Solovay's Completeness Without Fixed Points

Fedor Pakhomov(✉)

Steklov Mathematical Institute of Russian Academy of Sciences, Moscow, Russia
pakhfn@mi.ras.ru
http://www.mi.ras.ru/~pakhfn/

Abstract. In this paper we present a new proof of Solovay's theorem on arithmetical completeness of Gödel-Löb provability logic GL. Originally, completeness of GL with respect to interpretation of $\Box$ as provability in PA was proved by Solovay in 1976. The key part of Solovay's proof was his construction of an arithmetical evaluation for a given modal formula that made the formula unprovable in PA if it were unprovable in GL. The arithmetical sentences for the evaluations were constructed using certain arithmetical fixed points. The method developed by Solovay have been used for establishing similar semantics for many other logics. In our proof we develop new more explicit construction of required evaluations that doesn't use any fixed points in their definitions. To our knowledge, it is the first alternative proof of the theorem that is essentially different from Solovay's proof in this key part.

1 Introduction

The study of provability as a modality could be traced back to at least as early as Gödel work [Gö33]. Löb [Lö55] have proved a generalization of Gödel's Second Incompleteness Theorem that is now known as Löb's Theorem. In order to formulate his theorem Löb have stated conditions on provability predicates that are now known as Hilbert-Bernays-Löb derivability conditions. Despite Löb haven't mentioned the interpretation of a modality as a provability predicate there, his conditions essentially corresponded to the standard axiomatization of modal logic K4. Also note that arithmetical soundness of Gödel-Löb provability logic GL immediately follows from Löb's Theorem.

The axioms of modal system GL have first appeared in [Smi63]. Segerberg have shown that GL is Kripke-complete and moreover that it is complete with respect to the class of all finite transitive irreflexive trees [Seg71]. The arithmetical completeness of the system GL were established by Solovay [Sol76]. Solovay have proved that a modal formula φ is a theorem of GL iff for every arithmetical evaluation $f(x)$ the arithmetical sentence $f(\varphi)$ is provable in PA.

Latter modifications of Solovay's method were used in order to prove a lot of other similar results, we will mention just few of them. Japaridze have proved

F. Pakhomov—This work is supported by the Russian Science Foundation under grant 14-50-00005.

© Springer-Verlag GmbH Germany 2017
J. Kennedy and R.J.G.B. de Queiroz (Eds.): WoLLIC 2017, LNCS 10388, pp. 281–294, 2017.
DOI: 10.1007/978-3-662-55386-2_20

arithmetical completeness of polymodal provability logic GLP [Jap86]. Shavrukov [Sha88] and Berarducci [Ber90] have determined the interpretability logic of PA.

The key part of Solovay's proof was to show that in certain sense every finite GL-model is "embeddable" in arithmetic. Using the construction of "embeddings", it is easy to construct evaluations $f_\varphi(x)$ such that $\mathsf{PA} \nvdash f_\varphi(\varphi)$, for all GL-unprovable modal formulas φ. In order to construct the "embeddings", Solovay have used Diagonal Lemma to define certain primitive-recursive function (Solovay function), for every finite GL Kripke model. Then, using the functions, Solovay have defined the sentences that constituted the "embeddings".

de Jongh, Jumelet, and Montagna have shown that GL is complete with respect to Σ_1-provability predicates for theories $\mathsf{T} \supseteq \mathsf{I\Delta_0} + \mathsf{Exp}$ [dJJM91]. Their proof have avoided the use of Solovay functions, however, their construction still "emulated" Solovay's approach using individual sentences constructed by Diagonal Lemma.

In a discussion on FOM (Foundation of Mathematics mailing list) Shipman have asked a question about important theorems that have "essentially" only one proof [Shi09]. The example of Solovay's theorem were provided by Sambin. To the author knowledge, up to the date there were no proofs of Solovay's theorem that have avoided the central idea of Solovay's proof—the Solovay's method of constructing required sentences in terms of certain fixed points.

We note that completeness of some extensions of GL with respect to interpretations of $\Box$ that are similar to formalized provability were proved by the completely different methods. Solovay in his paper [Sol76] have briefly mentioned a method of determining modal logics of several natural interpretations of $\Box$ in set theory, namely for the interpretations of $\Box$ as "to be true in all transitive models" and as "to be true in all models $\mathbf{V}_\kappa$, where κ is an inaccessible cardinal" (there are more detailed proofs in Boolos book [Boo95, Chap. 13]). A modification of the method also have been used to show completeness of wide variety of extensions of GL with respect to artificially defined (not Σ_1) provability-like predicates [Pak16].

In the paper we present a new approach to the proof of arithmetical completeness theorem for GL. We introduce a different method of "embedding" of finite GL Kripke models. As the result, the completeness of GL is achieved with the use of evaluations given by more explicitly constructed and more "natural" sentences (in particular, we do not rely on Diagonal Lemma in the construction). In order to avoid potential misunderstanding, we note that despite the sentences from evaluations are given explicitly, our proof rely on Gödel's Second Incompleteness Theorem and the results by Pudlák [Pud86] that were proved with the use of Diagonal Lemma.

Now we will give an example of unprovable GL-formula φ and an evaluation $f(x)$ provided by our proof such that $\mathsf{PA} \nvdash f(\varphi)$. We consider the formula

$$\varphi \leftrightharpoons \Diamond v \rightarrow (\Diamond u \rightarrow \Diamond(v \wedge u)).$$

We use the following definitions for numerical functions in order to define the evaluation $f(x)$:

$$\exp(x) = 2^x, \ \log(x) = \max(\{y \mid \exp(y) \leq x\} \cup 0),$$

$$\exp^\star(x) = \underbrace{\exp(\exp(\ldots\exp(}_{x \text{ times}}0)\ldots)), \ \log^\star(x) = \max(\{y \mid \exp^\star(y) \leq x\} \cup 0)$$

