Saturated Models of universal theories

Jeremy Avigad

July 10, 2000

Technical Report No. CMU-PHIL-112

Philosophy

Methodology

Logic

Carnegie Mellon

Pittsburgh, Pennsylvania 15213

Saturated models of universal theories*

Jeremy Avigad

July 10, 2000

Abstract

A notion called *Herbrand saturation* is shown to provide the model-theoretic analogue of a proof-theoretic method, Herbrand analysis. This provides uniform and sometimes simplified model-theoretic proofs of a number of important conservation theorems, and helps clarify the relationship between semantic and syntactic methods. A constructive, algebraic variation of the new method is described, providing yet a third approach, which is finitary but retains the semantic flavor of the model-theoretic version.

1 Introduction

Many important theorems in proof theory are *conservation* theorems, which is to say, they have the following form: if a theory T_1 proves a sentence φ of a certain kind, then another theory T_2 proves φ as well, or, perhaps, a specified translation, φ'. Typically the foundational interest in such a theorem lies in the *reduction* of T_1 to T_2: though, on the surface, the principles embodied in T_1 may seem "stronger" or "more abstract" than those of T_2, the conservation theorem shows that there is at least a sense in which the stronger principles can be eliminated in certain contexts.

Some examples of interesting conservation theorems are the following:

1. WKL_0, a fragment of second-order arithmetic based on a weak version of König's lemma, and hence also $I\Sigma_1$, a fragment of first-order arithmetic based on Σ_1 induction, are conservative over Primitive Recursive Arithmetic (PRA) for Π_2 sentences.

2. S^1_2, a weak fragment of arithmetic, is conservative over PV, a theory of polynomial-time computable functions, for sentences that are $\forall\exists\Sigma^b_1$.

3. Σ^1_1-AC_0, a fragment of second-order arithmetic based on arithmetic induction, arithmetic comprehension, and arithmetic choice axioms, is conservative over first-order arithmetic (for sentences in the common language).

*This is not a final version; it has been submitted for publication.

4. For each $k \geq 0$, $B\Sigma_{k+1}$, a fragment of arithmetic based on Σ_{k+1} collection axioms, is conservative over $I\Sigma_k$ for Π_{k+2} sentences.

In all these examples, both proof-theoretic and model-theoretic proofs are available, and overall neither approach can claim a clear advantage. The conservation of $I\Sigma_1$ over PRA is due to Mints, Parsons, and Takeuti, independently, obtained by syntactic methods in each case. The first model-theoretic proof is due to Paris and Kirby, using the notion of a semi-regular cut. The conservation of S^1_2 over PV is due to Buss [8], also using a proof-theoretic argument; the first model-theoretic proof is due to Wilkie.

The other results were first obtained using model-theoretic methods. The conservation of WKL_0 over PRA is due to Friedman. The conservation of Σ^1_1-AC over PA is due to Barwise and Schlipf [6], using recursively saturated models. Finally, the conservation of $B\Sigma_{k+1}$ over $I\Sigma_k$ is due to Friedman and Paris independently, using compactness and an ultrapower construction, respectively. For these three cases, the first proof-theoretic proofs are due to Sieg. For model-theoretic proofs of the results just described, see [26, 19, 20].

In these examples, the relationship between the model-theoretic and proof-theoretic methods is not transparent. And while the model-theoretic methods used to obtain these results are varied (including the use of semiregular cuts, recursive saturation, ultrapowers, and so on), it turns out that, in contrast, a single proof-theoretic method suffices throughout. *Herbrand analysis*, developed most fully by Sieg in [24, 25], applies most directly to universally axiomatized theories; but by introducing appropriate Skolem functions, the methods can be used to obtain all the results described above. Buss' *witnessing method* [8, 10] is equally general, and, at the core, is very similar to Herbrand analysis.

In Section 3, I will define a notion called *Herbrand saturation*, and I will show that every universal theory has an Herbrand-saturated model. In Section 4, I will show that this notion provides, in a sense, a semantic version of Herbrand analysis, allowing one to carry out essentially the same arguments while avoiding the use of the cut-elimination theorem. In the case of bounded arithmetic, this construction has been used in Zambella [29], where it is attributed to unpublished work by Visser; see also [22, Section 7.6]. Section 4 simply notes that the general construction is widely applicable, a fact which provides uniform and sometimes simplified model-theoretic proofs of a number of conservation theorems, and clarifies the relationship between the model-theoretic and proof-theoretic approaches.

Of course, proof-theoretic methods are not only finitary but have the advantage of providing explicit translations between the two theories, information that is apparently lost in the model-theoretic versions. This information can, however, be recovered. In Section 5, I will provide a constructive, algebraic variation of the construction, which lies midway between the model-theoretic and proof-theoretic versions. Algebraic methods like these can be found in the work of Dragalin (e.g. [16, 17]), where they are used to obtain similar proof-theoretic results; the approach I take below stems more directly from ideas found in [1, 3, 4, 13, 14, 15]. Though the algebraic constructions maintain most

of the semantic flavor of the model-theoretic ones, in Section 6 I show that the method leads to finitary consistency proofs, as well as algorithms that translate proofs from one theory to the other.

I am grateful to Thierry Coquand for drawing my attention to [13, 15], for showing me an algebraic proof of the convervation of ACA_0 over PA, and for emphasizing the constructive nature of these methods.

2 Preliminaries

Unless I specify otherwise, the logic in question is always classical first-order logic with equality. A formula is said to be universal (resp. existential) if it consists of a string of universal (resp. existential) quantifiers, possibly empty, followed by a quantifier-free formula. The classes of formulae that are $\forall\exists$, $\exists\forall$, etc. are defined similarly. I will identify formulae that differ only in the names of their bound variables, and use the notation $\varphi[t/x]$ to denote the result of substituting t for x in φ, renaming bound variables if necessary. Once a formula has been introduced as $\varphi(x)$, $\varphi(t)$ then abbreviates $\varphi[t/x]$. I will use $\vec{x}$ and $\vec{t}$ to denote finite sequences of variables and terms, $\varphi[\vec{t}/\vec{x}]$ to denote the simultaneous substitution of $\vec{t}$ for $\vec{x}$, etc. If φ is a formula with free variables, saying that φ is provable or valid is equivalent to saying that its universal closure is. For convenience, I will assume that all the languages we consider have at least one constant.

