Bollettino dell'Unione Matematica Italiana
https://doi.org/10.1007/s40574-022-00325-9

Provability logic: models within models in Peano Arithmetic

Alessandro Berarducci[1] · **Marcello Mamino**[1]

Received: 27 August 2021 / Accepted: 15 April 2022 / Published online: 07 May 2022

Abstract
In 1994 Jech gave a model-theoretic proof of Gödel's second incompleteness theorem for Zermelo–Fraenkel set theory in the following form: ZF does not prove that ZF has a model. Kotlarski showed that Jech's proof can be adapted to Peano Arithmetic with the role of models being taken by complete consistent extensions. In this note we take another step in the direction of replacing proof-theoretic by model-theoretic arguments. We show, without the need of formalizing the proof of the completeness theorem within PA, that the existence of a model of PA of complexity Σ^0_2 is independent of PA, where a model is identified with the set of formulas with parameters which hold in the model. Our approach is based on a new interpretation of the provability logic of Peano Arithmetic where $\Box\phi$ is defined as the formalization of "ϕ is true in every Σ^0_2-model".

Keywords Provability logic · Peano Arithmetic · Incompleteness theorems · Modal logic

Contents

Partially supported by the Italian research project PRIN 2017, "Mathematical logic: models, sets, computability", Prot. 2017NWTM8RPRIN.

✉ Alessandro Berarducci
alessandro.berarducci@unipi.it

Marcello Mamino
marcello.mamino@unipi.it

[1] Dipartimento di Matematica, Università di Pisa, Largo Bruno Pontecorvo 5, 56127 Pisa, Italy

1 Introduction

The precise statement of Gödel's second incompleteness theorem, informally that PA cannot prove its own consistency, depends upon the choice of an arithmetization of the sentence "PA is consistent". Gödel, sketching the proof in his seminal 1931 paper [2], elected to formalize consistency as syntactic consistency. This is by no means the only reasonable choice, as demonstrated by Thomas Jech's remarkably short proof [4] of a model theorethic version of the theorem for ZF. Namely that ZF cannot prove that ZF has a model. For arithmetic, Jech shows how to transfer his ZF argument to PA by means of a conservativity result [4, Remark 2]. Then, work by Kotlarski adapts Jech's technique [7, §3.7] to obtain a direct proof: the idea is to replace models with complete theories and use the Hilbert-Bernays arithmetized completeness theorem. In this note, we take another step in the direction of replacing proof-theoretic by model-theoretic arguments: we will intend consistency to mean that PA has models of arithmetic complexity Σ^0_2.

Taking advantage of the fact that PA has partial truth predicates for formulas of bounded complexity, we show that the existence of a model of PA of complexity Σ^0_2 is independent of PA (Theorem 10.5), where a model is identified with the set of formulas with parameters which are true in the model. The presence of parameters is what makes it possible to express Tarski's truth conditions and do away with the arithmetized completeness theorem, as well as any formalized notion of syntactic consistency. For the reader that might be interested in comparing our approach to other proofs of Goëdel's incompleteness theorems, we may suggest [6, 7, 10].

In our approach, we first define a Π^0_3 predicate $\mathrm{MODEL}(x)$ expressing the fact that x is a code for a Σ^0_2-model of PA. We then consider an arithmetical interpretation of modal logic where $\Box\phi$ formalizes the fact that the formula ϕ holds in every Σ^0_2-model of PA. The formula $\Box\phi$ is in fact provably equivalent to the Σ^0_1 formalization of the provability predicate "PA $\vdash \phi$", but since in our formalization we want to avoid the syntactic notion of provability, we are not going to use this fact. Thus, on the face of it, $\Box\phi$ has complexity Π^0_4. Under our interpretation of the modal operator, $\neg\Box\perp$ says that there is a Σ^0_2-model of PA, and we will prove that this statement is independent of PA reasoning as follows. The crucial step is to verify Löb's derivability conditions [8] for our intepretation of the modal operator $\Box$, i.e. we need to prove:

1. $\mathrm{PA} \vdash \phi \implies \mathrm{PA} \vdash \Box\phi$
2. $\mathrm{PA} \vdash \Box\phi \to \Box\Box\phi$
3. $\mathrm{PA} \vdash \Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$

Here and throughout the paper we write $\mathrm{PA} \vdash \theta$ to mean that θ is true in any model of PA (and we write $M \models T$ to mean that M is a model of T). The modal counterparts of 1.–3. form the basis of the so called "provability logic" [1, 14]. From 1.–3. and the fixed point theorem one can derive $\mathrm{PA} \vdash \Box(\Box\phi \to \phi) \to \Box\phi$, whose modal counterpart is also an axiom of provability logic, see for instance [15].

Under our interpretation, the proof of 3. is straightforward. To prove 1. suppose there is a model X of PA where $\Box\phi$ fails. We need to find a model $Z \models \mathrm{PA}$ where ϕ fails. We can assume that X is countable and has domain $\mathbb{N}$. By definition there is $y \in X$ such that $X \models \mathrm{MODEL}(y)$ and $X \models$ "$y \models \neg\phi$", namely X thinks that y is a code of a Σ^0_2-model where ϕ fails. Given X and y we are able to construct a model $Z \models \mathrm{PA}$ (with domain $\mathbb{N}$) which satisfies exactly those formulas with parameters $\varphi[s]$ such that $X \models$ "$y \models \varphi[s]$". In particular $Z \models \neg\phi$, thus concluding the proof of 1.

Point 2. is the aritmetization of 1., namely we show that there is a function $x, y \mapsto {}^x y$ (of complexity Π^0_3) which maps, provably in PA, a code x of a Σ^0_2-model X and a y such that $X \models \mathrm{MODEL}(y)$, into a code of a Σ^0_2-model Z as above (the most delicate part is the mechanism to handle non-standard formulas with a non-standard number of parameters).

Granted the derivability conditions, we obtain the unprovability of $\neg\Box\perp$ by standard methods: we define G such that $\mathrm{PA} \vdash G \leftrightarrow \neg\Box G$ we show that G is unprovable and equivalent to $\neg\Box\perp$. Finally, we show

4. $\mathbb{N} \models \Box\phi \implies \mathrm{PA} \vdash \phi$

(the opposite direction follows from 1.) and we deduce that the negation of G is also unprovable, hence $\neg\Box\perp$ is independent of PA. This means that the existence of a model of complexity Σ^0_2 is independent of PA.

For the proof of 4. suppose that $\mathrm{PA} \nvdash \phi$. Then there is a Σ^0_2-model M of PA where ϕ fails (for a model-theoretic proof of this fact see Fact 4.6). A code $m \in \mathbb{N}$ of M witnesses the fact that $\mathbb{N} \not\models \Box\phi$.

2 Primitive recursive functions

The language of PA has function symbols $0, S, +, \cdot$ for zero, successor, addition, and multiplication. The axioms of PA are those of Robinson's arithmetic Q plus the first-order induction scheme. The standard model of PA is the set $\mathbb{N}$ of natural numbers with the usual interpretation of the symbols.

If t is a closed term of PA and M is a model of PA, let $t^M \in M$ be the value of t in M. If $n \in \mathbb{N}$, let $\overline{n} = S^n(0)$ be the numeral for n. In the standard model $\mathbb{N}$ the value of $\overline{n}$ is n. If $f : \mathbb{N} \to \mathbb{N}$ is a primitive recursive function then (using Gödel's β-function) f can be represented by a Σ^0_1-formula $\psi(x, y)$ of PA in such a way that, forall $m, n \in \mathbb{N}$ we have:

1. $f(m) = n \implies \mathrm{PA} \vdash \psi(\overline{m}, \overline{n})$
2. $f(m) \neq n \implies \mathrm{PA} \vdash \neg\psi(\overline{m}, \overline{n})$
3. $\mathrm{PA} \vdash \forall x \exists! y \psi(x, y)$

and similarly for n-ary functions. In the above situation we shall often write $f(x) = y$ as shorthand for the formula $\psi(x, y)$. Given a model M of PA, with our notational conventions, we have

$$f(m) = n \iff M \models f(\overline{m}) = \overline{n}.$$

We recall that an element of M is standard if it is the value of some numeral, i.e. it is of the form $\overline{n}^M$ for some $n \in \mathbb{N}$. If we identify $n \in \mathbb{N}$ with $\overline{n}^M \in M$, then 1.–3. say that ψ defines an extension of $f : \mathbb{N} \to \mathbb{N}$ to a function $f : M \to M$. In general ψ can be chosen to satisfy additional properties which depend on the way f is presented as a primitive recursive function. Consider for instance the function $f(n) = 2^n$ presented via the functional equations $2^0 = 1$ and $2^{n+1} = 2^n 2$. Then ψ can be chosen in such a way that $\mathrm{PA} \vdash \forall x (2^{x+1} = 2^x \cdot 2)$, where 2^x is defined within PA as the unique y such that $\psi(x, y)$. With this choice of ψ, in any model M of PA, the functional equation $2^{x+1} = 2^x 2$ continues to hold for non-standard values of x, thus $\psi(x, y)$ determines (by the induction scheme of PA) a unique definable extension of the function $n \in \mathbb{N} \mapsto 2^n \in \mathbb{N}$ to the non-standard elements. In general, two different presentations of the same primitive recursive function determine different extensions to the non-standard elements, unless PA is able to show that the two representations are equivalent.

