The Σ_1-Provability Logic of HA

Mohammad Ardeshir*
Department of Mathematical Sciences
Sharif University of Technology

S. Mojtaba Mojtahedi†
Department of Mathematics,
Statistics and Computer Science,
College of Sciences, University of Tehran

January 14, 2017

Abstract

In this paper we introduce a modal theory iH_σ which is sound and complete for arithmetical Σ_1-interpretations in HA, in other words, we will show that iH_σ is the Σ_1-provability logic of HA. Moreover we will show that iH_σ is decidable. As a by-product of these results, we show that $\mathsf{HA}+\Box\bot$ has de Jongh property.

Contents

*mardeshir@sharif.ir
†mojtaba.mojtahedi@ut.ac.ir

1 Introduction

As far as we know, there are at least two updated reliable sources [AB04, BV06] for current situation, historical background and motivations for *provability logic*. To be self-contained, in this introduction, we extract a brief backgrounds of provability logic from the mentioned sources for readers not much familiar with the subject.

Provability Logic is a modal logic in which the modal operator $\Box$ has intended meaning of provability in some formal system. Unlike the other realms of modal logic, e.g. temporal logic, epistemic logic and deontic logic, here in provability logic, we have a rational meaning for $\Box A$:

"A is provable in the system T"

The notion of provability logic goes back essentially to K. Gödel [Göd33] in 1933, where he intended to provide a semantics for Heyting's formalization of *intuitionistic logic* IPC. He defined a *translation*, or *interpretation* τ from the propositional language to the modal language such that

$$\mathsf{IPC} \vdash A \quad \Longleftrightarrow \quad \mathsf{S4} \vdash \tau(A).$$

The translation $\tau(A)$ adds a $\Box$ before each sub-formula of A. The idea behind this translation is hidden in the intuitionistic meaning of truth (the BHK interpretation): "The truth of a proposition coincides with its provability". Hence if one assumes $\Box A$ as "provability of A", then it is reasonable to add a $\Box$ behind each sub-formula and expect to have a correspondence between the intuitionistic propositional calculus IPC and some classical modal logic.

On the other hand, by works of Gödel in [Göd31], for each arithmetical formula A and recursively axiomatizable theory T (like PA), we can formalize the statement "there exists a proof in T for A" by a sentence of the language of arithmetic, i.e. $\mathsf{Prov}_T(\ulcorner A \urcorner) := \exists x\, \mathsf{Proof}_T(x, \ulcorner A \urcorner)$, where $\ulcorner A \urcorner$ is the code of A. Now the question is whether we can find some modal propositional theory such that the $\Box$ operator captures the notion of *provability* in classical mathematics. Let us restrict our attention to the part of mathematics known as Peano Arithmetic PA. Hence the question is to find some propositional modal theory $T_\Box$ such that:

$$T_\Box \vdash A \quad \Longleftrightarrow \quad \forall * \ \mathsf{PA} \vdash A^*$$

By $(\ )^*$, we mean a mapping from the modal language to the first-order language of arithmetic, such that

- p^* is an arithmetical first-order sentence, for any atomic variable p, and $\bot^* = \bot$,
- $(A \circ B)^* = A^* \circ B^*$, for $\circ \in \{\vee, \wedge, \to\}$,
- $(\Box A)^* := \exists x\, \mathsf{Proof}_{\mathsf{PA}}(x, \ulcorner A^* \urcorner)$.

It turned out that S4 is *not* a right candidate for interpreting the notion of *provability*, since $\neg\Box\bot$ is a theorem of S4, contradicting Gödel's second incompleteness theorem (Peano Arthmetic PA, does not prove its own consistency).

In 1976, R. Solovay [Sol76] proved that the right modal logic, in which the $\Box$ operator interprets the notion of *provability in* PA, is GL. This modal logic is well-known as the Gödel-Löb logic, and has the following axioms and rules:

- all tautologies of classical propositional logic,
- $\Box(A \to B) \to (\Box A \to \Box B)$,
- $\Box A \to \Box\Box A$,
- Löb's axiom (L): $\Box(\Box A \to A) \to \Box A$,
- Necessitation Rule: $A/\Box A$,
- Modus ponens: $(A, A \to B)/B$.

Theorem. (Solovay) *For any sentence A in the language of modal logic, $\mathsf{GL} \vdash A$ if and only if for all interpretations $(\ )^*$, $\mathsf{PA} \vdash A^*$.*

There are many open problems which could be assumed as a generalization of the above theorem. A list of such problems could be found in [BV06]. Also a live list of open problems could be found in the homepage of Lev Beklemishev[1].

The question of generalizing Solovay's result from classical theories to intuitionistic ones, such as the intuitionistic counterpart of PA, well-known as Heyting Arithmetic HA, proved to be remarkably difficult [AB04]. This problem was taken up by A. Visser, D. de Jongh and their students. The problem of axiomatizing the provability logic of HA remains a major open problem since the end of 70s [AB04]. Precisely speaking, the problem of the provability logic of HA is as follows:

$$\text{Find a modal theory } \mathsf{iH} \text{ such that:} \quad \mathsf{iH} \vdash A \qquad \Longleftrightarrow \qquad \forall * \ \mathsf{HA} \vdash A^*$$

Note that in the above statement of the provability logic of HA, we have $(\Box A)^* := \mathsf{Prov}_{\mathsf{HA}}(\ulcorner A^* \urcorner)$. The following list contains important results about the provability logic of HA with arithmetical nature:

- Myhill 1973 and Friedman 1975. $\mathsf{iH} \nvdash \Box(A \vee B) \to (\Box A \vee \Box B)$, [Myh73, Fri75]
- Leivant 1975. $\mathsf{iH} \vdash \Box(A \vee B) \to \Box(\boxdot A \vee \boxdot B)$, in which $\boxdot A$ is a shorthand for $A \wedge \Box A$, [Lei75]
- Visser 1981. $\mathsf{iH} \vdash \Box\neg\neg\Box A \to \Box\Box A$ and $\mathsf{iH} \vdash \Box(\neg\neg\Box A \to \Box A) \to \Box(\Box A \vee \neg\Box A)$, [Vis81, Vis82]
- Iemhoff 2001. Introduced a uniform axiomatization of all known axiom schemas of iH in an extended language with a bimodal operator $\rhd$. In her Ph.D. dissertation [Iem01], Iemhoff raised a conjecture that implies directly that her axiom system, iPH, restricted to the normal modal language, is equal to iH, [Iem01]
- Visser 2002. Introduced a decision algorithm for $\mathsf{iH} \vdash A$, for all A not containing any atomic variable. [Vis02]

In this paper, we introduce an axiomatization of a modal logic iH_σ and prove the following result which partially answers the question.

We first show that any TNNIL-proposition[2] A is in the Σ_1-provability logic of HA, iff $\mathsf{iGL}{+}\mathsf{CP} \vdash A$, where iGL is the intuitionistic Gödel-Löb's logic and CP is the *completeness principle* $B \to \Box B$ (we call this theory as LC). This fact in combination with the conservativity result of Theorem 4.24 and also some variant of Visser's NNIL-algorithm in [Vis02], implies that the Σ_1-provability logic of HA, is a decidable modal theory, that is called iH_σ here. More precisely, we find a system iH_σ such that

$$\mathsf{iH}_\sigma \vdash \mathsf{A} \qquad \Longleftrightarrow \qquad \forall * \ \mathsf{HA} \vdash \mathsf{A}^*,$$

[1] http://www.mi.ras.ru/~bekl

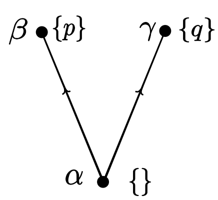
[2] We say that A is TNNIL, if any two nested occurrences of $\to$ in the left are separated by a $\Box$. For example $(p \to q) \to r$ and $\neg(p \to q)$ are not TNNIL, while $p \to q$ and $\Box(p \to q) \to r$ are TNNIL. Precise definition of TNNIL-propositions, a modal variant of NNIL-propositions [VvBdJRdL95], is in Section 4.1.2.

in which, $*$ range over all of the interpretations that p^* is a Σ_1-sentence for atomic variables p. The complete axiomatization of iH_σ is in Section 4.3. It is worth mentioning that a non-modal variant for all of the axioms of iH_σ, were already discovered by Visser in [Vis81, Vis82, Vis02]. He also showed in [Vis02] that those variant of axioms of iH_σ are sound for Σ_1 arithmetical interpretations in HA.

1.1 Inspiring examples

In the following four examples, we roughly explain the main roads in the paper. Before we continue with examples, let us review what we are going to do in this paper. Our main results are soundness and completeness theorems of iH_σ for arithmetical Σ_1-interpretations in HA. As usual, the difficult part is the completeness theorem. The soundness part is not problematic: some major part of soundness is already done by Visser [Vis02] and the rest (extended Leivant's principle) is done in Theorem 3.19. We are not going to talk about soundness in these examples. We explain how to refute some modal proposition A from the Σ_1-provability logic (and a fortiori from the provability logic) of HA, i.e. we will find some Σ_1-interpretation σ_{HA} such that $\mathsf{HA} \nvdash \sigma_{\mathsf{HA}}(A)$. The propositions which we treat here are $(p \to q) \vee (q \to p)$, $\Box(p \vee q) \to (\Box p \vee \Box q)$, $\neg\neg\Box(\neg\neg p \to p) \to \Box(\neg\neg p \to p)$ and finally $A := \Box(p \vee q) \to [(\Box p \to (p \vee q \vee \Box q)) \vee (\Box q \to (p \vee q \vee \Box p))]$.

Example 1.1. In this example we will show that how to refute the Dummet formula $A := (p \to q) \vee (q \to p)$ from the Σ_1-provability logic of HA. Since A is non-modal, we are actually faced with a special case of proving de Jongh property for HA with Σ_1-substitutions. C. Smoryńsk first discovered this result [Smo73a]. For a survey on the de Jongh property see [dJVV11]. Here we explain how to use Solovey's method [Sol76] in combination with Smorýnski's general method for defining first-order Kripke models of HA [Smo73b, page 372] to deduce the de Jongh property (with Σ_1-substitutions) for HA. We are not going to provide all details here, instead we explain the idea which motivated us to our main result Theorem 5.1. First we find a Kripke model $\mathcal{K}_0 \nVdash A$:

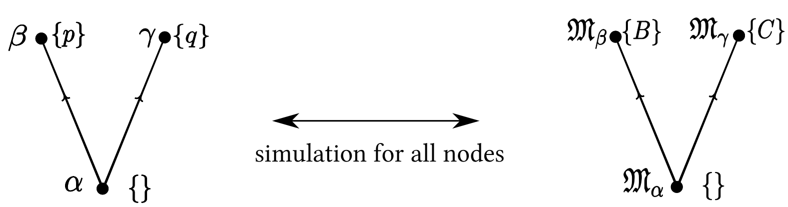


$$\beta \Vdash p \quad , \quad \beta \nVdash q \quad , \quad \gamma \Vdash q \quad , \quad \gamma \nVdash p$$

$$\alpha \leq \beta, \gamma \quad , \quad \alpha \nVdash p, q$$

In the left and right hand side of each node we wrote the name of that node and the set of atomic variables which are forced at that node, respectively. The precise definition of Kripke models for intuitionistic propositional logic IPC, came in Section 4.5.

Next we will find some arithmetical Σ_1-sentences B and C and also a first-order Kripke model $\mathcal{K}_1 \Vdash \mathsf{HA}$ such that $\mathcal{K}_1$ simulates $\mathcal{K}_0$, with B and C playing the role of p and q, respectively:



In the above picture, $\mathfrak{M}_\alpha$, $\mathfrak{M}_\beta$ and $\mathfrak{M}_\gamma$ are classical structures assigned to the corresponding nodes. For definition of intuitionistic first-order Kripke models, see Section 3.1.2.

To explain what are these classical structures and also what are the sentences B and C, we first define a recursive function F with the domain of natural numbers and with the range in the nodes of the Kripke model. Let us define:

$$B := \exists x(F(x) = \beta) \quad \text{and} \quad C := \exists x(F(x) = \gamma)$$

Since F is a recursive function, B and C are Σ_1 sentences. Moreover, for any $\delta \in \{\alpha, \beta, \gamma\}$, we assume the classical structures $\mathfrak{M}_\delta$ such that

$$\mathfrak{M}_\delta \models T_\delta \quad , \quad T_\delta := \mathsf{PA} + (\lim_{x\to\infty} F(x) = \delta)$$

In which $\lim_{x\to\infty} F(x) = \delta$ is defined as $\exists x \forall y \geq x F(y) = \delta$. The function F is defined as follows: $F(0) := \alpha_0$ and $F(n+1)$ is defined to be some node $\delta > F(n)$, if there exists some proof (in PA) with the Gödel number less than $n+1$ for the statement

$$\lim_{x\to\infty} F(x) \neq \delta$$

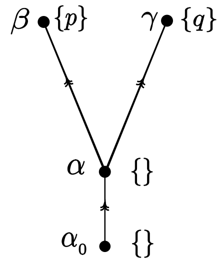
In other words, F climbs at stage $n+1$ to the node δ if there is a witness for the inconsistency of T_δ. This function is the same as the Solovey's function in [Sol76]. The recursive definition of F is such that although it is true that F is a constant function, PA can't prove it. Moreover, F is such that for any pair of nodes $\delta \lneq \delta'$, we have $T_\delta \vdash \mathsf{Con}(T_{\delta'})$, i.e.

$$\mathsf{PA} + \lim_{x\to\infty} F(x) = \delta \vdash \neg \mathsf{Prov}_{\mathsf{PA}}(\ulcorner \lim_{x\to\infty} F(x) \neq \delta' \urcorner)$$

This guaranties the existence of the classical structures $\mathfrak{M}_\delta \models T_\delta$, such that $\mathcal{K}_1$ is a first-order Kripke model of HA (see [Smo73a, Smo73b]). From $\mathfrak{M}_\beta \models T_\beta$, we can deduce that $\mathcal{K}_1, \beta \Vdash B$ and $\mathcal{K}_1, \beta \nVdash C$. From $\mathfrak{M}_\gamma \models T_\gamma$, we can deduce $\mathcal{K}_1, \gamma \Vdash C$ and $\mathcal{K}_1, \gamma \nVdash B$. These would imply that $\mathcal{K}_1, \alpha \nVdash (B \to C) \vee (C \to B)$, as desired.

Example 1.2. Let $A = \Box(p \vee q) \to (\Box p \vee \Box q)$. J. Myhill [Myh73] and H. Friedman [Fri75] have already shown that there exist some first-order arithmetical formulas B and C such that $\mathsf{HA} \nvdash \Box(B \vee C) \to (\Box B \vee \Box C)$, in other words, there exist some arithmetical substitution σ such that $\mathsf{HA} \nvdash \sigma_{\mathsf{HA}}(A)$, i.e. A does not belong to the provability logic of HA. However, their proof does not provide an explicit B and C. It only guarantees the existence of such arithmetical propositions. With the methods of this paper, we will find some explicit sentences B and C such that $\mathsf{HA} \nvdash \Box(B \vee C) \to (\Box B \vee \Box C)$.

As usual, we first find some Kripke model which refutes the proposition A. The Kripke models of intuitionistic modal logic have two relations: one for intuitionistic logic ($\leq$) which we illustrated it in the pictures with one arrow in the middle line, and another relation for modal connective ($\mathcal{R}$) which is illustrated with two arrows in middle line. All Kripke models in this paper have the following property: $\alpha \leq \beta \,\mathcal{R}\, \gamma$ implies $\alpha \,\mathcal{R}\, \gamma$. Also, in this example and other examples (Examples 1.1 to 1.4), the relations $\mathcal{R}$ and $\leq$ are transitive, $\mathcal{R} \subseteq \leq$ and moreover, $\mathcal{R}$ is irreflexive and $\leq$ is reflexive. In the pictures, we do not draw all the relations and always we assume the colsure of relations under the mentioned properties for the relations. For precise definition of Kripke semantics for intuitionistic modal logics, see Section 4.5. The Kripke counter-model for A is $\mathcal{K}_0$:

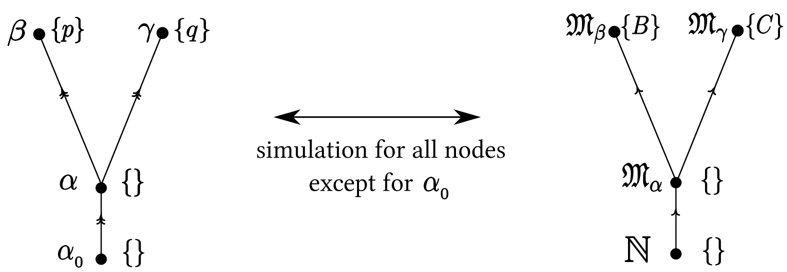


$$\beta \Vdash p \quad , \quad \beta \nVdash q \quad , \quad \gamma \Vdash q \quad , \quad \gamma \nVdash p$$

$$\alpha \, \mathcal{R} \, \beta,\gamma \quad , \quad \alpha \leq \beta, \gamma$$

$$\alpha_0 \, \mathcal{R} \, \alpha, \beta,\gamma \quad , \quad \alpha_0 \leq \alpha, \beta, \gamma \quad , \quad \alpha, \alpha_0 \nVdash p, q$$

As it may be observed, the node α_0 is not necessary. We add this extra (root) node, whenever we are not able to simulate the behaviour of the existing root of the tree. Theorem 4.26 ensures us that always for invalid propositions, such Kripke models exist. Next we will find some arithmetical sentences B and C and also a first-order Kripke model $\mathcal{K}_1 \Vdash \mathsf{HA}$ such that $\mathcal{K}_1$ simulates $\mathcal{K}_0$ with B and C playing the role of p and q, respectively:



In the above picture, $\mathfrak{M}_\alpha$, $\mathfrak{M}_\beta$ and $\mathfrak{M}_\gamma$ are classical structures assigned to the corresponding nodes and $\mathbb{N}$ indicates the standard model of arithmetic. To explain these classical structures and also the sentences B and C, we first define a recursive function F with the domain of natural numbers and with the range in the nodes of the Kripke model. Let us define:

$$B := \exists x(F(x) = \beta) \quad \text{and} \quad C := \exists x(F(x) = \gamma)$$

Moreover, for any $\delta \in \{\alpha, \beta, \gamma\}$, we assume the classical structures $\mathfrak{M}_\delta$ such that

$$\text{(1.1)} \quad \mathfrak{M}_\delta \models T_\delta, \quad T_\delta := \mathsf{PA} + (\lim_{x\to\infty} F(x) = \delta) + \mathsf{Prov}_{\mathsf{HA}}(\ulcorner \varphi_\delta \urcorner), \quad \varphi_\alpha := B \vee C, \quad \varphi_\beta := \varphi_\gamma := B \wedge C$$

Note that $\mathfrak{M}_\delta \models \mathsf{Prov}_{\mathsf{HA}}(\ulcorner \varphi_\delta \urcorner)$, and this implies $\mathcal{K}_1, \delta \Vdash \mathsf{Prov}_{\mathsf{HA}}(\ulcorner \varphi_\delta \urcorner)$. This means that the node δ in the first-order Kripke model forces the interpretation of those boxed propositions which are forced at δ in the propositional Kripke model $\mathcal{K}_0$. As we will see in Corollary 5.23:

$$\mathsf{PA} + (\lim_{x\to\infty} F(x) = \delta) \vdash \mathsf{Prov}_{\mathsf{HA}}(\ulcorner \varphi_\delta \urcorner)$$

Hence we may define T_δ simply as $\mathsf{PA} + (\lim_{x\to\infty} F(x) = \delta)$, instead of our previous definition in eq. (1.1).

The function F is defined as follows: $F(0) := \alpha_0$ and $F(n+1)$ is defined to be some node δ such that $F(n) \, \mathcal{R} \, \delta$, if there exists some proof (in PA) with the Gödel number less than $n+1$ for the statement

$$\neg[\lim_{x\to\infty} F(x) = \delta \wedge \mathsf{Prov}_{\mathsf{HA}}(\ulcorner \varphi_\delta \urcorner)]$$

Otherwise define $F(n+1) := F(n)$. In other words, F climbs at stage $n+1$ to the node δ if there is a witness for the inconsistency of T_δ with $\mathsf{Prov}_{\mathsf{HA}}(\ulcorner \varphi_\delta \urcorner)$. What is φ_δ? The proposition φ_δ is the conjunction of all propositions E such that $\Box E$ is forced at δ. Since the number of such propositions are infinite, we only take care of those E which are important to us, i.e. those which are a sub-formula of A. Without φ_δ, the function F becomes exactly what Solovay used to prove his completeness theorems for GL [Sol76], as we used in Example 1.1. This is not enough for our aim. We need to have $T_\alpha \vdash \mathsf{Prov}_{\mathsf{HA}}(\ulcorner B \vee C \urcorner)$ and generally, $T_\delta \vdash \mathsf{Prov}_{\mathsf{HA}}(\ulcorner \varphi_\delta \urcorner)$, which is not the case without the clause of φ_δ in the recursive definition of F.

By arguments in Section 5, there exists some classical structures $\mathfrak{M}_\delta \models T_\delta$ such that $\mathcal{K}_1$ is a first-order Kripke model of HA. This implies that $\mathfrak{M}_\beta \models B$, $\mathfrak{M}_\beta \not\models C$, $\mathfrak{M}_\gamma \models C$ and $\mathfrak{M}_\gamma \not\models B$. Since $T_\alpha \vdash \mathsf{Prov}_{\mathsf{HA}}(\ulcorner B \vee C \urcorner)$, we can deduce that $\mathfrak{M}_\alpha \models \mathsf{Prov}_{\mathsf{HA}}(\ulcorner B \vee C \urcorner)$. Also it is easy to show that for any first-order Kripke model $\mathcal{K} \Vdash \mathsf{HA}$, any node δ and arbitrary Σ_1-sentence E, we have

$$\mathcal{K}, \delta \Vdash E \qquad \Longleftrightarrow \qquad \mathfrak{M}_\delta \models E$$

Since $\mathsf{Prov}(x)$ is a Σ_1-predicate, we can deduce $\mathcal{K}_1, \alpha \Vdash \mathsf{Prov}_{\mathsf{HA}}(\ulcorner B \vee C \urcorner)$ and $\mathcal{K}_1, \alpha \nVdash \mathsf{Prov}_{\mathsf{HA}}(\ulcorner B \urcorner) \vee \mathsf{Prov}_{\mathsf{HA}}(\ulcorner C \urcorner)$. Hence $\mathcal{K}_1 \nVdash \mathsf{Prov}_{\mathsf{HA}}(\ulcorner B \vee C \urcorner) \to (\mathsf{Prov}_{\mathsf{HA}}(\ulcorner B \urcorner) \vee \mathsf{Prov}_{\mathsf{HA}}(\ulcorner C \urcorner))$, as desired. We will consider this proposition (A) again in Example 6.1 and refute it from Σ_1-provability logic of HA with the direct use of our main theorem in Section 5.

Example 1.3. In this example, we show that how to refute $A = \neg\neg\Box(\neg\neg p \to p) \to \Box(\neg\neg p \to p)$ from the provability logic of HA and also from the Σ_1-provability logic of HA. In this example, the TNNIL algorithm is involved.