I will say that a theory T is *universal* if it can be axiomatized by a universal set of sentences. Herbrand's theorem can be stated follows:

Theorem 2.1 (Herbrand's theorem) *Suppose T is a universal theory, and $T \vdash \exists\vec{y}\ \psi(\vec{x},\vec{y})$, where ψ is quantifier-free with the free variables shown. Then there are sequences of terms $\vec{t}_1(\vec{x}),\ldots,\vec{t}_k(\vec{x})$ with at most the free variables shown, such that*

$$T \vdash \psi(\vec{x},\vec{t}_1(\vec{x})) \vee \ldots \vee \psi(\vec{x},\vec{t}_k(\vec{x})).$$

In fact, the latter is provable from substitution instances of axioms of T and equality axioms, using only propositional logic.

By soundness and completeness, provability can be exchanged with semantic entailment in the statement of the theorem. Herbrand's theorem is an easy consequence of the cut-elimination theorem, but it has an easy model-theoretic proof as well: if the conclusion fails, then T is consistent with the set

$$\{\neg\psi(\vec{c},t(\vec{c})) \mid t(\vec{x}) \text{ a term in the language of } T\},$$

where $\vec{c}$ is a sequence of new constants; by completeness, T together with this set has a model; taking the submodel generated by the set of terms $\{t(\vec{c})\}$ yields a model of T in which $\exists\vec{y}\ \psi(\vec{c},\vec{y})$ is false. A refinement of this argument yields the second, stronger statement. Note that in particular, the theorem implies that if T is a universal theory and φ is quantifier free, then T proves φ if and

only if φ is provable from substitution instances of axioms of T and equality axioms, using only propositional logic.

Say that "T supports definition by cases" if for every sequence of terms $t_1(\vec{x}), \ldots, t_k(\vec{x})$ and quantifier-free formulae $\theta_1(\vec{x}), \ldots, \theta_{k-1}(\vec{x})$ there is a function symbol f such that T proves

$$f(\vec{x}) = \begin{cases} t_1(\vec{x}) & \text{if } \theta_1(\vec{x}) \\ t_2(\vec{x}) & \text{if } \neg\theta_1(\vec{x}) \wedge \theta_2(\vec{x}) \\ \vdots & \\ t_k(\vec{x}) & \text{otherwise.} \end{cases}$$

If T is a universal theory and T supports definition by cases, then Theorem 2.1 implies that whenever $T \vdash \forall\vec{x}\ \exists\vec{y}\ \psi(\vec{x},\vec{y})$ and ψ is quantifier-free, there is a sequence of function symbols $\vec{f}$ such that $T \vdash \forall\vec{x}\ \psi(\vec{x}, \vec{f}(\vec{x}))$.

If $\mathcal{M}$ is a structure for a language L, let $L(\mathcal{M})$ denote the language with additional constants to denote the elements of the universe of $\mathcal{M}$. A *type with parameters from* $\mathcal{M}$ is a set of sentences in an extension of $L(\mathcal{M})$ by finitely many constants. When the context is clear, I will say "type" instead of "type with parameters from $\mathcal{M}$." A type Γ is said to be *realized in* $\mathcal{M}$ if there is an interpretation of the additional constants by elements of the universe, making every sentence in Γ true in $\mathcal{M}$. A type is *universal* if all its sentences are universal, and a type is *principal* if, in fact, it consists of a single sentence. The *universal diagram* of $\mathcal{M}$ is the set of universal sentences of $L(\mathcal{M})$ that are true in $\mathcal{M}$.

3 Herbrand saturation

Definition 3.1 *Let $\mathcal{M}$ be a structure for a language L. $\mathcal{M}$ is* Herbrand saturated *if every principal universal type consistent with the universal diagram of $\mathcal{M}$ is realized in $\mathcal{M}$.*

Put slightly differently, the definition requires that any $\exists\forall$ sentence of $L(\mathcal{M})$ that is consistent with the universal diagram of $\mathcal{M}$ is true in $\mathcal{M}$. To compare this to the more traditional notion, recall that a model $\mathcal{M}$ is said to be *saturated* if every type in $L(\mathcal{M})$ that is consistent with the *complete* diagram of $\mathcal{M}$ is realized in $\mathcal{M}$. Here we only require that certain principal universal types are realized; but to be realized, the type has only to be consistent with the *universal* diagram of $\mathcal{M}$.[1]

Theorem 3.2 *Every consistent universal theory has an Herbrand-saturated model.*

Proof. Let L be the language of T, and let L_ω denote a new language with an additional sequence of new constant symbols $c_0, c_1, c_2, \ldots$. Let $\theta_1(\vec{x}_1, \vec{y}_1), \theta_2(\vec{x}_2, \vec{y}_2), \ldots$

[1] In fact, the obvious modification of Theorem 3.2 using a transfinite iteration would enable us to realize arbitrary universal types, not just the principal ones.

enumerate the quantifier-free formulae the new language. Recursively construct an increasing sequence of sets S_i of universal sentences, as follows. First, let S_0 be a set of universal axioms for T. At stage $i+1$, try to satisfy $\forall \vec{y}_{i+1}\ \theta_{i+1}(\vec{x}_{i+1}, \vec{y}_{i+1})$: pick a new sequence of constants $\vec{c}$ that do not occur in S_i or θ_{i+1}, and let

$$S_{i+1} = \begin{cases} S_i \cup \{\forall \vec{y}_{i+1}\ \theta_{i+1}(\vec{c}, \vec{y}_{i+1})\} & \text{if this is consistent} \\ S_i & \text{otherwise.} \end{cases}$$

By induction, each S_i is consistent, and hence so is their union, S_ω. Let $\mathcal{N}$ be a model of S_ω, and let $\mathcal{M}$ be the submodel of $\mathcal{N}$ whose universe is generated by the terms of L_ω; that is, $|\mathcal{M}| = \{t^{\mathcal{N}} \mid t \in L_\omega\}$. Since S_ω is a set of universal sentences, $\mathcal{M}$ is also a model of S_ω, and therefore a model of T.

Note that every element of the universe of $\mathcal{M}$ is denoted by one of the constants c_j. This is true because each element of the universe of $\mathcal{M}$ is denoted by a term t in L_ω; pick i such that θ_i is the formula $x = t$, so for some constant c the formula $c = t$ is an element of S_{i+1}.