A representation is natural if PA proves the validity of the same functional equations that are used in the presentation of the function in the metatheory.

We shall always assume that the primitive recursive functions we consider are represented in PA in a natural way. Given a formula $\phi(x)$ of PA and a primitive recursive function f, we will feel free to write $\phi(f(x))$ as a short-hand for the formula $\exists y(f(x) = y \wedge \phi(y))$, where "$f(x) = y$" stands for the formula $\psi(x, y)$ that we have chosen to represent f inside PA. So, for instance, it makes sense to write $\phi(2^x)$ although the language of PA does not have a symbol for the exponential function. Using similar conventions, we may act as if the language of PA had been enriched with a symbol for each primitive recursive function, or, more precisely, for each primitive recursive presentation of a function.

We fix an effective Gödel numbering of terms and formulas of PA and we write $\ulcorner\phi\urcorner \in \mathbb{N}$ for the Gödel number of ϕ. In the next section we will introduce various primitive recursive functions involved in the formalization of syntactic notion. We use $x_0, x_1, x_2, \dots$ as formal variables of PA, but we also use other letters (such as x, y, z, t) as metavariables standing for formal variables.

3 Arithmetization

The content of this section is entirely standard, but we include it to fix the notations.

Proposition 3.1 *There are primitive recursive functions* SUCC, PLUS, TIMES, VAR, *which are increasing in both arguments, such that:*

- $\mathrm{SUCC}(\ulcorner t\urcorner) = \ulcorner S(t)\urcorner$
- $\mathrm{PLUS}(\ulcorner t_1\urcorner, \ulcorner t_2\urcorner) = \ulcorner t_1 + t_2\urcorner$
- $\mathrm{TIMES}(\ulcorner t_1\urcorner, \ulcorner t_2\urcorner) = \ulcorner t_1 \cdot t_2\urcorner$
- $\mathrm{VAR}(i) = \ulcorner x_i\urcorner$

where t, t_1, t_2 are terms and $i \in \mathbb{N}$.

The above functions can be naturally represented in PA by Σ^0_1-formulas, so they have a natural extension (denoted by the same names) to non-standard models of PA. By formalizing the recursive definition of the class of terms inside PA we obtain:

Proposition 3.2 *There is a formula* $\mathrm{Tm}(x) \in \Sigma^0_1$ *such that* PA *proves that, for all x,* $\mathrm{Tm}(x)$ *holds if and only if one and only one of the following alternatives holds:*

- $\exists i \;\; x = \mathrm{VAR}(i)$
- $x = \overline{\ulcorner 0\urcorner}$
- $\exists a \;\; \mathrm{Tm}(a) \wedge x = \mathrm{SUCC}(a)$
- $\exists a, b \;\; \mathrm{Tm}(a) \wedge \mathrm{Tm}(b) \wedge x = \mathrm{PLUS}(a, b)$
- $\exists a, b \;\; \mathrm{Tm}(a) \wedge \mathrm{Tm}(b) \wedge x = \mathrm{TIMES}(a, b)$

Since the class of (codes of) terms is a primitive recursive, under the natural formalization both $\mathrm{Tm}(x)$ *and its negation are equivalent, in* PA, *to Σ^0_1-formulas.*

Corollary 3.3 *For every term t of* PA, $\mathrm{PA} \vdash \mathrm{Tm}(\overline{\ulcorner t\urcorner})$.

We have analogous propositions for the codes of formulas.

Proposition 3.4 *There are primitive recursive functions* NOT, AND, EXISTS, EQUALS, *which are increasing in both arguments, such that:*

- $\mathrm{NOT}(\ulcorner\phi\urcorner) = \ulcorner\neg\phi\urcorner$
- $\mathrm{AND}(\ulcorner\phi\urcorner, \ulcorner\psi\urcorner) = \ulcorner\phi \wedge \psi\urcorner$
- $\mathrm{EXISTS}(i, \ulcorner\phi\urcorner) = \ulcorner\exists x_i \phi\urcorner$
- $\mathrm{EQUALS}(\ulcorner t_1\urcorner, \ulcorner t_2\urcorner) = \ulcorner t_1 = t_2\urcorner$

where ϕ, ψ are formulas, t_1, t_2 are terms, and $i \in \mathbb{N}$.

The above functions can be naturally represented in PA by Σ^0_1-formulas, so they have a natural extension (denoted by the same names) to non-standard models of PA.

Proposition 3.5 *There is a formula* $\mathrm{Fm}(x) \in \Sigma^0_1$ *such that* PA *proves that, for all x,* $\mathrm{Fm}(x)$ *holds if and only if one and only one of the following alternatives holds:*

- $\exists a, b\ \mathrm{Tm}(a) \wedge \mathrm{Tm}(b) \wedge x = \mathrm{EQUALS}(a, b)$
- $\exists \phi\ \mathrm{Fm}(\phi) \wedge x = \mathrm{NOT}(\phi)$
- $\exists \phi, \psi\ \mathrm{Fm}(\phi) \wedge \mathrm{Fm}(\psi) \wedge x = \mathrm{AND}(\phi, \psi)$
- $\exists i, \phi\ \mathrm{Fm}(\phi) \wedge x = \mathrm{EXISTS}(i, \phi)$

Since the class of (codes of) formulas is primitive recursive, under the natural formalization both $\mathrm{Fm}(x)$ *and its negation are equivalent, in* PA, *to Σ^0_1-formulas.*

Corollary 3.6 *For every formula ϕ,* $\mathrm{PA} \vdash \mathrm{Fm}(\overline{\ulcorner\phi\urcorner})$.

Definition 3.7 If M is a model of PA and $\phi \in M$ is such that $M \models \mathrm{Fm}(\phi)$, we will say that ϕ is an arithmetized formula in the model M. Similarly, an arithmetized term of M is an element $a \in M$ such that $M \models \mathrm{Tm}(a)$.

If ψ is a formula of PA in the metatheory, then $\overline{\ulcorner\psi\urcorner}^M$ is an arithmetized formula of M, but if M is non-standard there are arithmetized formulas which are not of this form. Similarly, if t is a term of PA, then $\overline{\ulcorner t\urcorner}^M$ is a arithmetized term of M, and if M is non-standard it will also contain non-standard arithmetized terms.

4 Σ^0_2-models

In this section we define a Σ^0_2-model as a model M with domain $\mathbb{N}$ such that the set of formulas with parameters which are true in the model is Σ^0_2-definable (so the standard model $(\mathbb{N}, 0, S, +, \cdot)$ is not Σ^0_2-definable). We proceed below with the formal definitions.

An infinite sequence of natural numbers $(a_n)_n$ is finitely supported if there is $k \in \mathbb{N}$ such that $a_n = 0$ for all $n \geq k$. There is a bijection between natural numbers and finitely supported sequences of natural numbers: it suffices to map $s \in \mathbb{N}$ to the sequence of the exponents appearing in the prime factorization $\Pi_k p_k^{a_k}$ of $s + 1$ (where $p_0 = 2$, $p_1 = 3$, $p_2 = 5$ and in general p_k is the $k + 1$-th prime).

Definition 4.1 (PA) Given s, k, let $\mathrm{el}(s, k)$ be the least a such that p_k^{a+1} does not divide $s + 1$. According to the definition,

$$s + 1 = \Pi_k p_k^{\mathrm{el}(s,k)}$$

where $\Pi_k p_k^{\mathrm{el}(s,k)}$ can be regarded as a finite product since all but finitely many factors are equal to 1. Note that $\mathrm{el}(s, k)$ is a primitive recursive function of s, k.