The first thing is that we cannot directly refute A from the provability logic of HA as we did in Examples 1.1 and 1.2. The difficulty comes from the nested implications in the left hand side which are not separated by a $\Box$. Note that $\neg p$ is a shorthand for $p \to \bot$. To overcome this difficulty, we iteratively use Visser's NNIL (No Nested Implication to the Left) approximation [Vis02] in the modal language, i.e. inside any $\Box$ we compute the best NNIL approximation from below and replace it for the proposition. The approximated proposition for any modal proposition E is denoted in this paper by E^+. Some of Visser's NNIL approximations are [Vis02]:

$$(\neg\neg p)^+ = p \quad , \quad (\neg\neg p \to p)^+ = p \vee \neg p \quad , \quad ((p \to q) \to r)^+ = r \vee (p \wedge (q \to r))$$

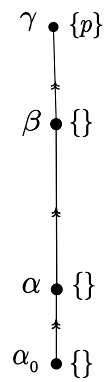
The process of computing the approximation $(.)^+$ is complicated and we do not precisely define it in this example. It is explained in details in Section 4.1. We may briefly describe it in the following way.

Let A be a non-modal proposition. Its NNIL approximation A^+, is some proposition with no nested implications to the left such that $\mathsf{IPC} \vdash A^+ \to A$, and for any other NNIL proposition B such that $\mathsf{IPC} \vdash B \to A$, we have $\mathsf{IPC} \vdash B \to A^+$. It is clear that, up to IPC-deductive equivalency, such an approximation is unique.

We have the following approximation for A:

$$A^+ = \Box(p \vee \neg p) \vee \neg\Box(p \vee \neg p)$$

Now we can handle this simplified proposition A^+ as we did in Examples 1.1 and 1.2. The following Kripke model $\mathcal{K}_0$ is a counter-model for A^+:



$\alpha_0, \alpha, \beta \nVdash p \qquad \gamma \Vdash p \qquad$ and

$\alpha \,\mathcal{R}\, \beta, \gamma \qquad \alpha \leq \beta, \gamma \qquad \beta \leq \gamma \qquad \beta \,\mathcal{R}\, \gamma$

$\alpha_0 \,\mathcal{R}\, \alpha, \beta, \gamma \qquad \alpha_0 \leq \alpha, \beta, \gamma$

One can define the recursive function F exactly the same as Example 1.2 with new definitions for $\varphi_{_\delta}$ and B:

$$\varphi_{_\alpha} := \top \quad , \quad \varphi_{_\beta} = \varphi_{_\gamma} := B \vee \neg B \quad , \quad B := (\exists x F(x) = \gamma)$$

Then we can define the first-order Kripke model $\mathcal{K}_1 \Vdash \mathsf{HA}$ which simulates $\mathcal{K}_0$ in a same way as Example 1.2. Then we can deduce that $\mathcal{K}_1 \nVdash \mathsf{Prov}_{_\mathsf{HA}}(\ulcorner B \vee \neg B \urcorner) \vee \neg \mathsf{Prov}_{_\mathsf{HA}}(\ulcorner B \vee \neg B \urcorner)$. Although we have refuted $\Box(p \vee \neg p) \to \neg\Box(p \vee \neg p)$ from the Σ_1-provability logic of HA, the proposition A is not refuted yet. But the key point here is that by Visser's Rule, we have $\mathsf{HA} \vdash \neg\neg\mathsf{Prov}_{_\mathsf{HA}}(\ulcorner \neg\neg B \to B \urcorner) \to \mathsf{Prov}_{_\mathsf{HA}}(\ulcorner \neg\neg B \to B \urcorner)$ if and only if $\mathsf{HA} \vdash \mathsf{Prov}_{_\mathsf{HA}}(\ulcorner \neg\neg B \to B \urcorner) \vee \neg \mathsf{Prov}_{_\mathsf{HA}}(\ulcorner \neg\neg B \to B \urcorner)$. And also by formalized Visser's Rule, we have:

$$\mathsf{HA} \vdash \mathsf{Prov}_{_\mathsf{HA}}(\ulcorner B \vee \neg B \urcorner) \leftrightarrow \mathsf{Prov}_{_\mathsf{HA}}(\ulcorner \neg\neg B \to B \urcorner)$$

This will finish the refutation process, i.e. $\mathsf{HA} \nvdash \neg\neg\mathsf{Prov}_{_\mathsf{HA}}(\ulcorner \neg\neg B \to B \urcorner) \to \mathsf{Prov}_{_\mathsf{HA}}(\ulcorner \neg\neg B \to B \urcorner)$. The Visser's Rule says that for any Σ_1-sentence B, we have $\mathsf{HA} \vdash \neg\neg B \to B$ iff $\mathsf{HA} \vdash B \vee \neg B$. The proof of this rule first appeared in [Vis81] (see Corollary 4.8 item 1).

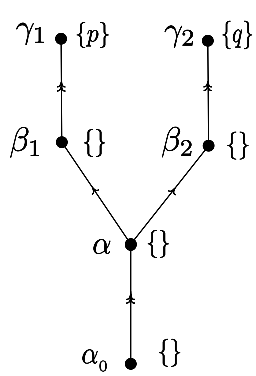
We will consider this proposition (A) again in Example 6.2 and refute it from Σ_1-provability logic of HA with the direct use of our main theorem in Section 5.

In all of the Examples 1.1 to 1.3, the relation for modal operator did not play an independent role, i.e. $\mathcal{R}$ and $\leq$ either where equal (Examples 1.2 and 1.3) or could be defined as equal relations (Example 1.1). This made too much simplifications in the definition of the recursive function F. In the following example, there exist some α and β such that $\alpha \leq \beta$ but it is not the case that $\alpha \,\mathcal{R}\, \beta$.

Example 1.4. In this example we refute the modal proposition

$$A := \Box(p \vee q) \to [(\Box p \to (p \vee q \vee \Box q)) \vee (\Box q \to (p \vee q \vee \Box p))]$$

Like Examples 1.1 and 1.2, we first find a Kripke counter-model $\mathcal{K}_0 \nVdash A$:



$\gamma_1 \Vdash p \qquad \gamma_2 \Vdash q$

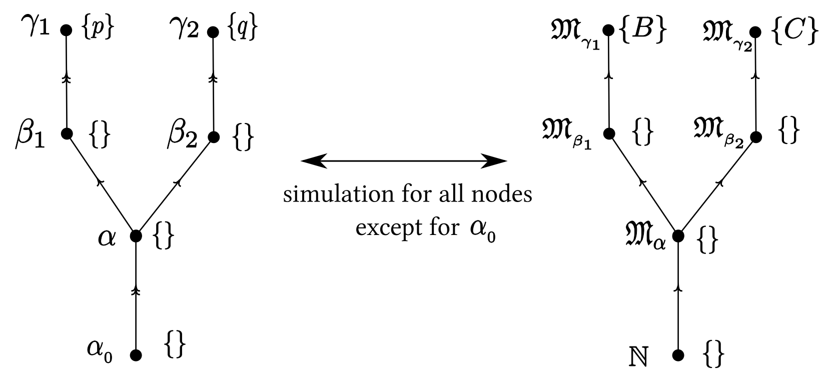
$\alpha_0, \alpha, \beta_1, \beta_2, \gamma_2 \nVdash p \qquad \alpha_0, \alpha, \beta_1, \beta_2, \gamma_1 \nVdash q$

$\alpha_0 \leq \alpha, \beta_1, \beta_2, \gamma_1, \gamma_2 \qquad \alpha_0 \,\mathcal{R}\, \alpha, \beta_1, \beta_2, \gamma_1, \gamma_2$

$\alpha \leq \beta_1, \beta_2, \gamma_1, \gamma_2 \qquad \alpha \,\mathcal{R}\, \gamma_1, \gamma_2 \qquad \beta_i \leq \gamma_i \qquad \beta_i \,\mathcal{R}\, \gamma_i$

$\alpha \not\mathcal{R} \beta_1 \qquad \alpha \not\mathcal{R} \beta_2$

We simulate this Kripke model with a first-order Kripke model $\mathcal{K}_1 \Vdash \mathsf{HA}$:



We define the Σ_1-sentences B and C and also the sentences φ_δ for any $\delta \neq \alpha_0$ like before:

$$B := (\exists x F(x) = \gamma_1) \qquad C := (\exists x F(x) = \gamma_2) \qquad \varphi_\alpha := B \vee C \qquad \varphi_{\beta_1} := B \qquad \varphi_{\beta_2} := C$$

$$\text{(1.2)} \qquad \varphi_{\gamma_1} := \varphi_{\gamma_2} := B \wedge C \qquad \mathfrak{M}_\delta \models T_\delta \qquad T_\delta := \mathsf{PA} + \lim_{x\to\infty} F(x) = \delta + \mathsf{Prov}_{\mathsf{HA}}(\ulcorner \varphi_\delta \urcorner)$$

The recursive definition of F is more complicated than previous examples. This is because we have really two different relations: $\leq$ and $\mathcal{R}$. The clause in recursive definition of F for the treatment of $\mathcal{R}$ is as before. For $\leq$ we use a variant of Berarducci's primitive recursive function in [Ber90] which he used for characterizing the interpretability logic of PA.
We define $F(0) := \alpha_0$. Assume that we have defined $F(n) := \delta$, and we will define $F(n+1) := \delta'$ if one of the following cases occurs, otherwise we define $F(n+1) := F(n) = \delta$.

- $\delta \,\mathcal{R}\, \delta'$ and there exists some witness (which is less than or equal to $n+1$) for the inconsistency of $T_{\delta'}$, or in other words, there exists some proof (in PA) with the Gödel number $\leq n+1$ for the statement
$$\neg[\lim_{x\to\infty} F(x) = \delta' \wedge \mathsf{Prov}_{\mathsf{HA}}(\ulcorner \varphi_{\delta'} \urcorner)]$$
- All of the following conditions hold:
  - $\delta \not\mathcal{R} \delta'$ and $\delta \leq \delta'$,
  - There exists some witness (which is less than or equal to $n+1$) for the inconsistency of $T_{\delta'}$,
  - The inconsistency rank of $T_{\delta'}$ (we call it $r(\delta', n+1)$) is less than the inconsistency rank of T_δ (we call it $r(\delta, n+1)$),
  - $F(r(\delta', n+1)) \,\mathcal{R}\, \delta$.

  The inconsistency rank of T_δ is defined to be the minimum k such that there exists a witness (less than or equal to $n+1$) for the inconsistency of
$$\mathsf{PA}_k + \lim_{x\to\infty} F(x) = \delta + \mathsf{Prov}_{\mathsf{HA}}(\ulcorner \varphi_\delta \urcorner)$$
  In above definition, PA_k is the theory $I\Sigma_1$ plus induction axiom for those formulas with Gödel number less than k.

The crucial fact about the function F is that F would not climb over tree (see Theorem 5.26). This fact is crucial for proving that the first-order Kripke model $\mathcal{K}_1 \Vdash \mathsf{HA}$ exists such that it fulfils the conditions in eq. (1.2). By Corollary 5.23, we have

$$T_\delta \vdash \mathsf{Prov}_{\mathsf{HA}}(\ulcorner \varphi_\delta \urcorner) \quad \text{for any } \delta \neq \alpha_0$$

To simplify notations, we use $\Box A$ instead of $\mathsf{Prov}_{\mathsf{HA}}(\ulcorner A \urcorner)$ for arithmetical formula A. Hence we have

$$\mathfrak{M}_\alpha \models \Box(B \vee C) \quad \mathfrak{M}_{\beta_1} \models \Box B \quad \mathfrak{M}_{\beta_2} \models \Box C \quad \mathfrak{M}_{\gamma_1} \models B \quad \mathfrak{M}_{\gamma_2} \models C$$

Moreover, we have

$$\mathfrak{M}_{\beta_1} \not\models B, C \qquad \mathfrak{M}_{\beta_2} \not\models B, C$$

We need two more conditions to deduce that

$$\mathcal{K}_1 \nVdash \Box(B \vee C) \to [(\Box B \to (B \vee C \vee \Box C)) \vee (\Box C \to (B \vee C \vee \Box B))]$$

These two conditions are $\mathfrak{M}_{\beta_1} \not\models \Box C$ and $\mathfrak{M}_{\beta_2} \not\models \Box B$. We will show in Theorem 5.15 that these conditions hold as well. The proof of this fact take up all Section 5.3 and there, we use Lemma 3.18 which is the essential result in Section 3.

Why $\leq$ is not treated like $\mathcal{R}$ in recursive definition of F? Because if we do so, we are not able to prove that the function F is constant (Theorem 5.26) and even the consistency of $L = \alpha_0$ and consequently the consistency of all the theories T_δ will be lost.

1.2 What happens in classical case

The main result of this paper in classical case, i.e. the Σ_1-provability logic of PA is already characterized by A. Visser [Vis81] and is remarkably simpler than the intuitionistic case. A. Visser showed:

$$\mathsf{GLV} \vdash A \qquad \Longleftrightarrow \qquad \forall * \ \mathsf{PA} \vdash A^*,$$

in which, $*$ ranges over all of the interpretations that p^* is a Σ_1-sentence for atomic variables p and GLV is GL plus the completeness axiom for atomic variables: $p \to \Box p$. For a proof of this fact see [Boo95] page 135. It is shown [AM15] that the provability logic of PA could be reduced to its Σ_1-provability logic.

1.3 Map of sections

Let us explain the content of sections and their interrelationship. All of the contents of this paper are minimally chosen for one major goal: *soundness and completeness of* iH_σ *for arithmetical* Σ_1-*interpretations*, i.e. Theorems 6.3 and 6.5. In Section 2, we give definitions of some elementary notions and also make some conventions. In Section 3, we gather all the required statements with arithmetical nature. Most of the lemmas and definitions are for proving a refinement of Leivant's principle in Lemma 3.18 (or its simplified form in Theorem 3.14). This will be used in Section 5. In Section 4, we collect all required notions with propositional nature. The most crucial fact we will show in this section is that in iH_σ (precise axiomatization of iH_σ will come in Section 4.3), one could transform any modal proposition A to another proposition A^+ with simpler form, which is called TNNIL in this paper. Roughly speaking, in a TNNIL-formula, every two nested implications in the left hand side are separated by a $\Box$. This is done in Theorem 4.18 and Corollary 4.19. Then we show that the theory LC (intuitionistic version of GL plus the axiom schema $A \to \Box A$) is TNNIL-conservative over iH_σ (Theorem 4.24). It turns out that LC and iH_σ actually prove same TNNIL modal propositions (Corollary 6.4). Moreover, it is shown in Theorem 4.26 that LC is sound and complete for a special class of finite Kripke models (perfect Kripke models). In Section 5, we show that one could transform a finite Kripke model of LC (with tree-frame) to a first-order Kripke model of HA (Theorem 5.1). This transformation is such that there is a natural correspondence between these two Kripke models. Finally in Section 6, we use the results of Sections 3, 4 and 5, to prove the soundness and completeness of iH_σ for arithmetical Σ_1-interpretations.

2 Definitions and conventions

The propositional non-modal language $\mathcal{L}_0$ contains atomic variables, $\vee, \wedge, \to, \bot$ and the propositional modal language, $\mathcal{L}_\Box$ has an additional operator $\Box$. In this paper, the atomic propositions (in modal or non-modal language) includes atomic variables and $\bot$. For an arbitrary proposition A, $\mathsf{Sub}(A)$ is defined to be the set of all sub-formulae of A, including A itself. We take $\mathsf{Sub}(X) := \bigcup_{A\in X} \mathsf{Sub}(A)$ for a set of propositions X. We use $\boxdot A$ as a shorthand for $A \wedge \Box A$. The logic IPC is intuitionistic propositional non-modal logic over usual propositional non-modal language. The theory $\mathsf{IPC}_\Box$ is the same theory IPC in the extended language of propositional modal language, i.e. its language is propositional modal language and its axioms and rules are same as IPC . Because we have no axioms for $\Box$ in $\mathsf{IPC}_\Box$, it is obvious that $\Box A$ for each A, behaves exactly like an atomic variable inside $\mathsf{IPC}_\Box$. First-order intuitionistic logic is denoted IQC and the logic CQC is its classical closure, i.e. IQC plus the principle of excluded middle. For a set of sentences and rules $\Gamma \cup \{A\}$ in the propositional non-modal, propositional modal or first-order language, $\Gamma \vdash A$ means that A is derivable from Γ in the system $\mathsf{IPC}, \mathsf{IPC}_\Box, \mathsf{IQC}$, respectively. For an arithmetical formula, $\ulcorner A \urcorner$ represents the Gödel number of A. For an arbitrary arithmetical theory T with a set of Δ_0- axioms, we have the Δ_0-predicate $\mathsf{Proof}_T(x, \ulcorner A \urcorner)$, that is a formalization of "x is the code of a proof for A in T". We also have the provability predicate $\mathsf{Prov}_T(\ulcorner A \urcorner) := \exists x\ \mathsf{Proof}_T(x, \ulcorner A \urcorner)$. The set of natural numbers is denoted by $\omega := \{0, 1, 2, \ldots\}$.

Definition 2.1. *Suppose T is a recursively enumerable (r.e.) arithmetical theory and σ is a substitution i.e. a function from atomic variables to arithmetical sentences. We define the interpretation σ_T which extend the substitution σ to all modal propositions A, inductively:*

- *$\sigma_T(A) := \sigma(A)$ for atomic A,*
- *σ_T distributes over $\wedge, \vee, \to$,*
- *$\sigma_T(\Box A) := \mathsf{Prov}_T(\ulcorner \sigma_T(A) \urcorner)$.*

We call σ a Σ_1-substitution, if for every atomic A, $\sigma(A)$ is a Σ_1-sentence. We also say that σ_T is a Σ_1-interpretation if σ is a Σ_1-substitution.

Definition 2.2. *The provability logic of a sufficiently strong theory T, is defined to be a modal propositional theory $\mathcal{PL}(T)$ such that $\mathcal{PL}(T) \vdash A$ iff for all arithmetical substitutions σ, $T \vdash \sigma_T(A)$. If we restrict the substitutions to Σ_1-substitutions, then the new modal theory is $\mathcal{PL}_\sigma(T)$.*

Lemma 2.3. *Let $A(p_1, \ldots, p_n)$ be a non-modal proposition with $p_i \neq p_j$ for all $0 < i < j \leq n$. Then for all modal sentences $B_1, \ldots, B_n$ we have:*

$$\mathsf{IPC} \vdash A \quad \textit{iff} \quad \mathsf{IPC}_\Box \vdash A[p_1|\Box B_1, \ldots, p_n|\Box B_n]$$

Proof. By simple inductions on complexity of proofs in IPC and $\mathsf{IPC}_\Box$. ❑

We define NOI (No Outside Implication) as the set of modal propositions A, such that any occurrence of $\to$ is in the scope of some $\Box$. To be able to state an extension of Leivant's Principle (that is adequate to axiomatize Σ_1-provability logic of HA) we need a translation on the modal language which we call *Leivant's translation.* We define it recursively as follows:

- $A^l := A$ for atomic or boxed A,
- $(A \wedge B)^l := A^l \wedge B^l$.
- $(A \vee B)^l := \boxdot A^l \vee \boxdot B^l$.
- $(A \to B)^l$ is defined by cases: If $A \in \mathsf{NOI}$, we define $(A \to B)^l := A \to B^l$, otherwise we define $(A \to B)^l := A \to B$.

Definition 2.4. *Minimal provability logic* iGL*, is the same as Gödel-Löb provability logic* GL *, with all tautologies of intuitionistic logic (in modal language) instead of tautologies of classical logic.* iK4 *is* iGL *without Löb's axiom. Note that we can get rid of the necessitation rule by adding* $\Box A$ *to the axioms, for each axiom* A *in the above list. We will use this fact later in this paper. We list the following axiom schemas:*

- *The Completeness Principle:* $\mathsf{CP} := A \to \Box A$.
- *Restriction of Completeness Principle to atomic formulae:* $\mathsf{CP_a} := p \to \Box p$, *for atomic* p.
- *Leivant's Principle:* $\mathsf{Le} := \Box(B \vee C) \to \Box(\Box B \vee C)$. *[Lei75]*
- *Extended Leivant's Principle:* $\mathsf{Le}^+ := \Box A \to \Box A^l$.

We define theories $\mathsf{LC} := \mathsf{iGL} + \mathsf{CP}$ *and* $\mathsf{LLe}^+ := \mathsf{iGL} + \mathsf{Le}^+ + \mathsf{CP_a}$. *Note that in the presence of* CP *and modus ponens, the necessitation rule is superfluous.*

3 Arithmetic

In this section, we gather some preliminaries from intuitionistic arithmetic. Mostly we will prove some refinements of well-known theorems such as: Π_2-conservativity of PA over HA, Gödel's diagnolization lemma and Σ_1-completeness of HA. Most of these preliminaries will be used to prove a refinement of Leivant's principle $\Box(A \vee B) \to \Box(\boxdot A \vee \boxdot B)$ in the technical Lemma 3.17. Theorem 3.14 states a simplified version of Lemma 3.17.

3.1 Some arithmetical preliminaries

The first-order language of arithmetic contains three functions (successor, addition and multiplication), one predicate symbol and a constant: $(S, +, ., \leq, 0)$. First-order intuitionistic arithmetic (HA) is the theory over IQC with the axioms:

Q1 $S(x) \neq 0$,

Q2 $S(x) = S(y) \to x = y$,

Q3 $y = 0 \vee \exists x\, S(x) = y$,

Q4 $x + 0 = x$,

Q5 $x + S(y) = S(x + y)$,

Q6 $x.0 = 0$,

Q7 $x.S(y) = (x.y) + x$,

Q8 $x \leq y \leftrightarrow \exists z(z + x = y)$,

Ind: For each formula $A(x)$:

$$\mathsf{Ind}(A, x) := \mathcal{UC}[A(0) \wedge \forall x(A(x) \to A(S(x)))] \to \forall x A(x)]$$

In which $\mathcal{UC}(B)$ is the universal closure of B.

Peano Arithmetic PA, has the same axioms of HA over CQC. We also define $x < y$ as $x \leq y \wedge x \neq y$. Let T be an r.e. theory with the set of axioms $A_1, A_2, \ldots$. It is known in the literature (see e.g. [Ber90, section 2.3] or [Vis02, section 8.1]) that T_n indicates the theory with the first n axioms of T, i.e. $A_1, \ldots A_n$. In the following notation, we order the axioms of HA and PA in a way which best fit the relevant lemmas and theorems in this paper.

Notation 3.1. *From now on, when we are working in first-order language of arithmetic, for a first-order sentence A, $\Box A$ and $\Box^+ A$ are shorthand for $\mathsf{Prov}_{\mathsf{HA}}(\ulcorner A \urcorner)$ and $\mathsf{Prov}_{\mathsf{PA}}(\ulcorner A \urcorner)$, respectively. Let $i\Sigma_1$ be the theory HA, where the induction principle is restricted to Σ_1-formulae. We also define the theories HA_x to be the theory with axioms of HA, in which the induction principle is restricted to formulas satisfying at least one of the following conditions:*

- *formulas of the form $(A \to B) \to B$ in which A and B are Σ_1.*
- *formulas with Gödel number less than x.*

We can define similar concept for PA_x. Note that classically, a formula of the form $(A \to B) \to B$ in which A and B are Σ_1, is equivalent to the Σ_1-formula $A \vee B$ and hence PA_0 is the well-known theory $I\Sigma_1$. We also define $\Box_x A$ and $\Box_x^+ A$ to be $\mathsf{Prov}_{\mathsf{HA}_x}(\ulcorner A \urcorner)$ and $\mathsf{Prov}_{\mathsf{PA}_x}(\ulcorner A \urcorner)$, respectively.

We recall that a function f on $\omega := \{0, 1, 2, \ldots\}$ is recursive iff there exists some Σ_1-formula $A_f(\bar{x}, y)$ such that $\mathbb{N} \models A_f(\bar{x}, y)$ iff $f(\bar{x}) = y$. It is called provably total in T, iff $T \vdash \forall \bar{x} \exists y A_f(\bar{x}, y)$.

