Provability algebras and proof-theoretic ordinals, I

Lev D. Beklemishev*

Steklov Mathematical Institute, Moscow
and Utrecht University
e-mail: Lev.Beklemishev@phil.uu.nl

February 28, 2001

Abstract

We suggest an algebraic approach to proof-theoretic analysis based on the notion of *graded provability algebra*, that is, Lindenbaum boolean algebra of a theory enriched by additional operators which allow for the structure to capture proof-theoretic (syntax-sensitive) information. We use this method to analyze Peano arithmetic and show how an ordinal notation system up to ϵ_0 can be recovered from the corresponding algebra in a canonical way. This method also establishes links between proof-theoretic ordinal analysis and the work which has been done in the last two decades on provability logic and reflection principles. Because of its abstract algebraic nature, we hope that it will also be of interest for non-prooftheorists.

1 Introduction

A well-known and old problem in proof theory is the one of canonical or 'natural' ordinal notations (see [10, 8, 7] for a discussion and [14] for a general background on ordinal analysis). Rather than being stated in a precise mathematical way, this problem is of a conceptual nature. Historically, primitive recursive ordinal notation systems were used to give consistency proofs for formal theories capturing significant parts of mathematics, such as Peano arithmetic or predicative analysis. Later it was realized that it is the property of a *description* of an ordinal notation system rather than just of its ordinal that accounts for the possibility of a consistency proof: for any sentence φ one can artificially define a primitive recursive ordering of type ω such that the corresponding induction schema implies consistency of φ. So, consistency proof results yield meaningful ordinals only for specific 'canonical' ordinal notation systems. A problem

*Supported by RFBR grants 98-01-00282, 98-01-00249.

of similar nature appears in the program of classification of (provably total) computable functions by means of subrecursive hierarchies.

The picture becomes more stable for Π^1_1 proof-theoretic ordinals which make sense for second and higher order systems: the notion of the proof-theoretic ordinal of a theory is correctly defined (e.g., as the supremum of order types of provably well-founded primitive recursive ordering relations). Thus, ordinal analysis in complexity Π^1_1 stands on a rigorous basis. Still, there remains a question what it means to *find* the proof-theoretic ordinal of a formal system? In what terms does it have to be *specified*? What does a primitive recursive description of this ordinal really tell us? This brings us back to the problem of choice of a specific ordinal notation system out of many.

Summarizing, at present we lack general criteria which would separate 'natural' from 'pathological' ordinal notation systems. This leaves the fundament of ordinal analysis (for logical complexities below Π^1_1) in an unsatisfactory state. It also makes proof-theoretic results too much dependent on the details of syntax and contributes to the much criticized *lack of modularity* of proof-theory [8].

The problem of canonical ordinal notations is sometimes formulated as the question whether a natural ordinal classification of *all* total computable functions exists. (A proof-theoretic analog of this problem, the question of classifying arithmetical sentences by hierarchies of iterated reflection principles, was first considered by Turing in [18]. To the best of my knowledge, Turing's was the first discussion of the problem of canonical ordinal notations, though the appropriate terminology has been coined later.)

Whether such a global classification exists is rather doubtful, at least at present there is no evidence supporting a possible positive solution of this question. However, in a significant number of cases there do exist positive solutions of a restricted version of this question — ordinal classifications of the classes of provably total computable functions of particular (mathematically and foundationally meaningful) formal theories.

In this paper we are concerned with the question of recovering an ordinal notation system, not just an ordinal, from a given theory. Formal theories are often identified with sets of formulas in a first order language. However, this point of view is too rough for our present goals: proof-theoretic ordinals are sensitive to the choice of particular proof systems, in other words, to the information of intensional character. The question arises, what kinds of data are relevant for a meaningful proof-theoretic analysis. What is the most general concept of a formal theory, which would make it possible to rigorously specify its canonical ordinal notation system?

To approach this question in a systematic way we pursue an algebraic view of proof theory provided by *provability algebras*, that is, by Lindenbaum boolean algebras enriched by additional provability operators. Pure Lindenbaum algebras of all reasonable theories are countable atomless boolean algebras. They are pairwise isomorphic[1] and therefore too poor to distinguish between, say, PA

[1] And even recursively isomorphic considered as numerated structures. This follows from M. B. Pour-El and S. Kripke [13].

and ZF, let alone to capture proof-theoretic ordinals. We show how to suitably enrich the expression power of these structures.

We consider the test case of PA and come up with a well-behaved notion of *graded provability algebra*. With this algebra a certain ordinal *characteristic* and the corresponding ordinal notation system is associated in a canonical way. A proof-theoretic analysis of PA, including Gentzen's consistency proof by transfinite induction up to ϵ_0 and the analysis of provably total recursive functions, is then obtained in a quite simple and abstract form. We return to the discussion of the general problem of recovering an ordinal notation system from a formal theory in the last section of this paper.

2 Background concepts

One of our aims is to show that these algebraic methods naturally arise from traditional proof-theoretic questions. Therefore, we present more background information than is technically necessary just for analyzing Peano arithmetic.

As our basic fragment of arithmetic we take *elementary arithmetic* EA. The precise formulation of EA is not important, for definiteness we specify the language of EA as that of Peano arithmetic augmented by a symbol exp for the function 2^x and a symbol $\leq$. Δ_0-formulas in the language of EA are those with all quantifier occurrences bounded by terms. Π_n- and Σ_n-formulas are obtained from Δ_0 by adding a quantifier prefix in the standard way. Axioms of EA consist of some minimal set of open defining axioms for all the symbols of the language and the induction schema for Δ_0-formulas. Peano arithmetic PA can be obtained from EA by adding the full induction schema. We shall also use an extension of EA by an axiom stating that the superexponentiation function $\exp^{(x)}(x)$ is total, denoted EA^+. EA and EA^+ are finitely axiomatizable fragments of primitive recursive arithmetic PRA. For most of the paper the reader can use PRA instead of EA or EA^+ while suitably weakening the formulations of the results. EA seems to be a natural lower bound for which our methods can work without substantial changes.

2.1 Provably total computable functions

One of the central notions in proof theory is that of a provably total computable function of a formal system. With a system T extending[2] EA we can associate the class $\mathcal{F}(T)$ of all functions $f : \mathbb{N}^k \to \mathbb{N}$ such that for some Σ_1-formula $\varphi(\vec{x}, y)$ there holds:

(i) $f(\vec{x}) = y \iff \mathbb{N} \vDash \varphi(\vec{x}, y)$;

(ii) $T \vdash \forall \vec{x}\, \exists y\, \varphi(\vec{x}, y)$.

Thus, the mapping $T \mapsto \mathcal{F}(T)$ sends sound (that is, true in the standard model $\mathbb{N}$) theories T to classes of number-theoretic functions. $\mathcal{F}(T)$ is closed under

[2] In this paper we only deal with extensions of EA in the same language.

composition and, for T containing EA, also contains all elementary functions $\mathcal{E}$. Also notice that the class $\mathcal{F}(T)$ only depends on the set of Π_2-consequences of T. Hence, if T is Π_2-conservative over U, then $\mathcal{F}(T) \subseteq \mathcal{F}(U)$. We write $T \subseteq_n U$ to denote Π_{n+1}-conservativity of T over U. $T \equiv_n U$ means $T \subseteq_n U$ and $U \subseteq_n T$.

The mapping $\mathcal{F}$ on Π_2-axiomatizable theories can be characterized in the following way. Consider a true Π_2-sentence $\pi = \forall x \exists y \varphi(x, y)$ with $\varphi(x, y)$ a Δ_0-formula. Let $f_\pi(x) = \mu y.\ \varphi(x, y)$ denote the *witnessing function* for π. The following basic result almost immediately follows from Herbrand's theorem (cf. [1] for details).