Now it is not difficult to show that $\mathcal{M}$ is Herbrand saturated. Suppose $\mathcal{M} \not\models \exists \vec{x}\ \forall \vec{y}\ \varphi(\vec{x}, \vec{y}, \vec{a})$, where φ is quantifier-free and $\vec{a}$ is a sequence of parameters from $\mathcal{M}$. We need to show that this formula is inconsistent with the universal diagram of $\mathcal{M}$. Let $\vec{d}$ be a sequence of constants in L_ω denoting the elements $\vec{a}$, choose i such that θ_{i+1} is the formula $\varphi(\vec{x}, \vec{y}, \vec{d})$, and let $\vec{c}$ be the constants used at stage $i+1$ in the construction. Then $\mathcal{M} \not\models \forall \vec{y}\ \varphi(\vec{c}, \vec{y}, \vec{d})$, and so, by the construction, the latter formula is inconsistent with S_i. Since $\vec{c}$ does not occur in S_i, the formula $\exists \vec{x}\ \forall \vec{y}\ \varphi(\vec{x}, \vec{y}, \vec{d})$ is also inconsistent with S_i. But, renaming $\vec{d}$ and the constants in S_i to the constants of $L(\mathcal{M})$ that name the same elements, S_i is a subset of the universal diagram of $\mathcal{M}$. □

If $\mathcal{M}$ is any model and S is a finite subset of its universal diagram, then S is also satisfied by the submodel of $\mathcal{M}$ generated by the elements mentioned in S. This can be used to show that the restriction to universal theories in Theorem 3.2 is necessary. For example, if T is the theory of dense linear orders with at least two points, and $\mathcal{M}$ is a model of T, then the $\exists\forall$ sentence asserting the existence of two points with nothing between them is consistent with the universal diagram of $\mathcal{M}$, but is inconsistent with T, and hence false in $\mathcal{M}$.

Given any model $\mathcal{M}$, one can let T be the universal diagram of $\mathcal{M}$ and apply Theorem 3.2. With this fact in mind, it is not hard to see that Theorem 3.2 implies (and is implied by) the statement that every model $\mathcal{M}$ has a Σ_1-elementary extension that is Herbrand saturated.

The following theorem describes a feature of Herbrand-saturated models that makes them useful: any $\forall\exists$ sentence true in such a model is "witnessed," in a strong way, by a finite set of terms with parameters.

Theorem 3.3 *Let $\mathcal{M}$ be an Herbrand-saturated structure for a language L. Suppose $\mathcal{M} \models \forall \vec{x}\ \exists \vec{y}\ \varphi(\vec{x}, \vec{y}, \vec{a})$, where $\varphi(\vec{x}, \vec{y}, \vec{z})$ is a quantifier-free formula in L, and $\vec{a}$ is a sequence of parameters from $\mathcal{M}$. Then there is a universal formula*

$\psi(\vec{z},\vec{w})$ with the free variables shown, and sequences of terms $\vec{t}_1(\vec{z},\vec{w}),\ldots,\vec{t}_k(\vec{z},\vec{w})$, such that $\mathcal{M} \models \exists\vec{w}\ \psi(\vec{a},\vec{w})$, and

$$\models \psi(\vec{z},\vec{w}) \rightarrow \varphi(\vec{x},\vec{t}_1(\vec{x},\vec{z},\vec{w}),\vec{z}) \vee \ldots \vee \varphi(\vec{x},\vec{t}_k(\vec{x},\vec{z},\vec{w}),\vec{z}).$$

Note that the last formula is valid, and hence provable in pure logic. In particular, the conclusion of the theorem implies that there is a sequence of parameters $\vec{b}$ such that $\forall\vec{x}\ (\varphi(\vec{x},\vec{t}_1(\vec{x},\vec{a},\vec{b}),\vec{a}) \vee \ldots \vee \varphi(\vec{x},\vec{t}_k(\vec{x},\vec{a},\vec{b}),\vec{a}))$ is true in $\mathcal{M}$.

The proof of the theorem is just an application of Herbrand's theorem.

Proof. If $\exists\vec{x}\ \forall\vec{y}\ \neg\varphi(\vec{x},\vec{y},\vec{a})$ is not true in $\mathcal{M}$, then it is inconsistent with the universal diagram of $\mathcal{M}$. This implies that there is a universal formula $\psi(\vec{z},\vec{w})$ of L, and a sequence of parameters $\vec{b}$ from $\mathcal{M}$, such that $\mathcal{M} \models \psi(\vec{a},\vec{b})$ and $\models \psi(\vec{a},\vec{b}) \rightarrow \exists\vec{y}\ \varphi(\vec{x},\vec{y},\vec{a})$. Replace the constants $\vec{a}$ and $\vec{b}$ by variables $\vec{z}$ and $\vec{w}$, note that the resulting formula is equivalent to an existential sentence, and apply Herbrand's theorem. □

Finally, the following theorem provides us with a recipe for proving conservation theorems.

Theorem 3.4 *Let T_2 be a universal theory and let T_1 be a theory in the language of T_2. If every Herbrand-saturated model of T_2 is also a model of T_1, then every $\forall\exists$ sentence provable in T_1 is also provable in T_2.*

Proof. Suppose every Herbrand-saturated model of T_2 is a model of T_1. Let $\varphi(\vec{x},\vec{y})$ be a quantifier-free formula in the language of T_2, with the free variables shown, and suppose that T_2 does not prove $\forall\vec{x}\ \exists\vec{y}\ \varphi(\vec{x},\vec{y})$. We will show that T_1 does not prove it either.

The second assumption implies that $T_2 \cup \{\forall\vec{y}\ \neg\varphi(\vec{d},\vec{y})\}$ is a consistent universal theory, where $\vec{d}$ is a sequence of new constants. By Proposition 3.2, there is an Herbrand-saturated model of this theory; but then the reduct of this model to the language of T_2 is an Herbrand-saturated model of T_2 satisfying $\exists\vec{x}\ \forall\vec{y}\ \neg\varphi(\vec{x},\vec{y})$. By our hypothesis, this is also a model of T_1, in which $\forall\vec{x}\ \exists\vec{y}\ \varphi(\vec{x},\vec{y})$ is false. □

4 Applications

In this section I will show that the notion of Herbrand saturation does much the same work that the methods of Herbrand analysis typically do. I will focus on the conservation of $I\Sigma_1$ over *PRA* as a prototypical case, and then briefly discuss the other conservation results mentioned in Section 1.