Remark 4.2 (PA) The coding of finitely supported sequences defined above is injective

$$s_1 = s_2 \quad \leftrightarrow \quad \forall k \ \ \mathrm{el}(s_1, k) = \mathrm{el}(s_2, k)$$

Proposition 4.3 *(PA) Given s, a, k, there is a unique t, denoted $s[a/k]$, such that* $\mathrm{el}(t, i) = \mathrm{el}(s, i)$ *for all $i \neq k$ and* $\mathrm{el}(t, k) = a$.

Note that $s[a/k]$ is a primitive recursive function of s, a, k.

We will consider countable models M of PA. We can assume that all such models have domain $\mathbb{N}$, but the intepretation of the function symbols $0, S, +, \cdot$ will in general differ from the standard one.

Definition 4.4 Let $M = (\mathbb{N};\ 0_M, S_M, +_M, \cdot_M)$ be a model of PA with domain $\mathbb{N}$. If ϕ is a formula in the language of PA and $s \in \mathbb{N}$ we write

$$M \models \phi[s]$$

to express the fact that ϕ holds in M in the environment coded by s, i.e. the environment which, for each i, assigns the value $\mathrm{el}(s, i)$ to the variable x_i. For simplicity we take as a basis of logical connectives $\neg, \wedge, \exists$ (negation, conjunction, existential quantification). The universal quantifier $\forall$ and the logical connectives $\wedge$ and $\rightarrow$ are defined in terms of $\neg, \wedge, \exists$ in the usual way. Tarski's truth conditions then take the following form:

- $M \models (\exists x_i \phi)[s] \iff$ there is $x \in \mathbb{N}$ such that $M \models \phi(s[x/i])$
- $M \models (\phi \wedge \psi)[s] \iff M \models \phi[s]$ and $M \models \psi[s]$
- $M \models (\neg\phi)[s] \iff M \not\models \phi[s]$
- $M \models (t_1 = t_2)[s] \iff \mathrm{val}(t_1, M, s) = \mathrm{val}(t_2, M, s)$

where $\mathrm{val}(t, M, s)$ is the value of the term t in the model M when variables are evaluated according to s, namely $\mathrm{val}(x_i, M, s) = \mathrm{el}(s, i)$.

If ϕ is closed (it has no free variables), then the validity of a formula ϕ in M does not depend on the environement: $M \models \phi[s] \iff M \models \phi[0]$. In this case we may write $M \models \phi$ for $M \models \phi[0]$. Occasionally we make use of the connective $\perp$ standing for "false". Thus for every M we have $M \not\models \perp$.

Definition 4.5 Let M be a model of PA with domain $\mathbb{N}$. We say that M is a Σ^0_2-model if the set of pairs $(\ulcorner\phi\urcorner, s) \in \mathbb{N} \times \mathbb{N}$ such that $M \models \phi[s]$ is an arithmetical set of complexity Σ^0_2.

For a technical reason, which will be clarified in the comments before Lemma 7.1, we assume that the constant 0 is interpreted in M with the element $0 \in \mathbb{N}$, namely $0_M = 0$.

We recall that a set of natural numbers is Δ^0_2 if both the set and its complement can be defined by a Σ^0_2-formula. Notice that a Σ^0_2-model is in fact automatically Δ^0_2. We will need the following fact.

Fact 4.6 *Let T be a recursively axiomatized theory without finite models. If T has a model, then T has a model whose elementary diagram has arithmetic complexity Δ^0_2.*

Fact 4.6 can be easily derived from the usual proof of the completeness theorem based on König's lemma, together with the observation that a recursive binary tree with an infinite path has a Δ^0_2 infinite path (see [5, 12]). We thank the anonymous referee for suggesting that it can also be derived model-theoretically from Skolem's proof of the existence of countable models as limits of finite models in [13] (see also p. 20-21 of [16] and related developments in [9, 11]). We include a model-theoretic proof below. We stress that Fact 4.6 will only be used in the metatheory, namely we do not need to formalize its proof within PA. Moreover, Fact 4.6 will only be used in the proof of PA $\nvdash \Box \perp$, but not in the proof of PA $\nvdash \neg\Box \perp$.

Proof of Fact 4.6 We can assume that T has a $\overrightarrow{\forall}\,\overrightarrow{\exists}$-axiomatization, namely it is axiomatized by formulas of the form $\forall\bar{x}\exists\bar{y}\theta(\bar{x},\bar{y})$ where θ is quantifier free and $\bar{x},\bar{y}$ are tuples of variables. We can reduce to this situation by expanding the language L of T with the introduction of a new predicate symbol $R_\varphi(\bar{x})$ for each L-formula $\varphi(\bar{x})$ together with the following axioms:

- $R_\varphi(\bar{x}) \leftrightarrow \varphi(\bar{x})$ for each atomic φ
- $R_{\neg\varphi}(\bar{x}) \leftrightarrow \neg R_\varphi(\bar{x})$
- $R_{\alpha\wedge\beta}(\bar{x}) \leftrightarrow R_\alpha(\bar{x}) \wedge R_\beta(\bar{x})$
- $R_{\exists y\varphi}(\bar{x}) \leftrightarrow \exists y R_\varphi(\bar{x}, y)$
- $R_{\forall y\varphi}(\bar{x}) \leftrightarrow \forall y R_\varphi(\bar{x}, y)$

(with implicit universal quantifiers over $\bar{x}$). After such a modification, we can assume that T has effective elimination of quantifiers, a $\overrightarrow{\forall}\,\overrightarrow{\exists}$-axiomatization, and is formulated in a relational language L (possibly with equality). We need to find a model of T whose atomic diagram is Δ^0_2 (the elementary diagram will then also be Δ^0_2 because T has effective elimination of quantifiers).

We will construct a Δ^0_2-model of T as a limit of finite models following the ideas of [11, 13] with suitable modifications to handle theories rather than single formulas. We need some definitions.

Let $S \subseteq L$ be a finite fragment of the language L.

An (S, m)-structure is a finite sequence of S-structures $\bar{M} = (M_0, M_1, \dots, M_m)$ such that M_ℓ is a substructure of $M_{\ell+1}$ for all $\ell < m$. Given another (S, m)-structure $\bar{N}$, we say that $\bar{N}$ is an m-substructure of $\bar{M}$ if N_ℓ is a substructure of M_ℓ for all $\ell \le m$.

Let $\varphi := \forall\bar{x}\exists\bar{y}\theta$ be a closed formula, with θ quantifier free. We say that φ is a (p, q)-formula if the number of $\forall$-quantifiers in φ is p and the number of $\exists$-quantifiers is q.

If $\bar{M}$ is a (S, m)-structure and φ is a closed (p, q)-formula in the language S, we say that $\bar{M}$ is an (S, m)-model of φ, if for all $\ell < m$ and for every $a_1, \dots, a_p \in \mathrm{dom}(M_\ell)$ there are $b_1, \dots, b_q \in \mathrm{dom}(M_{\ell+1})$ such that $M_{\ell+1} \models \theta(\bar{a}, \bar{b})$. Note that a $(S, 0)$-structure satisfies every closed formula.

We say that $\bar{M}$ is (p, q)-bounded if $|M_0| = 1$ and for all $\ell < m$, $|M_{\ell+1}| \le |M_\ell| + q\,|M_\ell|^p$. Note that if $\bar{M}$ is (p, q)-bounded, then it is (a, b)-bounded for all $a \ge p, b \ge q$.

The following facts follow easily from the definitions. The idea of the proof is as in [11, Claim 1.3] with minor adaptations.

1. If φ has a model, then for every $n \in \mathbb{N}$ φ has an (S, n)-model $\bar{M}$.
2. If $\varphi = \forall\bar{x}\exists\bar{y}\theta$ is a (p, q)-formula with an (S, n)-model $\bar{M}$, then φ has a (p, q)-bounded n-submodel $\bar{N}$. (Proof: Define N_ℓ by induction on ℓ. Pick an arbitardy element $a \in M_0$ and put $N_0 = \{a\}$. Given $\ell < n$, there are $|N_\ell|^p$ possible p-tuples $\bar{x}$ from N_ℓ. For each of them choose a q-tuple $\bar{y}$ from $M_{\ell+1}$ witnessing $\theta(\bar{x}, \bar{y})$ and put its elements in $N_{\ell+1}$.)

An (S, n)-structure $\bar{N} = (N_0, \dots, N_n)$ is called initial if N_n is a finite initial segment of $\mathbb{N}$ (we do not require that N_ℓ is initial for $\ell < n$).

