

# Partial Collapses of the $\Sigma_1$ Complexity Hierarchy in Models for Fragments of Bounded Arithmetic

Zofia Adamowicz

Institute of Mathematics, Polish Academy of Sciences

Śniadeckich 8, 00-950 Warszawa, Poland

z.adamowicz@impan.gov.pl

Leszek Aleksander Kołodziejczyk \*

Institute of Mathematics, Warsaw University

Banacha 2, 02-097 Warszawa, Poland

lak@mimuw.edu.pl

June 27, 2006

## Abstract

For any  $n$ , we construct a model of  $T_2^n + \neg exp$  in which each  $\exists s\Pi_{n+1}^b$  formula is equivalent to an  $\exists\Pi_n^b$  formula.

The central question in bounded arithmetic is whether there exists a model of the theory  $S_2$  in which the bounded formula hierarchy (or, in other words, the polynomial hierarchy) is infinite. However, it is also natural to ask whether there actually exists a model for  $S_2$ , or at least for a reasonably strong fragment of this theory, where the bounded formula hierarchy *collapses*. The answer to this question has remained equally elusive.

A somehow related, though perhaps more tractable problem is whether there is a model of bounded arithmetic *not satisfying exp* (the totality of the

---

\*Part of this work was carried out while the author was a Foundation for Polish Science (Fundacja na rzecz Nauki Polskiej) scholar.

exponential function) in which the  $\Sigma_1$  formula hierarchy collapses, that is, every  $\Sigma_1$  formula is equivalent to an  $\exists\Pi_m^b$  formula for some fixed  $m$ .

Note that by [GD82], any model which satisfies *exp* also satisfies Matijasevič's theorem, so in any such model a  $\Sigma_1$  formula is equivalent to a purely existential formula. Even without referring to Matijasevič's theorem, one can easily show that in any model of *exp*, the  $\Sigma_1$  hierarchy collapses to  $\exists\Delta_1^b$ : given a  $\Sigma_1$  formula  $\varphi = \exists x \psi$  with  $\psi$  bounded, we may existentially quantify an object so large in comparison to the arguments of  $\varphi$  and the hypothetical witness for  $\exists x$  that all the bounded quantifiers in  $\psi$  become sharply bounded (such an object may easily be  $\Delta_1^b$ -defined using the arguments and witness as parameters).

For models without *exp*, the situation is much less clear. In such “short” models, the difference between unbounded and bounded quantifiers is smaller than in structures satisfying *exp*, in the sense that even an unbounded quantifier does not have access to elements which are enormously large with respect to the parameters of a given formula. Thus, the method of collapsing  $\Sigma_1$  to  $\exists\Delta_1^b$  outlined in the previous paragraph breaks down. As a matter of fact, to the best of our knowledge there is no construction of a model of  $S_2 + \neg\text{exp}$ , nor even  $S_2^n$  or  $T_2^n$  ( $+ \neg\text{exp}$ ) for some  $n$ , in which the  $\Sigma_1$  hierarchy would be known to collapse.

In the present paper, we show that any finitely axiomatizable fragment of  $S_2$  has a “short” model in which a *partial* collapse of the  $\Sigma_1$  hierarchy occurs. Specifically, we show that for any  $n$ , there is a model  $\mathbf{M}$  of  $T_2^n$  with the following two properties. Firstly, there is an element  $a \in \mathbf{M}$  such that the standard iterations of  $\#$  on  $a$  are cofinal in  $\mathbf{M}$  (so, in particular,  $\mathbf{M} \models \neg\text{exp}$ ). Secondly, the model  $\mathbf{M}$  satisfies  $\exists s\Pi_{n+1}^b \equiv \exists\Pi_n^b$ , i.e. each  $\exists s\Pi_{n+1}^b$  formula is equivalent to a  $\exists\Pi_n^b$  formula. Unfortunately, we cannot tell whether the collapse extends further — it is conceivable that for  $m > n + 1$ , the  $\exists s\Pi_m^b$  formulae will be strictly more expressive than  $\exists\Pi_n^b$ .