It is well known that all primitive recursive functions are provably total in $I\Sigma_1$ with a Δ_0-formula as defining formula. So we may use primitive recursive function symbols in the language of arithmetic with their defining axioms (as far as we work in $I\Sigma_1$).

Lemma 3.2. *Let A, B be Σ_1-formulae such that $\mathsf{PA} \vdash A \to B$. Then $\mathsf{HA} \vdash A \to B$.*

Proof. Let $\mathsf{PA} \vdash A \to B$. Then as it is well known in classical logic, we have $\mathsf{PA} \vdash \neg A \vee B$. Since A and B are Σ_1, there are some Δ_0 formulas $A'(x)$ and $B'(y)$ such that $A = \exists x A'(x)$ and $B = \exists y B'(y)$. We may assume that x is not free in B' and y is not free in A'. Hence we may deduce that $\mathsf{PA} \vdash \exists y(\neg A'(x) \vee B'(y))$. By Π_2-conservativity of PA over HA [TvD88](3.3.4), we can deduce that $\mathsf{HA} \vdash \exists y(\neg A'(x) \vee B'(y))$. Then we may deduce that $\mathsf{HA} \vdash \exists y(A'(x) \to B'(y))$ and hence (since y is not free in A') $\mathsf{HA} \vdash A'(x) \to B$ and by generalization rule $\mathsf{HA} \vdash \forall x(A'(x) \to B)$. This implies that $\mathsf{HA} \vdash A \to B$ (since x is not free B). ❑

Lemma 3.3. *For any Δ_0-formula $A(\bar{x})$, we have $\mathsf{HA}_0 \vdash \forall \bar{x}(A(\bar{x}) \vee \neg A(\bar{x}))$.*

Proof. This is well-known in the literature. ❑

The Gödel-Gentzen translation associates a formula A^g for any formula A in a first-order language, and is defined inductively by the following items:

- $A^g := A$, for atomic A,
- $(A \wedge B)^g := A^g \wedge B^g$,
- $(A \vee B)^g := \neg(\neg A^g \wedge \neg B^g)$,
- $(A \to B)^g := A^g \to B^g$,
- $(\forall x A)^g := \forall x A^g$,
- $(\exists x A)^g := \neg\neg \exists x\, A^g$.

The Friedman translation associates a formula A^C, for an arbitrary formula C and A in first-order language. Roughly speaking, A^C is the result of adding C as a disjunct to all atomic sub-formulas of A. To define A^C, we assume that free variables of C do not appear as bound variables of A. It is obvious that we can always take care of this detail by renaming bound variables of A to fresh variables.

- $A^C := A \vee C$, for atomic A,
- $(A \wedge B)^C := A^C \wedge B^C$,

- $(A \vee B)^C := A^C \vee B^C$,
- $(A \to B)^C := A^C \to B^C$,
- $(\forall x A)^C := \forall x A^C$,
- $(\exists x A)^C := \exists x A^C$.

As shown in [TvD88], we have the following properties for Gödel-Gentzen and Friedman translations:

- For each Σ_1-formula A in the language of arithmetic, $\mathsf{HA} \vdash A^g \leftrightarrow \neg\neg A$ and $\mathsf{HA} \vdash A^C \leftrightarrow (A \vee C)$.
- For any A in the language of arithmetic, $\mathsf{CQC} \vdash A$ implies $\mathsf{IQC} \vdash A^g$.
- HA_0 is closed under Friedman's translation with respect to Σ_1-formulas. i.e. for any Σ_1-formula B and any A, $\mathsf{HA}_0 \vdash A$ implies $\mathsf{HA}_0 \vdash A^B$. Actually in [TvD88], this property is proved for HA instead of HA_0, but this case is very similar to that one.

We have the following variant of Lemma 3.2.

Lemma 3.4. *For any Σ_1-formula A, $\mathsf{PA}_0 \vdash A$ implies $\mathsf{HA}_0 \vdash A$. Hence for any Π_2-sentence A, $\mathsf{PA}_0 \vdash A$ implies $\mathsf{HA}_0 \vdash A$.*

Proof. First observe that $\mathsf{PA}_0 \vdash B$ implies $\mathsf{HA}_0 \vdash B^g$, by induction on proof of B in PA_0. We refer the reader to [TvD88] for a detailed proof of this fact for PA and HA instead of PA_0 and HA_0. It should only be noted that for any instance B of induction over Σ_1 formulae in PA_0, by definition of Gödel-Gentzen translation, B^g belongs to the axioms of HA_0. Hence, we have $\mathsf{HA}_0 \vdash \neg\neg A$, and thus $\mathsf{HA}_0 \vdash (\neg\neg A)^A$. This implies $\mathsf{HA}_0 \vdash A$, as desired. ❑

Consider the mapping:

$$F : n \mapsto A(S^n(0)) := A(\overbrace{S \ldots S}^{n \text{ times}}(0))$$

Let G be the primitive recursive function that assigns to n the Gödel number of $F(n)$. Instead of $G(x)$, we use the notation $\ulcorner A(\dot{x}) \urcorner$ which is common in the literature. We may omit the dot over variables when no confusion is likely.

Lemma 3.5. *For every formula $A(x, x_1 \ldots, x_n)$ with free variables exactly as shown, there exists a formula $B(x_1, \ldots, x_n)$ such that*

$$\mathsf{HA}_0 \vdash B(x_1, \ldots, x_n) \leftrightarrow A(\ulcorner B(\dot{x}_1, \ldots, \dot{x}_n) \urcorner, x_1, \ldots, x_n)$$

Moreover, if the formula A is Δ_0, then B is also Δ_0.

Proof. It is easy to see that the usual proof of the fixed point lemma holds in this setting. ❑

The following lemma states the Σ_1-completeness of HA_0.

Lemma 3.6. *HA_0 proves all true Σ_1 sentences. Moreover this argument is formalizable and provable in HA_0, i.e. for every Σ_1-formula $A(x_1, \ldots, x_k)$ we have $\mathsf{HA}_0 \vdash A(x_1, \ldots, x_k) \to \Box_0 A(\dot{x}_1, \ldots, \dot{x}_k)$.*

Proof. It is a well-known fact that any true (in the standard model $\mathbb{N}$) Σ_1-sentence is provable in $i\Sigma_1$. Moreover this argument is constructive and formalizable in $i\Sigma_1$. ❑

Lemma 3.7. *For every formula A, we have $\mathsf{PA} \vdash \forall x\ \Box^+(\Box^+_x A \to A)$ and $\mathsf{HA} \vdash \forall x\ \Box(\Box_x A \to A)$.*

Proof. The case of PA is well known. For the case HA, see [Smo73b] or [Vis02, Theorem 8.1]. ❑

3.1.1 Coding of finite sequences

We use some fixed method for encoding of finite sequences and use $\langle x_1, \ldots, x_n\rangle$ as the code of the finite sequence $(x_1, \ldots, x_n)$. We assume here that the encoding is a one-one correspondence between natural numbers and the associated finite sequences. For details on coding of finite sequences, we refer the reader to [Smo85], Chapter 0.

Let $x = \langle x_0, x_1, \ldots, x_n\rangle$ and $y = \langle y_0, y_1, \ldots, y_m\rangle$. The following notations are used in this paper:

- $\mathsf{lth}(x)$ is defined as the length of the sequence with the code x, i.e. here $\mathsf{lth}(x) := n + 1$,
- $x * y := \langle x_0, \ldots, x_n, y_0, \ldots, y_m\rangle$,
- $(x)_i$ is defined (if $i < \mathsf{lth}(x)$) as the i-th element in the sequence with the code x, i.e. here $(x)_i := x_i$. If also $i \geq \mathsf{lth}(x)$, we define $(x)_i := 0$,
- $\hat{x}$ is defined as the final element of the sequence with the code x, i.e. here $\hat{x} := (x)_{\mathsf{lth}(x) \dot{-} 1}$,
- x is an initial segment of y ($x \subseteq_{\mathsf{i}} y$) if $\mathsf{lth}(x) \leq \mathsf{lth}(y)$ and for all $j < \mathsf{lth}(x)$, we have $(x)_j = (y)_j$.

3.1.2 Kripke models of HA

A first-order Kripke model for HA is a triple $\mathcal{K} = (K, <, \mathfrak{M})$ such that:

- The frame of $\mathcal{K}$, i.e. $(K, <)$, is a non-empty partially ordered set,
- $\mathfrak{M}$ is a function from K to the first-order classical structures for the language of the arithmetic, i.e. $\mathfrak{M}(\alpha)$ is a first-order classical structure, for each $\alpha \in K$,
- For any $\alpha \leq \beta \in K$, $\mathfrak{M}(\alpha)$ is a weak substructure of $\mathfrak{M}(\beta)$.

For any $\alpha \in K$ and first-order formula $A \in \mathcal{L}_\alpha$ (the language of arithmetic augmented with constant symbols $\bar{a}$ for each $a \in |\mathfrak{M}(\alpha)|$), we define $\mathcal{K}, \alpha \Vdash A$ (or simply $\alpha \Vdash A$, if no confusion is likely) inductively as follows:

- For atomic A, $\alpha \Vdash A$ iff $\mathfrak{M}(\alpha) \models A$. Note that in the structure $\mathfrak{M}(\alpha)$, $\bar{a}$ is interpreted as a,
- $\mathcal{K}, \alpha \Vdash A \vee B$ iff $\mathcal{K}, \alpha \Vdash A$ or $\mathcal{K}, \alpha \Vdash B$,
- $\mathcal{K}, \alpha \Vdash A \wedge B$ iff $\mathcal{K}, \alpha \Vdash A$ and $\mathcal{K}, \alpha \Vdash B$,
- $\mathcal{K}, \alpha \Vdash A \to B$ iff for all $\beta \geq \alpha$, $\mathcal{K}, \beta \Vdash A$ implies $\mathcal{K}, \beta \Vdash B$,
- If $A = \forall x B$, $\alpha \Vdash A$ iff for all $\beta \geq \alpha$ and each $b \in |\mathfrak{M}(\beta)|$, we have $\beta \Vdash B[x : \bar{b}]$.

It is well-known in the literature that HA is complete for first-order Kripke models.

Lemma 3.8. *Let $\mathcal{K} = (K, <, \mathfrak{M})$ be a Kripke model of* HA *and A be an arbitrary Σ_1-formula. Then for each $\alpha \in K$, we have $\alpha \Vdash A$ iff $\mathfrak{M}(\alpha) \models A$.*

Proof. Use induction on the complexity of A to show that for each $\alpha \in K$, we have $\alpha \Vdash A$ iff $\mathfrak{M}(\alpha) \models A$. In the inductive step for $\to$ and $\forall$, use Lemma 3.3. ❑

3.2 q-Realizability and Leivant's principle

A variant of realizability introduced by Kleene, is q-realizability (see [TvD88]) which is defined inductively for arithmetical formula A as follows:

- $x \ \mathsf{q} \ A := A$ for atomic A.
- $x \ \mathsf{q} \ (A_1 \wedge A_2) := \mathsf{j}_1(x) \ \mathsf{q} \ A_1 \wedge \mathsf{j}_2(x) \ \mathsf{q} \ A_2$,
- $x \ \mathsf{q} \ (A_1 \vee A_2) := (\mathsf{j}_1(x) = 0 \rightarrow \mathsf{j}_2(x) \ \mathsf{q} \ A_1) \wedge (\mathsf{j}_1(x) \neq 0 \rightarrow \mathsf{j}_2(x) \ \mathsf{q} \ A_2)$,
- $x \ \mathsf{q} \ (A_1 \rightarrow A_2) := \forall y \, (y \ \mathsf{q} \ A_1 \rightarrow \exists u \, (\mathsf{T}xyu \wedge \mathsf{U}(u) \ \mathsf{q} \ A_2)) \wedge (A_1 \rightarrow A_2)$,
- $x \ \mathsf{q} \ \exists y A(y) := j_1(x) \ \mathsf{q} \ A(j_2(x))$,
- $x \ \mathsf{q} \ \forall y A(y) := \forall y \, \exists u \, (\mathsf{T}xyu \wedge \mathsf{U}(u) \ \mathsf{q} \ A(y))$

In the y above definition j_1 and j_2 are inverses for a one-to-one onto, pairing function, j, such that $x = \mathsf{j}(\mathsf{j}_1(x), \mathsf{j}_2(x))$. Also $\mathsf{T}xyu$ is Kleene's predicate formalizing "u is a computation for the Turing Machine with code x with input y", and U is the result extractor function, i.e. if u is a computation for a Turing Machine, then $\mathsf{U}(u)$ is its output.

Lemma 3.9. *For any formula A we have $\mathsf{HA}_0 \vdash x \, \mathsf{q} \, A \rightarrow A$.*

Proof. See [TvD88]. ❑

In the following, $\{x\}$ is partial recursive function of the Turing Machine with the code x. The notation $\{x\}y\downarrow$ means that "the function $\{x\}$ is defined on input y", or equivalently "the Turing machine with the code x halts on the input y". It is well known that $\{x\}y\downarrow$ is a Σ_1 sentence. We use terms which contain some Kleene's bracket notation. In that case, we use $t\downarrow$ to mean that all the brackets in t are defined (terminate).

One immediate consequence of q-realizability, is Church's Rule for HA:

Lemma 3.10. *For every formula $A(x, y)$, if $\mathsf{HA} \vdash \forall x \, \exists y \, A(x, y)$, then there exists some $n \in \omega$ such that $\mathsf{HA} \vdash \forall x \, (\{n\}(x)\downarrow \wedge A(x, \{n\}(x)))$.*

Proof. See [TvD88]. ❑

It is easy to observe that "$\mathsf{HA} \vdash A$" implies "there exists some n such that $\mathsf{HA} \vdash n \ \mathsf{q} \ A$"([TvD88]). The point of the following lemma is that we can refine the above statement in the following way. There exists some recursive function f such that "$\mathsf{HA}_m \vdash A(k_1, \ldots, k_l)$" implies "there exists some recursive function g such that $\mathsf{HA}_{f(m)} \vdash g(k_1, \ldots, k_l) \ \mathsf{q} \ A(k_1, \ldots, k_l)$". Moreover, we can formalize this statement in HA.

Lemma 3.11. *Suppose that $A(x_1, \ldots, x_m)$ is an arithmetical formula with free variables as shown. Then, there exists a provably (in HA) total recursive function f such that:*

$$\mathsf{HA} \vdash \Box_x A(\dot{x}_1, \ldots, \dot{x}_m) \rightarrow \exists z \ \Box_{f(x)}(\{\dot{z}\}\langle \dot{x}_1, \ldots, \dot{x}_m \rangle\downarrow \wedge \{\dot{z}\}\langle \dot{x}_1, \ldots, \dot{x}_m \rangle \ \mathsf{q} \ A(\dot{x}_1, \ldots, \dot{x}_m))$$

Proof. The proof is very similar to the proof of the soundness part of [TvD88, Theorem 4.10]. First define $f(n)$ in this way:

$$f(n) := \max(\{\ulcorner B^{\mathsf{q},x} \urcorner \mid \ulcorner B \urcorner < n, \ x \text{ is a free variable of } B\} \cup \{n\})$$

in which, $B^{\mathsf{q},x} := \{t(u)\}\langle x \rangle\downarrow \wedge \{t(u)\}\langle x \rangle \ \mathsf{q} \ B$, $u \neq x$ and $t(u)$ is a primitive recursive function that will be defined later in the proof. Let's fix some sequence of numbers $\boldsymbol{m}$. With induction on the complexity of the proof $\mathsf{HA}_n \vdash A(\boldsymbol{m})$, we show that (by $A(\boldsymbol{m})$, we mean $A[\boldsymbol{x} : \boldsymbol{m}]$) by

$$\mathsf{HA} \vdash \text{“}\mathsf{HA}_n \vdash A(\boldsymbol{m})\text{”} \rightarrow \exists z \ \text{“}\mathsf{HA}_{f(n)} \vdash \{z\}\langle \boldsymbol{m} \rangle\downarrow \wedge \{z\}\langle \boldsymbol{m} \rangle \ \mathsf{q} \ A(\boldsymbol{m})\text{”}$$

We only treat the case where A is an instance of induction schema. All the other cases are trivial. Assume that $\ulcorner B \urcorner < n$ and

$$A(\boldsymbol{m}) = (B[x:0] \wedge \forall x(B \to B[x:S(x)])) \to \forall x B$$

We should find some number $\{z\}\langle \boldsymbol{m} \rangle = k$ such that

$$\mathsf{HA}_{f(n)} \vdash k \ \mathsf{q} \ [(B(0) \wedge \forall x(B(x) \to B(x+1)) \to \forall x B]$$

By definition of q-realizability, we have:

$$k \ \mathsf{q} \ A(\boldsymbol{m}) = \overbrace{\forall u[u \ \mathsf{q} \ (B(0) \wedge \forall x(B(x) \to B(x+1)) \to (\{k\}(u)\downarrow \wedge \{k\}(u) \ \mathsf{q} \ \forall x B)]}^{C} \wedge A(\boldsymbol{m})$$

Since $f(n) \geq n$, we have $\mathsf{HA}_{f(n)} \vdash A(\boldsymbol{m})$. Hence it remains only to show that $\mathsf{HA}_{f(n)} \vdash C$. Define the primitive recursive function $t(u)$ in the following way. For any given u, $t(u)$ is the code of the Turing Machine that fulfills the following conditions:

$$\begin{cases} \{t(u)\}\langle 0 \rangle = j_1(u) \\ \{t(u)\}(x+1) = \{\{j_2(u)\}\langle x \rangle\}\langle \{t(u)\}\langle x \rangle \rangle \end{cases}$$

Finally, let k be the code of the Turing Machine that computes the primitive recursive function t. Now it is not difficult to observe that, by induction on $B^{\mathsf{q},x}$, one could deduce C in HA_0, and hence $\mathsf{HA}_{f(n)} \vdash C$. This implies $\mathsf{HA}_{f(n)} \vdash A(\boldsymbol{m})$, as desired. ❑

Lemma 3.12. *For every sentence A, there exists some provably (in HA) total recursive function h_A such that $\mathsf{HA} \vdash \forall x \, \Box_{h_A(x)}(\Box_{\dot{x}} A \to A)$.*

Proof. By Lemma 3.7 we have $\mathsf{HA} \vdash \forall x \, \exists y \, \Box_y(\Box_{\dot{x}} A \to A)$. Now we have the desired result by use of Lemma 3.10. ❑

Lemma 3.13. *Suppose that $A(x_1, \ldots, x_m)$ is a Σ_1-formula with variables as shown. Then there exists some $n_A \in \mathbb{N}$, such that*

$$\mathsf{HA} \vdash A(x_1, \ldots, x_m) \to (\{n_A\}\langle x_1, \ldots, x_m \rangle\downarrow \wedge \{n_A\}\langle x_1, \ldots, x_m \rangle \ \mathsf{q} \ A(x_1, \ldots, x_m))$$

Proof. This theorem for **r**-realizability instead of q-realizability is proved in [TvD88](Proposition 4.4.5). The proof for **q**-realizability is quite similar and we leave it to the reader. ❑

It is well-known that the disjunction property holds for IPC and HA, however it is also shown that in case of HA , the proof is not formalizable in HA, i.e. $\mathsf{HA} \nvdash \Box(A \vee B) \to (\Box A \vee \Box B)$. But this is not the end of story! Daniel Leivant in his PhD dissertation [Lei75] showed that $\mathsf{HA} \vdash \Box(A \vee B) \to \Box(A \vee \Box B)$. Albert Visser in an unpublished paper showed that we can extend Leivant's principle to the following version. For every Σ_1-sentence A, $\mathsf{HA} \vdash \Box(A \to (B \vee C)) \to \Box(A \to (\Box B \vee C))$. In the following lemma, we will show that we can find (constructively) from the code x of the proof of $A \to (B \vee C)$, some $f(x)$ such that $\Box(A \to (\Box_{f(x)} B \vee C))$ holds. Although the statement of this theorem would not be used later in this paper, we bring it here for better understanding of its generalization in a more technical lemma, i.e. Lemma 3.17.

Theorem 3.14. *For arbitrary sentences A, B, C such that $A \in \Sigma_1$, there exists a provably (in HA) total recursive function f such that*

$$\mathsf{HA} \vdash \Box_x(A \to (B \vee C)) \to \Box_{f(x)}(A \to (\Box_{f(x)} B \vee C))$$

Proof. First observe that, by Lemma 3.13, there exists some finite number $n_A \in \mathbb{N}$ such that $\mathsf{HA} \vdash A \to (\{n_A\}\langle\rangle\downarrow \wedge \{n_A\}\langle\rangle \ \mathsf{q}\ A)$. We set $t_0 := \{n_A\}\langle\rangle$. Hence there exists some $n_0 \in \mathbb{N}$ such that

$$\mathsf{HA} \vdash \Box_{n_0}(A \to (t_0\downarrow \wedge t_0 \ \mathsf{q}\ A)) \tag{3.1}$$

We work inside HA. Assume $\Box_x(A \to (B \vee C))$. By Lemma 3.11, there exists some z such that $\Box_{g_0(x)}(\{\dot{z}\}\langle\rangle\downarrow \wedge \{\dot{z}\}\langle\rangle \ \mathsf{q}\ (A \to (B \vee C)))$, in which g_0 is the recursive function provided by Lemma 3.11. We define $t_1 := \{\dot{z}\}\langle\rangle$ and hence we have $\Box_{g_0(x)} t_1\downarrow$. If we set $g_1(y) := g_0(y) + n_0$, by use of eq. (3.1), we can deduce $\Box_{g_1(x)}(A \to (t_0\downarrow \wedge \{t_1\}(t_0) \ \mathsf{q}\ (B \vee C)))$. We set $t_2 := \{t_1\}(t_0)$. Then, by definition of q-realizability, we have:

$$\Box_{g_1(x)}(A \to (t_2\downarrow \wedge (j_1(t_2) = 0 \to j_2(t_2) \ \mathsf{q}\ B) \wedge (j_1(t_2) \neq 0 \to j_2(t_2) \ \mathsf{q}\ C))).$$

Let $B' := (j_1(t_2) = 0) \to j_2(t_2) \ \mathsf{q}\ B$ and $C' := (j_1(t_2) \neq 0) \to j_2(t_2) \ \mathsf{q}\ C$. Then we have $\Box_{g_1(x)}(A \to B')$ and, hence, by Σ_1-completeness (Lemma 3.6), we can deduce $\Box_0\Box_{g_1(x)}(A \to B')$, that again by use of Lemma 3.6, implies $\Box_0(A \to \Box_{g_1(x)}B')$. Thus we have

$$\Box_{g_1(x)}(A \to (t_2\downarrow \wedge \Box_{g_1(x)}B' \wedge C'))$$

Again by Lemma 3.6 and Lemma 3.9, $\Box_{g_1(x)}(A \to (t_2\downarrow \wedge (j_1(t_2) = 0 \to \Box_{g_1(x)}B) \wedge (j_1(t_2) \neq 0 \to C)))$. Since atomic formulae are decidable in HA, so for any atomic formulae D, there exists some finite n_2 such that in HA_{n_2} we have decidability of D. Let $\mathsf{HA}_{n_2} + t_2\downarrow$ decide $j_1(t_2) = 0$. If we set $f(x) := g_1(x) + n_2$, we can deduce $\Box_{f(x)}(A \to (\Box_{f(x)}B \vee C))$, as desired. ❑

3.3 The extended Leivant's Principle

In this section, we study properties of the extended Leivant's principle, Le^+. We prove that for any Σ_1- substitution σ, $\mathsf{HA} \vdash \sigma_{\mathsf{HA}}(\mathsf{Le}^+)$.