We observe that, for fixed S, n, p, q, there are only finitely many (p, q)-bounded initial (S, n)-structures and that any (p, q)-bounded (S, n)-structures is isomorphic to an initial one.

Let $(\varphi_n)_{n\in\mathbb{N}}$ be a recursive enumeration of the axioms of T and let L_n be the language of $\varphi_0 \wedge \dots \wedge \varphi_n$ (a finite fragment of L). Let $a_n, b_n \in \mathbb{N}$ be such that φ_n is a closed (a_n, b_n)-formula. Let $P := (p_n)_{n\in\mathbb{N}}$ and $Q := (q_n)_{n\in\mathbb{N}}$ where $p_n := \max_{k\le n} a_k$ and $q_n := \sum_{k\le n} b_k$. Since $\varphi_0 \wedge \dots \wedge \varphi_n$ is equivalent to a (p_n, q_n)-formula in the language L_n, there is an initial (p_n, q_n)-bounded (L_n, n)-model $\bar{N}$ of $\varphi_0, \dots, \varphi_n$. We call such a structure a $T_{|n}$-model.

We say that a $T_{|n+1}$-model $\bar{M} = (M_0, \dots, M_{n+1})$ extends a $T_{|n}$-model $\bar{N} = (N_0, \dots, N_n)$, if for each $\ell \leq n$, N_ℓ is the L_n-reduct of a substructure of M_ℓ (which is a L_{n+1}-structure).

We define a finitely branching forest $M_T(P, Q)$ as follows. The roots of $M_T(P, Q)$ are the $T_{|0}$-models. For $n > 0$, the nodes of $M_T(P, Q)$ at level n are the $T_{|n}$-models which extend some node of $M_T(P, Q)$ at leven $n-1$. The extension relation turns $M_T(P, Q)$ into a finitely branching forest (we may make it into a finitely branching tree by adding a fictitious new root).

By induction on n one can show that every $T_{|n}$-model is isomorphic to a node of $M_T(P, Q)$ at level n. Assuming that T has a model, it follows that $M_T(P, Q)$ is infinite. Since moreover $M_T(P, Q)$ is recursive and finitely branching, $M_T(P, Q)$ has an infinite path of complexity Δ^0_2 (just take the left-most path with respect to some natural ordering). Let M be the union of the structures M_m such that there is an m-model of the form $\bar{M} = (M_0, M_1, \dots, M_m)$ in the path (the domain of M is the union of the domains, and the interpretation of each relation symbol $R \in L$ is the union of its interpretations in those M_m in which it is defined). Then M is a model of T whose atomic diagram has complexity Δ^0_2. □

5 Codes of models

In this section we define the notion of Σ^0_2-model and show that the set of codes of Σ^0_2-models is Π^0_3-definable (Proposition 5.7). This is related to the observation in [6] that the set of codes of consistent complete extensions of a recursively axiomatized theory is Π^0_3-definable. The difference is that our formulation does not involve the syntactic notion of consistency, which would require fixing a proof-system.

We need the fact that in PA there are Σ^0_n-truth predicates for Σ^0_n-formulas (see [3]). In particular we have:

Fact 5.1 *There is a formula* $\mathrm{Sat}_2(x_0, x_1) \in \Sigma^0_2$ *such that for every* $\psi(x_1) \in \Sigma^0_2$*,*

$$\mathrm{PA} \vdash \forall x_1 \;\; \mathrm{Sat}_2(\overline{\ulcorner \psi \urcorner}, x_1) \leftrightarrow \psi(x_1);$$

For our purposes we need a variation of Sat_2 which works for formulas in two variables and additional parameters as in the following corollary.

Corollary 5.2 *There is a formula* $\mathrm{Sat}(x_0, x_1, x_2) \in \Sigma^0_2$ *such that for every* $n \in \mathbb{N}$ *and every formula* $\psi(z_1, \dots, z_n, x, y) \in \Sigma^0_2$*,*

$$\mathrm{PA} \vdash \forall a_1, \dots, a_n, \;\exists c \;\forall x, y \;\; \mathrm{Sat}(c, x, y) \leftrightarrow \psi(a_1, \dots, a_n, x, y).$$

The idea is that c codes the predicate $\{(x, y) \mid \psi(a_1, \dots, a_n, x, y)\}$.

Proof We make use of the predicate Sat_2 of Fact 5.1 and of the coding of sequences in Definition 4.1. For simplicity we write $(s)_i$ for $\mathrm{el}(s, i)$. Let $\mathrm{Sat}(c, x, y)$ be the formula $\mathrm{Sat}_0((c)_0, f(c, x, y))$ where $f(c, x, y)$ is the least t such that:

- $(t)_0 = x$
- $(t)_1 = y$
- $\forall i > 0 \; (t)_{i+1} = (c)_i$

Now, given ψ, there is a Σ^0_2-formula $\theta_\psi(t)$ such that, in PA,

$$\theta_\psi(t) \leftrightarrow \psi((t)_2, \dots, (t)_{n+1}, (t)_0, (t)_1)$$

Reasoning in PA, given $a_1, \dots, a_n$, let c be minimal such that $(c)_0 = \ulcorner \theta_\psi \urcorner$, $(c)_1 = a_1, \dots, (c)_n = a_n$. Then

$$\begin{aligned} \mathrm{Sat}(c, x, y) &\leftrightarrow \mathrm{Sat}_2(\overline{\ulcorner \theta_\psi \urcorner}, f(c, x, y)) \\ &\leftrightarrow \theta_\psi(f(c, x, y)) \\ &\leftrightarrow \psi(a_1, \dots, a_n, x, y) \end{aligned}$$

Definition 5.3 Let M be a Σ^0_2-model of PA (Definition 4.5). Then by definition there is a Σ^0_2-formula $\psi_M(x_0, x_1)$ such that for all formulas ϕ of PA and all $s \in \mathbb{N}$,

$$M \models \phi[s] \iff \mathbb{N} \models \psi_M(\ulcorner \phi \urcorner, s)$$

Letting $m = \ulcorner \psi_M \urcorner$, this is equivalent to

$$M \models \phi[s] \iff \mathbb{N} \models \mathrm{Sat}(m, \ulcorner \phi \urcorner, s)$$

where $\mathbb{N}$ is the standard model of PA. If the above equivalence holds for all (ϕ, s) we say that m is a code for the model M.

Our next goal is to show that the set of codes of Σ^0_2-models is Π^0_3-definable. We want to do so avoiding any recourse to a proof-system.

Definition 5.4 We write ιy for "the unique y such that". When we write an expression like $f(x) = \iota y.P(x, y)$ we mean that f is the partial function defined as follows: if there is one and only one y such that $P(x, y)$, then $f(x)$ is such a y; in the opposite case $f(x)$ is undefined.

Definition 5.5 (PA) Given m, we define partial functions $0_m, s_m, +_m, \cdot_m$ (of arity 0, 1, 2, 2 respectively) as follows. Fix an arbitrary s (for instance $s = 0$).

- $0_m = \iota y.\ \mathrm{Sat}(m,\ \overline{\ulcorner 0 = x_0 \urcorner},\ s[y/0])$
- $S_m(a) = \iota y.\ \mathrm{Sat}(m,\ \overline{\ulcorner S(x_0) = x_1 \urcorner},\ s[a/0, y/1])$
- $a +_m b = \iota y.\ \mathrm{Sat}(m,\ \overline{\ulcorner x_0 + x_1 = x_2 \urcorner},\ s[a/0, b/1, y/2])$
- $a \cdot_m b = \iota y.\ \mathrm{Sat}(m,\ \overline{\ulcorner x_0 \cdot x_1 = x_2 \urcorner},\ s[a/0, b/1, y/2])$

We say that m is total if these functions are total, i.e. the various y always exist and are unique. Since Sat is Σ^0_2, "m is total" is a Π^0_3-definable predicate in m. If m is total we define a function VAL whose first argument satisfies the predicate $\mathrm{Tm}(x)$ as follows:

- $\mathrm{VAL}(\mathrm{VAR}(i), m, s) = \mathrm{el}(s, i)$
- $\mathrm{VAL}(\overline{\ulcorner 0 \urcorner}, m, s) = 0_m$
- $\mathrm{VAL}(\mathrm{SUCC}(a), m, s) = S_m(\mathrm{VAL}(a, m, s))$
- $\mathrm{VAL}(\mathrm{PLUS}(a, b), m, s) = \mathrm{VAL}(a, m, s) +_m \mathrm{VAL}(b, m, s)$
- $\mathrm{VAL}(\mathrm{TIMES}(a, b), m, s) = \mathrm{VAL}(a, m, s) \cdot_m \mathrm{VAL}(b, m, s)$

Note that VAL is Π^0_3-definable.