Our proof is based on classical logical methods rather than the computer science-inspired techniques common in the research on bounded arithmetic. We use the notion of  $\Sigma_{n+1}^b$ -maximal model, i.e. a model which, in a sense, satisfies as many  $s\Sigma_{n+1}^b$  formulae as possible. We show that “short”  $\Sigma_{n+1}^b$ -maximal models for  $T_2^n$  exist, and that in each such model, any  $s\Sigma_{n+1}^b$  formula  $\varphi(x)$  is equivalent to an infinite conjunction  $\bigwedge_{k \in \mathbb{N}} \text{Cons}_k(\varphi(x))$  of certain  $\Pi_{n+1}^b$  consistency statements. The statements are uniform enough for the infinite conjunction to be equivalent to a single  $\forall\Sigma_n^b$  formula. This gives  $s\Sigma_{n+1}^b \subseteq \forall\Sigma_n^b$ . Taking negations, we get  $s\Pi_{n+1}^b \subseteq \exists\Pi_n^b$ , and our result follows.

The paper is divided into four sections. Section 1 is preliminary. Section 2 describes the construction of  $\Sigma_{n+1}^b$ -maximal models. In section 3, we introduce our consistency statements, and in the last section we complete the proof of the main result.

## 1 Preliminaries

We assume familiarity with the basic notions and results of bounded arithmetic, which may be found in [HP93] or [Kra95]. Throughout the paper, we use the letter  $n$  to denote an arbitrary fixed natural number  $\geq 1$ , needed to specify the bounded arithmetic theories and formula classes we deal with (e.g.  $T_2^n$ ,  $\Sigma_n^b$ ,  $\Sigma_{n+1}^b$ , etc.).

Recall the difference between general  $\Sigma_n^b$  and strict  $\Sigma_n^b$ , or  $s\Sigma_n^b$ . The class  $\Sigma_n^b$  is defined as the closure of  $\Pi_{n-1}^b$  under connectives, sharply bounded quantification, and bounded existential quantification, whereas  $s\Sigma_n^b$  is the prenex version of  $\Sigma_n^b$ , i.e. it consists of formulae of the form

$$\exists x_1 \leq t_1 \forall x_2 \leq t_2 \dots Qx_n \leq t_n \psi,$$

where  $\psi$  is sharply bounded.  $S_2^n$ , and thus also  $T_2^n$ , proves that every  $\Sigma_n^b$  formula is equivalent to an  $s\Sigma_n^b$  formula, but it is unknown whether this holds also for weaker bounded arithmetic theories.

For any class of formulae  $\Gamma$ , the class  $\exists\Gamma$  consists of formulae from  $\Gamma$  preceded by a tuple of existential quantifiers. In  $\exists^b\Gamma$ , these initial existential quantifiers are additionally required to be bounded.  $\forall\Gamma$  and  $\forall^b\Gamma$  are defined analogously.

We want to consider fragments of diagrams of models, so given a model  $\mathbf{M}$ , we work not just with the usual bounded arithmetic language  $L_2$ , but also with its extension  $L(\mathbf{M})$  obtained by adding a constant symbol  $\underline{d}$  for every element  $d$  of  $\mathbf{M}$ . We will be particularly interested in the positive part of the  $\Pi_n^b$  diagram of  $\mathbf{M}$ , i.e.  $\text{Th}_{\Pi_n^b}(\mathbf{M}_{L(\mathbf{M})})$ , where  $\mathbf{M}_{L(\mathbf{M})}$  is the expansion of  $\mathbf{M}$  to the language  $L(\mathbf{M})$ .

We need to encode the extended language  $L(\mathbf{M})$  in arithmetic. For simplicity, we may let the first few odd numbers be the Gödel numbers of symbols of  $L_2$ , assign the number  $2d$  to the constant symbol  $\underline{d}$ , and treat formulae as sequences of symbols.

To code sequences, we can use any standard feasible sequence-coding method available in bounded arithmetic. If  $s$  is a sequence,  $\text{lh}(s)$  stands for

the length of  $s$  and  $(s)_i$  denotes the  $i$ -th element of  $s$ , for  $i = 0, \dots, \text{lh}(s) - 1$ . A “bar”, as in  $\bar{x}$  or  $\bar{d}$ , always denotes a tuple, and  $\underline{\bar{d}}$  denotes the tuple of constants for elements of  $\bar{d}$ .

