Annals of Pure and Applied Logic 48 (1990) 261–276
North-Holland

EXPONENTIATION AND SECOND-ORDER BOUNDED ARITHMETIC

Jan KRAJÍČEK*
Department of Mathematics, University of Illinois, 1409 West Green Street, Urbana, IL 61801, USA

Communicated by A. Nerode
Received 30 January 1989; revised June 1989

(1) $V_2^i \vdash A(a)$ iff for some term t: $S_2^i \vdash$ "$2^{t(a)}$ exists $\rightarrow A(a)$", a bounded first-order formula, $i \geq 1$.
(2) V_2^i (resp. V_2) is not Π_1^b-conservative over S_2^i (resp. over S_2).
(3) Any model of V_2 not satisfying Exp satisfies the collection scheme $B\Sigma_1^0$.
(4) V_3^1 is not Π_1^b-conservative over S_2.

Second-order bounded arithmetic V_2 and its fragments V_2^i were introduced in [1]. Here we investigate the relation of these systems to the first-order systems S_2 and S_2^i augmented by a limited use of exponentiation. The main connection is the following: For $A(a)$ a first-order bounded formula, V_2^i proves $A(a)$ iff S_2^i proves

$$\text{“}2^{t(a)} \text{ exists} \rightarrow A(a)\text{”} \quad \text{for some term } t(a).$$

From this we entail that V_2^i is not Π_1^b-conservative over S_2^i.

The connection between second-order systems and exponentiation is proved by a model-theoretic argument. This argument can be used to show that a model of V_2 not satisfying Exp must satisfy $B\Sigma_1^0$. This contributes to the question from [6] whether there is a model of $I\Delta_0 + \neg\text{Exp}$ not satisfying $B\Sigma_1^0$.

Finally we define a very weak provability notion for S_2 devised for a construction of a consistency statement which would separate S_2 and V_2^1. We do not succeed; however, the provability notion can be used to separate S_2 and $V_2^1 +$ "f is total", for any reasonably defined non-decreasing function f which eventually majorizes all $2^{|x|^k}$ $(k < \omega)$. In particular, V_3^1 is not Π_1^b-conservative over S_2.

I would like to thank G. Takeuti for some valuable comments and suggestions about the manuscript of this paper.

* On leave from: Mathematical Institute, Czechoslovak Academy of Sciences, Zintá 25, 115 67 Praha-1, Czechoslovakia.

0168-0072/90/$03.50

1. Preliminaries

For the definition of S_2, V_2 and their fragments see [1]—we assume knowledge of that paper.

$$L_2 = \{0, 1, s, +, \cdot, \lfloor x/2 \rfloor, |x|, x \# y, \leq, =\}$$

is the language of S_2. L_1 denotes the language L_2 without the function symbol #, i.e. $L_1 = L_2 \setminus \{\#\}$.

S_1 is a theory axiomatized by those axioms and rules of S_2 which do not contain #. In other words, S_1-proofs are S_2-proofs consisting only of L_1-formulas.

Definition 1.1. Let $A(\bar{a})$ be a bounded L_1-formula all whose free variables are among $\bar{a}$. By induction on the logical complexity of A we define an L_1-term $V_A(\bar{a})$:

(i) $A(\bar{a})$ is an atomic formula of the form $t_1(\bar{a}) = t_2(\bar{a})$ or $t_1(\bar{a}) \leq t_2(\bar{a})$. Put:

$$V_A(\bar{a}) := t_1(\bar{a}) + t_2(\bar{a}).$$

(ii) $A(\bar{a})$ is of the form $\neg B(\bar{a})$. Put:

$$V_A(\bar{a}) := V_B(\bar{a}).$$

(iii) $A(\bar{a})$ is of the form $B(\bar{a}) \wedge C(\bar{a})$, $B(\bar{a}) \vee C(\bar{a})$ or $B(\bar{a}) \supset C(\bar{a})$. Put:

$$V_A(\bar{a}) := V_B(\bar{a}) + V_C(\bar{a}).$$

(iv) $A(\bar{a})$ is of the form $\exists x \leq t(\bar{a})\, B(x, \bar{a})$ or $\forall x \leq t(a)\, B(x, \bar{a})$. Put:

$$V_A(\bar{a}) := V_B(x/t(\bar{a}), \bar{a}).$$

The intention of the definition is that in order to evaluate the truth value of $A(\bar{a})$ one has to compute only numbers $\leq V_A(\bar{a})$. The following is essentially a presentation of results of [2, 4].

Assume $A(\bar{a})$ has the form

$$\forall x_1 \leq t_1(\bar{a})\, \exists y_1 \leq s_1(\bar{a}, x_1) \cdots \forall x_k \leq t_k(\bar{a}, x_1, \ldots, x_{k-1}, y_1, \ldots, y_{k-1})$$
$$\exists y_k \leq s_k(\bar{a}, x_1, \ldots, x_k, y_1, \ldots, y_{k-1})\, B(\bar{a}, \bar{x}, \bar{y}),$$

B quantifier free. Then $A(\bar{a})$ is true iff there exist Skolem functions $f_1(\bar{a}, x_1), \ldots, f_k(\bar{a}, x_1, \ldots, x_k)$ such that:

(a) for $y_j := f_j(\bar{a}, x_1, \ldots, x_j)$ there holds:

if $x_1 \leq t_1(\bar{a})$ then y_1 is defined and $y_1 \leq s_1(\bar{a}, x_1)$,
if $x_2 \leq t_2(\bar{a}, x_1, y_1)$ then y_2 is defined and $y_2 \leq s_2(\bar{a}, x_1, x_2, y_1)$,
.

if $x_k \leq t_k(\bar{a}, x_1, \ldots, x_{k-1}, y_1, \ldots, y_{k-1})$ then y_k is defined and $y_k \leq s_k(\bar{a}, x_1, \ldots, x_k, y_1, \ldots, y_{k-1})$,

and

(b) $B(\bar{a}, \bar{x}, y_j/f_j)$ is true.

By (a) all these functions assume values $\leq V_A(\bar{a})$. Thus the k-tuple $(f_1, \ldots, f_k)$ can be coded $\leq 2^{V_A(\bar{a})^{2k}}$, in the sense of [4].

Thus $A(\bar{a})$ is true iff "$\exists f_1, \ldots, f_k \leq 2^{V_A(\bar{a})^{2k}}$, $f_1, \ldots, f_k$ are functions and (a), (b) above are satisfied".

As L_1-terms can be evaluated in S_2^1, there is a Δ_1^b (w.r.t. S_2^1) definition of truth for open L_1-formulas. In formula "..." above there is hidden universal quantification over x_j's and existential quantification over y_j's coming from (a) above. But as $x_j, y_j \leq V_A(\bar{a})$, $x_j, y_j \leq |2^{V_A(\bar{a})^{2k}}|$ so these quantifiers are sharply bounded. Thus the formula "..." above is Σ_1^b in S_2^1 in parameter $2^{V_A(\bar{a})^{2k}}$. As the same holds for $\neg A$, formula "..." is Δ_1^b in S_2^1.