Definition 5.6 (PA) We write $\mathrm{MODEL}(m)$ if m is total (Definition 5.5) and the conjunction of the universal closure of the following clauses holds, where the variables ϕ, ψ are relativized to the predicate Fm, the variables a, b are relativized to the predicate Tm, and the variables i, s are unrestricted.

- $0_m = 0$ (see Definition 4.5)
- $\mathrm{Sat}(m, \mathrm{EXISTS}(i, \phi), s) \leftrightarrow \exists x\ \mathrm{Sat}(m, \phi, s[x/i])$

- $\mathrm{Sat}(m, \mathrm{AND}(\phi, \psi), s) \leftrightarrow \mathrm{Sat}(m, \phi, s) \wedge \mathrm{Sat}(m, \psi, s)$
- $\mathrm{Sat}(m, \mathrm{NOT}(\phi), s) \leftrightarrow \neg\, \mathrm{Sat}(m, \phi, s)$
- $\mathrm{Sat}(m, \mathrm{EQUALS}(a, b), s) \leftrightarrow \mathrm{VAL}(a, m, s) = \mathrm{VAL}(b, m, s)$
- $\mathrm{Ax}_{\mathrm{PA}}(\phi) \rightarrow \mathrm{Sat}(m, \phi, s)$

Where $\mathrm{Ax}_{\mathrm{PA}}(x)$ is the natural formalization of "x is an axiom of PA".

Proposition 5.7 1. $\mathrm{MODEL}(m)$ *is a* Π^0_3*-formula in the free variable* m.
2. *If* M *is a* Σ^0_2*-model of* PA *and* m *is a code for* M (Definition 5.3)*, then* $\mathbb{N} \models \mathrm{MODEL}(m)$.
3. *If* $m \in \mathbb{N}$ *and* $\mathbb{N} \models \mathrm{MODEL}(m)$*, then there is a* Σ^0_2*-model* M *such that*

$$M \models \phi[s] \iff \mathbb{N} \models \mathrm{Sat}(m, \ulcorner \phi \urcorner, s)$$

for all ϕ, s.

If 3. holds, M is the (unique) model coded by m. So every Σ^0_2-model has a code, but different codes may code the same model.

Proof Point 1. is by inspection of the definition of $\mathrm{MODEL}(x)$. Indeed we have already observed that the totality condition in Definition 5.5 is Π^0_3. It is also clear that the negative occurrence of the subformula $\exists a \ \ \mathrm{Sat}(m, \phi, s[a/i])$ in Definition 5.6 is Π^0_3 and the other parts in the definition of $\mathrm{MODEL}(x)$ have lower complexity.

To prove 2. we recall that, by its very definition, $\mathrm{MODEL}(m)$ expresses the fact that the set $\{(\phi, s) \mid \mathrm{Sat}(m, \phi, s)\}$ satisfies Tarski's truth conditions for arithmetized formulas (standard or non-standard). When interpreted in the standard model $\mathbb{N}$, we only need to consider standard arithmetized formulas and (2) follows from the assumption that M is a model.

To prove 3., let $m \in \mathbb{N}$ be such that $\mathbb{N} \models \mathrm{MODEL}(m)$. Define M as the structure with domain $\mathbb{N}$ which interprets $0, S, +, \cdot$ as $0_m, S_m, +_m, \cdot_m$ respectively. By induction on the complexity of the formula ϕ we have $M \models \phi[s] \iff \mathbb{N} \models \mathrm{Sat}(m, \overline{\ulcorner \phi \urcorner}, s)$.

6 An anti-quote notation

Definition 6.1 If ϕ is a formula without free variables, we write $\mathrm{True}(x, \overline{\ulcorner \phi \urcorner})$ for $\mathrm{Sat}(x, \overline{\ulcorner \phi \urcorner}, 0)$ and observe that

$$\mathrm{PA} \vdash \mathrm{MODEL}(m) \rightarrow \forall s (\mathrm{True}(m, \overline{\ulcorner \phi \urcorner}) \leftrightarrow \mathrm{Sat}(m, \overline{\ulcorner \phi \urcorner}, s)),$$

i.e. PA proves that the truth of a closed formula in a model does not depend on the environment.

Definition 6.2 If $\psi(x_0, \ldots, x_n)$ is a formula of PA, we write

$$\mathrm{True}(m, \overline{\ulcorner \psi(\dot{a}_0, \ldots, \dot{a}_n) \urcorner})$$

for $\exists s \ \ \mathrm{el}(s, \overline{0}) = a_0 \wedge \cdots \wedge \mathrm{el}(s, \overline{n}) = a_n \wedge \mathrm{Sat}(m, \overline{\ulcorner \psi(x_0, \ldots, x_n) \urcorner}, s)$.

If $\mathrm{MODEL}(m)$ holds, $\mathrm{Sat}(m, \overline{\ulcorner \psi(\dot{a}_0, \ldots, \dot{a}_n) \urcorner})$ formalizes the fact that ψ holds in the model coded by m in the environment which assigns the value a_i to the variable x_i.

Intuitively $\overline{\ulcorner \ \urcorner}$ is a quote notation and the dot is an anti-quote. If an expression appears within the scope of $\overline{\ulcorner \ \urcorner}$ it is only its name that matters, not its value, but if we put a dot on it, it is its value that matters and not its name. The following remark will further clarify the issue.

Remark 6.3 Assume $\mathrm{MODEL}(m)$. If f is a primitive recursive function, there is a difference between $\mathrm{True}(m, \overline{\ulcorner \psi(\dot{f}(x)) \urcorner})$ and $\mathrm{True}(m, \overline{\ulcorner \psi(f(\dot{x})) \urcorner})$. In the first case we evaluate $f(x)$ outside of m and we intepret $\psi(x_0)$ in m in the environment $x_0 \mapsto f(x)$. In the second case we interpret the formula $\psi(f(x_0))$ in m in the environment $x_0 \mapsto x$. More precisely, PA proves that if $\mathrm{MODEL}(m)$ holds, then:

- $\mathrm{True}(m, \overline{\ulcorner \psi(\dot{f}(x)) \urcorner}) \leftrightarrow \exists s(\mathrm{el}(s, 0) = f(x) \wedge \mathrm{Sat}(m, \overline{\ulcorner \psi(x_0) \urcorner}, s))$
- $\mathrm{True}(m, \overline{\ulcorner \psi(f(\dot{x})) \urcorner}) \leftrightarrow \exists t(\mathrm{el}(t, 0) = x \wedge \mathrm{Sat}(m, \overline{\ulcorner \psi(f(x_0)) \urcorner}, t))$

For example, $\mathrm{True}(m, \overline{\ulcorner s(\dot{x}) = \dot{s}(x) \urcorner})$ might non hold when $x = 0$.

7 Coding environments

Given a finitely supported sequence $a_0, a_1, \dots a_n, \dots \in \mathbb{N}$, there is some $s \in \mathbb{N}$ which codes the given sequence in the sense that $\mathrm{el}(s, k) = a_k$ for all $k \in \mathbb{N}$. Now let M be a model of PA with domain $\mathbb{N}$.

The aim of this section is to construct a function Env which, given M and s, produces an element $\mathrm{Env}(s, M) \in M$ such that for all $k \in \mathbb{N}$

$$\mathrm{el}^M(\mathrm{Env}(s, M), \overline{k}^M) = \mathrm{el}(s, k) = a_k$$

In fact we will produce a Π^0_3-definable function env such that given s and a code m for a Σ^0_2-model M, yields $\mathrm{env}(s, m) = \mathrm{Env}(s, M)$.

To construct $\mathrm{Env}(s, M)$ we encounter a technical difficulty as we need $\mathrm{el}^M(\mathrm{Env}(s, M), \overline{k}^M) = 0$ for all large enough $k \in \mathbb{N}$. When M is isomorphic to $\mathbb{N}$ this implies $0^M = 0$, which is the technical condition required in Definition 4.5. A different approach would have been to code environments by finite sequences instead of finitely supported sequences. With this encoding the assumption $0^M = 0$ becomes unnecessary at the expense of complicating the definition of Tarski's semantics.

