FUNDAMENTA
MATHEMATICAE
171 (2002)

Herbrand consistency and bounded arithmetic

by

Zofia Adamowicz (Warszawa)

Abstract. We prove that the Gödel incompleteness theorem holds for a weak arithmetic $T_m = I\Delta_0 + \Omega_m$, for $m \geq 2$, in the form $T_m \nvdash \mathrm{HCons}(T_m)$, where $\mathrm{HCons}(T_m)$ is an arithmetic formula expressing the consistency of T_m with respect to the Herbrand notion of provability. Moreover, we prove $T_m \nvdash \mathrm{HCons}^{I_m}(T_m)$, where HCons^{I_m} is HCons relativised to the definable cut I_m of $(m-2)$-times iterated logarithms. The proof is model-theoretic. We also prove a certain non-conservation result for T_m.

In [PW] Paris and Wilkie asked the following question: does $I\Delta_0$ prove the cut free consistency of $I\Delta_0$? Here we solve (negatively) an analogous question with $I\Delta_0 + \Omega_m$, $m \geq 2$, in place of $I\Delta_0$. The theory $I\Delta_0 + \Omega_m$ can be considered as another version of bounded arithmetic and Herbrand provability is a version of cut free provability (it is defined and formalized in Section 2).

Pudlák [P] and Hájek–Pudlák [HP] proved Gödel's Incompleteness Theorem for weak arithmetic with the ordinary (Hilbert) notion of provability.

As Herbrand consistency of a theory is a weaker statement than its ordinary consistency, proving its unprovability in some theory is more difficult.

In [P] Pudlák also proves that theories of the form $I\Delta_0 + \Omega_m$ do prove their own Herbrand consistency relativised to a certain definable cut J_m. Here we show that they do not prove their own Herbrand consistency relativised to I_m. It follows that consistently $J_m \subsetneq I_m$.

Pudlák [P] (see also Hájek and Pudlák [HP]) in his proof uses a provability predicate Prov and its restriction Prov* to a definable initial segment and shows that Prov and Prov* satisfy some derivability conditions from which the main result is obtained in a routine way.

Our result for the case of $I\Delta_0 + \Omega_2$ $(m = 2)$ has been proved in [AZ] and for $m = 1$ in [A1]. In [AZ] we applied an idea similar to that of [P] and [HP].

2000 *Mathematics Subject Classification*: Primary 03F30.

Research supported in part by The State Committee for Scientific Research (Poland), KBN, grant number 5 P03A-037 20.

In this paper we give a different proof, much more model-theoretic than the former one.

The Paris–Wilkie problem has also been considered by Willard [W].

We use standard notation throughout. In particular, Δ_0 denotes the class of bounded arithmetical formulas and $I\Delta_0$ is the system of weak arithmetic with induction scheme for Δ_0 formulas only. $B\Sigma_1$ denotes the Σ_1 collection scheme. Addition and multiplication are regarded as relations.

Let $\omega_0(x) = x^2$ and $\omega_{m+1}(x) = 2^{\omega_m(\log x)}$ (arithmetical log denotes the integral part of the logarithm). The axiom Ω_m states the totality of the function ω_m. The axiom exp states the totality of the exponential function $y = 2^x$.

Generally formulas are always defined as elements of $\mathbb{N}$ or of a model M under consideration. In other words we identify formulas with their Gödel numbers.

Let Sat be a universal formula for Δ_0. Thus Sat is Σ_1 and

$$M \models \mathrm{Sat}(\varphi) \quad \text{iff} \quad M \models \varphi,$$

for $\varphi \in \Delta_0$, in every model M of $I\Delta_0 + \exp$.

For each $n \in \mathbb{N}$ let

$$\log^n M = \{a \in M : \exists b \in M\ (M \models (b = \exp^n(a)))\}.$$

Of course every $\log^n M$ is a definable initial segment of M ("$y = \exp(x)$" can be expressed by a Δ_0 formula—see [HP]). Thus we have $I_m^M = \log^{m-2} M$.

1. We shall express (in Sec. 2) the Herbrand consistency by a Π_1 formula $\mathrm{HCons}_m(\varphi)$ (φ is Herbrand consistent with T_m). We shall also use an auxiliary Π_1 formula $\mathrm{HCons}_m^{I_m}(\varphi)$, obtained from HCons_m by restriction of the initial quantifier to the definable segment I_m (in the standard model I_m is $\mathbb{N}$). The formula $\mathrm{HCons}_m^{I_m}$ will have the following property:

(∗) *For a bounded θ if*

$$T_m + \exists \overline{x} \in \log^{m+1}\ \theta(\overline{x}) + \mathrm{HCons}_m^{I_m}(\text{“}0 = 0\text{”})$$

is consistent then so is

$$T_m + \exists \overline{x} \in \log^{m+2}\ \theta(\overline{x}).$$

Note that $\mathrm{HCons}_m(\text{“}0 = 0\text{”})$ expresses "T_m is Herbrand consistent".

Now with HCons_m and $\mathrm{HCons}_m^{I_m}$ as above we can prove the announced result,

$$T_m \nvdash \mathrm{HCons}_m^{I_m}(\text{“}0 = 0\text{”}).$$

We need the following theorem:

1.1. Theorem. *For $m, n \in \mathbb{N}$ there is a bounded formula $\theta(\overline{x})$ (where $\overline{x}$ is a finite string of variables) such that*

$$I\Delta_0 + \Omega_n + \exists \overline{x} \in \log^m \ \theta(\overline{x})$$

is consistent and

$$I\Delta_0 + \Omega_n + \exists \overline{x} \in \log^{m+1} \ \theta(\overline{x})$$

is inconsistent.