Proposition 1 *$g \in \mathcal{F}(\mathsf{EA} + \pi)$ iff g can be obtained from elementary functions and f_π by composition.*

This induces a reducibility relation $\leq_c$ on the set of all functions:

$$g \leq_c f \iff g \text{ is obtained from } \mathcal{E} \cup \{f\} \text{ by composition.}$$

This reducibility can also be characterized in a more computational way as follows: $g \leq_c f$ iff g is computable in elementary time using only *boundedly many* queries to the oracle f. Boundedness means that the number of queries does not depend on the input $\vec{x}$ of the program for $g(\vec{x})$. This makes $\leq_c$ rather similar to the so-called bounded truth-table reducibility in recursion theory. For the uses of $\leq_c$ in fragments of arithmetic see [2, 4].

The transitive relation $\leq_c$ induces a degree structure on the set of functions: $f \sim_c g$ iff $f \leq_c g$ and $g \leq_c f$. Proposition 1 means that $\mathcal{F}(\mathsf{EA} + \pi)$ is precisely (the union of) the set of degrees below $\deg(f_\pi)$. f_π is a function with a Δ_0-graph and it is clear that any function with an elementary (Δ_0) graph has the form f_π, for some $\pi \in \Pi_2$, so we can restrict our attention to Δ_0-degrees.

Let us call a degree Δ_0 if it is generated by a function with a Δ_0-graph, and let $\mathcal{D}$ be the set of all Δ_0-degrees with the inherited ordering $\leq_c$. $\mathcal{D}$ forms a lattice, see L. Kristiansen [11] for a recursion-theoretic study of $\mathcal{D}$. It is known that on *increasing* functions with Δ_0-graph $\leq_c$ coincides with "elementary in" reducibility, however this fails without the condition of monotonicity.

2.2 Lindenbaum algebras

Let $\mathcal{L}_T$ denote the Lindenbaum boolean algebra of T, that is, the set of of all sentences modulo provable equivalence in T. It provides an algebraic view of proof-theoretic objects: *schemata* over T correspond to subsets of $\mathcal{L}_T$; extralogical *inference rules* correspond to *operators* acting on $\mathcal{L}_T$; deductively closed sets of formulas, usually called *extensions of* T, correspond to *filters* of $\mathcal{L}_T$, that is, subsets of $\mathcal{L}_T$ upward closed w.r.t. $\leq$ and closed under inf. (*Ideals* are defined dually.) If U is an extension of T, then $\mathcal{L}_U$ can be identified with the corresponding factoralgebra of $\mathcal{L}_T$.

The mapping $\mathcal{F}$ induces an order-inversing function

$$\hat{\mathcal{F}} : \{\text{true } \Pi_2\text{-sentences}\} \to \mathcal{D},$$

which is correctly defined on elements of $\mathcal{L}_{\mathsf{EA}}$ (by Proposition 1):

$$\hat{\mathcal{F}} : \pi \longmapsto \deg(f_\pi).$$

It establishes a good correspondence between recursion-theoretical and arithmetical objects in complexity Π_2: theories correspond to ideals in $\mathcal{D}$, inference rules correspond to *subrecursive operators* acting on $\mathcal{D}$.

In general, formulas of higher arithmetical complexity need not correspond to single elements of $\mathcal{D}$. The situation is described in terms of lattices of ideals. Let Ideals($\mathcal{D}$) denote the lattice of ideals of $\mathcal{D}$. $\mathcal{D}$ is canonically embeddable into Ideals($\mathcal{D}$), identified with the set of its own *principal ideals*. Principal ideal generated by a degree $\mathbf{a}$, that is, the set of all degrees below $\mathbf{a}$, will be denoted $\iota(\mathbf{a})$. Let TrueFilters($\mathcal{L}_{\mathsf{EA}}$) denote the set of all subfilters of the filter of true sentences in $\mathcal{L}_{\mathsf{EA}}$. Then $\hat{\mathcal{F}}$ naturally lifts to a mapping

$$\hat{\mathcal{F}} : \text{TrueFilters}(\mathcal{L}_{\mathsf{EA}}) \to \text{Ideals}(D),$$
$$T \longmapsto \mathcal{F}(T)/{\sim_c},$$

for any filter T.

2.3 Jumps

There are several natural subrecursive operators acting on $\mathcal{D}$. The most important one is the jump operator $'$, which sends

$$\mathbf{a} \in \mathcal{D} \quad \longmapsto \quad \mathbf{a}' = \deg(\text{universal function for } \iota(\mathbf{a})).$$

Notice that $\iota\mathbf{a}$ is generated by a single function f over $\mathcal{E}$. One can write out a primitive recursion schema in f defining the evaluation function $\mathsf{eval}_f(e, x)$ for terms composed of f and elementary functions (cf. [1]). Then for any f with a Δ_0-graph, substituting its definition in eval_f, we can define:

$$\deg(f)' = \deg(\mathsf{eval}_f).$$

Let $\mathbf{0}$ be the degree of any elementary function. Then the principal ideals generated by $\mathbf{0}$, $\mathbf{0}'$, $\mathbf{0}''$, ... coincide with the classes $\mathcal{E} = \mathcal{E}_3$, $\mathcal{E}_4$, $\mathcal{E}_5$, ... of the Grzegorczyk hierarchy. This follows from the old results of Kleene or, alternatively, can be considered as a definition of the Grzegorczyk hierarchy.

The arithmetical counterpart of $'$ acts on the whole of $\mathcal{L}_{\mathsf{EA}}$ and is known under the name *1-consistency* or *uniform* Π_2*-reflection* principle. Let n-$\mathsf{Con}(T)$ denote a natural formula expressing that the theory $T + Th_{\Pi_n}(\mathbb{N})$ is consistent. Notice that for elementarily presented theories T the formula n-$\mathsf{Con}(T)$ is Π_{n+1} and can be formulated using a truthdefinition for Π_n-sentences in EA. We assume that 0-$\mathsf{Con}(T) = \mathsf{Con}(T)$ is the usual Gödel's consistency assertion for T.

The *n-consistency operator* $\langle n \rangle_T : \mathcal{L}_T \to \mathcal{L}_T$ is defined by:

$$\deg(\varphi) \quad \longmapsto \quad \langle n \rangle_T \varphi = \deg(n\text{-}\mathsf{Con}(T + \varphi)),$$

where φ is any sentence. The following proposition shows that $\langle 1 \rangle_{\mathsf{EA}}$ precisely corresponds to $'$ under the mapping $\hat{\mathcal{F}}$.

Proposition 2 *For true* $\pi \in \Pi_2$, $\hat{\mathcal{F}}(\langle 1\rangle_{\mathsf{EA}}\pi)) = \hat{\mathcal{F}}(\pi)'$ in $\mathcal{L}_{\mathsf{EA}}$.

A proof of this proposition rests upon two separate observations. The first one is quite general. If T is an *elementarily presented* theory, that is, if it comes together with a Δ_0-definable proof predicate, then this induces a natural indexing on the set of its provably total computable functions. We call a pair $\langle e, p\rangle$ a *T-index* of a function $f(x)$ iff e codes a (Turing machine) program for f and p codes a T-proof of the totality of e, that is, of the Π_2-formula $\forall x \exists y\, \varphi_e(x) = y$, where φ_e is the function computed by a program e. (φ_e is usually formalized using Kleene's T-predicate.) Let $\psi_T(q, x)$ be the universal function associated with this indexing of $\mathcal{F}(T)$, that is, $\psi_T(q, x)$ returns y if $q = \langle e, p\rangle$ is a T-index and $\varphi_e(x) = y$. Then the statement of totality of ψ_T is EA-equivalent to 1-$\mathsf{Con}(T)$. In fact, it is almost literally the same formula. We conclude: $\mathcal{F}(\mathsf{EA} + 1\text{-}\mathsf{Con}(T))$ is the set of functions c-reducible to ψ_T.