## 2 $\Sigma_{n+1}^b$ -maximal models

**Definition 2.1.** Let  $T$  be a theory containing  $S_2^1$ . A model  $\mathbf{M} \models T$  is called  $\Sigma_{n+1}^b$ -maximal w.r.t.  $T$  if for any  $\mathbf{M}' \models T$ ,  $\mathbf{M} \preceq_{\Sigma_n^b} \mathbf{M}'$  implies  $\mathbf{M} \preceq_{\Sigma_{n+1}^b} \mathbf{M}'$ .

The notion of  $\Sigma_{n+1}^b$ -maximality is a suitably modified version of the concept of *maximality* or *1-closedness* (see e.g. [Ada91] or [AB01]), which is, in turn, an arithmetical version of the general model-theoretical concept of *existential closure*. A related notion was also used in [Bec04].

It is quite easy to show that  $\Sigma_{n+1}^b$ -maximal models w.r.t.  $T_2^n$  exist. The proof is a rather standard iterative argument:

**Lemma 2.2.** *Let  $i \leq n$  and let  $\mathbf{M}$  be a countable model of  $T_2^i$ . There exists a model  $\mathbf{M}_+ \models T_2^i$  such that:*

- (a)  $\mathbf{M} \preceq_{\Sigma_n^b} \mathbf{M}_+$ ;
- (b)  $\mathbf{M}_+$  is  $\Sigma_{n+1}^b$ -maximal w.r.t.  $T_2^i$ ;
- (c)  $\mathbf{M}$  is cofinal in  $\mathbf{M}_+$ .

*Proof.* Let a countable  $\mathbf{M} \models T_2^i$  be given. Let  $t_0, t_1, \dots$  be an enumeration of all triples

$$\langle m, \langle l_1, \dots, l_r \rangle, \varphi \rangle,$$

where  $m, r, l_1, \dots, l_r \in \mathbb{N}$ ,  $\varphi$  is an  $s\Sigma_{n+1}^b$  formula and the number of free variables in  $\varphi$  is  $r$ . We may assume w.l.o.g. that  $t_k = \langle m, \langle l_1, \dots, l_r \rangle, \varphi \rangle$  implies  $m \leq k$ .

We will use this enumeration to construct a chain  $\mathbf{M}_0 \preceq_{\Sigma_n^b} \mathbf{M}_1 \preceq_{\Sigma_n^b} \dots$  of countable models of  $T_2^n$ .

We take  $\mathbf{M}_0 := \mathbf{M}$ . Assume we have defined  $\mathbf{M}_0, \dots, \mathbf{M}_k$  and enumerations  $\{a_l^j : l \in \mathbb{N}\}$  of  $\mathbf{M}_j$  for  $j \leq k$ . Consider the triple  $t_k = \langle m, \langle l_1, \dots, l_r \rangle, \varphi \rangle$ . If there is a  $\Sigma_n^b$ -elementary extension  $\mathbf{M}'$  of  $\mathbf{M}_k$  satisfying  $T_2^i + \varphi(a_{l_1}^m, \dots, a_{l_r}^m)$ , take such a (countable)  $\mathbf{M}'$  and let  $\mathbf{M}_{k+1}$  be the initial segment of  $\mathbf{M}'$  determined by  $\mathbf{M}_k$  (note that  $\mathbf{M}_{k+1}$  is also a  $\Sigma_n^b$ -elementary extension of  $\mathbf{M}_k$ ; moreover,  $\mathbf{M}_{k+1}$  satisfies  $\varphi(a_{l_1}^m, \dots, a_{l_r}^m)$ , since  $\varphi$  is a bounded formula, so the

witness for its initial existential quantifier is small enough to fit into  $\mathbf{M}_{k+1}$ ). Otherwise let  $\mathbf{M}_{k+1} := \mathbf{M}_k$ . Fix any enumeration  $\{a_l^{k+1} : l \in \mathbb{N}\}$  of  $\mathbf{M}_{k+1}$ .

