

On the Structure of Initial Segments of Models of Arithmetic

Jan Krajíček and Pavel Pudlák

Mathematical Institute, Czechoslovak Academy of Sciences, Žitná 25, CS-11567 Praha 1
 Czechoslovakia

Abstract. For any countable nonstandard model M of a sufficiently strong fragment of arithmetic T , and any nonstandard numbers $a, c \in M$, $M \models c \leq a$, there is a model K of T which agrees with M up to a and such that in K there is a proof of contradiction in T with Gödel number $\leq 2^{a^c}$.

Introduction

For any M a model of arithmetic and $a \in M$, $M \upharpoonright a$ will denote the structure with the universe $\{i \in M \mid M \models i \leq a\}$, and with operations inherited from M . Thus $+$ and \cdot are only partial functions in $M \upharpoonright a$.

The general question that we study here is this: For which functions $f(x)$ the structure $M \upharpoonright a$ uniquely determines the structure $M \upharpoonright f(a)$?

“Determines” means “up to elementary equivalence” or equivalently as all $M \upharpoonright a$ are recursively saturated “up to isomorphism”.

It is easily seen that $M \upharpoonright a$ determines $M \upharpoonright a^k$, for any $k < \omega$. Paris and Dimitracopoulos [3] showed that $M \upharpoonright a$ does not determine $M \upharpoonright 2^a$. Namely, they proved that for any countable nonstandard model M of PA and any $a \in M$ nonstandard, there is K a model of PA such that $a \in K$, $M \upharpoonright a = K \upharpoonright a$, but $M \upharpoonright 2^{2^a} \not\cong K \upharpoonright 2^{2^a}$. Thus either $M \upharpoonright a$ does not determine $M \upharpoonright 2^a$ or $M \upharpoonright 2^a$ does not determine $M \upharpoonright 2^{2^a}$.

Later Hájek [3] found one Δ_0 formula $\Phi(x)$ such that for any $a \in M$ as above, there is K , $a \in K$, $K \upharpoonright a = M \upharpoonright a$ and $K \models \Phi(2^{2^a})$ and $M \models \neg \Phi(2^{2^a})$ or $K \models \neg \Phi(2^{2^a})$ and $M \models \Phi(2^{2^a})$. He has also shown that for any $a \in M$ as above, there is K such that $a \in K$, $M \upharpoonright a = K \upharpoonright a$ and

$$K \models \exists d < 2^{2^{2^a}} \text{Prf}_{\text{PA}}(d, 0 = 1).$$

Recently Solovay (unpublished manuscript) showed that there is a K such that even

$$K \models \exists d < 2^{2^a} \text{Prf}_{\text{PA}}(d, 0 = 1).$$

The main result of this paper shows that for any nonstandard number $c \in M \upharpoonright a$ there is a model K which agrees with M up to a and

$$K \models \exists d < 2^{a^c} \text{Prf}_{\text{PA}}(d, 0 = 1).$$

This improves the results of Hájek [2] and Solovay (unpublished manuscript). The result is derived from the bounds to the length of proofs of “finitistic consistency statements” in Pudlák [4, 5], which were initiated by Smoryński [7]. It should be stressed that here we have talked about the *Gödel numbers* of the proofs. Later we shall use also the *length of proofs* which means that the bounds will be less by one exponential.

These questions are closely connected with some problems in complexity theory (see [3]). Paris and Dimitracopoulos [3] proved for instance that if $M \upharpoonright a$ determines $M \upharpoonright a^{\log a}$ then $\text{PH} \neq \text{PSPACE}$. We shall prove two statements of this kind.

1. Preliminaries and Definitions

The proof of our result is based on an estimate proved in [4]. We shall recall this estimate and also some other definitions and facts.