In particular, for $m \in \mathbb{N}$ there is a bounded formula $\theta_m(\overline{x})$ such that

$$I\Delta_0 + \Omega_m + \exists \overline{x} \in \log^{m+1} \ \theta_m(\overline{x})$$

is consistent and

$$I\Delta_0 + \Omega_m + \exists \overline{x} \in \log^{m+2} \ \theta_m(\overline{x})$$

is inconsistent.

The theorem can be considered as a certain non-conservation result and may be interesting in its own right. We prove it later in this section.

Now the proof of the main result is as follows. Let θ_m be given by Theorem 1.1. We shall show that

$$T_m + \exists \overline{x} \in \log^{m+1} \ \theta_m(\overline{x}) + \mathrm{HCons}_m^{I_m}(\text{“}0 = 0\text{”})$$

is inconsistent.

Suppose that this theory is consistent. Then, by (∗),

$$T_m + \exists \overline{x} \in \log^{m+2} \ \theta_m(\overline{x})$$

is consistent. But this violates the choice of θ_m. Hence

$$T_m + \exists \overline{x} \in \log^{m+1} \ \theta_m(\overline{x}) + \mathrm{HCons}_m^{I_m}(\text{“}0 = 0\text{”})$$

is inconsistent. Thus, in view of the consistency of $T_m + \exists \overline{x} \in \log^{m+1} \ \theta_m(\overline{x})$, we have

$$T_m \nvdash \mathrm{HCons}_m^{I_m}(\text{“}0 = 0\text{”}),$$

which completes the proof.

We have shown that to prove our main result it is sufficient to construct formulas HCons_m, $\mathrm{HCons}_m^{I_m}$ with the properties stated above. This will be done in subsequent sections.

Now let us prove Theorem 1.1.

Let $m, n \in \mathbb{N}$. Suppose that for every bounded formula θ such that $I\Delta_0 + \Omega_n + \exists \overline{x} \in \log^m \ \theta(\overline{x})$ is consistent, $I\Delta_0 + \Omega_n + \exists \overline{x} \in \log^{m+1} \ \theta(\overline{x})$ is consistent. Fix a bounded formula θ_0 such that

$$I\Delta_0 + \Omega_n + \exists \overline{x} \in \log^m \ \theta_0(\overline{x})$$

is consistent, $\overline{x} = x_1, \ldots, x_k$. Hence

$$I\Delta_0 + \Omega_n + \exists \overline{x} \in \log^{m+1} \ \theta_0(\overline{x})$$

is consistent. Therefore

$$I\Delta_0 + \Omega_n + \exists y \in \log^m \ \exists \overline{x} \leq y \ \Big(\bigwedge_{i=1,\dots,k} y \geq 2^{x_i} \wedge \theta_0(\overline{x}) \Big)$$

is consistent.

Let $\theta_1(y)$ be $\exists \overline{x} \leq y \ (\bigwedge_{i=1,\dots,k} y \geq 2^{x_i} \wedge \theta_0(\overline{x}))$. Applying our supposition to θ_1 we infer

$$I\Delta_0 + \Omega_n + \exists y \in \log^{m+1} \ \exists \overline{x} \leq y \ \Big(\bigwedge_{i=1,\dots,k} y \geq 2^{x_i} \wedge \theta_0(\overline{x}) \Big)$$

is consistent. Hence

$$I\Delta_0 + \Omega_n + \exists \overline{x} \in \log^{m+2} \ \theta_0(\overline{x})$$

is consistent. Continuing we infer that

$$I\Delta_0 + \Omega_n + \exists \overline{x} \in \log^{m+n'} \ \theta_0(\overline{x})$$

is consistent for all $n' \in \mathbb{N}$. Thus there is a model M of $I\Delta_0$ and an $\overline{a} \in M$ such that $M \models \theta_0(\overline{a})$ and $M \models (\exp^{n'}(\max \overline{a})$ exists) for $n' \in \mathbb{N}$. Consider the initial segment M' of M determined by the elements $\exp^{n'}(\max \overline{a})$ for $n' \in \mathbb{N}$. Then $M' \models I\Delta_0 + \exp$ and $M' \models \theta_0(\overline{a})$. It follows that the theory $I\Delta_0 + \exp + \exists \overline{x} \ \theta_0(\overline{x})$ is consistent.

Thus, for every bounded θ, if $I\Delta_0 + \Omega_n + \exists \overline{x} \in \log^m \ \theta(\overline{x})$ is consistent then so is $I\Delta_0 + \exp + \exists \overline{x} \ \theta(\overline{x})$.

Also, for any bounded $\theta_1, \dots, \theta_l$, if

$$I\Delta_0 + \Omega_n + \bigwedge_{i=1,\dots,l} \exists \overline{x} \in \log^m \ \theta_i(\overline{x})$$

is consistent then so is

$$I\Delta_0 + \exp + \bigwedge_{i=1,\dots,l} \exists \overline{x} \ \theta_i(\overline{x})$$

because the sentence $\bigwedge_{i=1,\dots,l} \exists \overline{x} \ \theta_i(\overline{x})$ can be presented as

$$\exists \overline{x}_1, \dots, \overline{x}_l \bigwedge_{i=1,\dots,l} \theta_i(\overline{x}_i).$$

Let Σ_1^* denote the collection of all sentences of the form $\exists \overline{x} \in \log^m \ \theta(\overline{x})$, where θ is bounded. Let $T^* \subseteq \Sigma_1^*$ be maximal (with respect to Σ_1^*) consistent with $I\Delta_0 + \Omega_n$. It follows that $I\Delta_0 + \exp + T^*$ is consistent.