Define a translation $q_\sigma(A, x)$ recursively for a modal proposition A and a Σ_1-substitution σ, as follows:

- $q_\sigma(A, x) := \sigma_{\mathsf{HA}}(A)$, if A is atomic or boxed,
- $q_\sigma(A \wedge B, x) := q_\sigma(A, j_1(x)) \wedge q_\sigma(B, j_2(x))$,
- $q_\sigma(A \vee B, x) := (j_1(x) = 0 \to q_\sigma(A, j_2(x))) \wedge (j_1(x) \neq 0 \to q_\sigma(B, j_2(x)))$,
- if $A = B \to C$ and $B \in \mathsf{NOI}$, we define $q_\sigma(B \to C, x) := \sigma_{\mathsf{HA}}(B) \to (\{x\}(n_B)\downarrow \wedge q_\sigma(C, \{x\}(n_B)))$, in which n_B is as in Lemma 3.13. If $B \notin \mathsf{NOI}$, then define $q_\sigma(A, x) := \sigma_{\mathsf{HA}}(A)$.

Lemma 3.15. *Let A be a modal proposition and t be a term in first-order language of arithmetic which possibly contain Kleene's brackets. Then*

- $\mathsf{HA}_0 \vdash x \ \mathsf{q}\ \sigma_{\mathsf{HA}}(A) \to q_\sigma(A, x)$,
- $\mathsf{HA}_0 \vdash (t\downarrow \wedge q_\sigma(A, t)) \to \sigma_{\mathsf{HA}}(A)$,

Proof. Proof of both parts are by induction on the complexity of A. ❑

For the next lemma, we need some auxiliary notation $\sigma_l(A, x)$. Informally speaking, $\sigma_l(A, x)$ is going to be $\sigma_{\mathsf{HA}}(A^l)$ with one difference. The new added boxes in A^l should be interpreted as provability in HA_x. More precisely, we define it inductively as the following.

- A is atomic or boxed. $\sigma_l(A, x) := \sigma_{\mathsf{HA}}(A)$,
- $A = B \wedge C$. then $\sigma_l(A, x) := \sigma_l(B, x) \wedge \sigma_l(C, x)$,

- $A = B \vee C$. then $\sigma_{\imath}(A, x) := \boxdot_x \sigma_{\imath}(B, x) \vee \boxdot_x \sigma_{\imath}(C, x)$, in which $\boxdot_x D$ is defined as $D \wedge \Box_x D$,
- $A = B \to C$. Like the definition of A^l, we define $\sigma_{\imath}(A, x)$ by cases. If $B \in \mathsf{NOI}$, then we define $\sigma_{\imath}(A, x) := \sigma_{\mathsf{HA}}(B) \to \sigma_{\imath}(C, x)$, otherwise we define $\sigma_{\imath}(A, x) := \sigma_{\mathsf{HA}}(A)$.

Lemma 3.16. *Let A be a modal proposition. Then*

1. $\mathsf{HA}_0 \vdash (x \leq y \wedge \sigma_{\imath}(A, x)) \to \sigma_{\imath}(A, y)$,
2. $\mathsf{HA}_0 \vdash \sigma_{\imath}(A, x) \to \sigma_{\mathsf{HA}}(A^l)$,
3. $\mathsf{HA}_0 \vdash \sigma_{\imath}(A, x) \to \sigma_{\mathsf{HA}}(A)$.

Proof. Use induction on A. ❑

Lemma 3.17. *Let A be a modal proposition, D be any Σ_1-sentence and t be a term in first-order language of arithmetic which possibly contain Kleene's brackets. Then there exists a provably total recursive function f such that*

$$\mathsf{HA} \vdash \Box_x(D \to (t{\downarrow} \wedge q_\sigma(A, t)) \to \Box_{f(x)}(D \to \sigma_{\imath}(A, f(x)))$$

Proof. We use induction on A. For simplicity of notations, we assume here that t is a normal term. One can easily build the general case.

Atomic, Boxed or conjunction. Trivial.

Disjunction. Let $A = B \vee C$. Then by definition of q_σ, we have

$$\mathsf{HA} \vdash \Box_x(D \to q_\sigma(B \vee C, t)) \to [\Box_x((D \wedge j_1(t) = 0) \to q_\sigma(B, j_2(t))) \wedge \Box_x((D \wedge j_1(t) \neq 0) \to q_\sigma(C, j_2(t))]$$

Hence by the induction hypothesis, there exists functions g and h such that

$$\begin{aligned}\mathsf{HA} \vdash &\Box_x(D \to q_\sigma(B \vee C, t)) \to \\ &\Box_{g(x)}((D \wedge j_1(t) = 0) \to \sigma_{\imath}(B, g(x))) \wedge \Box_{h(x)}((D \wedge j_1(t) \neq 0) \to \sigma_{\imath}(C, h(x)))\end{aligned}$$

Let $f(x)$ be the maximum of $g(x)$ and $h(x)$. One can use the Σ_1-completeness of HA_0 (Lemma 3.6) and Lemma 3.16 to derive

$$\mathsf{HA} \vdash \Box_x(D \to q_\sigma(B \vee C, t)) \to \Box_{f(x)}(D \to (\boxdot_{f(x)} \sigma_{\imath}(B, f(x)) \vee \boxdot_{f(x)} \sigma_{\imath}(C, f(x))))$$

Implication. Assume that $A = B \to C$. If $B \notin \mathsf{NOI}$, by Lemma 3.15, we are done. So assume that $B \in \mathsf{NOI}$. By definition of q_σ, there exists some term t_1 such that

$$\mathsf{HA} \vdash \Box_x[D \to q_\sigma(B \to C, t)] \to \Box_x[(D \wedge \sigma_{\mathsf{HA}}(B)) \to (t_1{\downarrow} \wedge q_\sigma(C, t_1))]$$

Since $B \in \mathsf{NOI}$, $\sigma_{\mathsf{HA}}(B)$ is a Σ_1-formula. Hence by the induction hypothesis, there exists some function f such that

$$\mathsf{HA} \vdash \Box_x(D \to q_\sigma(A, t)) \to \Box_{f(x)}((D \wedge \sigma_{\mathsf{HA}}(B)) \to \sigma_{\imath}(C, f(x)))$$

This by definition of $\sigma_{\imath}(B \to C, f(x))$, implies the desired result. ❑

Lemma 3.18. *For any Σ_1-substitution σ and modal proposition A, there exists some provably total recursive function g such that $\mathsf{HA} \vdash \Box_x \sigma_{\mathsf{HA}}(A) \to \Box_{g(x)} \sigma_{\imath}(A, g(x))$.*

Proof. Work inside HA. Assume $\Box_x \sigma_{\mathsf{HA}}(A)$. By Lemma 3.11, there exists some y such that

$$\Box_{f_0(x)}(t{\downarrow} \wedge t \ \mathsf{q}\ \sigma_{\mathsf{HA}}(A))$$

in which $t := \{y\}\langle\rangle$ and f_0 is a provably total recursive function as stated in Lemma 3.11. Hence by the first item of Lemma 3.15, $\Box_{f_0(x)}(t{\downarrow} \wedge q_\sigma(A, t))$. Hence by Lemma 3.17, we have the function f such that $\Box_{f(f_0(x))} \sigma_{\imath}(A, f(f_0(x))$. ❑

Theorem 3.19. *For any Σ_1-substitution σ, we have* $\mathsf{HA} \vdash \sigma_{\mathsf{HA}}(\mathsf{Le}^+)$.

Proof. Let A be a modal proposition. We must show $\mathsf{HA} \vdash \Box\sigma_{\mathsf{HA}}(A) \to \Box\sigma_{\mathsf{HA}}(A^l)$. Now the desired result may be deduced by Lemma 3.18 and the second item of Lemma 3.16. ❑

Although there are other ways of proving the above theorem (see [Vis02] or [Iem01]), we need its major preliminary lemma (i.e. Lemma 3.18) in the proof of the completeness theorem. Specially, we use Lemma 3.18 in the proof of Lemma 5.12.

3.4 Interpretability

Let T and S be two first-order theories. Informally speaking, we say that T interprets S ($T \rhd S$) if there exists a translation from the language of S to the language of T such that T proves the translation of all of the theorems of S. For a formal definition see [Vis98]. It is well-known that for recursive theories T and S containing PA, the assertion $T \rhd S$ is formalizable in first-order language of arithmetic. For two arithmetical sentences A and B, we use the notation $A \rhd B$ to mean that $\mathsf{PA} + A$ interprets $\mathsf{PA} + B$. The following theorem due to Orey, first appeared in [Fef60].

Theorem 3.20. *For recursive theories T and S containing* PA*, we have:*

$$\mathsf{PA} \vdash (T \rhd S) \leftrightarrow \forall x\, \Box_T \mathsf{Con}(S^x),$$

in which S^x is the restriction of the theory S to axioms with Gödel number $\leq x$ and $\mathsf{Con}(U) := \neg \Box_U \bot$.

Proof. See [Fef60]. p.80 or [Ber90]. ❑

Convention. From Theorem 3.20, one can easily observe that $\mathsf{PA} \vdash (A \rhd B) \leftrightarrow \forall x\, \Box^+(A \to \neg\Box^+_x \neg B)$. So from now on, $A \rhd B$ means its Π_2-equivalent $\forall x\, \Box^+(A \to \neg\Box^+_x \neg B)$, even when we are working in weaker theories like HA. We remind the reader that $\Box^+$ stands for provability in PA.

4 Propositional modal logics

In this section, we collect all the required notions with propositional flavour. This section is mostly devoted to provide an axiomatic system for the Σ_1-provability logic of HA, i.e. iH_σ, and stating some of its essential properties that we need them later in the proof of soundness (Theorem 6.3) or completeness (Theorem 6.5) of iH_σ for arithmetical Σ_1-interpretations. The following are some of important results that will be used in the proof of completeness theorem.

- In Section 4.3, it is shown that the axiomatic system iH_σ is capable of simplifying any modal proposition to an equivalent TNNIL^- proposition (Corollary 4.19). This fact is useful for proof of the completeness theorem (Theorem 6.5).
- In Section 4.4, the TNNIL-conservativity of the theory LC over iH_σ (Theorem 4.24) is proved. This conservativity plays an important role in the proof of completeness theorem. As far as working with TNNIL-formulas, we get rid of all those complicated axioms of iH_σ and just use the more handful theory LC.
- In Section 4.5, we will prove the finite model property for the theory LC (Theorem 4.26). With the aid of our main theorem in next section (Theorem 5.1), such finite counter-models are used to be transformed to a first-order counter-models of HA.

4.1 The NNIL formulae and related topics

The class of *No Nested Implications to the Left*, NNIL formulae in a propositional language was introduced in [VvBdJRdL95], and more explored in [Vis02]. The crucial result of [Vis02] is providing an algorithm that as input, receives a non-modal proposition A and returns its best NNIL approximation A^* from below, i.e., $\mathsf{IPC} \vdash A^* \to A$ and for all NNIL formula B such that $\mathsf{IPC} \vdash B \to A$, we have $\mathsf{IPC} \vdash B \to A^*$. Also for all Σ_1-substitutions σ, we have $\mathsf{HA} \vdash \sigma_{\mathsf{HA}}(\Box A \leftrightarrow \Box A^*)$ [Vis02].

- In Section 4.1.1, we state Visser's NNIL-algorithm for computing A^*, and some of its useful properties.
- In Section 4.1.2, we explain the extension of this algorithm to the modal language (the TNNIL-algorithm), which computes A^+ and is essentially the same as the NNIL-algorithm with this extra rule: treat inside $\Box$ as a fresh proposition, i.o.w. in the inductive definition of the algorithm $(\Box A)^+ := \Box A^+$. Then we prove some useful properties of the TNNIL-algorithm: Lemma 4.7 and Corollary 4.8. The best feature of TNNIL-algorithm is that for all Σ_1-substitutions σ, we have $\mathsf{HA} \vdash \sigma_{\mathsf{HA}}(\Box A \leftrightarrow \Box A^+)$ (first part of Corollary 4.8).
- In Section 4.1.3, we define another algorithm TNNIL^- for computing A^-, which is essentially the same as the TNNIL-algorithm, with this minor difference: Only treat those sub-formulae which are boxed and leave the others. With this minor change, we even have a better feature for A^-, i.e., for all Σ_1-substitutions σ, we have $\mathsf{HA} \vdash \sigma_{\mathsf{HA}}(A \leftrightarrow A^-)$ (Lemma 4.10).

Now we define the class NNIL of modal propositions precisely by $\mathsf{NNIL} := \{A \mid \rho A \leq 1\}$, in which the complexity measure ρ, is defined inductively as follows:

- $\rho(\Box A) = \rho(p) = \rho(\bot) = \rho(\top) = 0$, for an arbitrary atomic variables p and modal proposition A,
- $\rho(A \wedge B) = \rho(A \vee B) = \max(\rho A, \rho B)$,
- $\rho(A \to B) = \max(\rho A + 1, \rho B)$,

In the following, we define another complexity measure $\mathfrak{o}(.)$ on modal propositions. We need this measure for termination of the NNIL-algorithm.

Definition 4.1. *Let D be a modal proposition. Let*

- $I(D) := \{E \in \mathsf{Sub}(D) \mid E$ *is an implication that is not in the scope of a* $\Box\}$.
- $\mathfrak{i}(D) := \max\{|I(E)| \mid E \in I(D)\}$, *where $|X|$ is the number of elements of X.*
- $\mathfrak{c}D :=$ *the number of occurrences of logical connectives which are not in the scope of a* $\Box$.
- $\mathfrak{d}D :=$ *the maximum number of nested boxes. To be more precise,*
  - $\mathfrak{d}D := 0$ *for atomic D,*
  - $\mathfrak{d}D := \max\{\mathfrak{d}D_1, \mathfrak{d}D_2\}$, *where $D = D_1 \circ D_2$ and $\circ \in \{\wedge, \vee, \to\}$,*
  - $\mathfrak{d}\Box D := \mathfrak{d}D + 1$,
- $\mathfrak{o}D := (\mathfrak{d}D, \mathfrak{i}D, \mathfrak{c}D)$.

We order the measures $\mathfrak{o}D$ lexicographically, i.e., $(d, i, c) < (d', i', c')$ iff $d < d'$ or $d = d', i < i'$ or $d = d', i = i', c < c'$.

For definition of NNIL-algorithm, we use the bracket notation $[A]B$ from [Vis02]. We also use a variant of this notation, $[A]'B$:

Definition 4.2. *For any two modal propositions A and B, we define $[A]B$ and $[A]'B$ by induction on the complexity of B:*

- $[A]B = [A]'B = B$, *for atomic or boxed* B,
- $[A](B_1 \circ B_2) = [A](B_1) \circ [A](B_2)$, $[A]'(B_1 \circ B_2) = [A]'(B_1) \circ [A]'(B_2)$ *for* $\circ \in \{\vee, \wedge\}$,
- $[A](B_1 \to B_2) = A \to (B_1 \to B_2)$, $[A]'(B_1 \to B_2) = A' \to (B_1 \to B_2)$, *in which* $A' = A[B_1 \to B_2 \mid B_2]$, *i.e., replace each outer occurrence of* $B_1 \to B_2$ *(by outer occurrence we mean that it is not in the scope of any* $\Box$*) in* A *by* B_2,

For a set X *of modal propositions, we also define* $[A]X := \bigvee_{B \in X} A[B]$ *and* $[A]'X := \bigvee_{B \in X} [A]'B$.

Remark 4.3. *It is easy to observe that* $[A]B$ *and* $[A]'B$ *are equivalent in* $\mathsf{IPC}_\Box$.

4.1.1 The NNIL-algorithm

For each modal proposition A, the proposition A^* is defined by induction on $\mathfrak{o}A$ as follows [Vis02]:

1. A is atomic or boxed, take $A^* := A$.
2. $A = B \wedge C$, take $A^* := B^* \wedge C^*$.
3. $A = B \vee C$, take $A^* := B^* \vee C^*$.
4. $A = B \to C$, we have several sub-cases. In the following, an occurrence of E in D is called an *outer occurrence*, if E is neither in the scope of an implication nor in the scope of a boxed formula.
   (a) C contains an outer occurrence of a conjunction. In this case, there is some formula $J(q)$ such that
      - q is a propositional variable not occurring in A.
      - q is outer in J and occurs exactly once.
      - $C = J[q|(D \wedge E)]$.

      Now set $C_1 := J[q|D], C_2 := J[q|E]$ and $A_1 := B \to C_1, A_2 := B \to C_2$ and finally, define $A^* := A_1^* \wedge A_2^*$.
   (b) B contains an outer occurrence of a disjunction. In this case, there is some formula $J(q)$ such that
      - q is a propositional variable not occurring in A.
      - q is outer in J and occurs exactly once.
      - $B = J[q|(D \vee E)]$.

      Now set $B_1 := J[q|D], B_2 := J[q|E]$ and $A_1 := B_1 \to C, A_2 := B_2 \to C$ and finally, define $A^* := A_1^* \wedge A_2^*$.
   (c) $B = \bigwedge X$ and $C = \bigvee Y$ and X, Y are sets of implications or atoms. We have several sub-cases:
      i. X contains atomic variables or boxed formula E. We set $D := \bigwedge(X \setminus \{E\})$ and take $A^* := E^* \to (D \to C)^*$.
      ii. X contains $\top$. Define $D := \bigwedge(X \setminus \{\top\})$ and take $A^* := (D \to C)^*$.
      iii. X contains $\bot$. Take $A^* := \top$.
      iv. X contains only implications. For any $D = E \to F \in X$, define

      $$B{\downarrow}D := \bigwedge((X \setminus \{D\}) \cup \{F\}).$$

      Let $Z := \{E \mid E \to F \in X\} \cup \{C\}$ and define:

      $$A^* := \bigwedge\{((B{\downarrow}D) \to C)^* | D \in X\} \wedge \bigvee\{([B]'E)^* \mid E \in Z\}$$

      We should show $\mathfrak{o}([B]'E) < \mathfrak{o}A$. For a proof of this fact see [Vis02].

Remark 4.4. In fact in [Vis02], the NNIL-algorithm is only for non-modal propositions. One may also compute the best NNIL-approximation for modal propositions, in the following way. Let A be a given modal proposition. Let $B_1, \ldots, B_n$ be all boxed sub-formulae of A which are not in the scope of any other boxes. Let $A'(p_1, \ldots, p_n)$ be the unique non-modal proposition such that $\{p_i\}_{1 \le i \le n}$ are fresh atomic variables not occurring in A and $A = A'[p_1|B_1, \ldots, p_n|B_n]$. Let $\gamma(A) := (A')^*[p_1|B_1, \ldots, p_n|B_n]$. Then it is easy to observe that $\mathsf{IPC}_\Box \vdash \gamma(A) \leftrightarrow A^*$.

The above defined algorithm is not deterministic, however from the following theorem we know that A^* is unique up to $\mathsf{IPC}_\Box$ equivalence. Notation $A \rhd_{\mathsf{IPC}_\Box, \mathsf{NNIL}} B$ (A, NNIL-preserves B) from [Vis02], means that for each NNIL modal proposition C, if $\mathsf{IPC}_\Box \vdash C \to A$, then $\mathsf{IPC}_\Box \vdash C \to B$, in which A, B are modal propositions.

Theorem 4.5. *For each modal proposition A,*

1. *The* NNIL *algorithm with input A terminates and the output formula A^*, is an* NNIL *proposition such that $\mathsf{IPC}_\Box \vdash A^* \to A$.*
2. $\mathsf{IPC}_\Box \vdash A^* \to B$ *iff* $A \rhd_{\mathsf{IPC}_\Box, \mathsf{NNIL}} B$.
3. *A^* is the best* NNIL *approximation of A from below i.e. $\mathsf{IPC}_\Box \vdash A^* \to A$ and for each* NNIL *proposition B, with $\mathsf{IPC}_\Box \vdash B \to A$, we have $\mathsf{IPC}_\Box \vdash B \to A^*$.*
4. $\mathsf{IPC}_\Box \vdash A_1 \to A_2$ *implies* $\mathsf{IPC}_\Box \vdash A_1^* \to A_2^*$.
5. $\mathsf{IPC}_\Box \vdash A \leftrightarrow B$ *implies* $\mathsf{IPC}_\Box \vdash A^* \leftrightarrow B^*$.
6. *For each Σ_1-substitution σ,* $\mathsf{HA} \vdash \Box\sigma_{\mathsf{HA}}(A) \leftrightarrow \Box\sigma_{\mathsf{HA}}(A^*)$.

Proof. 1. Direct consequence of [Vis02, Theorem 7.1]. First assume $A'[p_1, \ldots, p_n]$ be as in Remark 4.4. By [Vis02, Theorem 7.1], we have $\mathsf{IPC} \vdash (A')^* \to A'$, and hence by Lemma 2.3, $\mathsf{IPC}_\Box \vdash (A')^*[p_1|B_1, \ldots, p_n|B_n] \to A$.

2. Direct consequence of [Vis02, Theorem 7.2]. First suppose that $\mathsf{IPC}_\Box \vdash A^* \to B$. Let A', B' be non-modal propositions as defined in Remark 4.4, i.e, $A = A'[p_1|C_1, \ldots, p_n|C_n], B = B'[p_1|C_1, \ldots, p_n|C_n]$. Then by Lemma 2.3, $\mathsf{IPC} \vdash (A')^* \to B'$. Now by [Vis02, Theorem 7.2], we have $A' \rhd_{\mathsf{IPC}, \mathsf{NNIL}} B'$, and then by Lemma 2.3, $A \rhd_{\mathsf{IPC}_\Box, \mathsf{NNIL}} B$. For the proof of the other way around, note that all of the previous deductions are reversible.

3. Suppose $\mathsf{IPC}_\Box \vdash B \to A$ and B is NNIL. Since $\mathsf{IPC}_\Box \vdash A^* \to A^*$, from item 2 above, we get $A \rhd_{\mathsf{IPC}_\Box, \mathsf{NNIL}} A^*$. By $\mathsf{IPC}_\Box \vdash B \to A$ and $B \in \mathsf{NNIL}$, we have $\mathsf{IPC}_\Box \vdash B \to A^*$.

4. Suppose that $\mathsf{IPC}_\Box \vdash A_1 \to A_2$. By part 1, $\mathsf{IPC}_\Box \vdash A_1^* \to A_2$ and hence by item 3, $\mathsf{IPC}_\Box \vdash A_1^* \to A_2^*$.

5. Direct consequence of item 4.

6. First suppose that A is a non-modal proposition. Combining Theorem 10.2 and Corollary 7.2 from [Vis02], implies that $\mathsf{IPC} \vdash A^* \to B$ iff $A \mathrel{|\!\!\sim}^{\mathsf{HA}}_{\mathsf{HA}, \Sigma} B$, in which $A \mathrel{|\!\!\sim}^{\mathsf{HA}}_{\mathsf{HA}, \Sigma} B$ means that for each Σ_1-substitution σ, we have $\mathsf{HA} \vdash \Box\sigma_{\mathsf{HA}}(A) \to \Box\sigma_{\mathsf{HA}}(B)$. This implies that $\mathsf{HA} \vdash \Box\sigma_{\mathsf{HA}}(A) \leftrightarrow \Box\sigma_{\mathsf{HA}}(A^*)$. Now for a modal proposition A, suppose that $A'(p_1, \ldots, p_n)$ and $B_1, \ldots, B_n$ be such that $A = A'[p_1|B_1, \ldots, p_n|B_n]$, in which A' is a non-modal proposition and $p_1, \ldots, p_n$ are fresh atomic variables (not occurred in A). Let σ' be the substitution defined by $\sigma'(p_i) := \sigma_{\mathsf{HA}}(B_i)$, for each $1 \le i \le n$, and for any other atomic variable q, $\sigma'(q) = \sigma(q)$. Clearly, σ' is again a Σ_1-substitution and hence we have $\mathsf{HA} \vdash \Box\sigma'_{\mathsf{HA}}(A') \leftrightarrow \Box\sigma'_{\mathsf{HA}}((A')^*)$. This implies $\mathsf{HA} \vdash \Box\sigma_{\mathsf{HA}}(A) \leftrightarrow \Box\sigma_{\mathsf{HA}}(A^*)$.