Define  $\mathbf{M}_+$  as the union  $\bigcup_{k \in \mathbb{N}} \mathbf{M}_k$ . One may easily check that  $\mathbf{M}_+$  satisfies (a)-(c). In particular, to prove that  $\mathbf{M}_+ \models T_2^i$  (which is part of (b)), observe that  $\mathbf{M}_k \preceq_{\Sigma_n^b} \mathbf{M}_+$  for each  $k$ , so if there was a counterexample to  $\Sigma_i^b$  induction in  $\mathbf{M}$ , then by  $\Sigma_i^b$ -elementarity this counterexample would have already been contained in some  $\mathbf{M}_k$ .  $\square$

**Corollary 2.3.** *Let  $\mathbf{M}$  be any countable model of  $T_2^n$  containing an element  $a$  such that the standard iterations of  $\#$  on  $a$  are cofinal in  $\mathbf{M}$ . Then  $\mathbf{M}$  can be  $\Sigma_n^b$ -elementarily extended to a model  $\mathbf{M}_+$  which is  $\Sigma_{n+1}^b$ -maximal w.r.t.  $T_2^n$  and in which the standard iterations of  $\#$  on  $a$  remain cofinal.*

### 3 The consistency statements

In this section, we introduce the consistency statements whose conjunction is equivalent to an  $s\Sigma_{n+1}^b$  formula in an appropriately chosen model. We start by formulating a simple general observation on models of  $T_2^n$ .

**Lemma 3.1.** *Let  $\mathbf{M} \models T_2^n$ ,  $\bar{d} \in \mathbf{M}$ , and let  $\varphi(\bar{x})$  be an  $s\Sigma_{n+1}^b$  formula. Then  $\mathbf{M}$  can be  $\Sigma_n^b$ -elementarily extended to a model  $\mathbf{M}' \models T_2^n + \varphi(\bar{d})$  if and only if the theory  $T_2^n + \text{Th}_{\Pi_n^b}(\mathbf{M}_{L(\mathbf{M})}) + \varphi(\bar{d})$  is consistent.*

*Proof.* Both implications are straightforward. For the “only if” direction, simply observe that  $\mathbf{M}'$  (or strictly speaking, the expansion of  $\mathbf{M}'$  to  $L(\mathbf{M})$ ) must satisfy  $\text{Th}_{\Pi_n^b}(\mathbf{M}_{L(\mathbf{M})})$  as it is a  $\Sigma_n^b$ -elementary extension of  $\mathbf{M}$ .

For the “if” direction, consider any  $\mathbf{M}' \models T_2^n + \text{Th}_{\Pi_n^b}(\mathbf{M}_{L(\mathbf{M})}) + \varphi(\bar{d})$ .  $\mathbf{M}'$  is certainly an extension of  $\mathbf{M}$  (modulo the obvious embedding), so all that needs to be checked is that the extension is  $\Sigma_n^b$ -elementary. Any  $\Sigma_n^b$  formula satisfied in  $\mathbf{M}'$  is clearly satisfied in  $\mathbf{M}$ , since  $\mathbf{M}' \models \text{Th}_{\Pi_n^b}(\mathbf{M}_{L(\mathbf{M})})$ . On the other hand, by  $\Sigma_n^b$  replacement, every  $\Sigma_n^b$  formula is provably equivalent in  $T_2^n$  to an  $s\Sigma_n^b$ , and thus  $\exists^b \Pi_{n-1}^b$ , formula. Moreover, an  $\exists^b \Pi_{n-1}^b$  formula satisfied in  $\mathbf{M}$  will also be satisfied in  $\mathbf{M}'$ , since  $\mathbf{M}' \models \text{Th}_{\Pi_{n-1}^b}(\mathbf{M}_{L(\mathbf{M})})$  and the witnesses for the initial existential quantifiers are present in  $\mathbf{M}'$ .  $\square$