Let $T^{**} \subseteq \Sigma_1$ consist of those Σ_1 sentences ϕ of the form $\exists \overline{x} \ \theta(\overline{x})$, where θ is bounded, for which the sentence "$\exists \overline{x} \in \log^m \ \theta(\overline{x})$" is in T^*. We shall show that T^{**} is maximal consisting of Σ_1 sentences consistent with $I\Delta_0 + \exp$.

Let $\phi \in \Sigma_1$ of the form $\exists \overline{x}\ \theta(\overline{x})$ be such that $I\Delta_0 + \exp + T^{**} + \phi$ is consistent. Then

$$I\Delta_0 + \exp + T^{**} + \exists \overline{x} \in \log^m\ \theta(\overline{x})$$

is consistent. Hence

$$I\Delta_0 + \Omega_n + T^{*} + \exists \overline{x} \in \log^m\ \theta(\overline{x})$$

is consistent. Hence, by the maximality of T^*, the sentence "$\exists \overline{x} \in \log^m\ \theta(\overline{x})$" is in T^*, whence $\phi \in T^{**}$. It follows that T^{**} is maximal consisting of Σ_1 sentences consistent with $I\Delta_0 + \exp$.

Let $M \models I\Delta_0 + \exp + T^{**} + B\Sigma_1$ be such that $\Sigma_1(M)$ (the set of Σ_1 sentences true in M) is not coded in M. Such a model M exists (see [WP2], the proof of Theorem 9). Note that by the maximality of T^{**}, $\Sigma_1(M) = T^{**}$.

By another result of [WP2] (Theorem 5(2)), M has a proper end-extension to a model M' of $I\Delta_0 + \Omega_{n+1}$. By the maximality of T^* with respect to $I\Delta_0 + \Omega_n$ and Σ_1^*, we have

$$M' \models \phi \ \Leftrightarrow\ M \models \phi,$$

for every $\phi \in \Sigma_1^*$. Let $a \in M' \setminus M$. We thus have

$$M \models \phi \ \Leftrightarrow\ M' \models \phi^a,$$

for every $\phi \in \Sigma_1^*$.

Since every $\phi \in \Sigma_1$ is equivalent in M (and so in M') in a canonical way to an $\exists\Sigma_1^b$ sentence (via the Matiyasevich theorem) and $M' \models \Omega_1$, we may use the universal formula for $\exists\Sigma_1^b$ formulas available in M' to infer that

$$\{\phi \in \Sigma_1^* : M' \models \phi^a\}$$

is coded in M'. Here Σ_1^b denotes Buss's class (see [B], [HP]) and $\exists\Sigma_1^b$ denotes the class of formulas of the form $\exists \overline{x}\ \theta(\overline{x})$, where θ is Σ_1^b. The required universal formula can be built using the formula μ_1 from Theorem 4.18 of [HP] (see also the appendix of [A]). The notation ϕ^a denotes the formula obtained from ϕ by bounding its unbounded existential quantifiers to a.

But then T^* is coded in M' and consequently so is T^{**}; hence $\Sigma_1(M)$ is coded in M', whence it is coded in M, contradiction.

Thus the theorem has been proved.

2. Let us recall what we mean by Herbrand type provability of a sentence. Let φ be a sentence of the form

$$(2.1) \qquad \exists x_1\ \forall y_1 \ldots \exists x_m\ \forall y_m\ \overline{\varphi}(x_1, y_1, \ldots, x_m, y_m),$$

where $\overline{\varphi}$ is open.

Extend the language by new function symbols $f_1, \ldots, f_m$ such that f_k is of arity k. The symbol f_k can be treated as a symbol for a Skolem function for the kth existential quantifier in $\neg\varphi$. Let $\mathcal{T}$ be the set of terms of the extended language. We call $\widetilde{\varphi}(t_1,\ldots,t_m)$ a *Herbrand variant* of φ if $\widetilde{\varphi}$ is of the form

$$\overline{\varphi}(t_1,f_1(t_1),\ldots,t_m,f_m(t_1,\ldots,t_m))$$

for some $t_1,\ldots,t_m\in\mathcal{T}$.

We say that φ is *Herbrand provable* (in logic) if there is a finite $\mathcal{T}'\subseteq\mathcal{T}$ such that

$$\bigvee_{t_1,\ldots,t_m\in\mathcal{T}'}\widetilde{\varphi}(t_1,\ldots,t_m)$$

is a propositional tautology.

Assume now that $T=\{\phi_1,\phi_2,\ldots\}$ is a fragment of arithmetic and $+,\cdot$ are treated as relations. Assume that ϕ_j is of the form

$$\forall x_1\ \exists y_1\ldots\forall x_m\ \exists y_m\ \bar{\phi}_j(x_1,y_1,\ldots,x_m,y_m),$$

where $\bar{\phi}$ is open. We may assume that $m\le j$.

We are aiming at formulating Herbrand type consistency of T. To this end we need to extend the language by some function symbols s^j_k such that s^j_k is of arity k. The symbol s^j_k is a symbol for a Skolem function for the kth existential quantifier in ϕ_j. We have $k\le j$. Let the language so obtained be denoted by $\widetilde{L}$. Then, by the above definition, a Herbrand variant of $\neg\phi_j$ is a formula $\neg\widetilde{\phi}_j(t_1,\ldots,t_k)$ of the form

$$\neg\bar{\phi}_j(t_1,s^j_1(t_1),\ldots,t_k,s^j_k(t_1,\ldots,t_k)),$$

where $t_1,\ldots,t_k$ are terms of $\widetilde{L}$.

Then T is *Herbrand inconsistent* if a finite disjunction of some Herbrand variants

$$\neg\widetilde{\phi}_j(t_1,\ldots,t_k)$$

is provable in the propositional calculus.