Our second observation is that the T-indexing is equivalent to the Gödel numbering of terms composed from f and elementary functions provided T is finite and Π_2-axiomatized. This follows from Proposition 1, which can be formalized in EA or EA^+, respectively, depending on the choice of a cut-free Gentzen-type or a standard Hilbert-type proof system for T. Therefore the totality of ψ_T is EA^+-provably equivalent to the totality of eval_{f_T}, where f_T is the witnessing function for the axiom of T. Yet, any of these two statements already implies the totality of superexponentiation function (over EA). Hence, their equivalence is formalizable in EA irrespectively of the choice of the proof-system and the result follows by Proposition 1, q.e.d.

2.4 Provability algebras and logics

The structure of Lindenbaum boolean algebra $\mathcal{L}_T$ enriched by the operator $\langle n\rangle_T$ is called *the n-provability algebra of T* and denoted $\mathcal{M}^n_T$. We omit the subscript $_T$ whenever the underlying algebra is specified by the context. The dual operators $[n]\varphi = \neg\langle n\rangle\neg\varphi$ are called *n-provability* operators. $\langle 0\rangle$ is usually denoted $\Diamond$, $[0]$ is $\Box$. R. Magari [12] was the first to consider the structure $\mathcal{M}^0_T = (\mathcal{L}_T, \Box_T)$, which is now known as *Magari algebra of T*.

Notice that *terms* in the language of n-provability algebras can be identified with *propositional modal formulas*, that is, the formulas built up from propositional variables and $\top$ (truth) by boolean connectives and $\langle n\rangle$. *Identities* of $\mathcal{M}^n_T$ constitute the *n-provability logic* of T, that is, the set of modal formulas provable in T under every substitution of arithmetical sentences for propositional variables and the translation of $\langle n\rangle$ as $\langle n\rangle_T$. Solovay's theorem [17] characterizes the 0-provability logic of any sound theory T. Later G. Boolos and C. Smoryński realized that n-provability algebras share the same provability logic.[3]

Axioms: (i) Boolean tautologies;

(ii) $\Diamond(\varphi \vee \psi) \to (\Diamond\varphi \vee \Diamond\psi)$; $\neg\Diamond\neg\top$;

[3]Strictly speaking, they considered different n-provability operators, but the difference is not that essential here.

(iii) $\Diamond\varphi \to \Diamond(\varphi \wedge \neg\Diamond\varphi)$.

Rules: modus ponens, $\varphi \to \psi \vdash \Diamond\varphi \to \Diamond\psi$.

This logic is usually named **GL** after Gödel and Löb,[4] and by Solovay's theorem we have:

$$\mathbf{GL} \vdash \varphi(\vec{x}) \iff \mathcal{M}_T^0 \vDash \forall\vec{x}\,(\varphi(\vec{x}) = \top),$$

for any modal formula/term $\varphi(\vec{x})$. Notice that Löb's axiom (iii) generalizes Gödel's second incompleteness theorem.

GL has been thoroughly investigated: it is decidable, satisfies Craig's interpolation property, is sound and complete for the class of all transitive and conversely well-founded Kripke frames. It enjoys finite model property and a reasonable sequent-style proof system for **GL** is also known. See G. Boolos [5] for additional information on **GL** and V. Shavrukov [16] on Magari algebras.

2.5 Graded provability algebras

Mixing together different n-provability operators proves to be more interesting, but also more difficult to study. The structure $\mathcal{M}_T = (\mathcal{L}_T, \langle 0\rangle_T, \langle 1\rangle_T, \dots)$ is called the *graded provability algebra of* T.

Terms of the graded provability algebra correspond to propositional *polymodal formulas*. The identities of $\mathcal{M}_T$ constitute the system **GLP** (cf. [5]):

Axioms: (i) Axioms of **GL** for each operator $\langle n\rangle$;

(ii) $\langle n\rangle\varphi \to \langle m\rangle\varphi$, for $m \leq n$;

(iii) $\langle m\rangle\varphi \to [n]\langle m\rangle\varphi$, for $m < n$.

Rules: modus ponens, $\varphi \to \psi \vdash \langle n\rangle\varphi \to \langle n\rangle\psi$.

This system was first shown to be sound and complete by G. Japaridze in [6] (using a somewhat different notion of n-provability). We have:

$$\mathbf{GLP} \vdash \varphi(\vec{x}) \iff \mathcal{M}_T \vDash \forall\vec{x}\,(\varphi(\vec{x}) = \top),$$

for any sound theory T. Later K. Ignatiev [9] simplified his work and thoroughly investigated the modal logical properties of **GLP**. By now we know that **GLP** is sufficiently well-behaved: it is decidable, satisfies Craig's interpolation property. Unfortunately, it is not complete for any class of Kripke frames. Yet, it has a simple translation into a system **LN** obtained from **GLP** by replacing axioms (ii) by the weaker principle

$$[m]\varphi \to [n][m]\varphi, \text{ for } m \leq n.$$

LN is already sound and complete for a nice class of (finite) Kripke frames.

[4] **GL** is usually formulated in terms of the dual modality $\Box$. In our treatment the choice of $\Diamond$ seems to be more natural. Curiously, as communicated by A. Visser, if $\Diamond$ is basic and $\Box$ is treated as an abbreviation, the usual axioms of **GL** formulated in terms of $\Box$ do not suffice.

2.6 Ordinal characteristic

A graded provability algebra $\mathcal{M}_T$ and, in fact, any **GLP**-algebra $\mathcal{M}$, can be associated in a natural way a certain ordinal *characteristic*, $\mathrm{char}(\mathcal{M})$. The term 'characteristic' hints at the analogy with the notions like characteristic of a field.

$\mathrm{char}(\mathcal{M})$ is defined as follows. First, consider the *prime subalgebra* $\mathcal{P} \subset \mathcal{M}$, that is, the set of all elements of $\mathcal{M}$ generated from $\top$ by all functions of $\mathcal{M}$: boolean operations and $\langle n \rangle$, for all n. Elements of $\mathcal{P}$ can be identified with *letterless* formulas in the language of **GLP** (modulo **GLP**-provable equivalence). Let $\mathcal{P}^0$ be the set of all *consistent* elements of $\mathcal{P}$, that is, we throw out the element $\bot$ from $\mathcal{P}$. For any formulas $\varphi, \psi \in \mathcal{P}^0$ we define:

$$\varphi <_0 \psi \iff \mathbf{GLP} \vdash \psi \to \Diamond\varphi.$$

Clearly, $<_0$ is an irreflexive transitive relation (by Gödel's theorem). Moreover, K. Ignatiev [9] proved that $<_0$ is well-founded and the ordinal of $\mathcal{P}^0$ is ϵ_0. Thus, we may call the characteristic of $\mathcal{M}$ the ordinal of the ordering $<_0$ on $\mathcal{P}^0$. In this paper we analyse $<_0$ in some detail in order to establish direct links with proof-theoretic ordinal notation systems. In fact, only a certain part of $\mathcal{P}^0$ will be needed for the analysis of PA.

2.7 Reduction property

As mentioned above, already since Ignatiev's work it has been known that the logic **GLP** has a certain characteristic 'ordinal' equal to ϵ_0. This ordinal emerged from the normal form of letterless formulas of **GLP**. However, at that time the specialists on provability logic, including the author of the present paper, completely missed the relationship with the traditional proof-theoretic ordinals. In hindsight I see two factors that may have blurred our understanding of this relationship: on the one hand, the predominant interpretation of n-provability was the n-fold application of ω-rule above PA, which was clearly out of the suitable range of proof-theoretic strength. On the other hand, there was a missing link provided by the modern theory of iterated reflection principles (see [15, 3]). The next proposition proved in [4], Theorem 2, presents a key property of graded provability algebra of T which, however, cannot be stated just in terms of its identities.