Let us summarize the discussion in a lemma. For details of the truth definition see [2, 4, 7].

Lemma 1.2. *There exists a formula* TR(x, y, z) *which is* Δ_1^b *w.r.t.* S_2^1 *and is such that* S_2^1 *proves:*

"*if* $e \geq 2^{V_A(\bar{a})^{2k}}$ *then* TR$(A, \langle\bar{a}\rangle, e)$ *satisfies Tarski's truth conditions*".

More precisely, "..." *reads as follows:*

"*if* V_A *and* k *are defined from* A *as above and* $e \geq 2^{V_A(\bar{a})^{2k}}$ *then:*
if $A = \neg B$ *then* $\mathrm{TR}(A, \langle\bar{a}\rangle, e) \equiv \neg\mathrm{TR}(B, \langle\bar{a}\rangle, e)$,
if $A = B \wedge C$ *then* $\mathrm{TR}(A, \langle\bar{a}\rangle, e) \equiv \mathrm{TR}(B, \langle\bar{a}\rangle, e) \wedge \mathrm{TR}(C, \langle\bar{a}\rangle, e)$,
if $A = (\exists x \leq t(a)\, B(x, \bar{a}))$ *then* $\mathrm{TR}(A, \langle\bar{a}\rangle, e)$
$\equiv \exists x \leq \mathrm{val}(t(\bar{a}))\mathrm{TR}(B, \langle x, \bar{a}\rangle, e)$".

2. V_2^i and exponentiation

Consider first case $i = 1$. We define a theory $S_2^1 + 1$-Exp which is a special case of theories considered in [3].

Definition 2.1. For a formula $A(\bar{a})$,

d is an $S_2^1 + 1$-Exp-*proof of* $A(\bar{a})$ (denoted $d : S_2^1 + 1\text{-Exp} \vdash A(\bar{a})$)

iff

d is an S_2^1-proof of a sequent of the form: $t(\bar{a}) < |c| \rightarrow A(\bar{a})$,

c a free variable not occurring in t or A.

Definition 2.2. $\mathfrak{M} = (\mathfrak{M}_1, \mathfrak{M}_2)$ is a 1-*fold* model of S_2^1 iff
(i) $\mathfrak{M}_1 \models S_2^1$, $\mathfrak{M}_2 \models S_2^1$,
(ii) $\mathfrak{M}_1 \subseteq_e \mathfrak{M}_2$,
(iii) $2^{\mathfrak{M}_1} \subseteq \mathfrak{M}_2$ (i.e. $\forall m \in \mathfrak{M}_1\, \exists n \in \mathfrak{M}_2\; \mathfrak{M}_2 \models m < |n|$).

A 1-fold model $\mathfrak{M}$ is *large* 1-*fold* if there holds moreover:

(iv) $\exists c \in \mathfrak{M}_2\, \forall m \in \mathfrak{M}_1\, \exists n \in \mathfrak{M}_2\, \mathfrak{M}_2 \models m < |n|\ \&\ n < c$.

Lemma 2.3. *Let* $A(\bar{a})$ *be a bounded formula. Then* (i), (ii) *and* (iii) *are equivalent.*

(i) $S_2^1 + 1\text{-Exp} \vdash A(\bar{a})$.

(ii) *For any* 1-*fold model* $\mathfrak{M} = (\mathfrak{M}_1, \mathfrak{M}_2)$ *of* S_2^1, $\mathfrak{M}_1 \models \forall \bar{x}\, A(\bar{x})$.

(iii) *For any large* 1-*fold model* $\mathfrak{M} = (\mathfrak{M}_1, \mathfrak{M}_2)$, $\mathfrak{M}_1 \models \forall \bar{x}\, A(\bar{x})$.

Proof. (i)$\Rightarrow$(ii). Assume $S_2^1 + 1\text{-Exp} \vdash A(\bar{a})$, i.e.

$$S_2^1 \vdash t(\bar{a}) < |c| \to A(\bar{a}).$$

As $\mathfrak{M}_2 \models S_2^1$,

$$\mathfrak{M}_2 \models t(\bar{a}) < |c| \to A(\bar{a}).$$

By $2^{\mathfrak{M}_1} \subseteq \mathfrak{M}_2$ we have for any $\bar{m} \subseteq \mathfrak{M}$ an element $n \in \mathfrak{M}_2$ such that

$$\mathfrak{M}_2 \models t(\bar{m}) < |n|.$$

Thus for all $\bar{m} \subseteq \mathfrak{M}$, $\mathfrak{M}_2 \models A(\bar{m})$. As $\mathfrak{M}_1 \subseteq_e \mathfrak{M}_2$,

$$\mathfrak{M}_1 \models \forall \bar{x}\, A(\bar{x}).$$

Not (i)$\Rightarrow$not (iii). Assume that for any term $t(\bar{a})$, $S_2^1 + t(\bar{a}) < |c| + \neg A(\bar{a})$ is consistent. By compactness, the theory (with $\bar{a}$, c as constants)

$$S_2^1 + \neg A(\bar{a}) + \{t(\bar{a}) < |c| \mid t \text{ a term}\}$$

is consistent. Let $\mathfrak{M}_2$ be a model of this theory, $\bar{a}, c \subseteq \mathfrak{M}_2$. Define

$$\mathfrak{M}_1 = \{m \in \mathfrak{M}_2 \mid \text{for some term } t,\ \mathfrak{M}_2 \models m \leqslant t(\bar{a})\}.$$

Then the pair $(\mathfrak{M}_1, \mathfrak{M}_2)$ contradicts (iii). As (ii)$\Rightarrow$(iii) is trivial, we are done. □

Let $\Sigma_\infty^{0,b}$ denote the class $\bigcup_i \Sigma_i^b$, the class of first-order bounded formulas. $\Sigma_\infty^{0,b}$ is a proper subclass of $\Sigma_0^{1,b}$, the class of bounded second-order formulas without second-order quantifiers, cf. [1].

Lemma 2.4. *Let* $A(\bar{a})$ *be a* $\Sigma_\infty^{0,b}$-*formula. Then*

$$S_2^1 + 1\text{-Exp} \vdash A(\bar{a}) \quad \textit{iff} \quad V_2^1 \vdash A(\bar{a}).$$

Proof. Recall that V_2^1 is (fully) conservative over $\tilde{V}_2^1$ (a version of V_2^1 without second-order function variables), cf. [1].