Hence, T is *Herbrand consistent* if every finite conjunction of some $\widetilde{\phi}_j(t_1,\ldots,t_k)$ is consistent with the propositional calculus.

To formalize the property "T is Herbrand consistent" in arithmetic we have to encode the language $\widetilde{L}$ in arithmetic. So we number all terms of $\widetilde{L}$ in the following natural order. Let the constants $0,1$ be terms of rank 0 and let the terms of rank at most $i+1$ consist of all terms of rank at most i and of all terms of the form $s^j_k(t_1,\ldots,t_k)$ for $j\le i+1-(k-1)$, with $t_1,t_2,\ldots,t_k$ of rank $i-(k-1),i-(k-2),\ldots,i$ respectively. We number terms of rank 0, then of rank 1 etc. by consecutive natural numbers leaving some numbers not used. Let the numbers left aside serve to number logical symbols of the language $\widetilde{L}$. The exact form of our numbering is given below in this section. Then terms of rank at most i are numbered by numbers less than l_i, for some recursive function $i\mapsto l_i$.

As a matter of fact in our numbering every term has a lot of numbers. If $t_1, \ldots, t_k$ are of rank i_0 then they are also of rank at most i for every $i \geq i_0$, and so the term $s^j_k(t_1, \ldots, t_k)$ is a term of rank at most $i+1$ for every $i \geq i_0$.

There are recursive uniformly definable functions $S^{i,j}_k$ such that $S^{i,j}_k : [0, l_{i-(k-1)}) \times [0, l_{i-(k-2)}) \times \ldots \times [0, l_i) \to [0, l_{i+1})$ and the following holds: if the terms $t_1, \ldots, t_k$ are numbered by $a_1, \ldots, a_k$ then the term $s^j_k(t_1, \ldots, t_k)$ as a term of rank $i+1$ is numbered by $S^{i,j}_k(a_1, \ldots, a_k)$.

Let the encoded language be denoted by L^*. Let E_i denote the collection of encoded atomic and negated atomic formulas on terms of rank at most i.

We shall call a function $p : E_i \to \{0,1\}$ a T-*evaluation* of rank i if $p(\neg\varphi) = 1 - p(\varphi)$ for $\varphi \in E_i$. Each such p extends uniquely (in a routine way) to open sentences of L^* with terms $< l_i$. We assume further that $p(\varphi) = 1$ for every axiom of equality φ and that p makes

$$\bar{\phi}_j(t_1, s^j_1(t_1), \ldots, t_k, s^j_k(t_1, \ldots, t_k))$$

true for every Herbrand variant of ϕ_j with terms of rank at most i, i.e. p takes value 1 at the formula

$$\bar{\phi}_j(a_1, S^{i_1,j}_1(a_1), \ldots, a_k, S^{i_k,j}_k(a_1, \ldots, a_k))$$

of L^*, for $a_1 < l_{i_1}$, $a_2 < l_{i_2}, \ldots, a_k < l_{i_k}$, $j < i_1 < i_2 < \ldots < i_k < i$.

Note that every T-evaluation of rank $i+1$ makes true every conjunction of some $\widetilde{\phi}_j(t_1, \ldots, t_k)$ with $t_1, \ldots, t_k$ of rank at most $i-(k-1), i-(k-2), \ldots, i$ respectively.

Thus, T is Herbrand consistent if for every i there is a T-evaluation of rank i.

To be able to define all the required notions at stage i we need $\exp^3 i$ to exist. This is because the numbers l_i and E_i are roughly of size $\exp^2 i$ and any T-evaluation of rank i is roughly of size $\exp^3 i$.

The whole formalization is available in $I\Delta_0 + \Omega_m$. In particular we have a Δ_1 formula $V^T(p, i)$ expressing "p is a T-evaluation of rank i". Then we may formulate Hcons(T) as

$$\forall i \in \log^3 \ \exists p \ V^T(p, i).$$

This may be considered a weak form of Herbrand consistency, but it makes our negative results even stronger.

Here is the exact definition of our coding. Let M be a model of T_m. Define

$$l_0 = 2,$$
$$l_{i+1} = l_i + (i+1)l_i + il_il_{i-1} + \ldots + l_i \ldots l_0.$$

We have

$$l_i \leq 2^{2^i} \tag{2.2}$$

for each $i \in \log^3$. For,

$$l_{i+1} = l_i(1 + (i+1) + il_{i-1} + \ldots + l_{i-1} \ldots l_0) = l_i(1 + (i+1) + l_i - l_{i-1})$$

and hence, assuming (2.2) for a given $i > 0$, we obtain

$$l_{i+1} \le 2^{2^i}(1 + (i+1) + 2^{2^i} - l_{i-1}) = (2^{2^i})^2 + 2^{2^i}(1 + (i+1) - l_{i-1}) \le 2^{2^{i+1}}$$

since, obviously, $l_j \ge 1 + (j+2)$ for each $j \ge 1$.