(note that the functions $\exp^\star(x)$ and $\log^\star(x)$ are called *super exponentiation* and *super logarithmic* functions, respectively). The evaluation $f(x)$ is given as following:

$$f(v) \leftrightharpoons \exists x(\mathsf{Prf}(x, \ulcorner 0 = 1 \urcorner) \wedge \forall y < x(\neg\mathsf{Prf}(y, \ulcorner 0 = 1 \urcorner)) \wedge \log^\star(x) \equiv 0 \pmod 2),$$

$$f(u) \leftrightharpoons \exists x(\mathsf{Prf}(x, \ulcorner 0 = 1 \urcorner) \wedge \forall y < x(\neg\mathsf{Prf}(y, \ulcorner 0 = 1 \urcorner)) \wedge \log^\star(x) \equiv 1 \pmod 2).$$

We note that somewhat similar approach based on the parity of $\log^\star$ were used by Solovay in his letter to Nelson [Sol86]. Solovay proved that there are sentences F and G such that $\mathsf{I\Delta_0} + \Omega_1 + \mathsf{F}$ and $\mathsf{I\Delta_0} + \Omega_1 + \mathsf{G}$ are cut-interpretable in $\mathsf{I\Delta_0} + \Omega_1$, but $\mathsf{I\Delta_0} + \Omega_1 + \mathsf{F} \wedge \mathsf{G}$ isn't cut-interpretable in $\mathsf{I\Delta_0} + \Omega_1$. Also, Kotlarski in [Kot96] have used an explicit parity-based construction of a pair of sentences in order to give an alternative proof for Rosser's Theorem.

2 Preliminaries

Let us first define Gödel-Löb provability logic GL. The language of GL extends the language of propositional calculus with propositional constants $\top$ (truth) and $\bot$ (false) by the unary modal connective $\Box$. GL have the following Hilbert-style deductive system:

1. axiom schemes of classical propositional calculus PC;
2. $\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$;
3. $\Box(\Box\varphi \to \varphi) \to \Box\varphi$;
4. $\frac{\varphi \quad \varphi\to\psi}{\psi}$;
5. $\frac{\varphi}{\Box\varphi}$.

The expression $\Diamond\varphi$ is an abbreviation for $\neg\Box\neg\varphi$.

A set with a binary relation $(W, \prec)$ is called *irreflexive transitive tree* if

1. $\prec$ is a transitive irreflexive relation;
2. there is an element $r \in W$ that is called the *root* of $(W, \prec)$ such that the upward cone $\{a \mid r \prec a\}$ coincides with W;
3. for any element $w \in W$ the restriction of $\prec$ on the downward cone $\{a \mid a \prec w\}$ is a strict well-ordering order.

Segerberg [Seg71] have shown that the logic GL is complete with respect to the class of all finite irreflexive transitive trees.

Our proof relies on the results by Verbrugge and Visser [VV94] and indirectly on the results by Pudlák [Pud86]. This results are sensitive to details of formalization of some metamathematical notions. Thus unlike some other papers, where this kind of details could be safely be left unspecified, we will need to be more careful here.

We identify syntactical expressions with binary strings. We encode binary strings by positive integers numbers. A positive integer n of the form $1a_{k-1}\ldots a_0$ in binary notation encodes the binary string $a_{k-1}\ldots a_0$. We note that the binary logarithm $\log(n)$ of a number n coincides with the length of the binary string that the number n encodes. For a formula F the number n that encodes F is known as the *Gödel number* of F.

A *proof* of an arithmetical formula φ in an arithmetical theory T is a list of arithmetical formulas such that it ends with φ and every formula in the list is either an axiom of T, or is an axiom of predicate calculus, or is obtained by inference rules from previous formulas.

We will be interested in formalization of provability in the theory PA and its extensions by finitely many axioms. We take the standard axiomatization of PA (by axioms of Robinson arithmetic Q and the induction schema). We consider the natural axiomatization in arithmetic of the property of a number to be the Gödel number of some axiom of PA. For all extensions T of PA by finitely many axioms this gives us Δ_0-predicates $\mathsf{Prf}_{\mathsf{T}}(x,y)$ that are natural formalizations of "x is a proof of the formula with Gödel number y in the theory T" that is based on the definition of the notion of proof given above. And we obtain Σ_1-provability predicates

$$\mathsf{Prv}_{\mathsf{T}}(y) \leftrightharpoons \exists x \mathsf{Prf}_{\mathsf{T}}(x,y).$$

We will use effective binary numerals. The *n-th numeral* is defined as follows:

1. $\underline{0}$ is the term 0;
2. $\underline{1}$ is the term 1;
3. $\underline{2n}$ is the term $(1+1)\cdot\underline{n}$;
4. $\underline{2n+1}$ is the term $(1+1)\cdot\underline{n}+1$.

Clearly, the length of $\underline{n}$ is $\mathcal{O}(\log(n))$.

For an arithmetical formula F we denote by $\ulcorner\mathsf{F}\urcorner$ the n-th numeral, where n is the Gödel number of the formula F.

We denote by $\mathsf{Prv}(x)$ and $\mathsf{Prf}(x,y)$ the predicates $\mathsf{Prv}_{\mathsf{PA}}(x)$ and $\mathsf{Prf}_{\mathsf{PA}}(x,y)$.

An *arithmetical evaluation* is a function $f(x)$ from GL formulas to the sentences of the language of first-order arithmetic such that

1. $f(\varphi\land\psi) \leftrightharpoons f(\varphi)\land f(\psi)$;
2. $f(\varphi\lor\psi) \leftrightharpoons f(\varphi)\lor f(\psi)$;
3. $f(\neg\varphi) \leftrightharpoons \neg f(\varphi)$;
4. $f(\varphi\to\psi) \leftrightharpoons f(\varphi\to\psi)$;
5. $f(\top) \leftrightharpoons 0=0$;
6. $f(\bot) \leftrightharpoons 0=1$;
7. $f(\Box\varphi) \leftrightharpoons \mathsf{Prv}(\ulcorner f(\varphi)\urcorner)$.

Note that an arithmetical evaluation is uniquely determined by its values on propositional variables $u, v, \ldots$.