(1) Assume $V_2^1 \nvdash A(\bar{a})$, i.e. $\tilde{V}_2^1 \nvdash A(\bar{a})$. Thus there is a model $(\mathfrak{M}, \mathfrak{X})$ such that for some $\bar{m} \subseteq \mathfrak{M}$:

$$(\mathfrak{M}, \mathfrak{X}) \models \tilde{V}_2^1 + \neg A(\bar{m}).$$

Claim. *There is a model $\mathfrak{M}^1$ of S_2^1 such that $\mathfrak{M} \subseteq_e \mathfrak{M}^1$ and $2^{\mathfrak{M}} \subseteq \mathfrak{M}^1$.*

Proof of Claim. The idea—developed in [7]—is to use pairs of the form (a, α), $a \in \mathfrak{M}$, $\alpha \in \mathfrak{X}$, to code numbers with value $\sum_{i<a, i\in\alpha} 2^i$. We shall use only pairs (a, α) with α bounded, i.e. $u \in \alpha$ implies $u \leq v$ for some v and all u.

In [7] it was shown that there are $\Delta_1^{1,b}$-definable relations $R_=((a_1, \alpha_1), (a_2, \alpha_2))$, $R_\leq((a_1, \alpha_1), (a_2, \alpha_2))$ and for any f a function symbol of L_2, $F_f((a_1, \alpha_1), \ldots, (a_{n+1}, \alpha_{n+1}))$, where n is the arity of f, such that if $R_=$ resp. $R_\leq$ interprets "$(a_1, \alpha_1) = (a_2, \alpha_2)$" resp. "$(a_1, \alpha_1) \leq (a_2, \alpha_2)$" and F_f interprets "$f((a_1, \alpha_1), \ldots, (a_n, \alpha_n)) = (a_{n+1}, \alpha_{n+1})$" then $\mathring{U}_2^1$ proves the translation of BASIC and of the equality axioms.

As we deal only with pairs (a, α) such that α is bounded we need only bounded $\Delta_1^{1,b}$-CA and not full $\Delta_1^{1,b}$-CA of $\mathring{U}_2^1$. Any instance of bounded $\Delta_1^{1,b}$-CA:

$$\exists \sigma \forall x < a\, x \in \sigma \equiv A(x),$$

can be proved by $\Sigma_1^{1,b}$-IND on a, i.e. in V_2^1. Thus we can prove the translation of the basic properties of function and relation symbols of L_2, as well as the translations of axioms of BASIC and of equality axioms, in V_2^1.

In this translation the original numbers of $\mathfrak{M}$ are best represented as pairs $(|m|, \alpha_m)$ where $\alpha_m = \{i_0 < \cdots < i_k\}$ such that $m = 2^{i_0} + \cdots + 2^{i_k}$.

Let $\mathfrak{M}^1$ be the structure $\mathfrak{M} \times \mathfrak{X}/R_=$ with relation $\leq$ and functions $f \in L_2$ interpreted according to $R_\leq$ and F_f. We claim that $\mathfrak{M} \subseteq_e \mathfrak{M}^1$ (i.e. $\mathfrak{M}$ is isomorphic to an initial segment of $\mathfrak{M}^1$, $2^{\mathfrak{M}} \subseteq \mathfrak{M}^1$ and $\mathfrak{M}^1 \models S_2^1$).

For $\mathfrak{M} \subseteq_e \mathfrak{M}^1$ it is essentially only needed to prove:

$$(b, \beta)\, R_\leq (|a|, \alpha_a) \Rightarrow \exists c \leq a\, (b, \beta)\, R_= (|c|, \alpha_c)$$

and

$$F_f((|a_1|, \alpha_{a_1}), \ldots, (|a_{n+1}|, \alpha_{a_{n+1}})) \Rightarrow f(a_1, \ldots, a_n) = a_{n+1}.$$

This is proved by induction ($\Delta_1^{1,b}$-IND) on a resp. on $a_1 + \cdots + a_{n+1}$.

Condition $2^{\mathfrak{M}} \subseteq \mathfrak{M}^1$ is easy as the pair $(a+1, \{a\})$ represents a number greater than "$2^{(|a|, \alpha_a)}$". This is proved by induction ($\Delta_1^{1,b}$-IND) on a using the formula:

$$F_+((b+1, \{b\}), (b+1, \{b\}), (b+2, \{b+1\})).$$

To see that $\mathfrak{M}^1 \models S_2^1$ take a Σ_1^b-formula $A(a, \bar{b})$. We construct a translation of the formula A into a $\Sigma_1^{1,b}$-formula

$$A^*((a, \alpha), (\overline{b, \beta}))$$

such that for (m, μ) and (n_i, η_i) from $\mathfrak{M}^1$,

$$\mathfrak{M}^1 \models A((m, \mu), (\overline{n_1, \eta_i})) \quad \text{iff} \quad (\mathfrak{M}, \mathfrak{X}) \models A^*((m, \mu), (\overline{n_i, \eta_i})).$$

Translate functions and relations according to $R_=$, $R_\leq$ and F_f's. Translation $*$ commutes with propositional connectives. Quantifiers are translated as follows:

(a) $(\exists x \leq t(a_1, \ldots)\, B(x, a_1, \ldots))^*$

$$= \exists x \leq t^1(a_1, \ldots)\, \exists \sigma \text{ "}t((a_1, \alpha_1), \ldots) \geq (x, \sigma)\text{" } \wedge B_1^*$$

where term t^1 is chosen such that

$$b \leqslant t(\bar{a}) \to |b| \leqslant t^1(|a_1|, \ldots),$$

(b) $$(\forall x \leqslant |t(a_1, \ldots)|\, B(x, \bar{a}))^* = \forall x \leqslant t^1(\bar{a})\, \exists \sigma\, \text{“}(|x|, \sigma) = x\text{”} \wedge B^*((|x|, \sigma)),$$

(c) $$(\exists x \leqslant |t(\bar{a})|\, B(x, \bar{a}))^* = \exists x \leqslant t^1(\bar{a})\, \exists \sigma\, \text{“}(|x|, \sigma) = x\text{”} \wedge B^*((|x|, \sigma))$$

where t^1 in (b), (c) has the same properties as in (a).

To show that the translation in (b), (c) is correct one needs:

$$\forall x\, \exists \sigma\, \text{“}(|x|, \sigma) = x\text{”}.$$

This is proved by $\Sigma_1^{1,b}$-PIND on x.

We may assume that A is in a prenex form (as it is sufficient to verify PIND only for prenex formulas). A^* is then a $\Sigma_1^{1,b}$-formula.