❑

4.1.2 The TNNIL-algorithm

Definition 4.6. TNNIL *(Thoroughly* NNIL*) is the smallest class of propositions such that*

- TNNIL *contains all atomic propositions,*
- *if* $A, B \in \mathsf{TNNIL}$*, then* $A \vee B, A \wedge B, \Box A \in \mathsf{TNNIL}$*,*
- *if all* $\rightarrow$ *occurring in* A *are contained in the scope of a* $\Box$ *(or equivalently* $A \in \mathsf{NOI}$*) and* $A, B \in \mathsf{TNNIL}$*, then* $A \rightarrow B \in \mathsf{TNNIL}$*.*

Let TNNIL^- *indicates the set of all the propositions like* $A(\Box B_1, \ldots, \Box B_n)$*, such that* $A(p_1, \ldots, p_n)$ *is an arbitrary non-modal proposition and* $B_1, \ldots, B_n \in \mathsf{TNNIL}$*.*

Here we define A^+ to be the TNNIL-formula approximating A. The major difference between A^+ and A^* is that $\mathsf{IPC}_\Box \vdash A^+ \rightarrow A$ may not hold any more. Informally speaking, to find A^+, we first compute A^* and then replace all outer boxed formula $\Box B$ in A by $\Box B^+$. To be more accurate, we define A^+ by induction on $\mathfrak{d}A$. Suppose that for all B with $\mathfrak{d}B < \mathfrak{d}A$, we have defined B^+. Now suppose that $A'(p_1, \ldots, p_n)$ and $\Box B_1, \ldots, \Box B_n$ are such that $A = A'[p_1|\Box B_1, \ldots, p_n|\Box B_n]$, where A' is a non-modal proposition and $p_1, \ldots, p_n$ are fresh atomic variables (not occurred in A). It is clear that $\mathfrak{d}B_i < \mathfrak{d}A$ and then we can define $A^+ := (A')^*[p_1|\Box B_1^+, \ldots, p_n|\Box B_n^+]$.

Lemma 4.7. *For every modal proposition* A*,*

1. *If* $\mathsf{iGL} \vdash A$ *then* $\mathsf{iGL} \vdash A^+$*.*
2. *If* $\mathsf{iK4} \vdash A$ *then* $\mathsf{iK4} \vdash A^+$*.*

Proof. We prove the first part by induction on the complexity of proof $\mathsf{iGL} \vdash A$. Proof of the second part is similar to the first one.

- A is an axiom.
  - A is Löb's axiom, i.e., $A = \Box(\Box B \rightarrow B) \rightarrow \Box B$. Then $A^+ = \Box(\Box B^+ \rightarrow B^+) \rightarrow \Box B^+$, that is valid also in iGL.
  - $A = \Box B \rightarrow \Box\Box B$. Then $A^+ = \Box B^+ \rightarrow \Box\Box B^+$, that is valid in iGL.
  - $A = (\Box(B \rightarrow C) \wedge \Box B) \rightarrow \Box C$. Then $A^+ = (\Box(B \rightarrow C)^+ \wedge \Box B^+) \rightarrow \Box C^+$. On the other hand, $\mathsf{IPC}_\Box \vdash (B \wedge (B \rightarrow C)) \rightarrow C$ and hence $\mathsf{IPC}_\Box \vdash (B \wedge (B \rightarrow C))^* \rightarrow C^*$, by Theorem 4.5 item 4. Now we can infer $\mathsf{IPC}_\Box \vdash (B^+ \wedge (B \rightarrow C)^+) \rightarrow C^+$, by definition of TNNIL-algorithm and Lemma 2.3. Finally, by the necessitation rule in iGL, we have $\mathsf{iGL} \vdash (\Box B^+ \wedge \Box(B \rightarrow C)^+) \rightarrow \Box C^+$.
- A is a theorem of $\mathsf{IPC}_\Box$. Then $\mathsf{IPC}_\Box \vdash A^+$, by Theorem 4.5 item 5 and Lemma 2.3.
- $A = \Box B$ and A is derived by applying the necessitation rule. Let $\mathsf{iGL} \vdash B$. By induction hypothesis, $\mathsf{iGL} \vdash B^+$ and then $\mathsf{iGL} \vdash \Box B^+$.
- A is derived by modus ponens. Let $\mathsf{iGL} \vdash B$ and $\mathsf{iGL} \vdash B \rightarrow A$. From these, we have $\mathsf{iGL} \vdash B^+ \wedge (B \rightarrow A)^+$ and then $\mathsf{iGL} \vdash (B \wedge (B \rightarrow A))^+$. Since $\mathsf{IPC}_\Box \vdash (B \wedge (B \rightarrow A)) \rightarrow A$, then by Theorem 4.5 item 4 we have $\mathsf{IPC}_\Box \vdash (B \wedge (B \rightarrow A))^* \rightarrow A^*$. Then by Lemma 2.3, $\mathsf{IPC}_\Box \vdash (B \wedge (B \rightarrow A))^+ \rightarrow A^+$ and hence $\mathsf{iGL} \vdash A^+$ as desired.

❑

Corollary 4.8. *For any modal proposition* A*,*

1. *For all* Σ_1*-substitution* σ *we have* $\mathsf{HA} \vdash \Box\sigma_{\mathsf{HA}}(A) \leftrightarrow \Box\sigma_{\mathsf{HA}}(A^+)$ *and hence* $\mathsf{HA} \vdash \sigma_{\mathsf{HA}}(A)$ *iff* $\mathsf{HA} \vdash \sigma_{\mathsf{HA}}(A^+)$*,*

2. $\mathsf{iGL} \vdash A_1 \to A_2$ *implies* $\mathsf{iGL} \vdash A_1^+ \to A_2^+$, *and* $\mathsf{iK4} \vdash A_1 \to A_2$ *implies* $\mathsf{iK4} \vdash A_1^+ \to A_2^+$,

3. $\mathsf{iGL} \vdash A_1 \leftrightarrow A_2$ *implies* $\mathsf{iGL} \vdash A_1^+ \leftrightarrow A_2^+$, *and* $\mathsf{iK4} \vdash A_1 \leftrightarrow A_2$ *implies* $\mathsf{iK4} \vdash A_1^+ \leftrightarrow A_2^+$.

Proof. The first assertion can be deduced simply by induction on $\mathfrak{d}A$ and using Theorem 4.5 item 6.

To prove the second part, first note that by Theorem 4.5 item 4, if $\mathsf{IPC}_\Box \vdash A_1 \to A_2$, then $\mathsf{IPC}_\Box \vdash A_1^* \to A_2^*$. By Lemma 2.3, we can replace each outer occurrence of boxed formulae by arbitrary propositions, in particular, by their TNNIL approximations. We should take care of these replacements to be such that equal propositions be substituted by equal approximations and unequal propositions substituted by unequal ones. Then by definition of A_i^+, we have $\mathsf{IPC}_\Box \vdash A_1^+ \to A_2^+$.

Now suppose that $\mathsf{iGL} \vdash A_1 \to A_2$ ($\mathsf{iK4} \vdash A_1 \to A_2$). Let $A = A_1 \to A_2$. This implies $\mathsf{IPC}_\Box \vdash (A \wedge A_1) \to A_2$. Then $\mathsf{IPC}_\Box \vdash (A \wedge A_1)^+ \to A_2^+$, and hence by TNNIL-algorithm, $\mathsf{IPC}_\Box \vdash (A^+ \wedge A_1^+) \to A_2^+$. This implies $\mathsf{IPC}_\Box + A^+ \vdash A_1^+ \to A_2^+$ and by Lemma 4.7, $\mathsf{iGL} \vdash A_1^+ \to A_2^+$ ($\mathsf{iK4} \vdash A_1^+ \to A_2^+$).

Proof of the third part is a direct consequence of the second part. ❑

4.1.3 The TNNIL^--algorithm

Corollary 4.9. *There exists a* TNNIL^-*-algorithm such that for any modal proposition* A*, it halts and produces a proposition* $A^- \in \mathsf{TNNIL}^-$ *such that* $\mathsf{IPC}_\Box \vdash A^+ \to A^-$.

Proof. Let $A := B(\Box C_1, \ldots, \Box C_n)$, and $B(p_1, \ldots, p_n)$ is non-modal. Clearly such B exists. Then define $A^- := B(\Box C_1^+, \ldots, \Box C_n^+)$. Now definition of A^+ implies $A^+ = (A^-)^*$ and hence Theorem 4.5 item 1 implies that A^- has desired property. ❑

Lemma 4.10. *For each modal proposition* A *and* Σ_1*-substitution* σ, $\mathsf{HA} \vdash \sigma_{\mathsf{HA}} A \leftrightarrow \sigma_{\mathsf{HA}} A^-$.

Proof. Use definition of $(.)^-$ and Corollary 4.8 item 1. ❑

Remark 4.11. Note that $\mathsf{LC} \vdash A \leftrightarrow B$ does not imply $\mathsf{LC} \vdash A^+ \leftrightarrow B^+$. A counterexample is $A := \neg\neg p$ and $B := \neg \boxdot(\neg p)$. We have $A^+ = A^* = p$ and $B^+ = (\Box\neg p \to p)$. Now one can use Kripke models to show $\mathsf{LC} \nvdash (\Box\neg p \to p) \to p$.

Remark 4.12. In the algorithm produced for NNIL, let's change the step (1) in this way (and use new symbol $(.)^\dagger$ instead of $(.)^*$)

1. $A^\dagger := A$ for atomic A, and $(\Box B)^\dagger := \Box B^\dagger$,

Then the new algorithm also halts, and for any modal proposition A, we have $\mathsf{iK4} \vdash A^\dagger \leftrightarrow A^+$.

4.2 The Box Translation

The following definition of the box-translation, is essentially from [Vis82, Definition 4.1]. The box-translation extends the well-known Gödel-McKinsey-Tarski translation. In this subsection, we prove that iGL is closed under box-translation (Proposition 4.16).

Definition 4.13. *For every proposition* A *in the modal propositional language, we associate a proposition* $A^\Box$*, called the box-translation of* A*, in the following way:*

- $A^\Box := A \wedge \Box A$, *for atomic* A,
- $(A \circ B)^\Box := A^\Box \circ B^\Box$, *for* $\circ \in \{\vee, \wedge\}$,
- $(A \to B)^\Box := (A^\Box \to B^\Box) \wedge \Box(A^\Box \to B^\Box)$,
- $(\Box A)^\Box := \Box(A^\Box)$.

Lemma 4.14. *For any modal proposition* A*, we have* $\mathsf{iK4} \vdash A^\Box \to \Box A^\Box$.

Proof. Easy induction over the complexity of A. ❑

In the following lemma we state some properties of $\boxdot$.

Lemma 4.15. *For any modal proposition A, the following propositions are provable in* $\mathsf{iK4}$*:*

1. $\Box \boxdot A \leftrightarrow \Box A \leftrightarrow \boxdot \Box A$,
2. $\boxdot A^{\Box} \leftrightarrow A^{\Box}$.

Proof. The first part is easily deduced in $\mathsf{iK4}$. For the second part use Lemma 4.14. ❑

We say that a modal theory T is *closed under box-translation* if for every proposition A, $T \vdash A$ implies $T \vdash A^{\Box}$.

Proposition 4.16. *The theory* iGL *is closed under the box-translation.*

Proof. The proof can be carried out in three steps:

1. For any proposition A first we show that $\mathsf{IPC}_{\Box} \vdash A$ implies $\mathsf{iK4} \vdash A^{\Box}$. This can be done by a routine induction on the length of the proof in IPC. Note that for any axiom A of IPC, we have $\mathsf{iK4} \vdash A^{\Box}$. As for the rule of modus ponens, suppose that $\mathsf{IPC}_{\Box} \vdash A$ and $\mathsf{IPC}_{\Box} \vdash A \to B$. By induction hypothesis, then $\mathsf{iK4} \vdash A^{\Box}$ and $\mathsf{iK4} \vdash (A^{\Box} \to B^{\Box}) \wedge \Box(A^{\Box} \to B^{\Box})$ and so $\mathsf{iK4} \vdash B^{\Box}$.

2. Next observe that
$$(\Box A \to \Box\Box A)^{\Box} = \Box A^{\Box} \to \Box\Box A^{\Box}$$
and also
$$\mathsf{iK4} \vdash [(\Box(A \to B) \wedge \Box A) \to \Box B]^{\Box} \leftrightarrow [(\Box(A^{\Box} \to B^{\Box}) \wedge \Box A^{\Box}) \to \Box B^{\Box}]$$

3. Observe that the box translation of an instance of Löb's axiom L, is also an instance of L.

❑

4.3 Axiomatizing the TNNIL-algorithm

In this subsection we present axioms which we need for the TNNIL^{-}-algorithm $(.)^{-}$. More precisely, we will find some axiom set X such that $X \vdash A^{-} \leftrightarrow A$.

To do that, we use some relation $\blacktriangleright$ on modal propositions. A variant of this relation for non-modal case, first appeared in [Vis02]. The relation $\blacktriangleright$ is defined to be the smallest relation on modal propositions satisfying the following conditions:

A1. If $\mathsf{iK4} \vdash A \to B$, then $A \blacktriangleright B$,

A2. If $A \blacktriangleright B$ and $B \blacktriangleright C$, then $A \blacktriangleright C$,

A3. If $C \blacktriangleright A$ and $C \blacktriangleright B$, then $C \blacktriangleright A \wedge B$,

A4. If $A \blacktriangleright B$, then $\Box A \blacktriangleright \Box B$,

B1. If $A \blacktriangleright C$ and $B \blacktriangleright C$, then $A \vee B \blacktriangleright C$,

B2. Let X be a set of implications, $B := \bigwedge X$ and $A := B \to C$. Also assume that $Z := \{E | E \to F \in X\} \cup \{C\}$. Then $A \blacktriangleright [B]Z$,

B3. If $A \blacktriangleright B$, then for any atomic or boxed C we have $C \to A \blacktriangleright C \to B$.

Remark 4.17. *Let A, B and Z be as in* B2. *Then the relation $\rhd$ has the following additional property:*

$$A \rhd [B]'Z$$

The reason goes as follows. First observe, by induction on E and using A1-A3, that $[B]E \rhd [B]'E$. That by use of A1-A3 and B1, implies that $[B]Z \rhd [B]'Z$. Hence by A2 and B2, we have $A \rhd [B]'Z$.

The notation $A \rhd\lhd B$ means $A \rhd B$ and $B \rhd A$. Let us define the theory

$$\mathsf{iH}_\sigma := \mathsf{LLe}^+ + \mathsf{CP_a} + \{\Box \mathsf{A} \to \Box \mathsf{B} | \mathsf{A} \rhd \mathsf{B}\}$$

Note that by A1, the relation $\rhd$ contains all the pairs (A, B) such that $\mathsf{iK4} \vdash A \to B$. But it worth mentioning that the inclusion is strict. The axiom which makes $\rhd$ strictly superset of $\{(A, B) : \mathsf{iK4} \vdash A \to B\}$ is B2, i.e. in the absence of B2, the relation $\rhd$ is the same as provable implications in $\mathsf{iK4}$. However, with B2 the story is different, e.g. one can observe that $\neg\neg p \rhd p$, for any atomic p, holds while $\mathsf{iK4} \nvdash \neg\neg p \to p$.

Notation. In the rest of the paper, we use $A \equiv B$ as a shorthand for $\mathsf{iK4} \vdash A \leftrightarrow B$.

The following theorem, shows that A1-A4 and B1-B3, axiomatize the TNNIL algorithm:

Theorem 4.18. *For any modal proposition A, we have $A \rhd\lhd A^+$.*

Proof. We prove the desired result by induction on $\mathfrak{o}(A)$. Suppose we have the desired result for each proposition B with $\mathfrak{o}(B) < \mathfrak{o}(A)$. We treat A by the following cases.

1. (A1) A is atomic. Then $A^+ = A$ by definition, and the result holds trivially.

2. (A1-A4, B1) $A = \Box B, A = B \wedge C, A = B \vee C$. All these cases hold by induction hypothesis. In boxed case, we use induction hypothesis and A4. In conjunction, we use A1-A3 and in disjunction we use A1, A2 and B1.

3. $A = B \to C$. There are several sub-cases. Similar to the definition of the NNIL-algorithm, an occurrence of a sub-formula B of A is said to be an *outer occurrence* in A, if it is neither in the scope of a $\Box$ nor in the scope of $\to$.

   (a) (A1-A3) C contains an outer occurrence of a conjunction. We can treat this case using induction hypothesis and TNNIL-algorithm.

   (b) (A1-A3) B contains an outer occurrence of a disjunction. We can treat this case by induction hypothesis and TNNIL-algorithm.

   (c) $B = \bigwedge X$ and $C = \bigvee Y$, where X and Y are sets of implications, atoms and boxed formulae. We have several sub-cases:

      i. (A1, A2, B3) X contains atomic variables. Let p be an atomic variable in X. Set $D := \bigwedge(X \setminus \{p\})$. Then $A^+ \equiv p \to (D \to C)^+$. On the other hand, we have by induction hypothesis and A2 and B3, that $p \to (D \to C)^+ \rhd\lhd p \to (D \to C)$. Finally by A1 and A2, we have $A^+ \rhd\lhd A$.

      ii. (A1, A2, B3) X contains boxed formula. Similar to the previous case.

      iii. (A1, A2) X contains $\top$ or $\bot$. Trivial.

      iv. (A1-A3, B1-B3) X contains only implications. This case needs the axiom B2 and it seems to be the interesting case. we have:

$$A^+ \equiv \bigwedge \left\{ (B{\downarrow}D \to C)^+ \mid D \in X \right\} \wedge \bigvee \{([B]'E)^+ : E \in Z\}$$

By the argument in [Vis02], we have $\mathfrak{o}\,(B{\downarrow}D \to C) < \mathfrak{o}(A)$ and $\mathfrak{o}([B]'E) < \mathfrak{o}(A)$ and hence one can apply induction hypothesis on $B{\downarrow}D \to C$ and $[B]'E$. Then by induction hypothesis, A1-A3, B1 and B3, we have:

$$A^+ \blacktriangleright\!\!\blacktriangleleft \bigwedge \{B{\downarrow}D \to C \mid D \in X\} \wedge [B]'Z$$

First we show that for each $E \in Z$,

$$\mathsf{iK4} \vdash \Big(\bigwedge\{(B{\downarrow}D) \to C \mid D \in X\} \wedge [B]'E\Big) \to A \tag{4.1}$$

Since $[B]E$ and $[B]'E$ are $\mathsf{IPC}_\Box$-equivalent (Remark 4.3), it's enough to show that

$$\mathsf{iK4} \vdash \Big(\bigwedge\{(B{\downarrow}D) \to C \mid D \in X\} \wedge [B]E\Big) \to A \tag{4.2}$$

If $E = C$, we are done by $\mathsf{IPC}_\Box \vdash [B]C \to (B \to C)$. So let E be the antecedent of some $E \to F \in X$. We reason in $\mathsf{iK4}$. Assume $\bigwedge\{(B{\downarrow}D \to C \mid D \in X\}$, $[B]E$ and B as the hypothesis. We want to derive C. From B and $[B]E$, we derive E. Also from B, we derive $E \to F$, and so F. Hence we have $\bigwedge(X \setminus \{E \to F\}) \wedge F$, which implies C, as desired.
Now eq. (4.3) by use of A1 and A2 implies $A^+ \blacktriangleright A$.
To show the other way around, i.e. $A \blacktriangleright A^+$, we first show

$$A \blacktriangleright \bigwedge \{B{\downarrow}D \to C \mid D \in X\} \wedge [B]'Z \tag{4.3}$$

and then by use of induction hypothesis and A2, we can deduce $A \blacktriangleright A^+$, as desired. So it remains to show that eq. (4.3) holds. We have $\mathsf{IPC}_\Box \vdash A \to \bigwedge\{B{\downarrow}D \to C \mid D \in X\}$, and hence by A1, $A \blacktriangleright \bigwedge\{B{\downarrow}D \to C \mid D \in X\}$. On the other hand, by Remark 4.17, we have $A \blacktriangleright [B]'Z$. Now A3 implies eq. (4.3), as desired.

❑

Corollary 4.19. $\mathsf{iH}_\sigma \vdash \mathsf{A}^- \leftrightarrow \mathsf{A}$.

Proof. Let $A = B(\Box C_1, \Box C_2, \ldots, \Box C_n)$ where $B(p_1, \ldots, p_n)$ is a non-modal proposition. By definition of A^-, we have $A^- = B(\Box C_1^+, \ldots, \Box C_n^+)$. Then Theorem 4.18 implies that $\mathsf{iH}_\sigma \vdash \Box(\mathsf{C_i})^+ \leftrightarrow \Box\mathsf{C_i}$. Hence $\mathsf{iH}_\sigma \vdash \mathsf{A}^- \leftrightarrow \mathsf{A}$, as desired. ❑

4.4 TNNIL-Conservativity of LC over LLe$^+$

It is clearly the case that $\mathsf{LC} \supseteq \mathsf{LLe}^+$. One can use Kripke models (from the next section) to show $\neg\neg\Box\bot \in \mathsf{LC} \setminus \mathsf{LLe}^+$. This implies that the inclusion is strict. As we will see later in this section, LC and LLe^+ have the same TNNIL-theorems. To prove this, we need some lemmas.

Lemma 4.20. $\mathsf{iK4} + \mathsf{Le}^+ \vdash \mathsf{Le}$.

Proof. Assume some axiom instance of Le, $\Box(B \vee C) \to \Box(\Box B \vee C)$. Let $A := B \vee C$. By axiom schema Le^+, we have $\Box A \to \Box A^l$, which is $\Box(B \vee C) \to \Box(\boxdot B \vee \boxdot C)$. This implies (inside $\mathsf{iK4}$) $\Box(B \vee C) \to \Box(\Box B \vee C)$. ❑

Lemma 4.21. *For each modal proposition A,*

1. *If $A \in \mathsf{NOI}$, then $\mathsf{iK4} + \mathsf{CP_a} \vdash A \to \Box A$.*
2. $\mathsf{iK4} \vdash A^l \to A$.

3. *If* $A \in \mathsf{NOI}$*, then* $\mathsf{iK4} + \mathsf{CP_a} \vdash A^l \leftrightarrow A$*.*

4. $\mathsf{LLe}^+ \vdash \Box A^l \leftrightarrow \Box A$*.*

Proof. Proofs of items 1 to 3 are routine by induction on A. Item 4 is deduced from item 2, i.e. we have $\Box A^l \to \Box A$, by item 2 and $\Box A^l \leftarrow \Box A$ is exactly Le^+. ❑

Lemma 4.22. *For any* TNNIL *formula* A*, we have*

1. $\mathsf{LLe}^+ \vdash \boxdot A^l \leftrightarrow \boxdot A^\Box$*,*

2. $\mathsf{LLe}^+ \vdash A^\Box \to A$*,*

3. *If* $A \in \mathsf{NOI}$*, then* $\mathsf{LLe}^+ \vdash A^\Box \leftrightarrow A$*,*

4. $\mathsf{LLe}^+ \vdash \Box A \leftrightarrow \Box A^\Box$*.*

Proof. We prove all items by induction on the complexity of A, simultaneously. In the middle of proof, when we are using induction hypothesis of item $i, 1 \leq i \leq 4$, we mention the number in parenthesis that number and also when we deduce some item of lemma, we also mention the number of that part in parentheses as well.