**Corollary 3.2.** *Let  $\mathbf{M}$  be a  $\Sigma_{n+1}^b$ -maximal model of  $T_2^n$ ,  $\bar{d} \in \mathbf{M}$ , and let  $\varphi(\bar{x})$  be an  $s\Sigma_{n+1}^b$  formula. Then  $\mathbf{M} \models \varphi(\bar{d})$  if and only if the theory  $T_2^n + \text{Th}_{\Pi_n^b}(\mathbf{M}_{L(\mathbf{M})}) + \varphi(\bar{d})$  is consistent.*

In other words, in a  $\Sigma_{n+1}^b$ -maximal model the truth of an  $s\Sigma_{n+1}^b$  formula is reduced to consistency with the positive part of the  $\Pi_n^b$  diagram. It remains to show that in a model of an appropriate form, this consistency property can be expressed by a  $\forall\Sigma_n^b$  formula.

In the remaining part of this section, we assume that the model  $\mathbf{M}$  is as given by Corollary 2.3, i.e.  $\mathbf{M}$  is a  $\Sigma_{n+1}^b$ -maximal model of  $T_2^n$  containing an element  $a$  such that the standard iterations of  $\#$  on  $a$  are cofinal in  $\mathbf{M}$ .

Let  $\psi(\bar{x})$  be any formula. For any standard  $k$ , define the formula  $\text{Cons}_k(\psi(\bar{x}))$  as:

There is no proof of  $\neg\psi(\bar{x})$  from  $\text{Th}_{\Pi_n^b}(\mathbf{M}_{L(\mathbf{M})})$  containing at most  $|k|$  symbols and not containing constants for numbers greater than  $2^{|a|^k}$ .

Thus, each  $\text{Cons}_k$  expresses “partial” consistency with the  $\Pi_n^b$  diagram of  $\mathbf{M}$  (parametrized by  $k$ ). For our purposes, it is important to calculate the quantifier complexity of  $\text{Cons}_k$ .

**Lemma 3.3.** *Let  $\psi(\bar{x})$  be any formula.*

- (a) *For any  $k$ ,  $\text{Cons}_k(\psi(\bar{x}))$  can be formulated as a (strict)  $\Pi_{n+1}^b$  formula with  $a$  as an additional parameter;*
- (b)  *$\bigwedge_{k \in \mathbb{N}} \text{Cons}_k(\psi(\bar{x}))$  can be formulated as a  $\forall\Sigma_n^b$  formula with  $a$  as an additional parameter.*

*Proof.*  $\text{Cons}_k(\psi(\bar{x}))$  is:

$$\begin{aligned} & \forall s \left[ \text{“}s \text{ is a sequence of formulae”} \ \& \ \sum_{i < \text{lh}(s)} \text{lh}((s)_i) \leq |k| \right. \\ & \ \& \ \text{“no } (s)_i \text{ contains a constant for a number greater than } 2^{|a|^k}\text{”} \\ & \ \& \ \forall i < \text{lh}(s) \left( (s)_i \in T_2^n \vee \text{“}(s)_i \text{ is a true } \Pi_n^b \text{ formula”} \right) \\ & \ \vee \ \text{“}(s)_i \text{ is derived from previous elements of } s \text{ by an inference rule”} \\ & \ \Rightarrow (s)_{\text{lh}(s)-1} \neq \ulcorner \psi(\bar{x}) \urcorner \end{aligned}$$

Syntactic properties of formulae and proofs are  $\Delta_1^b$ , provably in  $S_2^1$ . Thus, in order to show that the formula above can be written in  $s\Pi_{n+1}^b$  form with  $a$  as a parameter, we need only to check two things: firstly, that the universal quantifier  $\forall s$  can be bounded, and secondly, that “ $(s)_i$  is a true  $\Pi_n^b$  formula” can be expressed in  $\Pi_n^b$ .

The quantifier  $\forall s$  refers only to sequences of formulae which together contain no more than  $|k|$  symbols and which do not contain constants for numbers greater than  $2^{|a|^k}$ . Each potential  $s$  is a sequence of length at most  $|k|$ . Moreover, each formula in a potential  $s$  is a sequence of length at most  $k$  whose elements are all bounded by  $2^{|a|^{k+1}}$  (the Gödel number of the constant for  $2^{|a|^k}$ ). Thus, each element of  $s$  can be bounded by roughly  $2^{|k| \cdot (|a|^k + 1)}$ , and  $s$  itself can be bounded by roughly  $2^{|k|^2 \cdot (|a|^k + 1)}$ , which can obviously be expressed by a term in  $a$ .