We will use $\top$, $\bot$, $\Box$, and $\Diamond$ within arithmetical formulas: the expression $\top$ is an abbreviation for $0 = 0$, the expression $\bot$ is an abbreviation for $0 = 1$, the expression $\Box\mathsf{F}$ is an abbreviation for $\mathsf{Prv}(\ulcorner \mathsf{F} \urcorner)$, and the expression $\Diamond\mathsf{F}$ is an abbreviation for $\neg\mathsf{Prv}(\ulcorner \neg\mathsf{F} \urcorner)$. The expressions of the form $\Box^n\mathsf{F}$ and $\Diamond^n\mathsf{F}$ are abbreviations for $\underbrace{\Box\Box\ldots\Box}_{n \text{ times}}\mathsf{F}$ and $\underbrace{\Diamond\Diamond\ldots\Diamond}_{n \text{ times}}\mathsf{F}$, respectively.

3 Proof of Solovay's Theorem

In the section we will just give a proof of "completeness part" of Solovay's theorem. Soundness of the logic GL essentially is due to Löb [Lö55] and we refer a reader to Boolos book [Boo95, Chap. 3] for a detailed proof.

Theorem 1. *If a modal formula φ is not provable in GL then there exists an arithmetical evaluation $f(x)$ such that $\mathsf{PA} \nvdash f(\varphi)$.*

Let us fix some modal formula φ that is not provable in GL. By Segerberg's result [Seg71], we can find a finite transitive irreflexive tree $\mathfrak{F} = (W, \prec)$ such that r is the root of $\mathfrak{F}$ and there is a model $\mathbf{M}$ on $\mathfrak{F}$ with $\mathbf{M}, r \nVdash \varphi$. For all the worlds a of $\mathfrak{F}$ we denote by $h(a)$ their "height":

$$h(a) = \sup(\{0\} \cup \{h(b) + 1 \mid a \prec b\}).$$

Let us assign arithmetical sentences C_a to all the worlds a of $\mathfrak{F}$. We put C_r to be $0 = 0$. We consider a non-leaf world a and assign sentences C_b to all its immediate successors b. Suppose $b_0, \ldots, b_n$ are all the immediate successors of a. We fix some enumeration $b_0, \ldots, b_n$ such that $h(b_n) = h(a) - 1$. For $i < n$ we put C_{b_i} to be the sentence

$$\begin{aligned}\exists x(\mathsf{Prf}_{\mathsf{PA}+\Diamond^{h(a)-1}\top}(x, \ulcorner 0 = 1 \urcorner) \wedge \forall y < x(\neg\mathsf{Prf}_{\mathsf{PA}+\Diamond^{h(a)-1}\top}(y, \ulcorner 0 = 1 \urcorner))\\ \wedge \log^\star(x) \equiv i \pmod{n+1}\\ \wedge \exists y < \exp(\exp(x))(\mathsf{Prf}_{\mathsf{PA}+\Diamond^{h(b_i)}\top}(y, \ulcorner 0 = 1 \urcorner))).\end{aligned}$$

The sentence C_{b_n} is

$$\Box^{h(a)}\bot \wedge \bigwedge_{i<n} \neg\mathsf{C}_{b_i}.$$

Note that $\mathsf{PA} \vdash \neg(\mathsf{C}_{b_i} \wedge \mathsf{C}_{b_j})$, for $i \neq j$ and

$$\mathsf{PA} \vdash \Box^{h(a)}\bot \leftrightarrow \bigvee_{i \leq n} \mathsf{C}_{b_i}.$$

We note that all C_{b_i} are PA-equivalent to Σ_1-sentences: it is obvious for $i \neq n$ and C_{b_n} is equivalent to Σ_1-sentence since it states that there is a $\mathsf{PA}+\Diamond^{h(a)-1}\top$-proof of $0 = 1$ and in addition it states that the least $\mathsf{PA} + \Diamond^{h(a)-1}\top$-proof of $0 = 1$ satisfy certain $\Delta_0(\exp)$-property.

We assign sentences F_a to all the worlds a of $\mathfrak{F}$. The sentence F_a is

$$\bigwedge_{b\preceq a} \mathsf{C}_b \wedge \Diamond^{h(a)}\top.$$

It is easy to see that the disjunction of all F_a's is provable in PA and any conjunction $\mathsf{F}_a \wedge \mathsf{F}_b$, for $a \neq b$, is disprovable in PA.

Lemma 1. *For any set of worlds A we have*

$$\mathsf{PA} + \Box^{h(r)+1}\bot \vdash \Diamond\Big(\bigvee_{a\in A} \mathsf{F}_a\Big) \leftrightarrow \bigvee_{b,\exists a\in A(b\prec a)} \mathsf{F}_b.$$

Let us first prove Theorem 1 using Lemma 1 and only then prove the lemma.

Proof. For a variable v we assign the evaluation $f(v)$:

$$\bigvee_{\mathbf{M},a\Vdash v} \mathsf{F}_a.$$

By induction on the length of modal formulas ψ we prove that

$$\mathsf{PA} + \Box^{h(r)+1}\bot \vdash f(\psi) \leftrightarrow \bigvee_{\mathbf{M},a\Vdash \psi} \mathsf{F}_a.$$

The only non-trivial case for the induction step is when the topmost connective of ψ is modality. Assume ψ is of the form $\Box\chi$. From inductive assumption we know that

$$\mathsf{PA} \vdash \Box^{h(r)+1}\bot \to (f(\chi) \leftrightarrow \bigvee_{\mathbf{M},a\Vdash \chi} \mathsf{F}_a).$$