Atomic: For atomic A, we have $A^l = A$ and $A^\Box = \boxdot A$, hence by properties of $\boxdot$ (Lemma 4.15 item 1), $\mathsf{iK4} \vdash \boxdot A^l \leftrightarrow \boxdot A^\Box$ (item 1) and $\mathsf{LLe}^+ \vdash A^\Box \leftrightarrow A$ (items 2 and 3), which by necessitation, that implies $\mathsf{LLe}^+ \vdash \Box A^\Box \leftrightarrow \Box A$ (item 4).

Boxed: Let $A = \Box B$. Then by definition, $A^l = A$ and $A^\Box = \Box B^\Box$. Hence by induction hypothesis (item 1), $\mathsf{LLe}^+ \vdash \boxdot B^l \leftrightarrow \boxdot B^\Box$. Then $\mathsf{LLe}^+ \vdash \boxdot\Box B^l \leftrightarrow \boxdot\Box B^\Box$. On the other hand, by Lemma 4.21 item 4, $\mathsf{LLe}^+ \vdash \Box B^l \leftrightarrow \Box B$. Hence $\mathsf{LLe}^+ \vdash \boxdot\Box B \leftrightarrow \boxdot\Box B^\Box$ (item 1). Also by induction hypothesis (item 4), $\mathsf{LLe}^+ \vdash \Box B \leftrightarrow \Box B^\Box$. Hence $\mathsf{LLe}^+ \vdash A \leftrightarrow A^\Box$ (items 2 and 3). That, by necessitation, implies $\mathsf{LLe}^+ \vdash \Box A \leftrightarrow \Box A^\Box$ (item 4).

Conjunction: This case is trivial.

Disjunction: Assume $A = B \vee C$. If $A \in \mathsf{NOI}$, then $B, C \in \mathsf{NOI}$ and hence induction hypothesis for B and C (item 3) implies $\mathsf{LLe}^+ \vdash A^\Box \leftrightarrow A$ (item 3). For the other parts, we have, by definition, $A^l = \boxdot B^l \vee \boxdot C^l$. Hence by induction hypothesis (item 1), $\mathsf{LLe}^+ \vdash A^l \leftrightarrow (\boxdot B^\Box \vee \boxdot C^\Box)$. Hence, by Lemma 4.15 item 2, we derive $\mathsf{LLe}^+ \vdash \boxdot A^l \leftrightarrow \boxdot A^\Box$ (item 1). To prove the item 2, note that, by induction hypothesis (item 2), $\mathsf{LLe}^+ \vdash B^\Box \to B$ and $\mathsf{LLe}^+ \vdash C^\Box \to C$. Hence $\mathsf{LLe}^+ \vdash (B \vee C)^\Box \to (B \vee C)$ (item 2). To prove item 4, we note that, by item 1 for A, we have $\mathsf{LLe}^+ \vdash \boxdot A^l \leftrightarrow \boxdot A^\Box$. Hence $\mathsf{LLe}^+ \vdash \Box\boxdot A^l \leftrightarrow \Box\boxdot A^\Box$, which implies $\mathsf{LLe}^+ \vdash \Box A^l \leftrightarrow \Box A^\Box$ (by Lemma 4.15 item 1). Now Lemma 4.21 item 4 implies $\mathsf{LLe}^+ \vdash \Box A \leftrightarrow \Box A^\Box$ (item 4).

Implication: Assume $A = B \to C$. Clearly $A \notin \mathsf{NOI}$ and $B \in \mathsf{NOI}$. We only show induction claim for item 1. The other items can be shown easily. By induction hypothesis (item 3), $\mathsf{LLe}^+ \vdash B^\Box \leftrightarrow B$, and also by Lemma 4.21 item 1, $\mathsf{LLe}^+ \vdash \boxdot B \leftrightarrow B$. Note that $\boxdot A^l = \boxdot(B \to C^l)$ and hence $\mathsf{LLe}^+ \vdash \boxdot A^l \leftrightarrow \boxdot(\boxdot B \to C^l)$. By Lemma 4.21 item 3, $\mathsf{LLe}^+ \vdash B^l \leftrightarrow B$, hence $\mathsf{LLe}^+ \vdash \boxdot A^l \leftrightarrow \boxdot(\boxdot B^l \to C^l)$. Now properties of $\boxdot$ implies that $\mathsf{LLe}^+ \vdash \boxdot A^l \leftrightarrow \boxdot(\boxdot B^l \to \boxdot C^l)$, and induction hypothesis (item 1), implies $\mathsf{LLe}^+ \vdash \boxdot A^l \leftrightarrow \boxdot(\boxdot B^\Box \to \boxdot C^\Box)$. This implies, again by properties of $\boxdot$, the desired result, $\mathsf{LLe}^+ \vdash \boxdot A^l \leftrightarrow \boxdot A^\Box$ (item 1). ❑

Lemma 4.23. *If* $\mathsf{LC} \vdash A$*, then* $\mathsf{iGL} \vdash A^\Box$*.*

Proof. From $\mathsf{LC} \vdash A$ we have $\mathsf{iGL} \vdash \bigwedge_i \boxdot(B_i \to \Box B_i) \to A$. Hence by Proposition 4.16, we have $\mathsf{iGL} \vdash [\bigwedge_i \boxdot(B_i \to \Box B_i) \to A]^\Box$. This implies $\mathsf{iGL} \vdash \bigwedge_i \boxdot(B_i^\Box \to \Box B_i^\Box) \to A^\Box$. Now Lemma 4.15 item 2 implies $\mathsf{iGL} \vdash A^\Box$, as desired. ❑

Theorem 4.24. LC *is* TNNIL*-conservative over* LLe^+*.*

Proof. Let $\mathsf{LC} \vdash A$. From Lemma 4.23 we have $\mathsf{iGL} \vdash A^{\Box}$ and hence $\mathsf{LLe}^{+} \vdash A^{\Box}$. Now Lemma 4.22 item 2 implies that $\mathsf{LLe}^{+} \vdash A$. ❑

4.5 Kripke semantics for LC

Let us first review results and notations from [Iem01] which will be used here. Assume two binary relations R and S on a set. Define $\alpha(R \circ S)\gamma$ iff there exists some β such that $\alpha R \beta$ and $\beta S \gamma$.

A Kripke model $\mathcal{K}$, for intuitionistic modal logic, is a quadruple $(K, <, \mathcal{R}, V)$, such that K is a set (we call its elements as nodes), $(K, <)$ is a partial ordering, $\mathcal{R}$ is a binary relation on K such that $(\leq \circ \mathcal{R}) \subseteq \mathcal{R}$, and V is a binary relation between nodes and atomic variables such that $\alpha V p$ and $\alpha \leq \beta$ implies $\beta V p$. Then we can extend V to the modal language with $\mathcal{R}$ corresponding to $\Box$ and $\leq$ for intuitionistic $\to$. More precisely, we define $\Vdash$ inductively as an extension of V as follows:

- $\mathcal{K}, \alpha \Vdash p$ iff $\alpha V p$, for atomic variable p,
- $\mathcal{K}, \alpha \Vdash A \vee B$ iff $\mathcal{K}, \alpha \Vdash A$ or $\mathcal{K}, \alpha \Vdash B$,
- $\mathcal{K}, \alpha \Vdash A \wedge B$ iff $\mathcal{K}, \alpha \Vdash A$ and $\mathcal{K}, \alpha \Vdash B$,
- $\mathcal{K}, \alpha \nVdash \bot$ and $\mathcal{K}, \alpha \Vdash \top$,
- $\mathcal{K}, \alpha \Vdash A \to B$ iff for all $\beta \geq \alpha$, $\mathcal{K}, \beta \Vdash A$ implies $\mathcal{K}, \beta \Vdash B$,
- $\mathcal{K}, \alpha \Vdash \Box A$ iff for all β with $\alpha \, \mathcal{R} \, \beta$, we have $\mathcal{K}, \beta \Vdash A$.

If also we assume that $\mathcal{R}$ is empty and restrict our attention to non-modal language, we have the usual Kripke models for intuitionistic (non-modal) logic. In the rest of paper, we may simply write $\alpha \Vdash A$ instead of $\mathcal{K}, \alpha \Vdash A$, if no confusion is likely. By an induction on the complexity of A, one can observe that $\alpha \Vdash A$ implies $\beta \Vdash A$ for all A and $\alpha \leq \beta$. We define the following notions.

- If $\alpha \leq \beta$, β is called to be above α and α is beneath β. If $\alpha \, \mathcal{R} \, \beta$, β is called to be a successor of α. We define $\mathcal{R}(\alpha)$ to be the set of all successors of α.
- A Kripke model is finite if its set of nodes is finite. A Kripke model is tree-frame if its set of nodes with ordering $\leq$ is a tree.
- A Kripke model $\mathcal{K} = (K, <, \mathcal{R}, V)$ is reverse well-founded iff K is well-founded with the ordering $\mathcal{R}^{-1}$.
- $\mathcal{K}$ is called *neat* iff $\alpha \, \mathcal{R} \, \gamma$ and $\alpha \leq \beta \leq \gamma$ implies $\alpha \, \mathcal{R} \, \beta$ or $\beta \, \mathcal{R} \, \gamma$.
- $\mathcal{K}$ is called *brilliant* iff $(\mathcal{R} \circ \leq) \subseteq \mathcal{R}$. ([Iem01])
- $\mathcal{K}$ is called *perfect* iff it is brilliant, reverse well-founded and $R \subseteq <$.
- Suppose X is a set of propositions that is closed under sub-formulae (we call such X to be adequate). An X-saturated set of propositions Γ with respect to some theory T is a subset of X that
  - For each $A \vee B \in X$, $T + \Gamma \vdash A \vee B$ implies $A \in \Gamma$ or $B \in \Gamma$.
  - For each $A \in X$, $T + \Gamma \vdash A$ implies $A \in \Gamma$.

Lemma 4.25. *Let $T \nvdash A$ and let X be an adequate set. Then there is an X-saturated set Γ such that $T \cap X \subseteq \Gamma \nvdash A$.*

Proof. See [Iem01]. ❑

Theorem 4.26. LC *is sound and complete for finite neat perfect Kripke models with tree frames.*

Proof. Soundness part can easily be proved by induction on the complexity of formulae. For the completeness, we first find some finite perfect Kripke counter-model for each A with $\mathsf{LC} \nvdash A$, and then convert it to a perfect Kripke model with finite tree frame.

Assume $\mathsf{LC} \nvdash A$. Let $\mathsf{Sub}(A)$ be the set of sub-formulae of A. Then define

$$X := \{B, \Box B \mid B \in \mathsf{Sub}(A)\}$$

It is obvious that X is a finite adequate set. We define $\mathcal{K} = (K, <, \mathcal{R}, V)$ as follows. Take K as the set of all X-saturated sets with respect to LC, and $\leq$ is the subset relation over K. Define $\alpha \,\mathcal{R}\, \beta$ iff for all $\Box B \in X$, $\Box B \in \alpha$ implies $B \in \beta$, and also there exists some $\Box C \in \beta \setminus \alpha$. Finally define $\alpha V p$ iff $p \in \alpha$, for atomic p.
It only remains to show that $\mathcal{K}$ is a finite perfect Kripke model that refutes A. To do this, we first show by induction on $B \in X$ that $B \in \alpha$ iff $\alpha \Vdash B$, for each $\alpha \in K$. The only non-trivial case is $B = \Box C$. Let $\Box C \notin \alpha$. We must show $\alpha \nVdash \Box C$. The other direction is easier to prove and we leave it to reader. Let $\beta_0 := \{D \in X \mid \alpha \vdash \Box D\}$. If $\beta_0, \Box C \vdash C$, then, by definition of β_0, we have $\alpha \vdash \Box \beta_0$ and hence by Löb's axiom, $\alpha \vdash \Box C$, contradicting $\Box C \notin \alpha$. Hence $\beta_0, \Box C \nvdash C$ and so there exists some X-saturated set β such that $\beta \nvdash C$, $\beta \supseteq \beta_0 \cup \{\Box C\}$. Hence $\beta \in K$ and $\alpha \,\mathcal{R}\, \beta$. Then by induction hypothesis, $\beta \nVdash C$ and hence $\alpha \nVdash \Box C$.

Since $\mathsf{LC} \nvdash A$, by Lemma 4.25, there exists some X-saturated set $\alpha \in K$ such that $\alpha \nvdash A$, and hence by the above argument we have $\alpha \nVdash A$.
$\mathcal{K}$ trivially satisfies all the properties of perfect Kripke model. As a sample, we show that why $\mathcal{R} \subseteq <$ holds. Assume $\alpha \,\mathcal{R}\, \beta$ and let $B \in \alpha$. If B is boxed formula, like C, then by definition, $C \in \beta$ and hence $\beta \vdash B$ and we are done. So assume B is not a boxed formula. Then by definition of X, we have $\Box B \in X$ and by the completeness axiom in LC, we have $\alpha \vdash \Box B$ and hence by definition of $\mathcal{R}$, it is the case that $B \in \beta$. This shows $\alpha \subseteq \beta$ and hence $\alpha \leq \beta$. But α is not equal to β, because $\alpha \,\mathcal{R}\, \beta$ implies existence of some $\Box C \in \beta \setminus \alpha$. Hence $\alpha < \beta$, as desired.

Now we explain how to convert $\mathcal{K}$ to a Kripke model $\mathcal{T} := (T, <_t, \mathcal{R}_t, V_t) \nVdash A$ with a neat tree frame. Let T be the set of all finite (excluding empty sequence) sequences $\langle \alpha_1, \ldots, \alpha_n \rangle$ such that $\alpha_1 < \ldots < \alpha_n$. Let $\leq_t$ be the initial segment relation. Then define $\langle \alpha_1, \ldots, \alpha_n \rangle \; \mathcal{R}_t \; \langle \alpha_1, \ldots, \alpha_{n+k} \rangle$ iff $\alpha_{n+i} \,\mathcal{R}\, \alpha_{n+i+1}$ for some $0 \leq i < k$. Finally, define $\langle \alpha_1, \ldots, \alpha_n \rangle \; V_t \; p$, for atomic p, iff $\alpha_n \; V \; p$. Now one can prove by induction on B, that for any $\alpha = \langle \alpha_1, \ldots, \alpha_n \rangle \in T$, $\mathcal{T}, \alpha \Vdash B$ iff $\mathcal{K}, \alpha_n \Vdash B$. Hence $\mathcal{T} \nVdash A$. ❑

Since LC has finite model property, as it is expected, we can easily deduce the decidability of LC:

Corollary 4.27. LC *is decidable.*

Proof. Let A be given. Assume that n is the number of elements of X defined in the above proof. It shows us that we should only check if for all Kripke models $\mathcal{K}$ with 2^n nodes (only over atomic variables that appear in A), we have $\mathcal{K} \Vdash A$. If that was the case, we say "yes" to $\mathsf{LC} \vdash A$?, otherwise the answer is "no" to $\mathsf{LC} \vdash A$?. ❑

Relation to intuitionistic non-modal Kripke models

The usual intuitionistic non-modal Kripke models are the same Kripke models as is defined above, without the additional relation $\mathcal{R}$. Extending it to all non-modal propositions is the same as the one for modal language. It is well-known that IPC is sound and complete for non-modal Kripke models. We have the following conservativity result.

Theorem 4.28. $\mathsf{LC} + \Box\bot$ *is conservative over* IPC *in non-modal language, i.e. for any non-modal proposition A, if* $\mathsf{LC} + \Box\bot \vdash A$, *then* $\mathsf{IPC} \vdash A$.

Proof. We reason contrapositively. Assume that $\mathsf{IPC} \nvdash A$. Then by completeness, there exists some non-modal Kripke model $\mathcal{K} = (K, \leq, V) \nVdash A$. Let $\mathcal{R} := \emptyset$ and $\mathcal{K}' := (K, \leq, \mathcal{R}, V)$. It is easy to observe that $\mathcal{K}'$ is a Kripke model and for all non-modal proposition B and $u \in K$, we have $\mathcal{K}, u \Vdash B$ iff $\mathcal{K}', u \Vdash B$. Hence we have $\mathcal{K}' \nVdash A$. Now soundness theorem (Theorem 4.26) implies $\mathsf{LC} + \Box\bot \nvdash A$. ❑

5 Transforming Kripke models

Smorýnski ([Smo73a]) showed that one could simulate the behaviour of a propositional non-modal Kripke model by a first-order Kripke model of HA. Also Solovay ([Sol76]) showed that one could simulate the behaviour of a Kripke model of classical modal logic inside PA. However, the combination of these two ideas could be assumed as major obstacle towards the characterization of the provability logic of HA. In this section, we will show that one could simulate the behaviour of perfect Kripke models by first-order Kripke models of HA. This would lead us to the characterization of the Σ_1-provability logic of HA. More precisely, we will prove the following theorem.

Theorem 5.1. *Let $\mathcal{K}_0 = (K_0, \mathcal{R}_0, <_0, V_0)$ be a finite neat perfect Kripke model with tree frame and $\Gamma \subseteq \mathsf{TNNIL}^-$ be a finite set. Then there exists some arithmetical Σ_1-substitution σ and a Kripke model $\mathcal{K}_1 = (K_0, <_0, \mathfrak{M})$ such that for all $A \in \Gamma$ and $\alpha \in K_0$ we have $\mathcal{K}_0, \alpha \Vdash A$ iff $\mathcal{K}_1, \alpha \Vdash \sigma_{\mathsf{HA}}(A)$.*

Before we continue with the rather long proof of Theorem 5.1, that will take up all of this section, let us explain the outline of the proof.

First we define a recursive function F (the Solovay function) with the domain of natural numbers. $F(0)$ is defined to be some fresh node α_0. The function F, always climbs over the frame $(K_0, \mathcal{R}_0, <_0)$, but it is reluctant to do so. It only goes to some node β at some stage $n+1$ (i.e. $F(n+1) = \beta$), if $n+1$ is a witness (in some sense which would be clarified in this section) for this statement

$$\text{“}F \text{ is not going to stay in } \beta \text{ forever or } \neg\sigma_{\mathsf{HA}}(\Box\varphi_\beta)\text{”}.$$

In this definition, φ_β is the conjunction of all sentences B such that $\Box B \in \mathsf{Sub}(\Gamma)$ and $\beta \Vdash \Box B$. Here, $\mathsf{Sub}(\Gamma)$ is the set of sub-formulae of some formula in Γ. The most interesting (and difficult part to prove as well) property of the function F is that this function actually (in the standard model of arithmetic $\mathbb{N}$) does not climb over tree at all, i.e. the function F is constant, $\mathbb{N} \models \forall x F(x) = \alpha_0$. In contrast with the classical case, proving this fact for the intuitionistic case is rather complicated.

Let $L \succcurlyeq \alpha$ denote “the function F would go above α or remain equal to α”. Then we define the substitution $\sigma(p) := \bigvee_{\alpha \Vdash p} L \succcurlyeq \alpha$. Then we define the I-frame $\mathcal{I} = (K_0, <_0, T)$, where T_α is defined to be PA plus the following statement: “The limit of the function F is α”. Finally, with the aid of Theorem 5.29 we find the desired Kripke model $\mathcal{K}_1$, by assigning an appropriate classical model of T_α to the node α. We will show that $T_\alpha \vdash \sigma_{\mathsf{HA}}(\Box\varphi_\alpha)$ (Corollary 5.23) and also $T_\alpha \vdash \sigma_{\mathsf{HA}}(\neg\Box B)$ for any $\Box B \in \mathsf{Sub}(\Gamma)$ and $\alpha \nVdash \Box B$ (Theorem 5.15). In this way, we can simulate the role of modal operator $\Box$ in the first-order Kripke model $\mathcal{K}_1$.

Notation. In the rest of this section, we fix the Kripke model $\mathcal{K}_0 = (K_0, \mathcal{R}_0, <_0, V_0)$ and the set $S := \{B \in \mathsf{Sub}(\Gamma) \mid B \in \mathsf{TNNIL}\}$. We also assume that $\alpha_0 \notin K_0$ and define

$$\mathcal{R} := \mathcal{R}_0 \cup \{(\alpha_0, \alpha) \mid \alpha \in K_0\} \qquad < := <_0 \cup \{(\alpha_0, \alpha) \mid \alpha \in K_0\} \qquad K := K_0 \cup \{\alpha_0\}$$

In other words, we add α_0 in beneath of all nodes of $\mathcal{K}_0$. Finally we define $\mathcal{K} := (K, \mathcal{R}, <, V_0)$.

5.1 Definition of the Solovay function

Solovay used some special recursive function (here we call it the *Solovay function*) to prove the completeness of GL (The Gödel-Löb logic) for arithmetical interpretations in PA (See [Sol76]). The Solovay function in [Sol76], is a function $G : \mathbb{N} \longrightarrow X$, in which X is a finite partially ordered set ordered by $\preccurlyeq$. The recursive definition of G is such that G climbs over X, i.e. $G(x) \preccurlyeq G(x+1)$ and moreover, it goes to some new node iff there exists a witness that G would not remain there. More precisely, $G(x+1) \neq G(x)$ iff $x+1$ is the code of a proof (in PA) for the fact that the limit of the function G is not $G(x+1)$. Although it is true (in the standard model) that G will not climb over X (i.e. G is a constant function), PA can't prove this fact. In this subsection, we define a similar recursive function (we call it F) for the proof of our main theorem (Theorem 5.1), and state and prove some of its properties.

For technical reasons, we first define the set of all codes of sequences $z = \langle F(0), \dots, F(x)\rangle$ by an arithmetical formula $\theta(z)$, and then define $\phi_\theta(x, y) := \exists z(\mathsf{lth}(z) = x + 1 \wedge \theta(z) \wedge \hat{z} = y)$ as the graph of a function F_θ and finally, let $F := F_\theta$. It is clear that we can also define $\theta_F(z)$ from the function F in the following way.

$$\theta_F(z) := \exists x(z = \langle F(0), \dots, F(x)\rangle) \text{ or equivalently } \theta_F(z) := \forall\, x < \mathsf{lth}(z)(F(x) = (z)_x)$$

To be able to speak about $\mathcal{K}$ inside HA, we need some conventions. Suppose that $K = \{\alpha_0, \alpha_1, \dots, \alpha_k\}$. Hence for each $\alpha \in K$, there exists a unique index $0 \leq i \leq k$ such that $\alpha = \alpha_i$. We define $\overline{\alpha}$ to be $\overline{i}$ ($\bar{n}$ is n-th numeral in the language of arithmetic, i.e. $\overline{i} := S^i(0)$). We may simply use α instead of $\overline{\alpha}$, if no confusion is likely. The following notations for arbitrary terms t and s in the language of arithmetic will be used later.

- $\overline{K}(t) := \bigvee_{\alpha \in K}(t = \overline{\alpha})$,
- $t \prec s := \bigvee_{\alpha \lneqq \beta}(t = \overline{\alpha} \wedge s = \overline{\beta})$, $t \preccurlyeq s := \bigvee_{\alpha \leq \beta}(t = \overline{\alpha} \wedge s = \overline{\beta})$,
- $t \,\overline{\mathcal{R}}\, s := \bigvee_{\alpha \mathcal{R} \beta}(t = \overline{\alpha} \wedge s = \overline{\beta})$,
- $\varphi_{\alpha} := \bigwedge_{B \in \mathsf{Sub}(\Gamma), B \in \mathsf{TNNIL}, \alpha \Vdash \Box B} B$.

In the following definition, $L_\theta = y$ as the arithmetical formula equivalent to "The limit of the function F_θ is equal to y". Similarly, define $\alpha \prec L_\theta$ and so on.