We shall identify syntactical expressions with 0–1 sequences or the positive integers which these sequences define in dyadic notation. Thus the *length*, denoted by $|\dots|$, of an expression is the number of bits. *Proofs* are sequence of formulas (not trees). The n -th *numeral*, denoted by \underline{n} , is the term defined inductively by $\underline{0} = 0$, $\underline{1} = 1$ and, for $n \geq 1$,

$$\underline{2n} = (1 + 1) \cdot n, \quad \underline{2n+1} = ((1 + 1) \cdot n) + 1$$

Thus $|\underline{n}| = O(\log n)$. All logarithms are to the base 2. The relation

$$T \upharpoonright^m A$$

denotes that there is a proof d of A in theory T with the length $|d| \leq m$. For a fixed recursive axiomatization of T this relation is recursive. However such information is not sufficient for deriving any bounds. Therefore we shall assume more: that T as a set of axioms is in NP. Then $T \upharpoonright^m A$ can be decided in nondeterministic time $p(m, |A|)$ for some polynomial p . Given an NP axiomatization of T , let $\text{Con}_T(a)$ be the “natural” formalization of:

$$\text{not } T \upharpoonright^a (0 = 1).$$

(For a more precise meaning of “natural” see [4].) Con_T denotes $\forall x \text{Con}_T(x)$.

We shall fix a sufficiently strong fragment of arithmetic T_0 and consider only theories T in the language of arithmetic L which contain T_0 . It is convenient to take $I\Delta_0 + \text{Exp}$ as T_0 (i.e., Peano arithmetic with induction restricted to bounded formulas plus an axiom saying that exponentiation is a total function), because in this theory the formalization of syntax is easy. However we could use a weaker theory. The exponentiation x^y is usually not included in the language of arithmetic. In order to simplify the exposition we shall include it in our language L , thus $L = \{0, 1, +, \cdot, \leq, x^y\}$.

Lemma 1.1. *There exists a polynomial $p(x)$ such that for any true statement A of the form $\underline{n} = \underline{m}$, $\underline{n} \neq \underline{m}$, $\underline{n} + \underline{m} = \underline{k}$, $\underline{n} + \underline{m} \neq \underline{k}$, $\underline{n} \cdot \underline{m} = \underline{k}$, $\underline{n} \cdot \underline{m} \neq \underline{k}$, $\underline{n} \leq \underline{m}$, $\underline{n} \not\leq \underline{m}$, $\underline{n}^m = \underline{k}$ or $\underline{n}^m \neq \underline{k}$,*

$$T_0 \mid \frac{p(|A|)}{} A$$

Proof. It is easily seen that the usual polynomial time algorithms for computing with integers in binary notation can be transformed into such proofs. \square

Lemma 1.2. *For every NP axiomatized theory T , $T_0 \subseteq T$ there exists $\varepsilon > 0$, $k < \omega$ such that T_0 proves the formalization of the following statement:*

“For any n, m , if there is no proof of contradiction in T with length $\leq n^{m \cdot k}$ then

(*) not $T \mid \frac{n^{\varepsilon \cdot m}}{} \text{Con}_T(n^m)$.”

Proof. Theorem 3.6 of [4] and Lemma 1.1 above give (*) under the assumption that T is consistent. The proof actually shows that if a proof of $\text{Con}_T(n^m)$ in T of length $\leq n^{\varepsilon \cdot m}$ is given (and if ε is small) then a proof of a contradiction in T can be constructed from it. (This is quite similar to the proof of the second incompleteness theorem.) It is a matter of a routine computation to show that such a proof of a contradiction has length polynomial in n^m . The metatheory that we need here is very weak: we need the Diagonalization Lemma and elementary transformations of proofs. Hence this proof can be formalized in T_0 . \square

Further we shall use the following notation. Given a model M we identify its elements with their names. For $b \in M$, $\text{Diag}(M \upharpoonright b)$ is the set of sentences of the form $e = f$, $e \neq f$, $e + f = g$, $e + f \neq g$, $e \cdot f = g$, $e \cdot f \neq g$, $e \leq f$, $e \not\leq f$, $e^f = g$, and $e^f \neq g$, for $e, f, g \leq b$, which are true in $M \upharpoonright b$; $\text{Th}(M \upharpoonright b)$ are all sentences in the language of arithmetic which are true in $M \upharpoonright b$ and where the operations are treated as ternary relations.

2. The Results

Theorem 2.1. *Let $T \supseteq T_0$ be a recursively axiomatizable theory, let $\text{Con}_T(x)$ be a formalization of the consistency of T up to the length x corresponding to some NP-enumeration of the axioms of T . Let M be a nonstandard countable model of T and c, a , nonstandard elements of M , $M \models c \leq a$. Then there exists a countable model K of T such that $a \in K$ and*

- (i) $M \upharpoonright a = K \upharpoonright a$,
- (ii) $M \upharpoonright 2^a \subseteq K$,
- (iii) $K \models \neg \text{Con}_T(a^c)$.