Assume:

$$\mathfrak{M}^1 \models A(0, \overline{(n, \eta)}) \wedge \forall x\, A(\lfloor x/2 \rfloor, \overline{(n, \eta)}) \to A(x, \overline{(n, \eta)}).$$

Then (let us forget the parameters $\overline{(n, \eta)}$)

$$(\mathfrak{M}, \mathfrak{X}) \models A^*((1, \{0\})) \wedge \forall (x, \sigma)\, A^*(\lfloor \text{“}(x, \sigma)/2\text{”} \rfloor) \to A^*((x, \sigma)).$$

Assume also:

(*) $$(\mathfrak{M}, \mathfrak{X}) \models \neg A^*((k, \kappa)),$$

for some $(k, \kappa) \in \mathfrak{M}^1$. As

$$\text{“}\lfloor (x, \kappa)/2 \rfloor = (x - 1, \kappa) \vee (x, \kappa) = (x - 1, \kappa)\text{”} \quad \text{and} \quad \text{“}(1, \kappa) = (1, \{0\})\text{”},$$

the formula above implies:

$$(\mathfrak{M}, \mathfrak{X}) \models A^*((1, \kappa)) \wedge \forall x \leqslant k\, A^*((x, \kappa)) \to A^*((x + 1, \kappa)).$$

Thus, by $\Sigma_1^{1,b}$-IND in $(\mathfrak{M}, \mathfrak{X})$, we have:

$$(\mathfrak{M}, \mathfrak{X}) \models A^*((k, \kappa)),$$

contradicting (*). So

$$(\mathfrak{M}, \mathfrak{X}) \models \forall x\, \forall \sigma\, A^*((x, \sigma)), \quad \text{i.e.} \quad \mathfrak{M}^1 \models \forall x\, A(x).$$

This proves the claim. (Let us remark that a different translation of bounded formulas was used in [7].) The claim together with Lemma 2.3 implies:

$$S_2^1 + 1\text{-Exp} \nvdash A(\bar{a}).$$

(2) Assume now $S_2^1 + 1\text{-Exp} \nvdash A(\bar{a})$. By Lemma 2.3 there is a large 1-fold model of S_2^1, $\mathfrak{M} = (\mathfrak{M}_1, \mathfrak{M}_2)$, such that for some $\bar{m} \subseteq \mathfrak{M}_1$,

$$\mathfrak{M}_1 \models \neg A(\bar{m}).$$

Let $c \in \mathfrak{M}_2$ witness condition (iv) of Definition 2.2, i.e. $2^{\mathfrak{M}_1} < c$.
Take $\mathfrak{X} = \{\alpha \subseteq \mathfrak{M}_1 \mid \alpha \text{ is coded (in } \mathfrak{M}_2) \leqslant c\}$.

Claim. $(\mathfrak{M}_1, \mathfrak{X}) \models V_2^1$.

Let B be a second-order bounded formula. Translate B into B^{**}, a $\Sigma_\infty^{0,b}$-formula with parameter c, as follows:

(a) First-order relations and functions are left unchanged.

(b) $x \in \sigma$ is translated as "x is an element of the set coded by σ", "$.\ .$" is a $\Sigma_\infty^{0,b}$-formula, cf. [2],

(c) ** commutes with Boolean connectives,

(d) $(\forall x \leqslant t\, B(x))^{**} = \forall x \leqslant |c|\, (x \leqslant t \rightarrow B(x))^{**}$,

$(\exists x \leqslant t\, B(x))^{**} = \exists x \leqslant |c|\, (x \leqslant t \wedge B(t))^{**}$,

(e) $(\exists \sigma\, B(\sigma))^{**} = \exists \sigma \leqslant c\, (B(\sigma))^{**}$,

$(\forall \sigma\, B(\sigma))^{**} = \forall \sigma \leqslant c\, (B(\sigma))^{**}$.

Now let A be a $\Sigma_1^{1,b}$-formula. So A^{**} is a Σ_1^b-formula. Assume:

$$(\mathfrak{M}_1, \mathfrak{X}) \models A(0) \wedge \forall x\, A((x)) \rightarrow A((x+1)).$$

Thus for all $u \in 2^{\mathfrak{M}_1}$,

$$\mathfrak{M}_2 \models A^{**}(0) \wedge (A^{**}(|u|) \rightarrow A^{**}(|u|+1)).$$

That is, for all $u \in 2^{\mathfrak{M}_1}$,

$$\mathfrak{M}_2 \models A^{**}(0) \wedge (A^{**}(|\lfloor u/2 \rfloor|) \rightarrow A^{**}(|u|)).$$

As A^{**} is Σ_1^b and $\mathfrak{M}_2 \models S_2^1$, we have for all $u \in 2^{\mathfrak{M}_1}$:

$$\mathfrak{M}_2 \models A^{**}(|u|).$$

But this implies:

$$(\mathfrak{M}_1, \mathfrak{X}) \models \forall x\, A(x), \quad \text{i.e.} \quad (\mathfrak{M}_1, \mathfrak{X}) \models S_2(\alpha) + \Sigma_1^{1,b}\text{-IND}.$$

It remains to show that

$$(\mathfrak{M}_1, \mathfrak{X}) \models \Sigma_0^{1,b}\text{-CA}.$$

This reduces to show that for any Σ_0^b-formula $A(a)$ (with parameters in $\mathfrak{M}_2$): there is $d \in \mathfrak{M}_2$, $\mathfrak{M}_2 \models d \leqslant c$ such that

$$\mathfrak{M}_2 \models \forall x \leqslant |c|,\ A(x) \equiv \text{"}x \text{ is an element of the set coded by } d\text{"}.$$

This is proved by PIND on a in the following Σ_1^b-formula:

$$\exists d \leqslant c\ \forall x \leqslant |a|\ A(x) \equiv \text{"}x \in d\text{"}.$$

This completes the proof of the claim.

From the claim it follows that $V_2^1 \nvdash A(\bar{a})$. This proves the lemma. □

In the same way the following theorem is proved ($S_2^i + 1$-Exp is defined completely analogically with Definition 2.1).

Theorem 2.5. *Let $i \geq 1$ and let $A(\bar{a})$ be a first-order formula without any occurrence of a second-order variable, i.e. a $\Sigma_\infty^{0,b}$-formula. Then*

$$V_2^i \vdash A(\bar{a}) \quad \textit{iff} \quad S_2^i + 1\text{-Exp} \vdash A(\bar{a}).$$

Corollary 2.6. *For all $i, j \geq$*

$$\textit{if } S_2^i = S_2^j \quad \textit{then} \quad V_2^i \underset{\Sigma_\infty^{0,b}}{\equiv} V_2^j.$$

Corollary 2.7. *For $i \geq 1$, V_2^i is not Π_1^b-conservative over S_2^i. Also V_2 is not Π_1^b-conservative over S_2.*

Proof. Theory $S_2^i + \text{Exp}$ is equal to the theory $I\Delta_0 + \text{Exp}$ of [5]. There it was shown that $I\Delta_0 + \text{Exp}$ proves certain consistency statements (i.e. Π_1^b-formulas) unprovable in $I\Delta_0 + \Omega_1$, which is equivalent to S_2. Hence in particular, $S_2^i + \text{Exp}$ is not Π_1^b-conservative over S_2^i which immediately implies that neither is $S_2^i + 1$-Exp. This entails the corollary. □

Corollary 2.7 extends a result from [7] where it was shown that $\mathring{V}_2^1$ is not conservative over S_2^1.