We use Lemma 1:

$$\begin{aligned}
\mathsf{PA} + \Box^{h(r)+1}\bot \vdash f(\Box\chi) &\leftrightarrow \Box(f(\chi)) \\
&\leftrightarrow \Box(\Box^{h(r)+1}\bot \wedge f(\chi)) \\
&\leftrightarrow \Box(\Box^{h(r)+1}\bot \wedge \bigvee_{\mathbf{M},a\Vdash \chi} \mathsf{F}_a) \\
&\leftrightarrow \Box(\bigvee_{\mathbf{M},a\Vdash \chi} \mathsf{F}_a) \\
&\leftrightarrow \Box(\neg \bigvee_{\mathbf{M},a\Vdash \neg\chi} \mathsf{F}_a) \\
&\leftrightarrow \neg\Diamond(\bigvee_{\mathbf{M},a\Vdash \neg\chi} \mathsf{F}_a) \\
&\leftrightarrow \neg \bigvee_{\mathbf{M},a\Vdash \Diamond\neg\chi} \mathsf{F}_a. \\
&\leftrightarrow \bigvee_{\mathbf{M},a\Vdash \Box\chi} \mathsf{F}_a.
\end{aligned}$$

Therefore,

$$\mathsf{PA} + \Box^{h(r)+1}\bot \vdash f(\varphi) \leftrightarrow \bigvee_{\mathbf{M},a \Vdash \varphi} \mathsf{F}_a.$$

Since $\mathbf{M}, r \nVdash \varphi$, we have $\mathsf{PA} + \Box^{h(r)+1}\bot + \mathsf{F}_r \vdash \neg f(\varphi)$. The sentence F_r is just equivalent to $\Diamond^{h(r)}\top$. Hence, by Gödel's Second Incompleteness Theorem for $\mathsf{PA} + \Diamond^{h(r)}\top$, the theory $\mathsf{PA} + \Box^{h(r)+1}\bot + \mathsf{F}_r$ is consistent. Therefore, $\neg f(\varphi)$ is consistent with PA and thus $\mathsf{PA} \nvdash f(\varphi)$.

In order to prove Lemma 1, clearly, it will be enough to prove the following two lemmas:

Lemma 2. *For any world a from $\mathfrak{F}$, we have*

$$\mathsf{PA} + \Box^{h(r)+1}\bot \vdash \Diamond\mathsf{F}_a \rightarrow \bigvee_{b \prec a} \mathsf{F}_b.$$

Proof. Let us reason in $\mathsf{PA} + \Box^{h(r)+1}\bot$. Assume $\Diamond\mathsf{F}_a$. We need to prove $\bigvee_{b \prec a} \mathsf{F}_b$. Let us denote by $r = c_0 \prec c_1 \prec \ldots \prec c_n = a$ the maximal chain from r to a. Let us find the greatest k such that C_{c_k} holds.

Note that for any $1 \le i \le n$ the sentence $\Box^{h(c_{i-1})}\bot$ implies C_{c_i}. Indeed, $\Box^{h(c_{i-1})}\bot$ implies that C_c for some immediate successor c of c_{i-1}. But since C_c is Σ_1 and we assumed $\Diamond\mathsf{F}_a$, we would have $\Diamond(\mathsf{F}_a \wedge \mathsf{C}_c)$, which is possible only for $c = c_i$.

By a simple check of cases $k = 0$ and $k \neq 0$ we obtain $\Box^{h(c_k)+1}\bot$. Therefore, for all $i < k$, we have $\Box^{h(c_i)}\bot$ and hence, for all $i \le k$, the sentence C_{c_i} holds. From $\Box(\mathsf{F}_a \rightarrow \Diamond^{h(a)}\top)$ and $\Diamond\mathsf{F}_a$ we derive $\Diamond^{h(a)+1}\top$. Thus, $\neg\mathsf{C}_a$ and hence $k < n$. Since $\Box^{h(c_k)}\bot$ implies $C_{c_{k+1}}$, we have $\Diamond^{h(c_k)}\top$. Therefore the sentence F_{c_k} holds and finally we derive $\bigvee_{b \prec a} \mathsf{F}_b$.

Lemma 3. *For any worlds $a \prec b$, we have* $\mathsf{PA} + \Box^{h(r)+1}\bot \vdash \mathsf{F}_a \rightarrow \Diamond\mathsf{F}_b$.

We will use model-theoretic methods in our proof of Lemma 3. More precisely, we will need to use within PA some facts that we will establish using model-theoretic methods. There is an approach to formalization in arithmetic of results obtained by model-theoretic methods that is based on the use of the systems of the second-order arithmetic. In particular there is a well-known system ACA_0 that is a conservative extension of PA. We will use the formalization of model-theoretic notions in systems of second-order arithmetic that could be found in Simpson book [Sim09, Sects. II.8 and IV.3].

The key model-theoretic result that we use is the Injecting Inconsistencies Theorem. We will use the version of the theorem that is a corollary of the version of the theorem that were proved by Visser and Verbrugge [VV94, Theorem 5.1]. Earlier similar results are due to Hájek, Solovay, Krajíček, and Pudlák [Há84, Sol89, KP89].

Definition 1. *Suppose* $\mathfrak{M}$ *is a model of* PA*. We denote by* $\mathfrak{M}\restriction a$ *the structure with the domain* $\{e \in \mathfrak{M} \mid \mathfrak{M} \models e \leq a\}$ *the constant* 0 *and partial functions* S, $+$, *and* $\cdot$ *induced by* $\mathfrak{M}$ *on the domain. For two structures* $\mathfrak{A}$ *and* $\mathfrak{B}$ *with the constant* 0 *and (maybe) partial functions* S, $+$, *and* $\cdot$ *we write*

1. $\mathfrak{A} \subseteq \mathfrak{B}$ *if the domain of* $\mathfrak{A}$ *is a subset of the domain of* $\mathfrak{B}$ *and for any arithmetical term* $t(x_1,\ldots,x_n)$ *and elements* $q_1,\ldots,q_n \in \mathfrak{A}$*:*
   (a) *if* p *is the value of* $t(q_1,\ldots,q_n)$ *in* $\mathfrak{B}$ *and* $p \in \mathfrak{A}$ *then the value of* $t(q_1,\ldots,q_n)$ *is defined in* $\mathfrak{A}$ *and is equal to* p*,*
   (b) *if* p *is the value of* $t(q_1,\ldots,q_n)$ *in* $\mathfrak{A}$ *then the value of* $t(q_1,\ldots,q_n)$ *is defined in* $\mathfrak{B}$ *and is equal to* p*;*
2. $\mathfrak{A} = \mathfrak{B}$ *if* $\mathfrak{A} \subseteq \mathfrak{B}$ *and* $\mathfrak{B} \subseteq \mathfrak{A}$*.*