Lemma 7.1 *Let M be a model of* PA *with domain $\mathbb{N}$. Given $s \in \mathbb{N}$, there is a unique t, denoted* $\mathrm{Env}(s, M)$*, such that:*

1. $\forall k < s\ \forall a, \mathbb{N} \models \mathrm{el}(s, k) = a \implies M \models \mathrm{el}(t, \overline{k}) = a$
2. $M \models \forall k \geq \overline{s}\ \mathrm{el}(t, k) = 0$

Note that for $k \geq s$, we have $\mathbb{N} \models \mathrm{el}(s, k) = 0$. It follows that for all $s, k \in \mathbb{N}$ we have $M \models \mathrm{el}(\mathrm{Env}(s, M), \overline{k}) = x_0$ in the environment $x_0 \mapsto \mathrm{el}(s, k)$, or in other words

$$\mathrm{el}^M(\mathrm{Env}(s, M), \overline{k}^M) = \mathrm{el}(s, k)$$

where the superscript indicates the model where el *and $\overline{k}$ are evaluated.*

Proof We will prove the following more general result: for all n there is a unique t such that:

1. $\forall k < n\ \forall a, \mathbb{N} \models \mathrm{el}(s, k) = a \implies M \models \mathrm{el}(t, \overline{k}) = a$
2. $M \models \forall k \geq \overline{n}\ \mathrm{el}(t, k) = 0$

Granted this, the lemma follows by taking $n = s$. To prove our claim we proceed by induction on n. For $n = 0$, we take $t = 0$. The inductive step follows from Proposition 4.3, which allows to modify a given coded sequence by changing any of its values.

Recalling the substitution function $s[z/k]$ from Proposition 4.3, the crucial property of Env is that it commutes with substitutions in the sense of the following proposition.

Proposition 7.2 *Let M be a model of* PA *with domain $\mathbb{N}$, then for all $z, s, k \in \mathbb{N}$*

$$M \models e_1[z/\overline{k}] = e_2$$

where $e_1 = \mathrm{Env}(s, M)$ *and* $e_2 = \mathrm{Env}(s[z/k], M)$.

We may write the proposition more perspicuosly as

$$M \models \mathrm{Env}(s, M)[z/\overline{k}] = \mathrm{Env}(s[z/k], M),$$

but note that $\mathrm{Env}(s, M)$ and $\mathrm{Env}(s[z/k], M)$ are defined outside of M, while $e_1[z/\overline{k}]$ depends on the intepretation of a Σ^0_1-formula inside M (the formula which defines the primitive recursive substitution function in Proposition 4.3).

Proof It suffices to show that for all $i \in M$,

$$M \models \mathrm{el}(\mathrm{Env}(s, M)[z/\overline{k}], i) = \mathrm{el}(\mathrm{Env}(s[z/k], M), i).$$

We distinguish three cases:

- $i = \overline{k}^M$
- $i = \overline{x}^M$ for some $x \in \mathbb{N}$ different from k
- i is a non-standard element of M

In the first case both sides of the equality to be proved are equal to z. In the second case they are both equal to $\mathrm{el}(s, x)$. In the third case they are both equal to 0.

In the rest of the section we formalize Lemma 7.5 and Proposition 7.2 inside PA. We need some definitions.

Definition 7.3 Let $\mathrm{num} : \mathbb{N} \to \mathbb{N}$ be the primitive recursive function $n \mapsto \ulcorner S^n(0) \urcorner$.

We can represent num inside PA, so it will make sense to apply it to non-standard elements of a model of PA.

Definition 7.4 (PA) Assuming $\mathrm{MODEL}(m)$, let $\mathrm{numv}(n, m) = \mathrm{VAL}(\mathrm{num}(n), m, 0)$ (the third argument of VAL codes the environment, which is irrelevant in this case).

If n is standard, then $\mathrm{numv}(n, m)$ is the value of the numeral $\overline{n}$ in the model coded by m.

We can now define a function env such that, if M is a Σ^0_2-model with code m, then $\mathrm{env}(s, m) = \mathrm{Env}(s, M)$.

Lemma 7.5 *(PA) Let m be such that* $\mathrm{MODEL}(m)$. *Given s, there is a unique t, denoted* $\mathrm{env}(s, m)$*, such that:*

1. $\forall k < s\ \ \mathrm{True}(m, \overline{\ulcorner \mathrm{el}(\dot{t}, \dot{\mathrm{numv}}(k, m)) = \dot{\mathrm{el}}(s, k) \urcorner})$
2. $\mathrm{True}(m, \overline{\ulcorner \forall k \geq \dot{\mathrm{numv}}(s, m)\ \ \mathrm{el}(\dot{t}, k) = 0 \urcorner})$.

Similarly to Proposition 7.1 *for all s, k we have*

$$\mathrm{True}(m, \overline{\ulcorner \mathrm{el}(\dot{\mathrm{env}}(s, m), \dot{\mathrm{numv}}(k, m)) = \dot{\mathrm{el}}(s, k) \urcorner}).$$

Proof By formalizing the proof of Lemma 7.1 in PA.

We can now give a formalized version of Proposition 7.2.

Proposition 7.6 *(PA)* $\forall m, z, k, s$, *if* $\mathrm{MODEL}(m)$, *then*

$$\mathrm{True}(m, \overline{\ulcorner \dot{\mathrm{env}}(s, m)[\dot{z}/\dot{\mathrm{numv}}(k, m)] = \dot{\mathrm{env}}(s[z/k], m) \urcorner})$$

Proof of Proposition 7.6 Work in PA and assume $\mathrm{MODEL}(m)$. Given z, k, s we need to show

$$\mathrm{True}(m, \overline{\ulcorner \dot{\mathrm{env}}(s, m)[\dot{z}/\dot{\mathrm{numv}}(k, m)] = \dot{\mathrm{env}}(s[z/k], m) \urcorner}).$$

By Remark 4.2 and the definition of $\mathrm{MODEL}(m)$, this is equivalent to

$$\forall i \;\; \mathrm{True}(m, \overline{\ulcorner \mathrm{el}(\dot{\mathrm{env}}(s, m)[\dot{z}/\dot{\mathrm{numv}}(k, m)], \dot{i}) = \mathrm{el}(\dot{\mathrm{env}}(s[z/k], m), \dot{i}) \urcorner})$$

We distinguish three cases:

- $i = \mathrm{numv}(k, m)$
- $i = \mathrm{numv}(x, m)$ for some $x \neq k$
- none of the above, namely i is a non-standard element of the model coded by m

In the first, case we have

$$\begin{aligned}
&\mathrm{True}(m, \overline{\ulcorner \mathrm{el}(\dot{\mathrm{env}}(s, m)[\dot{z}/\dot{\mathrm{numv}}(k, m)], \dot{i}) = \dot{z} \urcorner}) && \text{by Proposition 4.3}\\
&\mathrm{True}(m, \overline{\ulcorner \dot{z} = \dot{\mathrm{el}}(s[z/k], k) \urcorner}) && \text{by Lemma 7.5}\\
&\mathrm{True}(m, \overline{\ulcorner \dot{\mathrm{el}}(s[z/k], k) = \mathrm{el}(\dot{\mathrm{env}}(s[z/k], m), \dot{i}) \urcorner}) && \text{by Proposition 4.3}
\end{aligned}$$

and we conclude by transitivity of the equality inside the model coded by m.
In the second case, we have

$$\begin{aligned}
&\mathrm{True}(m, \overline{\ulcorner \mathrm{el}(\dot{\mathrm{env}}(s, m)[\dot{z}/\dot{\mathrm{numv}}(k, m)], \dot{i}) = \mathrm{el}(\dot{\mathrm{env}}(s, m), \dot{i}) \urcorner}) && \text{by Proposition 4.3}\\
&\mathrm{True}(m, \overline{\ulcorner \mathrm{el}(\dot{\mathrm{env}}(s, m), \dot{i}) = \dot{\mathrm{el}}(s, x) \urcorner}) && \text{by Lemma 7.5}\\
&\mathrm{True}(m, \overline{\ulcorner \dot{\mathrm{el}}(s, x) = \dot{\mathrm{el}}(s[z/k], x) \urcorner}) && \text{by Proposition 4.3}\\
&\mathrm{True}(m, \overline{\ulcorner \dot{\mathrm{el}}(s[z/k], x) = \mathrm{el}(\dot{\mathrm{env}}(s[z/k], m), \dot{i}) \urcorner}) && \text{by Lemma 7.5}
\end{aligned}$$

and we conclude again by transitivity of the equality.
In the third case,

$$\begin{aligned}
&\mathrm{True}(m, \overline{\ulcorner \mathrm{el}(\dot{\mathrm{env}}(s, m)[\dot{z}/\dot{\mathrm{numv}}(k, m)], \dot{i}) = \mathrm{el}(\dot{\mathrm{env}}(s, m), \dot{i}) \urcorner}) && \text{by Proposition 4.3}\\
&\mathrm{True}(m, \overline{\ulcorner \mathrm{el}(\dot{\mathrm{env}}(s, m), \dot{i}) = 0 \urcorner}) && \text{by Lemma 7.5}\\
&\mathrm{True}(m, \overline{\ulcorner 0 = \mathrm{el}(\dot{\mathrm{env}}(s[z/k], m), \dot{i}) \urcorner}) && \text{by Lemma 7.5}
\end{aligned}$$

and we conclude as above. □

Recalling that $s + 1 = \Pi_i p_i^{\mathrm{el}(s,i)}$ we can illustrate the definition of env by the following example.