Corollary 2.8. *For $i \geq 1$,*

$$V_2^i \underset{\Sigma_\infty^{0,b}}{\equiv} V_2^i(\text{BD}).$$

Proof. In the construction of $\mathfrak{M}^1$ in the proof of Lemma 2.4, only the assumption $\mathfrak{M} \vDash V_2^1(\text{BD})$ is actually used. □

Let $\sqrt{S_2^i}$ denote a theory arising from S_2^i by replacing Σ_i^b-PIND by the rule

$$\frac{A(\lfloor\sqrt{a}\rfloor),\ \Gamma \rightarrow \Delta,\ A(a)}{A(0),\ \Gamma \rightarrow \Delta,\ A(t)}$$

($\sqrt{S_2^i}$ implies the soundness of the following rule which may be called Σ_i^b-LLIND:

$$\frac{A(a),\ \Gamma \rightarrow \Delta,\ A(a+1)}{A(0),\ \Gamma \rightarrow \Delta,\ A(\|a\|)}$$

Then analogically with the first part of the proof of Lemma 2.4 we have (recall $U_2^1 \vdash \Delta_1^{1,b}$-IND, cf. [1]):

$$U_2^i + \text{bounded } \Delta_1^{1,b}\text{-CA} \vdash \sqrt{S_2^i} + 1\text{-Exp}.$$

Another way to generalize the method of this section is to consider higher-order extensions of S_2 (based on induction and appropriate comprehension axioms for variables of higher orders). Then similarly the bounded first-order consequences of the $(k+1)$-th order extension are characterized as $S_2 + k$-Exp. (That is, $S_2 + k\text{-Exp} \vdash A(a)$ iff

$$S_2 \vdash t(a) < |c_1|, c_1 < |c_2|, \ldots, c_{k-1} < |c_k| \to A(a).)$$

The characterization of bounded first-order consequences of fragments of such theories is more complicated and needs further weakenings of the induction rule in the line of LIND. LLIND. . . . .

3. Another corollary to the construction

In connection with the problem of existence of end-extensions to models of $\mathrm{I}\Delta_0$ the question whether there is a model of $\mathrm{I}\Delta_0$ satisfying neither Exp nor $\mathrm{B}\Sigma_1^0$ was posed in [6]. (Exp is a Π_2^0-formula $\forall x\, \exists y\, x = |y|$.) A modest contribution to this problem is the next theorem.

Theorem 3.1. *Let* $\mathcal{M} = (\mathfrak{M}, \mathfrak{X})$ *be a model of* V_2 *not satisfying* Exp, *i.e.*

$$\mathcal{M} \models V_2 + \neg\mathrm{Exp}.$$

Then $\mathfrak{M} \models \mathrm{B}\Sigma_1^0$.

Proof. Take model $\mathfrak{M}^1$ of S_2 (extending $\mathfrak{M}$) constructed from $(\mathfrak{M}, \mathfrak{X})$ in the proof of Lemma 2.4. Thus $\mathfrak{M} \subseteq_e \mathfrak{M}^1$ and $2^{\mathfrak{M}} \subseteq \mathfrak{M}^1$.

From the assumption $\mathfrak{M} \models \neg\mathrm{Exp}$ it follows that $\mathfrak{M}^1$ is a proper end-extension of $\mathfrak{M}$. It follows easily that $\mathfrak{M} \models \mathrm{B}\Sigma_1^0$, cf. [6]. □

4. A restricted provability notion

Definition 4.1. (1) D is a *restricted* S_2*-proof of* A (denoted $D : S_2 \vdash_{\mathrm{R}} A$) iff the following conditions hold: D is a 5-tuple $D = \langle d, \bar{w}, \bar{v}, \overline{d'}, \overline{d''}\rangle$, and:

(i) d is an S_1-proof of a sequent of the form:

$$2 \leq c_0, |a_0| \leq |c_0|, \ldots, |a_n| \leq |c_0|, |c_0|\,|c_0| \leq |c_1|,$$
$$|c_1|\,|c_0| \leq |c_2|, \ldots, |c_{j-1}|\,|c_0| \leq |c_j| \to A,$$

where $c_0, \ldots, c_j$ do not occur in A. In particular, A is an L_1-formula.

(ii) All formulas in d are bounded and in prenex form.

(iii) If $\bar{a} = (a_0, \ldots, a_n)$ and $\bar{c} = (c_0, \ldots, c_j)$ are all parameters of d (i.e. free variables of the end sequent) and $\bar{b} = (b_0, \ldots, b_k)$ are all other free variables occurring in d then it holds $(l, m \leq k)$:

(a) the sequents of d where b_l occurs form a connected subtree of d,

(b) if the elimination rule of b_l is below the elimination rule of b_m then $l<m$,

(c) the elimination rate of b_l is either $\forall\leqslant$:right, $\exists\leqslant$:left or PIND.

(iv) $\bar{w}=\langle w_0(\bar{a},\bar{c}),\dots,w_k(\bar{a},\bar{c})\rangle$ is a sequence of L_1-terms such that for $l\leqslant k$ it holds: if the elimination rule of b_l has the form:

$$\frac{A(|b_l/2|),\Gamma\to\Delta,A(b_l)}{A(0),\Gamma\to\Delta,A(t_l(\bar{a},\bar{c},b_0,\dots,b_{l-1}))}$$

or

$$\frac{b_l\leqslant t_l(\bar{a},\bar{c},b_0,\dots,b_{l-1}),A(b_l),\Gamma\to\Delta}{\exists x\leqslant t_l(\bar{a},\bar{c},b_0,\dots,b_{l-1})A(x),\Gamma\to\Delta}$$

or

$$\frac{b_l\leqslant t_l(\bar{a},\bar{c},b_0,\dots,b_{l-1}),\Gamma\to\Delta,A(b_l)}{\Gamma\to\Delta,\forall x\leqslant t_l(\bar{a},\bar{c},b_0,\dots,b_{l-1})A(x)}$$

then

$(*)_l$ $\qquad w_l(\bar{a},\bar{c})\geqslant t_l(\bar{a},\bar{c},b_0/w_0,\dots,b_{l-1}/w_{l-1}).$

(v) $\bar{d'}=\langle d'_0,\dots,d'_k\rangle$ is a sequence of proofs such that d'_l is a quantifier-free and induction-free S_1-proof of $(*)_l$.

(vi) $\bar{v}$ is a sequence of L_1-terms such that if a formula $B(\bar{a},\bar{b},\bar{c},\bar{x})$ (with $\bar{a},\bar{b},\bar{c},\bar{x}$ variables free in B) which occurs as a subformula in d (we consider B associated with its occurrence), then the sequence $\bar{v}$ contains the L_1-term $V_B(\bar{a},\bar{b},\bar{c},\bar{x})$ defined in Section 1.