Recall that $\neg \text{Con}_T(a^c)$ implies $\exists d < 2^{a^c}$, $\text{Prf}_T(d, 0 = 1)$.

Proof. Fix M, a, c satisfying the assumptions of the theorem. If $M \models \neg \text{Con}_T(a^c)$ for some $k < \omega$ then we can put $K = M$. So we can assume the opposite. Then choosing c possibly smaller we can meet the assumption of Lemma 1.2: in M there is no proof of contradiction in T with length $\leq a^{c \cdot k_0}$, for suitable $k_0 < \omega$.

Let $L(2^a)$ be the language L augmented with all $e \in M \upharpoonright 2^a$. An $L(2^a)$ formula A can be turned into an M formal formula in the language L by translating each e as the e -th numeral. This translation will be denoted by $t(A)$. Observe that for each (standard) formula of $L(2^a)$, $|t(A)| \leq k \cdot a$, for some $k < \omega$; the same is true for the translations of (standard) proofs.

Define an $L(2^a)$ theory S by

$$S := \{A \in L(2^a) \mid M \models T \upharpoonright^{a^k} t(A), \text{ for some } k < \omega\} + \neg \text{Con}_T(a^c).$$

Clearly $T \subseteq S$.

Claim 1. $\text{Diag} M(2^a) \subseteq S$.

This is an immediate corollary of Lemma 1.1.

Claim 2. S is consistent.

Proof of the claim. Suppose it is not. Then there is a proof d of

$$T + A_1 + \dots + A_n \vdash \text{Con}_T(a^c),$$

where

$$(*) \quad M \models [T \upharpoonright^{a^k} t(A_1), \dots, t(A_n)],$$

for some $k < \omega$. If we replace the constants in d by the corresponding numerals, we obtain an M -proof of

$$T + t(A_1) + \dots + t(A_n) \vdash \text{Con}_T(a^c)$$

of length $\leq \ell \cdot a$ for some $\ell < \omega$. Combining this M -proof with the M -proofs of $t(A_1), \dots, t(A_n)$ of $(*)$ we obtain

$$M \models [T \upharpoonright^{a^m} \text{Con}_T(a^c)],$$

for some $m < \omega$. But by Lemma 1.2 we have

$$M \models \neg [T \upharpoonright^{a^{e \cdot c}} \text{Con}_T(a^c)]$$

which is a contradiction, since

$$M \models a^m \leq a^{e \cdot c}$$

We shall use the Omitting Type Theorem [1] to construct the required model K . Let $\tau(x)$ be the type in $L(2^a)$ defined by

$$\tau(x) := \{x \leq a\} + \{x \neq e \mid e \in M \upharpoonright a\}.$$

Claim 3. S locally omits type $\tau(x)$.

Proof of the claim: Assume that for some $L(2^a)$ formula $A(x)$:

$$S \vdash A(x) \rightarrow x \leq a,$$

$$S \vdash A(x) \rightarrow x \neq e, \text{ for } e \leq a.$$

By the same argument as in Claim 2, we have

$$M \models [T \upharpoonright^{a^k} t(A(x)) \rightarrow x \leq a],$$

$$(**) \quad M \models [T \upharpoonright^{a^{k_e}} t(A(x)) \rightarrow x \neq e], \text{ for } e \leq a,$$

where $k, k_0, k_1, \dots, k_a < \omega$. Using induction in M we can take minimal $g \in M$ such that $(**)$ holds with k_e replaced by g for all $e \leq a$. Since all $k_e, e \leq a$, are standard, g is standard too (by underspill). Now we can combine the $a+2$ shortest proofs of $t(A(x)) \rightarrow x \leq a$ and $t(A(x)) \rightarrow x \neq e$ in M and we obtain

$$M \models [T \upharpoonright^{a^m} t(A(x)) \rightarrow (x \leq a \wedge x \neq 0 \wedge \dots \wedge x \neq a)]$$

for some $m < \omega$. On the other hand we have also

$$M \models [T \upharpoonright^{a^\ell} \neg(x \leq a \wedge x \neq 0 \wedge \dots \wedge x \neq a)],$$

for some $\ell < \omega$, thus for some $r < \omega$,

$$M \models [T \upharpoonright^{a^r} \neg t(A(x))],$$

i.e., $\forall x \neg A(x) \in S$. Hence $S + \exists x A(x)$ is inconsistent. This proves the claim.