Notice that $\mathcal{M}_T$ bears some additional structure, namely a family of distinguished subsets $P_0 \subset P_1 \subset \ldots \subseteq \mathcal{M}_T$, which correspond to Π_1, Π_2, ... (degrees of) sentences. Obviously, $\bigcup_{i \geq 0} P_i = \mathcal{M}$. We call this family the natural *stratification of* $\mathcal{M}$; the algebra $\mathcal{M}_T$ taken together with its natural stratification will also be called *graded provability algebra of* T. (The presence of the stratification is not a big deal, since the sets P_i are 'almost expressible' in the language of the algebra. Namely, by the so-called *Goldfarb's principle*, an element φ below $\langle n \rangle \top$ belongs to P_n iff $\varphi = \langle n \rangle \psi$ for some ψ.)

Stratification allows us to express the notion of Π_{n+1}-conservative extension of theories. Let U and V be filters in $\mathcal{M}$. We write $U \subseteq_n V$ iff $U \cap P_n \subseteq V$. $U \equiv_n V$ means $U \subseteq_n V$ and $V \subseteq_n U$. The same notation is also applied to

arbitrary *sets* of elements of $\mathcal{M}$ and means the corresponding relation between *filters* generated by those sets.

Proposition 3 *Assume T is a Π_{n+2}-axiomatized theory containing EA. Then for all $\varphi \in \mathcal{M}_T$, the following holds in $\mathcal{M}_T$:*

$$\{\langle n+1\rangle_T \varphi\} \equiv_n \{Q^n_k(\varphi) : k < \omega\},$$

where

$$\begin{aligned} Q^n_0(\varphi) &= \top, \\ Q^n_{k+1}(\varphi) &= \langle n\rangle_T (Q^n_k(\varphi) \wedge \varphi). \end{aligned}$$

Proof. Let T and φ be given. In the formulation of Theorem 2 in [3] take $U = T$ and $T = T + \varphi$. Then the set of Π_{n+1}-consequences of $T + \langle n+1\rangle_T \varphi$ can be described as the closure of T under the rule

$$\frac{\psi}{\langle n\rangle_T (\varphi \wedge \psi)}, \text{ where } \psi \in \Pi_{n+1},$$

which generates the elements $Q^n_k(\varphi)$. The theory $T + \{Q^n_k(\varphi) : k < \omega\}$ is also closed under this rule by monotonicity of $\langle n\rangle_T$, q.e.d.

Example: the formulas $Q^n_k(\top)$ are equivalent to $\langle n\rangle \ldots \langle n\rangle \top$ (k times). Thus, one corollary of the above proposition is that, under appropriate conditions, 1-consistency of T is Π_1-conservative over its ω times iterated consistency assertion. For background information and the uses of such results in fragments of PA see [3].

By Proposition 3, the filter generated by all Π_{n+1}-consequences of an element $\langle n+1\rangle_T \varphi \in \mathcal{M}_T$ of complexity Π_{n+2} can be generated by specific Π_{n+1}-elements $Q^n_k(\varphi)$. A remarkable property of this relationship, which is made an essential use of below, is that these elements are expressible in the language of graded provability algebra $\mathcal{M}_T$. Not all the algebras satisfying **GLP** have this property: e.g., if we throw away the operation $\langle 1\rangle$ from the structure, the logic and the characteristic of the algebra remain the same, but the Π_1-consequences of $\langle 2\rangle\top$ cannot be expressed in terms of $\langle 0\rangle$ alone. Thus, Proposition 3 expresses a specific kind of definitional completeness of $\mathcal{M}_T$. It can also be viewed as a reduction of $\langle n+1\rangle_T \varphi$ to formulas of lower arithmetical complexity, therefore we call this property of $\mathcal{M}_T$ *reduction property*.

We note that a proof of Theorem 2 in [3] can be obtained rather directly by cut-elimination in predicate logic (a model-theoretic proof is also possible). Hence, Proposition 3 is formalizable in EA^+.

In the remaining part of the paper we present the details of how the notions involved can be used to provide a proof-theoretic analysis of Peano arithmetic and its main fragments.

3 A provability-logical view of ϵ_0

Here we read up an ordinal notation system from a graded provability algebra. This is done by a partial normal form result, which will be sufficient for the purposes of this paper. This section is completely within propositional logic.

Let S denote the set of all words in the alphabet $\mathbb{N} = \{0, 1, \dots\}$, including the empty word Λ. S_n will denote the set of words in the alphabet $\{n, n+1, \dots\}$. To each element $\alpha = n_1 n_2 \dots n_k$ of S we associate its *modal interpretation*, that is, the variable-free modal formula

$$\langle n_1 \rangle \langle n_2 \rangle \cdots \langle n_k \rangle \top, \tag{1}$$

and its arithmetical interpretation α_T^* in $\mathcal{L}_T$, whenever an elementary presented extension T of EA is fixed. We do not distinguish between the word α and formula (1). We also let $\Lambda = \top = \Lambda_T^*$.

Below we use $\vdash$ to denote provability in **GLP**. We write $\alpha \sim \beta$ if $\vdash \alpha \leftrightarrow \beta$. $\alpha = \beta$ means graphical identity.

For each n there is an ordering $<_n$ on S defined by

$$\alpha <_n \beta \iff \vdash \beta \to \langle n \rangle \alpha.$$

It is immediately seen that $<_n$ is transitive and irreflexive. We shall later see that it is well-founded of depth ϵ_0. We shall also mainly consider the restriction of the ordering $<_n$ on S_n, where it can be shown to coincide with $<_0$.

A useful characteristic of a word $\alpha \in S$ is its *width* $w(\alpha)$, that is, the number of different letters occurring in it. We shall often define functions on words by induction on width.

Some of the elements of S are pairwise equivalent, so we first define a subset $NF \subset S$ of *normal forms*. These are (as we shall see later) isomorphic to Cantor normal forms of ordinals $< \epsilon_0$.

- Λ and any word of width 1 belongs to NF.
- Assume $w(\alpha) > 1$ and let n be the smallest letter in α. Then graphically $\alpha = \alpha_1 n \cdots n \alpha_k$, where all α_i do not contain n and hence $w(\alpha_i) < w(\alpha)$ for $1 \le i \le k$. Then $\alpha \in NF$ iff all $\alpha_i \in NF$ and, for all $1 \le i < k$, $\alpha_{i+1} \not<_{n+1} \alpha_i$. (Note that $\alpha_i \in S_{n+1}$.)

The following simple lemma will be often used.

Lemma 4 *(i) If $m < n$, then $\vdash \langle n \rangle \varphi \wedge \langle m \rangle \psi \leftrightarrow \langle n \rangle (\varphi \wedge \langle m \rangle \psi)$;*

(ii) If $\alpha \in S_{n+1}$, then $\vdash \alpha \wedge n\beta \leftrightarrow \alpha n \beta$.

Proof. Statement (i) is obvious from the axioms of **GLP**. Statement (ii) follows by repeated application of (i), q.e.d.