(vii) For $A(\bar{a},\bar{b},\bar{c})$ a formula of the form

$$Q_1x_1\leqslant t_1(\bar{a},\bar{b},\bar{c})\,Q_2x_2\leqslant t_2(\bar{a},\bar{b},\bar{c},x_1),\dots,$$
$$Q_rx_r\leqslant t_r(\bar{a},\bar{b},\bar{c},x_1,\dots,x_{r-1})\,C(\bar{a},\bar{b},\bar{c},\bar{x}),$$

C quantifier free which occurs in d, $\bar{v}$ contains terms $P_{A,i}(\bar{a},\bar{c})$ ($i=1,\dots,r$) and a term $q_A(\bar{a},\bar{c})$ such that the following holds:

$(**)_{A,1}$ $\qquad P_{A,1}(\bar{a},\bar{c})\geqslant t_1(\bar{a},b_l/w_l,\bar{c}),$

$(**)_{A,2}$ $\qquad P_{A,2}(\bar{a},\bar{c})\geqslant t_2(\bar{a},b_l/w_l,\bar{c},x_1/P_{A,1}),$

$\vdots$

$(**)_{A,r}$ $\qquad P_{A,r}(\bar{a},\bar{c})\geqslant t_r(\bar{a},b_l/w_l,\bar{c},x_1/P_{A,1},\dots,x_{r-1}/P_{A,r-1})$

and

$(***)_A$ $\qquad q_A(\bar{a},\bar{c})\geqslant V_C(\bar{a},b_l/w_l,\bar{c},x_j/P_{A,j}).$

(viii) $\bar{d''}$ is a sequence of quantifier-free and induction-free S_1-proofs and for all A occurring in d and $i\leqslant$ the quantifier complexity of A, $\bar{d''}$ contains proofs $d''_{A,i}$ of $(**)_{A,i}$ and d''_A of $(***)_A$.

(2) The number j in (i) (= the number of formulas of the form $|c_{i-1}|\,|c_0|\leqslant|c_i|$ in the antecedent of the end-sequent of d) is called *the dimension* of D and denoted $\dim(D)$.

(3) A restricted S_2-proof D is called *strictly restricted* if it holds:

$$\dim(D)\leqslant\|D\|.$$

We denote by $D:S_2 \vdash_{\mathrm{SR}} A$ the formula

"$D:S_2 \vdash_{\mathrm{R}} A$ and D is strictly restricted".

This provability notion is motivated by [3, 7]. Recall that by [2, 3, 5] the formula "$D:S_2 \vdash_{\mathrm{SR}} A$" can be chosen as L_1-formula, Δ_1^b w.r.t. S_2^1. We also assume that the formula is in a prenex form.

Lemma 4.2. *For any bounded L_1-formula it holds.*

$$S_2 \vdash A \quad \textit{iff} \quad S_2 \vdash_{\mathrm{SR}} A.$$

Proof. The 'if' part is true as S_2 proves sequents of the form

$$\to \exists x \; |c_{i-1}| \; |c_0| \leqslant x.$$

The 'only if' part follows essentially by cut-elimination and a compactness argument, cf. [3, 7]. Recall that PIND can be proved only from its instances for prenex formulas. □

The following two lemmas are usual probability conditions needed for the proof of Gödel's theorem.

Lemma 4.3. $S_2^1 \vdash (D:S_2 \vdash_{\mathrm{SR}} B) \to \exists D_1 \leqslant t(D)[(D_1:S_2 \vdash_{\mathrm{SR}} (D:S_2 \vdash_{\mathrm{SR}} B))]$ *for some fixed L_2-term t.*

Proof. Argue in S_2^1. As $D:S_2 \vdash_{\mathrm{SR}} B$ is a true $(\Delta_1^b \cap L_1)$-sentence there is an S_1-proof d of it with length

$$|d| \leqslant |(D:S_2 \vdash_{\mathrm{SR}} B)|^{k'} \leqslant |D|^k$$

for some fixed $k' \leqslant k < \omega$ — cf. [5]. Moreover, all formulas in d are substitution instances of formulas with Gödel number $\leqslant k''$, for some fixed $k'' < \omega$ (cf. [5]). Thus there exist terms $\bar{w}$, $\bar{v}$, instances of some standard iterations of terms, needed for the strictly restricted proof. (As d is an S_1-proof with empty antecedent, $\dim d = 0$.) □

Lemma 4.4.

$$S_2^1 \vdash [(D_1:S_2 \vdash_{\mathrm{SR}} A, \Gamma \to \Delta) \wedge (D_2:S_2 \vdash_{\mathrm{SR}} \Gamma' \to \Delta', A)] \to$$
$$\exists D_3 \leqslant t(D_1, D_2)\; (D_3:S_2 \vdash_{\mathrm{SR}} \Gamma, \Gamma' \to \Delta, \Delta'),$$

for some fixed L_2-term t.

Proof. Argue in S_2^1. There exists an obvious restricted proof D_3: join D_1 and D_2 by an application of cut-rule.

Clearly:

$$\dim D_3 \leqslant \dim D_1 + \dim D_2.$$

Thus it is sufficient to add to D_3 some dummy inferences to get D_4 such that the

Gödel number of D_4 fulfills

$$2(D_1 \# D_2)^2 \leqslant D_4.$$

So

$$\|D_4\| \geqslant \|2(D_1 \# D_2)^2\| \geqslant |2\,|D_1 \# D_2|\,| \geqslant 1 + \|D_1 \# D_2\| \geqslant 1 + |1 + |D_1|\,|D_2|\,| \geqslant 1 + (\|D_1\| + \|D_2\| - 1) \geqslant \|D_1\| + \|D_2\|.$$

Hence

$$\|D_4\| \geqslant \|D_1\| + \|D_2\| \geqslant \dim D_1 + \dim D_2 \geqslant \dim D_3 = \dim D_4.$$

Thus D_4 is the required strictly restricted proof. □

5. V_2^1 versus S_2

In this section we investigate the problem whether V_2^1 is conservative over S_2.

Definition 5.1. A *dyadic numeral of n*, denoted $\underline{n}$, is defined:

$$\underline{0} := 0, \quad \underline{1} := 1, \quad \underline{2} := (1+1),$$
$$\underline{2n} := (\underline{2} \cdot \underline{n}) \quad \text{and} \quad \underline{2n+1} := (\underline{2n} + 1).$$

In the following definition we assume that the formalization is based on dyadic numerals, i.e. Gödel numbers are represented by them. $\underset{\sim}{a}$ denotes a formalization of the dyadic numeral.

Definition 5.2. (a) $\mathrm{SRPr}(a, b)$ is an L_1-formula formalizing:

$$\text{“}\exists D \leqslant a\, (D : S_2 \vdash_{\mathrm{SR}} b)\text{”}.$$

Moreover, SRPr is Δ_1^b w.r.t. S_2^1 and a natural formalization—in the sense of [1, 5, 7]—such that S_2^1 can prove Lemmas 4.3 and 4.4 for the formalization. Cf. [3].

(b) $\mathrm{SRCon}(S_2)(a)$ is an L_1-formula defined as

$$\mathrm{SRCon}(S_2)(a) := \neg\mathrm{SRPr}(a, \ulcorner 0 = 1 \urcorner).$$

Thus $\forall x\, \mathrm{SRCon}(S_2)(x)$ expresses: “S_2 is strictly restricted consistent”.

Lemma 5.3. $S_2^1 + 1\text{-Exp} \vdash \mathrm{SRCon}(S_2)(a)$.