Example 7.7 Let $s + 1 = 2^7 3^5$ and let M be a Σ_2^0-model coded by m. Then $\mathrm{env}(s, m)$ is the unique element t such that $M \models x_2 + 1 = 2^{x_0} 3^{x_1}$ in the environment $x_0 \mapsto 7, x_1 \mapsto 5, x_2 \mapsto t$. Note that 7 and 5 are not necessarily equal to $\overline{5}^M$ and $\overline{7}^M$, so in general $M \not\models x_2 + 1 = 2^7 3^5$ in the environment $x_2 \mapsto t$.

We are now ready to prove Propositions 7.2 and 7.6.

8 A model within a model

Proposition 8.1 *Let X be a model of* PA *with domain $\mathbb{N}$. Given $y \in X$ such that $X \models$ MODEL(y), there is a model $Z \models$ PA with domain $\mathbb{N}$ such that*

$$Z \models \phi[s] \iff X \models \mathrm{Sat}(y, \overline{\ulcorner\phi\urcorner}, t)$$

where $t = \mathrm{Env}(s, X)$.

Proof Let X, y be as in the hypothesis. Let $\mathcal{Z}$ be the set of pairs (ϕ, s) such that $X \models \mathrm{Sat}(y, \overline{\ulcorner\phi\urcorner}, \mathrm{Env}(s, X))$. We need to prove that there is a model Z of PA with domain $\mathbb{N}$ such that $Z \models \phi[s] \iff (\phi, s) \in \mathcal{Z}$. To this aim we need to check Tarski's truth conditions and verify that $\mathcal{Z}$ contains the axioms of PA. The latter condition follows easily from the assumption $X \models \mathrm{MODEL}(y)$. Let us check the truth condition for negation:

$$\begin{aligned}
(\neg\phi, s) \in \mathcal{Z} &\leftrightarrow X \models \mathrm{Sat}(y, \overline{\ulcorner\neg\phi\urcorner}, \mathrm{Env}(s, X)) \\
&\leftrightarrow X \models \neg\,\mathrm{Sat}(y, \overline{\ulcorner\phi\urcorner}, \mathrm{Env}(s, X)) \\
&\leftrightarrow X \not\models \mathrm{Sat}(y, \overline{\ulcorner\phi\urcorner}, \mathrm{Env}(s, X)) \\
&\leftrightarrow (\phi, s) \notin \mathcal{Z}
\end{aligned}$$

where in the second equivalence we used the fact that $X \models \mathrm{MODEL}(y)$. Similarly, we can verify Tarski's truth condition for the quantifier $\exists$:

$$\begin{aligned}
(\exists x_k\phi, s) \in \mathcal{Z} &\leftrightarrow X \models \mathrm{Sat}(y, \overline{\ulcorner\exists x_k\phi\urcorner}, \mathrm{Env}(s, X))) \\
&\leftrightarrow X \models \exists x_0\,\mathrm{Sat}(y, \overline{\ulcorner\phi\urcorner}, \mathrm{Env}(s, X)[x_0/\overline{k}]) \\
&\leftrightarrow \exists z \in \mathbb{N}\ X \models \mathrm{Sat}(y, \overline{\ulcorner\phi\urcorner}, \mathrm{Env}(s, X)[z/\overline{k}]) \\
&\leftrightarrow \exists z \in \mathbb{N}\ X \models \mathrm{Sat}(y, \overline{\ulcorner\phi\urcorner}, \mathrm{Env}(s[z/k], X)) \\
&\leftrightarrow \exists z \in \mathbb{N}\ (\phi, s[z/k]) \in \mathcal{Z}
\end{aligned}$$

where in the fourth equivalence we used Proposition 7.2. We leave the other verifications to the reader. □

In the above proposition if X is a Σ^0_2-model, then Z is also Σ^0_2. In the rest of the section we prove that there is a definable function which computes a code ${}^x y$ of Z given y and a code x for X.

Proposition 8.2 *(PA) Given x, y, there is z such that for all ϕ, s,*

$$\mathrm{Sat}(z, \phi, s) \iff \mathrm{True}(x, \overline{\ulcorner\mathrm{Sat}(\dot{y}, \mathrm{nu\dot{m}v}(\phi, x), \mathrm{e\dot{n}v}(s, x))\urcorner})$$

We define ${}^x y$ as the minimal such z and observe that the function $x, y \mapsto {}^x y$ is Π^0_3-definable.

Proof Given x, y, the set

$$\{(\phi, s) \mid \mathrm{True}(x, \overline{\ulcorner\mathrm{Sat}(\dot{y}, \mathrm{nu\dot{m}v}(\phi, x), \mathrm{e\dot{n}v}(s, x))\urcorner})\}$$

is Σ^0_2-definable with parameters x, y, so by Corollary 5.2 there is some z which codes this set, and we take ${}^x y$ to be the minimal such z. It can be readily verified that $x, y \mapsto {}^x y$ is Π^0_3-definable.

Theorem 8.3 *(PA) If* $\mathrm{MODEL}(x)$ *and* $\mathrm{True}(x, \overline{\ulcorner\mathrm{MODEL}(\dot{y})\urcorner})$*, then* $\mathrm{MODEL}({}^x y)$.

Proof We need to show, inside PA, that the class of all pairs (ϕ, s) such that $\mathrm{Sat}({}^x y, \phi, s)$ satisfies Tarski's truth conditions and contains the arithmetized axioms of PA. The latter property is easy, so we limit ourself to verify the clauses for $\neg$ and $\exists$ in Tarski's truth conditions.

$$\begin{aligned}\mathrm{Sat}({}^x y, \mathrm{NOT}(\phi), s) &\leftrightarrow \mathrm{True}(x, \overline{\ulcorner \mathrm{Sat}(\dot{y}, \mathrm{nu\dot{m}v}(\mathrm{NOT}(\phi), x), \mathrm{e\dot{n}v}(s, x))\urcorner})\\ &\leftrightarrow \mathrm{True}(x, \overline{\ulcorner \neg\,\mathrm{Sat}(\dot{y}, \mathrm{nu\dot{m}v}(\phi), \mathrm{e\dot{n}v}(s, x))\urcorner})\\ &\leftrightarrow \neg\,\mathrm{True}(x, \overline{\ulcorner \mathrm{Sat}(\dot{y}, \mathrm{nu\dot{m}v}(\phi), \mathrm{e\dot{n}v}(s, x))\urcorner})\\ &\leftrightarrow \neg\,\mathrm{Sat}({}^x y, \phi, s)\end{aligned}$$

where in the second equivalence we used the fact that $\mathrm{True}(x, \mathrm{MODEL}(\dot{y}))$ and in the third we used the hypothesis $\mathrm{MODEL}(x)$. Similarly we have:

$$\begin{aligned}\mathrm{Sat}({}^x y, \mathrm{EXISTS}(k, \phi), s) &\leftrightarrow \mathrm{True}(x, \overline{\ulcorner \mathrm{Sat}(\dot{y}, \mathrm{nu\dot{m}v}(\mathrm{EXISTS}(k, \phi), x), \mathrm{e\dot{n}v}(s, x))\urcorner})\\ &\leftrightarrow \mathrm{True}(x, \overline{\ulcorner \exists x_0\,\mathrm{Sat}(\dot{y}, \mathrm{nu\dot{m}v}(\phi, x), \mathrm{e\dot{n}v}(s, x)[x_0/\mathrm{nu\dot{m}v}(k, x)])\urcorner})\\ &\leftrightarrow \exists z\,\mathrm{True}(x, \overline{\ulcorner \mathrm{Sat}(\dot{y}, \mathrm{nu\dot{m}v}(\phi, x), \mathrm{e\dot{n}v}(s, x)[\dot{z}/\mathrm{nu\dot{m}v}(k, x)])\urcorner})\\ &\leftrightarrow \exists z\,\mathrm{True}(x, \overline{\ulcorner \mathrm{Sat}(\dot{y}, \mathrm{nu\dot{m}v}(\phi, x), \mathrm{e\dot{n}v}(s[z/k], x))\urcorner})\\ &\leftrightarrow \exists z\,\mathrm{Sat}({}^x y, \phi, s[z/k])\end{aligned}$$

where the fourth equivalence makes use of the properties of env (Proposition 7.6).