The graph of the function l (as a function of i) is definable in T_m (with the help of an ordinary technique, see e.g. [WP1] or [HP]), so that the domain of l is an initial segment. From (2.2) it follows that l is defined at least on $\log^3$. An easy estimation shows that

$$\log^{m-1} M = \bigcup_{i \in \log^{m+1} M} [l_i, l_{i+1})$$

(where $[a,b)$ is the interval $\{x : a \le x < b\}$), but we do not use this fact. Define also (i, j will denote elements of $\log^m$ throughout)

$$\begin{aligned} (2.3) \quad S_k^{i,j}(a_1, \ldots, a_k) = l_i &+ (i+1)l_i + il_il_{i-1} \\ &+ \ldots + ((i+1) - (k-2))l_i \ldots l_{i-(k-2)} \\ &+ jl_i \ldots l_{i-(k-1)} + (a_1, \ldots, a_k)_i \quad \text{for } k \ge 2, \\ S_1^{i,j}(a_1) = l_i &+ jl_i + (a_1)_i \end{aligned}$$

for $1 \le k \le i$, $j \le i$ and $a_1 < l_{i-(k-1)}, \ldots, a_k < l_i$ (otherwise set $S_k^{i,j} = 0$; also let $S_1^{0,0}(0) = 2$ and $S_1^{0,0}(1) = 3$). Here $(a_1, \ldots, a_k)_i$ denotes the position (a number $\le l_i \ldots l_{i-(k-1)}$) of $(a_1, \ldots, a_k)$ in the lexicographical ordering of the product

$$[0, l_{i-(k-1)}) \times \ldots \times [0, l_i).$$

The graph of $S_k^{i,j}$ is (uniformly in i, j, k) definable in T_m. The values $S_k^{i,j}(a_1, \ldots, a_k)$ as in (2.3) fill the interval $[l_i, l_{i+1})$ for each $i \in \log^3 M$. Thus, (2.3) constitutes a numbering of $\log M$ (except 0 and 1).

The inner language L^*, encoded in T_m in the usual way, is obtained from the ordinary arithmetical language L (in which addition and multiplication are treated as relations) by adding elements $a \in \log$ as terms (except the $S_k^{i,j}(a_1, \ldots, a_k)$'s with $j = 0$, which may serve to define other primitive notions and formulas of L^*).

Let T be a set of sentences of L^*. An evaluation p on E_i is a T-evaluation if p satisfies the following condition (denoted briefly by $p \Vdash^* \varphi$):

(2.4) for each axiom φ of T in its prenex form, $\forall x_1 \, \exists y_1 \ldots \forall x_m \, \exists y_m \overline{\varphi}$, if φ has index $j \ge 1$ (in a fixed enumeration of T), then for all $i_1 < \ldots < i_m < i$ ($\varphi < i_1$) and arbitrary $a_1 < l_{i_1}, \ldots, a_m < l_{i_m}$, p assumes the value 1 at $\overline{\varphi}(a_1, S_1^{i_1,j}(a_1), \ldots, a_m, S_m^{i_m,j}(a_1, \ldots, a_m))$.

It is understood here that all terms occurring in the axioms of T are less than l_i. Notice that, roughly, we have $p \le 2^{l_i}$ for each evaluation p on E_i.

The condition (2.4) generalizes in a natural way as follows. For an evaluation p on E_i and a sentence φ of L^* as in (2.1), which contains parameters $\overline{c}$ and is of the form $\psi(\overline{c})$, where $\psi \in \mathbb{N}$ and $\overline{c} < l_j$, write $p \Vdash \varphi$ if

$$\forall i_1 \in [j+1, i)\ \forall a_1 < l_{i_1}\ \exists b_1 < l_{i_1+1} \ldots \quad \forall i_m \in [i_{m-1}+1, i)\ \forall a_m < l_{i_m}\ \exists b_m < l_{i_m+1}$$

such that p is 1 at

$$\overline{\varphi}(a_1, b_1, \ldots, a_m, b_m).$$

For open φ we assume that $p \Vdash \varphi$ if $p(\varphi) = 1$. Thus, we have $p \Vdash \varphi$ for each standard axiom φ of T and each T-evaluation p.

All quantifiers in the above definition are bounded by $l_i \in \log$ (i.e. $\exp(l_i)$ exists). Hence, using the universal formula Sat we can find a Δ_0 formula F with an additional parameter b (bounding the unrestricted quantifier in Sat) such that

$$p \Vdash \varphi \quad \text{iff} \quad F(p, i, \varphi, b)$$

for every evaluation p on E_i, standard φ with terms $< l_i$ and any b such that $b \ge 2^{l_i^{\varphi}}$ (cf. Lessan [L] and Theorem 2 of [DP]). It follows that

$$p \Vdash \varphi \quad \text{iff} \quad \forall b\ (b \ge 2^{l_i^{\varphi}} \Rightarrow F(p, i, \varphi, b)) \tag{2.5}$$

for every evaluation p on E_i and a standard sentence φ with terms $< l_i$.

Assume that T is Δ_0 definable in T_m. We construct a Δ_1 formula V^T such that

$$p \text{ is a } T\text{-evaluation on } E_i \quad \text{iff} \quad V^T(p, i) \tag{2.6}$$

iff $\forall b\ (b \ge 2^{\omega_1(l_i)} \Rightarrow V_0^T(p, i, b))$ with bounded V_0^T.

Let M be a (non-standard) model of T_m and let $i' = i + j \in \log^3 M$, where $j > \mathbb{N}$. Every T-evaluation $p \in M$ on E_{i+j}^M determines a model $M(p, i)$ as follows. Put

$$a =_p b \ \equiv\ p(\text{“}a = b\text{”}) = 1$$

for $a, b < l_{i+\mathbb{N}}$. Clearly, $=_p$ is an equivalence relation on the initial segment $[0, l_{i+\mathbb{N}})$ of M. Let

$$M(p, i) = \{[a] : a < l_{i+\mathbb{N}}\}$$

consist of equivalence classes and define

$$[a] + [b] = [c] \quad \text{iff} \quad p(\text{“}a + b = c\text{”}) = 1$$

and similarly for multiplication and ordering. It follows immediately that

$$M(p, i) \models \varphi \quad \text{iff} \quad p(\varphi) = 1$$

for arbitrary open φ with parameters $< l_{i+\mathbb{N}}$ (a is a name for $[a]$).