Proof. Assume $D : S_2 \vdash_{\mathrm{SR}} 0 = 1$, i.e. $D = \langle d, \bar{w}, \bar{v}, \overline{d'}, \overline{d''} \rangle$, where d is an S_1-proof of the sequent of the form:

$$2 \leqslant c_0,\ |a_0| \leqslant |c_0|, \ldots, |a_n| \leqslant |c_0|,$$
$$|c_0|\,|c_0| \leqslant |c_1|, \ldots, |c_{j-1}|\,|c_0| \leqslant |c_j| \rightarrow 0 = 1.$$

By substituting 0 for all a_i in the whole d, and adding a few inferences, we can assume that the end-sequent of d has the form:

$$2 \leq c_0,\ |c_0|\,|c_0| \leq |c_1|, \ldots, |c_{j-1}|\,|c_0| \leq |c_j| \to 0 = 1.$$

Call it $S(\bar{c})$.

Let $\bar{b} = b_0, \ldots, b_k$ be all non-parametrical free variables of d.

By soundness of the rules, using terms $\bar{w}$ guaranteeed by D, prove: "if $S(\bar{c})$ is not true, then there is an initial sequent $S_0(\bar{b}, \bar{c})$ and $b_l \leq w_l(\bar{c})$, $l = 0, \ldots, k$, such that $S_0(\bar{b}, \bar{c})$ is not true". As any initial sequent is true, so must be $S(\bar{c})$.

This argument can be formalized in S^1_2 using the truth definition of Lemma 1.2. Statement "..." is Σ^b_1 and is proved by induction on the number of inferences in d. The only point is to have a number so large that the truth definition of Lemma 1.2 can be applied to all formulas in d. Proofs $\overline{d''}$ are used to verify that terms $\bar{w}$ have been correctly chosen.

Terms in $\bar{v}$ and proofs in $\overline{d''}$ are used for the proof that terms V_A's are defined correctly. Also we have that

$$V_A(\bar{b}, \bar{c}) \leq q_A(\bar{c}) \quad \text{for } b_l \leq w_l(\bar{c}).$$

Thus q_A's, being coded in D, satisfy

$$|q_A| \leq |D|.$$

Also k of Lemma 1.2, i.e. the quantifier complexity of A, satisfies $k \leq |D|$. Thus the equality:

$$2^{q_A(\bar{c})^{2k}} \leq e$$

from Lemma 1.2 follows from:

$$2^{q_A(\bar{c})^{2|D|}} \leq e. \tag{†}$$

As for any L_1-form $t(\bar{c})$, $\mathrm{val}(t(c), \bar{c}) \leq \max(2, \bar{c})^{|t|}$ we have

$$\mathrm{val}(q_A, \bar{c}) \leq \langle \bar{c} \rangle^{|q_A|}.$$

So (†) follows from:

$$2^{\langle \bar{c} \rangle^{2|D|^2}} \leq e.$$

Thus we have:

$$S^1_2 \vdash \langle \bar{c} \rangle^2 \# D \# D < |e| \to [(D : S_2 \vdash_{\mathrm{SR}} 0 = 1) \supset \mathrm{TR}(S(\bar{c}), \langle \bar{c} \rangle, e)].$$

Define $t_0 := \underline{2}$, $t_i := 2(t_{i-1} \cdot t_{i-1})$, for $i \leq j$. Then for $i \leq j$

$$\mathrm{val}(t_i) = 2^{(2^{i+1}-1)} \leq 2^{(2^{\|D\|+1}-1)} \leq s(D),$$

where $s(x)$ is some term.

Thus t_i's can be defined in S^1_2 and $\langle t_0, \ldots, t_j \rangle$ can be coded $\leq s(D)^{\|D\|}$. We now substitute in $S(\bar{c})$, t_i's for c_i's. Adding a few (quantifier free and induction free) inferences we get a strictly restricted proof D' of $0 = 1$ with $\dim D' = 0$. (††)

implies, as $D' \leqslant D \cdot s(D)^{h\|D\|}$ for some $h < \omega$, that for some L_2-term $t(x)$ it holds:

$$S_2^1 \vdash t(D) < |e| \rightarrow [(D : S_2 \vdash_{\mathrm{SR}} 0 = 1) \rightarrow \mathrm{TR}(0 = 1, 0, e)],$$

i.e., using Lemma 1.2,

$$S_2^1 + 1\text{-Exp} \vdash 0 = 1.$$

This is a contradiction. □

Unfortunately we are not able to show that $S_2 \nvdash \mathrm{SRCon}(S_2)(a)$. The lemmas in Section 4 are the usual probability conditions needed for the Gödel theorem but the obstacle to the standard proof is that the strictly restricted provability is not — provably in S_2 — closed under the substitution of numerals for free variables, i.e. S_2 cannot prove that $S_2 \vdash_{\mathrm{SR}} A(a)$ implies $S_2 \vdash_{\mathrm{SR}} A(\underline{n})$, for all n. Thus we have to use another construction giving a weaker result.

Consider a Σ_1^b, #-free formula ϕ such that

$$S_1 \vdash \exists x\, \phi(x) \equiv (\exists d, d : S_2 \vdash_{\mathrm{R}} \neg\phi(a)).$$

By a standard argument (using Lemma 4.2) it follows:

$$S_2 \nvdash \neg\phi(a).$$

Let us look under which conditions V_2^1 could prove $\neg\phi(a)$. Assume $\mathcal{M} = (\mathfrak{M}, \mathfrak{X})$ is a model of V_2^1 and $\exists x\, \phi(x)$, i.e.

$$\mathcal{M} \models V_2^1 + \phi(m),$$

for some $m \in \mathfrak{M}$.

By the definition of ϕ and by Parikh's theorem there is $d \in \mathfrak{M}$, such that:

$$\mathfrak{M} \models d \leqslant m^r \wedge (d : S_2 \vdash_{\mathrm{R}} \neg\phi(a)),$$

for some fixed $r < \omega$.

The end-sequent of d has the form

$$2 \leqslant c_0\, |a| \leqslant |c_0|, \ldots, |c_{j-1}|\, |c_0| \leqslant |c_j| \rightarrow \neg\phi(a).$$

Adding some inferences to d we easily get a proof d_1 of:

$$2 \leqslant \underline{m}, |\underline{m}| \leqslant |\underline{m}|, \ldots, |c_{j-1}|\, |\underline{m}| \leqslant |c_j| \rightarrow \neg\phi(\underline{m}).$$

A code of such a proof d_1 will satisfy $d_1 \leqslant m^{c \cdot j^2}$ (m^r for d, $m^{O(j)}$ for the end-sequent and this itself j-times for its derivation). Thus to guarantee that d_1 exists we need the assumption $j \leqslant |m|^l$, for some $l < \omega$. Then $d_1 \leqslant t_1(m)$ can be assumed for some fixed term t_1.

