (O7) ensures that $a \in g(\rho + 1)$. Thus there exists $\delta < \rho + 1$ such that $a = g(\delta)$. Also $\alpha < \rho$. The desired function f can be defined by $f(\eta) := g(\delta + \eta)$ for $\eta < \alpha$. One easily verifies that $\mathfrak{D}_0[a, \alpha, f]$. □

Definition 5.4.

$\mathrm{Fld}(s, r) := \exists x\,[\langle x, s\rangle \in r \,\vee\, \langle s, x\rangle \in r]$.

$\mathrm{Lo}(a, r) := r \subseteq a \times a \,\wedge$ *“r is a linear ordering” (cf. 3.1).*

$\mathrm{Wf}(a, r) := r \subseteq a \times a \,\wedge\, \forall x[x \neq \varnothing \,\wedge\, x \subseteq a \rightarrow (\exists y \in x)\,(\forall z \in x)\,\langle z, y\rangle \notin r]$.

$\mathrm{Wo}(a, r) := \mathrm{Lo}(a, r) \,\wedge\, \mathrm{Wf}(a, r)$.

Definition 5.5. *(Axiom Beta) If r is a well-founded relation on a, i.e.* $\mathrm{Wf}(a, r)$, *then f is said to be a* collapsing for r *if* $\mathrm{Collab}(a, r, f)$ *holds, where*

$$\mathrm{Collab}(a, r, f) := \mathrm{Fun}(f) \,\wedge\, \mathrm{dom}(f) = a \,\wedge\, (\forall x \in a)\,(f(x) = \{f(y) \mid \langle y, x\rangle \in r\}).$$

Axiom Beta *(cf. [3, I.9.4]) is the assertion* $\forall u \forall v\,\exists f\,[\mathrm{Wf}(u, v) \rightarrow \mathrm{Collab}(u, v, f)]$.

Theorem 5.6. $\mathbf{KPl}^r$ *proves axiom Beta. Inspection of the usual proof actually shows that* $\mathbf{KPl}^r$ *proves something stronger, namely that if* $\mathrm{Wf}(a,r)$, $\langle a,r\rangle \in b$, *and* $\mathrm{Ad}(b)$, *then the function which is collapsing for* r *is also an element of* b.

Proof. The proof is just a slight variation of the standard proof from $\mathbf{KP}$ + (Σ_1-separation) in [3, Theorem 9.6]: Just do the definition of the function F inside an admissible set $\mathbb{A}$ which contains the well-founded relation r. Then Σ_1 separation can be replaced by Δ_0 separation involving $\mathbb{A}$ as a parameter. For more details see [32, Theorem 4.6]. □

Theorem 5.7. *Every instance of* (BI) *is a theorem of* $\mathbf{KPl}$ *(via the translation* * *of Definition 2.5).*

Proof. By means of axiom Beta every well-ordering $\prec$ on $\mathbb{N}$ is order-isomorphic to an ordinal α. As a result, the schema of transfinite on $\prec$ is implied by (FOUND). For more details see [32, Lemma 7.1]. □

Lemma 5.8 (Iterated inductive definitions in $\mathbf{AUT}$-$\mathbf{KPl}^r$ and $\mathbf{KPl}^*$.). *For* $A[\mathrm{P}^+, b, c, d, \vec{t}]$ *in* $\Delta_0(\mathrm{P}^+)$ *let*

$$\mathrm{Cl}^A_{\mathrm{N}}(a,b,c,\vec{t}) \;:=\; \forall j \in \mathrm{N}(A[a,b,c,j,\vec{t}] \to j \in a)$$

and $\mathrm{IT}^A_{\mathrm{N}}(r,a,\vec{t})$ *be the formula*

$$a \subseteq \mathrm{N}\times\mathrm{N} \wedge (\forall i \in \mathrm{N})[\mathrm{Cl}^A_{\mathrm{N}}((a)_i,(a)_{ri},i,\vec{t}) \wedge \forall z \subseteq \mathrm{N}(\mathrm{Cl}^A_{\mathrm{N}}(z,(a)_{ri},i,\vec{t}) \to (a)_i \subseteq z)],$$

where $(a)_i := \{k \in \mathrm{N} \mid \langle i,k\rangle \in a\}$ *and* $(a)_{ri} := \{k \in \mathrm{N} \mid (\exists m \in \mathrm{N})(\langle m,i\rangle \in r \wedge \langle m,k\rangle \in a)\}$.

(i) $\mathbf{AUT}$-$\mathbf{KPl}^r \vdash \mathrm{Wo}(\mathrm{N},r) \to \exists y\, \mathrm{IT}^A_{\mathrm{N}}(r,y,\vec{t})$.

(ii) $\mathbf{KPl}^* \vdash \mathrm{Wo}(\mathrm{N},r) \wedge r,t \in \mathrm{M} \to \exists y\, \mathrm{IT}^A_{\mathrm{N}}(r,y,\vec{t})$.

Proof. (ii): Suppose $\mathrm{Wo}(\mathrm{N},r)$ and $r,\vec{t} \in \mathrm{M}$. Let $S := \{k \in \mathrm{N} \mid \mathrm{Fld}(k,r)\}$. Then $S \in \mathrm{M}$ and $\mathrm{Wo}(S,r)$. By Theorem 5.6 there exists a function $h \in \mathrm{M}$ such that $\mathrm{Collab}(S,r,h)$. Then $\mathrm{rng}(h)$ is an ordinal $\alpha \in \mathrm{M}$, $h : S \to \alpha$ is bijective and, moreover, $(\forall ij \in S)(\langle i,j\rangle \in r \to h(i) < h(j))$. By (Ad^*) there exists a function f such that $\mathfrak{D}[\alpha,f]$. In particular, $f(0) \in \mathrm{M}$. Using axiom (Lim) there exists an admissible set K such that $\alpha, f, \mathrm{M} \in K$. Let $K_\beta := f(\beta)$ for $\beta < \alpha$. Within the admissible K we simultaneously define a function g with $\mathrm{dom}(g) = \alpha$ and a sequence of functions $(f_\beta)_{\beta<\alpha}$ by Σ recursion as follows:

$$f_\beta(\xi) := \{k \in \mathrm{N} \mid A[\bigcup\{f_\beta(\gamma) \mid \gamma < \xi\}, \bigcup\{g(\delta) \mid \delta < \beta\}, h^{-1}(\beta), k, \vec{t}]\},$$

$$g(\beta) := \bigcup \mathrm{rng}(f_\beta).$$

For $i \in \mathrm{N} \setminus S$ let f_i be a function with $\mathrm{dom}(f_i) = \{\xi \mid \xi \in \mathrm{M}\}$ defined by Σ recursion in K via

$$f_i(\xi) := \{k \in \mathrm{N} \mid A(\bigcup\{f_i(\gamma) \mid \gamma < \xi\}, \varnothing, i, k, \vec{t}]\}.$$

Also let

$$a := \{\langle i, k\rangle \mid i \in S \wedge k \in g(h(i))\} \cup \{\langle i, k\rangle \mid i \in \mathrm{N} \setminus S \wedge k \in \mathrm{rng}(f_i)\}.$$

By construction, $f, g, (f_\beta)_{\beta<\alpha}$, and a are elements of K. To begin with we show that for $i \in \mathrm{N} \setminus S$,

$$\mathrm{Cl}^A_{\mathrm{N}}((a)_i, (a)_{ri}, i, \vec{t}),$$
$$\forall z \subseteq \mathrm{N}\, [\mathrm{Cl}^A_{\mathrm{N}}(z, (a)_{ri}, i, \vec{t}) \to (a)_i \subseteq z].$$

Proof of (5.1): We have $(a)_{ri} = \varnothing$ since $i \in \mathrm{N} \setminus S$. As M is admissible, f_i is Σ definable in M. If $A[(a)_i, \varnothing, i, k, \vec{t}]$ holds for some $k \in \mathrm{N}$ then $j \in (a)_i \leftrightarrow \exists \xi \in \mathrm{M}\, (j \in f_i(\xi))$, and therefore, since $(a)_i$ occurs positively, utilizing Σ reflection in M we arrive at

$$\exists \delta \in \mathrm{M}\, A[\bigcup\{f_i(\xi) \mid \xi < \delta\}, \varnothing, i, k, \vec{t}],$$

thus $\exists \delta \in \mathrm{M}\, (k \in f_i(\delta))$, so $k \in (a)_i$. This verifies (5.1).

Proof of (5.1): let $z \subseteq \mathrm{N}$ and suppose $\mathrm{Cl}^A_{\mathrm{N}}(z, \varnothing, i, \vec{t})$. By transfinite induction on $\xi \in \mathrm{M}$ we show that $\forall \xi \in \mathrm{M}\, (f_i(\xi) \subseteq z)$, yielding (5.1). So suppose inductively that $\bigcup\{f_i(\gamma) \mid \gamma < \xi\} \subseteq z$. Then

$$f_i(\xi) = \{k \in \mathrm{N} \mid A[\bigcup\{f_i(\gamma) \mid \gamma < \xi\}, \varnothing, i, k, \vec{t}]\} \subseteq \{k \in \mathrm{N} \mid A[z, \varnothing, i, k, \vec{t}]\}.$$

As $\mathrm{Cl}^A_{\mathrm{N}}(z, \varnothing, i, \vec{t})$ holds, the latter implies $f_i(\xi) \subseteq z$.

Next we address the case when $i \in S$ and to this end show, by induction on $\beta < \alpha$, that

$$g \restriction \beta \in K_\beta.$$

Suppose that $g \restriction \delta \in K_\delta$ for all $\delta < \beta$. Then also $\forall \delta < \beta (g \restriction \delta \in K_\beta)$, and the sequence $(g \restriction \delta)_{\delta<\beta}$ is thus Σ definable in the admissible K_β. Note that since $\alpha \in \mathrm{M}$ and $\beta < \alpha$, we have $\beta \in K_\beta$. Using Σ replacement (cf. Theorem 2.23) inside K_β, we then have $(g \restriction \delta)_{\delta<\beta} \in K_\beta$. If β is a limit ordinal we have $g \restriction \beta = \bigcup\{g \restriction \delta \mid \delta < \beta\} \in K_\beta$. Suppose β is a successor $\rho + 1$. Then $g \restriction \rho$

and $\mathrm{dom}(f_\rho)$ are elements of K_β. Thus, by Σ recursion in K_β, f_ρ belongs to K_β, too. Therefore, $g \restriction \beta = g \restriction \rho \cup \{\langle \rho, \bigcup \mathrm{rng}(f_\rho)\rangle\} \in K_\beta$. As a result, transfinite induction on β establishes (5.1).

Now assume $i \in S$ and $i = h^{-1}(\beta)$. We have to show (5.1) and (5.1) for i. From (5.1) and the definition of a we get $(a)_{ri} \in K_\beta$. Also $(a)_i = \mathrm{rng}(f_\beta)$ and, for $\xi \in K_\beta$,

$$f_\beta(\xi) = \{k \in \mathrm{N} \mid A[\bigcup\{f_\beta(\gamma) \mid \gamma < \xi\}, (a)_{ri}, i, k, \vec{t}]\}.$$

So f_β is definable by Σ recursion in K_β. From $A[(a)_i, (a)_{ri}, i, k, \vec{t}]$ it follows, by Σ reflection in K_β, that

$$\exists \xi \in K_\beta\, A[\bigcup\{f_\beta(\gamma) \mid \gamma < \xi\}, (a)_{ri}, i, k, \vec{t}],$$

and hence $k \in (a)_i$, thereby showing (5.1) for $i \in S$. (5.1) can be shown for $i \in S$ in the same way as for $i \in \mathrm{N} \setminus S$.

(i) can be shown in the same way as (ii), except for a small change which consists in choosing an admissible set b such that $r, \vec{t} \in b$ and invoking Lemma 5.3 to ensure the existence of a function f with $\mathfrak{D}_0[b, \alpha, f]$. □

Theorem 5.9. *Via the translation $\,^*$ of definition 2.5 we have*

(i) Π^1_1-$\mathbf{TR}_0 \subseteq \mathbf{AUT}$-$\mathbf{KPl}^r$.

(ii) Π^1_1-$\mathbf{TR} \subseteq \mathbf{AUT}$-$\mathbf{KPl}^w$.

(iii) Π^1_1-$\mathbf{TR} + (\mathrm{BI}) \subseteq \mathbf{AUT}$-$\mathbf{KPl}$.

(iv) $\mathbf{ID}_{\prec^*} \overset{\wedge}{\longrightarrow} \mathbf{ID}^* + (\mathrm{BI}) \subseteq \mathbf{KPl}^*$

where in (iv) the first the translation ˆ stems from Definition 3.9.

Proof. (i) follows from Theorem 4.6(i) and Lemma 5.8(i). (ii) is an immediate consequence of (i) as does (iii) if viewed in conjunction with Theorem 5.7. It reamins to show (iv). For $Q[a, b, c]$ arithmetical, we have $r_i := \{\langle j, k\rangle \mid j, k \in \mathrm{N} \wedge Q[i, j, k]^*\} \in \mathrm{M}$ for $i \in \mathrm{N}$. Therefore Lemma 5.8(ii) and Theorem 5.7 imply that $\mathbf{ID}^* + (\mathrm{BI}) \subseteq \mathbf{KPl}^*$. The first entailment via ˆ is a consequence of Theorem 3.10 and Theorem 3.13(i). □

Finally we would like to find a set-theoretic pendant to Π^1_1-$\mathbf{TR} + \Sigma^1_2$-$\mathbf{AC}$. We take this as an opportunity to introduce a few more traditional axiom schemata considered in second order arithmetic (cf. [17]).

Definition 5.10. *Let $\mathcal{F}$ be a collection of formulae in $\mathcal{L}_2$.*

$$(\mathcal{F}\text{-CA}) := \{\exists Z \forall x\, [x \in Z \leftrightarrow F(x)] \mid F(a) \in \mathcal{F}\}.$$
$$(\mathcal{F}\text{-AC}) := \{\forall x \exists Y\, H(x, Y) \to \exists Z \forall x\, H(x, Z_x) \mid H(a, U) \in \mathcal{F}\}.$$
$$(\mathcal{F}\text{-DC}) := \{\forall X \exists Y\, G(X, Y) \to \forall W \exists Z\, [W = Z_{\bar{0}} \wedge \forall x\, G(Z_x, Z_{x+1})] \mid G(U, V) \in \mathcal{F}\}$$
$$(\Delta^1_2\text{-CA}) := \{\forall x\, [A(x) \leftrightarrow B(x)] \to \exists Z \forall x\, [x \in Z \leftrightarrow A(x)] \mid A(a) \in \Pi^1_2, B(a) \in \Sigma^1_2\}.$$

If (S) denotes any of the above schemata, then $\mathbf{S}$ stands for the theory $\mathbf{ACA} + (S)$.

The following well-known relationships can be found in [18, Theorem 2.3.1].

Theorem 5.11. *(i)* Σ^1_1-$\mathbf{AC}_0 \subseteq \Pi^1_1$-$\mathbf{CA}_0$.

(ii) Δ^1_2-$\mathbf{CA} = \Sigma^1_2$-$\mathbf{AC} = \Sigma^1_2$-$\mathbf{AC}$.

Theorem 5.12. Π^1_1-$\mathbf{TR} + \Sigma^1_2$-$\mathbf{AC} \subseteq \mathbf{KPi}^w + \mathbf{AUT}$-$\mathbf{KPl}^w$.

Proof. In view of Theorem 5.9(ii) it suffices to show Σ^1_2-$\mathbf{AC} \subseteq \mathbf{KPi}^w$. But this inclusion is a consequence of Theorem 5.11 and Theorem 7.2, a result we shall show later. □

Remark 5.13. *(i) Theorem 5.9 crucially uses Lemma 5.8 which is essentially a generalization of Gandy's Theorem (cf. [3, VI.2.6]) to the iterated scenario.*

(ii) Theories of iterated admissibility were also considered by Jäger in [30]. However, in the theories in [30] iterated admissibility is couched in terms of inference rules and they come also equipped with an extended Bar rule. As a result, they are different from the theories considered here. There are several conjectures about the proof-theoretic strength of such theories stated in [30]. These conjectures turn out to be true as they are corollaries of results in this paper. Details will be spelled out at the appropriate places.

6 Theories of iterated choices and Δ^1_2 comprehension

Let $\mathbf{N}$ be the standard structure of the natural numbers with language $\mathcal{L}_1$. Every level $\mathrm{L}(\alpha)_{\mathbf{N}}$ of the constructible hierarchy above $\mathbf{N}$ (for a precise definition see [3, II]) can be viewed as a structure of the language $\mathcal{L}^*$ wherein the predicate symbol Ad is interpreted by the class $\{\mathrm{L}(\beta)_{\mathbf{N}} \mid \beta < \alpha$ and $\mathrm{L}(\beta)_{\mathbf{N}}$ is admissible$\}$.

If $\mathbf{T}$ is a theory with language $\mathcal{L}^*$, then the structures $\mathrm{L}(\alpha)_{\mathbf{N}}$ satisfying and $\mathrm{L}(\alpha)_{\mathbf{N}} \models \mathbf{T}$ are said to be the *standard models* of $\mathbf{T}$.

The smallest standard model of the theories $\mathbf{AUT}$-$\mathbf{KPl}^r$, $\mathbf{AUT}$-$\mathbf{KPl}^w$ and $\mathbf{AUT}$-$\mathbf{KPl}$ is $\mathrm{L}(\mathrm{g}_1(0))_{\mathbf{N}}$ where the mapping $\xi \mapsto \mathrm{g}_0(\xi)$ enumerates the admissible ordinals $\geq \omega^{CK}_1$ and their limits, and (recursively) for $\alpha > 0$, $\xi \mapsto \mathrm{g}_\alpha(\xi)$ enumerates the common fixed points of all the functions g_β with $\beta < \alpha$.

All further $\mathcal{L}_2$-theories to be introduced in this section and in section 8 will comprise Δ^1_2-$\mathbf{CA}_0$ and will turn out to be subtheories of Δ^1_2-$\mathbf{CA}$ + (BI). On the set-theoretic side they correspond to theories in strength between $\mathbf{KPi}^r$ and $\mathbf{KPi}$. The difference in proof-theoretic strength between the latter two theories is enormous, albeit both theories have the same minimal standard model $\mathrm{L}(\iota_0)_{\mathbf{N}}$ with ι_0 being the least recursively inaccessible ordinal. As a result, the minimal standard model is hardly indicative of the proof-theoretic strength of these theories. A better measure is provided by the minimal Π_2-model.

Definition 6.1. $\mathrm{L}(\alpha)_{\mathbf{N}}$ *is a* Π_2-model *of a set theory* $\mathbf{T}$*, whenever*

$$\mathbf{T} \vdash F \Rightarrow \mathrm{L}(\alpha)_{\mathbf{N}} \models F$$

holds for all set-theoretic Π_2 *sentences.*

(The notion of a Π_2*-model appears to have been introduced in [32].)*

As far as the theories $\mathbf{AUT}$-$\mathbf{KPl}^r$, $\mathbf{AUT}$-$\mathbf{KPl}^w$ and $\mathbf{AUT}$-$\mathbf{KPl}$ are concerned, $\mathrm{L}(\mathrm{g}_1(0))_{\mathbf{N}}$ is also their minimal Π_2-model. The theories $\mathbf{T}$ with $\mathbf{KPi}^r \subseteq \mathbf{T} \subseteq \mathbf{KPi}$ we are going to study next, though, will have their minimal Π_2-model $\mathrm{L}(\alpha)_{\mathbf{N}}$ at an ordinal $\alpha \leq \Gamma^{\mathrm{g}}_0 := \min\{\rho \mid \mathrm{g}_\rho(0) = \rho\}$. The main cause for the widely diverging Π_2-models of such theories is to be found in the amount of induction principles they incorporate. In conjunction with stronger induction principles, the pivotal principle of Σ collection gives rise to recursion principles which allow one to prove the existence of ever larger admissible sets.

To analyze the gap between Π^1_1-$\mathbf{TR}$ and Δ^1_2-$\mathbf{CA}$ + (BI) we consider iterations of principles stronger than Π^1_1-comprehension.

Definition 6.2. *(i)* Δ^1_2-$\mathbf{TR}$ *is the theory* $\mathbf{ACA}$ *augmented by the schema of transfinite* Δ^1_2 *recursion,* (Δ^1_2-TR),

$$\forall R[\mathrm{WO}(R) \wedge \forall X \forall iy\, [B(X,i,y) \leftrightarrow A(X,i,y)] \rightarrow \exists Z \forall iy\, [y \in Z_i \leftrightarrow B(Z_{Ri},i,y)]]$$

with $B(U,a,b) \in \Pi^1_2$ *and* $A(U,a,b) \in \Sigma^1_2$.

(ii) Σ^1_2-$\mathbf{TRDC}$ *(*Π^1_1-$\mathbf{TRDC}$*, respectively) is the theory* $\mathbf{ACA}$ *augmented by the schema of transfinitely iterated* Σ^1_2 *(*Π^1_1*) dependent choices,* (Σ^1_2-TRDC) *(*Π^1_1-TRDC*, respectively),*

$$\forall R[\mathrm{WO}(R) \wedge \forall i \forall X \exists Y\, C(X,Y,i) \rightarrow \exists Z \forall i\, C(Z_{Ri},Z_i,i)]$$

where $C(U,V,a) \in \Sigma^1_2$ *(*$C(U,V,a) \in \Pi^1_1$*, respectively).*

As it turns out, Π^1_1 dependent choices are as strong as Σ^1_2 dependent choices.

Lemma 6.3. Π^1_1-$\mathbf{TRDC}_0 = \Sigma^1_2$-$\mathbf{TRDC}_0$.

Proof. we have to show "$\supseteq$". Let $C(U,V,a)$ be a formula $\exists W\, A(U,V,W,a)$ with $A(U,V,S,a)$ Π^1_1. Suppose that $\mathrm{WO}(R)$ and $\forall i \forall X \exists Y\, C(X,Y,i)$. Then also

$$\forall i \forall X \exists Y\, A(\{z \mid \langle z,\bar{0}\rangle \in X\}, \{z \mid \langle z,\bar{0}\rangle \in Y\}, \{z \mid \langle z,\bar{1}\rangle \in Y\}, i)$$

and by (Π^1_1-TRDC) there exists V such that

$$\forall i\, A(\{z \mid \langle z,\bar{0}\rangle \in V_{Ri}\}, \{z \mid \langle z,\bar{0}\rangle \in V_i\}, \{z \mid \langle z,\bar{1}\rangle \in V_i\}, i),$$

and hence

$$\forall i\, C(\{z \mid \langle z,\bar{0}\rangle \in V_{Ri}\}, \{z \mid \langle z,\bar{0}\rangle \in V_i\}, i).$$

Letting $V^* : \{\langle i,z\rangle \mid \langle i, \langle z,\bar{0}\rangle\rangle \in V\}$, (6.1) implies $\forall i\, C(V^*_{Ri}, V_i^*, i)$. □

Lemma 6.4. Δ^1_2-$\mathbf{CA}_0 \subseteq \Sigma^1_2$-$\mathbf{TRDC}_0$.

Let $F(a), \neg G(a) \in \Sigma^1_2$. Suppose that $\forall x\, [G(x) \leftrightarrow F(x)]$. Then

$$\forall x \forall X \exists Y\, [(F(x) \wedge Y = \{\bar{0}\}) \vee (\neg G(x) \wedge Y = \{\bar{1}\})].$$

Applying (Σ^1_2-TRDC) to (6.1) and the well ordering $\varnothing$, there exists a set V such that

$$\forall x\, [(F(x) \wedge V_x = \{\bar{0}\}) \vee (\neg G(x) \wedge V_x = \{\bar{1}\})].$$

With $V' := \{x \mid V_x = \{\bar{0}\}\}$ we obtain the desired $\forall x\, [x \in V' \leftrightarrow F(x)]$. □

Lemma 6.5. Δ^1_2-$\mathbf{TR}_0 \subseteq \Sigma^1_2$-$\mathbf{TRDC}_0$.

Proof. Let $B(U,a,b), \neg A(U,a,b) \in \Sigma^1_2$. Moreover, suppose that $\mathrm{WO}(R)$ and

$$\forall X \forall i y\, [B(X,i,y) \leftrightarrow A(X,i,y)].$$

By Lemma 6.4, (6.1) imples

$$\forall i \forall X \exists Y \forall y\, [B(X,i,y) \leftrightarrow y \in Y].$$

The formula $\forall y\, [B(X,i,y) \leftrightarrow y \in Y]$, in view of (6.1) and the fact that (Σ^1_2-AC) $\subseteq$ Σ^1_2-$\mathbf{TRDC}_0$, is equivalent to a Σ^1_2 formula. Hence, with (Σ^1_2-TRDC), from (6.1) we obtain

$$\exists Z \forall i y\, [B(Z_{Ri}, i, y) \leftrightarrow y \in Z_i].$$

□

To facilitate the interpretation of Σ^1_2-$\mathbf{TRDC}_0$ into set theory without choice, we first reduce this theory to a theory Π^1_1-$\mathbf{TRK}_0$.

Definition 6.6. Π^1_1-$\mathbf{TRK}$ *is the theory* Δ^1_2-$\mathbf{CA}$ + (Π^1_1-TRK) *where*

$$(\Pi^1_1\text{-TRK}) \qquad \forall R\,[\mathrm{WO}(R) \wedge \forall i \forall X \exists ! Y\, D(X,Y,i) \to \exists Z \forall i\, D(Z_{Ri}, Z_i, i)]$$

with $D(U,V,a) \in \Pi^1_1$.

Lemma 6.7. Π^1_1-$\mathbf{TRK}_0 = \Pi^1_1$-$\mathbf{TRDC}_0$.

For the proof of Lemma 6.7 we need to show that a certain result from descriptive set theory is provable in Δ^1_2-$\mathbf{CA}_0$.

Lemma 6.8 (Π^1_1 uniformization)**.** *For every* Π^1_1 *formula* $A[\vec{S}, V, \vec{a}\,]$ *there exists a* Π^1_1 *formula* $H[\vec{S}, V, \vec{a}\,]$ *such that provably in* Π^1_1-$\mathbf{CA}_0$ *we have*

(i) $\forall Y\,(H[\vec{S}, V, \vec{a}\,] \to A[\vec{S}, V, \vec{a}\,])$*;*

(ii) $\exists Y\, A[\vec{S}, Y, \vec{a}\,] \to \exists ! Y\, H[\vec{S}, Y, \vec{a}\,]$.

Proof. [65, Lemma VI.2.1] □

Proof of Lemma 6.7: "⊆" follows from Lemma 6.3 and 6.4. For "⊇" let $A[\vec{S}, U, V, b, \vec{a}\,] \in \Pi^1_1$ and assume

$$\mathrm{WO}(R) \wedge \forall i \forall X \exists Y\, A[\vec{S}, X, Y, i, \vec{a}\,].$$

By Lemma 6.8 there is a Π^1_1 formula $H[\vec{S}, U, V, b, \vec{a}\,]$ such that

$$\forall i \forall X \forall Y\,(H[\vec{S}, X, Y, i, \vec{a}\,] \to A[\vec{S}, X, Y, i, \vec{a}\,])$$
$$\forall i \forall X \exists ! Y\, H[\vec{S}, X, Y, i, \vec{a}\,].$$

With the aid of (Π^1_1-TRK), (6.1) yields

$$\exists Z \forall i\, H[\vec{S}, Z_{Ri}, Z_i, i, \vec{a}\,].$$

(6.1) and (6.1) imply $\exists Z \forall i\, A[\vec{S}, Z_{Ri}, Z_i, i, \vec{a}\,]$. □

Lemma 6.9. Π^1_1-$\mathbf{TRK}_0 \subseteq \Delta^1_2$-$\mathbf{TR}_0$.

Proof. Assume $\mathrm{WO}(R) \wedge \forall i \forall X \exists ! Y\, C(X,Y,i)$ for some Π^1_1 formula $C(U,V,a)$. Let $B(U,a,b) := \exists Y\,[C(U,Y,a) \wedge b \in Y]$ and $A(U,a,b) := \forall Y\,[C(U,Y,a) \to b \in Y]$. By assumption,

$$\forall X \forall i \forall y\,[B(X,i,y) \leftrightarrow A(X,i,y)].$$

Using (Δ^1_2-TR), (6.1) yields the existence of a set S such that for all i,

$$S_i = \{y \mid \exists Y\, [C(S_{Ri}, Y, i) \,\wedge\, y \in Y]\}.$$

As $\forall i \exists! Y\, C(S_{Ri}, Y, i)$ it follows that $\forall i\, C(S_{Ri}, S_i, i)$. □

By Lemmata 6.3, 6.5, 6.7, and 6.9 we have the following:

Theorem 6.10. $\Delta^1_2\text{-}\mathbf{TR}_0 = \Sigma^1_2\text{-}\mathbf{TRDC}_0 = \Pi^1_1\text{-}\mathbf{TRDC}_0 = \Pi^1_1\text{-}\mathbf{TRK}_0$.

Lemma 6.8 also yields the following.

Theorem 6.11. $\Sigma^1_2\text{-}\mathbf{AC}_0 = \Delta^1_2\text{-}\mathbf{CA}_0$.

Another natural route to approach $\Delta^1_2\text{-}\mathbf{CA}$ + (BI) from below is to consider restrictions of the bar induction schema (BI).

Definition 6.12. *If $\mathcal{F}$ is a collection of $\mathcal{L}_2$-formulas, we let*

$$(\mathcal{F}\text{-BI}) := \{\forall X\, [\mathrm{WO}(X) \to \mathrm{TI}(X, F)] \mid F(a) \in \mathcal{F}\}.$$

It worthwhile noting that (Π^1_2-BI) is already deducible in $\Delta^1_2\text{-}\mathbf{CA}$.

Theorem 6.13. $(\Pi^1_2\text{-BI}) \subseteq \Sigma^1_2\text{-}\mathbf{DC}_0 \subseteq \Delta^1_2\text{-}\mathbf{CA}$.

Proof. The second inclusion follows from Theorem 5.11(ii). To show the first inclusion we argue in $\Sigma^1_2\text{-}\mathbf{DC}_0$. Suppose we have a counter-example to (Π^1_2-BI). Then there is a formula $H(a) = \forall X\, A(X, a)$ with $A(X, a) \in \Sigma^1_1$, and there exists a well-ordering $\prec$ and a number k such that

$$\mathrm{PROG}(\prec, H) \,\wedge\, \neg H(k).$$

Let variables $f, g, h, \ldots$ range over functions from N^{N}, where we identify f with the set $\{\langle n, f(n)\rangle \mid n \in \mathrm{N}\}$.

Since $\neg H(k)$ holds, there exists a set S such that $\neg A(S, k)$. Let f' be defined by $f'(0) = k$ and

$$f'(x+1) = \begin{cases} 0 \text{ if } x \in S \\ 1 \text{ if } x \notin S\,. \end{cases}$$

Letting $N(h) := \{x \mid h(x+1) = 0\}$ for $h \in \mathrm{N}^{\mathrm{N}}$, we have

$$f'(0) = k \wedge \neg A(N(f'), k).$$

Since $\mathrm{PROG}(\prec, h)$ we get $\forall i\, [\exists X\, \neg A(X, i) \;\to\; \exists Y \exists j (j \;\prec\; i \;\wedge\; \neg A(Y, j))]$, whence

$$\forall f \exists g\, [\neg A(N(f), f(0)) \to g(0) \prec f(0) \,\wedge\, \neg A(N(g), g(0))].$$

Applying (Σ^1_2-DC) to (6.1), we obtain a function h such that

$$h_0 = f' \wedge \forall i\,[\neg A(N(h_i), h_i(0)) \to h_{i+1}(0) \prec h_i(0) \wedge \neg A(N(h_{i+1}), h_{i+1}(0))],$$

where h_i denotes the function $x \mapsto h(\langle i, x\rangle)$.

Using induction (for a Π^1_1 formula), (6.1) and (6.1) imply

$$\forall i\,[\neg A(N(h_i), h_i(0)) \wedge h_{i+1}(0) \prec h_i(i)],$$

violating the assumption that $\prec$ is a well-ordering. □

The dual formula class, though, provides a strengthening.

Theorem 6.14. Σ^1_2-$\mathbf{TRDC}_0 \subseteq \Delta^1_2$-$\mathbf{CA}_0 + (\Sigma^1_2$-BI$)$.

Proof. According to Theorem 6.10 it suffices to show (Π^1_1-TRK) $\subseteq \Delta^1_2$-$\mathbf{CA}_0 +$ (Σ^1_2-BI). So suppose we have a Π^1_1 formula $C(U, V, a)$ such that

$$\mathrm{WO}(R) \wedge \forall i \forall X \exists! Y\, C(X, Y, i).$$

Let $F(U, a) := \forall j\,[(jRa \vee j \equiv a) \to C(U_{Rj}, U_j, j)]$. As $\mathrm{WO}(R)$ we get

$$F(U, a) \wedge F(V, a) \to \forall j\,[(jRa \vee j \equiv a) \to U_j = V_j].$$

We show

$$\forall i \exists Z\, F(Z, i)$$

by induction on R. From $\forall x\,[xRi \to \exists Z\, F(Z, x)]$ it follows by (6.1) and with the help of (Δ^1_2-CA) that there exists a set S such that

$$S = \{\langle x, y\rangle \mid xRi \wedge \exists Z\,(F(Z, x) \wedge y \in Z_x)\} = \{\langle x, y\rangle \mid xRi \wedge \forall Z\,[F(Z, x) \to y \in Z_x)\}.$$

Moreover, owing to (6.1), there exists a set V such that $C(S_{Ri}, V, i)$.

Letting $S^* := S \cup \{\langle i, y\rangle \mid y \in V\}$ we have $F(S^*, i)$. Thus (6.1) follows by (Σ^1_2-BI).

In view of (6.1) and (6.1), we can apply (Δ^1_2-CA) to show that

$$U := \{\langle i, y\rangle \mid \exists Z\,[F(Z, i) \wedge y \in Z_i)\}$$

is set. Moreover, by (6.1) and (6.1), we also have $\forall i\, C(U_{Ri}, U_i, i)$, showing ($\Pi^1_1$-TRK). □

Corollary 6.15. (Π^1_2-BI) $\subseteq \Delta^1_2$-$\mathbf{CA}_0 +$ (Σ^1_2-BI).

Proof. This follows from Theorems 6.13 and 6.14. □

Remark 6.16. *We shall later see that* Σ^1_2-$\mathbf{TRDC}_0$ *and* Δ^1_2-$\mathbf{CA}_0 + (\Sigma^1_2\text{-BI})$ *have the same prooftheoretic ordinal. Using standard arguments this imples that both theories prove the same* Π^1_1 *sentences. Indeed, this result can be improved. Both theories prove the same* Π^1_3 *sentences, but this stronger conservativity result cannot be simply gleaned from the proof-theoretic ordinal. One has to scrutinize the whole series of reductions to arrive at it.*

7 Set theories with recursion schemata

As in the case of $\mathcal{L}_2$-theories of iterated Π^1_1-comprehension, one can also single out a set-theoretic counterpart to (Σ^1_2-TRDC). Ignoring the latter's choice aspects, Σ-recursion lends itself as a pendant to (Σ^1_2-TRDC). In order to interpret Σ^1_2-$\mathbf{TRDC}_0$ in $\mathbf{KPi}^r + (\Sigma\text{-REC})$, we need a "quantifier theorem" which reduces Σ^1_2 formulae of $\mathcal{L}_2$ to set-theoretic Σ_1 formulas, thereby reducing the number of unbounded set quantifiers by one. In the case of $\mathbf{ZF}$ this is a standard result. (cf. [12, CH.5,7.14]).

Theorem 7.1. *To any* Σ^1_2 $\mathcal{L}_2$*-formula* $B[\vec{a},\vec{U}\,]$ *one can assign a* Σ_1 *formula* $B_\sigma[\vec{a},\vec{b}\,]$ *of* $\mathcal{L}^*$ *such that*

$$\mathbf{KPl}^r \vdash \vec{a} \in \mathrm{N} \wedge b \subseteq \mathrm{N} \to (B[\vec{a},\vec{b}\,]^* \leftrightarrow B_\sigma[\vec{a},\vec{b}\,]).$$

Proof. The crucial step in the well known proof (usually carried out in $\mathbf{ZF}$) consists in realizing that via the Π^1_1 normal form (the equivalence (4.1) in the proof of Lemma 4.5), every Π^1_1 formula is equivalent to a Σ_1 formula exploiting axiom Beta, and consequently every Π^1_1 formula is Δ_1. Since axiom Beta is provable in $\mathbf{KPl}^r$ by Theorem 5.6, the desired result follows. For more details see [32, Theorem 7.1]. □

Theorem 7.2. Δ^1_2-$\mathbf{CA}_0 \subseteq \mathbf{KPi}^r$.

Immediate by the latter Theorem, using Δ separation (Theorem 2.22) in $\mathbf{KPi}^r$. □

Lemma 7.3 (Embedding Lemma for Π^1_1-$\mathbf{TRK}_0$)**.** Π^1_1-$\mathbf{TRK}_0 \subseteq \mathbf{KPi}^r + (\Sigma\text{-REC})$.