Lemma 5 *Let $\alpha = \alpha_1 n \alpha_2 n \cdots n \alpha_k$, where all $\alpha_i \in S_{n+1}$. If $\alpha_1 >_{n+1} \alpha_2$, then*

$$\alpha \sim \alpha_1 n \alpha_3 n \cdots n \alpha_k.$$

Proof. Let $\beta = \alpha_3 n \cdots n\alpha_k$. Since $\alpha_1, \alpha_2 \in S_{n+1}$,

$$\begin{array}{rcl} \vdash \alpha = \alpha_1 n \alpha_2 n \beta & \leftrightarrow & \alpha_1 \wedge n\alpha_2 n\beta \\ & \leftrightarrow & \alpha_1 \wedge \langle n \rangle (\alpha_2 \wedge n\beta). \end{array}$$

In our logic we also have

$$\begin{array}{rcl} \vdash \langle n+1 \rangle \alpha_2 \wedge n\beta & \rightarrow & \langle n+1 \rangle (\alpha_2 \wedge n\beta) \\ & \rightarrow & \langle n \rangle (\alpha_2 \wedge n\beta). \end{array}$$

Hence, if $\vdash \alpha_1 \rightarrow \langle n+1 \rangle \alpha_2$, then

$$\begin{array}{rcl} \vdash \alpha_1 n\beta & \leftrightarrow & \alpha_1 \wedge n\beta \\ & \leftrightarrow & \alpha_1 \wedge \langle n \rangle (\alpha_2 \wedge n\beta), \end{array}$$

q.e.d.

Corollary 6 *Every word $\alpha \in S$ can be brought into an equivalent normal form, that is, there is an $\alpha' \in NF$ such that $\alpha' \sim \alpha$.*

Also notice that the length of α' obtained by the repeated application of the previous lemma does not exceed the length of α.

Lemma 7 *Any two normal forms $\alpha, \beta \in S_n$ are $<_n$-comparable, that is,*

$$\alpha <_n \beta \text{ or } \beta <_n \alpha \text{ or } \beta = \alpha. \qquad (*)$$

Proof. We reason by induction on the joint width of α and β and w.l.o.g. assume $n = 0$. For unary words the claim is obvious, so we consider the case that $w(\alpha\beta) > 1$.

As before, α and β can be written in the form

$$\alpha = \alpha_k 0 \alpha_{k-1} 0 \cdots 0 \alpha_1, \qquad \beta = \beta_m 0 \beta_{m-1} 0 \cdots 0 \beta_1,$$

where all α_i and β_j do not contain 0. By the induction hypothesis we obtain

$$\alpha_1 <_1 \beta_1 \text{ or } \beta_1 <_1 \alpha_1 \text{ or } \beta_1 = \alpha_1.$$

Claim. If $\alpha_1 <_1 \beta_1$, then $\alpha <_0 \beta$. Symmetrically, if $\alpha_1 >_1 \beta_1$, then $\alpha >_0 \beta$.

We only prove the first part. Let $\overline{\alpha}_i = \alpha_i 0 \cdots 0 \alpha_1$. We prove by induction on i that $\overline{\alpha}_i <_1 \beta_1$, for all $i \leq k$. It is obvious that $\beta_1 \leq_0 \beta$ and $<_1$ is stronger than $<_0$, so the Claim will follow.

Notice that $\vdash \alpha_{i+1} 0 \overline{\alpha}_i \leftrightarrow (\alpha_{i+1} \wedge 0\overline{\alpha}_i)$. On the other hand, $\beta_1 >_1 \alpha_{i+1}$ by transitivity of $<_1$ and because $\alpha \in NF$. By the induction hypothesis we have $\beta_1 >_1 \overline{\alpha}_i$, hence

$$\begin{array}{rcl} \vdash \beta_1 & \rightarrow & \langle 1 \rangle \alpha_{i+1} \wedge \langle 0 \rangle \overline{\alpha}_i \\ & \rightarrow & \langle 1 \rangle (\alpha_{i+1} \wedge 0\overline{\alpha}_i) \\ & \rightarrow & \langle 1 \rangle \alpha_{i+1} 0 \overline{\alpha}_i, \end{array}$$

which proves the induction step.

Continuing the proof of Lemma 7 from the Claim we can conclude that the disjunction $(*)$ can only be false if $\alpha_1 = \beta_1$. In this case we have to compare α_2 and β_2 using the induction hypothesis again. Assume w.l.o.g. that $\alpha_2 <_1 \beta_2$. Then we have $\alpha_2 0 \alpha_1 <_1 \beta_2 0 \beta_1$, because

$$\vdash \beta_2 \wedge 0\alpha_1 \to \langle 1 \rangle (\alpha_2 \wedge 0\alpha_1).$$

Following the proof of the Claim we then obtain $\overline{\alpha}_i <_1 \beta_2 0 \beta_1 \leq_0 \beta$, for all $i > 1$. It follows that in this case $\alpha <_0 \beta$. Using the symmetry, the only remaining case is that both $\alpha_1 = \beta_1$ and $\alpha_2 = \beta_2$, and the reasoning can be continued. If $\alpha \neq \beta$, at the end we come to the situation when one of the two words, say α, is a proper end segment of the other. Then obviously $\beta >_0 \alpha$, q.e.d.

Notice that the above proof provides an effective comparison algorithm for words in *NF*. Hence we can also effectively check whether a given word belongs to *NF*.

Since $<_0$ is irreflexive we obtain

Corollary 8 *The normal form of a given word is graphically unique.*

Corollary 9 *Any words $\alpha, \beta \in S_n$ satisfy the trichotomy:*

$$\alpha <_n \beta \text{ or } \beta <_n \alpha \text{ or } \beta \sim \alpha.$$

This follows from the fact that the orderings $<_n$ respect the logical equivalence relation $\sim$. Thus, we also obtain

Corollary 10 *For any $\alpha, \beta \in S_n$, either $\vdash n\alpha \to n\beta$ or $\vdash n\beta \to n\alpha$ and this can be effectively decided.*

Lemma 11 *For all $\alpha, \beta \in S_n$ there is an (effectively constructible) $\gamma \in S_n$ such that $\vdash \gamma \leftrightarrow (\alpha \wedge \beta)$.*

Proof. We reason by induction on the width of $\alpha\beta$. We can write α and β in the form $\alpha = \alpha_1 n \alpha'$ and $\beta = \beta_1 n \beta'$ with $\alpha_1, \beta_1 \in S_{n+1}$. We then have

$$\vdash \alpha \wedge \beta \leftrightarrow \alpha_1 \wedge n\alpha' \wedge \beta_1 \wedge n\beta'.$$

From the previous corollary we know that either $\vdash n\alpha' \to n\beta'$ or $\vdash n\beta' \to n\alpha'$. Assume $n\alpha'$ is stronger. By the induction hypothesis we can find a $\gamma_1 \in S_{n+1}$ such that $\vdash \gamma_1 \leftrightarrow (\alpha_1 \wedge \beta_1)$, therefore

$$\begin{aligned} \vdash \alpha \wedge \beta &\leftrightarrow \alpha_1 \wedge \beta_1 \wedge n\alpha' \\ &\leftrightarrow \gamma_1 \wedge n\alpha' \\ &\leftrightarrow \gamma_1 n \alpha', \end{aligned}$$

which has the required form, q.e.d.

Now we are going to establish the correspondence between notations and ordinals. For each n define a function $o_n : S_n \to \epsilon_0$ by recursion on width and a subsidiary recursion on $\min(\alpha) - n$, where $\min(\alpha)$ denotes the minimal letter in α:

- If $\alpha = \underbrace{nn\ldots n}_{k \text{ times}}$, then $o_n(\alpha) = k$ $(k \geq 0)$.
- If $\alpha = \alpha_1 n \cdots n \alpha_k$ with all $\alpha_i \in S_{n+1}$ and not all of them empty, then $o_n(\alpha) = \omega^{o_{n+1}(\alpha_k)} + \cdots + \omega^{o_{n+1}(\alpha_1)}$.

Notice that $o_{n+1}(\alpha_i)$ is defined, because $w(\alpha_i) < w(\alpha)$ in case $k > 1$. If $k = 1$, then $\alpha_1 = \alpha \neq \Lambda$, hence $\min(\alpha_1) - (n+1) < \min(\alpha) - n$.

We also let $o(\alpha) = o_0(\alpha)$. Examples: $o(101) = o_1(212) = \omega + \omega$; $o(2101) = \omega + \omega^{o_1(21)} = \omega + \omega^{\omega^{o_2(\Lambda)+o_2(2)}} = \omega + \omega^{\omega^{0+1}} = \omega^\omega$.