We note that the definition actually could also be applied to models of $\mathsf{I\Delta_0}$*.*

We will show in Appendix B that the following theorem is formalizable in $\mathsf{ACA_0}$:

Theorem 2. *Let* T *be an extension of* PA *by finitely many axioms. Let* $\mathsf{Con_T}(x)$ *denote the formula* $\forall y(\log(y) \leq x \to \neg\mathsf{Prf_T}(y, \ulcorner 0=1 \urcorner))$*. Let* $\mathfrak{M}$ *be a nonstandard countable model of* T*. And let* q, p *be nonstandard elements of* $\mathfrak{M}$ *such that* $\mathfrak{M} \models q \leq p$ *and* $\mathfrak{M} \models \mathsf{Con_T}(p^k)$*, for all standard* k*. Then there exists a countable model* $\mathfrak{N}$ *of* T *such that* $p \in \mathfrak{N}$ *and*

1. $\mathfrak{M}\restriction p = \mathfrak{N}\restriction p$*;*
2. $\mathfrak{M}\restriction \exp(p^k) \subseteq \mathfrak{N}$*, for all standard* k*;*
3. $\mathfrak{N} \models \neg\mathsf{Con_T}(p^q)$*;*
4. $\mathfrak{N} \models \mathsf{Con_T}(p^k)$*, for all standard* k*.*

Let us now prove Lemma 3 using the formalization of Theorem 2.

Proof. It would be enough to prove the lemma for the case when b is an immediate successor of a. Indeed, after that we will be able to derive $\Diamond^n\mathsf{F}_b$ for any b, $a \prec b$, where n is the length of the maximal chain from a to b; next we could conclude that we have the required $\Diamond\mathsf{F}_b$.

Now let us consider the case when b is an immediate successor of a and is b_k in our fixed order $b_0,\ldots,b_n$ of the immediate successors of a.

For the rest of the proof we reason in $\mathsf{ACA_0} + \mathsf{F}_a + \Box^{h(r)+1}\bot$ in order to show that we have $\Diamond\mathsf{F}_b$; since $\mathsf{ACA_0}$ is a conservative extension of PA, this will conclude the proof.

Since we have $\Diamond^{h(a)}\top$, we could construct a model $\mathfrak{M}$ of $\mathsf{PA} + \Diamond^{h(a)-1}\top$. Suppose $v \in \mathfrak{M}$ is the least $\mathsf{PA} + \Diamond^{h(a)-1}\top$-proof of $0=1$ in $\mathfrak{M}$, if there exists one and an arbitrary nonstandard number, otherwise. Note that since we have $\Diamond^{h(a)}\top$, the element v couldn't be standard. Next we find some nonstandard $u \in \mathfrak{M}$ such that

1. $\mathfrak{M} \models \exp(\exp(u)) < v$,
2. $\mathfrak{M} \models \log^\star(u+1) \equiv k-1 \pmod{n+1}$,
3. $\mathfrak{M} \models \log^\star(u) \equiv k-2 \pmod{n+1}$.

We can find u with this properties since we know that the functions $\exp(x)$ and $\exp^\star(x)$ are total on standard natural numbers and hence we know that the functions $\log(x)$ and $\log^\star(x)$ map nonstandard elements to nonstandard elements in $\mathfrak{M}$.

Now we apply Theorem 2 to the model $\mathfrak{M}$ with $p = u$ and $q = \log(u) + 1$. We obtain a model $\mathfrak{M}'$ of $\mathsf{PA} + \Diamond^{h(a)-1}\top$ such that $\mathfrak{M} \upharpoonright u = \mathfrak{M}' \upharpoonright u$ and there is the least $\mathsf{PA} + \Diamond^{h(a)-1}\top$-proof $d \in \mathfrak{M}'$ of $0 = 1$ such that

$$\mathfrak{M}' \models u + 1 < u^2 < \log(d) \leq u^{\log(u)+1} \leq \exp((\log(u) + 1)^2) < \exp(u).$$

Thus,

$$\mathfrak{M}' \models \log^\star(d) \equiv k \pmod{n + 1}.$$

If $h(b) = h(a) - 1$, then we have constructed a model of $\mathsf{PA} + \mathsf{C}_b + \Diamond^{h(b)}\top$.

Assume $h(b) < h(a) - 1$. Clearly, there are no $\mathsf{PA} + \Diamond^{h(b)}\top$-proofs of $0 = 1$ in $\mathfrak{M}'$. We apply Theorem 2 to $\mathfrak{M}'$ with $p = d^{\log(d)+1}$ and $q = \log(d) + 1$. We obtain a model $\mathfrak{M}''$ of $\mathsf{PA} + \Diamond^{h(b)}\top$ such that

$$\mathfrak{M}' \upharpoonright d^{\log(d)+1} = \mathfrak{M}'' \upharpoonright d^{\log(d)+1},$$

there is a $\mathsf{PA} + \Diamond^{h(b)}\top$-proof of $0 = 1$ in $\mathfrak{M}''$ and for the least $\mathsf{PA} + \Diamond^{h(b)}\top$-proof $e \in \mathfrak{M}''$ of $0 = 1$ we have

$$\mathfrak{M}'' \models \log(e) \leq d^{(\log(d)+1)^2} \leq \exp((\log(d) + 1)^3) < \exp(d).$$

Since $\mathfrak{M}' \upharpoonright d^{\log(d)+1} = \mathfrak{M}'' \upharpoonright d^{\log(d)+1}$ and $\mathsf{Prf}(x, y)$ is a Δ_0 predicate, we see that d is the least $\mathsf{PA} + \Diamond^{h(a)-1}\top$-proof of $0 = 1$ in $\mathfrak{M}''$. Hence $\mathfrak{M}''$ is a model of $\mathsf{PA} + \mathsf{C}_b + \Diamond^{h(b)}\top$.