Now by the Omitting Type Theorem there exists a model $K \models S$ which omits $\tau(x)$. Since $\text{Diag}(M \upharpoonright^{2^a}) \subseteq S$ we can embed $M \upharpoonright^{2^a}$ into K and all the properties of K are clearly satisfied. \square

Remark. Observe that Theorem 2.1 implies that, in fact, we can have $M \upharpoonright^{2^{a^k}} \subseteq K$, for all $k < \omega$ simultaneously. To get this apply the theorem to c_1 and a^{c_2} instead of c and a , for some nonstandard c_1, c_2 s.t. $c_1 \cdot c_2 \leq c$.

The following theorem is a strengthening of a statement proved by Solovay (unpublished manuscript). He proved this result with the assumptions $2^{2^r} < a$ and $c = a \cdot \log(a)^{-1}$. Recall that $I\Sigma_i$ is the fragment of Peano arithmetic PA obtained by restricting the schema of induction to Σ_i formulas.

Theorem 2.2. *Let M be a countable nonstandard model of PA, let r, c, a be nonstandard elements of M , $M \models r, c \leq a$, and suppose $M \models \text{Con}_{I\Sigma_r}$. Then there exists a model K of PA such that $a \in K$ and*

- (i) $M \upharpoonright^a = K \upharpoonright^a$,
- (ii) $M \upharpoonright^{2^a} \subseteq K$,
- (iii) $K \models \neg \text{Con}_{I\Sigma_r}(a^c)$, (i.e., $K \models \exists d < 2^{a^c} \text{Prf}_{I\Sigma_r}(d, 0 = 1)$), and $K \models \text{Con}_{I\Sigma_{r-1}}$.

We need the following lemmas.

Lemma 2.3. *There exists $\varepsilon > 0$ such that for all $k, i \geq 1$ and $n \geq k$, i , it is not true that*

$$I\Sigma_i \upharpoonright^{n^{\varepsilon \cdot k}} \text{Con}_{I\Sigma_i}(n^k).$$

Proof-sketch. Theorem 3.6 of [4] does not say that there is a single $\varepsilon > 0$ for all $I\Sigma_i$. However in [5] the proof of this theorem was analyzed more closely. It was shown there (Lemma 2.1) that a lower bound to the length of proof of $\text{Con}_T(n)$ in T is a linear function of $f^{-1}(n)$ where $f(n)$ is any polynomial time computable function such that

$$(*) \quad T \upharpoonright^n A \text{ implies } T \upharpoonright^{f(n)} [T \upharpoonright^n A].$$

Thus it is sufficient to show that there is a *uniform polynomial* bound to (*) for $T = I\Sigma_i$, $i=0, 1, \dots$. This can be done easily, for example using the techniques of [5]. \square

Lemma 2.4. *There exists $k < \omega$ such that for all $i \geq 0$*

$$I\Sigma_{i+1} \vdash^{(i+2)^k} \text{Con}_{I\Sigma_i}.$$

Proof. The fact that $I\Sigma_{i+1}$ proves $\text{Con}_{I\Sigma_i}$ is well-known. Since we need such a proof with polynomial length we have to analyze it in more detail. First recall that there exists a finite set of axioms P^- such that each $I\Sigma_i$ can be axiomatized by P^- plus a *single* instance of the induction for Σ_i formulas. This instance of induction is the induction for a universal Σ_i formula. A universal Σ_i formula can be obtained from a universal Σ_0 formula $\Theta_0(x, y)$ by defining

$$\Theta_i(x, y) \equiv \exists z_1 \forall z_2 \dots Qz_i \Theta_0(x, \langle y, \langle z_1, \langle z_2, \dots \rangle \rangle \rangle),$$

where $\langle \dots, \dots \rangle$ is the usual pairing function. Θ_0 is Δ_1 over $I\Sigma_1$ and Θ_i are Σ_i for $i \geq 1$. All this is provable in $I\Sigma_1$. Moreover we can show in $I\Sigma_1$ that in a Gentzen style system the induction axiom can be replaced by the induction rule for $\Theta_i(x, y)$ and that free-cuts can be eliminated from the proofs. If we have such a free-cut-free proof of a contradiction, then all formulas in it are substitution instances of subformulas of P^- and Θ_i . Hence using the definition of the value of arithmetical terms (available in $I\Sigma_1$) we can write a formula $\text{Sat}_i(x, y)$ such that