Define $x \uparrow \alpha$ to be the result of replacing in α every letter n by $x+n$, where $x \in \mathbb{N}$. It is easy to see that $o(\alpha) = o_n(n \uparrow \alpha)$, for all n.

The following properties are straightforward from what we already know.

Lemma 12 *(i)* $\forall \alpha, \beta \in S_n (\alpha <_n \beta \Rightarrow o_n(\alpha) < o_n(\beta))$;

(ii) $\forall \alpha \in S_n \; \forall z < o_n(\alpha) \; \exists \beta <_n \alpha \; o_n(\beta) = z$;

(iii) $o_n : S_n \to \epsilon_0$ *is surjective.*

Proof. Notice that the algorithm of bringing a word $\alpha \in S_n$ into a normal form preserves the ordinal $o_n(\alpha)$. By induction on width it is then easy to see that for $\alpha, \beta \in NF$ the (term for the) ordinal $o_n(\alpha)$ is represented in Cantor normal form. The rules of comparison of α with β are the same as those for Cantor normal forms, which yields claim (i).

It is also clear that for any ordinal z in Cantor normal form we can find an $\alpha \in S_n \cap NF$ such that $o_n(\alpha) = z$. This yields claims (ii) and (iii), q.e.d.

Thus, we can consider the set of normal forms as a 1–1 ordinal notation system for ϵ_0. Properties (i) and (ii) together mean that o_n is a *p-morphism* from the structure $(S_n, <_n)$ to ϵ_0.

4 Proof-theoretic analysis

The reduction property allows to assign fundamental sequences to our ordinal notations. For $\alpha \in NF$ and any $k < \omega$ we define the elements $\alpha[k] \in NF$ as follows:

- If $\alpha = 0\beta$ then $\alpha[k] = \beta$.
- If $\alpha = \langle n+1 \rangle \beta$ then $\alpha[k]$ is the (uniquely defined) $\gamma \in NF$ such that $\gamma \sim Q^n_k(\beta)$.

Notice that such a γ exists by Lemma 11, because $Q^n_k(\beta)$ are built up from β only using $\langle n \rangle$ and $\wedge$. The reduction property now reads:

$$\{\alpha\} \equiv_n \{\alpha[k] : k < \omega\}, \qquad (**)$$

if $\alpha = \langle n+1 \rangle \beta$. Recall that this property holds in any graded provability algebra $\mathcal{M}_T$, where T is sound and Π_{n+2}-axiomatizable.

Lemma 13 *If $\alpha = \langle n+1\rangle\beta$ then:*

(i) $\alpha[k] <_0 \alpha$;

(ii) $m < k \to \alpha[m] <_0 \alpha[k]$;

(iii) $\forall\varphi <_0 \alpha \exists k\, \varphi <_0 \alpha[k]$.

Proof. Properties (i) and (ii) follow at once from the definition of the formulas $Q^n_k(\beta)$. To prove Part (iii) consider the graded provability algebra $\mathcal{M}_T$ for any T satisfying the conditions of Proposition 3. If $\vdash \alpha \to \Diamond\varphi$, then $\mathcal{M}_T \vDash \alpha^*_T \leq \Diamond_T\varphi^*_T$. Since $\Diamond_T\varphi^*_T$ is a Π_1-formula and $\mathcal{M}_T \vDash \{\alpha^*_T\} \equiv_0 \{\alpha[k]^*_T : k < \omega\}$ we conclude that $\mathcal{M}_T \vDash \alpha[k]^*_T \leq \Diamond_T\varphi^*_T$, for some k. Therefore we have $\varphi <_0 \alpha[k]$, as required. Thus, we used fathfulness of the arithmetical interpretation and the reduction property to prove a purely propositional fact (iii), q.e.d.

A good way of depicting the ordering $<_0$ is by means of a *reduction tree*: starting with a word $\alpha \in S$ we generate its immediate successors $\alpha[k]$, for all k. Thus, nonempty α has one successor β if $\alpha = 0\beta$, and ω successors, otherwise. Empty words are the leaves of the tree. Since $<_0$ is well-founded, every branch of the tree terminates. It is also clear that the height of the tree generated by α is precisely $o(\alpha)$: use Corollary 9 for the successor case and Part (iii) of Lemma 13 for the limit case.

As a corollary of the reduction property we also obtain

Lemma 14 *If T is a sound and Π_{n+2}-axiomatized extension of EA. Then*

$$\mathsf{EA}^+ \vdash \forall\alpha \in NF\,(\Diamond_T\alpha^*_T \leftrightarrow \forall k \Diamond_T \alpha[k]^*_T).$$

Proof. This follows from $(**)$ and (ii) of Lemma 13, both formalizable in EA^+, q.e.d.

Now we are ready to present Gentzen's result on the consistency strength of Peano arithmetic. Below we write $<$ instead of $<_0$. Let $[T, IR(\Pi_1, NF)]$ denote the theory

$$T + \{\forall\alpha \in NF\, \varphi(\alpha) : T \vdash \forall\alpha \in NF\, (\forall\beta < \alpha\, \varphi(\beta) \to \varphi(\alpha)),\ \varphi \in \Pi_1\}.$$

The following theorem shows that $\mathsf{Con(PA)}$ can be proved over PRA by a single application of Π_1 transfinite induction rule up to ϵ_0. (The latter can also be easily reduced to an application of the induction rule for a Δ_0 formula.) It also shows that $\mathsf{Con(PA)}$ is the strongest formula with this property.

Theorem 1 $[\mathsf{PRA}, IR(\Pi_1, NF)] \equiv \mathsf{PRA} + \mathsf{Con(PA)}$.

Proof. For the inclusion $(\supseteq)$ we choose $T = \mathsf{EA}^+$. Recall that $\mathsf{PA} \equiv T + \{\langle n\rangle_T\top : n < \omega\}$. The one-letter words $n \in NF$ are (provably) cofinal in NF, therefore by formalizing the above equivalence

$$\begin{aligned} \mathsf{EA} \vdash \mathsf{Con(PA)} &\leftrightarrow \forall n \Diamond_T \langle n\rangle_T\top \\ &\leftrightarrow \forall\alpha \in NF\, \Diamond_T\alpha^*_T. \end{aligned}$$

Now we notice that by Lemma 14

$$\begin{array}{lll} T \vdash \forall \beta < \alpha \, \Diamond_T \beta_T^* & \rightarrow & \forall k \Diamond_T \alpha[k]_T^* \\ & \rightarrow & \Diamond_T \alpha_T^*. \end{array}$$

Therefore, the induction rule is applicable and we obtain

$$[T, IR(\Pi_1, NF)] \vdash \forall \alpha \in NF \, \Diamond_T \alpha_T^* \vdash \mathsf{Con}(\mathsf{PA}).$$

For $(\subseteq)$ assume that $\varphi \in \Pi_1$ and

$$\mathsf{PRA} \vdash \forall \alpha \in NF \, (\forall \beta < \alpha \, \varphi(\beta) \rightarrow \varphi(\alpha)).$$

Then the same formula is provable in a certain finite fragment $T \subseteq \mathsf{PRA}$. We have

$$T \vdash \Diamond_T \alpha_T^* \rightarrow \forall \beta \, (\Box_T (\alpha_T^* \rightarrow \varphi(\dot{\beta})) \rightarrow \varphi(\beta)),$$

because $\varphi \in \Pi_1$. Hence, using our assumption

$$\begin{array}{lll} T \vdash \forall \beta < \alpha \, \Box_T (\Diamond_T \beta_T^* \rightarrow \varphi(\dot{\beta})) & \rightarrow & (\Diamond_T \alpha_T^* \rightarrow \forall \beta < \alpha \, \Box_T (\alpha_T^* \rightarrow \varphi(\dot{\beta}))) \\ & \rightarrow & (\Diamond_T \alpha_T^* \rightarrow \forall \beta < \alpha \, \varphi(\beta)) \\ & \rightarrow & (\Diamond_T \alpha_T^* \rightarrow \varphi(\alpha)). \end{array}$$

Applying now Löb's theorem (in the form of reflexive induction) in T we obtain

$$T \vdash \forall \alpha \in NF \, (\Diamond_T \alpha_T^* \rightarrow \varphi(\alpha)),$$

which implies

$$T + \forall \alpha \in NF \, \Diamond_T \alpha_T^* \vdash \forall \alpha \in NF \, \varphi(\alpha).$$

The premise is equivalent to $\mathsf{Con}(\mathsf{PA})$, so we have

$$\mathsf{PRA} + \mathsf{Con}(\mathsf{PA}) \vdash \forall \alpha \in NF \, \varphi(\alpha),$$

as required, q.e.d.