Thus, under no additional assumptions, we have a model of $\mathsf{PA}+\mathsf{C}_b+\Diamond^{h(b)}\top$. Since all C_c, for $c \preceq a$, are Σ_1-sentences, actually we have a model of $\mathsf{PA} + \mathsf{F}_b$. Therefore, $\Diamond\mathsf{F}_b$.

4 Conclusions

In the present paper we have gave a new method of constructing arithmetical evaluations of modal formulas from a given Kripke model and proved arithmetical completeness of GL with respect to provability in PA using the method. We consider the evaluations that have been constructed in the paper to be more "natural" than the evaluations provided by Solovay's proof.

We proved the theorem specifically for the standard provability predicate for PA. It is unclear to author, for what exact class of provability predicates our methods are applicable. The most essential limitation for our technique seems to be the fact that it relies on the formalized version of Theorem 2. It seems very likely that for theories that are stronger than PA one could apply our method with only minor adjustments. In particular, it seems that for a general result one would need to modify Prf-predicates while preserving Prv-predicate (up to provable equivalence) in order to ensure that [VV94, Theorem 5.1] is applicable.

For theories that are weaker than PA, there are more significant problems with adopting our technique. Namely, our technique essentially relies on formalized version of the Injecting Inconsistencies Theorem. And the proofs of stronger versions of this theorem [KP89, VV94] essentially rely on the Omitting Types Theorem. We have provided a proof of the Omitting Types Theorem in ACA_0 in Appendix A, but it is not clear whether it could be done in weaker systems. The author is not familiar with results that calibrate reverse mathematics strength of the required version of the Omitting Types Theorem. We note that reverse mathematics analysis of other version of Omitting Types Theorem have been done by Hirschfeldt et al. [HSS09], in particular from their results it follows that their version of the Omitting Types Theorem is not provable in WKL_0 but follows from RT^2_2 (and thus couldn't be equivalent to ACA_0 over RCA_0). But nevertheless, we conjecture that the same kind of evaluations as we have gave in Sect. 3 will provide completeness of GL for all finitely axiomatizable extensions of $\mathsf{I}\Delta_0 + \mathsf{Exp}$.

Also, since the technique that were introduced in the paper is significantly different from Solovay's technique, it seems plausible that it may give some advantage for some open problems, for which Solovay's method have been the "default approach" before (see [BV06] for open problems in provability logic).

Acknowledgments. I want to thank David Fernández-Duque and Albert Visser for their questions that were an important part of the reason why I have started the research on the subject. And I want to thank Paula Henk, Vladimir Yu. Shavrukov, and Albert Visser for their useful comments on an early draft of the paper.

A Formalization of the Omitting Types Theorem

In order to formalize Theorem 2 in ACA_0 we will first show that the Omitting Types Theorem is formalizable in ACA_0. We will adopt the proof from [CK90]. We remind a reader that we use the approach to formalization of model theory from Simpson book [Sim09].

Definition 2 (ACA_0)**.** *Let* T *be a first-order theory and* $\Sigma = \Sigma(x_1, \ldots, x_n)$ *be a set of formulas of the language of* T *that have no free variables other than* $x_1, \ldots, x_n$*. We say that* T locally omits Σ *if for every formula* $\varphi(x_1, \ldots, x_n)$ *at least one of the following fails:*

1. *the theory* $\mathsf{T} + \varphi$ *is consistent;*
2. *for all* $\psi \in \Sigma$ *we have* $\mathsf{T} \vdash \forall x_1, \ldots, x_n(\varphi \to \psi)$*.*

We say that a model $\mathfrak{M}$ *of* T omits Σ *if for any* $a_1, \ldots, a_n \in \mathfrak{M}$ *there is a formula* $\psi(x_1, \ldots, x_n) \in \Sigma$ *such that* $\mathfrak{M} \not\models \psi(a_1, \ldots, a_n)$*.*

Theorem 3 (ACA_0)**.** *Suppose* T *is a consistent theory that locally omits the set of formulas* $\Sigma(x_1, \ldots, x_n)$*. Then there is a model* $\mathfrak{M}$ *of* T *that omits the set* Σ*.*

Proof. We will follow the proof of [CK90, Theorem 2.2.9] but make sure that our arguments could be carried out in ACA_0.

We will prove the theorem for $n = 1$, i.e. $\Sigma = \Sigma(x)$. The case $n > 1$ could be proved essentially the same way, but the notations would be more complicated.

We extend the language of T by fresh constants $c_0, c_1, \ldots$. We arrange all sentences of the extended language in a sequence $\varphi_0, \varphi_1, \ldots$ (since we work in ACA_0 the formulas are encoded by Gödel numbers and we could arrange them by their Gödel numbers). We will construct a sequence of finite sets of sentences

$$\emptyset = \mathsf{U}_0 \subset \mathsf{U}_1 \subset \ldots \subset \mathsf{U}_m \subset \ldots$$

such that for every m we have the following:

1. U_m is consistent with T;
2. either $\varphi_m \in \mathsf{U}_{m+1}$ or $\neg\varphi_m \in \mathsf{U}_{m+1}$;
3. if φ_m is of the form $\exists x \psi(x)$ and $\varphi_m \in \mathsf{U}_{m+1}$ then $\psi(c_p) \in \mathsf{U}_{m+1}$, where c_p is the first c_i that doesn't occur in U_m or φ_m;
4. there is a formula $\chi(x) \in \Sigma$ such that $\neg\chi(c_m) \in \mathsf{U}_{m+1}$.