- (i) it has length polynomial in i ,
- (ii) the Tarski conditions for satisfiability for substitution instances of subformulas of P^- and Θ_i can be proved by a proof of length polynomial in i ,
- (iii) Sat_i is Σ_i .

Condition (ii) implies that Sat_i preserves logical rules; condition (iii) implies that Sat_i preserves the induction rule since we assume $I\Sigma_{i+1}$.

Now let a free-cut-free proof d be given and suppose that it uses only induction rule for Θ_i . The statement that a sequent in the proof is true for every interpretation of free variables is expressed by a Π_{i+1} formula. Since $I\Sigma_{i+1}$ proves III_{i+1} we can use the induction on the depth of the proof to show that every sequent is true.

The proof of $\text{Con}_{I\Sigma_i}$ in $I\Sigma_{i+1}$ described above will have three main parts. The first part is the transition from the usual system $I\Sigma_i$ into the free-cut-free system with the induction rule only for Θ_i . This can be done in $I\Sigma_1$ for all i at once, i.e., $x \rightarrow \Theta_x$ is a function definable in $I\Sigma_1$ and $I\Sigma_1$ proves the sentence “for all x , $\neg \text{Con}_{I\Sigma_x}$ implies that there is a free-cut-free proof of $0=1$ with induction rule applied to Θ_x only”. Thus this part has a constant length. The second part consists of the proof that Sat_i (for this particular i) preserves the rules; it has polynomial length by (i) and (ii). The third part is the proof that all the sequents in d are true. This part is uniform for all i in the sense that only the formula Sat_i is changing in it.

As Sat_i has polynomial length this part has also polynomial length. \square

Proof of Theorem 2.2. The proof is almost identical with the proof of Theorem 2.1, thus we state only the differences. In the proof we use $I\Sigma_r$ instead of T . Hence we define:

$$S := \{A \in L(2^a) \mid M \models [I\Sigma_r]^{a^k} t(A), \text{ for some } k < \omega\} + \neg \text{Con}_{I\Sigma_r}(a^r).$$

Since r is nonstandard, $\text{PA} \subseteq S$. By the formalization of Lemma 2.4 in PA and since $r \leq a$ we have also $\text{Con}_{T, r-1} \in S$. The consistency of S is proved using the formalization of Lemma 2.3. We needed this stronger lemma in order to obtain ε standard also for r nonstandard. The rest of the proof does not require any comments. \square

We are not able to show that the bound to the length of the proof of contradiction in K in Theorem 2.1 is optimal. However we shall prove that it is optimal under an assumption on complexity classes which apparently has not been ruled out.

In the sequel we assume that T, M, c and a satisfy the assumptions of Theorem 2.1. By these assumptions $\text{Con}_T(n^k)$ is a $\text{coNTIME}(2^{n^k})$ predicate, for some $k \geq m, m, k$ constants.

Proposition 2.5. *If T proves that $\text{coNTIME}(2^{n^k}) \subseteq \text{NTIME}(2^{n^\ell})$ for some $k, \ell < \omega$, and $M \models \text{Con}_T$, then there is no model $K \models T$ such that $M \upharpoonright a = K \upharpoonright a$, $M \upharpoonright 2^{a^\ell} \subseteq K$ and $K \models \neg \text{Con}_T(a^m)$.*

Proof. The assumption implies that $\text{Con}_T(x^m)$ can be expressed in T as an NP formula $A(2^{x^\ell})$. But for such a formula we have

$$M \models A(2^{a^\ell}) \text{ implies } K \models A(2^{a^\ell})$$

whenever $M \upharpoonright a = K \upharpoonright a$ and $M \upharpoonright 2^{a^\ell} \subseteq K$. \square

If $\text{coNTIME}(2^{n^k}) \subseteq \text{NTIME}(2^{n^\ell})$ is true, we can add it to T_0 as an axiom and thus we obtain a theory for which the bound cannot be improved. Under a different assumption the bound *can* be improved.