To state “ $(s)_i$  is a true  $\Pi_n^b$  formula”, we need to use the universal formula for  $\Pi_n^b$  formulae, available already in  $S_2^1$ .<sup>1</sup> It is a  $\Pi_n^b$  formula with an additional parameter, which depends on the size of the arguments and whose only role is to bound all the quantifiers in the universal formula. It is known that to determine the truth value of formulae of length smaller than  $|l|$  for elements smaller than  $b$ , this additional parameter may be set to  $2^{|b|^l}$ . Since we are only interested in the truth of formulae of length at most  $|k|$  for numbers not exceeding  $2^{|a|^k}$ , we may set the parameter to  $2^{|a|^{k^2}}$ . This completes the proof of (a).

For (b), formulate the infinite conjunction as:

$$\forall b \forall c \forall k [(b = 2^{|a|^k} \ \& \ c = 2^{|a|^{k^2+1}}) \Rightarrow \text{Cons}_k(\psi(\bar{x}))],$$

where  $\text{Cons}_k$  is as above, but with the  $2^{|a|^k}$  bound on the size of constant symbols replaced by  $b$ , and with the  $2^{|k|^2 \cdot (|a|^k + 1)}$  bound for the  $\forall s$  quantifier and the  $2^{|a|^{k^2}}$  bounding parameter in the  $\Pi_n^b$  universal formula replaced by  $c$ . By our assumption on  $\mathbf{M}$ , the elements  $2^{|a|^{k^2+1}}$  for standard  $k$  are cofinal in  $\mathbf{M}$ , so the conjunction will range exactly over  $k \in \mathbb{N}$ , as required. Altogether, the conjunction is  $\forall s \Pi_{n+1}^b$  and thus  $\forall \Sigma_n^b$ .  $\square$

## 4 The main result

We are now ready to state a theorem which will yield our main result (Theorem 4.2) as a simple corollary.

**Theorem 4.1.** *Let  $\mathbf{M}$  be a  $\Sigma_{n+1}^b$ -maximal model of  $T_2^n$  containing some  $a$  such that the standard iterations of  $\#$  on  $a$  are cofinal in  $\mathbf{M}$ . For any strict*

---

<sup>1</sup>Strictly speaking, in  $S_2^1$  the universal formula works only for  $s\Pi_n^b$  formulae, but it works for all  $\Pi_n^b$  formulae once  $\Sigma_n^b$  replacement is available.

$\Sigma_{n+1}^b$  formula  $\psi(\bar{x})$ ,

$$\mathbf{M} \models \psi(\bar{x}) \equiv \bigwedge_{k \in \mathbb{N}} \text{Cons}_k(\psi(\bar{x})).$$

*Proof.* Just note that for any  $\bar{d} \in \mathbf{M}$ ,  $\bigwedge_{k \in \mathbb{N}} \text{Cons}_k(\psi(\bar{d}))$  is equivalent to the consistency of  $\psi(\bar{d})$  with the positive part of the  $\Pi_n^b$  diagram of  $\mathbf{M}$ . Then apply Corollary 3.2.  $\square$

**Theorem 4.2.** *There exists a model of  $T_2^n$  in which every  $\exists s\Pi_{n+1}^b$  formula is equivalent to an  $\exists\Pi_n^b$  formula (with no additional parameters).*

*Proof.* It is enough to show that there is a model of  $T_2^n$  in which every  $s\Sigma_{n+1}^b$  formula is equivalent to a  $\forall\Sigma_n^b$  formula (with the same parameters).

Let  $\mathbf{M}_0$  be any countable model of  $T_2^n$  which contains a proof of the inconsistency of  $PA$ . Let  $a$  be the smallest such proof. Consider the submodel of  $\mathbf{M}_0$  determined by the cut  $\#^{\mathbb{N}}(a)$  and extend it  $\Sigma_n^b$ -elementarily to a  $\Sigma_{n+1}^b$ -maximal model as in Corollary 2.3. Let  $\mathbf{M}$  be this maximal model. Notice that in  $\mathbf{M}$ ,  $a$  is still the smallest inconsistency proof of  $PA$ , so it is  $\Pi_1^b$ -definable.