9 Löb's derivability conditions

Definition 9.1 Given a closed formula of PA, we let $\Box\phi$ be the formula $\forall x(\mathrm{MODEL}(x) \to \mathrm{True}(x, \overline{\ulcorner\phi\urcorner})$. Note that $\Box\phi$ has complexity Π^0_4.

The first three points of the following result correspond to Löb's derivability conditions in [8].

Theorem 9.2 *Let ϕ, ψ be closed formulas of* PA. *We have:*

1. *If* $\mathrm{PA} \vdash \phi$, *then* $\mathrm{PA} \vdash \Box\phi$
2. $\mathrm{PA} \vdash \Box\phi \to \Box\Box\phi$
3. $\mathrm{PA} \vdash \Box(\phi \to \psi) \to (\Box\phi \to \Box\psi)$
4. $\mathbb{N} \models \Box\phi \implies \mathrm{PA} \vdash \phi$

Proof 1. Suppose $\mathrm{PA} \nvdash \Box\phi$. Then there is a model $X \models \mathrm{PA}$ such that $X \models \neg\Box\phi$. By definition this means that there is $y \in X$ such that $X \models \mathrm{MODEL}(y)$ and $X \models \mathrm{True}(y, \overline{\ulcorner\neg\phi\urcorner})$. By Proposition 8.1 there is a model $Z \models \mathrm{PA}$ such that $Z \models \neg\phi$, so $\mathrm{PA} \nvdash \phi$.

2. We write $\Diamond\phi$ for $\neg\Box\neg\phi$ and observe that $\Diamond\phi$ is provably equivalent to $\exists x(\mathrm{MODEL}(x) \wedge \mathrm{True}(x, \psi))$. The statement to be proved is equivalent to $\mathrm{PA} \vdash \Diamond\Diamond\phi \to \Diamond\phi$. Now $\Diamond\Diamond\phi$ says that there exist x, y such that $\mathrm{MODEL}(x)$, $\mathrm{True}(x, \overline{\ulcorner\mathrm{MODEL}(\dot{y})\urcorner})$ and $\mathrm{True}(x, \overline{\ulcorner\mathrm{True}(\dot{y}, \overline{\ulcorner\phi\urcorner})\urcorner})$. On the other hand $\Diamond\phi$ says that there is z such that $\mathrm{MODEL}(z)$ and $\mathrm{True}(z, \overline{\ulcorner\phi\urcorner})$. To prove the implication one can take $z = {}^x y$ as defined in Proposition 8.2.
3. Clear from the definitions and the rules of predicate calculus, recalling that $\Box\theta$ stands for $\forall x\ (\mathrm{MODEL}(x) \to \mathrm{True}(x, \overline{\ulcorner\theta\urcorner}))$.

4. Suppose $PA \nvdash \phi$. By Fact 4.6 there is a Σ^0_2 model M satisfying $\neg\phi$. Let $m \in \mathbb{N}$ be a code for such a model. Then $\mathbb{N} \models \mathrm{MODEL}(m)$ and $\mathbb{N} \models \mathrm{True}(m, \overline{\ulcorner \neg\phi \urcorner})$. This is equivalent to $\mathbb{N} \models \neg\Box\phi$.

10 An undecidable formula

By the diagonal lemma given a formula $\alpha(x)$ in one free variable there is a closed formula β such that $\mathrm{PA} \vdash \beta \leftrightarrow \alpha(\overline{\ulcorner \beta \urcorner})$. Using the diagonal lemma we can define a formula G which says "I have no Σ^0_2-model", as in the definition below.

Definition 10.1 Let G be such that $\mathrm{PA} \vdash G \leftrightarrow \neg\Box G$.

Using Theorem 9.2 we deduce that G is undecidable and equivalent to $\neg\Box\perp$ by the standard arguments, see for instance [1]. We give the details below.

Lemma 10.2 $\mathrm{PA} \nvdash G$.

Proof Suppose $\mathrm{PA} \vdash G$. Then $\mathrm{PA} \vdash \Box G$ (Theorem 9.2). On the other hand by definition of G, $\mathrm{PA} \vdash \neg\Box G$, contradicting the consistency of PA.

Lemma 10.3 $\mathrm{PA} \vdash G \leftrightarrow \neg\Box\perp$.

Proof We use 1.–3. in Theorem 9.2. Reason in PA. If G holds, we get $\neg\Box G$ by definition of G. Since $\perp \rightarrow G$ is a tautology we obtain $\Box\perp \rightarrow \Box G$, hence $\neg\Box\perp$.

Now assume $\neg G$. By definition of G we get $\Box G$ and by point 2. in Theorem 9.2 $\Box\Box G$ follows. Moreover we have $\Box(\Box G \leftrightarrow \neg G)$ (apply the definition of G inside the $\Box$), so we get $\Box\neg G$. Since we also have $\Box G$, we obtain $\Box\perp$.

Lemma 10.4 $\mathrm{PA} \nvdash \neg G$.

Proof Suppose $\mathrm{PA} \vdash \neg G$. Then by definition of G, $\mathrm{PA} \vdash \Box G$, so $\mathbb{N} \models \Box G$ and by Theorem 9.2(4) $\mathrm{PA} \vdash G$, contradicting the consistency of PA.

We have thus obtained:

Theorem 10.5 $\neg\Box\perp$ *is independent of* PA*, namely* PA *does not prove that* PA *has a* Σ^0_2-*model.*

Acknowledgements We thank the anonymous referee for his comments related to Fact 4.6.

Funding Open access funding provided by Università di Pisa within the CRUI-CARE Agreement

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Boolos, G.S.: The Logic of Provability. Cambridge University Press, Cambridge (1994)
2. Gödel, K.: Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. Monatshefte für Mathematik und Physik **38–38**(1), 173–198 (1931)
3. Hájek, P., Pudlák, P.: Metamathematics of First-Order Arithmetic. Springer, Berlin (1993)
4. Jech, T.: On Gödel's second incompleteness theorem. Proc. Am. Math. Soc. **121**(1), 311–313 (1994)
5. Kleene, S.C.: Introduction to Methamatematics. North-Holland Publishing Co., Amsterdam (1952)
6. Kotlarski, H.: The incompleteness theorems after 70 years. Ann. Pure Appl. Logic **126**(1–3), 125–138 (2004)
7. Kotlarski, H.: A Model-Theoretic Approach to Proof Theory. Trends in Logic, vol. 51. Springer, Cham (2019)
8. Löb, S.: Solution of a Problem of Leon Henkin. J. Symbol. Logic **20**(2), 115–118 (1995)
9. Quinsey, J.E.: Some Problems in Logic: Applications of Kripke's Notion of Fulfilment. PhD thesis, Oxford 1980 (2019). arXiv:1904.10540
10. Richard, K.: Models of Peano Arithmetic. Oxford Logic Guides, vol. 15. Oxford University Press, New York (1991)
11. Shelah, S.: On Logical Sentences in PA. In Logic Colloquium '82, Studies in Logic and the Foundations of Mathematics. Elsevier B.V., Oxford, pp 145–160 (1984)
12. Shoenfield, J.R.: Degrees of models. J. Symbol. Logic **25**(3), 233–237 (1960)
13. Skolem, T.: Some remarks on axiomatized set theory. In: From Frege to Goedel: A Source Book in Mathematical Logic, 1879–1931. Harvard University Press, Cambridge, p. 1967 (1922)
14. Solovay, R.M.: Provability interpretations of modal logic. Israel J. Math. **25**(3–4), 287–304 (1976)
15. Verbrugge, R.: Provability logic. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy (Fall 2017 Edition). https://plato.stanford.edu/archives/fall2017/entries/logic-provability/
16. Wang, H.: A survey of Skolem's Work in Logic. Selected Works in logic. Preface to Thoralf Skolem. Universitetsforlaget, Oslo (1970)