Proof. By Theorem 7.1 it suffices to show (Π^1_1-TRK) $\subseteq \mathbf{KPi}^r + (\Sigma\text{-REC})$. Let $A[a,U,V,\vec{d},\vec{S}\,] \in \Sigma^1_2$. Let $j_1,\ldots,j_k \in \mathrm{N}$, $s_1,\ldots,s_n \subseteq \mathrm{N}$ and, letting $B(a,b,c) := (A[a,b,c,\vec{j},\vec{s}\,])^*$, assume that

$$\mathrm{Wo}(r,\mathrm{N}) \wedge \forall i \in \mathrm{N}\, \forall x \subseteq \mathrm{N}\, \exists! y\, [y \subseteq \mathrm{N} \wedge B(i,x,y)].$$

We have to show that

$$\exists z \subseteq \mathrm{N} \times \mathrm{N}\, (\forall i \in \mathrm{N})\, B(i, (z)_{ri}, (z)_i)$$

(with $(z)_{ri}$ and $(z)_i$ being defined as in Lemma 5.8). By Theorem 7.1 there exists a Σ_1 formula $B_\sigma(a, b, c)$ such that

$$\forall i \in \mathrm{N}\, \forall x \subseteq \mathrm{N}\, \forall y \subseteq \mathrm{N}\, [B(i, x, y) \leftrightarrow B_\sigma(i, x, y)].$$

Letting $S := \{i \in \mathrm{N} \mid \mathrm{Fld}(i, r)\}$, by Theorem 5.6 there exists a function h such that h is collapsing for r, i.e. $\mathrm{Collab}(S, r, h)$. Whence h is a bijection from S onto $\alpha_h := \mathrm{rng}(h)$ satisfying $\forall ij\, [\langle i, j\rangle \in r \to h(i) < h(j)]$. Let

$$F(\beta, a, b) := F_0(\beta, a, b) \vee F_1(\beta, a, b),$$
$$F_0(\beta, a, b) := (\alpha_h \leq \beta \vee \bigcup \mathrm{rng}(a) \nsubseteq \mathrm{N}) \wedge b = \varnothing,$$
$$F_1(\beta, a, b) := \beta < \alpha_h \wedge \bigcup \mathrm{rng}(a) \subseteq \mathrm{N} \wedge b \subseteq \mathrm{N} \wedge B_\sigma(h^{-1}(\beta), \bigcup \mathrm{rng}(a), b).$$

In $\mathbf{KPl}^r$, $F(\beta, a, b)$ is equivalent to a Σ formula. In view of (7.1) and (7.1) we have $\forall\beta\forall x\exists! y\, F(\beta, x, y)$, whence, using ($\Sigma$-REC), there exists a function f, such that

$$\mathrm{dom}(f) = \alpha_h \wedge (\forall \beta < \alpha_h)\, F(\beta, f \restriction \beta, f(\beta))).$$

From (7.1) we get $(\forall \beta < \alpha_h)\, [\bigcup \mathrm{rng}(f \restriction \beta) \subseteq \mathrm{N} \wedge f(\beta) \subseteq \mathrm{N}$ by (Δ_0-FOUND). Whence from (7.1) we can conclude that for all $i \in S$,

$$\bigcup \mathrm{rng}(f \restriction h(i)) \subseteq \mathrm{N} \wedge f(h(i)) \subseteq \mathrm{N} \wedge B_\sigma(i, \bigcup \mathrm{rng}(f \restriction h(i)), f(h(i))).$$

With $X := \{\langle i, j\rangle \mid i \in S \wedge j \in f(h(i))\}$, (7.1) and (7.1) yield

$$(\forall i \in S)\, B(i, (X)_{ri}, (X)_i) \wedge X \subseteq \mathrm{N} \times \mathrm{N}.$$

Moreover, (7.1) implies $(\forall i \in \mathrm{N} \setminus S)\, \exists! y\, [y \subseteq \mathrm{N} \wedge B(i, \varnothing, y)]$, so that with the help of Σ replacement there exists a function g satisfying

$$\mathrm{dom}(g) = \mathrm{N} \setminus S \wedge (\forall i \in \mathrm{N} \setminus S)\, [g(i) \subseteq \mathrm{N} \wedge B(i, \varnothing, g(i))].$$

Letting $Y := X \cup \{\langle i, j\rangle \mid i \in \mathrm{N} \setminus S \wedge j \in g(i)\}$, (7.1) and (7.1) entail that

$$Y \subseteq \mathrm{N} \times \mathrm{N} \wedge (\forall i \in \mathrm{N})\, B(i, (Y)_{ri}, (Y)_i),$$

confirming (7.1). □

From Theorem 6.10 and Lemma 7.3 we get the following.

Theorem 7.4. Σ^1_2-$\mathbf{TRDC}_0 = \Delta^1_2$-$\mathbf{TR}_0 \subseteq \mathbf{KPi}^r + (\Sigma\text{-REC})$.

The theories Σ^1_2-$\mathbf{TRDC}_0$ and Σ^1_2-$\mathbf{TRDC}$ will later be interpreted in a semi-formal system of ramified set theory. This, however, will only provide a partial interpretation for Σ_1 formulae with free variables. To bring about this interpretation it is technically advisable to reduce $(\Sigma\text{-REC})$ to a simpler schema of Ad-valued recursion on ordinals.

Definition 7.5. *For $F(a,b,c)$ a Δ_0 formula, we denote by $\mathcal{C}^F(\alpha,f)$ the formula*

$$\begin{aligned}&\mathrm{Ord}(\alpha) \wedge \mathrm{Fun}(f) \wedge \mathrm{dom}(f) = \alpha \wedge \\ &(\forall\beta < \alpha)\,[\mathrm{Ad}(f(\beta)) \wedge F(\beta, f \restriction \beta, f(\beta)) \wedge (\forall x \in f(\beta))(\mathrm{Ad}(x) \to \neg F(\beta, f \restriction \beta, x))]\end{aligned}$$

Note that $\mathcal{C}^F(\alpha,f)$ is also Δ_0. By $(\mathrm{Ad\text{-}REC})$ *we denote the schema*

$$\forall\beta\forall x\exists y[\mathrm{Ad}(y) \wedge F(\beta,x,y)] \to \forall\alpha\exists f\, \mathcal{C}^F(\alpha,f)$$

where $F(a,b,c)$ is Δ_0.

As in Lemma 5.2 one proves

Lemma 7.6. $\mathbf{KPl}^r \vdash \mathcal{C}^F(\alpha,g) \wedge \mathcal{C}^F(\alpha,g) \to f = g$.

The "trick" of replacing the premiss $\forall\beta\forall x\exists!y\, F(\beta,x,y)$ by $\forall\beta\forall x\exists y[\mathrm{Ad}(y) \wedge F(\beta,x,y)]$ allows one to relinquish one's hold on the uniqueness requirement for y since admissible sets are well-ordered on the basis of $\mathbf{KPl}^r$.

Theorem 7.7. $\mathbf{KPi}^r + (\mathrm{Ad\text{-}REC}) = \mathbf{KPi}^r + (\Sigma\text{-REC})$.

Proof. For a proof see [43, Satz 5.7]. A proof will also be supplied in the sequel to this paper. □

8 Systems with Bar rules and other induction principles

An alternative to restricting the schema (BI) to specific syntactic complexity classes of formulae (as in $(\mathcal{F}\text{-BI})$) consists in directing the attention to the well-ordering over which transfinite induction is allowed in that one requires them to be provably well-ordered.

Definition 8.1. *(i) The* Bar rule, BR, *is the rule of inference*

$$\frac{\mathrm{WO}(\prec)}{\mathrm{TI}(\prec, F)}$$

with $\prec$ being a primitive recursive relation and $F(a)$ any formula of $\mathcal{L}_2$.

(ii) $\mathrm{BR}(\text{impl-}\Sigma^1_2)$ *is the rule*

$$\frac{\exists! X\,(\mathrm{WO}(X)\ \wedge\ G[X])}{\forall X\,(\mathrm{WO}(X)\ \wedge\ G[X] \to \mathrm{TI}(X,H))}$$

where $G[U]$ *is a* Σ^1_2 *formula (without additional parameters) and* $H(a)$ *is an arbitrary* $\mathcal{L}_2$ *formula.*

(iii) $\mathrm{BI}(\text{impl-}\Sigma^1_2)$ *denotes the schema*

$$\exists! X\,(\mathrm{WO}(X)\ \wedge\ G[X]) \to \forall X\,(\mathrm{WO}(X)\ \wedge\ G[X] \to \mathrm{TI}(X,H))$$

where $G[U]$ *is a* Σ^1_2 *formula (without additional parameters) and* $H(a)$ *is an arbitrary* $\mathcal{L}_2$ *formula.*

The Quantifier Theorem 7.1 and Axiom Beta suggest set-theoretic equivalences to the foregoing induction principles.

Definition 8.2. *(i)* $\mathrm{FOUNDR}(\text{impl-}\Sigma(\mathrm{M}))$ *is the rule of inference*

$$\frac{\exists! x\,(x \in \mathrm{M}\ \wedge\ F[x]^{\mathrm{M}})}{\forall x[x \in \mathrm{M} \wedge\ F[x]^{\mathrm{M}}\ \wedge\ \forall y(\forall z \in y\, H(z) \to H(y)) \to (\forall y \in x)\, H(y)]}$$

with $F[a]$ *a* Σ *formula and* $H(a)$ *any formula of* $\mathcal{L}^*$.

(ii) $\mathrm{FOUNDR}(\text{impl-}\Sigma)$ *is the rule of inference*

$$\frac{\exists! x\, F[x]}{\forall x[F[x]\ \wedge\ \forall y(\forall z \in y\, H(z) \to H(y)) \to (\forall y \in x)\, H(y)]}$$

with $F[a]$ *a* Σ *formula and* $H(a)$ *any formula of* $\mathcal{L}^*$.

(iii) $\mathrm{FOUND}(\text{impl-}\Sigma)$ *denotes the schema*

$$\exists! x\, F[x] \to \forall x[F[x]\ \wedge\ \forall y(\forall z \in y\, H(z) \to H(y)) \to (\forall y \in x)\, H(y)]$$

where $F[a]$ *is a* Σ *formula and* $H(a)$ *is any formula of* $\mathcal{L}^*$.

Remark 8.3. *The rule* $\mathrm{BR}(\text{impl-}\Sigma^1_2)$ *is, on the basis of* Δ^1_2-$\mathbf{CA}$*, much stronger than the rule* BR *whereas* $\mathrm{BR}(\text{impl-}\Sigma^1_2)$ *is still much weaker than* (BI). *The difference in strength between* (BI) *and* $\mathrm{BR}(\text{impl-}\Sigma^1_2)$ *is of course owed to the fact that the first is a rule while the second is a schema. But one can say something more illuminative about it. As it turns out,* $\mathrm{BR}(\text{impl-}\Sigma^1_2)$ *and* $\mathrm{BI}(\text{impl-}\Sigma^1_2)$ *are of the same strength (on the basis of* Δ^1_2-$\mathbf{CA}$*), in actuality the theories* Δ^1_2-$\mathbf{CA}+\mathrm{BR}(\text{impl-}\Sigma^1_2)$

and Δ^1_2-$\mathbf{CA}$ + $\mathrm{BI}(\text{impl-}\Sigma^1_2)$ *prove the same* Π^1_1 *statements. Thus the main difference between* $\mathrm{BR}(\text{impl-}\Sigma^1_2)$ *and* (BI) *is to be found in the premiss of* $\mathrm{BI}(\text{impl-}\Sigma^1_2)$ *requiring the well-ordering to be describable via a* Σ^1_2 *formula without parameters. Analogous remarks apply to the corresponding set-theoretic principles. The theme is explored in more detail in [46].*

The next lemma relates (in a weak sense) the $\mathcal{L}_2$ versions of Definition 8.1 to their set-theoretic counterparts.

Lemma 8.4. *(i)* $\Delta^1_2\text{-}\mathbf{CA} + \mathrm{BR}(\text{impl-}\Sigma^1_2) \subseteq \mathbf{KPl}^w + \mathrm{FOUNDR}(\text{impl-}\Sigma) = \mathbf{KPi}^r + \mathrm{FOUNDR}(\text{impl-}\Sigma)$.

(ii) $\Sigma^1_2\text{-}\mathbf{TRDC} + \mathrm{BR} \subseteq \mathbf{KPi}^w + (\Sigma\text{-REC}) + \mathrm{FOUNDR}(\text{impl-}\Sigma(\mathrm{M}))$.

(iii) $\Sigma^1_2\text{-}\mathbf{TRDC} + \mathrm{BR}(\text{impl-}\Sigma^1_2) \subseteq \mathbf{KPi}^w + (\Sigma\text{-REC}) + \mathrm{FOUNDR}(\text{impl-}\Sigma)$.

Proof. (i) The first identity is obvious since $\mathrm{FOUNDR}(\text{impl-}\Sigma)$ implies all instances of $(\mathrm{IND})^*$. Let $T := \Delta^1_2\text{-}\mathbf{CA} + \mathrm{BR}(\text{impl-}\Sigma^1_2)$ and $T' := \mathbf{KPi}^w + \mathrm{FOUNDR}(\text{impl-}\Sigma)$. We want to show

$$T \vdash A \;\Rightarrow\; T' \vdash A^*$$

by induction on the length of the derivation in T. Owing to Theorem 7.2 it suffices to assume that A is the consequence of an inference $\mathrm{BR}(\text{impl-}\Sigma^1_2)$. Then A is of the form

$$\forall X[\mathrm{WO}(X) \wedge F[X] \to \mathrm{TI}(X, H)]$$

with $F[U] \in \Sigma^1_2$. Moreover, inductively we have

$$T' \vdash (\exists! X(\mathrm{WO}(X) \wedge F[X]))^*.$$

We now argue in T'. By Theorem 7.1 there exists a Σ formula $F'[a]$ such that

$$\forall x \subseteq \mathrm{N}\,(F'[x] \leftrightarrow \mathrm{Wo}(\mathrm{N}, x) \wedge F[x]^*).$$

Let r be the unique well-ordering on N which satisfies $F'[r]$. Via Axiom Beta there exist a unique ordinal α and order isomorphism between r and α. As a result, α has an implicit Σ definition, so that with the help of $\mathrm{FOUNDR}(\text{impl-}\Sigma)$ we have transfinite induction on α for arbitrary formulae. Via the order isomorphism f we then obtain A^*.

The proof of (iii) is analogous to (i), using Lemma 7.3.

(ii) is also proved similarly. The only extra consideration one has to employ is the following. For a primitive recursive well-ordering $\prec$ we have $r := \{\langle i, j\rangle \mid i \prec$

$j\} \in \mathrm{M}$ and therefore the function f which is collapsing for r is an element of M, thus r is order isomorphic to an ordinal in M, which possesses an implicit $\Sigma(\mathrm{M})$ definition. □Below we shall list some results whose proofs are too long to be incorporated in the first part of this paper. They will be supplied in the second part.

Theorem 8.5. *(i)* $\mathbf{AUT\text{-}KPl}^r$*,* $\mathbf{KPi}^w + \mathrm{FOUND}(\text{impl-}\Sigma)$*, and* $\mathbf{KPi}^w + \mathrm{FOUNDR}(\text{impl-}\Sigma)$ *prove the same* Σ_1 *sentences.*

(ii) $\mathbf{KPi}^w + (\Sigma\text{-REC}) + \mathrm{FOUND}(\text{impl-}\Sigma)$ *and* $\mathbf{KPi}^w + (\Sigma\text{-REC}) + \mathrm{FOUNDR}(\text{impl-}\Sigma)$ *prove the same* Σ_1 *sentences.*

Proof. See [43], Satz 6.5. The proof (which is long) will be in incorporated in the second part of this paper. □

The following two results show that the strength of $(\Sigma\text{-FOUND})$ is already encapsulated in $(\Sigma\text{-REC})$.

Theorem 8.6. $\mathbf{KPi}^r + (\Sigma\text{-FOUND})$ *and* $\mathbf{KPi}^r + (\Sigma\text{-REC})$ *prove the same* Π_2 *sentences.*

Proof. See [43], Satz 7.1. The proof (which is long) will be in incorporated in the second part of this paper. □

Theorem 8.7. $\mathbf{KPi}^w + (\Sigma\text{-FOUND})$ *and* $\mathbf{KPi}^w + (\Sigma\text{-REC})$ *prove the same* Π_2 *sentences.*

Proof. See [43], Satz 7.20. The proof will be in the second part of this paper. □The next result shows

Theorem 8.8. $\mathbf{AUT\text{-}KPl}^r + \mathbf{KPi}^r$ *and* $\mathbf{AUT\text{-}KPl}^r$ *prove the same* Π_2 *sentences.*

Proof. See [43], Satz 7.20. The proof will be in the second part of this paper. □

Theorem 8.9. *For every* Π_2 *sentence* F*,*

$$\mathbf{KPi}^w + \mathrm{FOUND}(\text{impl-}\Sigma) \vdash F \;\Rightarrow\; \mathbf{AUT\text{-}KPl}^r \vdash F.$$

Proof. See [43], Satz 7.22. The proof will be in the second part of this paper. □

Theorem 8.10. *For every* Σ *sentence* G*,*

$$\mathbf{AUT\text{-}KPl}^r \vdash G \;\Rightarrow\; \mathbf{KPi}^w + \mathrm{FOUND}(\text{impl-}\Sigma) \vdash G.$$

Proof. See [43], Satz 7.23. The proof will be in the second part of this paper. □

II. WELL-ORDERING PROOFS

An ordinal α is said to be *provable* in a theory T (whose language encompasses $\mathcal{L}_2$) if there exists a recursive well-ordering $\prec$ whose order-type is α such that $\mathrm{T} \vdash \mathrm{WO}(\prec)$. In this chapter we try to give lower bounds for the provable ordinals of the various theories introduced in chapter I. That the results are indeed optimal will be shown in chapter III which will form the main chunk of the sequel to the present paper.

9 The functions φ_α and Φ_α

$\alpha, \beta, \gamma, \delta, \xi, \zeta, \rho$ will always denote ordinals. λ will be reserved for limit ordinals. Let $\alpha \mapsto \omega^\alpha$ be the ordinal function which enumerates the additive principal ordinals, i.e. the ordinals $\alpha > 0$ satisfying $(\forall \eta < \alpha)\, \eta + \alpha = \alpha$. This function is also a normal function since it is strictly increasing $\alpha < \beta \Rightarrow \omega^\alpha < \omega^\beta$ and satisfies $\omega^\lambda = \sup\{\omega^\eta \mid \eta < \lambda\}$.

Definition 9.1. *Inductive definition of the classes* $\mathrm{Cr}(\alpha)$*:*

1. $\mathrm{Cr}(0$ *is the class of additive principal ordinals.*
2. φ_α *is the function that enumerates* $\mathrm{Cr}(\alpha)$*, i.e.* $\varphi_\alpha(\xi)$ *is the* ξ*th member of* $\mathrm{Cr}(\alpha)$.
3. $\mathrm{Cr}(\alpha + 1) = \{\rho \mid \varphi_\alpha(\rho) = \rho\}$.
4. $\mathrm{Cr}(\lambda) = \bigcap\{\mathrm{Cr}(\xi) \mid \xi < \alpha\}$.

Definition 9.2. *Inductive definition of the classes* $\mathrm{Kr}(\alpha)$*:*

1. $\mathrm{Kr}(0)$ *is the class of uncountable cardinals.*
2. Φ_α *is the function that enumerates* $\mathrm{Kr}(\alpha)$.
3. $\mathrm{Kr}(\alpha + 1) = \{\rho \mid \Phi_\alpha(\rho) = \rho\}$.
4. $\mathrm{Kr}(\lambda) = \bigcap\{\mathrm{Kr}(\xi) \mid \xi < \alpha\}$.

On account of their definitions, the classes $\mathrm{Cr}(\alpha)$ and $\mathrm{Kr}(\alpha)$ are unbounded and closed in the ON (:=the class of ordinals) and thus every function φ_α and Φ_α is a normal function f, i.e. strictly increasing and continuous ($f(\lambda) = \sup\{f(\xi) \mid \xi < \lambda\}$ for limits λ).

In what follows we write $\varphi\alpha\beta$ for $\varphi_\alpha(\beta)$ and $\Phi\alpha\beta$ for $\Phi_\alpha(\beta)$.

The following three lemmas are proved for φ in [61, section 13], but the same proof works for Φ as well.

Lemma 9.3. *Let f be one of the functions φ or Φ. Suppose that $\alpha = f\gamma\delta$ and $\beta = f\xi\eta$.*

(i) $\alpha = \beta$ *holds if and only if one of the following three statements holds:*

1. $\gamma < \xi$ *and* $\delta = f\xi\eta$.
2. $\gamma = \xi$ *and* $\delta = \eta$.
3. $\xi < \gamma$ *and* $f\gamma\delta = \eta$.

(ii) $\alpha < \beta$ *holds if and only if one of the following three statements holds:*

1. $\gamma < \xi$ *and* $\delta < f\xi\eta$.
2. $\gamma = \xi$ *and* $\delta < \eta$.
3. $\xi < \gamma$ *and* $f\gamma\delta < \eta$.

Lemma 9.4. *(i)* $\varphi\alpha 0 < \varphi\beta 0 \Leftrightarrow \Phi\alpha 0 < \Phi\beta 0 \Leftrightarrow \alpha < \beta$.

(ii) $\alpha, \beta \leq \varphi\alpha\beta$ *and* $\alpha, \beta \leq \Phi\alpha\beta$.

Lemma 9.5. *For every* $\rho \in \mathrm{Cr}(0)$ *(*$\rho \in \mathrm{Kr}(0)$*) there exist unique ordinals* β, γ *such that* $\gamma < \rho$ *and* $\rho = \varphi\beta\gamma$ *(*$\rho = \Phi\beta\gamma$*).*

Definition 9.6. *(i)* $\alpha =_{nf} \varphi\beta\gamma \quad :\Leftrightarrow \quad \alpha = \varphi\beta\gamma$ *and* $\beta, \gamma < \alpha$.

(ii) $\alpha =_{nf} \Phi\beta\gamma \quad :\Leftrightarrow \quad \alpha = \Phi\beta\gamma$ *and* $\beta, \gamma < \alpha$.

(iii) $\alpha =_{nf} \alpha_1 + \ldots + \alpha_n \quad :\Leftrightarrow \quad \alpha = \alpha_1 + \ldots + \alpha_n,\ \alpha_1, \ldots, \alpha_n \in \mathrm{Cr}(0)$ *and* $\alpha > \alpha_1 \geq \ldots \geq \alpha_n$.

The normal forms of Definition 9.6 are unique representations of ordinals owing to Lemma 9.3.

Definition 9.7. *(i)* $\mathrm{SC} := \{\alpha \mid \varphi\alpha 0 = \alpha\}$.

(ii) $\Gamma_0^\Phi := \min\{\alpha \mid \Phi\alpha 0 = \alpha\}$.

Lemma 9.8. $\Gamma_0^\Phi = \sup\{\rho_n \mid n < \omega\}$ *where* $\rho_0 = \Phi 00$ *and* $\rho_{n+1} = \Phi\rho_n 0$.

Proof. As in [61, Theorem 14.16]. □

By $\mathfrak{R}$ we shall denote the class of uncountable regular cardinals. $\alpha \mapsto \Omega_\alpha$ is the mapping which enumerates the class $\mathfrak{R}_0 := \mathrm{Kr}(0) \cup \{0\}$. In more traditional notation we have $\Omega_\alpha = \aleph_\alpha$ for all $\alpha > 0$. The regular uncountable cardinals $< \Gamma_0^\Phi$ can be characterized as follows:

Theorem 9.9. *If* $\kappa \in \mathfrak{R}$ *and* $\kappa < \Gamma_0^\Phi$ *then there exists a unique* ξ *such that* $\kappa = \Omega_{\xi+1}$.

Proof. Let $\kappa \in \mathfrak{R}$ and $\kappa < \Gamma_0^\Phi$. By Lemma 9.4, $\kappa \leq \Phi\kappa 0 < \Phi(\kappa+1)0$. Hence there is largest ordinal β such that $\kappa \in \mathrm{Kr}(\beta)$. Thus $\kappa = \Phi\beta\delta$ for some $\delta < \kappa$. If δ were a limit we would have $\kappa = \sup\{\Phi\beta\xi \mid \xi < \delta\}$ and κ would be singular. As a result, $\kappa = \Phi\beta(\eta+1)$ for some η or $\kappa = \Phi\beta 0$.

If $\beta = \kappa = \Phi\beta 0$ one could show, by induction on n, utilizing Lemma 9.3(ii), that $\rho_n < \kappa$, contradicting $\kappa < \Gamma_0^\Phi$. Hence $\beta < \kappa$. Now one could show the cofinality of κ to be the same as that of β if $\kappa = \Phi\beta 0$ and β were a limit, making κ singular. Likewise, if $\beta = \zeta+1$ and $\kappa = \Phi\beta 0$ one could show that the cofinality of κ is ω, and similarly if $\kappa = \Phi\beta(\eta+1)$ and $\beta = \zeta+1$ the cofinality of κ would be ω, too. As a result, since κ is regular $> \omega$ we must have $\beta = 0$. Therefore $\kappa = \Omega_1$ or $\kappa = \Omega_{\xi+1}$, where $\xi = \eta+1+1$ if $\eta < \omega$ and $\xi = \eta$ otherwise. □

In what follows, the properties of the functions φ and Φ exhibited in this section will be used frequently and mostly tacitly.

10 The set of ordinals, $\mathrm{OT}(\Phi)$

This section introduces an ordinal representation system sufficient unto the task of expressing the proof-theoretic ordinals of all the theories considered so far. There will be no proofs in this section since they would be similar (with minor modifications) to those in [9] or [53]. **ZFC** will suffice as a background theory for showing the existence of the various functions.

We use the following conventions: (α,β), $(\alpha,\beta]$, $[\alpha,\beta)$, and $[\alpha,\beta]$ denote the intervals of ordinals between α and β in the obvious sense. For a set of ordinals A we use the abbreviations $A < \alpha \;:=\; (\forall\eta \in A)\,\eta < \alpha$ and $A \leq \alpha \;:=\; (\forall\eta \in A)\,\eta \leq \alpha$. Variables ν, μ, τ are understood to range over elements from $\mathfrak{R}_0$.

Definition 10.1. *By recursion on α we define the sets of ordinals ordinals $\mathrm{C}_\nu(\alpha)$ and the ordinals $\psi\nu\alpha$. The sets $\mathrm{C}_\nu(\alpha)$ themselves are defined inductively by the following clauses:*

($\mathrm{C}_\nu 1$) $[0,\nu] \subseteq \mathrm{C}_\nu(\alpha)$.

($\mathrm{C}_\nu 2$) $\xi,\eta \in \mathrm{C}_\nu(\alpha) \Rightarrow \xi+\eta \in \mathrm{C}_\nu(\alpha)$.

($\mathrm{C}_\nu 3$) $\xi,\eta \in \mathrm{C}_\nu(\alpha) \Rightarrow \varphi\xi\eta \in \mathrm{C}_\nu(\alpha)$.

($\mathrm{C}_\nu 4$) $\xi,\eta \in \mathrm{C}_\nu(\alpha) \Rightarrow \Phi\xi\eta \in \mathrm{C}_\nu(\alpha)$.

($\mathrm{C}_\nu 5$) $\xi < \alpha$ *and* $\xi,\mu \in \mathrm{C}_\nu(\alpha) \Rightarrow \psi\mu\xi \in \mathrm{C}_\nu(\alpha)$.

($\mathrm{C}_\nu 6$) $\psi\nu\alpha = \min\{\eta \mid \eta \notin \mathrm{C}_\nu(\alpha)\}$.

Definition 10.2. *(i)* $\alpha^+ := \min\{\kappa \in \mathfrak{R} \mid \alpha < \kappa\}$.

(ii) $\mathrm{S}(\alpha) := \min\{\mu \in \mathfrak{R}_0 \mid \alpha < \mu^+\}$.

Proposition 10.3. *(i)* $\alpha \leq \beta \Rightarrow \mathrm{C}_\nu(\alpha) \subseteq \mathrm{C}_\nu(\beta)$.

(ii) $\psi\nu\alpha \in (\nu, \nu^+)$.

(iii) $\nu < \Gamma_0^\Phi \Rightarrow \mathrm{C}_\nu(\alpha) \subseteq \Gamma_0^\Phi$.

(iv) $\psi\nu\alpha \in \mathrm{SC}$.

(v) $\psi\nu\alpha \notin \mathfrak{R}_0$.

(vi) $\psi\nu\alpha = \mathrm{C}_\nu(\alpha) \cap \nu^+$.

Proposition 10.4. *Let* $\alpha \in \mathrm{C}_\nu(\alpha)$ *and* $\beta \in \mathrm{C}_\nu(\beta)$.

(i) $\psi\nu\alpha = \psi\mu\beta$ *if and only if* $\nu = \mu$ *and* $\alpha = \beta$.

(i) $\psi\nu\alpha < \psi\mu\beta$ *if and only if* $\nu < \mu$ *or* $\nu = \mu \wedge \alpha < \beta$.

Definition 10.5. $\alpha =_{nf} \psi\nu\beta \;:\Leftrightarrow\; (\alpha = \psi\nu\beta \wedge \beta \in \mathrm{C}_\nu(\beta)$.

Definition 10.6. *The set of ordinals* $\mathrm{OT}(\Phi)$ *and the complexity* $\mathrm{G}\alpha < \omega$ *for* $\alpha \in \mathrm{OT}(\Phi)$ *are defined inductively by the following clauses:*

($\mathfrak{T}1$) $0 \in \mathrm{OT}(\Phi)$ *and* $\mathrm{G}(0) = 0$.

($\mathfrak{T}2$) $\alpha =_{nf} \alpha_1 + \ldots + \alpha_n \wedge \alpha_1, \ldots, \alpha_n \in \mathrm{OT}(\Phi) \Rightarrow$
$\alpha \in \mathrm{OT}(\Phi) \wedge \mathrm{G}\alpha = \max\{\mathrm{G}\alpha_1, \ldots, \mathrm{G}\alpha_n\} + 1$.

($\mathfrak{T}3$) $\alpha =_{nf} \varphi\beta\gamma \wedge \beta, \gamma \in \mathrm{OT}(\Phi) \Rightarrow \alpha \in \mathrm{OT}(\Phi) \wedge \mathrm{G}\alpha = \max\{\mathrm{G}\beta, \mathrm{G}\gamma\} + 1$.

($\mathfrak{T}4$) $\alpha =_{nf} \Phi\beta\gamma \wedge \beta, \gamma \in \mathrm{OT}(\Phi) \Rightarrow \alpha \in \mathrm{OT}(\Phi) \wedge \mathrm{G}\alpha = \max\{\mathrm{G}\beta, \mathrm{G}\gamma\} + 1$.

($\mathfrak{T}5$) $\alpha =_{nf} \psi\nu\gamma \wedge \nu, \gamma \in \mathrm{OT}(\Phi) \Rightarrow \alpha \in \mathrm{OT}(\Phi) \wedge \mathrm{G}\alpha = \max\{\mathrm{G}\nu, \mathrm{G}\gamma\} + 1$.

It follows from Lemma 9.5 and Proposition 10.3(iv),(v) that every ordinal $\alpha \in \mathrm{OT}(\Phi)$ enters $\mathrm{OT}(\Phi)$ owing to exactly one of the rules ($\mathfrak{T}1$)-($\mathfrak{T}5$). As a result the inductive definition of $\mathrm{OT}(\Phi)$ is deterministic, thus $\mathrm{G}\alpha$ is well-defined.

Theorem 10.7. $\mathrm{OT}(\Phi) = \mathrm{C}_0(\Gamma_0^\Phi)$.

Every element of $\mathrm{OT}(\Phi)$ can be uniquely named via a term built up from the "symbols" $0, +, \varphi, \Phi, \psi$. At this point we have not yet established that thereby $\mathrm{OT}(\Phi)$ with its ordering gives rise to a decidable well ordering. This can be achieved by showing that questions such as whether $\gamma < \Phi\beta\gamma$ in $(\mathfrak{T}4)$ and whether $\beta \in \mathrm{C}_\nu(\beta)$ in $(\mathfrak{T}5)$ can be decided. To this end we exhibit several lemmata which will entail the decidabilty of $(\mathrm{OT}(\Phi), <)$.

Definition 10.8. *The set of ordinals* $\mathrm{K}_\nu\alpha$ *for* $\alpha \in \mathrm{OT}(\Phi)$ *and* $\nu \in \mathfrak{R}_0$ *are defined inductively by the following clauses:*

$(\mathrm{K}_\nu 1)$ $\mathrm{K}_\nu 0 = \varnothing$.

$(\mathrm{K}_\nu 2)$ $\mathrm{K}_\nu\alpha = \bigcup\{\mathrm{K}_\nu\alpha_j \mid j = 1, \ldots, n\}$ *if* $\alpha =_{nf} \alpha_1 + \ldots + \alpha_n$.

$(\mathrm{K}_\nu 3)$ $\mathrm{K}_\nu\alpha = \mathrm{K}_\nu\beta \cup \mathrm{K}_\nu\gamma$ *if* $\alpha =_{nf} \varphi\beta\gamma$ *or* $\alpha =_{nf} \Phi\beta\gamma$.

$(\mathrm{K}_\nu 4)$ *Let* $\alpha =_{nf} \psi\mu\beta$.

$$\mathrm{K}_\nu\alpha = \begin{cases} \varnothing & \textit{if} \quad \mu < \nu \\ \{\beta\} \cup \mathrm{K}_\nu\beta \cup \mathrm{K}_\nu\mu & \textit{if} \quad \nu \leq \mu. \end{cases}$$

Lemma 10.9. *For* $\alpha \in \mathrm{OT}(\Phi)$ *we have* $\alpha \in \mathrm{C}_\nu(\beta) \Leftrightarrow \mathrm{K}_\nu\alpha < \beta$.

Definition 10.10. *Sets* $\mathrm{e}(\alpha)$ *and* $\mathrm{E}(\alpha)$ *are defined inductively as follows:*

1. $\mathrm{e}(0) = \mathrm{E}(0) = \varnothing$.
2. $\mathrm{e}(0) = \mathrm{E}(0) = \varnothing$ *if* $\alpha =_{nf} \alpha_1 + \ldots + \alpha_n$.
3. $\mathrm{e}(\alpha) = \{\beta\}$ *and* $\mathrm{E}(\alpha) = \varnothing$ *if* $\alpha =_{nf} \varphi\beta\gamma$.
4. $\mathrm{e}(\alpha) = \{\alpha\}$ *and* $\mathrm{E}(\alpha) = \{\beta\}$ *if* $\alpha =_{nf} \Phi\beta\gamma$.
5. $\mathrm{e}(\alpha) = \{\alpha\}$ *and* $\mathrm{E}(\alpha) = \varnothing$ *if* $\alpha =_{nf} \psi\nu\beta$.

Lemma 10.11. *Let* $\alpha, \beta, \gamma \in \mathrm{OT}(\Phi)$.

(i) If $\alpha = \varphi\beta\gamma$ *then* $\alpha =_{nf} \varphi\beta\gamma \Leftrightarrow [\mathrm{e}(\gamma) \leq \beta \wedge (\beta \notin \mathrm{SC} \vee \gamma > 0)]$.

(ii) If $\alpha = \Phi\beta\gamma$ *then* $\alpha =_{nf} \Phi\beta\gamma \Leftrightarrow \mathrm{E}(\gamma) \leq \beta$.

Proof. We only remark that it is essential for (ii) to hold that $\beta < \Phi\beta 0$ holds for all $\beta \in \mathrm{OT}(\Phi)$ by Theorem 10.7. □

Definition 10.12. *A coding function*

$$\ulcorner\ \urcorner : \mathrm{OT}(\Phi) \longrightarrow \mathbb{N}$$

is defined as follows: 1. $\ulcorner 0 \urcorner = (0)$. *2.* $\ulcorner \alpha \urcorner = (1, \ulcorner \alpha_1 \urcorner, \ldots, \ulcorner \alpha_n \urcorner)$ *if* $\alpha =_{nf} \alpha_1 + \ldots + \alpha_n$. *3.* $\ulcorner \alpha \urcorner = (2, \ulcorner \beta \urcorner, \ulcorner \gamma \urcorner)$ *if* $\alpha =_{nf} \varphi\beta\gamma$. *4.* $\ulcorner \alpha \urcorner = (3, \ulcorner \beta \urcorner, \ulcorner \gamma \urcorner)$ *if* $\alpha =_{nf} \Phi\beta\gamma$. *5.* $\ulcorner \alpha \urcorner = (4, \ulcorner \nu \urcorner, \ulcorner \gamma \urcorner)$ *if* $\alpha =_{nf} \psi\nu\gamma$. *Here* $(\ldots)$ *stands for some fixed primitive recursive coding of tuples of natural numbers.*

Let

$$\ulcorner \mathrm{OT}(\Phi) \urcorner := \{\ulcorner \alpha \urcorner \mid \alpha \in \mathrm{OT}(\Phi)\}$$

and define an ordering $\prec$ *on* $\mathbb{N}$ *via*

$$n \prec m \ :\Leftrightarrow\ \exists \alpha, \beta \in \mathrm{OT}(\Phi)\, (\alpha < \beta \wedge n = \ulcorner \alpha \urcorner \wedge m = \ulcorner \beta \urcorner).$$

If one now combines Lemma 9.3, Proposition 10.4, Lemma 10.9 and Lemma 10.11 one sees that $\ulcorner \mathrm{OT}(\Phi) \urcorner$ is a primitive recursive set equipped with a primitive recursive ordering $\prec$ such that $(\mathrm{OT}(\Phi), <)$ and $\ulcorner \mathrm{OT}(\Phi) \urcorner, \prec)$ are isomorphic.

In what follows we shall no longer distinguish between $(\mathrm{OT}(\Phi), <)$ and its arithmetization $\ulcorner \mathrm{OT}(\Phi) \urcorner, \prec)$. Via this identification, SC beomes a primitive recursive predicate and the functions $\mathrm{S}, \mathrm{K}, \mathrm{G}, \mathrm{e}, \mathrm{E}, \xi \mapsto \omega^\xi, \alpha \mapsto \Omega_\alpha, \varphi, \Phi, \psi$ can be viewed as primitive recursive functions acting on $\ulcorner \mathrm{OT}(\Phi) \urcorner$. In particular, all these relations and functions are definable in the language of arithmetic, $\mathcal{L}_1$.