Definition 5.2. *Let $\theta(z)$ be a Σ_1-formula in the language of arithmetic. Then*

- *$L_\theta = y$ is a shorthand for $\exists u \forall z(\theta(u * z) \rightarrow \hat{z} = y)$ in which $\hat{z}$ is the final element of the sequence with the code z,*
- *For each $\alpha \in K$, $\alpha \preccurlyeq L_\theta$, $\alpha \prec L_\theta$ and $\alpha \,\mathcal{R}\, L_\theta$ are shorthands for $\bigvee_{\alpha \leq \beta} \exists x(\theta(x) \wedge \hat{x} = \beta)$, $\bigvee_{\alpha \lneqq \beta} \exists x(\theta(x) \wedge \hat{x} = \beta)$ and $\bigvee_{\alpha \mathcal{R} \beta} \exists x(\theta(x) \wedge \hat{x} = \beta)$, respectively,*
- *The arithmetical substitution σ is defined on propositional variable p by*

$$\sigma(p) := \bigvee_{\beta \Vdash p} \beta \preccurlyeq L_\theta.$$

and finally, we extend σ to all propositions by interpreting $\Box$ as provability in HA , i.e., $\sigma_\theta := \sigma_{\mathsf{HA}}$, in which σ_{HA} is defined from σ as in Definition 2.1,

- *Let g be a recursive function with $\theta_g(z)$ as the formula $\exists x(z = \langle g(0), \dots, g(x)\rangle)$. We define $L_g = y$, $L_g \succ \alpha$, $L_g \succeq \alpha$, $\alpha \,\mathcal{R}\, L_g$ and σ_g to be $L_{\theta_g} = y$, $L_{\theta_g} \succ \alpha$, $L_{\theta_g} \succeq \alpha$, $\alpha \,\mathcal{R}\, L_{\theta_g}$ and σ_{θ_g}, respectively.*

Following Berarducci ([Ber90]), we define a primitive recursive function as follows:

$$r_\theta(\bar{\alpha}, x) = \min\left(\{k \mid \exists u \leq x \mathsf{Proof}_{\mathsf{PA}_k}(u, \ulcorner \neg(L = \alpha \wedge \Box \sigma_\theta(\varphi_\alpha))\urcorner)\} \cup \{x + 1\}\right)$$

Note that $r_\theta(x, y)$ depends also on a Σ_1-formula which is appeared in the subscript of σ. We also should note that $L = \alpha$ is defined in reference to θ as well. We may omit subscripts of the interpretation σ_θ and the function r_θ when no confusion is likely.

A variant of this function was first appeared in [Ber90], to define Solovay functions for characterizing interpretability logic of PA. It is easy to observe that $r(\alpha, x)$ is always equal or less than $x+1$, and $r(\alpha, x) \leq x$ iff

$$\exists y \leq x \mathsf{Proof}_{\mathsf{PA}}(y, \ulcorner \neg(L = \alpha \wedge \Box \sigma_F(\varphi_\alpha)) \urcorner)$$

Now we are in a position to define the Solovay-like function for $\mathcal{K}$. Informally speaking, $F : \mathbb{N} \to K$ is defined in such a way that fulfils the following conditions. $F(0) := \alpha_0$, and

$$F(x+1) := \begin{cases} \beta & \text{if } (x+1)_0 = \langle 1, \beta \rangle, F(x) R \beta \text{ and } r_F(\beta, x+1) \leq x+1, \\ \gamma & \text{if } (x+1)_0 = \langle 2, \gamma \rangle, \neg F(x) R \gamma \text{ and } F(x) \leq \gamma \text{ and} \\ & \quad r_F(\gamma, x+1) < r_F(F(x), x+1) \text{ and } \quad F(r_F(\gamma, x+1)) \, \mathcal{R} \, \gamma, \\ F(x) & \text{otherwise.} \end{cases} \tag{5.1}$$

As it is clear from the definition, F is used in its own definition, i.e. we are in a loop. This will be overcome by the Diagonalization lemma. To be able to define F, we first define $\theta(z)$ and then define $F(x) = y$ (the graph of the function F) as

$$\exists z (\mathsf{lth}(z) = x + 1 \wedge \theta(z) \wedge (z)_x = y)$$

By Diagonalization lemma (Lemma 3.5), we find a Δ_0 formula $\theta(y)$ such that

$$\mathsf{HA}_0 \vdash \theta(y) \leftrightarrow (\mathsf{lth}(y) \geq 1 \wedge (y)_0 = \overline{\alpha_0} \wedge \forall x < \mathsf{lth}(y)(x \neq 0 \to \chi(x, y))) \tag{5.2}$$

in which $\chi(x, y)$ is defined as disjunction of the following three formulae:

$$\chi_1 := \bigvee_{\beta \in K} [(x)_0 = \langle 1, \bar{\beta} \rangle \wedge (y)_x = \bar{\beta} \wedge (y)_{x \dot{-} 1} \, \bar{\mathcal{R}} \, \bar{\beta} \wedge r(\bar{\beta}, x) \leq x]$$

$$\chi_2 := \bigvee_{\beta \in K} [(x)_0 = \langle 2, \bar{\beta} \rangle \wedge (y)_x = \bar{\beta} \wedge \neg (y)_{x \dot{-} 1} \, \bar{\mathcal{R}} \, \bar{\beta} \wedge (y)_{x \dot{-} 1} \preccurlyeq \bar{\beta} \wedge r((y)_x, x) < r((y)_{x \dot{-} 1}, x))$$
$$\wedge (y)_{r((y)_x, x)} \, \bar{\mathcal{R}} \, (y)_x)]$$

$$\chi_3 := [(y)_x = (y)_{x \dot{-} 1}] \wedge \bigwedge_{\beta \in K} \neg [(x)_0 = \langle 1, \bar{\beta} \rangle \wedge (y)_{x \dot{-} 1} \, \bar{\mathcal{R}} \, \bar{\beta} \wedge r(\bar{\beta}, x) \leq x] \wedge$$
$$\bigwedge_{\beta \in K} \neg [(y)_{x \dot{-} 1} \preccurlyeq \bar{\beta} \wedge \neg (y)_{x \dot{-} 1} \, \bar{\mathcal{R}} \, \bar{\beta} \wedge (x)_0 = \langle 2, \bar{\beta} \rangle \wedge r(\bar{\beta}, x) < r((y)_{x \dot{-} 1}, x) \wedge (y)_{r(\bar{\beta}, x)} \, \bar{\mathcal{R}} \, \bar{\beta})]$$

In the above formulae, $r(x, y)$ is $r_\theta(x, y)$. Now we show that a provably total recursive function F can be defined from $\theta(y)$.

Lemma 5.3. *The formula θ is Δ_0 and*

1. $\mathsf{HA}_0 \vdash (\mathsf{lth}(y_1) \neq 0 \wedge \theta(y_1 * y_2)) \to \theta(y_1)$,
2. $\mathsf{HA}_0 \vdash (\theta(y_1) \wedge \theta(y_2) \wedge \mathsf{lth}(y_1) = \mathsf{lth}(y_2)) \to y_1 = y_2$,
3. $\mathsf{HA}_0 \vdash \forall x \exists y (\mathsf{lth}(y) = x + 1 \wedge \theta(y))$.

Proof. It is not difficult to observe that the first item holds by definition of θ in eq. (5.2). To prove the other items, it is enough to show $\mathsf{HA}_0 \vdash \forall x \exists ! y (\mathsf{lth}(y) = x + 1 \wedge \theta(y))$, in which $!\exists$, as usual, is the *uniqueness* existential quantifier. This can be simply done by induction on x. ❑

Now, let us define $\phi(x, y) := \exists z (\theta(z) \wedge \mathsf{lth}(z) = x + 1 \wedge \hat{z} = y)$. Note that $\phi(x, y)$ is actually a Δ_0 formula. The reason is the following. we can bound existential quantifier by the primitive recursive function $h(z)$ with the following primitive recursive definition:

- $h(0) := \langle k\rangle$, in which k is the number of nodes of Kripke model,
- $h(z+1) := h(z) * \langle k\rangle$.

Hence $\mathsf{HA}_0 \vdash \phi(x,y) \leftrightarrow \exists z \leq h(z)[\mathsf{lth}(z) = x+1 \wedge (z)_{x+1} = y \wedge \phi(z)]$.

Notation 5.4. The above lemma (Lemma 5.3) says that $\phi(x,y)$ is the graph of a Δ_0- function F. In the rest of the paper, we use F as a function symbol with the graph $\phi(x,y)$. We use σ and L instead of σ_θ and L_θ, respectively. For simplicity of notations, when we work in the first-order language of arithmetic, instead of $\sigma_{\mathsf{HA}}(B)$, we may use the notation B. For instance assume that p is an atomic variable in the propositional language. When we write down the formula $\mathsf{HA} \vdash \Box(\Box p \to p) \to \Box p$, we actually mean $\mathsf{HA} \vdash \sigma_{\mathsf{HA}}(\Box(\Box p \to p) \to \Box p)$. This abuse of notations, wipes out many unimportant symbols from the rest of Section 5.

One can observe that the function F fulfils the recursive conditions of eq. (5.1).

5.2 Elementary properties of the Solovay function

In this part, we will see some elementary properties of the function F.

Lemma 5.5. *The function F has the following properties:*

1. $\mathsf{HA}_0 \vdash \forall x, y(F(x) \preccurlyeq F(x+y))$,
2. *For any $\alpha \in K$,* $\mathsf{PA} \vdash \exists x F(x) = \alpha \to \bigvee_{\alpha \leq \beta} L = \beta$,
3. *For any $\alpha \in K$,* $\mathsf{PA} \vdash \alpha \prec L \leftrightarrow \bigvee_{\alpha < \beta} L = \beta$ *and* $\mathsf{PA} \vdash \alpha \,\bar{\mathcal{R}}\, L \leftrightarrow \bigvee_{\alpha \mathcal{R} \beta} L = \beta$.

Proof. 1. By recursive definition of F, $\mathsf{HA}_0 \vdash F(x) \preccurlyeq F(x+1)$. Let $A(y) := \bar{F}(x) \preccurlyeq \bar{F}(x+y)$ and use induction on y in $A(y)$.

2. We prove this fact by induction (in meta-language) on the tree $(K, \lneqq)$ with reverse order. Suppose that for all $\beta \gneqq \alpha$, we have $\mathsf{PA} \vdash \exists x F(x) = \beta \to \bigvee_{\beta \leq \gamma} L = \gamma$. Then

$$\mathsf{PA} \vdash F(x) = \alpha \to (\forall y \geq x F(y) = \alpha \vee \exists y \geq x F(y) \neq \alpha)$$

By part 1 and definition of $L = \alpha$, we get $\mathsf{PA} \vdash F(x) = \alpha \to (L = \alpha \vee \exists y \geq x(\alpha \prec F(y)))$. Now induction hypothesis implies $\mathsf{PA} \vdash F(x) = \alpha \to \bigvee_{\beta \geq \alpha} L = \beta$.

3. Proof of this part is an immediate consequence of part 2 and perfectness of $\mathcal{K}$.

❑

Lemma 5.6. *For any $\alpha, \beta \in K$ with $\alpha \,\mathcal{R}\, \beta$, $\mathsf{HA}_0 \vdash L = \alpha \to \neg\Box^+\neg(L = \beta \wedge \Box\varphi_\beta)$.*

Proof. We argue inside HA_0. Assume $L = \alpha$ and $\mathsf{Proof}_{\mathsf{PA}}(x, \ulcorner\neg(L = \beta \wedge \Box\varphi_\beta)\urcorner)$. Let $y > x$ such that $(y+1)_0 = \langle 2, \beta\rangle$. Then because $L = \alpha$, we have $F(y) = \alpha$. On the other hand, by recursive definition of F, $F(y+1) := \beta$, a contradiction. ❑

Lemma 5.7. *For any $\delta, \alpha, \beta \in K$ with $\delta \,\mathcal{R}\, \alpha \leq \beta$, $\mathsf{HA}_0 + L = \delta \vdash (L = \alpha \wedge \Box\varphi_\alpha) \rhd (L = \beta \wedge \Box\varphi_\beta)$.*

Proof. If $\alpha \,\mathcal{R}\, \beta$, by Lemma 5.6, $\mathsf{HA}_0 \vdash L = \alpha \to \neg\Box^+\neg(L = \beta \wedge \Box\varphi_\beta)$. So

$$\mathbb{N} \models \Box^+(L = \alpha \to \neg\Box^+\neg(L = \beta \wedge \Box\varphi_\beta))$$

and hence by Lemma 3.6 (Σ_1-completeness of HA_0), we can deduce $\mathsf{HA}_0 \vdash L = \alpha \rhd (L = \beta \wedge \Box\varphi_\beta)$. So assume $\alpha \,\not{\mathcal{R}}\, \beta$ and $\alpha \neq \beta$. By definition of $A \rhd B$, we must show

$$\mathsf{HA}_0 + L = \delta \vdash \forall x \Box^+[(L = \alpha \wedge \Box\varphi_\alpha) \to \neg\Box^+_x\neg(L = \beta \wedge \Box\varphi_\beta)].$$

We work inside HA_0. Assume $L = \delta$ and fix some large enough x such that $F(x) = \delta$. Then for each $u \leq x$, we have $F(u)\,\mathcal{R}\,\beta$. Now work in the scope of $\Box^+$. By Σ-completeness of PA, we have $\forall u \leq xF(u)\,\mathcal{R}\,\beta$. Assume $L = \alpha$, $\Box\varphi_{\alpha}$ and $\Box^+_x\neg(L = \beta \wedge \Box\varphi_{\beta})$. We should deduce $\bot$. By $\Box^+_x\neg(L = \beta \wedge \Box\varphi_{\beta})$, for sufficiently large y (larger than $\langle 2, \beta\rangle * z$, in which z is a proof code in PA_x for $\neg(L = \beta \wedge \Box\varphi_{\beta})$), we have $r(\beta, y) \leq x$. If $r(\alpha, y) \leq r(\beta, y)$, then $\Box^+_x\neg(L = \alpha \wedge \Box\varphi_{\alpha})$, and hence by Lemma 3.7, we have $\neg(L = \alpha \wedge \Box\varphi_{\alpha})$, a contradiction. If $r(\alpha, y) > r(\beta, y)$, since $r(\beta, y) \leq x$, then $F(r(\beta, y))\,\mathcal{R}\,\beta$. So by recursive definition of F, there exists some $z \geq y$ such that $F(z) = \beta$, contradicting $L = \alpha$. ❑

5.3 Deciding the boxed formulas

In this subsection, we will show that $\mathsf{HA} + L = \alpha + \Box\varphi_{\alpha}$ can decide boxed propositions in $\mathsf{Sub}(\Gamma)$. More precisely, for all $\Box B \in \mathsf{Sub}(\Gamma)$ and $\alpha \in K$,

$$\begin{cases} \mathsf{HA} + L = \alpha + \Box\varphi_{\alpha} \vdash \Box B & \text{if } \alpha \Vdash \Box B \\ \mathsf{HA} + L = \alpha + \Box\varphi_{\alpha} \vdash \neg\Box B & \text{if } \alpha \nVdash \Box B \end{cases}$$

Note that by definition of φ_{α}, if $\alpha \Vdash \Box B$, then B is a conjunct of φ_{α}. Hence in case $\alpha \Vdash \Box B$, we obviously have $\mathsf{HA} + \Box\varphi_{\alpha} \vdash \Box B$. Moreover we will show in Section 5.4 (Corollary 5.23) that $\mathsf{HA} \vdash L = \alpha \to \Box\varphi_{\alpha}$ for $\alpha \in K_0$, and then the following improvement of the above equation holds:

$$\begin{cases} \mathsf{HA} + L = \alpha \vdash \Box B & \text{if } \alpha \Vdash \Box B \\ \mathsf{HA} + L = \alpha \vdash \neg\Box B & \text{if } \alpha \nVdash \Box B \end{cases}$$

Lemma 5.8. *Let $B \in \mathsf{Sub}(\Gamma)$ be such that all occurrences of $\to$ in B are in the scope of some $\Box$ ($B \in \mathsf{NOI}$), $\alpha \in K$ and $\alpha \Vdash B$. Then $\mathsf{HA}_0 \vdash (L = \alpha \wedge \Box\varphi_{\alpha}) \to B$. Moreover, this argument is formalizable and provable in HA_0, i.e. $\mathsf{HA}_0 \vdash \Box_0((L = \alpha \wedge \Box\varphi_{\alpha}) \to B)$.*

Proof. One can prove $\mathsf{HA}_0 \vdash (L = \alpha \wedge \Box\varphi_{\alpha}) \to B$, by induction on the complexity of B. Then by Lemma 3.6, we derive its formalized form in HA_0. ❑

Notation 5.9. *We say that $\alpha \nVdash_{\mathsf{max}} A$ if $\alpha \nVdash A$ and for all $\beta \gneqq \alpha$ we have $\beta \Vdash A$.*

We have the following observations:

- $\alpha \nVdash_{\mathsf{max}} B \to C$ iff "$\alpha \Vdash B$ and $\alpha \nVdash_{\mathsf{max}} C$",
- $\alpha \nVdash_{\mathsf{max}} B \vee C$ implies "$\alpha \nVdash B$ and $\alpha \nVdash C$",
- $\alpha \nVdash_{\mathsf{max}} B \wedge C$ iff "$\alpha \nVdash_{\mathsf{max}} B$ or $\alpha \nVdash_{\mathsf{max}} C$".

Let A be a TNNIL-formula such that $\alpha \nVdash_{\mathsf{max}} A$. In Lemma 5.12 and Lemma 5.13, we need to put $\Box_x$ before all occurrences of subformulas B in the right of $\to$, when it is not the case that $\alpha \nVdash_{\mathsf{max}} B$. This is the content of the following definition.

Definition 5.10. Let A be a modal proposition, $\alpha \in K$ and x be a variable. We define the first-order sentence $d(A, \alpha, x)$, by induction on A. If this is not the case that $\alpha \nVdash_{\mathsf{max}} A$, then we define $d(A, \alpha, x) := \Box_x\sigma_{\mathsf{HA}}(A)$, and if $\alpha \nVdash_{\mathsf{max}} A$, we define the formula $d(A, \alpha, x)$ by cases:

- A is atomic or boxed. $d(A, \alpha, x) := \sigma_{\mathsf{HA}}(A)$,
- $A = B \to C$. Define $d(A, \alpha, x)$ by cases. If $B \notin \mathsf{NOI}$, then let $d(A, \alpha, x) := \sigma_{\mathsf{HA}}(A)$, otherwise let $d(A, \alpha, x) := \sigma_{\mathsf{HA}}(B) \to d(C, \alpha, x)$,
- $A = B \wedge C$. If $\alpha \nVdash_{\mathsf{max}} B$ then $d(A, \alpha, x) := d(B, \alpha, x)$, else $d(A, \alpha, x) := d(C, \alpha, x)$,

- $A = B \vee C$. $d(A, \alpha, x) := d(B, \alpha, x) \vee d(C, \alpha, x)$.

In the following lemma, we use definition of $\sigma_{\imath}(A, x)$ from Section 3.3:

Lemma 5.11. *Let A be a modal proposition, $\alpha \in K$ such that $\alpha \nVdash_{\max} A$. Then*

$$\mathsf{HA}_0 \vdash \sigma_{\imath}(A, x) \to d(A, \alpha, x)$$

Proof. Use induction on A. ❑

Lemma 5.12. *Let A be a modal proposition. Then there exists some provably (in HA) total recursive function g_A such that for any $\alpha \in K$ with $\alpha \nVdash_{\max} A$ we have*

$$\mathsf{HA} \vdash \Box_x A \to \Box_{g_A(x)} d(A, \alpha, g_A(x))$$

Proof. Use Lemma 3.18 and Lemma 5.11. ❑

Let $\mathsf{HA} \vdash A$, for arbitrary A in the language of arithmetic. Then by the compactness theorem, one could deduce that $\mathsf{HA}_n \vdash A$ for some $n \in \omega$. In the following definition of the n_1 and n_2, we make use of this fact. Define $m \in \omega$ as the maximum of the following n_i's:

- n_1. By Lemma 5.12 and the compactness theorem, we can find some n_1 such that for each $B \in \mathsf{Sub}(\Gamma)$, g_B is provably total in HA_{n_1}.
- n_2. For each $\alpha \in K$ and $B \in \mathsf{Sub}(\Gamma)$ such that $\alpha \nVdash_{\max} B$, by Lemma 5.12 and the compactness theorem, there exists some n such that HA_n proves the desired sentence of the Lemma. Let n_2 be the maximum of such n.
- n_3. By Lemma 3.12, for each $\alpha \in K$, there exists some provably (in HA) total function h_α, such that $h_\alpha(x) \geq x$ and $\mathsf{HA} \vdash \Box_{h_\alpha(x)}(\Box_x \neg(L = \alpha \wedge \Box\varphi_\alpha) \to \neg(L = \alpha \wedge \Box\varphi_\alpha))$. Hence by the compactness theorem, there exists some $n_\alpha \in \omega$ such that h_α is provably total in HA_{n_α} and

  $$\mathsf{HA}_{n_\alpha} \vdash \Box_{h_\alpha(x)}(\Box_x \neg(L = \alpha \wedge \Box\varphi_\alpha) \to \neg(L = \alpha \wedge \Box\varphi_\alpha))$$

  Let $n_3 := \max\{n_\alpha | \alpha \in K\}$.

Then define $\hat{g}_B(x)$ as the maximum of $g_B(x)$, m and x. Assume some $B \in \mathsf{Sub}(\Gamma)$. We define the provably (in HA_m) total recursive function f_B, by induction on the complexity of B:

$$f_B(x) := \begin{cases} \max(X) & \text{if } X = \{h_\alpha(f_C(\hat{g}_B(x))) \mid C \in \mathsf{Sub}(B), C \neq B, \alpha \in K\} \neq \emptyset \\ \hat{g}_B(x) & \text{else} \end{cases}$$

where h_α is as we stated in definition of n_3. From the above definition, one can observe that for each atomic $C \in \mathsf{Sub}(\Gamma)$, the set X is empty. Hence we have $f_C(x) = x$. Since each non-atomic formula B has some atomic sub-formula C, one can deduce that $f_B(x) \geq \hat{g}_B(x) \geq x, m$. Moreover, all of the above functions are provably total in HA_m.

Lemma 5.13. *Let $B, E \in \mathsf{Sub}(\Gamma) \cap \mathsf{TNNIL}$ and $\beta \in K$, such that $\beta \nVdash_{\max} B$, $\beta \nVdash_{\max} E$ and $B \in \mathsf{Sub}(E)$. Then*

$$\mathsf{HA}_m \vdash [F(f_E(x))\, \mathcal{R}\, \beta \wedge \Box_x E] \to \Box_{f_E(x)}\big((L = \beta \wedge \Box\varphi_\beta) \to \neg d(B, \beta, \hat{g}_E(x))\big) \tag{5.3}$$

Proof. We prove eq. (5.3) by induction on the complexity of B. As induction hypothesis, assume that for any sub-formula C of B ($C \neq B$) and any $E' \in \mathsf{Sub}(\Gamma) \cap \mathsf{TNNIL}$ and $\gamma \in K$, such that $C \in \mathsf{Sub}(E')$ and $\gamma \nVdash_{\max} C, E'$, we have

$$\mathsf{HA}_m \vdash \big(F(f_{E'}(x))\, \mathcal{R}\, \gamma \wedge \Box_x E'\big) \to \Box_{f_{E'}(x)}\big((L = \gamma \wedge \Box\varphi_\gamma) \to \neg d(C, \gamma, \hat{g}_{E'}(x))\big)$$

We consider different cases.