Also, directly from the above definition, we obtain the following:

(2.7) If $p \Vdash \varphi(c_1, \ldots, c_n)$, then $M(p,i) \models \varphi([c_1], \ldots, [c_n])$

for arbitrary standard φ with parameters $c_1, \ldots, c_n < l_{i+\mathbb{N}}$. In particular, $M(p,i)$ is a model of $T \cap \mathbb{N}$ for every T-evaluation p. The converse of (2.7) is in general not true.

Therefore the formula $\mathrm{HCons}_m(\varphi)$, expressing the Herbrand consistency of φ with T_m, can be assumed to have the form

$$\forall i \in \log^3 \ \exists p \ V^{T_m+\varphi}(p,i).$$

More precisely, $\mathrm{HCons}_m(\varphi)$ looks like

(2.8) $\forall y \ \forall i \le y \ [i \le \log^3 y \wedge y \ge 2^{\omega_1(l_i)} \Rightarrow \exists p \le y \ \ V_0^{T_m+\varphi}(p,i,y)]$.

Finally, $\mathrm{HCons}(T_m)$ is $\mathrm{HCons}_m(\text{“}0 = 0\text{”})$.

3. In order to prove that HCons_m and $\mathrm{HCons}_m^{I_m}$ have the required properties we need some auxiliary lemmas.

Lemma 3.2 and Corollary 3.3 show that the models $M(p,i)$ are end-extensions of the initial segment $\le^M i$ of M. Theorem 3.4 is the main step in proving $(*)$ of the introduction. It shows that $M(p,i)$ is a stretching of M in the sense that an element i of $\log^{m+1} M$ gets an additional exponent in $M(p,i)$ (falls into $\log^{m+2} M(p,i)$). Finally we prove $(*)$ of the introduction.

3.1. Definition. Let $M \models T_m$ be given and $i_0 \in \log^3 M$. Let p be a T_m-evaluation on E_{i_0}. For $i < i_0$ we define a numeral $\underline{i}$ determined by p. The sentence $\forall x \ \exists y \ (y = x+1)$ is an axiom of T_m and we may assume that this is the first axiom in a fixed enumeration of T_m. It follows that

$$\forall a < l_i \ \exists b < l_{i+1} \ p \Vdash (b = a+1)$$

for all $i < i_0$. Hence there exists a sequence $\langle c_i : i < i_0 \rangle$ of names such that

$$p \Vdash (c_0 = 0) \quad \text{and} \quad p \Vdash (c_{i+1} = c_i + 1) \text{ for all } i < i_0.$$

Let $\underline{i} = c_i$ for $i < i_0$.

In the next lemma and corollary we shall show that $\underline{i}$ is a name of the ith integer in the models $M(p,j)$ with $j < i_0 - \mathbb{N}$, in the case where i_0 is non-standard.

3.2. Lemma. *Let p, i_0 be as before. If, for some name a, $p \Vdash (a \le \underline{i})$, then there is a $j \le i$ such that $p \Vdash (a = \underline{j})$. Moreover,*

$(**)$ $\varphi(i_1, \ldots, i_n)$, *where* $i_1, \ldots, i_n < i_0$, *implies* $p \Vdash \varphi(\underline{i}_1, \ldots, \underline{i}_n)$,

for open φ all of whose terms are as indicated.

Proof. Induction on $i < i_0$. For $i = 0$ we have $p \Vdash (a \le \underline{i})$, whence $p \Vdash (a \le 0)$. Since the sentence $\forall x \ (x \le 0 \Rightarrow x = 0)$ can be assumed to be the axiom of T_m, we get $p \Vdash (a = 0)$, whence $p \Vdash (a = \underline{0})$. In the inductive step we apply, in a similar way, the axiom

$$\forall x, y, z\ (y = z + 1 \wedge x \leq y \Rightarrow x = y \vee x \leq z)$$

to $p \Vdash (a \leq \underline{i+1})$, i.e. to $p \Vdash (a = \underline{i} + 1)$, and obtain $p \Vdash (a = \underline{i+1})$ or $p \Vdash (a \leq \underline{i})$. In the latter case we use the inductive assumption to infer $p \Vdash (a = \underline{j})$ for some $j \leq i$.

For the second assertion of the lemma we prove first $p \Vdash (\underline{i} + \underline{j} = \underline{i+j})$ for all i, j such that $i + j < i_0$. We apply induction on j. Since p evaluates $\underline{i}+\underline{0}$ as $\underline{i}+0$, the axiom $\forall x\ (x+0 = x)$ yields immediately $p \Vdash (\underline{i}+\underline{0} = \underline{i})$. For the inductive step, notice that p evaluates $\underline{i}+\underline{j+1}$ as $\underline{i}+\underline{j}+1$ and hence as $\underline{i+j}+1$, by the inductive assumption. On the other hand $p \Vdash (\underline{i+j}+1 = \underline{i+j+1})$, by definition of the numerals, which yields the required result. In a similar way we prove $p \Vdash (\underline{i} \cdot \underline{j} = \underline{i \cdot j})$ for all i, j such that $i, j < i_0$, and also $p \Vdash (\underline{i} < \underline{j})$ whenever $i \leq j$. This shows that $(**)$ holds for all atomic (and therefore also for all open) sentences φ, which finishes the proof of the lemma.

We have the following immediate corollary:

3.3. Corollary. *Let M be a model of T_m, $i_0 \in \log^3 M$ and $p \in M$ a T_m-evaluation on E_{i_0+j}, where $j > \mathbb{N}$. Then the initial segment $\leq i_0$ of M is isomorphically embeddable into $M(p, i_0)$ as an initial segment. Consequently, if $a_1, \ldots, a_k \in M$, $a_1, \ldots, a_k \leq i_0$, $\varphi(x_1, \ldots, x_k)$ is bounded* (*with $+$ and $\cdot$ treated as relations*) *and*

$$M \models \varphi(a_1, \ldots, a_k),$$

then $M(p, i_0) \models \varphi([\underline{a_1}], \ldots, [\underline{a_k}])$.