Convention 10.13. *Lower case Greek letters* $\alpha, \beta, \gamma, \delta, \xi, \eta, \sigma, \zeta, \vartheta$ *will range over arbitrary elements of* $\mathrm{OT}(\Phi)$ *for the remainder of this paper while* ν, μ, τ *will be reserved for elements of* $\mathrm{OT}(\Phi) \cap \mathfrak{R}_0$. *Quantifiers* $\forall \alpha, \exists \alpha, \ldots$ *will exclusively range over elements of* $\mathrm{OT}(\Phi)$*, too.*

11 Distinguished sets

By a well-ordering proof in a given theory T we mean a proof formalizable in T which shows that a certain ordinal representation system (or a subset of it) is well-ordered. The notion of a *distinguished set* (of ordinals) (in German: ausgezeichnete Menge) will be central to carrying out well ordering proofs in the various subtheories of second order arithmetic introduced in earlier sections. A theory of distinguished sets developed for this purpose emerged in the works of Buchholz and Pohlers [4, 6, 7].

As a base theory in which all the results of this section can be proved one can take Π^1_1-$\mathbf{CA}_0$. It also worthwhile to point out that all the proofs work when the underlying logic is changed to intuitionistic logic. The principle of excluded third gets applied only to decidable properties (actually primitive recursive predicates). Thus all the proofs can be formalized in Π^1_1-$\mathbf{CA}^i_0$, the intuitionistic version of Π^1_1-$\mathbf{CA}_0$.

We introduce another operation on $\mathrm{OT}(\Phi)$ which will play an important role in the remainder of this paper.

Definition 11.1. *The* strongly critical subterms of level μ *of* α *are defined inductively as follows:*

1. $\mathrm{SC}_\mu(0) = \varnothing$.
2. $\mathrm{SC}_\mu(\alpha) = \{\alpha\}$ *if* $\alpha \in \mathrm{SC} \cap \mu^+$.
3. $\mathrm{SC}_\mu(\alpha) = \bigcup\{\mathrm{SC}_\mu(\alpha_i) \mid i = 1, \ldots, n\}$ *if if* $\alpha =_{nf} \alpha_1 + \ldots + \alpha_n$.
4. $\mathrm{SC}_\mu(\alpha) = \mathrm{SC}_\mu(\beta) \cup \mathrm{SC}_\mu(\gamma)$ *if* $\alpha =_{nf} \varphi\beta\gamma$.
5. $\mathrm{SC}_\mu(\alpha) = \mathrm{SC}_\mu(\beta) \cup \mathrm{SC}_\mu(\gamma)$ *if* $\alpha =_{nf} \Phi\beta\gamma$ *and* $\mu^+ \leq \alpha$.
6. $\mathrm{SC}_\mu(\alpha) = \mathrm{SC}_\mu(\beta) \cup \mathrm{SC}_\mu(\gamma)$ *if* $\alpha =_{nf} \psi\nu\gamma$ *and* $\mu^+ \leq \alpha$.

Definition 11.2. *Let* $U \subseteq \mathrm{OT}(\Phi)$ *and* $F(a)$ *be an* $\mathcal{L}_2$*-formula.*

(i) $U \cap \alpha := \{\eta \in U \mid \eta < \alpha\}$.

(ii) $U \cap \alpha \subseteq F \;:\Leftrightarrow\; (\forall \eta \in U \cap \alpha)\, F(\eta)$.

(iii) $\mathrm{Prg}(U, F) \;:\Leftrightarrow\; \forall \eta \in U\, [U \cap \eta \subseteq F \to F(\eta)]$.

(iv) $\mathrm{W}[U] := \{\eta \in U \mid \forall Y[\mathrm{Prg}(U, Y) \to U \cap \eta \subseteq Y]\}$.

(v) $\mathrm{M}^U_\mu := \{\eta < \mu^+ \mid (\forall \eta \in U \cap \mu)\mathrm{SC}_\nu(\eta) \subseteq U\}$.

(vi) $\mathrm{W}^U_\mu := \mathrm{W}[\mathrm{M}^U_\mu]$.

Remark 11.3. *(i) If* $<_U$ *denotes the restriction of* $<$ *to* U *and* $F_U(a)$ *is the formula* $a \in U \to F(a)$ *then* $\mathrm{Prg}(U, F) \leftrightarrow \mathrm{PROG}(<_U, F_U)$ *holds with* PROG *defined in Definition 3.1.*

(ii) M^U_μ *is always a set by arithmetical comprehension. To show that* $\mathrm{W}[U]$ *and* W^U_μ *are sets one can use* Π^1_1 *comprehension.* $\mathrm{W}[U]$ *and* W^U_μ *can also be shown to be sets in any theory which proves that the accessible part of an ordering* R *on* $\mathbb{N}$ *(where* R *is assumed to be a set) is a set. A case in point is constructive Zermelo-Fraenkel set theory with the regular extension axiom,* $\mathbf{CZF}$+REA *(see [1,2]). Actually the fragment* $\mathbf{CZF}^r$+REA *of* $\mathbf{CZF}$+REA

suffices. Here $\mathbf{CZF}^r$ *denotes* $\mathbf{CZF}$ *with* $\in$*-induction restricted to bounded formulae. To place this theory into perspective,* $\mathbf{CZF}^r + \mathrm{REA}$ *and* Π^1_1-$\mathbf{CA}_0$ *are of the same strength.*

The next Lemma lists basic properties of $\mathrm{W}[U]$, M^U_μ and W^U_μ.

Lemma 11.4. *(i)* $\mathrm{Prg}(U, S) \to \mathrm{W}[U] \subseteq S$.

(ii) $\mathrm{Prg}(U, \mathrm{W}[U])$.

(iii) $U \subseteq V \wedge \mathrm{Prg}(U, S) \to \mathrm{Prg}(V, \{\eta \mid \eta \in U \to \eta \in S\})$.

(iv) $\mathrm{Prg}(\mathrm{W}[U], S) \to \mathrm{W}[U] \subseteq S$.

(v) $\mathrm{W}[\mathrm{W}[U]] = \mathrm{W}[U]$.

(vi) $\mathrm{W}[U \cap \alpha] \subseteq \mathrm{W}[U]$.

(vii) $\mathrm{W}[U \cap \alpha] \subseteq \mathrm{W}[U]$.

(viii) $\alpha \in \mathrm{W}^U_\mu \leftrightarrow \alpha \in \mathrm{M}^U_\mu \wedge \mathrm{M}^U_\mu \cap \alpha \subseteq \mathrm{W}^U_\mu$.

Proof. (i) and (vii) are immediate by going back to the definitions.

(ii) Let $\alpha \in U$ and $U \cap \alpha \subseteq \mathrm{W}[U]$. By (i) we have $U \cap \alpha \subseteq S$ for every S satisfying $\mathrm{Prg}(U, S)$. Thence $\alpha \in \mathrm{W}[U]$.

(iii) Assume that $U \subseteq V$ and $\mathrm{Prg}(U, S)$ hold and also that $\alpha \in V$ and

$$V \cap \alpha \subseteq \{\eta \mid \eta \in U \to \eta \in S\}.$$

Then $U \cap \alpha = U \cap V \cap \alpha \subseteq S$, thus $\alpha \in U \to \alpha \in S$, i.e. $\alpha \in \{\eta \mid \eta \in U \to \eta \in S\}$.

(iv) Suppose $\mathrm{Prg}(W[U], S)$. (iii) implies $\mathrm{Prg}(U, \{\eta \mid \eta \in \mathrm{W}[U] \to \eta \in S\})$. Therefore, by (i), we also have $\mathrm{W}[U] \subseteq \{\eta \mid \eta \in \mathrm{W}[U] \to \eta \in S\}$, and hence $\mathrm{W}[U] \subseteq S$.

(v) $\mathrm{W}[\mathrm{W}[U]] \subseteq \mathrm{W}[U]$ holds by definition. Using (ii) we have $\mathrm{Prg}(\mathrm{W}[U], \mathrm{W}[\mathrm{W}[U]])$, hence, by (iv), $\mathrm{W}[U] \subseteq \mathrm{W}[\mathrm{W}[U]]$.

(vi) From

$$\eta \in U \cap \alpha \wedge \forall Y (\mathrm{Prg}(U \cap \alpha, Y) \to U \cap \alpha \cap \eta \subseteq Y)$$

we deduce that $\forall Y (\mathrm{Prg}(U, Y) \to U \cap \eta \subseteq Y)$, thence $\eta \in \mathrm{W}[U]$.

(viii) By (ii) we have $\mathrm{Prg}(\mathrm{M}^U_\mu, \mathrm{W}^U_\mu)$. W^U_μ is also a set. Thus (viii) follows. □

Definition 11.5. *(i) A set* $U \subseteq \mathrm{OT}(\Phi)$ *is said to be* distinguished *if (D1) and (D2) are satisfied:*

(D1) $(\forall\alpha \in U)\, \mathrm{S}\alpha \in U$.

(D2) $(\forall\mu \in U)\, U \cap \mu^+ = \mathrm{W}^U_\mu$.

(ii) We shall use the abbreviation $\mathrm{Ds}(U)$ *to convey that* U *is a distinguished set. Variables* P *and* Q *will always refer to distinguished sets.*

(iii) $\mathfrak{W} := \{\eta \mid \exists X\, [\mathrm{Ds}(X) \wedge \eta \in X]\}$.
Note that $\mathfrak{W}$ *cannot be shown to be a set in our background theory* Π^1_1-$\mathbf{CA}_0$ *(nor actually in any of the other theories we investigate in this paper).*

Lemma 11.6. *Recall that the letters* Q *and* P *are reserved for distinguished sets.*

(i) $Q \subseteq \mathrm{W}[Q]$ *and hence* $Q = \mathrm{W}[Q]$

(ii) $\mathrm{Prg}(Q, V) \to Q \subseteq V$.

Proof. (i) Let $\alpha \in Q$. Then $\mathrm{S}\alpha \in Q$ by (D1) and hence $Q \cap \alpha^+ = \mathrm{W}^Q_{\mathrm{S}\alpha}$. So by Lemma 11.4(v),(vi) we arrive at $\alpha \in Q \cap \alpha^+ = \mathrm{W}^Q_{\mathrm{S}\alpha} = \mathrm{W}[\mathrm{W}^Q_{\mathrm{S}\alpha}] = \mathrm{W}[Q \cap \alpha^+] \subseteq \mathrm{W}[Q]$.

(ii) is an immediate consequence of (i) and Lemma 11.4(i). □

Owing to Lemma 11.6(ii) we have transfinite induction over $<_Q := < \cap (Q \times Q)$ for arbitrary sets. Thus if we want to show that $Q \subseteq V$ holds for a set V it suffices to prove that

$$\forall\beta(\beta \in Q \wedge Q \cap \beta \subseteq V \to \beta \in V).$$

Specifically we have $\mathrm{WO}(<_Q)$.

Lemma 11.7. *(i)* $\nu \leq \mu \wedge \beta \in \mathrm{SC}_\mu(\alpha) \to \mathrm{SC}_\nu(\beta) \subseteq \mathrm{SC}_\nu(\alpha)$.

(ii) $\alpha \in Q \wedge \mu \in Q \to \mathrm{SC}_\mu(\alpha) \subseteq Q$.

(iii) $\mu \in \mathrm{M}^Q_\mu \to (\forall\nu \in Q)\mathrm{SC}_\mu(\alpha) \subseteq Q$.

(iv) $\mu \in \mathrm{M}^Q_\mu \wedge \mu \leq Q \to \mu \in Q$.

Proof. (i) follows by induction on $\mathrm{G}\alpha$

(ii) 1. Suppose $\mu < \mathrm{S}\alpha$. Then (D1) and (D2) imply that $\alpha \in Q \cap \alpha^+ = \mathrm{W}^Q_{\mathrm{S}\alpha} \subseteq \mathrm{M}^Q_{\mathrm{S}\alpha}$. As $\mu \in Q \cap \mathrm{S}\alpha$ we see that $\mathrm{SC}_\mu(\alpha) \subseteq Q$ by definition of $\mathrm{M}^Q_{\mathrm{S}\alpha}$.
2. Suppose $\mu \geq \mathrm{S}\alpha$. From (D2) it then follows that $\alpha \in \mathrm{W}^Q_\mu \subseteq \mathrm{M}^Q_\mu$. For $\nu \in Q \cap \mu$ we thus have $\mathrm{SC}_\nu(\alpha) \subseteq Q$, and from (i) we conclude that $(\forall\beta \in \mathrm{SC}_\mu(\alpha))\, \mathrm{SC}_\nu(\beta) \subseteq Q$. Therefore $\mathrm{SC}\mu(\alpha) \subseteq \{\alpha\} \cup \mathrm{M}^Q_\mu \cap \alpha$. By Lemma 11.4(viii) we get $\mathrm{SC}_\mu(\alpha) \subseteq \mathrm{W}^Q_\mu \subseteq Q$ as $\alpha \in \mathrm{W}^Q_\mu$.

(iii) will be proved by transfinite induction on Q (i.e. $<_Q$).
1. If $\nu \in Q \cap \mu^+$ then the desired assertion follows in the case $\nu < \mu$ from the

definition of M^Q_μ and in the case $\nu = \mu$ from (ii).
2. If $\nu \in Q$ and $\mu < \nu$ then by induction hypothesis we have $(\forall \tau \in Q \cap \nu)\mathrm{SC}_\tau(\mu) \subseteq Q$. and consequently $\mu \in \mathrm{M}^Q_\nu$. From $\nu \in Q \cap \nu^+ = \mathrm{W}^Q_\nu$ we obtain by Lemma 11.4(viii) that $\mathrm{M}^Q_\nu \cap \nu \subseteq \mathrm{W}^Q_\nu$, whence $\mu \in \mathrm{W}^Q_\nu$. Since $\mathrm{SC}_\nu(\mu) \subseteq \{\mu\}$ we arrive at the desired assertion.

(iv) follows directly from (iii). □

Lemma 11.8. $Q \cap \mu^+ \subseteq \mathrm{W}^Q_\mu$.

Proof. Let $\alpha \in Q \cap \mu^+$. Then $\alpha \in \mathrm{W}^Q_{\mathrm{S}\alpha}$ and so by Lemma 11.4(viii), $\mathrm{M}^Q_{\mathrm{S}\alpha} \cap \alpha \subseteq \mathrm{W}^Q_{\mathrm{S}\alpha}$. In view of Lemma 11.4(vi) it suffices to show that $\alpha \in \mathrm{W}[\mathrm{M}^Q_\mu \cap \alpha^+]$. Lemma 11.7(ii) yields $\alpha \in \mathrm{M}^Q_\mu \cap \alpha^+$. Using Lemma 11.4(iii), $\mathrm{Prg}(\mathrm{M}^Q_\mu \cap \alpha^+, U)$ implies

$$\mathrm{Prg}(\mathrm{M}^Q_{\mathrm{S}\alpha}, \{\eta \mid \eta \in \mathrm{M}^Q_\mu \cap \alpha^+ \to \eta \in U\}),$$

and further, by Lemma 11.4(i),

$$\mathrm{M}^Q_\mu \cap \alpha \subseteq \mathrm{M}^Q_{\mathrm{S}\alpha} \cap \alpha \subseteq \mathrm{W}^Q_{\mathrm{S}\alpha} \subseteq \{\eta \mid \eta \in \mathrm{M}^Q_\mu \cap \alpha^+ \to \eta \in U\},$$

thence $\mathrm{M}^Q_\mu \cap \alpha^+ \cap \alpha \subseteq U$. This shows $\alpha \in \mathrm{W}[\mathrm{M}^Q_\mu \cap \alpha^+]$. □

Proposition 11.9. $\mu \in \mathrm{M}^Q_\mu \wedge \mathrm{M}^Q_\mu \cap \mu \subseteq Q \to \mu \in \mathrm{W}^Q_\mu \wedge \mathrm{Ds}(\mathrm{W}^Q_\mu)$.

Proof. By Lemma 11.8, $\mathrm{M}^Q_\mu \cap \mu \subseteq Q$ implies $\mathrm{M}^Q_\mu \cap \mu = \mathrm{W}^Q_\mu \cap \mu$. Thus, by Lemma 11.4(viii), $\mu \in \mathrm{M}^Q_\mu$ implies $\mu \in \mathrm{W}^Q_\mu$.

Next we show that W^Q_μ is a distinguished set.

Ad (D1): If $\alpha \in \mathrm{W}^Q_\mu \cap \mu$ then $\mathrm{S}\alpha \in Q \cap \mu \subseteq \mathrm{W}^Q_\mu \cap \mu$. From $\alpha \in \mathrm{W}^Q_\mu$ and $\mu \leq \alpha$ we obtain $\mathrm{S}\alpha = \mu \in \mathrm{W}^Q_\mu$.

Ad (D2): For $\tau \leq \mu$ we have $(*)\quad \mathrm{W}^Q_\mu \cap \tau = Q \cap \tau$ since $\mathrm{M}^Q_\mu \cap \mu \subseteq Q$ yields $\mathrm{W}^Q_\mu \cap \tau \subseteq Q$, and so, by Lemma 11.8, $Q \cap \tau \subseteq \mathrm{W}^Q_\mu$ holds. Now let $P := \mathrm{W}^Q_\mu$ and suppose $\nu \in P$. By $(*)$, we then have $P \cap \nu = Q \cap \nu$, and thus, by Lemma 11.4(viii), $(**)\quad \mathrm{W}^P_\nu = \mathrm{W}^Q_\nu$. For $\nu < \mu$, $(*)$ entails $\nu \in Q$ and therefore $\mathrm{W}^P_\nu = \mathrm{W}^Q_\nu = Q \cap \nu^+ \overset{(*)}{=} \mathrm{W}^Q_\mu \cap \nu^+ = P \cap \nu^+$. If $\nu = \mu$, then $(**)$ yields $\mathrm{W}^Q_\mu = P = P \cap \mu^+$. □

Vacuously $\varnothing$ is a distinguished set. Proposition 11.9 yields the existence of non-trivial distinguished sets. For example, $\mathrm{W}^\varnothing_0$ is a distinguished set.

Lemma 11.10. $\mathrm{Prg}(P \cup Q, U) \to P \cup Q \subseteq U$.

Proof. Suppose $\mathrm{Prg}(P \cup Q, U)$. Then we have

$$P \cap \alpha \subseteq U \to \mathrm{Prg}(Q, \{\eta \mid \eta < \alpha \to \eta \in U\}), \text{ and}$$

$$P \cap \alpha \subseteq U \ \wedge \ Q \cap \alpha \subseteq U \ \wedge \ \alpha \in P \to \alpha \in U.$$

Therefore, by Lemma 11.6(ii), we have

$$P \cap \alpha \subseteq U \ \wedge \ \alpha \in P \to \alpha \in U,$$

i.e. $\mathrm{Prg}(P, U)$ holds, and consequently $P \subseteq U$ by Lemma 11.6(ii). Similarly one shows that $Q \subseteq U$. □

Lemma 11.11. $\mu \in P \cup Q \ \wedge \ \mu \leq P \ \wedge \ \mu \leq Q \to P \cap \mu^+ = Q \cap \mu^+$.

Proof. We use induction on $P \cup Q$, i.e. Lemma 11.10. Let $\mu \in P$ and suppose $\mu \leq Q$. The induction hypothesis yields $P \cap \mu = Q \cap \mu$ and, by Lemma 11.4(vii), we conclude that $\mu \in P \cap \mu^+ = \mathrm{W}_\mu^P = \mathrm{W}_\mu^Q \subseteq \mathrm{M}_\mu^Q$, and hence $\mu \in Q$ by Lemma 11.7(iv). As a result, $P \cap \mu^+ = \mathrm{W}_\mu^P = \mathrm{W}_\mu^Q = Q \cap \mu^+$. The same arguments can be used if $\mu \in Q$ and $\mu \leq P$. □

Proposition 11.12. $\alpha \in Q \to Q \cap \alpha^+ = \mathfrak{W} \cap \alpha^+$.

Proof. Let $\alpha \in Q$. $Q \cap \alpha^+ \subseteq \mathfrak{W} \cap \alpha^+$ is obvious by definition of $\mathfrak{W}$. Let $\eta \in \mathfrak{W} \cap \alpha^+$. Then there exists a distinguished set P such that $\eta \in P \cap \alpha^+$. Thus $\mathrm{S}\eta \in P \cup Q$, $\mathrm{S}\eta \leq \eta \in P$ and $\mathrm{S}\eta \leq \alpha \in Q$. Therefore $\eta \in P \cap \eta^+ = Q \cap \eta^+ \subseteq Q \cap \alpha^+$ using Lemma 11.11. □

Next we study closure properties shared by all distinguished sets.

Proposition 11.13. *(i)* $\alpha, \beta \in Q \to \alpha + \beta \in Q$.

(ii) $\alpha, \beta \in \mathfrak{W} \to \alpha + \beta \in \mathfrak{W}$.

Proof. (ii) is an immediate consequence of (i) in view of Proposition 11.12. In the proof of (i) let $X := \mathrm{M}_{\mathrm{S}\alpha}^Q$, $Y := \mathrm{W}_{\mathrm{S}\alpha}^Q$ and $U := \{\xi \mid \alpha + \xi \in Y\}$. Suppose $\alpha, \beta \in Q$. If $\mathrm{S}\alpha < \mathrm{S}\beta$ then $\alpha + \beta = \beta \in Q$. Now assume $\mathrm{S}\beta \leq \mathrm{S}\alpha$. Then we have $Q \cap \alpha^+ = Y$ and $\alpha, \beta \in Y$. Moreover we have

$$\eta \in X \ \wedge \ X \cap \eta \subseteq U \to \alpha + \eta \in X \ \wedge \ X \cap (\alpha + \eta) \subseteq Y,$$

so that with Lemma 11.4(viii) we get $\eta \in X \ \wedge \ X \cap \eta \subseteq U \to \alpha + \eta \in Y$. As a result, $\mathrm{Prg}(X, U)$ holds, and thus $Y \subseteq U$ by Lemma 11.4(i), hence $\alpha + \beta \in Y \subseteq Q$. □

Lemma 11.14. *Letting* $\mathfrak{F}(\alpha, \beta)$ *be the formula*

$$\alpha, \beta \in Q \ \wedge \ (\forall \xi \in Q \cap \alpha)(\forall \eta \in Q)(\varphi \xi \eta \in Q) \ \wedge \ (\forall \eta \in Q \cap \beta)(\varphi \alpha \eta \in Q),$$

the following are true:

(i) $\mathfrak{F}(\alpha,\beta) \wedge \mu = \max\{\mathrm{S}\alpha, \mathrm{S}\beta\} \wedge \gamma \in \mathrm{M}^Q_\mu \cap \varphi\alpha\beta \to \gamma \in Q$.

(ii) $\mathfrak{F}(\alpha,\beta) \to \varphi\alpha\beta \in Q$.

Proof. We show (i) by induction on $\mathrm{G}\gamma$. $\mathfrak{F}(\alpha,\beta)$ implies $\alpha,\beta \in Q \cap \mu^+ = \mathrm{W}^Q_\mu$. We distinguish cases according to the shape of γ. The assertion is trivially true if $\gamma = 0$. Let $\gamma =_{nf} \gamma_1 + \ldots + \gamma_n$. Then $\gamma_1, \ldots, \gamma_n \in \mathrm{M}^Q_\mu \cap \varphi\alpha\beta$, and thus by the induction hypothesis, $\gamma_1, \ldots, \gamma_n \in Q$, so $\gamma \in Q$ by Proposition 11.13. If $\gamma \in \mathrm{SC}$ then $\gamma \leq \alpha \vee \gamma \leq \beta$, and therefore, as $\alpha, \beta \in \mathrm{W}^Q_\mu$ and $\gamma \in \mathrm{M}^Q_\mu$, it follows from Lemma 11.4(viii) that $\gamma \in \mathrm{W}^Q_\mu \subseteq Q$.

The last case to consider is when $\gamma =_{nf} \varphi\xi\eta$ for some ξ, η. Then $\xi, \eta \in \mathrm{M}^Q_\mu \cap \varphi\alpha\beta$ and the induction hypothesis yields $\xi, \eta \in Q$. If $\xi \leq \alpha$ then $\gamma \in Q$ follows from $\mathfrak{F}(\alpha,\beta)$. If $\alpha < \xi$ then $\gamma < \beta$ must hold, and with the aid of Lemma 11.4(viii) we conclude that $\gamma \in Q$.

(ii) By (i) we have

$$\mathfrak{F}(\alpha,\beta) \wedge \mu = \{\mathrm{S}\alpha, \mathrm{S}\beta\} \to \mathrm{M}^Q_\mu \cap \varphi\alpha\beta \subseteq \mathrm{W}^Q_\mu.$$

By Lemma 11.7(ii) we also have

$$\mathfrak{F}(\alpha,\beta) \wedge \mu = \max\{\mathrm{S}\alpha, \mathrm{S}\beta\} \to \varphi\alpha\beta \in \mathrm{M}^Q_\mu.$$

Thus, by Lemma 11.4(viii),

$$\mathfrak{F}(\alpha,\beta) \wedge \mu = \max\{\mathrm{S}\alpha, \mathrm{S}\beta\} \to \varphi\alpha\beta \in \mathrm{W}^Q_\mu,$$

and hence $\mathfrak{F}(\alpha,\beta) \to \varphi\alpha\beta \in Q$. □

Proposition 11.15. *(i)* $\alpha, \beta \in Q \to \varphi\alpha\beta \in Q$.

(ii) $\alpha, \beta \in \mathfrak{W} \to \varphi\alpha\beta \in \mathfrak{W}$.

Proof. Again, by Proposition 11.12, (ii) is an immediate consequence of (i). Let $\alpha \in Q$, $U := \{\xi \mid (\forall\eta \in Q)(\varphi\xi\eta \in Q)\}$ and $V := \{\eta \mid \varphi\alpha\eta \in Q\}$. Lemma 11.14(ii) yields

$$(\forall\xi \in Q \cap \alpha)(\forall\eta \in Q)(\varphi\xi\eta \in Q) \to \mathrm{Prg}(Q, V)$$

and hence, using Lemma 11.6(ii),

$$(\forall\xi \in Q \cap \alpha)(\forall\eta \in Q)(\varphi\xi\eta \in Q) \to Q \subseteq V.$$

The latter implies $\mathrm{Prg}(Q, U)$, whence $Q \subseteq U$. □

Corollary 11.16. *(i)* $\mathrm{S}\alpha \leq \mu \wedge \mu \in Q \wedge \mathrm{SC}_{\mu}(\alpha) \subseteq Q \to \alpha \in Q$.

(ii) $\mathrm{S}\alpha \leq \mu \wedge \mu \in \mathfrak{W} \wedge \mathrm{SC}_{\mu}(\alpha) \subseteq Q \to \alpha \in \mathfrak{W}$.

Proof. This follows from Propositions 11.13 and 11.15. □

Lemma 11.17. *(i)* $\beta \in Q \wedge \alpha \in \mathrm{M}^{Q}_{\mathrm{S}\beta} \cap \beta \to \alpha \in Q$.

(ii) $\beta \in \mathfrak{W} \wedge \alpha \in \mathrm{M}^{Q}_{\mathrm{S}\beta} \cap \beta \to \alpha \in \mathfrak{W}$.

(i) $\beta \in Q$ implies $\beta \in Q \cap \beta^{+} = \mathrm{W}^{Q}_{\mathrm{S}\beta}$. Therefore, by Lemma 11.4(viii), $\alpha \in \mathrm{W}^{Q}_{\mathrm{S}\beta} \subseteq Q$. (ii) is an immediate consequence of (i). □

Definition 11.18. $\mathfrak{B}^{Q}_{\mu} := \{\alpha \mid (\forall \tau \in Q \cap \mu)[\mathrm{K}_{\tau}\alpha < \alpha \to \psi\tau\alpha \in Q\}$.

Lemma 11.19. *Assume* $\alpha \in \mathrm{M}^{Q}_{\mu}$, $\mathrm{M}^{Q}_{\mu} \cap \alpha \subseteq \mathfrak{B}^{Q}_{\mu}$, $\nu \in Q \cap \mu$, $\mathrm{K}_{\nu}\alpha < \alpha$ *and* $\gamma \in \mathrm{M}^{Q}_{\nu} \cap \psi\nu\alpha$. *Then* $\gamma \in Q$.

Proof. We proceed by induction on $\mathrm{G}\gamma$.
If $\gamma \leq \nu$ then $\gamma \in Q$ by Lemma 11.17(i). Now let $\nu < \gamma$.
1. $\gamma =_{nf} \gamma_1 + \ldots + \gamma_n$ By the induction hypothesis we get $\gamma_1, \ldots, \gamma_n \in Q$ and hence $\gamma \in Q$ by Lemma 11.13.
2. $\gamma =_{nf} \varphi\xi\eta$. By the induction hypothesis we get $\xi, \eta \in Q$ and hence $\gamma \in Q$ by Lemma 11.15.
3. $\gamma =_{nf} \Phi\xi\eta$. Then we would have $\gamma \leq \nu$ since $\gamma < \nu^{+}$, but this we ruled out. So this case cannot occur.
4. $\gamma =_{nf} \psi\nu\eta$. Then $\eta < \alpha$. By Lemma 11.7(i), $\gamma \in \mathrm{M}^{Q}_{\nu}$ entails that

$$(\forall \tau \in Q \cap \nu)(\forall \beta \in \mathrm{SC}_{\nu}(\eta))\, \mathrm{SC}_{\tau}(\beta) \subseteq \mathrm{SC}_{\tau}(\eta) \subseteq Q.$$

Since $\mathrm{SC}_{\nu}(\eta) < \psi\nu\eta < \psi\nu\alpha$, the latter entails that $\mathrm{SC}\nu(\eta) \subseteq \mathrm{M}^{Q}_{\nu} \cap \psi\nu\alpha$, and therefore, by the induction hypothesis, $\mathrm{SC}_{\nu}(\eta) \subseteq Q$. As a result we have shown that

$$(\forall \tau \leq \nu)[\tau \in Q \cap \mu \to \mathrm{SC}_{\tau}(\eta) \subseteq Q].$$

Via a subsidiary induction on Q we shall show that

$$(\forall \tau \in Q \cap \mu)\, \mathrm{SC}_{\tau}(\eta) \subseteq Q.$$

Let $\tau \in Q \cap \mu$. In view of (11.1) we may assume that $\nu < \tau$. The subsidiary induction hypothesis yields $(\forall \tau' \in Q \cap \tau)\, \mathrm{SC}_{\tau'}(\eta) \subseteq Q$, which implies $\mathrm{SC}_{\tau}(\eta) \subseteq \mathrm{M}^{Q}_{\tau}$. Since $\nu < \tau$, $\mathrm{K}_{\nu}\eta < \eta$ and $\mathrm{K}_{\nu}\alpha < \alpha$ hold, we conclude that $\mathrm{K}_{\tau}\eta < \eta$,

$\mathrm{K}_\tau\alpha < \alpha$ and $\mathrm{SC}_\tau(\eta) < \psi\tau\eta < \psi\tau\alpha$. Therefore we have $\mathrm{SC}_\tau(\eta) \subseteq \mathrm{M}_\tau^Q \cap \psi\tau\alpha$ and consequently, by applying the main induction hypothesis, $\mathrm{SC}_\tau(\eta) \subseteq Q$. This completes the proof of (11.1).

Finally, from (11.1) we conclude that $\eta \in \mathrm{M}_\mu^Q \cap \alpha \subseteq \mathfrak{B}_\mu^Q$, yielding $\gamma = \psi\nu\eta \in Q$. □

Lemma 11.20. $\mathrm{Prg}(\mathrm{M}_\mu^Q, \mathfrak{B}_\mu^Q)$.

Proof. Let $\alpha \in \mathrm{M}_\mu^Q$ and $\mathrm{M}_\mu^Q \cap \alpha \subseteq \mathfrak{B}_\mu^Q$. We have to show $\alpha \in \mathfrak{B}_\mu^Q$. So suppose $\nu \in Q \cap \mu$ and $\mathrm{K}_\nu\alpha < \alpha$. By Lemma 11.19 we have $\mathrm{M}_\nu^Q \cap \psi\nu\alpha \subseteq \mathrm{W}_\nu^Q$. For $\tau \in Q \cap \nu$ it holds $\mathrm{SC}_\tau(\psi\nu\alpha) = \mathrm{SC}_\tau(\nu) \cup \mathrm{SC}_\tau(\alpha)$ and therefore, using Lemma 11.7(ii), $\mathrm{SC}_\tau(\psi\nu\alpha) \subseteq Q$ since $\nu \in Q$ and $\alpha \in \mathrm{M}_\mu^Q$. Thus $\psi\nu\alpha \in \mathrm{M}_\nu^Q$, so that by Lemma 11.4(viii) we have $\psi\nu\alpha \in \mathrm{W}_\nu^Q \subseteq Q$. This shows $\alpha \in \mathfrak{B}_\mu^Q$. □

Lemma 11.21. *(i)* $\alpha, \nu \in Q \wedge \mathrm{K}_\nu\alpha < \alpha \to \psi\nu\alpha \in Q$.

(ii) $\alpha, \nu \in \mathfrak{W} \wedge \mathrm{K}_\nu\alpha < \alpha \to \psi\nu\alpha \in \mathfrak{W}$.

Proof. (ii) is a consequence of (i). For (i), let $\tau := \max\{\mathrm{S}\alpha, \mathrm{S}\nu\}$ and $\mu := \tau^+$. By Lemmata 11.20 and 11.4(i), we have $\mathrm{W}_\mu^Q \subseteq \mathfrak{B}_\mu^Q$. Therefore, since $\tau \in Q$, we have $Q \cap \mu \subseteq \mathfrak{B}_\mu^Q$, and hence $\psi\nu\alpha \in Q$. □

Lemma 11.22. $(\forall j \in U)\mathrm{Ds}(Q_j) \to \mathrm{Ds}(\bigcup\{Q_j \mid j \in U\})$.

Proof. Suppose $\mathrm{Ds}(Q_j)$ holds for all $j \in U$. Using arithmetical comprehension,

$$Z := \bigcup\{Q_j \mid j \in U\}$$

is a set. If $\alpha \in Z$ there exists $j \in U$ such that $\alpha \in Q_j$, thus $\mathrm{S}\alpha \in Q_j \subseteq Z$, showing that Z satisfies (D1). To verify (D2), suppose $\mu \in Z$. Then $\mu \in Q_i$ for some $i \in U$. Owing to Proposition 11.12 it follows that

$$\mathfrak{W} \cap \mu^+ = Q_i \cap \mu^+ \subseteq Z \cap \mu^+ \subseteq \mathfrak{W} \cap \mu^+,$$

and thus $Q_i \cap \mu^+ = Z \cap \mu^+$. By applying Lemma 11.4(vii), we see that $\mathrm{W}_\mu^Z = \mathrm{W}_\mu^{Q_i} = Q_i \cap \mu^+ = Z \cap \mu^+$. □

12 Well-ordering proofs in $\Pi_1^1\text{-TR}_0$, $\Pi_1^1\text{-TR} + \Delta_2^1\text{-CA}$ and $\Delta_2^1\text{-CA} + \mathrm{BR}(\text{impl-}\Sigma_2^1)$.

Lemma 12.1. $\nu < \mathrm{S}\alpha \to \mathrm{SC}_\nu(\mathrm{S}\alpha) \subseteq \mathrm{SC}_\nu(\alpha)$.

Proof. We use induction on $\mathrm{G}\alpha$.
1. If $\alpha =_{nf} \alpha_1 + \ldots + \alpha_n$ or $\alpha =_{nf} \varphi\xi\beta$ the assertion follows immediately from the induction hypothesis.
2. $\alpha =_{nf} \psi\mu\beta$. Then $\mathrm{S}\alpha = \mu$ and $\mathrm{SC}_\nu(\mu) \subseteq \mathrm{SC}_\nu(\alpha)$.
3. $\alpha =_{nf} \Phi\xi\beta$. Then $\mathrm{S}\alpha = \alpha$. □

Proposition 12.2. Π^1_1-$\mathbf{TR}_0 \vdash \forall\alpha(\alpha \in \mathfrak{W} \to \Omega_\alpha \in \mathfrak{W})$.

Proof. We argue informally in Π^1_1-$\mathbf{TR}_0$. Let $\alpha \in \mathfrak{W}$. Then there exists a distinguished set Q such that $\alpha \in Q$. By Lemma 11.6(ii), $<\restriction Q$ is a well-ordering, thus, using (Π^1_1-TR), there exists a set X such that for all $\beta \in Q$,

$$X_\beta = \mathrm{W}^{X_{Q\beta}}_{\Omega_\beta} \cup Q \;\text{ where }\; X_{Q\beta} := \bigcup\{X_\eta \mid \eta \in Q \cap \beta\} \;\text{ and }\; X_\eta := \{z \mid \langle\eta, z\rangle \in X\}.$$

We now show by induction on Q that for all $\beta \in Q$,

$$\Omega_\beta \in X_\beta \;\wedge\; \mathrm{Ds}(X_\beta) \;\wedge\; X_{Q\beta} \subseteq X_\beta.$$

Let $\beta \in Q$. The induction hypothesis, in conjunction with Lemma 11.22, yields

$$\mathrm{Ds}(X_{Q\beta}) \;\wedge\; (\forall\xi \in Q \cap \beta)\;(\Omega_\xi \in X_{Q\beta}).$$

As $0 \in \mathrm{W}^{\varnothing}_0 \subseteq Q$, we have $\Omega_0 = 0 \in X_0$ and hence (12.1) holds when $\beta = 0$. Now let $0 < \beta$. If $\nu \in X_{Q\beta} \cap \Omega_\beta$ we can use Lemma 11.7(ii) to conclude that $\mathrm{SC}_\nu(\Omega_\beta) = \mathrm{SC}_\nu(\beta) \subseteq X_{Q\beta}$ since $\beta \in Q \subseteq X_{Q\beta}$. This shows

$$\Omega_\beta \in \mathrm{M}^{X_{Q\beta}}_{\Omega_\beta}.$$

Now let $\delta \in \mathrm{M}^{X_{Q\beta}}_{\Omega_\beta} \cap \Omega_\beta$ and $\mathrm{S}\delta = \Omega_\sigma$. We want to show $\delta \in X_{Q\beta}$. We may assume that $\beta < \Omega_\beta$ since otherwise we have $\beta = \Omega_\beta$ and thus $\mathrm{M}^{X_{Q\beta}}_{\Omega_\beta} \cap \Omega_\beta = \mathrm{M}^Q_\beta \cap \beta \subseteq Q \subseteq X_{Q\beta}$ using Lemmata 11.11, 11.4(vii) and 11.17(i).