On the other side, as ϕ is Σ_1^b, $\phi(\underline{m})$ implies that there is a restricted proof d_2 of $\phi(\underline{m})$ (cf. the proof of Lemma 4.3). Again we may assume $d_2 \leqslant t_2(m)$, t_2 some term. Moreover — by the proof of Lemma 4.3 — $\dim d_2 = 0$.

Joining proofs d_1 and d_2 by the cut-rule gives a proof d_3 of:

$$2 \leq \underline{m},\ |\underline{m}| \leq |\underline{m}|, \ldots, |c_{j-1}|\ |\underline{m}| \leq |c_j| \rightarrow .$$

Again $d_3 \leq t_3(m)$, for some fixed term t_3 (cf. Lemma 4.4) obtained from t_1, t_2.

Proof d_3 is not restricted as its end-sequent does not have the appropriate form. We construct from d_3 a restricted proof d_4 by cutting out formulas $2 \leq \underline{m}$, $|\underline{m}| \leq |\underline{m}|$ from the antecedent and replacing formulas $|\underline{m}|\ |\underline{m}| \leq |c_1|$ and $|c_{i-1}|\ |\underline{m}| \leq |c_i|$ there by cedents:

$$2 \leq c_{-k},\ |c_{-k}|\ |c_{-k}| \leq |c_{-k+1}|, \ldots, |c_{-1}|\ |c_{-k}| \leq |c_0|,$$
$$|c_0|\ |c_{-k}| \leq |c_0^1|, \ldots, |c_0^{k-1}|\ |c_{-k}| \leq |c_1|$$

and

$$2 \leq c_{-k},\ |c_{i-1}|\ |c_{-k}| \leq |c_{i-1}^1|,\ |c_{i-1}^1|\ |c_{-k}| \leq |c_{i-1}^2|, \quad , |c_{i-1}^{k-1}|\ |c_{-k}| \leq |c_i|,$$

respectively, where $k := \|m\|$.

This is done quite straightforwardly, the resulting restricted proof d_4 satisfies:

$$d_4 \leq d_3^{c \cdot j}, \qquad \dim d_4 = j \cdot \|m\|,$$

and its end-sequent has the form:

$$2 \leq c_{-k}, \quad , |c_{i-1}^{l-1}|\ |c_{-k}| \leq |c_{i-1}^l|, \quad , |c_{j-1}^{k-1}|\ |c_{-k}| \leq |c_j| \rightarrow \qquad (\dagger)$$

Now we would like to get a contradiction by taking the truth definition of Lemma 2.1 and as in the proof of Lemma 5.3 show that (†) is true for some c_i^l's satisfying the antecedent.

The simplest choice for values of c_i^l's is obviously

$$c_{-k} := 2^{2^1} - 1,\ c_{-k+1} := 2^{2^2} - 1, \ldots, c_j :\equiv 2^{2^{((j+1)\|m\|+1)}} - 1.$$

Hence the whole $((j+1)\ \|m\| + 1)$-tuple would be coded below $u = 2^{2^{((j+1)\cdot\|m\|+2)}}$. Having such u, the use of the truth definition entails the contradiction.

Let us summarize the discussion. We took a diagonal formula ϕ and from the assumption $\mathcal{M} \models \phi(\underline{m})$ we have derived a contradiction under the assumption:

$$\mathcal{M} \models 2^{2^{((\dim d+1)\cdot\|m\|+2)}} \quad \text{exists},$$

where $d \leq m^r$ is the restricted proof of $\neg\phi(a)$ guaranteed by $\phi(m)$. Hence we have to put a suitable restriction on the dimension of proof d, say $\dim d \leq f(d)$. To have an analog of Lemma 4.2 valid we need that $f(d)$ is non-decreasing and eventually greater than any $j < \omega$. To have an analog Lemma 4.3 valid, i.e.

$$((d : S_2 \vdash_{\mathrm{R}} B) \wedge \dim d \leq f(d))$$
$$\rightarrow \exists d_1,\ d_1 : S_2 \vdash_{\mathrm{R}} \ulcorner((\underline{d} : S_2 \vdash_{\mathrm{R}} \underline{B}) \wedge \dim \underline{d} \leq f(\underline{d}))\urcorner,$$

we need that the relation $y \leq f(x)$ is Σ_1^b and #-free definable in S_2^1. An analog of Lemma 4.4 is used only for d_1, d_2 where $\dim d_2 = 0$ and so is always valid.

Furthermore, we used the assumption $j = \dim d \leqslant |d|$ to guarantee the existence of proofs $d_1, \ldots, d_4$. Thus the following assumptions on $f(d)$ are sufficient to carry out the argument:

(i) $S_2^1 \vdash f(d) \leqslant |d|$.
(ii) The graph of f has a Σ_1^b, #-free definition in S_2^1.
(iii) $S_2^1 \vdash$ "f is non-decreasing".
(iv) For all $j < \omega$, $S_2^1 \vdash \exists x\, \forall y > x\, \underline{j} < f(y)$.

We can now state the theorem.

Theorem 5.4. *Let f be a function satisfying the assumptions* (i)–(iv) *above, and define*:

$$g(x) := 2^{2^{(\|x\| \cdot f(x))}}.$$

Then $V_2^1 +$ "g is total" is not Π_1^b-conservative over S_2. In particular, V_3^1 is not Π_1^b-conservative over S_2.

Proof (sketch). The assumptions (i)–(iv) posed on the function $f(x)$ guarantee that we can carry out the argument above. In particular we need that

$$2^{2^{((\dim d+1)\cdot \|d\|+2)}} \quad \text{exists.}$$

This follows from the assumption

$$g(d) = 2^{2^{f(d)\cdot \|d\|}} \quad \text{exists.}$$

The particular case $g(x) \sim x \#_3 x$ is obtained for $f(x) \sim \|x\|$ □

Observe that $g(x)$ can be much slower than $x \#_3 x$; take e.g. $f(x) := \log_2^*(x)$.

References

[1] S. Buss, Bounded Arithmetic (Bibliopolis, Naples, 1986).
[2] H. Gaifman and C. Dimitracopoulos, Fragments of Peano's arithmetic and the MRDP theorem, in: Logic and Algorithmic, Monogr. 30 de L'Enseignement Mathématique, Geneve (1982) 187–206.
[3] J. Krajíček, Π_1-conservativeness in systems of Bounded Arithmetic, unpublished.
[4] J. Paris and C. Dimitracopoulos, Truth definitions for Δ_0 formulae, in: Logic and Algorithmic, Monogr. 30 de L'Enseignement Mathématique, Geneve (1982) 317–330.
[5] A. Wilkie and J. Paris, On the scheme of induction for bounded arithmetic formulas, Ann. Pure Appl. Logic 35(3) (1987) 261–302.
[6] A. Wilkie and J. Paris, The end extension problem, Proc. Int. Congr. LMPS '87, Moscow.
[7] G. Takeuti, Bounded arithmetic and truth definition, Ann. Pure Appl. Logic 39(1) (1988) 75–104.