The set of primitive recursive functions is the smallest set of functions (of various arities) from the natural numbers to the natural numbers, containing the constant zero, projections, and the successor function, and closed under composition and primitive recursion. The language of Primitive Recursive Arithmetic, or *PRA*, has a symbol for each primitive recursive function. The axioms of *PRA*

consist of quantifier-free defining equations for these functions, and a schema of induction for quantifier-free formulae. A relation is said to be primitive recursive if and only if its characteristic function is, and it is not hard to show that the primitive recursive relations are closed under Boolean operations and bounded quantification. Induction is then provably equivalent to the schema

$$\forall y\ (\varphi(0) \wedge \forall x < y\ (\varphi(x) \rightarrow \varphi(x+1)) \rightarrow \varphi(y)),$$

where φ is quantifier-free (or even atomic), possibly with free variables other than the one shown. Using these facts, one can show that *PRA* has a universal axiomatization.

A formula in the language of arithmetic is said to be Δ_0, or *bounded*, if all the quantifiers are bounded, and Σ_1 if it is of the form $\exists \vec{y}\ \varphi(\vec{y}, \vec{z})$, where φ is Δ_0. $I\Sigma_1$ denotes the fragment of Peano Arithmetic in which induction is restricted to Σ_1 formulae.

Theorem 4.1 *$I\Sigma_1$ is conservative over PRA for Π_2 sentences.*

Proof. Let $\mathcal{M}$ be an Herbrand-saturated model of *PRA*. By Proposition 3.4, we only need to show that $\mathcal{M}$ satisfies the schema of Σ_1 induction. Over *PRA*, every Σ_1 formula $\eta(x, \vec{z})$ is equivalent to one of the form $\exists y\ \varphi(x, y, \vec{z})$, where φ is quantifier-free; so it suffices to consider induction for formulae of that form.

To that end, suppose $\vec{a}$ is a sequence of parameters in $\mathcal{M}$, and $\mathcal{M}$ satisfies the induction hypotheses:

- $\exists y\ \varphi(0, y, \vec{a})$
- $\forall x\ (\exists y\ \varphi(x, y, \vec{a}) \rightarrow \exists y\ \varphi(x+1, y, \vec{a}))$.

We need to show that $\mathcal{M}$ satisfies $\forall x\ \exists y\ \varphi(x, y, \vec{a})$. First, observe that the second formula is equivalent to $\forall x, y\ \exists y'\ (\varphi(x, y, \vec{a}) \rightarrow \varphi(x+1, y', \vec{a}))$. Using Theorem 3.3 and the fact that *PRA* supports definition by cases, we have that there are sequences of parameters $\vec{b}_1$ and $\vec{b}_2$ and function symbols $f(\vec{z}, \vec{w}_1)$ and $g(x, y, \vec{z}, \vec{w}_2)$ such that $\mathcal{M}$ satisfies

- $\varphi(0, f(\vec{a}, \vec{b}_1), \vec{a})$ and
- $\forall x, y\ (\varphi(x, y, \vec{a}) \rightarrow \varphi(x+1, g(x, y, \vec{a}, \vec{b}_2), \vec{a}))$.

Let $\vec{b}$ denote the concatenation of $\vec{b}_1$ and $\vec{b}_2$, and let $\vec{w}$ denote the concatenation of $\vec{w}_1$ and $\vec{w}_2$. Let $h(x, \vec{z}, \vec{w})$ be the function symbol of *PRA* with defining equations

$$\begin{aligned} h(0, \vec{z}, \vec{w}) &= f(\vec{z}, \vec{w}_1) \\ h(x+1, \vec{z}, \vec{w}) &= g(x, h(x, \vec{z}, \vec{w}), \vec{z}, \vec{w}_2). \end{aligned}$$

Then $\mathcal{M}$ satisfies

- $\varphi(0, h(0, \vec{b}, \vec{a}), \vec{a})$ and

- $\forall x\ (\varphi(x, h(x, \vec{b}, \vec{a}), \vec{a}) \rightarrow \varphi(x, h(x+1, \vec{b}, \vec{a}), \vec{a}))$.

Since $\mathcal{M}$ is a model of PRA and hence satisfies quantifier-free induction, we have $\mathcal{M} \models \forall x\ \varphi(x, h(x, \vec{a}, \vec{b}), \vec{a})$, and hence $\mathcal{M} \models \forall x\ \exists y\ \varphi(x, y, \vec{a})$, as desired. □

The argument for the conservation of S_2^1 over PV is similar. It is convenient to take the first-order version of PV to be the theory CPV of [12], and then one only needs to show that Σ_1^b polynomial induction holds in any Herbrand-saturated model. The proof parallels the one above, except one uses bounded recursion on notations in place of primitive recursion.

At this point, it should be clear to readers familiar with [24] how one can adapt the arguments found there to obtain model-theoretic proofs of the other conservation results mentioned in the introduction. As a result, I will give only a few rough indications, and refer the reader to [24] for details (but see the comments below for some corrections).

To prove that WKL_0 is conservative over PRA for Π_2 sentences, use a many-sorted "second-order" version of PRA, denoted PRA_2, with function variables of the various aritites. Take composition and primitive recursion to be operations on the function sorts. Since any quantifier-free proof in PRA_2 of a formula without function variables is essentially a proof in PRA, it suffices to show that every Herbrand-saturated model of PRA_2 is a model of WKL_0.

So let $\mathcal{M}$ be such a model, and let g represent a binary tree in $\mathcal{M}$. Suppose, in $\mathcal{M}$, there is no infinite path through g; this means that for every infinite binary sequence f, there is an x such that f has left the tree by level x. We need to show that g is finite. By Herbrand saturation, there is a term $t(f)$ with parameters from $\mathcal{M}$, such that for each infinite binary sequence f, f has left the tree by level $t(f)$. By induction on terms one can show that there is a term b majorizing $t(f)$, provably in PRA_2; in other words, b does not involve f, and PRA_2 proves

$$\forall x\ (f(x) \leq 1) \rightarrow t(f) \leq b.$$

This implies that, in $\mathcal{M}$, g has no nodes at level b. Hence g is finite. (This is essentially the argument in [24]; but page 69 of [21] corects some errors in that account.)