We will give the definition that will determine unique sequence $\mathsf{U}_0, \mathsf{U}_1, \ldots$. We want to make sure that for our definition of the sequence $\mathsf{U}_0, \mathsf{U}_1, \ldots$, the property of a number x to be the code of the sequence $\langle \mathsf{U}_0, \mathsf{U}_1, \ldots, \mathsf{U}_y \rangle$ is expressible by a formula without second-order quantifiers. If we will ensure this, then we will be able to construct a set that encodes the sequence $\mathsf{U}_0, \mathsf{U}_1, \ldots, \mathsf{U}_m, \ldots$ using the arithmetic comprehension.

Let us define U_{m+1} in terms of U_m. If φ_m is consistent with $\mathsf{T} \cup \mathsf{U}_m$ then we put σ_m to be φ_m. Otherwise we put σ_m to be $\neg\varphi_m$. If σ_m is φ_m and is of the form $\exists x \psi(x)$ then we put ξ_m to be $\psi(c_p)$, where c_p is the first c_i that doesn't occur in U_m or φ_m. Otherwise, we put ξ_m to be equal to σ_m. We choose the formula $\chi(x)$ with the smallest Gödel number such that $\chi(x) \in \Sigma$ and $\mathsf{T} \nvdash \bigwedge \mathsf{U}_m \to \chi(c_m)$. We put $\mathsf{U}_{m+1} = \mathsf{U}_m \cup \{\xi_m, \sigma_m, \chi(c_m)\}$.

It is easy to see that for this definition, indeed, we could express by a formula without second-order quantifiers the property of a number x to be the code of the sequence $\langle \mathsf{U}_0, \mathsf{U}_1, \ldots, \mathsf{U}_y \rangle$. By a trivial induction on y we could prove that for every y the said sequence exists and unique. Thus, we have obtained the sequence $\mathsf{U}_0, \mathsf{U}_1, \ldots, \mathsf{U}_m, \ldots$ encoded by a set.

Now, using the definition of the sequence, we could easily prove that the sequence satisfy the conditions 1, 2, 3, and 4.

We consider the union $\mathsf{T} \cup \bigcup\limits_{i \in \mathbb{N}} \mathsf{U}_i = \mathsf{T}'$. By condition 1. the theory T' is consistent. By condition 2. the theory T' is complete. By condition 3. the theory T' gives the truth definition with Tarski conditions for a model with the domain $\{c_0, c_1, \ldots\}$; this gives us a model $\mathfrak{M}$ of T' with the domain $\{c_0, c_1, \ldots\}$. By condition 4. The model $\mathfrak{M}$ omits the set Σ.

B Formalization of the Injecting Inconsistencies Theorem

Now we are going to check that Theorem 2 is provable in ACA_0. Below we assume that a reader is familiar with the paper [VV94] and we will use some notions from the paper without giving the definitions here.

Theorem 4. *Let* $\mathsf{R} \subset \mathsf{I\Delta}_0 + \Omega_1$ *be a finitely axiomatizable theory. Then* ACA_0 *proves the following:*

Let $\mathsf{T} \supseteq \mathsf{I\Delta}_0 + \Omega_1$ *be a* Σ^b_1*-axiomatized theory for which the small reflection principle is provable in* R*. Let* $\mathsf{Con}_\mathsf{T}(x)$ *denote the formula* $\forall y(\log(y) \le x \to \neg\mathsf{Prf}_\mathsf{T}(y, \ulcorner 0 = 1 \urcorner))$*. Let* $\mathfrak{M}$ *be a non-standard model of* T *and let* c*,* a *be non-standard elements of* $\mathfrak{M}$ *such that* $\mathfrak{M} \models c \le a$*,* $\exp(a^c) \in \mathfrak{M}$*, and* $\mathfrak{M} \models \mathsf{Con}_\mathsf{T}(a^k)$*, for all standard* k*. Then there exists a model* $\mathfrak{K}$ *of* T *such that* $a \in \mathfrak{K}$ *and*

1. $\mathfrak{M} \upharpoonright a = \mathfrak{K} \upharpoonright a$*;*
2. $\mathfrak{M} \upharpoonright \exp(a^k) \subseteq \mathfrak{K}$*, for all standard* k*;*
3. $\mathfrak{K} \models \neg\mathsf{Con}_\mathsf{T}(a^c)$*;*
4. *for all standard* k *we have* $\mathfrak{K} \models \mathsf{Con}_\mathsf{T}(a^k)$*;*
5. $\mathfrak{K} \models \exp(a^c) \downarrow$*.*

Proof. Essentially, we just need to formalize the proof of [VV94, Theorem 5.1] in ACA_0. The only difference between our formulation and the formulation by Visser and Verbrugge is that we have replaced the requirement that the small reflection principle is provable in $\mathsf{I\Delta}_0 + \Omega_1$ with a stronger requirement that states that the small reflection principle is provable in R. First, we show how to formalize the proof itself and then explain why the results used in the proof are formalizable in ACA_0.

The only non-trivial part of the formalization of the proof itself is the issue with the lack of truth definition for the cut

$$\mathfrak{N} = \{u \in \mathfrak{M} \mid u < \exp(a^k), \text{ for some standard } k\}$$

of $\mathfrak{M}$. However, for the purposes of the proof, it would be enough for $\mathfrak{N}$ to be a weak model (i.e. poses truth definition only for axioms, [Sim09, Definition II.8.9]). Moreover, unlike the original proof of Visser and Verbrugge, we just need $\mathfrak{N}$ to be a weak model of $\mathsf{R} + \mathsf{B\Sigma}_1$ rather than a model of $\mathsf{B\Sigma}_1 + \Omega_1$. And since R is externally fixed finitely axiomatizable theory, we could create the required truth definition straightforward using arithmetical comprehension. Other parts of the proof could be formalized without any complications.

The proof of [VV94, Theorem 5.1] used Wilkie and Paris result [WP89, Theorem 1], Pudlák results from [Pud86], and the Omitting Types Theorem. We have already formalized the Omitting Types Theorem in Appendix A. The proof of [WP89, Theorem 1] is trivial and could be easily formalized in ACA_0. The technique of [Pud86] is purely finitistic and thus could be easily formalized in ACA_0.