Case 1: $\mathrm{S}\delta \leq \mathrm{S}\beta$ or there exists $\xi \in Q \cap \beta$ such that $\mathrm{S}\delta \leq \Omega_\xi$. Then, by Corollary 11.16, we obtain $\delta \in X_{Q\beta}$.

Case 2: $(\forall\xi \in Q \cap \beta)(\Omega_\xi < \Omega_\sigma)$ and $\mathrm{S}\beta < \mathrm{S}\delta = \Omega_\sigma$. In this case we have $\mathrm{S}\beta \in X_{Q\beta} \cap \Omega_\beta$, thus, using Lemma 12.1, we arrive at

$$\mathrm{SC}_{\mathrm{S}\beta}(\sigma) = \mathrm{SC}_{\mathrm{S}\beta}(\Omega_\sigma) \subseteq \mathrm{SC}_{\mathrm{S}\beta}(\delta) \subseteq X_{Q\beta},$$

and hence $\mathrm{SC}_{\mathrm{S}\beta}(\sigma) \subseteq X_{Q\beta} \cap (\mathrm{S}\beta)^+$. An application of Lemma 11.11 yields $\mathrm{SC}_{\mathrm{S}\beta}(\sigma) \subseteq Q$, and since $\sigma < \beta$ and $\mathrm{S}\sigma \leq \mathrm{S}\beta$ we conclude that $\sigma \in Q \cap \beta$ by employing Lemma 11.16. However, this is an impossibility since we assumed that $(\forall\xi \in Q \cap \beta)(\Omega_\xi < \Omega_\sigma)$. Thus Case 2 is ruled out.

In sum, we have shown that

$$\mathrm{M}_{\Omega_\beta}^{X_{Q\beta}} \subseteq X_{Q\beta}.$$

In view of the Lemmata 11.9 and 11.22 we can deduce $\Omega_\beta \in X_\beta \wedge \mathrm{Ds}(X_\beta)$ from (12.1 and (12.1). Moreover, by Lemma 11.22, we have $X_{Q\beta} \cap \Omega_\beta \subseteq \mathrm{W}_{\Omega_\beta}^{X_{Q\beta}}$, and hence

$$X_{Q\beta} = (X_{Q\beta} \cap \Omega_\beta) \cup Q \subseteq X_\beta.$$

This completes the proof of (12.1). Letting $Z := \bigcup\{X_\beta \mid \beta \in Q\}$, we can use Lemma 11.22 and (12.1) to conclude that $\mathrm{Ds}(Z)$ and $(\forall\beta \in Q)\,(\Omega_\beta \in Z)$, hence $\Omega_\alpha \in \mathfrak{W}$. □

Corollary 12.3. *Let $\mathfrak{E}[U, \beta, \gamma, Q]$ be the Π^1_1 formula $\gamma \in Q \vee \gamma \in \mathrm{W}^U_{\Omega_\beta}$. Put $\Xi_0 := 1$ and $\Xi_{n+1} := \Omega_{\Xi_n}$. Let $\mathbf{T}$ be the theory Π^1_1-$\mathbf{CA}_0$ plus the additional rule*

$$\frac{\exists! Q\,(F[Q] \wedge \mathrm{Ds}(Q))}{\forall P\,(F[P] \wedge \mathrm{Ds}(P) \to \exists X (\forall\beta \in P)\forall\gamma(\gamma \in X_\beta \leftrightarrow \mathfrak{E}[X_{P\beta}, \gamma, P]))}$$

with the proviso that $F[Q]$ is an arithmetical formula.

For all n we then have

$$\mathbf{T} \vdash \Xi_n \in \mathfrak{W}.$$

Proof. We proceed by metainduction on n. For $n = 0$ this obvious. Let $n = m+1$. By the the induction hypothesis, we have $\mathbf{T} \vdash \Xi_m \in \mathfrak{W}$. Let $\mu := \Xi_m$. Arguing in $\mathbf{T}$, there exists a distinguished set Q such that $\mu \in Q$ and $Q = Q \cap \mu^+$. Owing to Lemma 11.11, Q is uniquely determined via this description. Thus $\exists! P\, F[P]$, where $F[P] := (\mu \in P \wedge P \cap \mu^+ = P)$. Since μ can be described via an arithmetical formula, too, we can use the above rule to infer that there exists a set X such that $(\forall\beta \in Q)\forall\gamma(\gamma \in Z_\beta \leftrightarrow \mathfrak{E}[X_{Q\beta}, \gamma, Q])$. Inspection of the proof of Proposition 12.2 shows that the existence of X is what is needed to conclude that $\Omega_\mu \in \mathfrak{W}$, i.e. $\Xi_n \in \mathfrak{W}$. □

Corollary 12.4. *For all n, Δ^1_2-$\mathbf{CA} + \mathrm{BR}(\text{impl-}\Sigma^1_2) \vdash \Xi_n \in \mathfrak{W}$.*

Proof. As a corollary of the proof of Theorem 6.14 one has that the theorems of Δ^1_2-$\mathbf{CA} + \mathrm{BR}(\text{impl-}\Sigma^1_2)$ are closed under the inference rule (12.1). Thus, by Corollary 12.3, the claim is true. □

Lemma 12.5. *Let $\mathbf{T}^*$ be the theory $\mathbf{KPl}^r$ augmented by the rule*

$$\frac{\exists! \alpha\, A[\alpha]}{\forall\beta\forall x\,(A[\beta] \to \exists f\, \mathfrak{D}_0[x, \beta, f])}$$

for every Σ *formula* $A[\beta]$ *and* $\mathfrak{D}_0[x, \beta, f]$ *be defined as in Lemma 5.3.*
With $\mathbf{T}$ *being the theory of Corollary 12.3 we then have*

$$\mathbf{T} \subseteq \mathbf{T}^*.$$

(To avoid possible confusion I hasten to remark that quantifiers $\forall\beta, \exists\beta, \ldots$ *in theories with language* $\mathcal{L}_2$ *are still supposed to range over* $\mathrm{OT}(\Phi)$ *while the same quantifiers in the context of* $\mathcal{L}^*$*-theories are supposed to range over set-theoretic ordinals.)*

Proof. It is easy to show that Π^1_1-$\mathbf{CA}_0 \subseteq \mathbf{KPl}^r$: By Lemma 2.5, more precisely (4.1), Π^1_1 formulae are equivalent to formulae saying that certain arithmetical relations (which may contain set parameters) are well-founded, and thus, by Theorem 5.6, they are Δ_1 on any admissible set which houses the parameters of this formula. Therefore in $\mathbf{KPl}^r$ one has comprehension for Π^1_1 formulas. (see Theorem 5.6). So it suffices to establish the closure of the $\mathbf{T}^*$-provable formulae under the rule (12.1) (modulo the *-translation). Suppose

$$\mathbf{T}^* \vdash (\exists! Q(F[Q] \wedge \mathrm{Ds}(Q)))^*.$$

Since Π^1_1 formulae are provably Δ_1 in $\mathbf{KPl}^r$ and the formula $\mathrm{Ds}(Q)$ is arithmetical in Π^1_1, $\mathrm{Ds}(Q)$ is provably Δ_1 in $\mathbf{KPl}^r$. Moreover, by Theorem 5.6, Q is order-isomorphic to an ordinal α which will then have a provable Σ_1 definition in $\mathbf{T}^*$. By rule (12.1) there exist a function f with $\mathfrak{D}_0[Q, \alpha, f]$. Picking an admissible set K with $Q, \alpha, f \in K$, we can now proceed as in the proof of Lemma 5.8 to arrive at the conclusion of the rule (12.1). □Adding Δ^1_2-$\mathbf{CA}$ to Π^1_1-$\mathbf{TR}$ enables to show that much bigger ordinals belong to $\mathfrak{W}$.

Lemma 12.6. Π^1_1-$\mathbf{TR} + \Delta^1_2$-$\mathbf{CA} \vdash (\forall\delta < \psi 00)\, [\Phi 1\delta \in \mathfrak{W} \to \Phi 1(\delta + 1) \in \mathfrak{W}]$.

Proof. Let $\delta < \psi 00$ and suppose that $\Phi 1\delta \in \mathfrak{W}$. By employing arithmetical comprehension there exists a function $f : \mathbb{N} \longrightarrow \mathrm{OT}(\Phi)$ such that $f(0) = \Phi 1\delta$ and $f(k+1) = \Omega_{f(k)}$. Using Proposition 12.2 and (IND) we obtain

$$(\forall k \in \mathbb{N})\exists X\, [\mathrm{Ds}(X) \wedge f(k) \in X \wedge f(k) < \Phi 1(\delta + 1)].$$

Since by Lemma 6.11 (Σ^1_2-AC) is available in our background theory, we may infer from (12.1) the existence of a set Y such that

$$(\forall k \in \mathbb{N})[\mathrm{Ds}(Y_k) \wedge f(k) \in Y_k].$$

Letting $Z := \bigcup\{Y_k \mid k \in \mathbb{N}\}$ (which is a set by arithmetical comprehension), we conclude with the help of Lemma 11.22 that Z is a distinguished set. Using induction on $\mathrm{G}\alpha$ one easily establishes that

$$(\forall\alpha < \Phi 1(\delta + 1))(\exists k \in \mathbb{N})\, \alpha < f(k).$$

Using (Π^1_1-CA), $U := \mathrm{W}^Z_{\Phi 1(\delta+1)}$ is a set.

If $\nu \in Z \cap \Phi 1(\delta+1)$ then $\mathrm{SC}_\nu(\Phi 1(\delta+1)) = \mathrm{SC}_\nu(1) \cup \mathrm{SC}_\nu(\delta+1) = \varnothing$, and therefore $\Phi 1(\delta+1) \in \mathrm{M}^Z_{\Phi 1(\delta+1)}$. If $\beta \in \mathrm{M}^Z_{\Phi 1(\delta+1)} \cap \Phi 1(\delta+1)$ then, by (12.1), there exists $\nu \in Z \cap \Phi 1(\delta+1)$ with $\mathrm{S}\beta \leq \nu$, whence, by Corollary 11.16(i), $\beta \in Z$. Thus, in the light of Proposition 11.9, the foregoing observations show that $\Phi 1(\delta+1) \in U$ and $\mathrm{Ds}(U)$, whence $\Phi 1(\delta+1) \in \mathfrak{W}$. □

Lemma 12.7. *Let* $\omega_0 := \varphi 00$, $\omega_{n+1} := \varphi 0\omega_n$ *and* $varepsilon_0 := \varphi 10$. *Then, for all* $n < \omega$,

$$\Pi^1_1\text{-}\mathbf{TR} + \Delta^1_2\text{-}\mathbf{CA} \vdash (\forall \alpha < \omega_n)\, \Phi 1\alpha \in \mathfrak{W}.$$

Proof. For every (meta) n,

$$\mathbf{ACA} \vdash (\forall \alpha < \omega_n)[(\forall \delta < \alpha)F(\delta) \to F(\alpha)] \to (\forall \alpha < \omega_n)\, F(\alpha)$$

for every $\mathcal{L}_2$ formula $F(\alpha)$.

Therefore it suffices to infer $\Phi 1\alpha \in \mathfrak{W}$ from the assumptions $\alpha < \omega_n$ and $(\forall \delta < \alpha)\, \Phi 1\delta \in \mathfrak{W}$.

If $\alpha = \gamma + 1$ for some γ then $\Phi 1\alpha \in \mathfrak{W}$ is a consequence of 12.6. For $\alpha = 0$ note that $\Phi 10 \in \mathfrak{W}$ holds by employing a modification of the proof of 12.6 whereby one defines $f : \mathbb{N} \longrightarrow \mathrm{OT}(\Phi)$ by $f(0) = \Omega_1$ and $f(k+1) = \Omega_{f(k)}$.

Now assume that α is a limit. By assumption we have $(\forall \delta < \alpha)\exists X\, (\Phi 1\delta \in X \wedge \mathrm{Ds}(X))$. Applying ($\Sigma^1_2$-AC) we find a set Y such that

$$(\forall \delta < \alpha)[\Phi 1\delta \in Y_\delta \wedge \mathrm{Ds}(Y_\delta)].$$

Letting $Z := \bigcup\{Y_\delta \mid \delta < \alpha\}$ and $U := \mathrm{W}^Z_{\Phi 1\alpha}$, 11.22 tells us that Z is a distinguished set. For $\nu \in Z \cap \Phi 1\alpha$ we have $\mathrm{SC}_\nu(\Phi 1\alpha) = \varnothing$ as $\alpha < \psi 00$; and hence $\Phi 1\alpha \in \mathrm{M}^Z_{\Phi 1\alpha}$. For every $\beta \in \mathrm{M}^Z_{\Phi 1\alpha} \cap \Phi 1\alpha$ there exists $\gamma < \alpha$ with $\mathrm{S}\beta \leq \Phi 1\gamma$, and thus, using 11.16(i), it follows that $\beta \in Z$. Thus, applying 11.9, the foregoing yields that $\Phi 1\alpha \in U \wedge \mathrm{Ds}(U)$, thereby verifying $\Phi 1\alpha \in \mathfrak{W}$. □

Lemma 12.8. *For* $\alpha \in \mathrm{OT}(\Phi)$ *let* $<_\alpha$ *be the restriction of* $<$ *to ordinals* $< \alpha$, *i.e.* $\beta <_\alpha \gamma \Leftrightarrow \beta < \gamma < \alpha$. *We shall write* $\mathrm{WO}(\alpha)$ *rather than* $\mathrm{WO}(<_\alpha)$. *Then:*

$$\Pi^1_1\text{-}\mathbf{CA}_0 \vdash \alpha \in \mathfrak{W} \wedge \alpha < \Omega_1 \to \mathrm{WO}(\alpha).$$

Proof. Let $\alpha \in \mathfrak{W} \cap \Omega_1$. Then there exists a distinguished set Q such that $\alpha \in Q \cap \Omega_1$. Since $\mathrm{S}\alpha = 0 \in Q$, it follows that $\alpha \in Q \cap 0^+ = \mathrm{W}[\{\eta \mid \eta < \Omega_1\}]$, and hence $\mathrm{WO}(\alpha)$. □

Lemma 12.9. *With* Ξ_n *being defined as in 12.3, the following hold:*

(i) $\Xi_n < \Xi_{n+1}$ *and* $\mathrm{K}_0\Xi_n = \varnothing$*, hence* $\psi 0\Xi_n \in \mathrm{OT}(\Phi)$.

(ii) For every $\alpha < \Phi 10$ *there exists* n *such that* $\Xi_n > \alpha$.

(iii) For every $\beta < \psi 0(\Phi 10)$ *there exists* n *such that* $\beta < \psi 0\Xi_n$.

Proof. (i) can be easily shown by induction on n. (ii) follows by induction on $\mathrm{G}\alpha$, while (iii) follows from (ii) using induction on $\mathrm{G}\beta$. □

Definition 12.10. *Let* $\mathbf{T}$ *be a theory whose language is* $\mathcal{L}_2$ *or* $\mathcal{L}^*$. *We say that an ordinal* α *is* provable in $\mathbf{T}$ *if there exists a primitive recursive well-ordering whose order-type is* α *such that* $\mathbf{T} \vdash \mathrm{WO}(\prec)$.

The proof-theoretic ordinal *of* $\mathbf{T}$ *is the least ordinal not provable in* $\mathbf{T}$*, or, equivalently, it is the supremum of the provable ordinals of* $\mathbf{T}$. *We denote this ordinal by* $|\mathbf{T}|$.

Theorem 12.11. *(i)* $\psi 0(\Phi 10) \leq |\Pi^1_1\text{-}\mathbf{TR}_0|$.

(ii) $\psi 0(\Phi 10) \leq |\Delta^1_2\text{-}\mathbf{CA} + \mathrm{BR}(\text{impl-}\Sigma^1_2)|$.

(iii) Letting $\mathbf{T}$ *be any of the theories of 12.3 or 12.5 it holds that* $\psi 0(\Phi 10) \leq |T|$.

(iv) $\psi 0(\Phi 1\varepsilon_0) \leq |\Pi^1_1\text{-}\mathbf{TR} + \Delta^1_2\text{-}\mathbf{CA}|$.

Proof. (i) follows from 12.2, 12.9 and 12.8. (ii) is a consequence of 12.4, 12.9 and 12.8. (iii) follows from 12.3 and 12.5 using 12.9 and 12.8. (iv) is a consequence of 12.7, 12.9 and 12.8. □

13 Well-ordering proofs in Π^1_1-TR and Π^1_1-TR + (BI)

Building on 12.2, we will prove lower bounds for the theories mentioned in this section's title. We will also use techniques which were developed in [6] and [7], paragraph 13.

Using (BI) we can strengthen 11.6 as follows.

Lemma 13.1. *For every* $\mathcal{L}_2$ *formula* $F(a)$*,*

$$\Pi^1_1\text{-}\mathbf{CA} + (\mathrm{BI}) \vdash \mathrm{Prg}(\mathfrak{W}, F) \rightarrow \mathfrak{W} \subseteq F.$$

Proof. By 11.6 we have $\forall X(\mathrm{Prg}((Q, X) \rightarrow Q \subseteq X)$, which in the presence of (BI) yields $\mathrm{Prg}(Q, F) \rightarrow Q \subseteq F$ for every $\mathcal{L}_2$ formula $F(a)$ (cf. [15, Lemma 1.6.3]). Assuming $\mathrm{Prg}(\mathfrak{W}, F)$ and $\alpha \in \mathfrak{W}$, we use 11.12 to infer the existence of a distinguished set P with $\alpha \in P$ and $\mathfrak{W} \cap \alpha^+ = P \cap \alpha^+$. Therefore we have $\mathrm{Prg}(P, F)$, so $P \subseteq F$, and thence $F(\alpha)$. □

With the help of 13.1 we can strengthen some of the results of section 11. Using (BI), the proof of 11.19 carries over to $\mathfrak{W}$, yielding the following strengthening of 11.20.

Lemma 13.2. Π^1_1-$\mathbf{CA}$ + (BI) $\vdash \mathrm{Prg}(\mathrm{M}^{\mathfrak{W}}_{\mu}, \mathfrak{B}^{\mathfrak{W}}_{\mu})$.

For the next Lemma the employment of (IND) is crucial.

Lemma 13.3. Π^1_1-$\mathbf{TR} \vdash \mathrm{M}^{\mathfrak{W}}_{\Phi 10} \cap \Phi 10 = \mathfrak{W} \cap \Phi 10$.

Proof. Let f be the primitive recursive function $f : \omega \longrightarrow \mathrm{OT}(\Phi)$ defined by $f(0) = 1$ and $f(k+1) = \Omega_{f(k)}$. With help of (IND), 12.2 and 12.9(ii) yield

$$(\forall k < \omega)\, f(k) \in \mathfrak{W} \;\wedge\; (\forall \alpha < \Phi 10)(\exists k < \omega)\, \alpha < f(k).$$

Let $\xi \in \mathrm{M}^{\mathfrak{W}}_{\Phi 10} \cap \Phi 10$. Then, according to (13.1), there exists $k < \omega$ with $\mathrm{S}\xi \leq f(k)$. By 11.16 we then get $\xi \in \mathfrak{W} \cap \Phi 10$. Conversely, if $\xi, \mu \in \mathfrak{W} \cap \Phi 10$ we have $\mathrm{SC}_{\mu}(\xi) \subseteq \mathfrak{W}$ by 11.7(ii), whence $\xi \in \mathrm{M}^{\mathfrak{W}}_{\Phi 10} \cap \Phi 10$. □

Definition 13.4. *By $\mathfrak{I}(U, \alpha)$ we shall refer to the schema*

$$\mathrm{Prg}(U, F) \rightarrow \alpha \in U \;\wedge\; U \cap \alpha \subseteq F$$

where $F(a)$ is an arbitrary formula of $\mathcal{L}_2$.

Lemma 13.5. Π^1_1-$\mathbf{TR}$ + (BI) $\vdash \mathfrak{I}(\mathrm{M}^{\mathfrak{W}}_{\Phi 10}, (\Phi 10) + 1)$.

Proof. Let $X := \mathrm{M}^{\mathfrak{W}}_{\Phi 10}$ and $\tau := \Phi 10$. According to 13.3 we have $X \cap \tau = \mathfrak{W} \cap \tau$ which implies

$$\mathrm{Prg}(X, F) \rightarrow \mathrm{Prg}(\mathrm{W}, \{\xi \mid \xi < \tau \rightarrow F(\xi)\}),$$

and which, with the help of 13.1, implies $\mathrm{Prg}(X, F) \rightarrow \mathfrak{W} \cap \tau \subseteq F$. The latter yields

$$\mathrm{Prg}(X, F) \rightarrow X \cap \tau \subseteq F.$$

Since also $\tau, \tau + 1 \in X$, the desired assertion follows. □

Definition 13.6. *For every formula $F(a)$ we define the "Gentzen jump"*

$$F^j(\gamma) \;:=\; \forall \delta\, [\mathrm{M}^{\mathfrak{W}}_{\Phi 10} \cap \delta \subseteq F \rightarrow \mathrm{M}^{\mathfrak{W}}_{\Phi 10} \cap (\delta + \omega^{\gamma}) \subseteq F].$$

Lemma 13.7. *The following are deducible in* Π^1_1-$\mathbf{TR}$*:*

(i) $F^j(\gamma) \rightarrow \mathrm{M}^{\mathfrak{W}}_{\Phi 10} \cap \omega^{\gamma} \subseteq F$.

(ii) $\mathrm{Prg}(\mathrm{M}^{\mathfrak{W}}_{\Phi 10}, F) \rightarrow \mathrm{Prg}(\mathrm{M}^{\mathfrak{W}}_{\Phi 10}, F^j)$.

Proof. (i) is obvious. (ii) Let $M := \mathrm{M}^{\mathfrak{W}}_{\Phi 10}$. Then $M \cap (\delta + \omega^\gamma) \subseteq F$ is to proved under the assumptions (a) $\mathrm{Prg}(M, F)$, (b) $\gamma \in M \wedge M \cap \gamma \subseteq F^j$ and (c) $M \cap \delta \subseteq F$. So let $\eta \in M \cap (\delta + \omega^\gamma)$.

1. $\eta < \delta$: Then $F(\eta)$ is a consequence of (c).
2. $\eta = \delta$: Then $F(\eta)$ follows from (c) and (a).
3. $\delta < \eta < \delta + \omega^\gamma$: Then there exist $\gamma_1, \ldots, \gamma_k < \gamma$ such that $\eta = \delta + \omega^{\gamma_1} + \ldots + \omega^{\gamma_k}$ and $\gamma_1 \geq \ldots \geq \gamma_k$. $\eta \in M$ implies $\gamma_1, \ldots, \gamma_k \in M \cap \gamma$. Through applying (b) and (c) we obtain $M \cap (\delta + \omega^{\gamma_1}) \subseteq F$. By iterating this procedure we eventually arrive at $F(\delta + \omega^{\gamma_1} + \ldots + \omega^{\gamma_k})$, so $F(\eta)$ holds.

□

Lemma 13.8. *Let* $\delta_0 := (\Phi 10) + 1$, $\delta_{n+1} := \omega^{\delta_n}$ *and* $M := \mathrm{M}^{\mathfrak{W}}_{\Phi 10}$. *Then:*

$$\Pi^1_1\text{-}\mathbf{TR} + (\mathrm{BI}) \vdash \mathfrak{I}(M, \delta_n).$$

Proof. Proof by meta-induction on n. For $n = 0$ this follows from 13.5. Now let $n = m + 1$. Inductively we have $\mathrm{Prg}(M, F^j) \rightarrow F^j(\delta_m)$ for every formula $F(a)$. An application of 13.7 yields $\mathrm{Prg}(M, F) \rightarrow M \cap \delta_n \subseteq F$. Since trivially $\delta_n \in M$, we have shown $\mathfrak{I}(M, \delta_n)$. □

Theorem 13.9. $\psi 0 \varepsilon_{(\Phi 10)+1} \leq |\Pi^1_1\text{-}\mathbf{TR} + (\mathrm{BI})|$.

Proof. 13.2 and 13.8 yield $\delta_n \in \mathfrak{B}^{\mathfrak{W}}_{\Phi 10}$, and consequently $\psi 0 \delta_n \in \mathfrak{W}$. Since $\sup\{\psi 0 \delta_n \mid n < \omega\} = \psi 0 \varepsilon_{(\Phi 10)+1}$ the proof is completed. □

We now come to the well-ordering proof for Π^1_1-$\mathbf{TR}$. Since (BI) is not available in this theory, 13.1 cannot be exploited to prove that $\mathfrak{I}(\mathrm{M}^{\mathfrak{W}}_{\Phi 10}, \Phi 10)$ holds. However, Π^1_1-$\mathbf{TR}$ proves $(\forall \alpha < \Phi 10)(\exists k < \omega)(\alpha < f(k) \wedge f(k) \in \mathfrak{W})$ (where f was defined in 13.3), establishing $\psi 0((\Phi 10) \cdot \varepsilon_0)$ as a lower bound for this theory.

Convention: For the remainder of this section we let $\mathfrak{f} := \Phi 10$.

Lemma 13.10. *Multiplication* $\alpha \cdot \beta$ *of ordinals from* $\mathrm{OT}(\Phi)$ *can be easily defined via the normal forms of* α *and* β. *For* $\alpha \leq \varepsilon_0$ *we have:*

(i) $\mathrm{K}_0(\mathfrak{f} \cdot \alpha) = \varnothing$.

(ii) $\nu < \mathfrak{f} \rightarrow \mathrm{SC}_\nu(\mathfrak{f} \cdot \alpha) = \varnothing$.

(iii) $\beta < \mathfrak{f} \cdot \alpha \rightarrow (\exists \xi < \alpha)(\exists \delta < \mathfrak{f})(\beta = \mathfrak{f} \cdot \xi + \delta)$.

(iv) $\beta < \psi 0(\mathfrak{f} \cdot \varepsilon_0) \rightarrow (\exists \xi < \varepsilon_0)\, \beta < \psi 0(\mathfrak{f} \cdot \xi)$.

Proof. The proofs consist of simple calculations and in (iii) and (iv) involve inductions on $\mathrm{G}\beta$. □

Definition 13.11.

$$\begin{aligned}\mathfrak{H}(\delta) &:= \delta \leq \varepsilon_0 \wedge \\ &\qquad (\forall \mu \in \mathfrak{W} \cap \mathfrak{f})(\forall \eta, \nu \in \mathfrak{W} \cap \mu^+)[\mathrm{K}_\nu \eta < \mathfrak{f} \cdot \delta + \eta \rightarrow \psi\nu(\mathfrak{f} \cdot \delta + \eta) \in \mathfrak{W}],\\ \mathfrak{A}^\delta(\alpha, \mu, \nu) &:= \delta < \varepsilon_0 \wedge \mu \in \mathfrak{W} \cap \mathfrak{f} \wedge \alpha, \nu \in \mathfrak{W} \cap \mu^+ \wedge \mathrm{K}_\nu \alpha < \mathfrak{f} \cdot \delta + \alpha \wedge \\ &\qquad (\forall \eta \in \mathfrak{W} \cap \alpha)(\forall \tau' \in \mathfrak{W} \cap \mu^+)[\mathrm{K}_{\tau'} \eta < \mathfrak{f} \cdot \delta + \eta \rightarrow \psi\tau'(\mathfrak{f} \cdot \delta + \eta) \in \mathfrak{W}].\end{aligned}$$

Lemma 13.12. Π^1_1-**TR** $\vdash (\forall \xi < \delta)\mathfrak{H}(\xi) \wedge \mathfrak{A}^\delta(\alpha, \mu, \nu) \rightarrow (\forall \gamma \in \mathrm{M}^{\mathfrak{W}}_\nu \cap \psi\nu(\mathfrak{f}\delta + \alpha))(\gamma \in \mathfrak{W})$.

Proof. Assume the antecedent of the implication we have to verify. Let $\gamma \in \mathrm{M}^{\mathfrak{W}}_\nu \cap \psi\nu(\mathfrak{f}\delta + \alpha)$. We shall carry out an induction on $\mathrm{G}\gamma$ in order to show $\gamma \in \mathfrak{W}$, by distinguishing between the different shapes γ might assume. We shall write $\mathfrak{f}\delta$ for $\mathfrak{f} \cdot \delta$.

1. $\gamma \leq \nu$: Then $\gamma \in \mathfrak{W}$ follows from 11.17(ii). Henceforth assume $\gamma > \nu$.

2. $\gamma =_{nf} \gamma_1 + \ldots + \gamma_n$ or $\gamma =_{nf} \varphi\gamma_1\gamma_2$. Then $\gamma_j \in \mathrm{M}^{\mathfrak{W}}_\nu \cap \psi\nu(\mathfrak{f}\delta + \alpha)$ and therefore, by the inductive assumption, $\gamma_j \in \mathfrak{W}$, thus $\gamma \in \mathfrak{W}$ by 11.13 and 11.15, respectively.

3. $\gamma =_{nf} \psi\nu(\mathfrak{f}\delta + \alpha')$ and $\alpha' < \alpha$: Let $\gamma' := \mathfrak{f}\delta + \alpha'$. Since $\gamma \in \mathrm{M}^{\mathfrak{W}}_\nu$, 11.7(i) entails that
$$(\forall \tau \in \mathfrak{W} \cap \nu)(\forall \xi \in \mathrm{SC}_\nu(\gamma'))\, \mathrm{SC}_\tau(\xi) \subseteq \mathfrak{W}.$$
The latter implies $\mathrm{SC}_\nu(\gamma') \subseteq \mathrm{M}^{\mathfrak{W}}_\nu \cap \psi\nu\gamma' \subseteq \mathrm{M}^{\mathfrak{W}}_\nu \cap \psi\nu(\mathfrak{f}\delta + \alpha)$. Thus
$$\mathrm{SC}_\nu(\gamma') \subseteq \mathfrak{W}$$
by the induction hypothesis. Next we show via a subsidiary induction on Q that for every distinguished set Q with $\mu \in Q$,
$$(\forall \tau \in Q \cap \mu^+)\, \mathrm{SC}_\tau(\gamma') \subseteq Q.$$
We shall frequently use the fact that $\mathfrak{W} \cap \mu^+ = Q \cap \mu^+$ holds (by 12.11). If $\tau = \nu$ then this follows from (13.1). If $\tau < \nu$ then $\mathrm{SC}_\tau(\gamma') \subseteq \mathrm{SC}_\tau(\gamma) \subseteq \mathfrak{W} \cap \mu^+ \subseteq Q$ since $\gamma \in \mathrm{M}^{\mathfrak{W}}_\nu$.

Now assume that $\nu < \tau \leq \mu$. Since $\mathrm{SC}_\tau(\gamma') < \psi\tau\gamma' < \psi\tau(\mathfrak{f}\delta + \alpha)$, the subsidiary induction hypothesis yields $\mathrm{SC}_\tau(\gamma') \subseteq \mathrm{M}_\tau^{\mathfrak{W}} \cap \psi\tau(\mathfrak{f}\delta + \alpha)$. Moreover, $\mathrm{K}_\tau\alpha \subseteq \mathrm{K}_\nu\alpha < \mathfrak{f}\delta + \alpha$. Therefore $\mathfrak{A}^\delta(\alpha, \mu, \nu)$ and consequently, by the main induction hypothesis, $\mathrm{SC}_\tau(\gamma') \subseteq \mathfrak{W} \cap \mu^+ \subseteq Q$. This completes the proof of (13.1). As a result, $(\forall\tau \in \mathfrak{W} \cap \mu^+)\, \mathrm{SC}_\tau(\gamma') \subseteq \mathfrak{W}$. In combination with 11.16 the latter entails $\alpha' \in \mathfrak{W}$. Finally, $\mathfrak{A}^\delta(\alpha, \mu, \nu)$ and $\alpha' \in \mathfrak{W} \cap \alpha$ imply $\gamma \in \mathfrak{W}$.

4. $\gamma =_{nf} \psi\nu(\mathfrak{f}\delta' + \alpha')$, $\delta' < \delta$ and $\alpha' < \mathfrak{f}$: Let $\gamma' := \mathfrak{f}\delta' + \alpha'$. Let Q be a distinguished set. Via a subsidiary induction on Q we shall show that

$$(\forall\tau \in Q \cap \mathfrak{f})\, \mathrm{SC}_\tau(\gamma') \subseteq Q.$$

For $\tau \leq \nu$ this follows as in the previous case. Let $\nu < \tau < \mathfrak{f}$. Since $\mathrm{SC}_\tau(\gamma') < \psi\tau\gamma'$ and $\psi\tau\gamma' < \psi\tau(\mathfrak{f}\delta)$ the subsidiary induction hypothesis yields $\mathrm{SC}_\tau(\gamma') \subseteq \mathrm{M}_\tau^{\mathfrak{W}} \cap \psi\tau(\mathfrak{f}\delta)$, so that, owing to $\mathfrak{A}^\delta(0, \tau, \tau)$ and the main induction hypothesis, we arrive at $\mathrm{SC}_\tau(\gamma') \subseteq \mathfrak{W} \cap \tau^+ \subseteq Q$. This concludes the proof of (13.1).

(13.1) implies $(\forall\tau \in \mathfrak{W} \cap \mathfrak{f})\, \mathrm{SC}_\tau(\alpha') \subseteq \mathfrak{W}$, thence $\alpha' \in \mathrm{M}_{\mathfrak{f}}^{\mathfrak{W}} \cap \mathfrak{f}$. Via 13.3 we thus infer $\alpha' \in \mathfrak{W}$. Since $\delta' < \delta$ we also have $\mathfrak{H}(\delta')$ and therefore $\gamma \in \mathfrak{W}$.

□

Lemma 13.13. Π^1_1-**TR** $\vdash \delta < \varepsilon_0 \wedge (\forall\xi < \delta)\mathfrak{H}(\xi) \rightarrow \mathfrak{H}(\delta)$.

Proof. Assume $\delta < \varepsilon_0$ and $(\forall\xi < \delta)\mathfrak{H}(\xi)$. From $\mathfrak{A}^\delta(\alpha, \mu, \nu)$ and $\alpha, \nu \in Q$ we can infer $\psi\nu(\mathfrak{f}\delta + \alpha) \in \mathrm{M}_\nu^Q$ and with the help of 13.12 also $\mathrm{M}_\nu^Q \cap \psi\nu(\mathfrak{f}\delta + \alpha) \subseteq Q$, and hence $\psi\nu(\mathfrak{f}\delta + \alpha) \in Q$ by 11.4(viii). This shows

$$\mathfrak{A}^\delta(\alpha, \mu, \nu) \rightarrow \psi\nu(\mathfrak{f}\delta + \alpha) \in \mathfrak{W}.$$

Let $\mu \in \mathfrak{W} \cap \mathfrak{f}$. We want to show

$$(\forall\alpha, \nu \in \mathfrak{W} \cap \mu^+)\, [\mathrm{K}_\nu\alpha < \mathfrak{f}\delta + \alpha \rightarrow \psi\nu(\mathfrak{f}\delta + \alpha) \in \mathfrak{W}].$$

So let Q be a distinguished set with $\mu \in Q$. Since $\mathfrak{W} \cap \mu^+ = Q \cap \mu^+$ it suffices to show that if $\alpha, \nu \in Q \cap \mu^+$ and $\mathrm{K}_\nu\alpha < \mathfrak{f}\delta + \alpha$ hold true then $\psi\nu(\mathfrak{f}\delta + \alpha) \in Q$. We use induction on Q with α being the variable of induction. By induction hypothesis we then have

$$(\forall\eta \in Q \cap \alpha)(\forall\tau \in Q \cap \mu^+)[\mathrm{K}_\tau\eta < \mathfrak{f}\delta + \eta \rightarrow \psi\tau(\mathfrak{f}\delta + \eta) \in Q].$$

But then (13.1) implies $\psi\nu(\mathfrak{f}\delta + \alpha) \in Q$. □

Theorem 13.14. $\psi 0((\Phi 10)\cdot\varepsilon_0) \leq |\Pi^1_1\text{-}\mathbf{TR}|$.

Proof. Given $\beta < \psi 0((\Phi 10)\cdot\varepsilon_0)$ there exists (by 13.10(iv)) ω_n such that $\beta < \psi 0((\Phi 10)\cdot\omega_n)$. Since in $\Pi^1_1\text{-}\mathbf{TR}$ we have full transfinite induction on the initial segment of ordinals $\leq \omega_n$, Lemma 13.13 yields $\Pi^1_1\text{-}\mathbf{TR} \vdash \mathfrak{H}(\omega_n)$. Thus, using 11.21 and 12.8, we obtain

$$\Pi^1_1\text{-}\mathbf{TR} \vdash \mathrm{WO}(\psi 0((\Phi 10)\cdot\omega_n)),$$

which implies $\psi 0((\Phi 10)\cdot\varepsilon_0) \leq |\Pi^1_1\text{-}\mathbf{TR}|$. □

14 Well-ordering proofs in $\Sigma^1_2\text{-}\mathrm{TRDC_0}$ and $\Sigma^1_2\text{-}\mathrm{TRDC}$.

We start with the key lemma for all of the remaining well-ordering proofs.

Lemma 14.1.