To prove that $B\Sigma_{k+1}$ is conservative over $I\Sigma_k$ for Π_{k+2} sentences, embed $I\Sigma_k$ in a universal theory with Skolem functions returning least witnesses to Σ_k formulae. With these Skolem functions, Σ_k and Π_k formulae in the language of arithmetic are equivalent to formulae in the new language that are quantifier-free. Let $\mathcal{M}$ be an Herbrand-saturated model of this theory. Suppose $\varphi(x, y)$ is a Π_k formula with parameters in $\mathcal{M}$, such that the antecedent of the collection axiom, $\forall x < a\ \exists y\ \varphi(x, y)$, is true in $\mathcal{M}$. By Herbrand-saturation, there is a sequence of terms such that

$$\forall x < a\ (\varphi(x, t_1(x)) \vee \ldots \vee \varphi(x, t_k(x)))$$

is true in $\mathcal{M}$. Using strong Σ_k collection, derivable in $I\Sigma_k$, one can prove that the values of $t_1, \ldots, t_k$ are bounded, for values of x less than a. (The argument in

[24] is not quite right, but can be repaired with the use of strong Σ_k collection, as in [19, Section 1.63]. For other proof-theoretic proofs of this conservation result, see [9] and [7].)

Finally, to prove that Σ^1_1-AC_0 is conservative over Peano Arithmetic, embed PA in a second-order universal theory with function symbols. In this theory, allow operations on the function sorts that define new functions by composition, and operations μ that define new functions by minimization:

$$f(x, \vec{y}) = 0 \rightarrow f(\mu(f)(\vec{y})) = 0 \wedge \mu(f)(\vec{y}) \leq x.$$

With these μ operations, every arithmetic formula, possibly involving function variables, is equivalent to a formula that is quantifier-free. Let $\mathcal{M}$ be any Herbrand-saturated model of this theory. Suppose $\forall x\ \exists f\ \varphi(x, f)$ holds in $\mathcal{M}$, where φ is arithmetic. By Herbrand saturation, there is a sequence of terms $t_1(x), \ldots, t_k(x)$, such that

$$\forall x\ (\varphi(x, t_1(x)) \vee \ldots \vee \varphi(x, t_k(x)))$$

is true in $\mathcal{M}$. From $t_1, \ldots, t_k$ it is not difficult to obtain a term s such that $\mathcal{M}$ satisfies $\forall x\ \varphi(x, s_x)$, as required.

It seems worth mentioning that by combining the notion of Herbrand saturation with the methods of [9] and [2] one can carry out the ordinal analysis of, say, Peano arithmetic, without relying on cut-elimination. For example, if α is infinite and closed under multiplication, an Herbrand-saturated model of a suitable theory of $<\alpha$-recursion yields a model of Π_1 transfinite induction below α; and an Herbrand-saturated model of a suitable Skolemized version of Π_n transfinite induction below ω^α yields a model of Π_{n+1} transfinite induction below α. For another model-theoretic approach to ordinal analysis, see [5].

5 An algebraic version

There is a more direct way of obtaining the model $\mathcal{M}$ constructed in the proof of Theorem 3.2: given S_ω, let $\hat{S}$ be a maximally consistent extension, and "read off" a model from that. If we allow ourselves to be content with a Boolean-valued model instead of a traditional two-valued one, we can avoid the use of the maximally consistent extension. Instead of enumerating constants and formulae, we can build our model generically, using conditions to represent finite portions of S_ω and reasoning about what, on the basis of such a condition, is forced to be true in the maximal extension. In order to render our proofs entirely constructive, we can even omit the "consistency check" used in the proof of Theorem 3.2; we need only accept the fact that some of our conditions will force falsity.

In this section I will provide a constructive proof of the conservation of $I\Sigma_1$ over PRA, based on these ideas. In the next section I will make the sense in which the proof is constructive more precise. It will be clear, I hope, that the method can be adapted to the other conservation theorems as well, or to a general proof-theoretic analogue of Theorem 3.4.

Let L be the language of PRA, and let L_ω be the language with infinitely many new constant symbols $a, b, c, \ldots$. A *condition* is simply a finite set of universal sentences of L_ω. The definition below describes a relationship between conditions p and sentences φ of L_ω, where "p forces φ" means, intuitively, that on the basis of p we can determine that φ will necessarily be true in the model we are constructing. In fact, we will describe this relationship in two steps: first we will use the double-negation translation to translate each sentence φ to a negative sentence φ^N, and then we will say what it means for a condition to force a sentence of that form. Of course, these two steps can be combined to yield a forcing relation for classical logic; but given the nonstandard treatment of falsity (and hence negation) the one-step version would be difficult to work with.

Let us take the formulae of intuitionstic logic to be built up using the connectives $\forall$, $\exists$, $\wedge$, $\vee$, $\rightarrow$, and $\bot$, with $\neg\varphi$ taken to abbreviate $\varphi \rightarrow \bot$. A formula in this language is said to be *negative* if it does not involve $\exists$ and $\vee$. The Gödel-Gentzen double-negation translation for classical logic takes classical formulae φ to negative formulae φ^N, mapping atomic formula θ to $\neg\neg\theta$, $\bot$ to $\bot$, $\varphi \vee \psi$ to $\neg(\neg\varphi^N \wedge \neg\psi^N)$, $\varphi \wedge \psi$ to $\varphi^N \wedge \psi^N$, $\varphi \rightarrow \psi$ to $\varphi^N \rightarrow \psi^N$, $\exists x\, \varphi$ to $\neg\forall x\, \neg\varphi^N$, and $\forall x\, \varphi$ to $\forall x\, \varphi^N$. If Γ is a set of sentences, then Γ^N denotes the set of their $\cdot^N$-translations.

Theorem 5.1 *If Γ proves φ classically, then Γ^N proves φ^N in the negative fragment of minimal logic.*

Minimal logic can be described as the subsystem of intuitionistic logic obtained by leaving out the rule *ex falso sequitur quodlibet*, "from $\bot$ conclude anything."

Definition 5.2 *If θ is an atomic sentence of L_ω, define $p \Vdash \theta$ to mean $PRA \cup p \vdash \theta$. Extend the forcing notion to arbitrary negative formulae in the language of PRA inductively, via the following clauses:*

$$\begin{aligned} p \Vdash (\theta \wedge \eta) &\equiv p \Vdash \theta \text{ and } p \Vdash \eta \\ p \Vdash (\theta \rightarrow \eta) &\equiv \text{for every condition } q \supseteq p, \text{ if } q \Vdash \theta, \text{ then } q \Vdash \eta \\ p \Vdash \forall x\, \theta(x) &\equiv \text{for every closed term } t \text{ of } L_\omega,\ p \Vdash \theta(t) \end{aligned}$$

A formula ψ is said to be forced, written $\Vdash \psi$, if $\emptyset \Vdash \psi$.

Notes. 1. I am taking $\bot$ to be an atomic formula, so $p \Vdash \bot$ means that $PRA \cup p$ is inconsistent.