Recall that $I_m = \log^{m-2} M$.

Note that in the presence of Ω_m, the segment $\log^{m+1} M$ is closed under addition. For, we have

$$\begin{aligned} \exp^{m+1}(2a) &= \exp^m(2^{2a}) = \exp^m((2^a)^2) \\ &= \exp^m(\omega_0(\exp(a))) = \exp^{m-1}(\omega_1(\exp^2(a))) \\ &= \exp^{m-2}(\omega_2(\exp^3(a))) = \ldots = \omega_m(\exp^{m+1}(a)). \end{aligned}$$

So, if $\exp^{m+1}(a)$ exists in a model of $I\Delta_0 + \Omega_m$, then (by Ω_m), $\omega_m(\exp^{m+1}(a))$ exists, and thus $\exp^{m+1}(2a)$ exists. To see that $\exp^{m+1}(a + b)$ exists provided $\exp^{m+1}(a)$ and $\exp^{m+1}(b)$ exist, we show that $\exp^{m+1}(2\max(a, b))$ exists, and then using the Δ_0 minimum principle we infer the existence of $\exp^{m+1}(a + b)$.

It follows that I_m is closed under ω_2.

The following theorem implies the property $(*)$ of Section 1.

3.4. Theorem. *Let M be a model of T_m and $i_0 \in \log^{m+1} M$, $i_0 > \mathbb{N}$. Let $p \in M$ be a T_m-evaluation on E_{2i_0}. Then the model $M(p, i_0)$ satisfies*

$$T_m + [\underline{i}_0] \in \log^{m+2}.$$

Proof. Since Ω_m is an axiom of T_m we have $p \Vdash (\forall x \, \exists y \; y = \omega_m(x))$ and so

$$\forall a < l_i \; \exists b < l_{i+1} \; p \Vdash (b = \omega_m(a))$$

for each $i < i_0$. From (2.5) it follows that, for a fixed φ, the relation $p \Vdash \varphi$ is Δ_0 over M. Thus, there is a (code of a) sequence $\langle w_i : i \le i_0 \rangle \in M$ of names satisfying

$$\forall i < i_0 \; p \Vdash (w_{i+1} = \omega_m(w_i)) \quad \text{and} \quad p \Vdash (w_0 = \exp^m 2).$$

Clearly there is a standard n_0 (depending on the position of Ω_m in the enumeration of axioms) such that $w_i < l_{i+n_0}$ for each $i \le i_0$.

Provably in T_m, we have

$$\exp^{m+2}(k) = \omega_m^k(\exp^m 2) \tag{3.6}$$

for each $k \in \log^{m+2}$ (the superscript k denotes the kth iteration). This can be proved in T_m by straightforward induction on $l \le k$ applied to the formula $\exp^{m+2}(l) = \omega_m^l(\exp^m 2)$ which can be bounded by $\omega_m(\exp^{m+2}(k))$.

In fact the right hand side of (3.6), i.e. $y = \omega_m^k(\exp^m 2)$ can be defined by an arithmetical formula with the help of the Gödel β-function: let $\psi(x, y, a, b)$ be

$$\beta(a, b, 0) = \exp^m 2 \wedge \beta(a, b, x) = y \wedge \forall i < x \; \beta(a, b, i+1) = \omega_m(\beta(a, b, i))$$

where $\beta(a, b, i) = r$ stands for

$$\exists q \; (a = q(b(i+1)+1) + r \wedge r < b(i+1)+1).$$

Now, $y = \omega_m^x(\exp^m 2)$ can be defined by the formula $\exists a, b \; \psi(x, y, a, b)$.

In order to find a small enough name for a sequence corresponding to the w_is, let $\mathfrak{M}$ be the model $M(p, i_0)$ determined by p over M and consider the sequence s of iterations

$$s = \langle \exp^m 2, \omega_m(\exp^m 2), \ldots, \omega_m^{[\underline{k}]}(\exp^m 2) \rangle$$

of ω_m in $\mathfrak{M}$, where $[\underline{k}]$ is the maximal j with the property $\omega_m^j(\exp^m 2) \le [w_{i_0}]$ in $\mathfrak{M}$. Since the length and terms of s are relatively small, a standard reasoning shows that s has a β-code (a, b) in $\mathfrak{M}$, i.e.

$$\forall i \le [\underline{k}] \; \beta(a, b, i) = \omega_m^i(\exp^m 2)$$

in $\mathfrak{M}$. Since $\mathfrak{M} = M(p, i_0)$, the elements a, b have names A and B, respectively, with $A, B < l_{i_0+n_1}$ (for some standard $n_1 \in \mathbb{N}$).

Moreover, there are names $q_i, r_i < l_{i_0+n_i}$ for an $n_i \in \mathbb{N}$ such that

$$p \Vdash (A = q_i(B(i+1)+1) + r_i \wedge r_i < B(i+1)+1), \tag{3.7}$$

for each $i \le k$.

We shall show that there is a sequence $\langle q_i, r_i : i \le k\rangle$ in M such that $q_i, r_i < l_{i_0+n_i}$ for an $n_i \in \mathbb{N}$ and (3.7) holds.

For, we have in M

$$\forall i \le k\ \exists q_i, r_i\ p \Vdash (A = q_i(B(i+1)+1) + r_i \wedge r_i < B(i+1)+1).$$

Choose now q_i, r_i in M, for $i \le k$, so that q_i, r_i satisfy (3.7) and the least j such that $q_i, r_i < l_{i_0+j}$ is the least possible j for which suitable q_i, r_i exist. Then $j \in \mathbb{N}$ and the sequence $\langle q_i, r_i : i \le k\rangle$ is Δ_0 definable in M, so it is in M.