$\Sigma^1_2\text{-}\mathbf{TRDC}_0 \vdash \eta \in \mathfrak{W} \wedge (\forall \xi \in \mathfrak{W}\cap\eta)(\forall \alpha \in \mathfrak{W})(\Phi\xi\alpha \in \mathfrak{W}) \to (\forall \beta \in \mathfrak{W})(\Phi\eta\beta \in \mathfrak{W})$.

Proof. We shall argue on the basis of $\Sigma^1_2\text{-}\mathbf{TRDC}_0$. Suppose $\eta \in \mathfrak{W}$ and

$$(\forall \xi \in \mathfrak{W}\cap\eta)(\forall \alpha \in \mathfrak{W})(\Phi\xi\alpha \in \mathfrak{W}).$$

Let $\beta \in \mathfrak{W}$. Pick a distinguished set Q with $\eta, \beta \in Q$. For every distinguished set X we then have

$$(\forall \xi \in Q\cap\eta)(\forall \alpha \in X)\exists Y[\mathrm{Ds}(Y) \ \wedge \ \Phi\xi\alpha \in Y].$$

Thus, with the help of (Σ^1_2-AC) we find a set U such that

$$(\forall \xi \in Q\cap\eta)(\forall \alpha \in X)[\mathrm{Ds}(U_{\langle\xi,\alpha\rangle}) \ \wedge \ \Phi\xi\alpha \in U_{\langle\xi,\alpha\rangle}].$$

Letting

$$U^* \ := \ \bigcup\{U_{\langle\xi,\alpha\rangle} \mid \xi \in Q\cap\eta \ \wedge \ \alpha \in X\}$$

we have $\mathrm{Ds}(U^*)$ (by 11.22) and also $(\forall \xi \in Q\cap\eta)(\forall \alpha \in X)(\Phi\xi\alpha \in U^*)$. For an arbitrary distinguished set P the foregoing considerations imply that

$$\begin{aligned}&(\forall i < \omega)\forall X\exists Y[(i = 0 \to Y = P) \ \wedge \\ &\qquad (i > 0 \ \wedge \ \mathrm{Ds}(X) \to [\mathrm{Ds}(Y) \ \wedge \ (\forall \xi \in Q\cap\eta)(\forall \alpha \in X)(\Phi\xi\alpha \in Y)])].\end{aligned}$$

By applying (Σ^1_2-TRDC) (in actuality (Σ^1_2-DC) suffices) to (14.1) we can draw the existence of a set Z satisfying $Z_0 = P$ and for all $i > 0$,

$$\mathrm{Ds}(\bigcup\{Z_j \mid j < i\}) \ \to \ \mathrm{Ds}(Z_i) \wedge (\forall \xi \in Q\cap\eta)(\forall \alpha \in \bigcup\{Z_j \mid j < i\})\, \Phi\xi\alpha \in Z_i.$$

Induction on i in conjunction with 11.22 yields $\mathrm{Ds}(Z_i)$ for all i. Note that this induction is permissible in our background theory since $\{i < \omega \mid \mathrm{Ds}(Z_i)\}$ is a set by $(\Delta^1_2\text{-CA})$. Letting $P^* := \bigcup\{Z_i \mid i < \omega\}$ we have

$$\mathrm{Ds}(P^*) \,\wedge\, P \subseteq P^* \,\wedge\, (\forall \xi \in Q \cap \eta)(\forall \alpha \in P^*)\, \Phi\xi\alpha \in P^*.$$

Thus we showed that for all $\gamma \in Q$ and for all X there exists Y such that

$$\mathrm{Ds}(X) \to \exists Z[\mathrm{Ds}(Z) \,\wedge\, Q \cup X \subseteq Z \,\wedge\, (\forall \xi \in Q \cap \eta)(\forall \alpha \in Z)(\Phi\xi\alpha \in Z) \,\wedge\, Y = \mathrm{W}^{Z}_{\Phi\eta}$$

The latter formula is equivalent to a Σ^1_2 formula (using $(\Sigma^1_2\text{-AC})$), hence an via an application of $(\Sigma^1_2\text{-TRDC})$, with $< \cap (Q \times Q)$ being the well-ordering, there exists a set R such that

$$(\forall \gamma \in Q)[\mathrm{Ds}(R_{Q\gamma}) \,\to\, \exists Z[\mathrm{Ds}(Z) \,\wedge\, Q \cup R_{Q\gamma} \subseteq Z \,\wedge\, (\forall \xi \in Q \cap \eta)(\forall \alpha \in Z)(\Phi\xi\alpha \in Z) \,\wedge\, R_\gamma = \mathrm{W}^Z_{\Phi\eta\gamma}]],$$

where $R_{Q\gamma} := \bigcup\{R_\delta \mid \delta \in Q \cap \gamma\}$. By induction on Q we shall show that

$$(\forall \gamma \in Q)[\mathrm{Ds}(R_\gamma) \,\wedge\, \Phi\eta\gamma \in R_\gamma].$$

So assume inductively that $(\forall \delta \in Q \cap \gamma)[\mathrm{Ds}(R_\delta) \,\wedge\, \Phi\eta\gamma \in R_\delta]$. This implies $\mathrm{Ds}(R_{Q\gamma})$ and, in view of (14.1), there exists a set Z satisfying the following:

(a) $\mathrm{Ds}(Z)$;

(b) $Q \cup R_{Q\gamma} \subseteq Z$;

(c) $(\forall \xi \in Q \cap \eta)(\forall \alpha \in Z)(\Phi\xi\alpha \in Z)$;

(d) $(\forall \delta \in Q \cap \gamma)\Phi\eta\delta \in Z$;

(e) $R_\gamma = \mathrm{W}^Z_{\Phi\eta\gamma}$.

If $\gamma = \Phi\eta\gamma$ we have $\Phi\eta\gamma = \gamma \in Z \cap \gamma^+ = \mathrm{W}^Z_\gamma = R_\gamma$, which implies $\mathrm{Ds}(R_\gamma)$ and $\Phi\eta\gamma \in R_\gamma$.

Next assume that $\gamma < \Phi\eta\gamma$. If $\nu \in Z \cap \Phi\eta\gamma$ then $\mathrm{SC}_\nu(\Phi\eta\gamma) \subseteq \mathrm{SC}_\nu(\eta) \cup \mathrm{SC}_\nu(\gamma) \subseteq Z$ by 11.7(ii) since $\eta, \gamma \in Q \subseteq Z$. Therefore we have

$$\Phi\eta\gamma \in \mathrm{M}^Z_{\Phi\eta\gamma}.$$

We will also show that

$$\mathrm{M}^Z_{\Phi\eta\gamma} \cap \Phi\eta\gamma \subseteq Z.$$

Let $\rho \in \mathrm{M}^Z_{\Phi\eta\gamma} \cap \Phi\eta\gamma$. We shall employ induction on $\mathrm{G}\rho$ to show that $\rho \in Z$. If $\rho \notin \mathrm{SC}$ then $\rho \in Z$ follows from the inductive assumption by means of 11.13 and 11.15. Now suppose $\rho \in \mathrm{SC}$. If there exists $\nu \in Z \cap \Phi\eta\gamma$ with $\mathrm{S}\rho \leq \nu$ then $\mathrm{SC}_\nu(\rho) = \{\rho\} \subseteq Z$. Thus, in addition, we may assume that

$$\rho \in \mathrm{SC} \wedge (\forall \nu \in Z \cap \Phi\eta\gamma)(\nu < \mathrm{S}\rho).$$

We will distinguish several cases.

1. $\rho =_{nf} \psi\mu\zeta$: Then we have $\mu \in \mathrm{M}^Z_{\Phi\eta\gamma} \cap \Phi\eta\gamma$ by (14.1) since $\mu = \mathrm{S}\rho$. Applying the induction hypothesis we obtain $\mu \in Z$ which contradicts (14.1). Thus this case is ruled out.

2. $\rho =_{nf} \Phi\zeta\sigma$: Then (14.1) in conjunction with the induction hypothesis yields $\zeta, \sigma \in Z$.
   (i) $\zeta < \eta$: Then we have $\zeta \in \mathfrak{W} \cap \eta = Q \cap \eta$ by 11.12 since $\eta \in Q$. Whence $\rho \in Z$ holds owing to (c).
   (ii) $\zeta = \eta$ and $\sigma < \gamma$: Then, using (d), from $\sigma \in \mathfrak{W} \cap \gamma = Q \cap \gamma$ we obtain $\rho \in Z$.
   (iii) $\eta < \zeta$: In this case $\rho < \gamma$ must hold. Since $\gamma < \Phi\eta\gamma$ holds by assumption, $\rho \in Z$ follows with the aid of 11.16(i) since in this case we have $\mathrm{S}\gamma \in Q \cap \Phi\eta\gamma \subseteq Z \cap \Phi\eta\gamma$.

This completes the proof of (14.1). Applying (14.1), (14.1) and (e) in conjunction with 11.9, we conclude that $\mathrm{Ds}(R_\gamma) \wedge \Phi\eta\gamma \in R_\gamma$, thereby finishing the poof of (14.1). Finally, since $\beta \in Q$, (14.1) enables us to conclude that $\Phi\eta\beta \in \mathfrak{W}$. □

Corollary 14.2. *For any (meta) n,* Σ^1_2-$\mathbf{TRDC}_0 \vdash (\forall\alpha \in \mathfrak{W})\, \Phi n\alpha \in \mathfrak{W}$.

Proof. Use meta-induction on n. 12.2 yields the induction base while 14.1 provides the induction step. □

Corollary 14.3. *For any (meta) n,* Σ^1_2-$\mathbf{TRDC} \vdash (\forall\xi \leq \omega_n)(\forall\alpha \in \mathfrak{W})\, \Phi\xi\alpha \in \mathfrak{W}$.

Proof. In Σ^1_2-$\mathbf{TRDC}$ one has full induction for arbitrary formulae over any segment ω_n. Thus the assertion follows from 14.1. □

Theorem 14.4. *(i)* $\psi 0(\Phi\omega 0) \leq |\Sigma^1_2\text{-}\mathbf{TRDC}_0|$.

(ii) $\psi 0(\Phi\varepsilon_0 0) \leq |\Sigma^1_2\text{-}\mathbf{TRDC}|$.

Proof. (i) and (ii) are consequences of 14.2 and 14.3, respectively, by also enlisting the help of 11.21 and 12.8. □

15 Well-ordering proofs in $\mathbf{\Sigma^1_2}$-$\mathbf{TRDC + BR}$ and $\mathbf{\Sigma^1_2}$-$\mathbf{TRDC + BR(impl\text{-}\Sigma^1_2)}$.

Definition 15.1. *Let* $\boldsymbol{\vartheta}_0 := \Omega_1$, $\boldsymbol{\zeta}_0 := \psi 0 \boldsymbol{\vartheta}_0$, $\boldsymbol{\vartheta}_{n+1} := \Phi \boldsymbol{\zeta}_n 0$, $\boldsymbol{\zeta}_{n+1} := \psi 0 \boldsymbol{\vartheta}_{n+1}$.

Lemma 15.2. *(i) For all* n: $\mathrm{K}_0 \boldsymbol{\vartheta}_n < \boldsymbol{\vartheta}_n$, $\boldsymbol{\vartheta}_n < \boldsymbol{\vartheta}_{n+1}$ *and* $\boldsymbol{\zeta}_n =_{nf} \psi 0 \boldsymbol{\vartheta}_n$.

(ii) For every $\alpha < \Phi \Omega_1 0$ *there exists* n *such that* $\alpha < \boldsymbol{\vartheta}_n$.

(iii) For every $\beta < \psi 0 (\Phi \Omega_1 0)$ *there exists* n *such that* $\beta < \boldsymbol{\zeta}_n$.

Proof. We show (i) by induction on n. This is obvious when $n = 0$. Let $n = m + 1$. By the induction hypothesis we have $\mathrm{K}_0 \boldsymbol{\vartheta}_n = \mathrm{K}_0 \boldsymbol{\zeta}_m = \{\boldsymbol{\vartheta}_m\} \cup \mathrm{K}_0 \boldsymbol{\vartheta}_m < \boldsymbol{\vartheta}_n$, and consequently $\boldsymbol{\zeta}_n =_{nf} \psi 0 \boldsymbol{\vartheta}_n$, $\boldsymbol{\zeta}_m < \boldsymbol{\zeta}_n$ and $\boldsymbol{\vartheta}_n = \Phi \boldsymbol{\zeta}_m 0 < \Phi \boldsymbol{\zeta}_n 0 = \boldsymbol{\vartheta}_{n+1}$.

(ii): We use induction on $\mathrm{G}\alpha$. First suppose $\alpha =_{nf} \Phi \xi \eta$. Then, by induction hypothesis, there exist $n, n' < \omega$ such that $\xi < \boldsymbol{\vartheta}_n$ and $\eta < \boldsymbol{\vartheta}_{n'}$. Letting $k := \max(n, n') + 1$ it follows by (i) that $\alpha < \boldsymbol{\vartheta}_k$. In all other cases the assertion follows directly from the induction hypothesis.

(iii) is easily shown by induction on $\mathrm{G}\beta$ making use of (ii). □

Lemma 15.3. *For all (meta)* n, Σ^1_2-$\mathbf{TRDC} + \mathrm{BR} \vdash \boldsymbol{\zeta} \in \mathfrak{W}$.

Proof. We use (meta) induction on n. For $n = 0$ this is a consequence of 12.2 and 11.21. If $n = m + 1$ then the induction hypothesis yields that $\boldsymbol{\zeta}_m \in \mathfrak{W}$ is deducible in the theory and therefore, by 12.8, $\mathrm{WO}(\boldsymbol{\zeta}_m)$ holds. The segment below $\boldsymbol{\zeta}_m$ is then a primitive recursive provable well-ordering, thus an application of BR yields $\Phi \boldsymbol{\zeta}_m 0 = \boldsymbol{\vartheta}_n \in \mathfrak{W}$. Consequently, using 15.2 and 11.21, we have the derivability of $\psi 0 \boldsymbol{\vartheta}_n = \boldsymbol{\zeta}_n \in \mathfrak{W}$. □

Lemma 15.4. *For all (meta)* n, Σ^1_2-$\mathbf{TRDC} + \mathrm{BR}(\text{impl-}\Sigma^1_2) \vdash \boldsymbol{\rho}_n \in \mathfrak{W}$, *where* $\boldsymbol{\rho}_0 := \Phi 00$ *and* $\boldsymbol{\rho}_{m+1} := \Phi \boldsymbol{\rho}_m 0$.

Proof. We use (meta) induction on n. Let's denote the above theory by $\mathbf{T}$. The case $n = 0$ follows from 12.2. Let $n = m + 1$. The induction hypothesis yields $\mathbf{T} \vdash \boldsymbol{\rho}_m \in \mathfrak{W}$. Owing to 11.11, provably in $\mathbf{T}$ there exists a distinguished set Q such that $\boldsymbol{\rho}_m \in Q$ and $Q = Q \cap \boldsymbol{\rho}_m^+$. With the formula

$$F[U] := \exists P\, [\mathrm{Ds}(P) \wedge \boldsymbol{\rho}_m \in P \wedge U = \{\langle \alpha, \beta \rangle \mid \alpha, \beta \in P \wedge \alpha < \beta \wedge \beta < \boldsymbol{\rho}_m^+\}]$$

it thus holds that

$$\mathbf{T} \vdash \exists! X (\mathrm{WO}(X) \wedge F[X]).$$

Let $G(\xi) := (\forall\alpha \in \mathfrak{W})\, \Phi\xi\alpha \in \mathfrak{W}$ and $\boldsymbol{\tau} := \boldsymbol{\rho}_m^+$. Since $F[U]$ is (provably in $\mathbf{T}$) equivalent to a Σ^1_2 formula, via an application of $\mathrm{BR}(\text{impl-}\Sigma^1_2)$ to (15.1), $\mathbf{T}$ proves transfinite induction on $\mathfrak{W} \cap \boldsymbol{\tau}$. In particular,

$$\mathbf{T} \vdash \mathrm{Ds}(Q) \wedge \boldsymbol{\rho}_m \in Q \wedge (\forall\eta \in Q \cap \boldsymbol{\tau})[Q \cap \eta \subseteq G \to G(\eta)] \to (\forall\eta \in Q \cap \boldsymbol{\tau})G(\eta).$$

In conjunction with the induction hypothesis and 14.1, (15.1) implies $\mathbf{T} \vdash \boldsymbol{\rho}_n \in \mathfrak{W}$. □

Theorem 15.5. *(i)* $\psi 0(\Phi\Omega_1 0) \leq |\Sigma^1_2\text{-}\mathbf{TRDC} + \mathrm{BR}|$.

(ii) $\psi 0\Gamma^\Phi_0 \leq |\Sigma^1_2\text{-}\mathbf{TRDC} + \mathrm{BR}(\text{impl-}\Sigma^1_2)|$, *where* $\psi 0\Gamma^\Phi_0 := \mathrm{OT}(\Phi) \cap \Omega_1$.

Proof. (i) follows from 15.3, 15.2(iii), and 12.8. (ii) follows from 15.4, 11.21, 12.8 and 9.8. □

16 Prospectus

The lower bounds for the proof-theoretic ordinals of theories considered in this article turn out to be sharp. Proofs of upper bounds, though, will only appear in the second part of this paper which is devoted to ordinal analysis. We will finish this paper by listing all theories and their proof-theoretic ordinals.

(i) $|\mathbf{ID}_{\prec^*}| \leq |\mathbf{ID}^* + (\mathrm{BI})| \leq |\mathbf{KPl}^*| \leq \psi 0\varepsilon_{(\Phi 0\Omega_1)+1}$.

(ii) $|\Pi^1_1\text{-}\mathbf{TR}_0| = |\mathbf{AUT\text{-}ID}^{pos}_0| = |\mathbf{AUT\text{-}ID}^{mon}_0| = |\Pi^1_1\text{-}\mathbf{TR}_0 + \Delta^1_2\text{-}\mathbf{CA}_0| = |\mathbf{AUT\text{-}KPl}^r| =$

$|\mathbf{AUT\text{-}KPl}^r + \mathbf{KPi}^r| = |\mathbf{KPi}^w + \mathrm{FOUNDR}(\text{impl-}\Sigma)| = |\mathbf{KPi}^w + \mathrm{FOUND}(\text{impl-}\Sigma)| =$

$|\Delta^1_2\text{-}\mathbf{CA} + \mathrm{BI}(\text{impl-}\Sigma^1_2)| = |\Delta^1_2\text{-}\mathbf{CA} + \mathrm{BR}(\text{impl-}\Sigma^1_2)| = \psi 0(\Phi 10)$.

(iii) $|\Pi^1_1\text{-}\mathbf{TR}| = |\mathbf{AUT\text{-}ID}^{pos}| = |\mathbf{AUT\text{-}ID}^{mon}| = |\mathbf{AUT\text{-}KPl}^w| = \psi 0((\Phi 10) \cdot \varepsilon_0)$.

(iv) $|\Pi^1_1\text{-}\mathbf{TR} + (\mathrm{BI})| = |\mathbf{AUT\text{-}ID}^{pos}_2| = |\mathbf{AUT\text{-}ID}^{mon}_2| = |\mathbf{AUT\text{-}KPl}| = \psi 0\varepsilon_{(\Phi 10)+1}$.

(v) $|\Pi^1_1\text{-}\mathbf{TR} + \Delta^1_2\text{-}\mathbf{CA}| = |\Pi^1_1\text{-}\mathbf{TR} + \Sigma^1_2\text{-}\mathbf{AC}| = |\mathbf{AUT\text{-}KPl}^w + \mathbf{KPi}^w| = \psi 0(\Phi 1\varepsilon_0)$.

(vi) $|\Delta^1_2\text{-}\mathbf{TR}_0| = |\Sigma^1_2\text{-}\mathbf{TRDC}_0| = |\Delta^1_2\text{-}\mathbf{CA}_0 + (\Sigma^1_2\text{-}\mathrm{BI})| = |\mathbf{KPi}^r + (\Sigma\text{-}\mathrm{FOUND})| =$

$|\mathbf{KPi}^r + (\Sigma\text{-}\mathrm{REC})| = \psi 0(\Phi\omega 0)$.

(vii) $|\Delta^1_2\text{-}\mathbf{TR}| = |\Sigma^1_2\text{-}\mathbf{TRDC}| = |\Delta^1_2\text{-}\mathbf{CA} + (\Sigma^1_2\text{-BI})| = |\mathbf{KPi}^w + (\Sigma\text{-FOUND})| =$

$|\mathbf{KPi}^w + (\Sigma\text{-REC})| = \psi 0(\Phi\varepsilon_0 0)$.

(viii) $|\Delta^1_2\text{-}\mathbf{TR} + \text{BR}(\text{impl-}\Sigma^1_2)| = |\Delta^1_2\text{-}\mathbf{TR} + \text{BI}(\text{impl-}\Sigma^1_2)| = |\Sigma^1_2\text{-}\mathbf{TRDC} + \text{BR}(\text{impl-}\Sigma^1_2)| =$

$|\Sigma^1_2\text{-}\mathbf{TRDC} + \text{BI}(\text{impl-}\Sigma^1_2)| = |\mathbf{KPi}^w + (\Sigma\text{-REC}) + \text{FOUNDR}(\text{impl-}\Sigma)| =$

$|\mathbf{KPi}^w + (\Sigma\text{-REC}) + \text{FOUND}(\text{impl-}\Sigma)| = \psi 0\Gamma^\Phi_0$.

(ix) $|\Delta^1_2\text{-}\mathbf{TR} + \text{BR}| = |\Sigma^1_2\text{-}\mathbf{TRDC} + \text{BR}| = |\mathbf{KPi}^w + (\Sigma\text{-REC}) + \text{FOUNDR}(\text{impl-}\Sigma(\text{M}))| =$

$\psi 0(\Phi\Omega_1 0)$.

(x) $|\Pi^1_1\text{-}\mathbf{TR} + \text{BR}| = |\mathbf{AUT\text{-}KPl}^w + \text{FOUNDR}(\text{impl-}\Sigma(\text{M}))| = \psi 0((\Phi 10) \cdot \Omega_1)$.

(xi) $|\Pi^1_1\text{-}\mathbf{TR} + \text{BR}(\text{impl-}\Sigma^1_2)| = |\mathbf{AUT\text{-}KPl}^w + \text{FOUNDR}(\text{impl-}\Sigma)| = \psi 0\omega^{(\Phi 10)+(\Phi 10)}$.

(xii) $|\Pi^1_1\text{-}\mathbf{TR} + \Delta^1_2\text{-}\mathbf{CA} + \text{BR}| = |\mathbf{AUT\text{-}KPl}^w + \mathbf{KPi}^w + \text{FOUNDR}(\text{impl-}\Sigma(\text{M}))| = \psi 0(\Phi 1\Omega_1)$.

(xiii) $|\Pi^1_1\text{-}\mathbf{TR} + \Delta^1_2\text{-}\mathbf{CA} + \text{BR}(\text{impl-}\Sigma^1_2)| = |\mathbf{AUT\text{-}KPl}^w + \mathbf{KPi}^w + \text{FOUNDR}(\text{impl-}\Sigma)| = \psi 0(\Phi 20)$.

References

[1] P. Aczel: *The Type Theoretic Interpretation of Constructive Set Theory: Inductive Definitions*, in: Marcus, R. B. et al. (eds), *Logic, Methodology, and Philosopy of Science VII* (North–Holland, Amsterdam, 1986).

[2] P. Aczel, M. Rathjen: *Notes on constructive set theory*, Technical Report 40, Institut Mittag-Leffler (The Royal Swedish Academy of Sciences, 2001). *http://www.ml.kva.se/preprints/archive2000-2001.php*

[3] J. Barwise: *Admissible Sets and Structures* (Springer, Berlin 1975).

[4] W. Buchholz: *Normalfunktionen und konstructive Systeme von Ordinalzahlen*, in: Lecture Notes in Mathematics 500 (Springer, Berlin, 1975).

[5] W. Buchholz: *Eine Erweiterung der Schnitteliminationsmethode*, Habilitationsschrift (München 1977).

[6] W. Buchholz, W. Pohlers: *Provable wellorderings of formal theories for transfinitely iterated inductive definitions*, Journal of Symbolic Logic 43 (1978) 118–125.

[7] W. Buchholz: *Ordinal analysis of* ID_ν, in [10] (Springer, Berlin, 1981) 234–260.

[8] W. Buchholz: *A new system of proof–theoretic ordinal functions,* Ann. Pure Appl. Logic **32** (1986) 195–207.

[9] W. Buchholz, K. Schütte: *Proof theory of impredicative subsystems of analysis* (Bibliopolis, Naples, 1988).

[10] W. Buchholz, S. Feferman, W. Pohlers, W. Sieg: *Iterated inductive definitions and subsystems of analysis* (Springer, Berlin, 1981).

[11] W. Buchholz and K. Schütte: *Proof theory of impredicative subsystems of analysis* (Bibliopolis, Naples, 1988).

[12] F.R. Drake: *How recent work in mathematical logic relates to the foundations of mathematics*, The Mathematical Intelligencer vol. 7, no. 4 (1985) 27–35.

[13] S. Feferman: *Systems of predicative analysis*, Journal of Symbolic Logic 29 (1964) 1–30.

[14] S. Feferman: *Systems of predicative analysis II. Representations of ordinals*, Journal of Symbolic Logic 33 (1968) 193–220.

[15] S. Feferman: *Formal theories for transfinite iterations of generalized inductive definitions and some subsystems of analysis*, in: J. Myhill, A. Kino, R.F. Vesley (eds.): *Proof theory and Intuitionism* (North Holland, Amsterdam, 1970) 303–326.

[16] S. Feferman: *A Language and Axioms for Explicit Mathematics*, Lecture Notes in Math. 450 (Springer, Berlin, 1975), 87–139.

[17] S. Feferman: *Theories of finite type related to mathematical practice*, in: J. Barwise (ed.): *Handbook of Mathematical Logic* (North Holland, Amsterdam 1977) 913–971.

[18] S. Feferman, W. Sieg: *Iterated inductive definitions and subsystems of analysis*, in: W. Buchholz, S. Feferman, W. Pohlers, W. Sieg: *Iterated inductive definitions and subsystems of analysis* (Springer, Berlin, 1981) 16–77.

[19] H. Friedman: *Iterated inductive definitions and* Σ^1_2-AC, in: J. Myhill, A. Kino, R.F. Vesley (eds.): *Proof theory and Intuitionism* (North Holland, Amsterdam, 1970) 435–442.

[20] H. Friedman, N. Robertson, P. Seymour: *The metamathematics of the graph minor theorem*, Contemporary Mathematics 65 (1987) 229–261.

[21] R.O. Gandy: *Proof of Mostowski's conjecture*, Bulletin of the Poslish Axademy of Science 8 (1960) 265-299.

[22] G. Gentzen: *Die Widerspruchsfreiheit der reinen Zahlentheorie.* Mathematische Annalen 112 (1936) 493–565.

[23] D. Hilbert: *Die Grundlegung der elementaren Zahlentheorie.* Mathematische Annalen 104 (1931).

[24] D. Hilbert and P. Bernays: *Grundlagen der Mathematik II.* (Springer, Berlin, 1938).

[25] P.G. Hinman: *Recursion-theoretic hierarchies.* (Springer, Berlin, 1978).

[26] D. Isles: *Regular ordinals and normal forms,* in: A. Kino, J. Myhill, R.E. Vesley (eds.): Intuitionism and proof theory (North-Holland, Amsterdam, 1968) 288–300.

[27] G. Jäger: *Die konstruktible Hierarchie als Hilfsmittel zur beweistheoretischen Untersuchung von Teilsystemen der Analysis.* Dissertation, (Universität München, 1979).

[28] G. Jäger: *Beweistheorie von KPN*, Archiv f. Math. Logik 2 (1980) 53–64.

[29] G. Jäger: *Zur Beweistheorie der Kripke–Platek Mengenlehre über den natürlichen Zahlen*, Archiv f. Math. Logik 22 (1982) 121–139.

[30] G. Jäger: *Iterating admissibility in proof theory*, in: J. Stern (ed.): *Proceedings of the Herbrand Logic Colloquium '81*, (North-Holland, Amsterdam, 1982) 137–146.

[31] G. Jäger: *A well-ordering proof for Feferman's theory* T_0, Archiv f. Math. Logik 23 (1983) 65–77.

[32] Jäger, G.: *Theories for admissible sets: a unifying approach to proof theory* (Bibliopolis, Naples, 1986).

[33] G. Jäger and W. Pohlers: *Eine beweistheoretische Untersuchung von* Δ^1_2-$\mathbf{CA}+\mathbf{BI}$ *und verwandter Systeme*, Sitzungsberichte der Bayerischen Akademie der Wissenschaften, Mathematisch–Naturwissenschaftliche Klasse (1982).

[34] S.C. Kleene: *Arithmetical predicates and function quantifiers*, Transactions of the American Mathematical Society 79 (1955) 312–340.

[35] G. Kreisel: *On the interpretation of non-finitist proofs II,* Journal of Symbolic Logic 17 (1952) 43–58.

[36] G. Kreisel: *Mathematical significance of consistency proofs.* Journal of Symbolic Logic 23 (1958) 155–182.

[37] G. Kreisel: *Generalized inductive definitions,* in: Stanford Report on the Foundations of Analysis (Mimeographed, Stanford, 1963) Section III.

[38] P. Martin-Löf: *Intuitionistic Type Theory*, (Bibliopolis, Naples 1984).

[39] H. Pfeiffer: *Ausgezeichnete Folgen für gewisse Abschnitte der zweiten und weiterer Zahlklassen* (Dissertation, Hannover, 1964).

[40] W. Pohlers: *Cut elimination for impredicative infinitary systems, part I: Ordinal analysis of* ID_1, Arch. f. Math. Logik 21 (1981) 69–87.

[41] W. Pohlers: *Proof–theoretical analysis of* ID_ν *by the method of local predicativity*, in: W. Buchholz, S. Feferman, W. Pohlers, W. Sieg: *Iterated inductive definitions and subsystems of analysis* (Springer, Berlin, 1981) 261–357.

[42] W. Pohlers: *Cut elimination for impredicative infinitary systems, part II: Ordinal analysis for iterated inductive definitions*, Arch. f. Math. Logik 22 (1982) 113–129.

[43] M. Rathjen: *Untersuchungen zu Teilsystemen der Zahlentheorie zweiter Stufe und der Mengenlehre mit einer zwischen* Δ^1_2-CA *und* $\Delta^1_2-CA+BI$ *liegenden Beweisstärke*, PhD thesis, University of Münster, 1988.

[44] M. Rathjen: *Ordinal notations based on a weakly Mahlo cardinal*, Archive for Mathematical Logic 29 (1990) 249–263.

[45] M. Rathjen: *Proof-Theoretic Analysis of KPM*, Arch. Math. Logic 30 (1991) 377–403.

[46] M. Rathjen: *The role of parameters in bar rule and bar induction,* Journal of Symbolic Logic 56 (1991) 715–730.

[47] M. Rathjen: *Fragments of Kripke-Platek set theory with infinity,* in: P. Aczel, H. Simmons, S. Wainer (eds.): *Proof Theory* (Cambridge University Press, Cambridge, 1992) 251–273.

[48] M. Rathjen: *How to develop proof–theoretic ordinal functions on the basis of admissible sets.* Mathematical Quarterly 39 (1993) 47–54.

[49] M. Rathjen and A. Weiermann: *Proof–theoretic investigations on Kruskal's theorem,* Annals of Pure and Applied Logic 60 (1993) 49–88.

[50] M. Rathjen: *Admissible proof theory and beyond,* in: Logic, Methodology and Philosophy of Science IX (D. Prawitz, B. Skyrms and D. Westerstahl, eds.), Elsevier Science B.V. (1994) 123–147.

[51] M. Rathjen: *The strength of some Martin–Löf type theories.* Archive for Mathematical Logic 33 (1994) 347–385.

[52] M. Rathjen: *Collapsing functions based on recursively large ordinals: A well–ordering proof for KPM.* Archive for Mathematical Logic 33 (1994) 35–55.

[53] M. Rathjen: *Proof theory of reflection.* Annals of Pure and Applied Logic 68 (1994) 181–224.

[54] M. Rathjen: *Recent advances in ordinal analysis:* Π^1_2*-CA and related systems.* Bulletin of Symbolic Logic 1 (1995) 468–485.

[55] M. Rathjen: *The higher infinite in proof theory,* in: J.A. Makowsky and E.V. Ravve (eds.): *Logic Colloquium '95.* Lecture Notes in Logic, vol. 11 (Springer, New York, Berlin, 1998) 275–304.

[56] D. Schmidt: *Well-partial orderings and their maximal order types*, Habilitationsschrift (Heidelberg, 1979) 77 pages.

[57] K. Schütte: *Beweistheoretische Erfassung der unendlichen Induktion in der Zahlentheorie*, Mathematische Annalen 122 (1951) 369–389.

[58] K. Schütte: *Beweistheorie* (Springer, Berlin, 1960).

[59] K. Schütte: *Eine Grenze für die Beweisbarkeit der transfiniten Induktion in der verzweigten Typenlogik*, Archiv für Mathematische Logik und Grundlagenforschung 67 (1964) 45–60.

[60] K. Schütte: *Predicative well-orderings*, in: Crossley, Dummett (eds.), Formal systems and recursive functions (North Holland, 1965) 176–184.

[61] K. Schütte: *Proof Theory* (Springer, Berlin, 1977).

[62] H. Schwichtenberg: *Proof theory: Some applications of cut-elimination*. In: J. Barwise (ed.): *Handbook of Mathematical Logic* (North Holland, Amsterdam, 1977) 867–895.

[63] J. Shoenfield: *Mathematical logic* (Addison-Wesley, Reading Mass., 1967).

[64] S. Simpson: *Nichtbeweisbarkeit von gewissen kombinatorischen Eigenschaften endlicher Bäume*, Archiv f. Math. Logik 25 (1985) 45–65.

[65] S.G. Simpson: *Subsystems of Second Order Arithmetic* (Springer-Verlag, Berlin, Heidelberg, 1999)

[66] G. Spector: *Inductively defined sets of numbers*, in Infinitistic methods, Proceedings of the Warszaw Symposium (Pergamon Press, Oxford, 1961) 97–102.

[67] G. Takeuti: *Consistency proofs of subsystems of classical analysis*, Ann. Math. 86, 299–348.

[68] G. Takeuti, M. Yasugi: *The ordinals of the systems of second order arithmetic with the provably Δ^1_2–comprehension and the Δ^1_2–comprehension axiom respectively*, Japan J. Math. 41 (1973) 1–67.

[69] G. Takeuti: *Proof theory*, second edition (North Holland, Amsterdam, 1987).

[70] O. Veblen: *Continuous increasing functions of finite and transfinite ordinals,* Trans. Amer. Math. Soc. 9 (1908) 280–292.

[71] A. Weiermann: *Beweistheoretische Untersuchungen zur Theorie* $\mathbf{ID}_{\prec^*}$, (Diploma Thesis, Universität Münster, 1987).

[72] A. Weiermann: *How to characterize provably total functions by local predicativity,* Journal of Symbolic Logic 61 (1996) 52–69.

Weak Theories of Operations and Types

Thomas Strahm

Institut für Informatik und angewandte Mathematik,
Universität Bern,
Neubrückstrasse 10,
CH-3012 Bern, Switzerland
strahm@iam.unibe.ch

Abstract This is a survey paper on various weak systems of Feferman's explicit mathematics and their proof theory. The strength of the systems considered in measured in terms of their provably terminating operations typically belonging to some natural classes of computational time or space complexity.

Keywords: Proof theory, Feferman's explicit mathematics, applicative theories, higher types, types and names, partial truth, feasible operations

1 Introduction

In this article we survey recent results about a proof-theoretic approach to computational complexity via theories of operations and types in the sense of Feferman's explicit mathematics. The latter framework was introduced by Feferman [19–21] in the early 1970s. Beyond its original aim to provide a basis for Bishop-style constructivism, the explicit framework has gained considerable importance in proof theory in connection with the proof-theoretic analysis of subsystems of second order arithmetic and set theory as well as for studying the proof theory of abstract computations.

It is this latter focus which is most important in the present article. The operational or applicative core of explicit mathematics includes forms of combinatory logic and hence comprises a computationally complete functional language with the full defining power of the untyped lambda calculus. In this sense it is more expressive than standard arithmetical systems.

Apart from *operations or rules*, the second basic entity in explicit mathematics are *types*, which can be thought of as successively generated collections of operations. In addition, and this is essential in the explicit approach, extensional types are represented (or named) by intensional operations, uniformly in their parameters. This interplay of operations and types on the level of representations makes explicit mathematics extremely powerful.

As an alternative means of enhancing the first order part of explicit mathematics, we will also consider extensions of applicative theories by a partial truth predicate, leading to an expressive language embodying naive set theory. In this connection, we will review work done by Cantini.

Let us briefly outline the content of the paper. We omit references and credits and refer the reader to the corresponding sections of the paper.

We start off in Section 2 by introducing the first order applicative framework being based on the logic of partial terms. We define the basic theory of operations and words B and introduce two bounded induction schemas on the binary words.

In Section 3 we provide a review of function algebra characterizations of complexity classes and introduce four bounded applicative systems, PT, PTLS, PS, and LS, whose provably total functions coincide with the functions computable in polynomial time, simultaneously polynomial time and linear space, polynomial space, and linear space. We briefly address the lower and upper bound arguments for these systems. In particular, we outline a specific combination of partial cut elimination and a realizability interpretation.

Section 4 addresses higher type issues of the first-order system PT. It is a distinguished advantage of applicative theories that they allow for a very intrinsic and direct discussion of higher type aspects, since higher types arise naturally in the untyped setting. It makes perfect sense to consider the class of higher type functionals which are provably total in a given applicative system. We will discuss the relationship between PT and the Melhorn-Cook-Urquhart basic feasible functionals BFF.