2. Since the sentences in p are universal, for atomic θ we have that p forces θ if and only if there is a quantifier-free (or even propositional) proof of θ from substitution instances of p and the axioms of PRA. Indeed, we could have defined $p \Vdash \theta$ that way, and then, with some additional care, we could avoid uses of Herbrand's theorem below.

3. Since the extra constants of L_ω are not mentioned in the axioms of PRA, they can be treated like variables; that is, if $p(\vec{x}) \cup \{\theta(\vec{x})\}$ is a set of formulae in the language of PRA and $\vec{c}$ is a sequence of new constants, then $PRA \cup p(\vec{x})$

proves $\theta(\vec{x})$ if and only if $PRA \cup p(\vec{c})$ proves $\theta(\vec{c})$. I will use this fact below without mentioning it explicitly.

The definition of forcing for atomic formulae has two special properties: first, it is monotone in p; and second, if $p \Vdash \bot$, then $p \Vdash \theta$ for any atomic formlua θ. The next two lemmata hold for any forcing relation with these properties.

Lemma 5.3 *The forcing relation defined above is monotone for all sentences of L_ω: if ψ is any formula, $p \Vdash \psi$, and $q \supseteq p$, then $q \Vdash \psi$. Also, if $p \Vdash \bot$, then $p \Vdash \psi$.*

Proof. An easy induction on ψ. □

Lemma 5.4 *Suppose ψ is a negative formula, and ψ is provable intuitionistically. Then ψ is forced.*

Proof. Take intuitionistic logic to be given by a system of natural deduction (as in, say, [28]), and prove the following by induction on derivations: if $\Gamma \cup \{\varphi\}$ is a finite set of negative formulae of L_ω with free variables among $\vec{x}$, and φ is provable from Γ intutionistically, then for every condition p and sequence of terms $\vec{t}$, if $p \Vdash \theta[\vec{t}/\vec{x}]$ for every θ in Γ, then $p \Vdash \psi[\vec{t}/\vec{x}]$. □

Ultimately, our goal is to show that the double-negation translations of the axioms of $I\Sigma_1$ are forced.

Lemma 5.5 *Let φ be a quantifier-free sentence of L_ω. Then $p \Vdash \varphi$ if and only if $PRA \cup p \vdash \varphi$.*

Proof. By definition, this holds when φ is atomic. For the general case, use induction on φ. □

Lemma 5.6 *Let $\varphi(\vec{x})$ be a quantifier-free formula of L_ω with the free variables shown.*

1. $\{\forall \vec{x}\, \varphi(\vec{x})\} \Vdash \forall \vec{x}\, \varphi(\vec{x})$.
2. *If $p \Vdash \neg\forall \vec{x}\, \neg\varphi(\vec{x})$ and $\vec{c}$ is a list of the new constants appearing in $p \cup \{\varphi(\vec{x})\}$, then there are function symbols $\vec{f}$ of PRA such that $p \Vdash \varphi(\vec{f}(\vec{c}))$.*

Proof. For the first statement, we need to show that if $\vec{t}$ is any sequence of closed terms of L_ω then $PRA \cup \{\forall \vec{x}\, \varphi(\vec{x})\}$ proves $\varphi(\vec{t})$. This is easy.

For the second statement, suppose $p \Vdash \neg\forall \vec{x}\, \neg\varphi(\vec{x})$ and let $\vec{c}$ be a list of the new constants appearing in $p \cup \{\varphi(\vec{x})\}$. By the first clause, we have $\{\forall \vec{x}\, \neg\varphi(\vec{x})\} \Vdash \neg\varphi(\vec{x})$, and then by the definition of forcing for a negation, we have that $p \cup \{\forall \vec{x}\, \neg\varphi(\vec{x})\} \Vdash \bot$. But this means that $PRA \cup p \cup \{\forall \vec{x}\, \neg\varphi(\vec{x})\}$ is inconsistent, and hence $PRA \cup p$ proves $\exists \vec{x}\, \varphi(\vec{x})$. Applying Herbrand's theorem and using the fact that PRA supports definition by cases, we have a sequence of function symbols $\vec{f}$ such that $PRA \cup p$ proves $\varphi(\vec{f}(\vec{c}))$. By Lemma 5.5, this is equivalent to $p \Vdash \varphi(\vec{f}(\vec{c}))$. □

Lemma 5.7 *Let $\varphi(x)$ be a quantifier-free formula of L_ω with the free variable shown, and let t be a closed term of L_ω. Then $\{\varphi(t)\} \Vdash \neg\forall x\, \neg\varphi(x)$.*

Proof. Suppose $p \supseteq \{\varphi(t)\}$ and $p \Vdash \forall x\, \varphi(x)$. Then $p \Vdash \varphi(t)$ and $p \Vdash \neg\varphi(t)$, and hence $p \Vdash \bot$. □

Lemma 5.8 *The double-negation translation of each axiom of $I\Sigma_1$ is forced.*

Proof. Lemma 5.5 takes care of the quantifier-free axioms, so we only have to worry about the $\cdot^N$-translations of Σ_1 induction. Let $\eta(x, \vec{a})$ be a formula of the form $\neg\forall y\, \neg\varphi(x, y, \vec{a})$, where $\varphi(x, y, \vec{a})$ is a quantifier-free formula of L_ω with the free variable and extra constant symbols shown. It suffices to show that for any condition p, if $p \Vdash \eta(0, \vec{a}) \wedge \forall x\, (\eta(x, \vec{a}) \to \eta(x+1, \vec{a}))$, then $p \Vdash \forall x\, \eta(x, \vec{a})$.

To that end, suppose $p(\vec{a}, \vec{b})$ is a condition with at most the new constant symbols shown, such that

- $p(\vec{a}, \vec{b}) \Vdash \neg\forall y\, \neg\varphi(0, y, \vec{a})$, and
- $p(\vec{a}, \vec{b}) \Vdash \forall x\, (\neg\forall y\, \neg\varphi(x, y, \vec{a}) \to \neg\forall y\, \neg\varphi(x+1, y, \vec{a}))$.

Let c and d be additional new constants. Using Lemmata 5.6 and 5.7, there are function symbols f and g such that

- $p(\vec{a}, \vec{b}) \Vdash \varphi(0, f(\vec{a}, \vec{b}), \vec{a})$, and
- $p(\vec{a}, \vec{b}) \cup \{\varphi(c, d, \vec{a})\} \Vdash \varphi(c+1, g(c, d, \vec{a}, \vec{b}), \vec{a})$.