By Theorem 4.1, every  $s\Sigma_{n+1}^b$  formula  $\psi(\bar{x}, \bar{d})$ , where  $\bar{d} \in \mathbf{M}$  are the parameters, is equivalent in  $\mathbf{M}$  to  $\bigwedge_{k \in \mathbb{N}} \text{Cons}_k(\psi(\bar{x}, \bar{d}))$ . In Lemma 3.3, we showed that  $\bigwedge_{k \in \mathbb{N}} \text{Cons}_k(\psi(\bar{x}, \bar{d}))$  can be expressed as a  $\forall\Sigma_n^b$  formula with  $a$  as the only additional parameter. But  $a$  is  $\Pi_1^b$ -definable, so we can reformulate  $\bigwedge_{k \in \mathbb{N}} \text{Cons}_k(\psi(\bar{x}, \bar{d}))$  as a  $\forall\Sigma_n^b$  formula with no new parameters.  $\square$

**Remark.** In the proof above, the assumption that the initial model  $\mathbf{M}_0$  satisfies  $\neg\text{Cons}_{PA}$  was only needed to get an appropriately large  $\Pi_1^b$ -definable element in  $\mathbf{M}$  — which is, in turn, only needed to avoid introducing an additional parameter.

**Acknowledgement and remark.** The authors would like to thank the anonymous referee for pointing out a simpler proof of the main result. The original proof was based on the same strategy as the present one, but instead of the ordinary notion of consistency, it used a variant of consistency with respect to Herbrand proofs (i.e. proofs in the propositional calculus).

The formalization of the Herbrand notion of consistency is somewhat technical and tedious, so its elimination allowed us to make the paper shorter and hopefully more readable. On the other hand, it should be noted that Herbrand consistency, like other notions of cut-free consistency, is, in general,

much better-behaved in weak arithmetic theories than ordinary Hilbert- or Gentzen-style consistency, and its use may lead to stronger results.

For example, one may apply the  $\forall\Sigma_{n+1}^b$  conservativity of  $S_2^{n+1}$  over  $T_2^n$  to show that  $\exists s\Pi_{n+1}^b \equiv \exists\Pi_n^b$  holds in *any countable*  $\Sigma_{n+1}^b$ -maximal model of  $T_2^n$ , and not just in models containing an element whose  $\#$ -closure is cofinal. The argument is quite similar in spirit to the one presented here, but it seems that the notion of consistency it involves has to be cut-free, since in order to invoke conservativity one must deal with proofs from a nonstandard definable cut (and by a well-known result of [Pud85], no reasonable theory proves its own ordinary consistency even restricted to a cut).

The authors also thank Konrad Zdanowski for reading and commenting on a preliminary version of this paper.

## References

- [AB01] Z. Adamowicz and T. Bigorajska, *Existentially closed structures and Gödel's second incompleteness theorem*, Journal of Symbolic Logic **66** (2001), 349–356.
- [Ada91] Z. Adamowicz, *On maximal theories*, Journal of Symbolic Logic **56** (1991), 885–890.
- [Bec04] A. Beckmann, *Preservation theorems and restricted consistency statements in bounded arithmetic*, Annals of Pure and Applied Logic **126** (2004), 255–280.
- [GD82] H. Gaifman and C. Dimitracopoulos, *Fragments of Peano's arithmetic and the MDRP theorem*, Logic and Algorithmic, Monographie de l'Enseignement Mathématique, Université Genève, 1982, pp. 187–206.
- [HP93] P. Hájek and P. Pudlák, *Metamathematics of first-order arithmetic*, Springer-Verlag, 1993.
- [Kra95] J. Krajíček, *Bounded arithmetic, propositional logic, and complexity theory*, Cambridge University Press, 1995.
- [Pud85] P. Pudlák, *Cuts, consistency statements, and interpretations*, Journal of Symbolic Logic **50** (1985), 423–441.