In Section 5 we introduce a theory PET of polynomial time operations with explicit and variable types which is formulated in the full language of explicit mathematics and embodies a weak form of elementary type comprehension. The provably total operations of PET are still the polynomial time computable functions on binary words. We will also consider various extensions of PET by choice, quantification, uniformity and join principles.

In Section 6 we review work by Cantini on extensions of weak applicative theories by forms of self-referential truth with choice and uniformity, which has been essential in obtaining results about corresponding extensions of the system PET.

Finally, in Section 7 we address self-applicative systems proposed by Cantini and Calamai in the realm of so-called implicit computational complexity in the sense of Bellantoni, Cook and Leivant. It turns out that forms of safe induction formulated in a modal language provide very natural applicative characterizations of the functions computable in polynomial time and polynomial space.

We conclude this article by some comments on the relationship between primitive recursion and positive induction.

2 The axiomatic framework

In this section we first describe the informal setting of applicative systems and briefly motivate their underlying logic of partial terms. Then we outline the basic applicative language and theory of operations and words and mention some of its basic consequences and models. We conclude this section by specifying two important induction principles.

2.1 The informal applicative setting

Let us assume that we are given an untyped universe of operations or rules, which can be freely applied to each other. Self-application is meaningful, though not necessarily total. The computational engine of these rules is given by a partial combinatory algebra, featuring partial versions of Curry's combinators k and s. In addition, there is a ground "urelement" structure of the binary words or strings with certain natural operations on them.

Let $\mathbb{W}$ denote the set of (finite) binary words. We will consider the following operations on $\mathbb{W}$:

- s_0 and s_1: binary successors on $\mathbb{W}$ with predecessor p_W
- s_ℓ: (unary) lexicographic successor on $\mathbb{W}$ with predecessor p_ℓ
- $*$: word concatenation
- $\times$: word multiplication

Here s_ℓ denotes the successor in the ordering $<_\ell$ which orders words by length and words of the same length lexicographically. Moreover, $x \times y$ signifies the length of y fold concatenation of x with itself.

2.2 The logic of partial terms

All our theories considered in this survey are based on the classical logic of partial terms (LPT) due to Beeson and Feferman. It is is a modification of first-order predicate logic taking into account partial functions, cf. Beeson [1, 2] and Troelstra and van Dalen [52] for details. It is assumed that variables range over defined objects only. (Composed) terms do not necessarily denote and $t\downarrow$ signifies that t has a value or t denotes. The usual quantifier axioms of predicate logic are modified, e.g. we have

$$A(t) \ \wedge \ t\downarrow \ \rightarrow \ (\exists x)A(x)$$

Moreover, strictness axioms claim that subterms of a defined term are defined and that terms occurring in true positive atoms are defined.

For an excellent survey of logics of definedness the reader is referred to Feferman [22]. Feferman distinguishes between logics of existence and logics of partial terms in the above-explained sense, whereas the former were pioneered by Scott [41]. On the other hand, the pseudo-applicative terms used in Feferman [19, 21] may be considered as precursors to the logic of partial terms.

2.3 The basic applicative language

Our basic language L is a first order language for the logic of partial terms which includes:

- variables $a, b, c, x, y, z, u, v, f, g, h, \ldots$
- constants k, s, p, $\mathsf{p_0}$, $\mathsf{p_1}$, $\mathsf{d_W}$, ϵ, $\mathsf{s_0}$, $\mathsf{s_1}$, $\mathsf{p_W}$, s_ℓ, p_ℓ, $\mathsf{c}_\subseteq$, $\mathsf{l_W}$, $\ldots$
- relation symbols $=$ (equality), $\downarrow$ (definedness), W (binary words)
- arbitrary term application $\circ$

The meaning of the constants will become clear in the next paragraph.

The terms $(r, s, t, \ldots)$ and formulas $(A, B, C, \ldots)$ of L are defined in the expected manner. We assume the following standard abbreviations and syntactical conventions:

$$\begin{aligned} t_1 t_2 \ldots t_n &:= (\ldots(t_1 \circ t_2) \circ \cdots \circ t_n) \\ t_1 \simeq t_2 &:= t_1\downarrow \ \vee \ t_2\downarrow \ \rightarrow \ t_1 = t_2 \\ t \in \mathsf{W} &:= \mathsf{W}(t) \\ t : \mathsf{W}^k \rightarrow \mathsf{W} &:= (\forall x_1 \ldots x_k \in \mathsf{W}) t x_1 \ldots x_k \in \mathsf{W} \end{aligned}$$

$$t : \mathsf{W}^{\mathsf{W}} \times \mathsf{W} \to \mathsf{W} := (\forall f \in \mathsf{W} \to \mathsf{W})(\forall x \in \mathsf{W}) tfx \in \mathsf{W}$$

Finally, let us write $\overline{w}$ for the canonical closed L term denoting the binary word $w \in \mathbb{W}$.

2.4 The basic theory of operations and words B

The applicative base theory B has been introduced in Strahm [47, 48]. Its logic is the *classical* logic of partial terms. The non-logical axioms of B include:

- partial combinatory algebra:

$$\mathsf{k}xy = x, \qquad \mathsf{s}xy\downarrow \;\wedge\; \mathsf{s}xyz \simeq xz(yz)$$

- pairing p with projections p_0 and p_1
- defining axioms for the binary words W with ϵ, the successors s_0, s_1, s_ℓ an the predecessor p_{W} and and p_ℓ
- definition by cases d_{W} on W
- initial subword relation $\mathsf{c}_\subseteq$, tally length of words l_{W}

These axioms are fully spelled out in [47, 48]. The term $(t_1, t_2, \ldots, t_n)$ for n-tupling is defined as usual by iterating the pairing operation p.

Let us turn to the crucial consequences of the axioms about a partial combinatory algebra. For proofs of these standard results, the reader is referred to Beeson [1] or Feferman [19].

Lemma 2.1 (Explicit definitions and fixed points)**.**

1. *For each* L *term* t *there exists an* L *term* $(\lambda x.t)$ *so that*

$$\mathsf{B} \vdash (\lambda x.t)\downarrow \;\wedge\; (\lambda x.t)x \simeq t$$

2. *There is a closed* L *term* fix *so that*

$$\mathsf{B} \vdash \mathsf{fix}g\downarrow \;\wedge\; \mathsf{fix}gx \simeq g(\mathsf{fix}g)x$$

Let us quickly remind the reader of two standard models of B, namely the recursion-theoretic model PRO and the term model $\mathcal{M}(\lambda\eta)$. For an extensive discussion of many more models of the applicative basis, the reader is referred to Beeson [1] and Troelstra and van Dalen [53].

Example 2.2 (Recursion-theoretic model PRO)**.** Take the universe of binary words and interpret application $\circ$ as partial recursive function application in the sense of ordinary recursion theory.

Example 2.3 (The open term model $\mathcal{M}(\lambda\eta)$)**.** Take the universe of open λ terms and consider the usual reduction of the extensional untyped lambda calculus $\lambda\eta$, augmented by suitable reduction rules for the constants other than k and s. Interpret application as juxtaposition. Two terms are equal if they have a common reduct and W denotes those terms that reduce to a "standard" word $\overline{w}$.

2.5 Natural induction principles

We have not yet specified induction principles on the binary words W; these are of course crucial for our proof-theoretic characterizations of complexity classes below.

We call an L formula *positive* if it is built from the atomic formulas by means of disjunction, conjunction as well as existential and universal quantification over individuals. We let Pos stand for the collection of positive formulas. Further, an L formula is called W *free*, if the relation symbol W does not occur in it.

Most important in the sequel are the so-called *bounded (with respect to* W*) existential formulas* or $\Sigma^{\mathsf{b}}_{\mathsf{W}}$ *formulas* of L. A formula $A(f, x)$ belongs to the class $\Sigma^{\mathsf{b}}_{\mathsf{W}}$ if it has the form $(\exists y \leq fx)B(f, x, y)$ for $B(f, x, y)$ a *positive and* W *free* formula. It is important to note here that bounded quantifiers range over W, i.e., $(\exists y \leq fx)B(f, x, y)$ stands for

$$(\exists y \in \mathsf{W})[y \leq fx \ \wedge \ B(f, x, y)].$$

Further observe that the matrix B of a $\Sigma^{\mathsf{b}}_{\mathsf{W}}$ formula can have unrestricted existential and universal individual quantifiers, not ranging over W, however.

Below we will distinguish usual notation induction on binary words and the corresponding "slow" induction principle with respect to the lexicographic successor s_ℓ.

$\Sigma^{\mathsf{b}}_{\mathsf{W}}$ **notation induction on** W**:**

For each $\Sigma^{\mathsf{b}}_{\mathsf{W}}$ formula $A(x) \equiv (\exists y \leq fx)B(f, x, y)$,

$$\begin{aligned} f : \mathsf{W} \to \mathsf{W} \ \wedge \ A(\epsilon) \ \wedge \ (\forall x \in \mathsf{W})(A(x) \to A(\mathsf{s}_0 x) \wedge A(\mathsf{s}_1 x)) \\ \to (\forall x \in \mathsf{W})A(x) \end{aligned} \tag{$\Sigma^{\mathsf{b}}_{\mathsf{W}}$-$\mathsf{I}_{\mathsf{W}}$}$$

$\Sigma^{\mathsf{b}}_{\mathsf{W}}$ **lexicographic induction on** W**:**

For each $\Sigma^{\mathsf{b}}_{\mathsf{W}}$ formula $A(x) \equiv (\exists y \leq fx)B(f, x, y)$,

$$f : \mathsf{W} \to \mathsf{W} \ \wedge \ A(\epsilon) \ \wedge \ (\forall x \in \mathsf{W})(A(x) \to A(\mathsf{s}_\ell x)) \\ \to (\forall x \in \mathsf{W})A(x) \qquad (\Sigma^{\mathsf{b}}_{\mathsf{W}}\text{-}\mathsf{I}_\ell)$$

It is now easy, by making use of the fixed point theorem and $\Sigma^{\mathsf{b}}_{\mathsf{W}}$ notation induction on W, to show the existence of a type two functional for bounded recursion on notation, provably in $\mathsf{B} + (\Sigma^{\mathsf{b}}_{\mathsf{W}}\text{-}\mathsf{I}_{\mathsf{W}})$. This is the content of the following lemma whose detailed proof can be found in Strahm [48].

Lemma 2.4 (Bounded recursion on notation)**.** *There exists a closed* L *term* r_{W} *so that* $\mathsf{B} + (\Sigma^{\mathsf{b}}_{\mathsf{W}}\text{-}\mathsf{I}_{\mathsf{W}})$ *proves*

$$f : \mathsf{W} \to \mathsf{W} \ \wedge \ g : \mathsf{W}^3 \to \mathsf{W} \ \wedge \ b : \mathsf{W}^2 \to \mathsf{W} \ \to$$

$$\begin{cases} \mathsf{r}_{\mathsf{W}} fgb : \mathsf{W}^2 \to \mathsf{W} \ \wedge \\ x \in \mathsf{W} \ \wedge \ y \in \mathsf{W} \ \wedge \ y \neq \epsilon \ \wedge \ h = \mathsf{r}_{\mathsf{W}} fgb \to \\ \qquad hx\epsilon = fx \ \wedge \ hxy = gxy(hx(\mathsf{p}_{\mathsf{W}} y)) \mid bxy \end{cases}$$

Here $t \mid s$ is t if $t \leq s$ and s otherwise.

Similarly, bounded lexicographic recursion is derivable in $\mathsf{B} + (\Sigma^{\mathsf{b}}_{\mathsf{W}}\text{-}\mathsf{I}_\ell)$, see Strahm [48] for details.

3 Characterizing complexity classes

We now turn to the characterization of complexity classes by means of our applicative systems. We start our discussion by reviewing some function algebra characterizations of complexity classes and then propose four applicative systems, PT, PTLS, PS, and LS, whose provably total functions coincide with the functions computable in *polynomial time*, *simultaneously polynomial time and linear space*, *polynomial space*, and *linear space*. We sketch lower and upper bounds for these proof-theoretic characterizations.

3.1 Four function algebras

In this subsection we review know recursion-theoretic characterizations of various classes of computational complexity. Our main interest in the sequel are the functions on $\mathbb{W}$ which are computable on a Turing machine in *polynomial time*, *simultaneously polynomial time and linear space*, *polynomial space*, and *linear space*.

In the following we let FPTIME, FPTIMELINSPACE, FPSPACE, and FLINSPACE denote the respective classes of functions on binary words $\mathbb{W}$. For an extensive discussion of recursion-theoretic or function algebra characterizations of complexity classes the reader is referred to the survey article Clote [15].

In the following we use the notation of Clote [15] for a compact representation of function algebras. Accordingly, we call (partial) mappings from functions on $\mathbb{W}$ to functions on $\mathbb{W}$ *operators*. If $\mathcal{X}$ is a set of functions on $\mathbb{W}$ and OP is a collection of operators, then $[\mathcal{X}; \mathsf{OP}]$ is used to denote the smallest set of functions containing $\mathcal{X}$ and closed under the operators in OP. We call $[\mathcal{X}; \mathsf{OP}]$ a *function algebra*. Our crucial examples of operators in the sequel are *bounded recursion on notation* BRN and *bounded lexicographic recursion* BRL, cf. Strahm [48] for details. A further operator is the *composition operator* COMP. Below we also use I for the usual collection of projection functions and we simply write ϵ for the 0-ary function being constant to the empty word ϵ.

We are now ready to state the function algebra characterizations of the four complexity classes which are relevant in this paper. The characterization of FPTIME is due to Cobham [16]. The delineations of FPTIMELINSPACE and FPSPACE are due to Thompson [51]. Finally, the fourth assertion of our theorem is due to Ritchie [38]. For a uniform presentation of all these results we urge the reader to consult Clote [15].

Theorem 3.1. *We have the following function algebra characterizations of the complexity classes mentioned above:*

1. $[\epsilon, \mathsf{I}, \mathsf{s}_0, \mathsf{s}_1, *, \times; \mathsf{COMP}, \mathsf{BRN}] =$ FPTIME.
2. $[\epsilon, \mathsf{I}, \mathsf{s}_0, \mathsf{s}_1, *; \mathsf{COMP}, \mathsf{BRN}] =$ FPTIMELINSPACE.
3. $[\epsilon, \mathsf{I}, \mathsf{s}_\ell, *, \times; \mathsf{COMP}, \mathsf{BRL}] =$ FPSPACE.
4. $[\epsilon, \mathsf{I}, \mathsf{s}_\ell, *; \mathsf{COMP}, \mathsf{BRL}] =$ FLINSPACE.

We now turn to the proof-theoretic characterization of the above four complexity classes by means of suitable applicative theories.

3.2 Provably total functions

Let us first start with a formal definition of the notion of *provably total function* of a given L theory.

Definition 3.2. A function $F : \mathbb{W}^n \to \mathbb{W}$ is called *provably total in an* L *theory* T, if there exists a closed L term t_F such that

(i) $\mathsf{T} \vdash t_F : \mathsf{W}^n \to \mathsf{W}$ and, in addition,

(ii) $\mathsf{T} \vdash t_F \overline{w}_1 \cdots \overline{w}_n = \overline{F(w_1, \ldots, w_n)}$ for all $w_1, \ldots, w_n$ in $\mathbb{W}$.

The notion of a provably total word function is divided into two conditions (i) and (ii). The first condition (i) expresses that t_F is a total operation from W^n to W, *provably in the* L *theory* T. Condition (ii), on the other hand, claims that t_F indeed represents the given function $F : \mathbb{W}^n \to \mathbb{W}$, for each fixed word w in $\mathbb{W}$.

In the sequel, let $\boldsymbol{\tau}(\mathsf{T}) = \{F : F \text{ provably total in } \mathsf{T}\}$.

3.3 Four applicative systems

In the following we write $\mathsf{B}(*)$ for the extension of B by the obvious axioms about word concatenation on W, namely the standard recursive defining equations and the totality of $*$ on W. We assume that $*$ is a new constant of our applicative language L. Similarly, $\mathsf{B}(*, \times)$ extends $\mathsf{B}(*)$ by the standard axioms about word multiplication. For details, see Strahm [48].

Depending on whether we include $(\Sigma^{\mathsf{b}}_{\mathsf{W}}\text{-}\mathsf{I}_{\mathsf{W}})$ or $(\Sigma^{\mathsf{b}}_{\mathsf{W}}\text{-}\mathsf{I}_{\ell})$, and whether we assume as given only word concatenation or both word concatenation and word multiplication, we can now distinguish the following four applicative theories PT, PTLS, PS, and LS:

$$\mathsf{PT} := \mathsf{B}(*, \times) + (\Sigma^{\mathsf{b}}_{\mathsf{W}}\text{-}\mathsf{I}_{\mathsf{W}}) \qquad \mathsf{PTLS} := \mathsf{B}(*) + (\Sigma^{\mathsf{b}}_{\mathsf{W}}\text{-}\mathsf{I}_{\mathsf{W}})$$
$$\mathsf{PS} := \mathsf{B}(*, \times) + (\Sigma^{\mathsf{b}}_{\mathsf{W}}\text{-}\mathsf{I}_{\ell}) \qquad \mathsf{LS} := \mathsf{B}(*) + (\Sigma^{\mathsf{b}}_{\mathsf{W}}\text{-}\mathsf{I}_{\ell})$$

We note that a preliminary, more restrictive version of the system PT has previously been studied in Strahm [46] and Cantini [12].

In the sequel let us briefly sketch the lower and upper bound arguments for our applicative systems, which are worked out in full detail in Strahm [48].

3.4 Lower bounds

The lower bounds for our four applicative systems directly follow from Theorem 3.1 and the crucial Lemma 2.4, respectively its variant for bounded lexicographic recursion.

Theorem 3.3. *We have the following lower bounds:*

1. FPTIME $\subseteq \boldsymbol{\tau}(\mathsf{PT})$.

2. FPTIMELINSPACE $\subseteq \boldsymbol{\tau}(\mathsf{PTLS})$.

3. FPSPACE $\subseteq \boldsymbol{\tau}(\mathsf{PS})$.

4. FLINSPACE $\subseteq \boldsymbol{\tau}(\mathsf{LS})$.

Let us close this paragraph with the following remarks:

Remarks 3.4. 1. Ferreira's system PTCA^+ ([24], [25]) is directly contained in PT, where PTCA^+ corresponds to Buss' S^1_2 ([5]).

2. The Melhorn-Cook-Urquhart basic feasible functionals resp. the system PV^ω ([18]) are directly contained in PT (see Section 4).

3.5 Partial cut elimination

In order to extract computational content from proofs, we need a sequent-style reformulation of our systems and a preparatory partial cut-elimination result. It is employed in order to show that as far as the computational content of our systems is concerned, we can restrict ourselves to positive derivations, i.e., sequent style proofs using positive formulas only. Moreover, we will establish upper bounds directly for an extension of our systems by the axioms of *totality of application* and *extensionality of operations*:

Totality of application:

$$(\forall x)(\forall y)(xy\downarrow) \qquad \textbf{(Tot)}$$

Extensionality of operations:

$$(\forall f)(\forall g)[(\forall x)(fx \simeq gx) \rightarrow f = g] \qquad \textbf{(Ext)}$$

Observe that in the presence of the totality axiom, the logic of partial terms reduces to ordinary classical predicate logic. Accordingly, if T denotes one of the systems PT, PTLS, PS, or LS, then we write T^+ for the system T based on ordinary classical logic with equality and augmented with the axiom of extensionality.

In the following we let $\Gamma, \Delta, \Lambda, \dots$ range over finite *sequences* of formulas; a *sequent* is a formal expression of the form $\Gamma \Rightarrow \Delta$. As usual, the natural interpretation of the sequent $A_1, \dots, A_n \Rightarrow B_1, \dots, B_m$ is $(A_1 \wedge \cdots \wedge A_n) \rightarrow (B_1 \vee \cdots \vee B_m)$.

It is now a matter of routine to spell out a sequent-style version of our four applicative systems so that all the main formulas of axioms and rules are *positive*. Hence, partial cut-elimination is applicable in order to show that cuts can be restricted to positive formulas. In the following we write $\mathsf{T}^+ \vdash_{\star} \Gamma \Rightarrow \Delta$ to express that $\Gamma \Rightarrow \Delta$ has a proof in T^+ where all cut formulas are positive.

3.6 Realizability

In a second crucial step we use a notion of *realizability for positive formulas* in the standard open term model of our systems: quasi cut-free positive sequent derivations of PT, PTLS, PS, and LS are suitably realized by word functions in FPTIME, FPTIMELINSPACE, FPSPACE, and FLINSPACE, respectively, thus yielding the desired computational information concerning the provably total functions of these systems.

The notion of realizability as well as the style and spirit of our realizability theorems are related to the work of Leivant [31], Schlüter [40], and Cantini [13], all three in the context of FPTIME. However, in contrast to these papers, we work in a bounded unramified setting. Moreover, and this is similar to [13, 40], we are able to realize directly quasi cut-free positive derivations in the *classical* sequent calculus. Finally, in order to find our realizing functions, we can make direct use of the function algebra characterizations of FPTIME, FPTIMELINSPACE, FPSPACE, and FLINSPACE given in Theorem 3.1; hence, direct reference to a machine model is not needed.

In fact, the above mentioned literature on realizability in an applicative context, especially in the classical setting, is clearly related to and inspired by older work on *witnessing* that has been used in classical fragments of arithmetic. In particular, Buss' witnessing technique (cf. Buss [5–7]) has been employed with great success in a variety of contexts.

We are now ready to turn to realizability. Our realizers $\rho, \sigma, \tau, \ldots$ are simply elements of the set $\mathbb{W}$ of binary words. We presuppose a low-level pairing operation $\langle \cdot, \cdot \rangle$ on $\mathbb{W}$ with associated projections $(\cdot)_0$ and $(\cdot)_1$; for definiteness, we assume that $\langle \cdot, \cdot \rangle$, $(\cdot)_0$, and $(\cdot)_1$ are in FPTIMELINSPACE. Further, for each natural number i let us write i_2 for the binary notation of i.

Since we are only interested in realizing *positive* derivations, we need to define realizability for positive formulas only.

Definition 3.5. The notion $\rho \ \mathbf{r} \ A$ ("ρ realizes A") for $\rho \in \mathbb{W}$ and A a positive formula, is given inductively in the following manner:

$$\begin{array}{lll}
\rho \ \mathbf{r} \ \mathsf{W}(t) & \text{if} & \mathcal{M}(\lambda\eta) \models t = \overline{\rho}, \\
\rho \ \mathbf{r} \ (t_1 = t_2) & \text{if} & \rho = \epsilon \text{ and } \mathcal{M}(\lambda\eta) \models t_1 = t_2, \\
\rho \ \mathbf{r} \ (A \wedge B) & \text{if} & \rho = \langle \rho_0, \rho_1 \rangle \text{ and } \rho_0 \ \mathbf{r} \ A \text{ and } \rho_1 \ \mathbf{r} \ B, \\
\rho \ \mathbf{r} \ (A \vee B) & \text{if} & \rho = \langle i, \rho_0 \rangle \text{ and either } i = 0 \text{ and } \rho_0 \ \mathbf{r} \ A \text{ or} \\
& & \qquad\qquad\qquad i = 1 \text{ and } \rho_0 \ \mathbf{r} \ B, \\
\rho \ \mathbf{r} \ (\forall x)A(x) & \text{if} & \rho \ \mathbf{r} \ A(u) \text{ for a fresh variable } u, \\
\rho \ \mathbf{r} \ (\exists x)A(x) & \text{if} & \rho \ \mathbf{r} \ A(t) \text{ for some term } t.
\end{array}$$

If Δ denotes a sequence $A_1, \dots, A_n$, then $\rho \, \mathbf{r} \, \Delta$ iff $\rho = \langle i_2, \rho_0 \rangle$ for some $1 \le i \le n$ and $\rho_0 \ \mathbf{r} \ A_i$.

The next main lemma about the realizability of quasi-normal PT^+ derivations immediately entails that the provably total functions of PT^+ are computable in polynomial time. The lemma is proved in all detail in Strahm [48].

In the formulation of the lemma, we need the following notation. For an L formula A we write $A[\vec{u}]$ in order to express that all the free variables occurring in A are contained in the list $\vec{u}$. The analogous convention is used for finite sequences of L formulas.

Lemma 3.6 (Realizability for PT^+)**.** *Let $\Gamma \Rightarrow \Delta$ be a sequent of formulas in* Pos *with $\Gamma = A_1, \dots, A_n$ and assume that* $\mathsf{PT}^+ \mid\!\frac{\;}{\star}\; \Gamma[\vec{u}] \Rightarrow \Delta[\vec{u}]$*. Then there exists a function $F : \mathbb{W}^n \to \mathbb{W}$ in* FPTIME *so that we have for all terms $\vec{s}$ and all $\rho_1, \dots, \rho_n \in \mathbb{W}$:*

$$\textit{For all } 1 \le i \le n : \rho_i \ \mathbf{r} \ A_i[\vec{s}] \quad \Longrightarrow \quad F(\rho_1, \dots, \rho_n) \ \mathbf{r} \ \Delta[\vec{s}].$$

Analogous realizability results hold for the systems PTLS, PS, and LS, cf. [48] for details.

3.7 The main theorem concluded

We are now able to piece together the results of Sections 3.4, 3.5 and 3.6 and obtain the following main theorem.

Theorem 3.7. *We have the following characterizations:*

1. $\boldsymbol{\tau}(\mathsf{PT}) = \text{FPTIME}$.

2. $\tau(\mathsf{PTLS}) =$ FPTIMELINSPACE.

3. $\tau(\mathsf{PS}) =$ FPSPACE.

4. $\tau(\mathsf{LS}) =$ FLINSPACE.

In the next section we turn to some higher types aspects of the system PT.

4 Higher type issues

In the last two decades intense research efforts have been made in the area of so-called higher type complexity theory and, in particular, feasible functionals of higher types. This research is still ongoing and it is not yet clear what the right higher type analogue of the polynomial time computable functions is. Most prominent in the previous research is the class of so-called *basic feasible functionals* BFF, which has proved to be a very robust class with various kinds of interesting characterizations.

The basic feasible functionals of type 2, BFF_2, were first studied in Melhorn [34]. More than ten years later in 1989, Cook and Urquhart [18] introduced the basic feasible functionals at all finite types in order to provide functional interpretations of feasibly constructive arithmetic; in particular, they defined a typed formal system PV^ω and used it to establish functional and realizability interpretations of an intuitionistic version of Buss' theory S^1_2. The basic feasible functionals BFF are exactly those functionals which can be defined by PV^ω terms. Subsequently, much work has been devoted to BFF, cf. e.g. Cook and Kapron [17, 30], Irwin, Kapron and Royer [27], Pezzoli [37], Royer [39], and Seth [42].

In the following let us briefly discuss the relationship of PV^ω with our first-order applicative theory PT.

4.1 Higher types in the language $\mathbf{L}$

The collection $\mathcal{T}$ *of finite type symbols* $(\alpha, \beta, \gamma, \ldots)$ is inductively generated by the usual clauses, (i) $0 \in \mathcal{T}$, (ii) if $\alpha, \beta \in \mathcal{T}$, then $(\alpha \times \beta) \in \mathcal{T}$, and (iii) if $\alpha, \beta \in \mathcal{T}$, then $(\alpha \to \beta) \in \mathcal{T}$. Hence, we have product and function types as usual. Observe, however, that in our setting the ground type 0 stands for the set of binary words and not for the set of natural numbers. We use the usual convention and write $\alpha_1 \to \alpha_2 \to \cdots \to \alpha_k$ instead of $(\alpha_1 \to (\alpha_2 \to \cdots \to (\alpha_{k-1} \to \alpha_k) \cdots))$.

The abstract *intensional type structure* $\langle(\mathsf{IT}_\alpha,=)\rangle_{\alpha\in\mathcal{T}}$ in the applicative language L is now given by inductively defining the formula IT_α as follows:

$$
\begin{aligned}
x \in \mathsf{IT}_0 &:= x \in \mathsf{W},\\
x \in \mathsf{IT}_{\alpha\times\beta} &:= \mathsf{p}_0 x \in \mathsf{IT}_\alpha \ \wedge\ \mathsf{p}_1 x \in \mathsf{IT}_\beta \ \wedge\ \mathsf{p}(\mathsf{p}_0 x)(\mathsf{p}_1 x) = x,\\
x \in \mathsf{IT}_{\alpha\to\beta} &:= (\forall y \in \mathsf{IT}_\alpha)(xy \in \mathsf{IT}_\beta).
\end{aligned}
$$

Equality in IT_α is simply the restriction of equality in PT. Alternatively, one can consider an extensional type structure, cf. [48, 53].

4.2 The system PV^ω

PV^ω is a typed formal system whose terms denote exactly the basic feasible functionals. PV^ω includes:

- the simply typed lambda calculus over the base type of binary words
- basic operations on words, essentially the base operations of PT
- a type two functional for bounded recursion on notation
- notation induction on binary words for Σ^b_1 or *NP* formulas

For an exact definition, cf. e.g. Strahm [48]. We observe that due to Lemma 2.4, we indeed have a type two functional for bounded recursion on notation which has the correct type, provably in PT. Using the intensional type structure $\langle(\mathsf{IT}_\alpha,=)\rangle_{\alpha\in\mathcal{T}}$ sketched above, it is then a matter of routine to check that PV^ω can be directly interpreted in PT. This shows that the basic feasible functionals in all finite types are provably total in PT.

The question arises whether indeed the BFFs are *exactly* the provably total functionals of PT. This question has been answered in the positive for the *type two* BFFs in Strahm [49] by using an extension of the realizability argument sketched above. Moreover, it follows from the work in Cantini [14] that this result holds with respect to arbitrary finite types if one considers an intuitionistic version of PT. Therefore we can summarize:

Theorem 4.1. *1. The system PV^ω is contained in PT; i.e., the basic feasible functionals in all finite types are provably total in PT.*

2. The provably total type 2 functionals of PT coincide exactly with the basic feasible functionals of type 2.

Let us conclude this section with the following conjecture.

Conjecture 4.2. *The classical theory* PT *characterizes the basic feasible functionals in all finite types.*

5 Adding types and names

In this section, we will describe PET, a theory of **p**olynomial time operations with **e**xplicit **t**ypes. The theory PET is an extension of the applicative base theory $\mathsf{B}(*, \times)$ by means of a natural restriction of elementary comprehension, which is one of the crucial principles of explicit mathematics, see Feferman [19, 21]. Below we will use the language of explicit mathematics due to Jäger [28] which is based on a so-called naming relation $\Re$. The type existence axioms are naturally presented by means of a finite axiomatisation in the spirit of Feferman and Jäger [23]. The theory PET has been introduced in Spescha and Strahm [45].

5.1 The informal setting of types and names

Types in explicit mathematics are collections of operations and must be thought of as being generated successively from preceding ones. In contrast to the restricted character of operations, types can have quite complicated defining properties. What is essential in the whole explicit mathematics approach, however, is the fact that types are again represented by operations or, as we will call them in this case, *names*. Thus each type U is named or represented by a name u; in general, U may have many different names or representations. It is exactly this interplay between operations and types on the level of names which makes explicit mathematics extremely powerful and, in fact, witnesses its explicit character.

Types are extensional and have (explicit) names which are intensional. The names are generated via uniform operations and the link to the types they are referring to is established by the naming relation $\Re$. The element relation $\in$ is also a relation between an individual and a type, expressing that the individual is a member of the type. The formalization of explicit mathematics using a naming relation $\Re$ is due to Jäger [28].

5.2 The language of types and names

The language $\mathbb{L}$ is a two-sorted language extending L by

- type variables $U, V, W, X, Y, Z, \ldots$

- binary relation symbols $\Re$ (naming) and $\in$ (elementhood)
- new (individual) constants w (initial segment of W), id (identity), dom (domain), un (union), int (intersection), and inv (inverse image)

The *formulas* $(A, B, C, \ldots)$ of $\mathbb{L}$ are built from the atomic formulas of L as well as formulas of the form

$$(s \in X), \quad \Re(s, X), \quad (X = Y)$$

by closing under the boolean connectives and quantification in both sorts. The formula $\Re(s, X)$ reads as "the individual s is a name of (or represents) the type X".

We use the following abbreviations:

$$\Re(s) := (\exists X)\Re(s, X),$$
$$s \mathrel{\dot{\in}} t := (\exists X)(\Re(t, X) \ \wedge \ s \in X).$$

5.3 The theory PET

The following axioms state that each type has a name, that there are no homonyms and that equality of types is extensional.

Ontological axioms:

$$(\exists x)\Re(x, X) \tag{O1}$$
$$\Re(a, X) \ \wedge \ \Re(a, Y) \ \rightarrow \ X = Y \tag{O2}$$
$$(\forall z)(z \in X \leftrightarrow z \in Y) \ \rightarrow \ X = Y \tag{O3}$$

In the sequel we let $\mathsf{W}_a(x)$ stand for $\mathsf{W}(x) \ \wedge \ x \leq a$. The following axioms provide a finite axiomatization of a restricted form of the schema of elementary comprehension.

Type existence axioms:

$$a \in \mathsf{W} \rightarrow \Re(\mathsf{w}(a)) \ \wedge \ (\forall x)(x \mathrel{\dot{\in}} \mathsf{w}(a) \leftrightarrow \mathsf{W}_a(x)) \tag{$\mathbf{w}_a$}$$

$$\Re(\mathsf{id}) \ \wedge \ (\forall x)(x \mathrel{\dot{\in}} \mathsf{id} \leftrightarrow (\exists y)(x = (y, y))) \tag{\textbf{id}}$$

$$\Re(a) \rightarrow \Re(\mathsf{inv}(f, a)) \ \wedge \ (\forall x)(x \mathrel{\dot{\in}} \mathsf{inv}(f, a) \leftrightarrow fx \mathrel{\dot{\in}} a) \tag{\textbf{inv}}$$

$$\Re(a) \ \wedge \ \Re(b) \rightarrow \Re(\mathsf{un}(a, b)) \ \wedge \ (\forall x)(x \mathrel{\dot{\in}} \mathsf{un}(a, b) \leftrightarrow (x \mathrel{\dot{\in}} a \ \vee \ x \mathrel{\dot{\in}} b)) \tag{\textbf{un}}$$

$$\Re(a) \ \wedge \ \Re(b) \to \Re(\mathsf{int}(a,b)) \wedge \ (\forall x)(x \mathrel{\dot\in} \mathsf{int}(a,b) \leftrightarrow (x \mathrel{\dot\in} a \ \wedge \ x \mathrel{\dot\in} b)) \qquad \textbf{(int)}$$

$$\Re(a) \to \Re(\mathsf{dom}(a)) \ \wedge \ (\forall x)(x \mathrel{\dot\in} \mathsf{dom}(a) \leftrightarrow (\exists y)((x,y) \mathrel{\dot\in} a)) \qquad \textbf{(dm)}$$

In contrast to the usual formulation of elementary comprehension in explicit mathematics (cf. e.g. Feferman and Jäger [23]), we do not claim that the collection of binary words forms a type, but merely that for each word a, the collection $\{x \in \mathsf{W} : x \le a\}$ forms a type, uniformly in a. In addition, there are no complement types. The remaining type existence axioms are identical to the ones in [23].

Finally, the principle of type induction along W reads in the expected manner.

Type induction on W:

$$\epsilon \in X \ \wedge \ (\forall x \in \mathsf{W})(x \in X \to \mathsf{s}_0 x \in X \ \wedge \ \mathsf{s}_1 x \in X) \to (\forall x \in \mathsf{W})(x \in X)$$

The theory PET is defined to be the extension of the first-order applicative theory $\mathsf{B}(*,\times)$ by

- the ontological axioms
- the above type existence axioms
- type induction on W

In Spescha and Strahm [45] it is shown that the finite axiomatisation of type existence in PET gives rise to a natural restriction of the well-known schema of elementary comprehension in explicit mathematics.

5.4 The proof-theoretic strength of PET

Let PT^- be PT without universal quantifiers in induction formulas. Clearly, PT^- proves the totality of the polynomial time computable functions, since it is strong enough to represent bounded recursion on notation in the form of a type two functional (cf. Lemma 2.4). Indeed, PET is a conservative extension of PT^- as is shown in Spescha and Strahm [45].

Theorem 5.1. *We have the following proof-theoretic results:*

1. PET *is a conservative extension of* PT^-.
2. $\boldsymbol{\tau}(\mathsf{PT}^-) = \text{FP}\textsc{time}$.

The lower bound uses a rather involved embedding of PT^- into PET. The interpretation uses a bootstrapping functional mapping each operation f on W to an operation f^* such that $f^* x = \max_{y \subseteq x} fy$.

For the proof of the upper bound one starts off from a model of PT^- and extends it to a model of PET satisfying the same first order sentences. The construction is carried out in stages by defining the set of names and their extensions successively. Then one can show that the so-obtained model enjoys type induction.

For full details of these arguments, see Spescha and Strahm [45].

5.5 Extensions of PET

In addition to the principles $(\mathbf{Tot})$ and $(\mathbf{Ext})$ discussed above, Cantini [14] has considered a form of positive choice in the context of PT with a partial truth predicate (cf. Section 6) and shows that this principle does not increase the proof-theoretic strength. Cantini's result can be used to show that the following form of the axiom of choice formulated in the language $\mathbb{L}$ does not increase the strength of PET.

Positive axiom of choice:

$$(\forall x \in \mathsf{W})(\exists y \in \mathsf{W}) A(x,y) \to (\exists f : \mathsf{W} \to \mathsf{W})(\forall x \in \mathsf{W}) A(x, fx) \qquad \textbf{(AC)}$$

where $A(x,y)$ is a positive elementary formula.

Cantini has also shown in [14] that adding a uniformity principle for positive formulas of L yields an extension of PT whose provably total functions are still the functions computable in polynomial time. In our context, we can state Cantini's principle as follows.