Replace $\vec{a}$, $\vec{b}$, c, and d by variables $\vec{z}$, $\vec{w}$, x, and y respectively, and let $\psi(\vec{w}, \vec{z})$ be the conjunction of the formulae in $p(\vec{w}, \vec{z})$. By Lemma 5.5, we have that *PRA* proves that $\psi(\vec{w}, \vec{z})$ implies

- $\varphi(0, f(\vec{w}, \vec{z}), \vec{z})$, and
- $\varphi(x, y, \vec{z}) \to \varphi(x+1, g(x, y, \vec{w}, \vec{z}), \vec{z})$.

As in Section 4, there is a function symbol h such that *PRA* proves

$$\psi(\vec{w}, \vec{z}) \to \forall x\, \varphi(x, h(x, \vec{w}, \vec{z}), \vec{w}).$$

Substituting the constants back for the variables and using Lemma 5.5, we have

$$p(\vec{a}, \vec{b}) \Vdash \forall x\, \varphi(x, h(x, \vec{a}, \vec{b}), \vec{a})$$

and hence, by Lemma 5.7,

$$p(\vec{a}, \vec{b}) \Vdash \forall x\, \neg\forall y\, \neg\varphi(x, y, \vec{a}),$$

as desired. □

Putting this all together, we have another proof that $I\Sigma_1$ is conservative over *PRA* for Π_2 sentences.

Proof (of Theorem 4.1). Suppose $I\Sigma_1$ proves $\forall x\ \exists y\ \varphi(x,y)$, where φ is quantifier-free. Then there is a conjunction α of finitely many axioms of $I\Sigma_1$ such that $\alpha \to \forall x\ \exists y\ \varphi(x,y)$ is provable in classical first-order logic, and so $\alpha^N \to \forall x\ \neg\forall y\ \neg\varphi^N(x,y)$ is provable intuitionistically. By Lemma 5.8, α^N is forced, and hence so is $\forall x\ \neg\forall y\ \neg\varphi^N(x,y)$. Using Lemmata 5.5 and 5.6, there is a function symbol f such that *PRA* proves $\forall x\ \varphi^N(x, f(x))$. Since *PRA* proves that φ^N is equivalent to φ, we can conclude that *PRA* proves $\forall x\ \exists y\ \varphi(x,y)$. □

Notes. 1. One can view the forcing relation given above as providing a description of truth in an associated Kripke model, provided we allow the case that some nodes force falsity; the universe at each node consists of the set of closed terms of L_ω. In fact, this structure models an intuitionistic version of $I\Sigma_1$ together with Markov's principle, and this theory, in turn, interprets $I\Sigma_1$ via the double-negation interpretation. These facts are implicit in the argument above; for a presentation that makes them more explicit, see [1].

2. Alternatively, to each formula φ, we can assign the set of conditions $[\![\varphi]\!] = \{p \mid p \Vdash \varphi^N\}$, where φ^N is the double-negation translation of φ. From that point of view, what we are doing is assigning to each formula φ a truth value in a complete Boolean algebra consisting of "regular" sets of conditions; for a presentation along these lines, see [13, 14].

6 Finitary proofs of the conservation theorems

From a foundational point of view, we would like to know that our conservation results can be established in a weak theory; and, given a proof in the stronger theory T_1, it would be nice to know how to go about *finding* a corresponding proof in T_2. In this section, I will show that the methods described in the last section yield proofs that are finitary, which is to say, they can be carried out in primitive recursive arithmetic (and, in fact, in a fragment thereof).[2] One can use this fact to obtain specific algorithms for carrying out the translations. Once again, I will focus on the conservation of $I\Sigma_1$ over *PRA* as a protoypical case.

To begin with, we need a weak fragment of arithmetic in which one can comfortably develop syntactic notions. To that end, we will use an axiomatization of the *elementary recursive functions*, i.e. the smallest set of functions containing zero, successor, addition, multiplication, and exponentiation, and closed under composition and bounded recursion. The set of elementary recursive functions is a subset of the set of primitive recursive functions, and every elementary function is bounded by some fixed iterate of the exponential function. The theory *ERA* is the analogue of *PRA* for the elementary functions; it can alternatively be viewed as a Skolem extension of $I\Delta_0(exp)$ (also known as *EFA*) that is universally axiomatizable.

[2] Another method of obtaining finitary proofs of conservation results like the ones we have been studying has recently been sketched by Friedman [18].

In *ERA*, one can formalize the notions of a term and a formula in the language of *PRA*, the notion of a proof from the axioms of *PRA*, and the notion of a condition. To each negative formula $\varphi(x_1, \ldots, x_k)$ in the language of *PRA*, the clauses of Definition 5.2 associate a formula $\Psi_\varphi(y, x_1, \ldots, x_k)$ of *ERA*; if $\ulcorner p \urcorner$ is a number coding a condition and $\ulcorner t_1 \urcorner, \ldots, \ulcorner t_k \urcorner$ are numbers coding terms of *PRA*, $\Psi_\varphi(\ulcorner p \urcorner, \ulcorner t_1 \urcorner, \ldots, \ulcorner t_k \urcorner)$ asserts that p forces $\varphi(t_1, \ldots, t_k)$. The quantifier complexity of Ψ_φ increases with φ, so one cannot hope to find a single formula $\Psi(\ulcorner \varphi \urcorner, \ulcorner p \urcorner, \ulcorner t_1 \urcorner, \ldots, \ulcorner t_k \urcorner)$ that captures the notion uniformly. But for each fixed proof d of a Π_2 sentence θ in *PRA*, Section 4 shows us how to find a proof, in *ERA*, that θ is forced; and hence a proof, in *ERA*, of the existential assertion

$$\exists d' \ (d' \text{ is a proof of } \theta \text{ in } PRA).$$

And since this construction is syntactic, it can be carried out in a finitary metatheory. In other words, *PRA* proves the following: if θ is any Π_2 sentence provable in $I\Sigma_1$, then *ERA* proves that θ is provable in *PRA*. These ideas lead to the finitary consistency proofs we are after.