An easy induction in M shows that

$$p \Vdash (r_i = w_i) \tag{3.8}$$

for each $i \le k$. For, assume (3.8) for a given $i < k$. Thus

$$\mathfrak{M} \models [r_i] = [w_i].$$

By construction of the w's, $p \Vdash (w_{i+1} = \omega_m(w_i))$. Hence $[w_{i+1}] = \omega_m([r_i]) = [r_{i+1}]$ in $\mathfrak{M}$, which proves (3.8).

In particular we have

$$[r_k] = [w_k].$$

Suppose $k < i_0$. Then $p \Vdash (w_{k+1} = \omega_m(w_k))$, whence in $\mathfrak{M}$,

$$\begin{aligned}\omega_m^{[\underline{k+1}]}(\exp^m 2) &= \omega_m(\omega_m^{[\underline{k}]}(\exp^m 2)) = \omega_m([r_k]) \\ &= \omega_m([w_k]) = [w_{k+1}] \le [w_{i_0}],\end{aligned}$$

which contradicts the maximality of k. Hence $k = i_0$, and therefore (3.8) holds for each $i \le i_0$.

Note that

$$\mathfrak{M} \models [r_i] = \omega_m^{[\underline{i}]}(\exp^m 2),$$

by the choice of a, b and A, B. Hence

$$\mathfrak{M} \models [w_{i_0}] = \omega_m^{[\underline{i}_0]}(\exp^m 2) = \exp^{m+2}[\underline{i}_0].$$

Thus the proof of the theorem is complete.

Now we shall show $(*)$ of Section 1. Consider first a model M of

$$T_m + \exists \overline{x} \in \log^{m+1}\ \varphi(\overline{x}) + \mathrm{HCons}^{I_m}(\text{“}0 = 0\text{”}).$$

Let $\overline{a} \in \log^{m+1} M$, $\overline{a} = a_1, \ldots, a_k$, be such that $M \models \varphi(\overline{a})$. Let $i_0 = \max \overline{a}$. Since $\log^{m+1} M$ is closed under addition we infer

$$M \models \exists p\ V^{T_m}(p, 2i_0).$$

Fix p. By Corollary 3.3, $M(p, i_0) \models \varphi([\underline{a}_1], \ldots, [\underline{a}_k])$. By Theorem 3.4,

$$M(p, i_0) \models T_m + \varphi([\underline{a}_1], \ldots, [\underline{a}_k]) + [\underline{a}_1], \ldots, [\underline{a}_k] \in \log^{m+2}.$$

Hence the theory

$$T_m + \exists \overline{x} \in \log^{m+2}\ \varphi(\overline{x})$$

is consistent and $(*)$ follows.

Added in proof. Recently two new manuscripts on a similar subject have appeared: [W1]—a solution of the original version of the Paris–Wilkie problem, and [S]—a new partial solution.

References

[A] Z. Adamowicz, *A contribution to the end-extension problem and the* Π_1 *conservativeness problem*, Ann. Pure Appl. Logic 61 (1993), 3–48.

[A1] —, *On Tableau consistency in weak theories*, circulated manuscript, 1996; preprint 618, Inst. Math., Polish Acad. Sci., 2001.

[AZ] Z. Adamowicz and P. Zbierski, *On Herbrand consistency in weak arithmetic*, Arch. Math. Logic 40 (2001), 399–413.

[B] S. R. Buss, *Bounded Arithmetic*, Bibliopolis, 1986.

[DP] C. Dimitracopoulos and J. Paris, *Truth definitions for* Δ_0 *formulae*, in: Logic and Algorithmic, Monograph. Enseign. Math. 30, Univ. Genève, 1982, 317–329.

[HP] P. Hájek and P. Pudlák, *Metamathematics of First-Order Arithmetic*, Perspectives in Mathematical Logic, Springer, Berlin, 1993.

[L] H. Lessan, *Models of arithmetic*, dissertation, Manchester.

[PW] J. Paris and A. Wilkie, Δ_0 *sets and induction*, in: Open Days in Model Theory (Jadwisin, 1981), W. Guzicki *et al.* (eds.), Leeds Univ. Press, 1981, 237–248.

[P] P. Pudlák, *Cuts, consistency statements and interpretations*, J. Symbolic Logic 50 (1985), 423–441.

[S] S. Salehi, *Herbrand consistency in arithmetic with bounded induction*, PhD thesis, Inst. Math., Polish Acad. Sci., submitted.

[W] D. Willard, *The semantic Tableau version of the second incompleteness theorem extends almost to Robinson's arithmetic Q*, in: Automated Reasoning with Semantic Tableaux and Related Methods, Lecture Notes in Comput. Sci. 1847, Springer, 2000, 415–430.

[W1] —, *How to extend the semantic Tableaux and cut-free versions of the second incompleteness theorem almost to Robinson's arithmetic Q*, J. Symbolic Logic, to appear.

[WP1] A. Wilkie and J. Paris, *On the scheme of induction for bounded arithmetic formulas*, Ann. Pure Appl. Logic 35 (1987), 261–302.

[WP2] —, —, *On the existence of end extensions of models of bounded induction*, in: Logic, Methodology and Philosophy of Science, VIII (Moscow, 1987), Stud. Logic Found. Math. 126, North-Holland, 1989, 143–161.

Institute of Mathematics
Polish Academy of Sciences
Śniadeckich 8
00-950 Warszawa, Poland
E-mail: zosiaa@impan.gov.pl

Received 6 November 2000;
in revised form 21 June 2001