Positive uniformity principle:

$$(\forall x)(\exists y \in \mathsf{W}) A(x,y) \to (\exists y \in \mathsf{W})(\forall x) A(x,y) \qquad \textbf{(UP)}$$

where $A(x,y)$ is positive elementary.

The principle $(\mathbf{UP})$ leads to a very natural extension of PET by adding a type existence axiom for universal quantification; this axiom is the natural dual analogue of the domain type present in PET.

Universal quantification:

$$\Re(a) \to \Re(\mathsf{all}(a)) \;\wedge\; (\forall x)(x \mathrel{\dot\in} \mathsf{all}(a) \leftrightarrow (\forall y)((x,y) \mathrel{\dot\in} a)) \qquad \textbf{(all)}$$

The presence of the axiom $(\mathbf{all})$ makes the type existence axioms more symmetric, i.e. the types are generated from base types (initial segments of W and the identity type) by closing under domains, unions, intersections, existential quantification (inverse image) and universal quantification.

In order to see that $(\mathbf{all})$ does not increase the proof-theoretic strength of PET, one shows that any model of $\mathsf{PT} + (\mathbf{UP})$ can be extended to a a model of $\mathsf{PET} + (\mathbf{all})$. The presence of $(\mathbf{UP})$ is pivotal in the treatment of $(\mathbf{all})$. For a complete exposition of these results, see Spescha and Strahm [45].

Theorem 5.2. *The provably total functions of* PET *augmented by any combination of the principles* $(\mathbf{all})$, $(\mathbf{UP})$, $(\mathbf{AC})$, $(\mathbf{Tot})$, *and* $(\mathbf{Ext})$ *coincide with the polynomial time computable functions.*

The next natural step is to add the so-called *Join axiom*, which constructs disjoint unions of types named by an operation; it has been widely studied for many systems of explicit mathematics. The Join axioms are given by the following assertions (**J.1**) and (**J.2**) (j denotes a new constant).

Join axioms:

$$\Re(a) \ \wedge \ (\forall x \dot{\in} a)\Re(fx) \rightarrow \Re(\mathsf{j}(a,f)) \qquad (\mathbf{J.1})$$

$$\Re(a) \ \wedge \ (\forall x \dot{\in} a)\Re(fx) \rightarrow (\forall x)(x \dot{\in} \mathsf{j}(a,f) \leftrightarrow \Sigma(f,a,x)) \qquad (\mathbf{J.2})$$

where $\Sigma(f,a,x)$ is the formula

$$(\exists y)(\exists z)(x = (y,z) \ \wedge \ y \dot{\in} a \ \wedge \ z \dot{\in} fy)$$

In Spescha [43] and Spescha and Strahm [44] the realizability interpretation of the first order language L is extended to the language of types and names $\mathbb{L}$. In combination with a partial cut elimination argument, it is possible to show that the *intuitionistic* version of PET plus the Join axioms can be realized using polynomial time computable functions. Currently, work of Probst is underway in order to extend this result to classical logic.

6 Partial truth

In this section we address some interesting extensions of PT^+ which have been proposed and studied by Cantini [14]. The idea is to augment PT^+ by

- a (form of) self-referential truth (à la Aczel, Feferman, Kripke, etc.), providing a fixed point theorem for predicates

- an axiom of choice for operations and a uniformity principle, restricted to positive conditions

These extensions do not alter the proof-theoretic strength of PT, a fact that has been heavily used in the previous section in studying extensions of our theory PET.

In the following let us briefly report on some of the many results obtained in Cantini [14]. For a thorough exposition of frameworks for truth and abstraction based on combinatory logic, cf. Cantini [10] and Kahle [29].

6.1 The language L_{T}

The (first order) language L_{T} is an extension of the language L by

- a new unary predicate symbol T for *truth*
- new individual constants $\dot{=}$, $\dot{\mathsf{W}}$, $\dot{\wedge}$, $\dot{\vee}$, $\dot{\forall}$, $\dot{\exists}$

For each positive formula A of L_{T} we can inductively define a term $[A]$ whose free variables are exactly the free variables of A:

$$\begin{aligned}
[t = s] &:= (\dot{=}ts) \\
[\mathsf{T}(t)] &:= t \\
[s \in \mathsf{W}] &:= \dot{\mathsf{W}}s \\
[A \wedge B] &:= \dot{\wedge}\,[A][B] \\
[A \vee B] &:= \dot{\vee}\,[A][B] \\
[(\forall x)A] &:= \dot{\forall}(\lambda x.[A]) \\
[(\exists x)A] &:= \dot{\exists}(\lambda x.[A])
\end{aligned}$$

We have that $\lambda x.[A]$ can be interpreted as the propositional function defined by the formula A. We can now interpret the language of naive set theory by defining $x \in a$ as $\mathsf{T}(ax)$ and understand $\{x : A\}$ as $\lambda x.[A]$.

6.2 The truth axioms

The truth axioms for the positive fragment of L_{T} spell out the expected clauses according to the reductionist semantics as follows:

Truth axioms:

$$\mathsf{T}(\dot{=}xy) \leftrightarrow x = y$$

$$\begin{aligned}
\mathsf{T}(\dot{\mathsf{W}}x) &\leftrightarrow \mathsf{W}(x)\\
\mathsf{T}(x \mathbin{\dot\wedge} y) &\leftrightarrow \mathsf{T}(x) \;\wedge\; \mathsf{T}(y)\\
\mathsf{T}(x \mathbin{\dot\vee} y) &\leftrightarrow \mathsf{T}(x) \;\vee\; \mathsf{T}(y)\\
\mathsf{T}(\dot\forall f) &\leftrightarrow (\forall x)\mathsf{T}(fx)\\
\mathsf{T}(\dot\exists f) &\leftrightarrow (\exists x)\mathsf{T}(fx)
\end{aligned}$$

One of the many interesting consequences of these axioms is a second recursion or fixed point theorem for positive predicates, which can be obtained by lifting the fixed point theorem for combinatory logic (cf. Lemma 2.1) to the truth-theoretic language, cf. Cantini [10, 14].

6.3 Adding positive choice and uniformity

We can formulate positive choice and uniformity principles in the language L_{T} as follows:

Positive choice and uniformity in L_{T}:

$$(\forall x \in \mathsf{W})(\exists y \in \mathsf{W})\mathsf{T}(axy) \to (\exists f : \mathsf{W} \to \mathsf{W})(\forall x \in \mathsf{W})\mathsf{T}(ax(fx)) \qquad (\mathbf{AC})$$

$$(\forall x)(\exists y \in \mathsf{W})\mathsf{T}(axy) \to (\exists y \in \mathsf{W})(\forall x)\mathsf{T}(axy) \qquad (\mathbf{UP})$$

One of the numerous results obtained in Cantini [14] is stated in the following theorem. It has been used in Spescha and Strahm [44] in order to show the conservativity of various extensions of PET.

Theorem 6.1. $\boldsymbol{\tau}(\mathsf{PT}^+ + \text{truth axioms} + \mathbf{AC} + \mathbf{UP}) = \mathrm{FP}\textsc{time}$.

The proof methods used by Cantini include a subtle internal forcing semantics, non-standard variants of realizability and partial cut elimination properties. The forcing interpretation is very elegant and makes direct use of the truth predicate T.

7 Safe induction

Apart from the world of bounded recursion schemas, bounded arithmetic and bounded applicative theories there is the realm of so-called *tiered systems* in the sense of Cook and Bellantoni (cf. e.g. [3]) and Leivant (cf. e.g. [31, 32]). Crucial for this approach to characterizing complexities is a strictly predicative regime which distinguishes between different uses of variables in induction and recursion schemas, thus severely restricting the definable or provably total functions in various unbounded formalisms.

Unarguably, the tiered approach to complexity has led to numerous highly interesting and intrinsic recursion-theoretic and also proof-theoretic characterizations of complexity classes, which might lead to new subrecursive programming paradigms. Also, higher type issues have recently been a subject of interest in this area, cf. e.g. Bellantoni, Niggl, Schwichtenberg [4], Hofmann [26], and Leivant [33],

Finally, the tiered approach has provided neat distinctions between slow growing and fast growing proof theories, see e.g. Wainer [54] and Ostrin and Wainer [36].

Below let us briefly address some recent work along the lines of implicit characterizations in the context of untyped applicative theories based on classical logic.

7.1 Polynomial time

In our applicative setting the above-mentioned "predicativization" amounts to distinguishing between (at least) two sorts or types of binary words W_0 and W_1, say, where induction over W_1 is allowed for formulas which are positive and do not contain W_1, cf. Cantini [13] for such systems.

A more elegant viewpoint of the predicative regime is to consider a modal framework. Extend the language L by a modal operator $\Box$ and let $\Box$ obey the laws of an S4 modality. Let $t \in \Box\mathsf{W}$ stand for $\Box(t \in \mathsf{W})$. Then W and $\Box\mathsf{W}$ play the role of normal and safe strings in the Bellantoni-Cook sense, respectively. We call a formula *positive safe* if it is positive and does not involve the $\Box$ operator. Accordingly, we can formulate the following natural induction principle.

Positive safe notation induction:

For each positive safe formula $A(x)$,

$$A(\epsilon) \;\wedge\; (\forall x \in \Box\mathsf{W})(A(x) \to A(\mathsf{s}_0 x) \;\wedge\; A(\mathsf{s}_1 x)) \;\to\; (\forall x \in \Box\mathsf{W})A(x)$$

Let PR^μ denote the extension of the applicative theory B based on the classical modal predicate logic S4 and the schema of positive safe notation induction. The notion of a provably total word function can be suitably adapted for PR^μ, taking into account the two sorts W and $\Box\mathsf{W}$, cf. [13] for details.

We are ready to state the following theorem, which is proved in Cantini [13] by making use of cut elimination and realizability by Cook-Bellantoni functions.

Theorem 7.1. $\boldsymbol{\tau}(\mathsf{PR}^\mu) = \text{FP}\textsc{time}$.

7.2 Polynomial space

More recently, Calamai and Cantini [8, 9] have proposed an extension of PR^{μ}, termed PR^{μ}_{p}, where induction is strengthened to so-called positive safe tree induction with pointers. The principle is inspired by Oitavem's recent tiered characterization of FPSPACE in [35].

Positive safe tree induction with pointers:

For each positive safe formula $A(x, y)$,

$$(\forall p \in \Box \mathsf{W}) A(\epsilon, p) \ \wedge$$

$$(\forall x \in \Box \mathsf{W})(\forall p \in \Box \mathsf{W})(A(x, \mathsf{s}_0 p) \ \wedge \ A(x, \mathsf{s}_1 p) \rightarrow A(\mathsf{s}_0 x, p) \ \wedge \ A(\mathsf{s}_1 x, p))$$

$$\rightarrow (\forall x \in \Box \mathsf{W})(\forall p \in \Box \mathsf{W}) A(x, p)$$

The proof of the theorem below of Calamai and Cantini makes use of cut elimination and realizability by functions in Oitavem's function algebra with pointers and tree recursion.

Theorem 7.2. $\boldsymbol{\tau}(\mathsf{PR}^{\mu}_{p}) =$ FPSPACE.

8 Conclusion

In this article we have considered a number of applicative theories (with and without types or self-referential truth, with and without modality) whose induction principles are formulated for a suitable subclass of positive formulas.

Regarding induction for *arbitrary positive formulas*, say in the first order language L, one captures exactly the primitive recursive functions. For definiteness, let $(\mathsf{Pos}\text{-}\mathsf{I}_{\mathsf{W}})$ denote induction on W for formulas in Pos. Then $\boldsymbol{\tau}(\mathsf{B} + (\mathsf{Pos}\text{-}\mathsf{I}_{\mathsf{W}}))$ coincides with the primitive recursive functions. This result was first established by Cantini in [11] using asymmetric interpretation and formalized semantics in $\mathsf{I}\Sigma_1$ and can be considered as a generalization to the applicative context of the well-known Parsons-Mints-Takeuti theorem. The characterization theorem can also be established by the realizability techniques presented in this article (cf. Cantini [14], Strahm [48]). However, it has to be mentioned that Cantini's original result [11] is even a bit stronger, since negated equations in induction formulas are allowed.

We conclude this article by mentioning that the realizability techniques of this paper have recently been helpful in the context of abstract many sorted algebras with non-computable equality in establishing a further generalization of the Parsons-Mints-Takeuti theorem, cf. Strahm and Zucker [50].

References

[1] BEESON, M. J. *Foundations of Constructive Mathematics: Metamathematical Studies*. Springer, Berlin, 1985.

[2] BEESON, M. J. Proving programs and programming proofs. In *Logic, Methodology and Philosophy of Science VII*, Barcan Marcus et. al., Ed. North Holland, Amsterdam, 1986, pp. 51–82.

[3] BELLANTONI, S., AND COOK, S. A new recursion-theoretic characterization of the poly-time functions. *Computational Complexity 2* (1992), 97–110.

[4] BELLANTONI, S., NIGGL, K.-H., AND SCHWICHTENBERG, H. Higher type recursion, ramification and polynomial time. *Annals of Pure and Applied Logic 104*, 1–3 (2000), 17–30.

[5] BUSS, S. R. *Bounded Arithmetic*. Bibliopolis, Napoli, 1986.

[6] BUSS, S. R. The witness function method and fragments of Peano arithmetic. In *Proceedings of the Ninth International Congress on Logic, Methodology and Philosophy of Science, Uppsala, Sweden, August 7–14, 1991*, D. Prawitz, B. Skyrms, and D. Westerståhl, Eds. Elsevier, North Holland, Amsterdam, 1994, pp. 29–68.

[7] BUSS, S. R. First-order proof theory of arithmetic. In *Handbook of Proof Theory*, S. R. Buss, Ed. Elsevier, 1998, pp. 79–147.

[8] CALAMAI, G. *Proof-theoretic contributions to computational complexity*. PhD thesis, University of Siena, 2008.

[9] CANTINI, A. A footnote on the Parsons-Mints-Takeuti theorem. Talk at *Recent Trends in Proof Theory*, Bern, July 2008.

[10] CANTINI, A. *Logical Frameworks for Truth and Abstraction*. North-Holland, Amsterdam, 1996.

[11] CANTINI, A. Proof-theoretic aspects of self-referential truth. In *Tenth International Congress of Logic, Methodology and Philosophy of Science, Florence, August 1995*, Maria Luisa Dalla Chiara et. al., Ed., vol. 1. Kluwer, September 1997, pp. 7–27.

[12] CANTINI, A. Feasible operations and applicative theories based on $\lambda\eta$. *Mathematical Logic Quarterly 46*, 3 (2000), 291–312.

[13] CANTINI, A. Polytime, combinatory logic and positive safe induction. *Archive for Mathematical Logic 41*, 2 (2002), 169–189.

[14] CANTINI, A. Choice and uniformity in weak applicative theories. In *Logic Colloquium '01*, M. Baaz, S. Friedman, and J. Krajíček, Eds., vol. 20 of *Lecture Notes in Logic*. Association for Symbolic Logic, 2005.

[15] CLOTE, P. Computation models and function algebras. In *Handbook of Computability Theory*, E. Griffor, Ed. Elsevier, 1999, pp. 589–681.

[16] COBHAM, A. The intrinsic computational difficulty of functions. In *Logic, Methodology and Philosophy of Science II*. North Holland, Amsterdam, 1965, pp. 24–30.

[17] COOK, S. A., AND KAPRON, B. M. Characterizations of the basic feasible functionals of finite type. In *Feasible Mathematics*, S. R. Buss and P. J. Scott, Eds. Birkhäuser, Basel, 1990, pp. 71–95.

[18] COOK, S. A., AND URQUHART, A. Functional interpretations of feasibly constructive arithmetic. *Annals of Pure and Applied Logic 63*, 2 (1993), 103–200.

[19] FEFERMAN, S. A language and axioms for explicit mathematics. In *Algebra and Logic*, J. Crossley, Ed., vol. 450 of *Lecture Notes in Mathematics*. Springer, Berlin, 1975, pp. 87–139.

[20] FEFERMAN, S. Recursion theory and set theory: a marriage of convenience. In *Generalized recursion theory II, Oslo 1977*, J. E. Fenstad, R. O. Gandy, and G. E. Sacks, Eds., vol. 94 of *Stud. Logic Found. Math.* North Holland, Amsterdam, 1978, pp. 55–98.

[21] FEFERMAN, S. Constructive theories of functions and classes. In *Logic Colloquium '78*, M. Boffa, D. van Dalen, and K. McAloon, Eds. North Holland, Amsterdam, 1979, pp. 159–224.

[22] FEFERMAN, S. Definedness. *Erkenntnis 43* (1995), 295–320.

[23] FEFERMAN, S., AND JÄGER, G. Systems of explicit mathematics with non-constructive μ-operator. Part II. *Annals of Pure and Applied Logic 79*, 1 (1996), 37–52.

[24] FERREIRA, F. *Polynomial Time Computable Arithmetic and Conservative Extensions*. PhD thesis, Pennsylvania State University, 1988.

[25] FERREIRA, F. Polynomial time computable arithmetic. In *Logic and Computation, Proceedings of a Workshop held at Carnegie Mellon University, 1987*, W. Sieg, Ed., vol. 106 of *Contemporary Mathematics*. American Mathematical Society, Providence, Rhode Island, 1990, pp. 137–156.

[26] HOFMANN, M. Type systems for polynomial-time computation. Habilitation Thesis, Darmstadt, 1999. Appeared as LFCS Technical Report ECS-LFCS-99-406.

[27] IRWIN, R., KAPRON, B., AND ROYER, J. On characterizations of the basic feasible functionals, Part I. *Journal of Functional Programming 11* (2001), 117–153.

[28] JÄGER, G. Induction in the elementary theory of types and names. In *Computer Science Logic '87*, E. Börger, H. Kleine Büning, and M.M. Richter, Eds., vol. 329 of *Lecture Notes in Computer Science*. Springer, Berlin, 1988, pp. 118–128.

[29] KAHLE, R. *The Applicative Realm*. Habilitation Thesis, Tübingen, 2007. Appeared in Textos de Mathemática 40, Departamento de Mathemática da Universidade de Coimbra, Portugal, 2007.

[30] KAPRON, B., AND COOK, S. A new characterization of type 2 feasibility. *SIAM Journal on Computing 25* (1996), 117–132.

[31] LEIVANT, D. A foundational delineation of poly-time. *Information and Computation 110* (1994), 391–420.

[32] LEIVANT, D. Ramified recurrence and computational complexity I: Word recurrence and poly-time. In *Feasible Mathematics II*, P. Clote and J. Remmel, Eds. Birkhäuser, 1994, pp. 320–343.

[33] LEIVANT, D. Implicit computational complexity for higher type functionals (Extended abstract). In *CSL '02*, J. Bradfield, Ed., vol. 2471 of *Lecture Notes in Computer Science*. Springer, 2002, pp. 367–381.

[34] MELHORN, K. Polynomial and abstract subrecursive classes. *Journal of Computer and System Science 12* (1976), 147–178.

[35] OITAVEM, I. Characterizing PSPACE with pointers. *Mathematical Logic Quarterly 54*, 3 (2008), 323 – 329.

[36] OSTRIN, G., AND WAINER, S. S. Elementary arithmetic. *Annals of Pure and Applied Logic 133* (2005), 275–292.

[37] PEZZOLI, E. On the computational complexity of type 2 functionals. In *Computer Science Logic '97*, vol. 1414 of *Lecture Notes in Computer Science*. Springer, 1998, pp. 373–388.

[38] RITCHIE, R. W. Classes of predictably computable functions. *Transactions of the American Mathematical Society 106* (1963), 139–173.

[39] ROYER, J. Semantics vs. syntax vs. computations: Machine models for type-2 polynomial-time bounded functionals. *Journal of Computer and System Science 54* (1997), 424–436.

[40] SCHLÜTER, A. An extension of Leivant's characterization of poly-time by predicative arithmetic. Preprint, Stanford Univeristy, 1995. 13 pages.

[41] SCOTT, D. S. Identity and existence in intuitionistic logic. In *Applications of Sheaves*, M. P. Fourman, C. J. Mulvey, and D. S. Scott, Eds. Springer, Berlin, 1979. Lecture Notes in Mathematics 753.

[42] SETH, A. *Complexity Theory of Higher Type Functionals*. PhD thesis, Tata Institute of Fundamental Research, Bombay, 1994.

[43] SPESCHA, D. *Weak systems of explicit mathematics*. PhD thesis, Universität Bern, 2009.

[44] SPESCHA, D., AND STRAHM, T. Realizability in weak systems of explicit mathematics. In preparation.

[45] SPESCHA, D., AND STRAHM, T. Elementary explicit types and polynomial time operations. *Mathematical Logic Quarterly 55*, 3 (2009), 245–258.

[46] STRAHM, T. Polynomial time operations in explicit mathematics. *Journal of Symbolic Logic 62*, 2 (1997), 575–594.

[47] STRAHM, T. *Proof-theoretic Contributions to Explicit Mathematics*. Habilitationsschrift, University of Bern, 2001.

[48] STRAHM, T. Theories with self-application and computational complexity. *Information and Computation 185* (2003), 263–297.

[49] STRAHM, T. A proof-theoretic characterization of the basic feasible functionals. *Theoretical Computer Science 329* (2004), 159–176.

[50] STRAHM, T., AND ZUCKER, J. Primitive recursive selection functions for existential assertions over abstract algebras. *Journal of Logic and Algebraic Programming 76*, 2 (2008), 175 – 197.

[51] THOMPSON, D. B. Subrecursiveness: machine independet notions of computability in restricted time and storage. *Mathematical Systems Theory 6* (1972), 3–15.

[52] TROELSTRA, A., AND VAN DALEN, D. *Constructivism in Mathematics*, vol. I. North-Holland, Amsterdam, 1988.

[53] TROELSTRA, A., AND VAN DALEN, D. *Constructivism in Mathematics*, vol. II. North Holland, Amsterdam, 1988.

[54] WAINER, S. S. Provable recursiveness and complexity. In *Logic Colloquium '01*, M. Baaz, S. Friedman, and J. Krajíček, Eds., vol. 20 of *Lecture Notes in Logic*. Association for Symbolic Logic, 2005.

Computing Bounds from Arithmetical Proofs

Stanley S. Wainer

Leeds, UK

Abstract We explore the role of the function $a + 2^b$, and its generalizations to higher number classes, in providing complexity bounds for the provably computable functions across a broad spectrum of theories, based on a "predicative" induction scheme and ranging in strength between polynomial-time arithmetic and Π^1_1-CA_0. The resulting "fast-growing" sub-recursive hierarchy forges a direct link between proof theory and various combinatorial independence results. As illustration, the final section treats Friedman's "miniaturized" Kruskal Theorem for labelled trees, by showing directly that the appropriate bounding function for Π^1_1-CA_0 has a "bad" computation sequence.

1 The Fast Growing Hierarchy

If $B_b(a) = a + 2^b$ then B satisfies the recursion:

$$B_0(a) = a + 1 \qquad B_{b+1}(a) = B_b(B_b(a))\,.$$

This is the beginning of a version of the (so-called) Fast Growing Hierarchy:

$$B_\alpha(a) = \begin{cases} a+1 & \text{if } \alpha = 0 \\ B_{\alpha'}(B_{\alpha'}(a)) & \text{if } \alpha = \alpha' + 1 \\ B_{\alpha_a}(a) & \text{if } \alpha \text{ is a limit.} \end{cases}$$

Note the dependence on chosen fundamental sequences $\{\alpha_a\}$ to limits α.

We could, more suggestively, write $B_\alpha(a)$ as $a \oplus 2^\alpha$ where, if $\lambda_0, \lambda_1, \lambda_2, \ldots$ is a given fundamental sequence to limit λ, then $a \oplus \lambda$ is defined by the diagonalization $a \oplus \lambda_a$. This is a quite natural way to extend number-theoretic hierarchies into the transfinite, and in the case of the B_α functions there is a deep proof-theoretic involvement which we attempt to bring out here.

Firstly, a basic arithmetical context serves to illustrates this connection. The theory EA(I;O) is a stripped-down variant of the "ramified" theories of Leivant [8] who also highlights $a + 2^b$ as a crucial example; it imports to a formal theory the normal/safe variable discipline of Bellantoni–Cook [2]. For a more detailed proof-theoretic analysis see Ostrin–Wainer [9], [10].

2 Input-Output Arithmetic EA(I;O)

- EA(I;O) has the language of arithmetic, with (quantified, "output") variables $a, b, c, \ldots$.
- In addition there are numerical constants ("inputs") $x, y, z, \ldots$.
- It has symbols for the zero, successor, predecessor, addition, subtraction and multiplication functions, given by their usual defining axioms. Also there is a pairing function $\pi(a,b)$ $(:= 1/2(a+b)(a+b+1)+a+1)$ with inverses π_0, π_1 from which sequence numbers can be constructed using $\pi(s,a)$ to append a to s, and deconstructed by functions $(s)_i$ extracting the i-th component. All of these initial functions are quadratically bounded. However, the stock of basic symbols is enlarged further by the addition of the exponential function $B_b(a) = a + 2^b$ with defining equations as above. The given relations are $=$ and $\leq$.
- Only "basic" terms (those built out of variables and constants by application of the unary term constructors alone: successor, predecessor, π_0 and π_1 – nothing else) are allowed as witnessing or instantiating terms in the $\exists, \forall$ quantifier rules.
- The induction axioms are:

  $$A(0) \wedge \forall a(A(a) \rightarrow A(a+1)) \ \rightarrow \ A(t)$$

  where $t = t(x)$ is a *closed basic term* controlling induction-length. Note that if $A(a)$ is progressive then so is $\forall b \leq a.A(b) \equiv \forall b(b \leq a \rightarrow A(b))$, and so a more revealing instance of induction is

  $$A(0) \wedge \forall a(A(a) \rightarrow A(a+1)) \ \rightarrow \ \forall b \leq t.A(b)\,.$$

  In other words, EA(I;O) is really a theory of bounded induction, the (implicit) bounds being closed terms $t(x)$ dependent on inputs x which (because they are constants) cannot be universally quantified, and later re-instantiated, once introduced. Call this "input" (or "predicative") induction.

Definition 2.1. Write $t\downarrow$ for $\exists a(t = a)$.

Note 2.2. If t is not basic one cannot pass directly from $t = t$ to $t\downarrow$ because only basic terms are allowed as witnesses in the $\exists$ rule. However, the usual equality axioms allow one to derive immediately in EA(I;O):

$$t\downarrow \wedge A(t) \rightarrow \exists a A(a)$$

and the dual

$$t\downarrow \wedge \forall a A(a) \to A(t)\,.$$

Thus "defined" terms may witness existential quantifiers or instantiate universal ones.

Example 2.3. Some basic illustrations of induction complexity in EA(I;O):

- From $b+c=d$ we immediately get $b+(c+1)=d+1$ and since the term $d+1$ is basic, then $b+c=d \to \exists a(b+(c+1)=a)$. Hence $b+c\downarrow\to b+(c+1)\downarrow$.

  Therefore $b+x\downarrow$ by Σ_1-induction "up to" x.

  Then $\forall b(b+x\downarrow)$.

- One can then induct on the formula $b+x\cdot c\downarrow$ because $b+x\cdot c=d \to b+x\cdot(c+1)=d+x$ and $d+x\downarrow$ by the above. So $b+x\cdot c\downarrow\to b+x\cdot(c+1)\downarrow$. Clearly $b+x\cdot 0\downarrow$ because b is basic, and hence another application of Σ_1 input induction gives either $b+x^2\downarrow$ or $b+x\cdot y\downarrow$.

  Then $\forall b(b+x^2\downarrow)$, and similarly $\forall b(b+x^3\downarrow)$, $\forall b(b+x^4\downarrow)$ etc.

- Exponential requires a Π_2 induction on $\forall a(a+2^b\downarrow)$ since by two calls on the premise, and making crucial use of the above note, $\forall a(a+2^b\downarrow)\to\forall a(a+2^b+2^b\downarrow)$. Therefore $\forall a(a+2^b\downarrow)$ is progressive in b and by input induction, $\forall a(a+2^x\downarrow)$, i.e. $B_x(a)\downarrow$. In particular with $a=x$ we obtain $B_x(x)=x+2^x\downarrow$ which could be written $B_\omega(x)\downarrow$, choosing the identity function as the "standard" fundamental sequence to ω.

- To carry this a stage further we use Gentzen's method for proving transfinite induction below ε_0, but now the context is simpler. Consider the Π_3 formula:

  $$\forall b(\forall a(a+2^b\downarrow)\to b+2^c\downarrow\wedge\forall a(a+2^{b+2^c}\downarrow))\,.$$

  This is progressive in c, the base case $c=0$ being the progressiveness of $\forall a(a+2^b\downarrow)$ just shown. Therefore a Π_3 input induction proves the formula with $c:=t$ for any closed basic term t. By choosing $t=x$ and instantiating $b:=0$ one obtains $\forall a(a+2^{2^x}\downarrow)$. Instantiating $b:=x$ yields $\forall a(a+2^{x+2^x}\downarrow)$, or alternatively $\forall a(B_{B_\omega(x)}(a)\downarrow)$. Then since $B_\omega(x)\downarrow$ one could instantiate $a:=B_\omega(x)$ to obtain

  $$x+2^x+2^{x+2^x}=B_{B_\omega(x)}(B_\omega(x))=B_\omega(B_\omega(x))\downarrow\ .$$

  Higher levels of induction would then prove higher exponential stack-heights (hence higher iterates of B_ω) to be defined, provided they are applied to inputs.

These arguments can be generalized, as in Ostrin–Wainer [9], [10]. For any formula $A(a)$ of EA(I;O), let $Prog\, A(a)$ express its progressiveness in the variable a, i.e. $A(0) \wedge \forall a(A(a) \rightarrow A(a+1))$.

Lemma 2.4. *For any closed term t on inputs $\vec{x}$, and any formula A, one can prove in EA(I;O) that t $\downarrow$ and $Prog\, A(a) \rightarrow \forall a \leq t.A(a)$.*

Proof. If t is basic then the lemma merely restates input induction. For more complex closed terms one proceeds by induction on their build–up.

For example, if $t = t_1 + t_2$, first consider the formula $A(b) \rightarrow b{+}a \downarrow \wedge A(b{+}a)$. Then $Prog\, A(a)$ implies that this too is progressive in a. Therefore by applying the induction hypothesis for t_2 to it, one concludes $A(b) \rightarrow \forall a \leq t_2.(b + a \downarrow \wedge A(b{+}a))$. But the induction hypothesis for t_1 gives $Prog\, A(a) \rightarrow \forall b \leq t_1.A(b)$. Hence

$$Prog\, A(a) \rightarrow \forall b \leq t_1.\forall a \leq t_2.(b + a \downarrow \wedge A(b + a))$$

from which follows $Prog\, A(a) \rightarrow \forall a \leq t_1 + t_2.A(a)$, and also $t_1 + t_2 \downarrow$ by applying it to any provably progressive formula.

A similar argument deals with the case $t = t_2 \cdot t_1$ but here one uses the formula $Prog\, A(a) \rightarrow \forall b(A(b) \rightarrow A(b{+}t_2))$ derived as above. From this one easily proves $Prog\, A(a) \rightarrow Prog\,(t_2 \cdot a \downarrow \wedge A(t_2 \cdot a))$ because if $t_2 \cdot a = b$ then $t_2 \cdot (a+1) = b + t_2$. Now one applies the induction hypothesis for t_1 to conclude $t_2 \cdot t_1 \downarrow$ and $Prog\, A(a) \rightarrow A(t_2 \cdot t_1)$. If this is applied instead to $A'(a) \equiv \forall b \leq a.A(b)$ then $Prog\, A(a) \rightarrow Prog\, A'(a)$ and the desired result follows for $t = t_2 \cdot t_1$.

For $t = t_1 + 2^{t_2}$ consider the formula $\forall a(A(a) \rightarrow a + 2^b \downarrow \wedge A(a + 2^b))$ and note that its progressiveness in b is implied by the progressiveness of A in a. By the induction hypothesis for t_2 we then have

$$Prog\, A(a) \rightarrow \forall b \leq t_2.\forall a(A(a) \rightarrow a + 2^b \downarrow \wedge A(a + 2^b))\,.$$

By the induction hypothesis for t_1 we can instantiate $a := t_1$ and, since then obtain $Prog\, A(a) \rightarrow t_1{+}2^{t_2} \downarrow \wedge A(t_1{+}2^{t_2})$. Hence, as before, EA(I;O) proves $t_1{+}2^{t_2} \downarrow$ and $Prog\, A(a) \rightarrow \forall a \leq t_1 + 2^{t_2}.A(a)$.

For other term constructs note that, just as addition depends on iterating the successor, one could equally well iterate the predecessor to deal with subtraction of terms, or iterate π_0 to decode initial segments of sequences and hence, by π_1, locate their components $t = (t_1)_{t_2}$. Such terms (on inputs only) are then provably defined, and also provably bounded by t_1. Thus since $Prog\, A(a) \rightarrow \forall a \leq t_1.A(a)$ by the induction hypothesis, one may instantiate $a := (t_1)_{t_2}$ to obtain $Prog\, A(a) \rightarrow A((t_1)_{t_2})$. Applying this instead to $A'(a) \equiv \forall b \leq a.A(b)$ one then gets $Prog\, A(a) \rightarrow \forall a \leq (t_1)_{t_2}.A(a)$ as required.

Using these results, Spoors [12] shows that $\mathrm{I}\Delta_0 + \exp$ can be embedded in EA(I;O). For each formula $B(\vec{c})$ of $\mathrm{I}\Delta_0 + \exp$ let $B[\vec{t}]$ be the formula of EA(I;O) which results by first forming its universal closure, and then bounding all previously unbounded universal quantifiers by the closed terms $\vec{t} = t_1, t_2, \ldots$ successively.

Theorem 2.5 (Spoors)**.** *If $\mathrm{I}\Delta_0 + \exp \vdash B(\vec{a})$ then for any (long enough) sequence $\vec{t}$ of closed terms, EA(I;O) $\vdash B[\vec{t}]$.*

Definition 2.6. A function $f : \mathrm{N}^k \to \mathrm{N}$ is *provably recursive* or *provably computable* in EA(I;O) if its graph has a Σ_1 defining formula $F(\vec{c}, a)$ such that on inputs $\vec{c} = \vec{x}$, EA(I;O) proves $\exists a F(\vec{x}, a)$.

Theorem 2.7. *The provably recursive functions of EA(I;O) are exactly the elementary functions.*

Proof. By Spoors' result, any Σ_1 theorem $\exists a F(\vec{c}, a)$ of $\mathrm{I}\Delta_0 + \exp$ gets embedded into EA(I;O) as $\forall \vec{c} \leq \vec{t}. \exists a F(\vec{c}, a)$. By choosing $\vec{t} = \vec{x}$ and instantiating at $\vec{c} := \vec{x}$ one immediately obtains EA(I;O) $\vdash \exists a F(\vec{x}, a)$. Thus every provably recursive function of $\mathrm{I}\Delta_0 + \exp$ becomes provably recursive in EA(I;O). But the provably recursive functions of $\mathrm{I}\Delta_0 + \exp$ are just the elementary functions.

Conversely we must show that only elementary functions are provably recursive in EA(I;O). This is fairly easy to see, and illustrates the role of B in computing bounds for existential witnesses. Briefly, the procedure goes thus:

(i) *Witnesses for Σ_1 theorems $\exists a F(n, a)$, proved by Σ_1-inductions up to $x := n$, are bounded by B_h where $h = \log n$.*

This is because for fixed n, any input induction up to $x := n$ can be unravelled, inside EA(I;O), to a binary tree of cuts of height $\log n$. If it's a Σ_1-induction on $\exists a F(c, a)$ a typical cut at height $h + 1$ in this tree will have essentially the form:

$$\frac{\exists a F(c, a) \to \exists a F(c', a) \quad \exists a F(c', a) \to \exists a F(c'', a)}{\exists a F(c, a) \to \exists a F(c'', a)}$$

where the premises are at height h. Now assume, inductively on h, that B_h bounds witnesses for both premises, i.e. if $F(c, a)$ holds (in the standard model) then $F(c', a')$ holds for some a' computable in $B_h(a)$-many steps, and similarly for c' to c''. Composing B_h will then yield a bound $B_{h+1} = B_h \circ B_h$ for the conclusion at height $h + 1$.

(ii) *Witnesses for Σ_1 theorems $\exists a F(n, a)$, proved by Π_2-inductions up to $x := n$, are bounded by $B_{2^{h \cdot d}}$ where $h = \log n$. Higher levels of induction complexity require iterated exponentials $2^{2^{h \cdot d}}$ etcetera.*

To see this, suppose EA(I;O) $\vdash \exists a F(x, a)$. Partial cut-elimination yields a "free-cut-free" proof, so only cuts on induction formulas remain. Let d be the height of this proof. Then after unravelling all inductions in favour of iterated cuts, up to the maximum input $x := n$, the height of the resulting (induction–free) proof-tree will be of the order of $\log n \cdot d$. If all cuts are Σ_1, part (i) above applies immediately to give polynomial complexity bounds $B_{\log n \cdot d}(a) = a + 2^{\log n \cdot d} = a + n^d$. Note that unary, rather than binary, representation of numerals here entails a polynomial in n, not $\log n$; hence "linear space" complexity rather than polytime. For Π_2 inductions one must first reduce all cuts to Σ_1 form before part (i) can be used. But since all inductions have been eliminated, standard Gentzen cut-reduction applies, and the price to be paid is a further exponential increase in proof–height. Thus the complexity bounds will now be of order $B_{2^{\log n \cdot d}}(a) = a + 2^{n^d}$. Higher levels of induction would require further rounds of cut-reduction, yielding iterated exponential bounds.

This completes the proof because functions computable within (finitely) iterated exponential bounds are elementary.