Replacing consistency by n-consistency in the above agrument yields

Theorem 2 $[\mathsf{PRA}, IR(\Pi_{n+1}, NF)] \equiv \mathsf{PRA} + n\text{-}\mathsf{Con}(\mathsf{PA})$.

The result also extends to iterated applications of the rule, so we obtain the following statement.

Theorem 3 *The closure of* PRA *under* $IR(\Pi_{n+1}, NF)$ *is equivalent to*

$$\mathsf{PRA} + n\text{-}\mathsf{Con}(\mathsf{PA}) + n\text{-}\mathsf{Con}(\mathsf{PA} + n\text{-}\mathsf{Con}(\mathsf{PA})) + \ldots$$

Proof. This can be proved by induction on the number of iterated applications of the rule. Basis of induction is the content of the previous theorem. To treat the induction step let U_k denote $[\ldots[\mathsf{PRA}, IR(\Pi_{n+1}, NF)], \ldots, IR(\Pi_{n+1}, NF)]$ (k times). By the induction hypothesis U_k is equivalent to k times iterated consistency assertion for PA (over PRA). Taking $T = U_k$ for the base theory, we only need to prove that

$$[T, IR(\Pi_{n+1}, NF)] \equiv T + n\text{-}\mathsf{Con}(T + \mathsf{PA}).$$

We work in the graded provability algebra of T and observe that

$$T + \mathsf{PA} \equiv T + \{\langle m\rangle_T \top : n \leq m < \omega\}, \qquad (***)$$

because T is a Π_{n+1}-axiomatized theory. Formalizing $(***)$ as before we obtain that $n\text{-}\mathsf{Con}(T + \mathsf{PA})$ is equivalent to $\forall \alpha \in NF \cap S_n\ \langle n\rangle_T \alpha^*_T$. The latter formula follows by one application of the transfinite induction rule over T for the ordering $\leq_n$ restricted to $S_n \cap NF$, but the latter is provably isomorphic to the ordering $\leq_0$ on NF.

The argument for the converse inclusion in the proof of Theorem 1 also remains essentially unchanged, q.e.d.

From this result we may conclude that the the closure of PRA under Π_n transfinite induction rule up to ϵ_0 is a Π_{n+1}-regular theory (in the sense of [3]) and its Π_{n+1}-ordinal equals $\epsilon_0 \cdot \omega$.

We also obtain similar analysis for fragments $I\Sigma_m$ of PA (we take EA^+ as a base theory, because PRA is too close to $I\Sigma_1$).

Theorem 4 $[EA^+, IR(\Pi_{n+1}, \omega_{m+1})] \equiv EA^+ + n\text{-}Con(I\Sigma_m)$.

Here we use the segment of NF up to the ordinal notation $m+1$ corresponding to $I\Sigma_m \equiv \mathsf{EA}^+ + \langle m+1\rangle_{\mathsf{EA}^+}\top$. ω_{m+1} is just a more suggestive way of writing the word $m + 1 \in NF$.

As a corollary we infer Gentzen's proof in PA of transfinite induction up to any ordinal below ϵ_0.

Corollary 15 *For any m, PA proves full transfinite induction schema for the initial segment of NF up to ω_m.*

Proof. Let $\varphi \in \Pi_n$ be an induction formula. The formula

$$\forall \gamma < \omega_m\ (\forall \beta < \gamma\ \varphi(\beta) \to \varphi(\gamma)) \to \varphi(\alpha)$$

has complexity Π_{n+1} and applying the induction rule to it yields the instance of the induction schema for φ. Yet, from the previous theorem we know that $[\mathsf{PRA}, IR(\Pi_{n+1}, \omega_m)] \subseteq \mathsf{PA}$, q.e.d.

5 Transfinitely iterated consistency assertions

Here we establish the connection of our method with the results in [3], which allows for a refined ordinal analysis of theories in all complexities Π^0_n.

To the partial ordering $(S_n, <_n)$ and any elementary presented initial theory T we can naturally associate the progression of *iterated reflection principles* T^n_α, $\alpha \in S_n$, defined by the following equivalence (inside EA):

$$T^n_\alpha \equiv T + \{n\text{-}\mathsf{Con}(T^n_\beta) : \beta <_n \alpha,\ \alpha, \beta \in S_n\}.$$

It is known that the theories T^n_α are uniquely defined up to EA-provable equivalence [3]. (Notice, however, that unlike in [3] the theories T^n_α now use different orderings for different n and orderings $<_n$ are partial, but this is only a matter of technical convenience.) We also let T_α denote T^0_α.

The following main relationship is obtained almost for free.

Theorem 5 *Suppose T is an elementary presented Π_{n+1}-axiomatizable extension of EA^+. Provably in EA^+,*

$$\forall \alpha \in S_n,\ T + \alpha^*_T \equiv_n T^n_\alpha.$$

Proof. We use reflexive induction on α in T, that is, we prove

$$\mathsf{EA}^+ \vdash \forall \beta <_n \alpha\ \Box_{\mathsf{EA}^+}(T + \beta^*_T \equiv_n T^n_\beta) \to T + \alpha^*_T \equiv_n T^n_\alpha.$$

The claim then follows by Löb's theorem in EA^+. We prove the equivalence $T + \alpha^*_T \equiv_n T^n_\alpha$ by presenting an informal argument in EA^+.

For the inclusion $(\supseteq)$ notice that by the definition of $<_n$, if $\beta <_n \alpha$ then $T \vdash \alpha^*_T \to \langle n \rangle_T \beta^*_T$. By the reflexive induction hypothesis we have

$$\mathsf{EA}^+ \vdash \langle n \rangle_T \beta^*_T \to n\text{-}\mathsf{Con}(T^n_\beta).$$

It follows that for all $\beta <_n \alpha$,

$$T \vdash \alpha^*_T \to n\text{-}\mathsf{Con}(T^n_\beta),$$

hence $T + \alpha^*_T \vdash T^n_\alpha$.

For the inclusion $(\subseteq_n)$ we reason as follows. Assume $T + \alpha^* \vdash \pi$ with $\pi \in \Pi_n$. If $\alpha = n\beta$, then we have $T + \langle n \rangle_T \beta^*_T \vdash \pi$, which implies $T + n\text{-}\mathsf{Con}(T^n_\beta) \vdash \pi$ by the induction hypothesis. Since obviously $\beta <_n \alpha$ we conclude $T^n_\alpha \vdash \pi$.

If $\alpha = \langle m+1 \rangle \beta$ with $m \geq n$, then we have

$$\{\langle m+1 \rangle_T \beta^*_T\} \equiv_m \{Q^m_k(\beta^*_T) : k < \omega\}$$

by reduction property. As before, we can effectively find equivalent words $Q^m_k(\beta) \sim \gamma_k \in S_m$. Thus, by reduction property $T + (\gamma_k)^*_T \vdash \pi$, for some k. Since obviously $\gamma_k <_n \alpha$, this implies $T^n_{\gamma_k} \vdash \pi$ by the induction hypothesis, q.e.d.