Lemma 6.1 *PRA proves the* Π_2 *soundness of ERA.*

Proof. First, note that *PRA* proves the cut-elimination theorem, and hence Herbrand's theorem. Now argue in *PRA*. Suppose *ERA* proves a Π_2 sentence $\forall x \, \exists y \, \varphi(x, y)$. Then there is a term $t(x)$ of *ERA* and a propositional proof d of $\varphi(x, t(x))$ from substitution instances of axioms of *ERA* and equality axioms. Let n be any natural number, and $\bar{n}$ the corresponding numeral. By induction on d, $\varphi(\bar{n}, t(\bar{n}))$ is true. □

Theorem 6.2 *PRA proves that* $I\Sigma_1$ *is conservative over PRA for* Π_2 *sentences.*

Proof. Argue in *PRA*. If there a proof of a Π_2 sentence θ in $I\Sigma_1$, then *ERA* proves that there is a proof of θ in *PRA*. By the Σ_1 soundness of *ERA*, there really is such a proof. □

Notes. 1. In fact, the methods of Section 5 yield constructive proofs, which is to say, in the proof of Theorem 6.2 it is sufficient to use *ERA* with first-order *intuitionistic* logic. With that restriction, it might be more natural to use normalization instead of cut-elimination in the proof of Lemma 6.1.

2. To prove Lemma 6.1, and hence the conservation theorem, one only needs a theory strong enough to prove the cut-elimination or normalization theorem, and then evaluate terms and quantifier-free formulae of *ERA*. As a result, for the finitary metatheory it suffices to use either $I\Delta_0(superexp)$ or EFA^*, which assert the totality of an iterated exponential function.

3. In fact, in the finitary metatheory (be it *PRA*, EFA^*, etc.) we can prove the stronger conservation result: if $I\Sigma_1$ proves $\forall x \, \exists y \, \varphi(x, y)$ and φ is quantifier free, then there is a function symbol f and a propositional proof of $\varphi(x, f(x))$ from instances of equality axioms and axioms of *PRA*.

4. As noted above, in place of *ERA*, one only needs a theory strong enough to handle syntactic operations. One could simply construct a many-sorted theory with sorts for terms, formulae, finite sets of formulae, and proofs in *PRA*, with function symbols and quantifier-free axioms describing the requisite constructions.

5. If one applies modified realizability (see [27]) to the proofs constructed in the intuitionistic first-order version of *ERA* (or the syntactic theory just described), one obtains instead a typed lambda term denoting the desired proof. Thus we have a uniform (and efficient) procedure which assigns to each proof d of a Π_2 sentence in $I\Sigma_1$ a typed lambda term T_d denoting the corresponding proof in *PRA*, where T_d involving only syntactic constructions at the base types. Normalizing this term produces the desired proof. See [3] for a more detailed discussion along these lines.

5. Using Solovay's method of "shortening of cuts" (see [23]) one can show that, in general, the use of cut-elimination or normalization, with the potential superexponential increase in the length of proofs, cannot be avoided.

References

[1] Jeremy Avigad. Interpreting classical theories in constructive ones. To appear in the *Journal of Symbolic Logic.*

[2] Jeremy Avigad. Ordinal analysis without proofs. Submitted.

[3] Jeremy Avigad. Algebraic proofs of cut elimination. Submitted.

[4] Jeremy Avigad and Jeffrey Helzner. Transfer principles for intuitionistic nonstandard arithmetic. Submitted.

[5] Jeremy Avigad and Richard Sommer. A model-theoretic approach to ordinal analysis. *Bulletin of Symbolic Logic*, 3:17–52, 1997.

[6] Jon Barwise and J. S. Schlipf. On recursively saturated models of arithmetic. In D. Saracino and V. B. Weispfennig, editors, *Model theory and algebra: a memorial tribute to Abraham Robinson*, Lecture Notes in Mathematics #498, pages 42–55. Springer, 1975.

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

[8] Samuel Buss. *Bounded Arithmetic.* Bibliopolis, 1986.

[9] Samuel Buss. The witness function method and provably recursive functions of Peano arithmetic. In D. Westerståhl D. Prawitz, B. Skyrms, editor, *Proceedings of the Ninth International Congress on Logic, Methodology, and Philosophy of Science*, pages 29–68. Elsevier North-Holland, 1994.

[10] Samuel Buss. First-order proof theory of arithmetic. In Buss [11].

[11] Samuel Buss, editor. *The Handbook of Proof Theory.* North-Holland, 1998.

[12] Stephen Cook and Alasdair Urquhart. Functional interpretations of feasibly constructive arithmetic. *Annals of Pure and Applied Logic*, 63:103–200, 1993.

[13] Thierry Coquand. Two applications of boolean models. *Archive for Mathematical Logic*, 37:143–147, 1997.

[14] Thierry Coquand and Martin Hofmann. 'A new method fo establishing conservativity of classical systems over their intuitionistic version. *Mathematical Structures in Computer Science*, 9:323–333, 1999.

[15] Thierry Coquand and Jan Smith. An application of constructive completeness. In S. Berardi and M. Copo, editors, *Proccedings of the Workshop TYPES '95*, Lecture Notes in Computer Science 1158, pages 76–84. Springer, 1996.

[16] Albert Dragalin. *Mathematical Intuitionism: Introduction to Proof Theory.* Translations of mathematical monographs. American Mathematical Society, 1988.

[17] Albert Dragalin. Explict algebraic models for constructive and classical theories with nonstandard elements. *Studia Logica*, 53:33–61, 1995.

[18] Harvey Friedman. Finitist proofs of conservation. Posting to the Foundations of Mathematics forum, http://www.math.psu.edu/simpson/fom, September 29, 1999.

[19] Petr Hájek and Pavel Pudlák. *Metamathematics of first-order arithmetic.* Springer, 1993.

[20] Richard Kaye. *Models of Peano Arithmetic.* Clarendon, Oxford, 1991.

[21] Ulrich Kohlenbach. Mathematically strong subsystems of analysis with low rate of growth of provably recursive functionals. *Archive for Mathematical Logic*, 36:31–71, 1996.

[22] Jan Krajíček. *Bounded Arithmetic, Propositional Logic, and Complexity Theory.* Cambridge University, 1995.

[23] Pavel Pudlák. The lengths of proofs. In Buss [11].

[24] Wilfried Sieg. Fragments of arithmetic. *Annals of Pure and Applied Logic*, 28:33–72, 1985.

[25] Wilfried Sieg. Herbrand analyses. *Archive for Mathematical Logic*, 30:409–441, 1991.

[26] Stephen Simpson. *Subsystems of Second-Order Arithmetic.* Springer, 1998.

[27] A. S. Troelstra. Realizability. In Buss [11].

[28] A. S. Troelstra and Dirk van Dalen. *Constructivism in Mathematics: An Introduction*, volume 1. North-Holland, 1988.

[29] Domenico Zambella. Notes on polynomial bounded arithmetic. *Journal of Symbolic Logic*, 61:942–966, 1996.