3 Adding an Inductive Definition

Definition 3.1. ID_1(I;O) is obtained from EA(I;O) by adding, for each uniterated positive inductive form $F(X, a)$, a new predicate P, and Closure and Least-Fixed-Point axioms:

$$\forall a(F(P, a) \to P(a))$$

$$\forall a(F(A, a) \to A(a)) \to \forall a(P(a) \to A(a))$$

for each formula A.

Example 3.2. Associate the predicate N with the inductive form:

$$F(X, a) \;:\equiv\; a = 0 \vee \exists b(X(b) \wedge a = b + 1).$$

In this way we immediately capture full Peano Arithmetic, as in Wainer–Williams [15], for the Least-Fixed-Point axiom interprets the full induction scheme of PA in ID_1(I;O) as:

$$A(0) \wedge \forall a(A(a) \to A(a+1)) \;\to\; \forall a(N(a) \to A(a)).$$

Furthermore, by similar arguments to those already used, one easily proves the progressiveness in b of the formulas $\forall a(N(a) \to a + b \downarrow \wedge N(a + b))$ and $\forall a(N(a) \to a \cdot b \downarrow \wedge N(a \cdot b))$, so by the Least-Fixed-Point axiom, ID_1(I;O) proves $\forall a, b(N(a) \wedge N(b) \to a + b \downarrow \wedge N(a + b))$ and $\forall a, b(N(a) \wedge N(b) \to a \cdot b \downarrow \wedge N(a \cdot b))$. Hence, by relativising all quantifiers to N, one interprets PA in ID_1(I;O):

Theorem 3.3. *If PA* $\vdash A(\vec{a})$ *then ID*$_1$*(I;O)* $\vdash N(\vec{a}) \to A^N(\vec{a})$.

Then, since $N(x)$ is an immediate consequence of input induction, and since ID$_1$(I;O) proves, for a bounded formula A, that its interpretation A^N entails A itself,

Corollary 3.4. *If PA* $\vdash A(\vec{a})$ *with* A *a* Σ_1*-formula then, replacing* $\vec{a}$ *by inputs* $\vec{x}$*, ID*$_1$*(I;O)* $\vdash A(\vec{x})$.

The provably recursive functions of PA are therefore provably recursive (on inputs) in ID$_1$(I;O). To show the converse we need an ordinal analysis of ID$_1$(I;O), and this can be done by following Buchholz's Ω-rule treatment of classical ID theories as in [3], [4]. However the uncountable ordinal bounds which necessarily appear there are now replaced by countable ones.

3.1 Unravelling LFP-Ax by Buchholz' Ω-Rule

We are still working in the I/O context, so can fix $\vec{x} := \vec{n}$ and unravel inductions into iterated cuts as before. However the resulting ID$_1$(I;O)-derivations will be further complicated by the presence of Least-Fixed-Point axioms. These must be "unravelled" as well, before we can read off bounds. To do this, ID$_1$(I;O) is embedded into an infinitary system ID$_1$(I;O)$^\infty$ of Tait-style sequents

$$n : I; m : O \vdash^\alpha \Gamma$$

where n bounds the input values, m declares a bound on initial output parameters, and the ordinal heights α can, for present purposes, be restricted below ε_0, with standard fundamental sequences. For shorthand we write simply $n; m \vdash^\alpha \Gamma$.

Most of the rules are unsurprising and we don't list them, but the $\exists$-rule has two premises:

$$\frac{n; m \vdash^\beta m' \qquad n; m \vdash^\beta A(m'), \Gamma}{n; m \vdash^\alpha \exists a A(a), \Gamma}$$

Here the left-hand premise "computes" witness m' from m according to the axiom $n; m \vdash^\alpha m'$ if $m' \leq m + 1$, and the computation rule:

$$\frac{n; m \vdash^\beta m'' \qquad n; m'' \vdash^\beta m'}{n; m \vdash^\alpha m'}$$

which also applies with m' replaced by any set of formulas Γ. The universal quantifier is introduced by the ω-rule and the Closure axiom of an inductive definition is re-written as a rule.

In all of these the declared input n remains fixed, and controls the ordinal assignment in the following way: if $n; m' \vdash^{\beta} \Gamma'$ is a premise of a rule with conclusion $n; m \vdash^{\alpha} \Gamma$ then $\beta \in \alpha[n]$ where $\alpha[n] = \varnothing$ if $\alpha = 0$, or $\beta[n] \cup \{\beta\}$ if $\alpha = \beta + 1$, or $\alpha_n[n]$ if α is a limit. Thus, while input n is fixed, derivations are in fact finite because $n; m \vdash^{\alpha} \Gamma$ is equivalent to $n; m \vdash^{G_\alpha(n)} \Gamma$ where $G_\alpha(n) = |\alpha[n]|$ is the "slow growing" hierarchy. However the final (crucial) rule to be added allows inputs to change, and it is this that causes growth in the extracted bounds.

Buchholz' Ω-Rule

$$\frac{n; m \vdash^{\lambda_0} P(k), \Gamma_0 \qquad m' \vdash^{h}_{0} P(k), \Delta \;\Rightarrow\; \max(n,h); \max(m,m') \vdash^{\lambda_h} \Gamma_1, \Delta}{n; m \vdash^{\lambda} \Gamma_0, \Gamma_1}$$

where Δ is any set of positive-in-P formulas, $\vdash^{h}_{0}$ signifies a cut-free derivation of finite height h (therefore independent of input n), and λ is a limit. The map $h \mapsto \lambda_h$ is then a measure of the uniformity in the transformation $\Rightarrow$.

Lemma 3.5. *The Ω rule proves the Least-Fixed-Point axioms with height $\omega + 3$.*

Proof. The gist of it is this – following Buchholz [3], [4].

For the left-hand premise of the Ω-rule choose $n; m \vdash^{0} P(k), \neg P(k)$.

For the right-hand premise, first assume $m' \vdash^{h}_{0} P(k), \Delta$. Each step of this (direct, cut-free) proof can be mimicked to derive $\max(m,m') \vdash^{d+h} \neg\forall a(F(A,a) \to A(a)), A(k), \Delta$ for some fixed d depending only on the size of the given formula A. This establishes the right-hand premise with $\lambda = \omega$, choosing the almost–standard fundamental sequence $\omega_h = d + h$ instead of $\omega_h = h$.

Therefore the Ω-rule gives $n; m \vdash^{\omega} \neg\forall a(F(A,a) \to A(a)), \neg P(k), A(k)$. This is for arbitrary k and so by $\vee$ and $\forall$ rules one obtains the Least-Fixed-Point axiom with proof-height $\omega + 3$.

Theorem 3.6. *ID_1(I;O) is embeddable into ID_1(I;O)$^\infty$ with derivation heights $<$ $\omega \cdot 2$.*

3.2 Cut Elimination and Collapsing in ID_1(I;O)$^\infty$

As usual, Gentzen-style cut-reduction increases height exponentially, and so by the embedding, every theorem of ID_1(I;O) is cut–free derivable in ID_1(I;O)$^\infty$ with ordinal height $< \varepsilon_0$.

Lemma 3.7 (Collapsing)**.** *Given a cut-free ID_1(I;O)$^\infty$ derivation $n; m \vdash^{\alpha}_{0} \Gamma$ where Γ contains only positive occurrences of inductively defined predicates P, then $m \vdash^{h}_{0} \Gamma$ where $h < B_{\alpha+1}(n)$ for $n > 0$.*

Proof. By induction on α. Suppose $n; m \vdash_0^\alpha \Gamma$ comes about by an application of the Ω-rule from premises

$$n; m \vdash_0^{\alpha_0} P(k), \Gamma_0 \text{ and } m' \vdash_0^h P(k), \Delta \Rightarrow \max(n,h); \max(m, m') \vdash_0^{\alpha_h} \Gamma_1, \Delta .$$

Then, applying the induction hypothesis to the left premise, $m \vdash_0^h P(k), \Gamma_0$ where $h < B_{\alpha_0+1}(n)$. Applying the right premise, with this h and $\Delta = \Gamma_0$, $m' = m$, we obtain $\max(n, h); m \vdash_0^{\alpha_h} \Gamma$. The induction hypothesis can now be applied to this and yields $m \vdash_0^{h'} \Gamma$ where $h' < B_{\alpha_h+1}(\max(n, h))$. Hence, replacing h by its bound $B_{\alpha_0+1}(n)$, and using basic monotonicity properties of the B hierarchy, we get the required

$$h' < B_{\alpha_h+1}(B_{\alpha_0+1}(n)) \le B_{\alpha_{h+1}}(B_{\alpha_1}(n)) \le B_\alpha B_\alpha(n) = B_{\alpha+1}(n) .$$

All other cases are straightforward because if $n; m \vdash_0^\alpha \Gamma$ arises by any other rule from premises $n; m' \vdash_0^\beta \Gamma'$ then the induction hypothesis gives $m' \vdash_0^{h'} \Gamma'$ with $h' < B_{\beta+1}(n)$ and then the rule may be re-applied to give $m \vdash_0^h \Gamma$ with $h = B_{\beta+1}(n)$. But since $\beta \in \alpha[n]$ we then have $h \le B_\alpha(n) < B_{\alpha+1}(n)$ and this completes the proof.

Theorem 3.8. *ID_1(I;O) has the same provably recursive functions as PA.*

Proof. We have already shown that every function provably recursive in PA is provable also in ID_1(I;O). Conversely, suppose $f(\vec{x})$ has a Σ_1 graph $\exists b F(\vec{x}, a, b)$ such that $\exists a, b F(\vec{x}, a, b)$ is provable in ID_1(I;O). Then by the Embedding and Cut Elimination, there is an α below ε_0 such that for all $\vec{n}$, there is a $\text{ID}_1(\text{I;O})^\infty$ derivation of $\max \vec{n}; 0 \vdash_0^\alpha \exists a, b F(\vec{n}, a, b)$. By the Collapsing Lemma we then have $0 \vdash_0^h \exists a, b F(\vec{n}, a, b)$ with $h < B_{\alpha+1}(\max \vec{n})$. Since this derivation is cut-free and of finite height, our original bounding principle applies to give witnesses $a, b \le B_h(\max \vec{n})$ satisfying $F(\vec{n}, a, b)$. Therefore the value of $f(\vec{n})$ is computable by bounded search and will be elementary in any such bound. Replacing h by $B_{\alpha+1}(\max \vec{n})$, this bound becomes $\le B_h(h) = B_\omega(h) < B_\omega \circ B_{\alpha+1}(\max \vec{n}) \le B_{\alpha+2}(\max \vec{n})$. Hence f, being elementary in a level of the fast-growing hierarchy below ε_0, is provably recursive in PA.

4 Generalizing to $\text{ID}_{<\omega}$

Williams' thesis [16] generalizes the foregoing to theories of finitely iterated inductive definitions ID_i(I;O), still retaining the input/output discipline. As the inductive definitions are iterated, the higher levels of Ω-rules needed to unravel them are controlled by "tree-ordinals" in successively higher number-classes $\Omega_0 = \mathrm{N} \subset \Omega_1 = \Omega \subset \Omega_2 \subset \ldots \subset \Omega_i$. These are generated inductively by:

$$\alpha \in \Omega_{i+1} \text{ if } \alpha = 0 \vee \exists \beta \in \Omega_{i+1}(\alpha = \beta + 1) \vee \exists j \leq i(\alpha : \Omega_j \to \Omega_{i+1})\,.$$

Each Ω_{i+1} is partially ordered by the "sub-tree" ordering $\prec$. Furthermore, all tree-ordinals used here will be "structured" in the sense that all limit sub-trees $\lambda : \Omega_j \to \Omega_{i+1}$ are monotone with respect to $\prec$, when restricted to structured elements of Ω_j. For more on tree-ordinals and their uses in this context, see e.g. Fairtlough–Wainer [5], Wainer [14].

The infinitary system $\mathrm{ID}_{i+1}(\mathrm{I;O})^\infty$ then has (Tait-style) sequents of the form

$$\gamma_i : \Omega_i, \ldots, \gamma_1 : \Omega_1, n : I; m : O \vdash^\alpha \Gamma$$

where $\alpha \in \Omega_{i+1}$, and this is abbreviated $\vec{\gamma}, n; m \vdash^\alpha \Gamma$. The rules are generalized versions of the rules for $\mathrm{ID}_1(\mathrm{I;O})^\infty$. The underlying ordinal assignment principle, for all but the Ω_j rules, is that if $\vec{\gamma}, n; m \vdash^\alpha \Gamma$ is the conclusion of a rule with premises $\vec{\gamma'}, n'; m' \vdash^\beta \Gamma'$ then $\beta \in \alpha[\vec{\gamma}, n]$ where $\alpha[\vec{\gamma}, n] = \varnothing$ if $\alpha = 0$, $= \beta[\vec{\gamma}, n] \cup \{\beta\}$ if $\alpha = \beta + 1$, and $= \alpha_{\gamma_j}[\vec{\gamma}, n]$ if $\alpha : \Omega_j \to \Omega_{i+1}$ is a "limit". Note that if $\alpha \in \Omega_j$ for some $j \leq i$ then $\alpha[\vec{\gamma}, n] = \alpha[\gamma_{j-1}, \ldots, n]$ and so the initial declared parameters $\gamma_i : \Omega_i, \ldots, \gamma_j : \Omega_j$ become redundant.

There are Buchholz Ω_j-rules for each $j = 1, \ldots, i+1$, with P_j being the predicate defined by a j-times iterated induction (allowing negative occurrences of $P_{j'}$ for $j' < j$). The Ω_{i+1} rule takes two premises:

$$\vec{\gamma}, n; m \vdash^{\lambda_0} P_{i+1}(k), \Gamma_0$$

and, for all $\delta \in \Omega_i$ and all sets Δ of positive-in-P_{i+1} formulas,

$$\vec{\gamma}, n'; m' \vdash^\delta_0 P_{i+1}(k), \Delta \;\Rightarrow\; \vec{\gamma}(\gamma_i := \delta), \max(n, n'); \max(m, m') \vdash^{\lambda_\delta} \Gamma_1, \Delta\,.$$

The conclusion is $\vec{\gamma}, n; m \vdash^\lambda \Gamma_0, \Gamma_1$.

Collapsing from one level $i+1$ down to the one below is then computed in terms of higher-level extensions of the B_α hierarchy: $\varphi^{(i)}_\alpha(\beta)$ for $\alpha \in \Omega_{i+1}, \beta \in \Omega_i$ defined by

$$\varphi^{(i)}_\alpha(\beta) = \begin{cases} \beta + 1 & \text{if } \alpha = 0 \\ \varphi^{(i)}_{\alpha'} \circ \varphi^{(i)}_{\alpha'}(\beta) & \text{if } \alpha = \alpha' + 1 \\ \varphi^{(i)}_{\alpha_\beta}(\beta) & \text{if } \alpha : \Omega_i \to \Omega_{i+1} \\ \xi \mapsto \varphi^{(i)}_{\alpha_\xi}(\beta) & \text{if } \alpha : \Omega_j \to \Omega_{i+1} \text{with } j < i\,. \end{cases}$$

Note that if $\alpha \in \Omega_i$ then $\varphi^{(i)}_\alpha(\beta) = \beta + 2^\alpha$, so taking $\omega_i \in \Omega_{i+1}$ to be the identity function on Ω_i one obtains $\varphi^{(i)}_{\omega_i}(\beta) = \varphi^{(i)}_\beta(\beta) = \beta + 2^\beta$ which serves as a bound for each round of cut reduction.

Lemma 4.1 (Cut reduction). *If $\vec{\gamma}, n; m \vdash^\alpha \Gamma$ in ID$_{i+1}$(I;O)$^\infty$ with cut rank $r+1$ then $\vec{\gamma}, n; m \vdash^{\alpha'} \Gamma$ with cut rank r where $\alpha' = \varphi^{(i+1)}_{\omega_{i+1}}(\alpha)$.*

Lemma 4.2 (Collapsing). *If $\vec{\gamma}, n; m \vdash^\alpha_0 \Gamma$ in ID$_{i+1}$(I;O)$^\infty$ where Γ is positive in P_{i+1} then $\vec{\gamma}, n; m \vdash^\delta_0 \Gamma$ where $\delta \prec \varphi^{(i)}_{\alpha+1}(\gamma_i) \in \Omega_i$. Note that since $\delta \in \Omega_i$ no Ω_{i+1} rules remain.*

Proof. The proof goes as before for ID$_1$(I;O)$^\infty$, all cases being straightforward except for the Ω_{i+1} rule. In that case one may apply the induction hypothesis to the first premise, yielding

$$\vec{\gamma}, n; m \vdash^\delta_0 P_{i+1}(k), \Gamma_0$$

where $\delta \prec \varphi^{(i)}_{\lambda_0+1}(\gamma_i)$. Next apply the second premise to transform this into a derivation

$$\vec{\gamma}(\gamma_i := \delta), n; m \vdash^{\lambda_\delta}_0 \Gamma$$

and note that the declared parameters $\vec{\gamma}(\gamma_i := \delta), n; m$ can be "weakened" to $\vec{\gamma}(\gamma_i := \delta'), n; m$ with $\delta' = \varphi^{(i)}_\lambda(\gamma_i)$ because $\delta \prec \varphi^{(i)}_{\lambda_{\gamma_i}}(\gamma_i) = \varphi^{(i)}_\lambda(\gamma_i)$ provided $1 \preceq \gamma_i$. Now the induction hypothesis can again be applied to give $\vec{\gamma}, n; m \vdash^{\delta''}_0 \Gamma$ since the first declared parameter γ_i is immaterial as the ordinal bound $\delta'' \in \Omega_i$. We then have

$$\delta'' \prec \varphi^{(i)}_{\lambda_\delta+1}(\delta') \preceq \varphi^{(i)}_{\lambda_{\delta'}}(\delta') = \varphi^{(i)}_\lambda(\delta') = \varphi^{(i)}_\lambda(\varphi^{(i)}_\lambda(\gamma_i)) = \varphi^{(i)}_{\lambda+1}(\gamma_i)$$

as required.

Definition 4.3. The countable tree-ordinals $\tau_i \in \Omega_1$ are

$$\tau_1 \;=\; \varphi^{(1)}_\omega(\omega);\quad \tau_2 \;=\; \varphi^{(1)}_{\varphi^{(2)}_\omega(\omega_1)}(\omega);\quad \tau_3 \;=\; \varphi^{(1)}_{\varphi^{(2)}_{\varphi^{(3)}_\omega(\omega_2)}(\omega_1)}(\omega)\ \text{ etc}\,.$$

Then $\tau_1 = \omega + 2^\omega$, τ_2 is a version of ε_0 and τ_3 is a version of the Bachmann–Howard ordinal. B_{τ_i} is in fact a functor on the category N with τ_{i+1} its direct limit, or conversely, B_{τ_i} is the "slow growing" collapse of τ_{i+1} (see Wainer [13], [14]).

To see how these ordinals arise, suppose for example that ID$_2$(I;O) $\vdash A(x,a)$ where A contains no inductive predicates. This embeds into ID$_2$(I;O)$^\infty$ as $\omega : \Omega_1, n : I; m : O \vdash^{\omega_1+d} A(n,m)$, for all $n; m$, say with cut rank r. The $\omega_1 + d$ can be weakened to a d-times iterate of $\varphi^{(2)}_{\omega_2}$ applied on ω_1. A further r-times iterate then yields a bound α for a cut-free derivation of $A(n,m)$, namely

$$\alpha = \varphi^{(2)}_{\omega_2} \circ \ldots \circ \varphi^{(2)}_{\omega_2}(\omega_1) \prec \varphi^{(2)}_{\omega_2+2^{(d+r)}}(\omega_1) = \varphi^{(2)}_{\varphi^{(3)}_{(d+r)}(\omega_2)}(\omega_1)\,.$$

Collapsing provides a countable bound $\varphi^{(1)}_{\alpha+1}(\omega) \prec \varphi^{(1)}_{\varphi^{(2)}_{\varphi^{(3)}_{(d+r)}(\omega_2)}(\omega_1)}(\omega) \prec \tau_3$.

Theorem 4.4 (Williams). *In summary:*

- *Classical ID$_i$ is interpretable in ID$_{i+1}$(I;O).*
- *Its ordinal bound is τ_{i+2} in the notation above.*
- *ID$_{i+1}$(I;O) has the same provably recursive functions as ID$_i$.*
- *The provably recursive functions are those computable within B_α-bounded resource, for $\alpha < \tau_{i+2}$.*
- *ID$_{<\omega}$ and ID$_{<\omega}$(I;O) are mutually interpretable.*

Remark 4.5. Another way to view the relationship between ID$_i$ and ID$_i$(I;O) is in terms of the hierarchies which generate their provably recursive functions. For ID$_i$ it's the fast growing hierarchy below τ_{i+2}, whereas for ID$_i$(I;O) it is the slow growing hierarchy below that same ordinal. Arai [1] was the first to analyse ID theories in this light.

5 An Independence Result – Kruskal's Theorem

Kruskal's Theorem states that every infinite sequence $\{T_i\}$ of finite trees has an $i < j$ such that T_i is embeddable in T_j. By "finite tree" is meant a rooted (finite) partial ordering in which the nodes below any given one are totally ordered. An embedding of T_i into T_j is then just a one-to-one function from the nodes of T_i to nodes of T_j preserving infs (greatest lower bounds).

Friedman showed this theorem to be independent of the theory ATR$_0$ and went on, in [6], [7], to develop a significant extension of it which is independent of Π^1_1-CA$_0$. The Extended Kruskal Theorem concerns finite trees in which the nodes carry labels from a fixed finite list $\{0, 1, 2, \ldots, k\}$. By a more delicate argument, he proved that for any k, every infinite sequence $\{T_i\}$ of finite $\leq k$-labelled trees has an embedding $T_i \hookrightarrow T_j$ where $i < j$. However the notion of embedding is now more complex. $T_i \hookrightarrow T_j$ means that there is an embedding f in the former sense, but which also preserves labels and satisfies the "gap condition" which states: if node x comes immediately below node y in T_i, and if z is an intermediate node strictly between $f(x)$ and $f(y)$ in T_j, then the label of z must be $\geq$ the label of $f(y)$.

Both of these statements are Π^1_1, but Friedman showed that they can be miniaturized to an arithmetical Π^0_2 form which still reflects the proof-theoretic strength of the original results. See Simpson [11] for an excellent exposition.

The Miniaturized Kruskal Theorem for labelled trees runs as follows: For any number c and fixed k there is a number $K_k(c)$ so large that for every sequence $\{T_i\}$ of finite $\leq k$-labelled trees of length $K_k(c)$, and where each T_i is bounded in size by $\|T_i\| \leq c \cdot (i+1)$, there is an embedding $T_i \hookrightarrow T_j$ with $i < j$. In fact we shall consider a slight variant of this - where the size restriction $\|T_i\| \leq c \cdot (i+1)$ is weakened to $\|T_i\| \leq c \cdot 2^i$. Friedman showed that, by slowing down the sequence, 2^i may be replaced by $i+1$ without affecting the result's proof theoretic strength. An application of König's Lemma proves that the miniaturized version is a consequence of the full theorem.

In this section we give a proof that the Miniaturized Kruskal Theorem for labelled trees is independent of $\mathrm{ID}_{<\omega}$. Since Π^1_1-CA_0 is conservative over $\mathrm{ID}_{<\omega}$ for arithmetical sentences, both the miniaturized and the full Kruskal theorems are therefore independent of Π^1_1-CA_0 . Our proof again serves to illustrate the fundamental role played by the B-hierarchy. It consists in showing directly that the computation sequence for the "slow-growing" function $G_{\tau_k}(n) = |\tau_k[n]|$ is bad (i.e. has no embeddings). Since, by Wainer [13], $G_{\tau_k}(n) = B_{\tau_{k-1}}(n)$ it follows that for all k, n, $B_{\tau_{k-1}}(n) < K_k(c_k(n))$ for a suitably small $c_k(n)$. Therefore from the last section one sees immediately that the function K cannot be provably recursive in $\mathrm{ID}_{<\omega}$.

5.1 φ-terms, trees and i-sequences

Henceforth we shall regard the φ-functions as function symbols and use them, together with the constants $0, \omega_j$, to build terms. Each such term will of course denote a (structured) tree ordinal, but it is important to lay stress, in this section, upon these terms rather than the tree ordinals which they denote.

Definition 5.1. An *i-term*, for $i > 0$, is either ω_{i-1} or else of the form $\varphi^{(i)}_\alpha(\beta)$ (alternatively written $\varphi^{(i)}(\alpha, \beta)$) where β is an i-term and α is a j-term with $j \leq i+1$. (0-terms are just numerals $\bar{n}$ built from 0 by repeated applications of the successor $\varphi^{(0)}$ which has no subscript.) Note that each i-term may be viewed as a finite labelled tree whose root has label i, whose left hand subtree is the tree α and whose right hand subtree is the tree β. The tree ω_{i-1} consists of a single node labelled i, and the zero tree is the single node labelled 0. We often indicate the level i of a term γ by writing γ^i. Thus as tree ordinals, $\omega_{i-1} \preceq \gamma^i \in \Omega_i$.

Definition 5.2. For each $\leq i$-term γ and $i-1$-term ξ^{i-1} (assuming $i > 1$) we denote the term $\varphi_\gamma^{(i-1)}(\xi)$ by simply $\gamma(\xi)$ (or $\bar{n}+1$ if $i = 1$ and $\xi = \bar{n}$). With association to the left, a typical i-term then would be written as

$$\nu(\xi^{i_r})(\xi^{i_{r-1}})\dots(\xi^{i_1})(\xi^i)$$

where ν (the "indicator") is either 0 or an ω_j. In particular, the tree-ordinal τ_k may be written

$$\tau_k = \varphi^{(1)}(\varphi^{(2)}(\dots\varphi^{(k)}(\omega_0, \omega_{k-1})\dots, \omega_1), \omega_0)$$

and can then be denoted $\omega_0(\omega_{k-1})(\omega_{k-2})\dots(\omega_0)$.

Definition 5.3. The *computation sequence* starting with τ_k and fixed input n is the sequence of 1-terms and numerals generated according to the computation rules for the φ-functions, as follows:

$$\gamma = \nu(\xi^{i_r})(\xi^{i_{r-1}})\dots(\xi^{i_1})(\xi^1)$$

reduces (or rewrites) in one step to

$$\delta = \begin{cases} \xi^{i_r}(\xi^{i_r}(\xi^{i_{r-1}}))\dots(\xi^{i_1})(\xi^1) & \text{if } \nu = 0, \\ \xi^{i_j}(\xi^{i_r})(\xi^{i_{r-1}})\dots(\xi^{i_1})(\xi^1) & \text{if } \nu = \omega_{i_j} \text{ and } i_j < i_{j+1}, \dots, i_r, \\ \bar{n}(\xi^{i_r})(\xi^{i_{r-1}})\dots(\xi^{i_1})(\xi^1) & \text{if } \nu = \omega_0. \end{cases}$$

If $\gamma = \omega_0$ it reduces to $\bar{n}$, then to $\overline{n-1}$ etc. until it reaches 0 and stops. We henceforth omit the overbar from numerals.

Definition 5.4. Let *level* i of the computation sequence be what remains after stripping away, from each term of the form $\gamma(\xi^{i-1})\dots(\xi^1)$, the outermost $(\xi^{i-1})\dots(\xi^1)$, thus leaving γ alone. Now suppose γ occurs in level i of the computation sequence from τ_k and n (thus γ is a j-term for some $j \leq i$). Then the *i-sequence* from that occurrence of γ consists of all succeeding level i terms as far as the first zero. Write $\gamma \to^i \delta$ to indicate that γ precedes (or is) δ in the same i-sequence. Note that there is just one 1-sequence – the computation sequence itself.

Lemma 5.5. *One can show, for each fixed τ_k and n:*

- *The computation sequence starting with τ_k and n is finite.*

- *The length of the computation sequence is greater than the number of successor ordinals encountered in the reduction process, i.e. greater than the cardinality of the set of tree-ordinals $\tau_k[n]$, which by definition is exactly $G_{\tau_k}(n)$.*
- *The r-th member of the computation sequence from τ_k and n is bounded in size by $c_k(n) \cdot 2^r$ where $c_k(n)$ is* $\max(2k+1, n)$.
- *Each i-sequence is non-repeating and non-increasing with respect to the tree-ordinals denoted.*

5.2 The computation sequence is bad

Definition 5.6. $\gamma \hookrightarrow^+ \delta$ means that, as labelled trees, $\gamma \hookrightarrow \delta$ (i.e. γ is embeddable in δ, preserving labels, infs and satisfying the gap condition) and furthermore, if γ is a j'-term, the embedding does not completely embed γ inside any j-subterm of δ where $j < j'$.

Lemma 5.7. *Fix τ_k and n. Then for each i with $1 \leq i \leq k+1$ and every term δ, if $\gamma \to^i \delta$ and $\gamma \hookrightarrow^+ \delta$ then γ and δ are identical.*

Proof. By induction on i from $k+1$ down to 1, and within that an induction over the term or tree δ, and within that a subinduction over γ.

For the basis $i = k+1$, the $k+1$-sequences are just descending sequences of integers $\leq n$, so no term can be $\hookrightarrow^+$ embedded in any follower.

Now suppose $1 \leq i < k$ and assume the result for $i+1$. We proceed by induction on the term δ. If $\delta = \omega_j$ or 0 and $\gamma \hookrightarrow^+ \delta$ the only possibility is γ is δ. Suppose then, that δ is of the form $\varphi^{(j)}_\alpha(\beta)$. Then γ cannot be $\omega_{j'}$ for any $j' \varepsilon j$ because $\gamma \hookrightarrow^+ \delta$, and it cannot be $\omega_{j'}$ with $j' < j$ because none of its successors in the i-sequence could then be j-terms. Thus γ is also of the form $\varphi^{(j')}_{\alpha'}(\beta')$. By $\gamma \hookrightarrow^+ \delta$ we have $j' \leq j$ and by $\gamma \to^i \delta$ we have $j' \varepsilon j$, so $j' = j$. Also, we cannot have $\beta' \to^i \delta$ for otherwise, by the gap condition, $\gamma \hookrightarrow^+ \delta$ implies $\beta' \hookrightarrow^+ \delta$, so by the sub-induction hypothesis β' and δ would be identical, and then γ would contain δ as a proper sub-term, contradicting $\gamma \hookrightarrow \delta$.

The situation then, is this: $\gamma = \varphi^{(j)}_{\alpha'}(\beta')$, $\delta = \varphi^{(j)}_\alpha(\beta)$, $\gamma \to^i \delta$ and $\gamma \hookrightarrow^+ \delta$. Furthermore β must be of the form $\varphi^{(j)}_{\alpha_r} \ldots \varphi^{(j)}_{\alpha_2} \varphi^{(j)}_{\alpha_1}(\beta')$ where, as tree ordinals, $\alpha \prec \alpha_r \prec \ldots \prec \alpha_2 \prec \alpha_1 \prec \alpha'$.

Now there are four possible ways in which γ can embed in δ, only two of which actually happen.

Case 1. $\gamma \hookrightarrow^+ \beta$. Then $\gamma \to^i \delta \to^i \beta$ belong to the same i-sequence, so by the induction hypothesis γ is then identical to β. Therefore the ordinal denoted by γ

is strictly less than the ordinal of δ. But this is impossible because i-sequences are non-increasing.

Case 2. $\gamma \hookrightarrow^+ \alpha$. Then let η denote the smallest j-subterm of α such that $\gamma \hookrightarrow^+ \eta$. This occurrence of η in the subscript α of δ must be created anew as the i-sequence proceeds from γ to δ. The only way this can happen is that at some intervening stage a $\varphi^{(j)}_{\alpha''}(\beta'')$ occurs, where the indicator ν of α'' is ω_j. The next stage replaces ν by β'' and then β'' reduces to a j-subterm of α which contains η. Call this j-subterm η'. But this reduction from β'' to η', although it occurs at the level of $\varphi^{(j)}$-subscripts, must also occur in level i itself, and within the same i-sequence. Hence $\gamma \to^i \varphi^{(j)}_{\alpha''}(\beta'') \to^i \beta''$ and $\beta'' \to^i \eta'$. Also $\gamma \hookrightarrow^+ \eta'$. Thus by the induction hypothesis, η' being a proper subterm of δ, we have γ identical to η', and since the i-sequence is ordinally non-increasing this means that the ordinal of γ is not greater than the ordinal of β''. This is impossible however, because $\gamma \to^i \varphi^{(i)}_{\alpha''}(\beta'')$ and so γ is ordinally greater than β''.

Case 3. $\gamma \hookrightarrow^+ \delta$ where the embedding takes the root of γ to the root of δ and $\alpha' \hookrightarrow \beta$ and $\beta' \hookrightarrow \alpha$. By the gap condition, since β' and β are j-terms, α must be either a j-term or a $j+1$-term and α' a j'-term with $j' \leq j$. But since α' comes before α in the reduction sequence, α' cannot be a j'-term and α a j-term where $j' < j$. Therefore both α' and α are j-terms. Now the only way in which α' could arise as a $\varphi^{(j)}$-subscript is by means of an earlier diagonalization at level i: $\varphi^{(j)}_{\omega_j}(\xi) \to^i \varphi^{(j)}_{\xi}(\xi)$ followed by further reductions to $\varphi^{(j)}_{\alpha'}(\beta')$ where $\beta' = \varphi^{(j)}_{\alpha_r} \dots \varphi^{(j)}_{\alpha_1}(\xi)$ and ξ reduces to α'. However this reduction between level-j subscripts must occur also at level i and consequently $\beta' \to^i \xi \to^i \alpha' \to^i \alpha$. Because of the gap condition, $\beta' \hookrightarrow \alpha$ implies $\beta' \hookrightarrow^+ \alpha$, and so by the induction hypothesis, β' and α are identical. But this means that α' and α are identical, and so γ and δ are identical at level i.

Case 4. $\gamma \hookrightarrow^+ \delta$ where the embedding takes the root of γ to the root of δ and $\alpha' \hookrightarrow \alpha$ and $\beta' \hookrightarrow \beta$. If $j = i$ then $\alpha' \to^{i+1} \alpha$. Since, by the gap condition, $\alpha' \hookrightarrow^+ \alpha$, it follows from the induction hypothesis for $i+1$ that α' and α are identical. If $j < i$ then, as before, the reduction $\alpha' \to \alpha$ takes place also in level i so the sub-induction hypothesis implies again that α' and α are identical. Therefore γ and δ are identical too and this completes the proof.

Theorem 5.8. *The computation sequence from τ_k and n is a bad sequence, and therefore its length is bounded by the Kruskal function $K_k(c_k(n))$. Hence $G_{\tau_k}(n) < K_k(c_k(n))$.*

Proof. Apply the lemma with $i = 1$, noting that if γ and δ are 1-terms then $\gamma \hookrightarrow \delta$ automatically implies $\gamma \hookrightarrow^+ \delta$ since 1-terms never get inserted inside numerals.

Thus if γ came before δ in the computation sequence we could not have $\gamma \hookrightarrow \delta$ because then they would be identical, contradicting the previous lemma which says there can be no repetitions.

Corollary 5.9. *Neither Kruskal's theorem for labelled trees, nor its miniaturized version, is provable in* Π^1_1-CA_0.

References

[1] T. Arai, *A slow growing analogue of Buchholz' proof*, Annals of Pure and Applied Logic 54 (1991), 101-120.

[2] S. Bellantoni, S. Cook, *A new recursion theoretic characterization of the polytime functions*, Computational Complexity 2 (1992), 97-110.

[3] W. Buchholz, *An independence result for (*Π^1_1*-CA)+BI*, Annals of Pure and Applied Logic 33 (1987), 131-155.

[4] W. Buchholz, S. Feferman, W. Pohlers, W. Sieg, *Iterated Inductive Definitions and Subsystems of Analysis: Recent Proof Theoretical Studies*, Springer Lecture Notes in Mathematics 897, Springer-Verlag (1981).

[5] M.V. Fairtlough, S.S. Wainer, *Hierarchies of provably recursive functions*, in S. Buss (Ed) Handbook of Proof Theory, Elsevier Science BV (1998), 149-207.

[6] H. Friedman, *Beyond Kruskal's theorem I - III*, unpublished manuscripts, Ohio State University (1982).

[7] H. Friedman, N. Robertson, P. Seymour, *The metamathematics of the graph minor theorem*, in S.G. Simpson (Ed) Logic and Combinatorics, AMS Contemporary Mathematics 65 (1987), 229-261.

[8] D. Leivant, *Intrinsic theories and computational complexity*, in D. Leivant (Ed) Logic and Computational Complexity, Springer Lecture Notes in Computer Science 960, Springer-Verlag (1995), 177-194.

[9] G.E. Ostrin, S.S. Wainer, *Proof theoretic complexity*, in H. Schwichtenberg, R. Steinbrüggen (Eds) Proof and System Reliability, Kluwer Academic (2002), 369-397.

[10] G.E. Ostrin, S.S. Wainer, *Elementary arithmetic*, Annals of Pure and Applied Logic 133 (2005), 275-292.

[11] S.G. Simpson, *Nonprovability of certain combinatorial properties of finite trees*, in L. Harrington, M. Morley, A. Scedrov, S.G. Simpson (Eds) Harvey Friedman's Research in the Foundations of Mathematics, North-Holland Studies in Logic (1985), 87-117.

[12] E. Spoors, *Notes on EA(I;O)*, University of Leeds (2009).

[13] S.S. Wainer, *Slow growing versus fast growing*, Journal of Symbolic Logic 54 (1989), 608-614.

[14] S.S. Wainer, *Accessible recursive functions*, Bulletin of Symbolic Logic 5 (1999), 367-388.

[15] S.S. Wainer, R.S. Williams, *Inductive definitions over a predicative arithmetic*, Annals of Pure and Applied Logic 136 (2005), 175-188.

[16] R.S. Williams, *Finitely iterated inductive definitions over a predicative arithmetic*, Ph.D. dissertation, University of Leeds (2004).