Notice that the previous theorem does not depend on the property of the ordinal notation system being actually well-founded.

Now we are going from nonlinear orderings $<_n$ to linear ones. Let $\hat{o}(\alpha)$ denote the normal form of a word $\alpha \in S$. Notice that $\hat{o}$ is a p-morphism of S_n onto $S_n \cap NF$. The following lemma shows that progressions of iterated reflection principles connected by a p-morphism are equivalent.

Assume we are given two elementary orderings $(D_1, <_1)$ and $(D_2, <_2)$. Assume further that $h : D_1 \to D_2$ is an EA-provable p-morphism, that is, properties (i) and (ii) of Lemma 12 for h hold provably in EA. Then we have

Lemma 16 *Provably in EA,*

$$\forall \alpha \in D_1,\ T^n_\alpha \equiv T^n_{h(\alpha)}.$$

Here the progression on the left hand side is defined for the ordering $(D_1, <_1)$, and the one on the right hand side for $(D_2, <_2)$.

Proof. Both inclusions are straightforward by reflexive induction in EA using (i) and (ii), q.e.d.

As a corollary we obtain that T^n_α is equivalent to a linear progression of order type ϵ_0 defined just on the set of normal words (or, equivalently, on the set of Cantor normal forms of ordinals).

As a direct corollary of Theorem 5 we obtain Π^0_{n+1} proof-theoretic analysis of PA in the sense of [3]. Notice that this simplifies the treatment in [3].

Theorem 6 $\mathsf{PA} \equiv_n \bigcup_{\alpha \in NF} (\mathsf{EA}^+)^n_\alpha$.

For $n = 1$ together with Corollary 3.3 of [3] this yields a characterization of provably total recursive functions of PA. Let $\mathcal{F}_\alpha$ denote the α-th class of the Kleene hierarchy, that is, the set of functions c-reducible to α-times iterated jump of $\mathbf{0}$ (see [3] for an alternative definition). These classes are correctly defined, once an elementary ordinal notation system is fixed.

Corollary 17 $\mathcal{F}(\mathsf{PA}) = \bigcup_{\alpha \in NF} \mathcal{F}_\alpha$.

6 Discussion

We have obtained a proof-theoretic analysis of Peano arithmetic. What has been achieved by using the language of provability algebras? What does it tell us about the problem of canonical ordinal notation systems?

First of all, we have seen that the structure of graded provability algebra for EA already bears the structure of ϵ_0. That is, the standard ordinal notation system (not just the ordinal itself) has been extracted from this algebra in a satisfyingly canonical way. Secondly, our analysis of PA was based on the identification of PA with a specific filter on that algebra. This embedding is canonical

by the general correspondence *theories — filters* between proof-theoretic and algebraic frameworks. The remaining question is: was our choice of this particular algebra canonical?

Basically, we have made the following choices: base theory (EA), its provability predicate, stratification, n-provability predicates. The choice of base theory and its provability predicate is not really important. We can consider them fixed once and for all. The more important choice of the stratification was more or less dictated by the arithmetical *language* with its natural quantifier complexity levels. Also notice that our interpretation of operators $\langle n \rangle$ happens to be the *weakest possible* for which the axioms of **GLP** hold.

What if we work in a more expressive language, say, in that of set-theory? Then we also have an obvious stratification by quantifier complexity (Lévy hierarchy). The axioms of **GLP** hold, the characteristic is, therefore, equal to ϵ_0. However, for this stratification we do not have the crucial reduction property (see Section 2.7). Alternatively, the arithmetical complexity levels are expressible in the set-theoretic language and the usual stratification as well exists, but then ZF (and actually already much weaker systems) considered as a filter will not be generated by expressible elements. Thus, in both cases the structure is expressively too poor.

This suggests that the minimal graded provability algebra considered in this paper should be extended by additional operations corresponding to yet stronger provability concepts. A natural source of such concepts, in agreement with the current trends in proof theory, are the set-theoretical reflection principles. We plan to introduce relevant algebraic structures in a subsequent paper.

Of course, at this stage it would be premature to estimate the potential usefulness of our approach for the important open problems of ordinal analysis of strong systems, such as full second order arithmetic or ZF. Yet, we believe that it helps to clarify the more general question, where the canonical ordinal notation systems come from and in what terms proof-theoretic ordinals of a formal system could or should be specified.

It is also worth noting that in the area of provability logic this approach draws attention away from completeness results a lá Solovay towards the development of letterless fragments and normal form results for polymodal provability logics. A goal here would be to find appropriate generalization(s) of the notion of graded provability algebra and to develop a definability theory for it that could serve as an abstract framework for ordinal analysis. Within this framework it would be also be interesting to find a general argument explaining why prime subalgebras have to be well-founded. Recall that Löb's axiom actually means well-foundedness in Kripke frames.

References

[1] L.D. Beklemishev. Induction rules, reflection principles, and provably recursive functions. *Annals of Pure and Applied Logic*, 85:193–242, 1997.

[2] L.D. Beklemishev. A proof-theoretic analysis of collection. *Archive for Mathematical Logic*, 37:275–296, 1998.

[3] L.D. Beklemishev. Proof-theoretic analysis by iterated reflection. Preprint, `http://wwwmath.uni-muenster.de/logik/publ/pre/6.html`, 1999.

[4] L.D. Beklemishev. On the schema of induction for decidable predicates. Logic Group Preprint Series 205, University of Utrecht, June 2000. `http://www.phil.uu.nl/preprints.html`.

[5] G. Boolos. *The Logic of Provability.* Cambridge University Press, Cambridge, 1993.

[6] G.K. Dzhaparidze. *Modal Logical Means of Investigation of Provability.* PhD thesis, Moscow State University, 1986. In Russian.

[7] S. Feferman. Three conceptual problems that bug me. Lecture text for 7-th Scandinavian Logic Symposium, ftp://math.stanford.edu/pub/papers/feferman/conceptualprobs.ps.gz, 1996.

[8] J.-Y. Girard. *Proof theory and logical complexity, Vol. 1.* Bibliopolis, Napoli, 1987.

[9] K.N. Ignatiev. On strong provability predicates and the associated modal logics. *Journal of Symbolic Logic*, 58:249–290, 1993.

[10] G. Kreisel. Wie die Beweistheorie zu ihren Ordinalzahlen kam und kommt. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 78(4):177–223, 1977.

[11] L. Kristiansen. Subrecursive degrees and fragments of Peano arithmetic. Manuscript, 1998.

[12] 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.

[13] Pour-El M.B. and Kripke S. Deduction-preserving "recursive isomorphisms" between theories. *Fundamenta Mathematicae*, 61:141–163, 1967.

[14] W. Pohlers. Subsystems of set theory and second order number theory. In S.R. Buss, editor, *Handbook of Proof Theory*, pages 210–335. Elsevier, North-Holland, Amsterdam, 1998.

[15] U.R. Schmerl. A fine structure generated by reflection formulas over Primitive Recursive Arithmetic. In M. Boffa, D. van Dalen, and K. McAloon, editors, *Logic Colloquium'78*, pages 335–350. North Holland, Amsterdam, 1979.

[16] V.Yu. Shavrukov. Subalgebras of diagonalizable algebras of theories containing arithmetic. *Dissertationes Mathematicae*, 323, 1993.

[17] R.M. Solovay. Provability interpretations of modal logic. *Israel Journal of Mathematics*, 28:33–71, 1976.

[18] A.M. Turing. System of logics based on ordinals. *Proc. London Math. Soc.*, ser. 2, 45:161–228, 1939.