Now we want to derive the formalization of Theorem 2 from Theorem 4. In order to do it, we first need to fix some finite fragment $\mathsf{R} \subset \mathsf{I\Delta}_0 + \Omega_1$. And next we need to show in ACA_0 that all the extensions of PA by finitely many axioms are Σ_1^b-axiomatizable extensions of $\mathsf{I}\Delta_0 + \Omega_1$ for which R proves the small reflection principle. Obviously, extensions of PA by finitely many axioms are Σ_1^b-axiomatizable (and it could be checked in ACA_0).

In [VV94, Theorem 4.20] it were established that $\mathsf{I}\Delta_0 + \Omega_1$ proves small reflection principle for $\mathsf{I}\Delta_0 + \Omega_1$. By inspecting the proof, it is easy to see that it is possible to use only finitely many axioms of $\mathsf{I}\Delta_0 + \Omega_1$ in order to prove all the instances of the small reflection principle. Now we will indicate how to modify the proof of [VV94, Theorem 4.20] in order to prove in a finite fragment of $\mathsf{I}\Delta_0 + \Omega_1$ all the instances of the small reflection principle for all the extensions of PA by finitely many axioms. Actually, the only part of the proof that should be changed is [VV94, Lemma 4.16] that were needed to deal with the schema of Δ_0-induction schema in the case of $\mathsf{I}\Delta_0 + \Omega_1$-provability. For our adaptation we need to replace it with the analogous lemma that will deal with schema of full induction in the case of provability in PA. This analogous lemma could be proved essentially in the same way as [VV94, Lemma 4.16] itself with the only difference that the last part of the proof that were reducing an instance of induction schema to an instance of Δ_0-induction schema will not be needed any longer. This concludes the proof of Theorem 2 in ACA_0.

References

[Ber90] Berarducci, A.: The interpretability logic of Peano arithmetic. J. Symbolic Logic **55**(3), 1059–1089 (1990)

[Boo95] Boolos, G.: The Logic of Provability. Cambridge University Press, Cambridge (1995)

[BV06] Beklemishev, L., Visser, A.: Problems in the logic of provability. In: Gabbay, D.M., Goncharov, S.S., Zakharyaschev, M. (eds.) Mathematical Problems from Applied Logic I. International Mathematical Series, pp. 77–136. Springer, New York (2006)

[CK90] Chang, C.C., Keisler, H.J.: Model Theory, vol. 73. Elsevier, Amsterdam (1990)

[dJJM91] de Jongh, D., Jumelet, M., Montagna, F.: On the proof of Solovay's theorem. Studia Logica: Int. J. Symbolic Logic **50**(1), 51–69 (1991)

[Gö33] Gödel, K.: Ein Interpretation des intuitionistischen Aussagenkalküls. In Ergebnisse eines mathematischen Kolloquiums, vol. 4, pp. 39–40. Oxford (1933). Reprinted: An Interpretation of the Intuitionistic Propositional Calculus, Feferman, S, ed. Gödel Collected Works I publications, 1929–1936

[Há84] Hájek, P.: On a new notion of partial conservativity. In: Börger, E., Oberschelp, W., Richter, M.M., Schinzel, B., Thomas, W. (eds.) Computation and Proof Theory. LNM, vol. 1104, pp. 217–232. Springer, Heidelberg (1984). doi:10.1007/BFb0099487

[HSS09] Hirschfeldt, D., Shore, R., Slaman, T.: The atomic model theorem and type omitting. Trans. Am. Math. Soc. **361**(11), 5805–5837 (2009)

[Jap86] Japaridze, G.K.: The modal logical means of investigation of provability. Thesis in Philosophy, in Russian, Moscow (1986)

[Kot96] Kotlarski, H.: An addition to Rosser's theorem. J. Symbolic Logic **61**(1), 285–292 (1996)

[KP89] Krajíček, J., Pudlák, P.: On the structure of initial segments of models of arithmetic. Arch. Math. Logic **28**(2), 91–98 (1989)

[Lö55] Löb, M.H.: Solution of a problem of Leon Henkin. J. Symbolic Logic **20**(02), 115–118 (1955)

[Pak16] Pakhomov, F.: Semi-provability predicates and extensions of GL. In: 11th International Conference on Advances in Modal Logic, Short Presentations, pp. 110–115 (2016)

[Pud86] Pudlák, P.: On the length of proofs of finitistic consistency statements in first order theories. In: Logic Colloquium, vol. 84, pp. 165–196. Amsterdam, North-Holland (1986)

[Seg71] Segerberg, K.: An essay in classical modal logic. Ph.D. thesis, Uppsala:: Filosofiska Föreningen och Filosofiska Institutionen vid Uppsala Universitet (1971)

[Sha88] Shavrukov, V.Y.: The logic of relative interpretability over Peano arithmetic. Technical report, (5) (1988). Moscow: Steklov Mathematical Institute (in Russian)

[Shi09] Shipman, J.: Only one proof, 2009. FOM mailing list. http://cs.nyu.edu/pipermail/fom/2009-August/013994.html

[Sim09] Simpson, S.G.: Subsystems of Second Order Arithmetic, vol. 1. Cambridge University Press, Cambridge (2009)

[Smi63] Smiley, T.J.: The logical basis of ethics. Acta Philosophica Fennica **16**, 237–246 (1963)

[Sol76] Solovay, R.M.: Provability interpretations of modal logic. Israel J. Math. **25**, 287–304 (1976)

[Sol86] Solovay, R.M.: 12 May 1986. Letter by R. Solovay to E. Nelson

[Sol89] Solovay, R.M.: Injecting inconsistencies into models of PA. Ann. Pure Appl. Logic **44**(1–2), 101–132 (1989)

[VV94] Verbrugge, R., Visser, A.: A small reflection principle for bounded arithmetic. J. Symbolic Logic **59**(03), 785–812 (1994)

[WP89] Wilkie, A., Paris, J.: On the existence of end extensions of models of bounded induction. Stud. Logic Found. Math. **126**, 143–161 (1989)