Proposition 2.6. *If T proves that $\Delta_0 \subseteq \text{NP}$ then there exists a model K of T such that $a \in K$ and*

- (i) $M \upharpoonright a = K \upharpoonright a$,
- (ii) $K \models \neg \text{Con}_T(\log(a)^c)$.

Proof. Take $c_1, c_2 \leq \log(a)$ nonstandard such that $c_1 \cdot c_2 \leq c$. Apply Theorem 2.1 to c_1 and $\log(a)^{c_2}$. Thus we obtain a model $K \models T$ such that

- (i) $M \upharpoonright \log(a)^{c_2} = K \upharpoonright \log(a)^{c_2}$,
- (ii) $M \upharpoonright 2^{\log(a)^{c_2}} \subseteq K$,
- (iii) $K \models \neg \text{Con}_T(\log(a)^{c_1 \cdot c_2})$ hence $K \models \neg \text{Con}_T(\log(a)^c)$.

The condition (ii) implies that

$$\begin{aligned} M \models A(a) \text{ implies } K \models A(a), \text{ if } A(x) \text{ is NP,} \\ K \models B(a) \text{ implies } M \models B(a), \text{ if } B(x) \text{ is coNP.} \end{aligned}$$

The assumption $\Delta_0 \subseteq \text{NP}$ implies that each Δ_0 -formula $C(y)$ can be expressed as both an NP formula and a coNP formula. Thus

$$\text{Th}(M \upharpoonright a) = \text{Th}(K \upharpoonright a).$$

Since $M \upharpoonright a$ and $K \upharpoonright a$ share a common initial segment of a nonstandard length they have the same standard system

$$\text{SSy}(M \upharpoonright a) = \text{SSy}(K \upharpoonright a).$$

It follows by a well-known theorem (see e.g., [6]) that since they are recursively saturated, they are isomorphic. Thus we can identify these initial segments of M and K . \square

Observe that the assumption $\text{NP} = \text{coNP}$ implies the assumptions of Propositions 2.5 and 2.6.

It is natural to pose the following question: is it possible to improve the bound to a proof of contradiction in K in Theorem 2.1 if the requirement $M \upharpoonright 2^a \subseteq K$ is dropped?

References

1. Chang, C.C., Keisler, H.J.: *Model theory*. Amsterdam: North-Holland 1977
2. Hájek, P.: On a new notion of partial conservativity. In: Boerger, E., Oberschelp, W., Richter, M.M., Schinzel, B., Thomas, W. (eds.) *Logic Colloquium 83*, vol. 2, pp. 217–232. Berlin Heidelberg New York: Springer 1983
3. Paris, J.B., Dimitracopoulos, C.: Truth definitions for Δ_0 formulae, *Logic and Algorithmic (An Intr. Symp. held in honour of E. Specker, Zürich 1980)*, Monogr. No. 30 de L'Enseignement Mathématique, Genève (1982), pp. 317–329
4. Pudlák, P.: On the length of proofs of finitistic consistency statements in first order theories. In: Paris, J.B., Wilkie, A.J., Wilmers, G.M. (eds.) *Logic Colloquium 84*, pp. 165–196. Amsterdam: North-Holland 1986
5. Pudlák, P.: Improved bounds to the length of proofs of finitistic consistency statements, *Logic and Combinatorics*. In: Simpson, S.G. (ed.) *Contemporary Mathematics*, vol. 65, pp. 309–332 (1987)
6. Smoryński, C.: Lectures on nonstandard models of arithmetic. In: *Logic Colloquium 82*, pp. 1–70. Lolli, G., Longo, G., Marcja, A. (eds.) Amsterdam: North-Holland 1984
7. Smoryński, C.A.: Nonstandard models and related developments. In: Havington, L.A., Morley, M., Scechov, A., Simpson, S.G. (eds.), *Harvey Friedman's Research on the Foundations of Mathematics*. Amsterdam: North-Holland 1985

Received May 30, 1988/in revised form November 14, 1988