- B is atomic. Then $d(B,\beta,\hat{g}_{_B}(x))=\sigma(B)$ and the desired result holds by definition of the substitution σ and $\beta\nVdash B$ and also by Lemma 5.5 item 1,

- $B=\Box C$. Then $d(B,\beta,\hat{g}_{_B}(x))=\sigma_{\mathsf{HA}}(B)=\Box\sigma_{\mathsf{HA}}(C)$. Since $\beta\nVdash_{\max}\Box C$, there exists some γ such that $\beta\,\mathcal{R}\,\gamma\nVdash_{\max} C$. Then, by induction hypothesis,

$$\mathsf{HA}_m\vdash(F\,(f_{_C}\,(x))\;\mathcal{R}\,\gamma\wedge\Box_x C)\to\Box_{f_C(x)}\left((L=\gamma\wedge\Box\varphi_{_\gamma})\to\neg d(C,\gamma,\hat{g}_{_C}(x))\right)$$

By Lemma 5.12, we have $\mathsf{HA}_m\vdash\Box_x C\to\Box_{f_C(x)}d(C,\gamma,\hat{g}_{_C}(x))$. Hence

$$\mathsf{HA}_m\vdash(L=\beta\wedge\Box C)\to\Box\neg\left(L=\gamma\wedge\Box\varphi_{_\gamma}\right)$$

By Lemma 5.6, we have $\mathsf{HA}_m\vdash\neg\,(L=\beta\wedge\Box C)$. Hence by Lemma 3.6, $\mathsf{HA}_0\vdash\Box_m\neg\,(L=\beta\wedge\Box C)$. Since $f_{_B}(x)\geq m$, we have $\mathsf{HA}_m\vdash\Box_{f_B(x)}\left((L=\beta\wedge\Box\varphi_{_\beta})\to\neg d(B,\beta,\hat{g}_{_B}(x))\right)$, which implies eq. (5.3).

- $B=C\to D$. In this case $\beta\Vdash C\in\mathsf{NOI}$, $\beta\nVdash_{\max} D$ and $d(B,\beta,\hat{g}_{_B}(x))=\sigma_{\mathsf{HA}}(C)\to d(D,\beta,\hat{g}_{_B}(x))$. Hence, by induction hypothesis,

$$\mathsf{HA}_m\vdash(F\,(f_{_E}\,(x))\;\mathcal{R}\,\beta\wedge\Box_x E)\to\Box_{f_E(x)}\left((L=\beta\wedge\Box\varphi_{_\beta})\to\neg d(D,\beta,\hat{g}_{_E}(x))\right)$$

Then by Lemma 5.8,

$$\mathsf{HA}_m\vdash(F\,(f_{_E}\,(x))\;\mathcal{R}\,\beta\wedge\Box_x E)\to\Box_{f_E(x)}\left((L=\beta\wedge\Box\varphi_{_\beta})\to\neg(C\to d(D,\beta,\hat{g}_{_E}(x))\right)$$

- $B=C\wedge D$. Since $\beta\nVdash_{\max} B$, either $\beta\nVdash_{\max} C$ or $\beta\nVdash_{\max} D$ holds. We only treat the case that $\beta\nVdash_{\max} C$. The other case is similar. Assume that $\beta\nVdash_{\max} C$. Then by definition, $d(B,\beta,y)=d(C,\beta,y)$. Now the induction hypothesis for C, directly implies the desired result, i.e. eq. (5.3).

- $B=C\vee D$. This case is the interesting one. We have 4 sub-cases: (1) $\beta\nVdash_{\max} C$ and $\beta\nVdash_{\max} D$, (2) not $\beta\nVdash_{\max} C$ and $\beta\nVdash_{\max} D$, (3) $\beta\nVdash_{\max} C$ and not $\beta\nVdash_{\max} D$, (4) not $\beta\nVdash_{\max} C$ and not $\beta\nVdash_{\max} D$. We only treat the case (3) here. Other cases can be treated similarly. Assume that the case (3) occurs. By definition, $d(B,\beta,\hat{g}_{_E}(x))=d(C,\beta,\hat{g}_{_E}(x))\vee\Box_{\hat{g}_E(x)}D$. From the induction hypothesis for C,

$$\mathsf{HA}_m\vdash(F(f_{_E}(x))\,\mathcal{R}\,\beta\wedge\Box_x E)\to\Box_{f_E(x)}\left((L=\beta\wedge\Box\varphi_{_\beta})\to\neg d(C,\beta,\hat{g}_{_E}(x))\right)\tag{5.4}$$

So it is enough to show that

$$\mathsf{HA}_m\vdash(F(f_{_E}(x))\,\mathcal{R}\,\beta\wedge\Box_x E)\to\Box_{f_E(x)}\Big((L=\beta\wedge\Box\varphi_{_\beta})\to\neg\Box_{\hat{g}_E(x)}D\Big)\tag{5.5}$$

Since $\beta\nVdash D$ and not $\beta\nVdash_{\max} D$, there exists some $\gamma\gneqq\beta$ such that $\gamma\nVdash_{\max} D$. If $\beta\,\mathcal{R}\,\gamma$, then we can repeat the reasoning as in the case $B=\Box C$. So assume that $\beta\not\mathcal{R}\gamma$. By the induction hypothesis for D and γ, we have

$$\mathsf{HA}_m\vdash(F(f_{_D}(x))\,\mathcal{R}\,\gamma\wedge\Box_x D)\to\Box_{f_D(x)}\left((L=\gamma\wedge\Box\varphi_{_\gamma})\to\neg d(D,\gamma,\hat{g}_{_D}(x))\right)$$

On the other hand, by Lemma 5.12, we have

$$\mathsf{HA}_m\vdash\Box_x D\to\Box_{f_D(x)}d(D,\gamma,\hat{g}_{_D}(x))$$

Hence

$$\mathsf{HA}_m\vdash(F(f_{_D}(x))\,\mathcal{R}\,\gamma\wedge\Box_x D)\to\Box_{f_D(x)}\neg(L=\gamma\wedge\Box\varphi_{_\gamma})\tag{5.6}$$

We argue inside HA_m. Assume $F(f_{_E}(x))\,\mathcal{R}\,\beta$ and $\Box_x E$. Since $f_{_E}(x) \geq f_{_D}(\hat{g}_{_E}(x))$, by the assumption of $F(f_{_E}(x))\,\mathcal{R}\,\beta$, we have $F(f_{_D}(\hat{g}_{_E}(x)))\,\mathcal{R}\,\gamma$, and by Lemma 3.6, we get $\Box_m(F(f_{_D}(\hat{g}_{_E}(x)))\,\mathcal{R}\,\gamma)$. Hence if we replace $\hat{g}_{_E}(x)$ for x in eq. (5.6), we may deduce

$$\Box_m\left(\Box_{\hat{g}_{_E}(x)} D \to \Box_{f_{_D}(\hat{g}_{_E}(x))}\neg(L = \gamma \wedge \Box\varphi_{_\gamma})\right) \tag{5.7}$$

Now we work inside $\Box_{f_{_E}(x)}$. We have $F(f_{_E}(x))\,\mathcal{R}\,\beta$. Assume $\Box_{\hat{g}_{_E}(x)} D$ and $L = \beta$ and $\Box\varphi_{_\beta}$. We should deduce $\bot$. From $\Box_{\hat{g}_{_E}(x)} D$ and eq. (5.7), we have $\Box_{t(x)}\neg(L = \gamma \wedge \Box\varphi_{_\gamma})$, in which $t(x) := f_{_D}(\hat{g}_{_E}(x))$. So there exists some y_1 such that $\mathsf{Proof}_{\mathsf{HA}_{t(x)}}(y_1, \ulcorner\neg(L = \gamma \wedge \varphi_{_\gamma})\urcorner)$. Also by $L = \beta$, there exists some $y_2 \geq y_1$ such that $\forall z \geq y_2 F(z) = \beta$. Let some y greater than $\langle 2, \gamma\rangle^* y_2$ and $t(x)$. If $r(\beta, y+1) \leq t(x)$, then $\Box_{t(x)}\neg(L = \beta \wedge \Box\varphi_{_\beta})$. Now, since $f_{_E}(x) \geq h_\beta(t(x))$ and we are working in $\Box_{f_{_E}(x)}$, by Lemma 3.12, we have $\neg(L = \beta \wedge \Box\varphi_{_\beta})$ and hence $\bot$. If $t(x) < r(\beta, y+1)$, since $r(\gamma, y+1) \leq t(x)$, by recursive definition of F, then $F(y+1) = \gamma$, which contradicts with $L = \beta$.

❑

Corollary 5.14. *For each $B \in \mathsf{Sub}(\Gamma) \cap \mathsf{TNNIL}$ and $\beta \in K$ such that $\beta \nVdash_{\mathsf{max}} B$,*

$$\mathsf{HA}_m \vdash (F(f_{_B}(x))\,\mathcal{R}\,\beta \wedge \Box_x B) \to \Box_{f_{_B}(x)}\neg\left(L = \beta \wedge \Box\varphi_{_\beta}\right)$$

Proof. Use Lemma 5.13 for $E = B$. Then by eq. (5.3) and Lemma 5.12, one can deduce the desired result. ❑

Theorem 5.15. *For each $B \in \mathsf{Sub}(\Gamma) \cap \mathsf{TNNIL}$ and $\alpha \in K$ such that $\alpha \nVdash \Box B$,*

$$\mathsf{HA} \vdash L = \alpha \to \neg\Box B$$

Proof. From $\alpha \nVdash \Box B$, we conclude that there exists some $\beta \in K$ such that $\alpha\,\mathcal{R}\,\beta$ and $\beta \nVdash_{\mathsf{max}} B$. Now Corollary 5.14 implies $\mathsf{HA} \vdash (L = \alpha \wedge \Box B) \to \Box\neg(L = \beta \wedge \Box\varphi_{_\beta})$. On the other hand, by Lemma 5.6, $\mathsf{HA} \vdash L = \alpha \to \neg\Box\neg(L = \beta \wedge \Box\varphi_{_\beta})$. Hence $\mathsf{HA} \vdash (L = \alpha \wedge \Box B) \to \bot$, as desired. ❑

5.4 The Solovay function is a constant function

In this subsection, we will show that $L = \alpha_0$ is a true statement in the standard model (Theorem 5.26). This fact is necessary for showing that for any $\alpha \in K$, the theory $L = \alpha + \mathsf{PA}$ is consistent.

Lemma 5.16. *For each $\alpha \lneqq \beta \in K$ with $\alpha \not\mathcal{R} \beta$,*

$$\mathsf{HA} \vdash \exists x F(x) = \alpha \to \Box^+\neg(L = \beta \wedge \Box\varphi_{_\beta})$$

Proof. By Π_2 conservativity of PA over HA, it is enough to prove the above assertion in PA instead of HA. We work inside PA. Fix some x such that $F(x) = \alpha$. Then for each $y \leq x$, we have $F(y) \preccurlyeq \alpha$. Now, work inside $\Box^+$. Assume $L = \beta$ and $\Box\varphi_{_\beta}$. Then there exists some minimum z such that $F(z+1) = \beta$. So there exists some δ such that $F(z) = \delta$. Since $F(x) = \alpha$, we have $\beta \gneqq \delta \geq \alpha$. Hence $\delta \not\mathcal{R} \beta$. So by recursive definition of F, $r(\beta, z+1) < r(\delta, z+1)$ and $F(r(\beta, z+1))\,\mathcal{R}\,\beta$. Since $\alpha \not\mathcal{R} \beta$, we have $F(r(\beta, z+1)) \ngeqq F(x) = \alpha$, which implies $r(\beta, z+1) < x$. Since $x \leq z$, we have $r(\beta, z+1) < z$ and hence $\Box_x\neg(L = \beta \wedge \Box\varphi_{_\beta})$. Thus by Lemma 3.7, $\neg(L = \beta \wedge \Box\varphi_{_\beta})$, that is a contradiction. ❑

Lemma 5.17. *For any $\beta \in K$ and $B \in \mathsf{Sub}(\Gamma) \cap \mathsf{TNNIL}$,*

- *if $\beta \Vdash B$, then $\mathsf{HA} \vdash (L = \beta \wedge \Box\varphi_{_\beta}) \to B$,*

- *if* $\beta \nVdash B$ *and any occurrence of* $\to$ *in* B *is in the scope of some* $\Box$ ($B \in \mathsf{NOI}$), *then* $\mathsf{HA} \vdash (L = \beta \wedge \Box\varphi_{\beta}) \to \neg B$.

Proof. We prove both items by induction on the complexity of B.

- B is atomic. Then, by definition of the substitution σ, $\mathsf{HA} \vdash B \leftrightarrow \bigvee_{\gamma \Vdash B} \exists x F(x) = \gamma$. If $\beta \Vdash B$, then $\mathsf{HA} \vdash L = \beta \to B$. If $\beta \nVdash B$, then for each $\gamma \Vdash B$, we have $\gamma \nleq \beta$, and hence by Lemma 5.5 item 1, $\mathsf{HA} \vdash L = \beta \to \neg \exists x F(x) = \gamma$. Hence $\mathsf{HA} \vdash L = \beta \to \neg B$.

- B is a conjunction or disjunction. We have the desired conclusions by the induction hypotheses.

- $B = \Box C$. First assume $\beta \Vdash \Box C$. Then, by definition of φ_{β}, C is a conjunct of φ_{β}, and then $\mathsf{HA} \vdash (L = \beta \wedge \Box\varphi_{\beta}) \to B$. For the other side, assume $\beta \nVdash \Box C$. Then Theorem 5.15 implies $\mathsf{HA} \vdash L = \beta \to \neg\Box C$.

- $B = C \to D$. Since B is TNNIL, we have $C \in \mathsf{NOI}$. First assume that $\beta \Vdash C \to D$. If $\beta \Vdash C$, then $\beta \Vdash D$, and hence by the induction hypothesis,

  $$\mathsf{HA} \vdash (L = \beta \wedge \Box\varphi_{\beta}) \to (C \to D).$$

  If $\beta \nVdash C$, then again by the induction hypothesis, $\mathsf{HA} \vdash (L = \beta \wedge \Box\varphi_{\beta}) \to \neg C$, and hence $\mathsf{HA} \vdash (L = \beta \wedge \Box\varphi_{\beta}) \to (C \to D)$.

❑

Lemma 5.18. *Let* $\alpha \in K$ *and for each* $\beta \geq \alpha$, *we have* $\mathsf{HA} \vdash \beta \mathcal{R} L \to \varphi_{\beta}$. *Then for each* $\beta \geq \alpha$ *and* $\gamma \gneq \beta$ *such that* $\beta \not\mathcal{R} \gamma$, *we have*

$$\mathsf{HA} \vdash \exists x F(x) = \beta \to \Box^{+} L \neq \gamma$$

Proof. Fix some $\beta \geq \alpha$. We use induction on γ. Suppose that for each $\gamma_0 \gneq \gamma \gneq \beta$ with $\beta \not\mathcal{R} \gamma_0$, we have $\mathsf{HA} \vdash \exists x F(x) = \beta \to \Box^{+} L \neq \gamma_0$. Then

$\mathsf{HA} \vdash \exists x F(x) = \beta \to \Box^{+}\neg(L = \gamma \wedge \Box\varphi_{\gamma})$	Lemma 5.16
$\mathsf{HA} \vdash \exists x F(x) = \beta \to \Box^{+}((\exists x F(x) = \gamma \wedge \Box\varphi_{\gamma}) \to L \neq \gamma)$	
$\mathsf{HA} \vdash \exists x F(x) = \beta \to \Box^{+}((\exists x F(x) = \gamma \wedge \Box\varphi_{\gamma}) \to \gamma \mathcal{R} L)$	induction hypothesis and neatness
$\mathsf{HA} \vdash \exists x F(x) = \beta \to \Box(\exists x F(x) = \gamma \to (\Box\varphi_{\gamma} \to \varphi_{\gamma}))$	hypothesis of lemma and Lemma 3.2
$\mathsf{HA} \vdash \exists x F(x) = \beta \to \Box(\exists x F(x) = \gamma \to \Box\varphi_{\gamma})$	Löb's axiom, Σ_1-completeness of HA

This in combination with $\mathsf{HA} \vdash \exists x F(x) = \beta \to \Box^{+}\neg(L = \gamma \wedge \Box\varphi_{\gamma})$ implies

$$\mathsf{HA} \vdash \exists x F(x) = \beta \to \Box^{+} L \neq \gamma$$

❑

Lemma 5.19. *For any* $\gamma \in K_0$, $\mathsf{PA} \vdash \exists x F(x) = \gamma \to \Box^{+}\neg(L = \gamma \wedge \Box\varphi_{\gamma})$.

Proof. We work inside PA. Assume $\exists x F(x) = \gamma$. There exists a minimum x_0 such that $F(x_0) = \gamma$. Then by recursive definition of F, we have $F(x) \prec F(x_0)$ for all $x < x_0$, and $F(x_0 \dot{-} 1) = \beta$, and one of the following cases holds:

1. $\beta \mathcal{R} \gamma$ and $r(\gamma, x_0) \leq x_0$, by definition of r, we can deduce

   $$\exists x \leq x_0 \, \mathsf{Proof}_{\mathsf{PA}}(x, \ulcorner \neg(L = \gamma \wedge \varphi_{\gamma}) \urcorner)$$

   and then $\Box^{+}\neg(L = \gamma \wedge \Box\varphi_{\gamma})$.

2. $\beta \not\mathcal{R} \gamma, \beta \prec \gamma$ and $r(\gamma, x_0) < r(\beta, x_0)$. Because $r(\beta, x_0) \leq x_0 + 1$, we can deduce $r(\gamma, x_0) \leq x_0$. By repeating the above argument, we get $\Box^+ \neg(L = \gamma \wedge \Box\varphi_\gamma)$.

❑

Lemma 5.20. *Let $\beta \in K_0$ and for each $\gamma \geq \beta$, we have $\mathsf{HA} \vdash \gamma \mathcal{R} L \to \varphi_\gamma$. Then for each $\gamma \geq \beta$, we have $\mathsf{HA} \vdash \exists x F(x) = \gamma \to \Box\varphi_\gamma$.*

Proof. By Lemma 5.19, $\mathsf{PA} \vdash \exists x F(x) = \gamma \to \Box^+ \neg(L = \gamma \wedge \Box\varphi_\gamma)$. Then Lemma 3.2 implies

$$\mathsf{HA} \vdash \exists x F(x) = \gamma \to \Box^+ \neg(L = \gamma \wedge \Box\varphi_\gamma)$$

Hence

	$\mathsf{HA} \vdash \exists x F(x) = \gamma \to \Box^+(\Box\varphi_\gamma \to L \neq \gamma)$
(by Σ_1-completeness and Lemma 5.5)	$\to \Box^+(\Box\varphi_\gamma \to L \succ \gamma)$
(by Lemma 5.18 and Lemma 5.5)	$\to \Box^+(\Box\varphi_\gamma \to \gamma \mathcal{R} L)$
(by Π_2-conservativity of PA over HA)	$\to \Box(\Box\varphi_\gamma \to \gamma \mathcal{R} L)$
(by hypothesis)	$\to \Box(\Box\varphi_\gamma \to \varphi_\gamma)$
(Löb's axiom)	$\to \Box\varphi_\gamma$

❑

Lemma 5.21. *Let $\beta \in K_0$ and let for each $\gamma \geq \beta$, we have $\mathsf{HA} \vdash \gamma \mathcal{R} L \to \varphi_\gamma$. Then for each $\gamma \geq \beta$ and $B \in \mathsf{Sub}(\Gamma) \cap \mathsf{TNNIL}$, $\gamma \Vdash B$ implies $\mathsf{HA} \vdash \exists x F(x) = \gamma \to B$.*

Proof. We prove this by induction on the frame $(K, \lneqq)$ with reverse order. Let some $\gamma \geq \beta$ and as the (first) induction hypothesis, assume that for each $\gamma_0 \gneqq \gamma$ and $B \in \mathsf{Sub}(\Gamma) \cap \mathsf{TNNIL}$, if $\gamma_0 \Vdash B$, then $\mathsf{HA} \vdash \exists x F(x) = \gamma_0 \to B$. We will show that for each $B \in \mathsf{Sub}(\Gamma) \cap \mathsf{TNNIL}$, if $\gamma \Vdash B$, then $\mathsf{HA} \vdash \exists x F(x) = \gamma \to B$. We prove this by a (second) induction on the complexity of $B \in \mathsf{Sub}(\Gamma) \cap \mathsf{TNNIL}$. Let some $B \in \mathsf{Sub}(\Gamma) \cap \mathsf{TNNIL}$ and $\gamma \Vdash B$ and as the (second) induction hypothesis, assume that for each $C \in \mathsf{Sub}(\Gamma) \cap \mathsf{TNNIL}$ with lower complexity than B (i.e. C is a strict sub-formula of B) such that $\gamma \Vdash C$, we have $\mathsf{HA} \vdash \exists x F(x) = \gamma \to C$. We will show $\mathsf{HA} \vdash \exists x F(x) = \gamma \to B$. We have following cases.

- B is atomic. It is trivial by definition of B.
- B is conjunction or disjunction. The result follows easily by (second) induction hypothesis.
- $B = \Box C$. Suppose that $\gamma \Vdash \Box C$. Then by definition of φ_γ, we have $\mathsf{HA} \vdash \Box\varphi_\gamma \to \Box C$. Now the result is a consequence of Lemma 5.20.
- $B = C \to D \in \mathsf{Sub}(\Gamma) \cap \mathsf{TNNIL}$ and C does not have an occurrence of implication which is not in the scope of any box. Suppose that $\gamma \Vdash C \to D$. There are two sub-cases.
  1. $\gamma \Vdash C$. Then, by (second) induction hypothesis, we can derive $\mathsf{HA} \vdash \exists x F(x) = \gamma \to D$, and hence $\mathsf{HA} \vdash \exists x F(x) = \gamma \to (C \to D)$.
  2. $\gamma \nVdash C$. Then Lemma 5.17 implies that $\mathsf{HA} \vdash (L = \gamma \wedge \Box\varphi_\gamma) \to \neg C$, and by Lemma 5.20, $\mathsf{HA} \vdash L = \gamma \to \neg C$. Now we have $\mathsf{HA} \vdash C \to L \neq \gamma$ and hence

     $$\mathsf{HA} \vdash (\exists x F(x) = \gamma \wedge C) \to L \neq \gamma$$

     So $\mathsf{PA} \vdash (\exists x F(x) = \gamma \wedge C) \to L \neq \gamma$. Then by Lemma 5.5 item 3

     $$\mathsf{PA} \vdash (\exists x F(x) = \gamma \wedge C) \to L \succ \gamma$$

On the other hand, C is implication-free and C is Σ_1, so by Lemma 3.2, we have $\mathsf{HA} \vdash (\exists x F(x) = \gamma \wedge C) \to L \succ \gamma$. For arbitrary $\gamma_0 \gneqq \gamma$, we have $\gamma_0 \Vdash C \to D$. So by the first induction hypothesis, we can derive $\mathsf{HA} \vdash \exists x F(x) = \gamma_0 \to (C \to D)$. By definition of $\gamma \prec L$, we have $\mathsf{HA} \vdash \gamma \prec L \to (C \to D)$. Hence $\mathsf{HA} \vdash (\exists x F(x) = \gamma \wedge C) \to (C \to D)$, which implies $\mathsf{HA} \vdash \exists x F(x) = \gamma \to (C \to D)$.

❑

Lemma 5.22. *For any $\alpha \in K$, $\mathsf{HA} \vdash \alpha \mathcal{R} L \to \varphi_{\alpha}$.*

Proof. Our proof is by reverse induction on the frame $(K, \lneqq)$. As the induction hypothesis, assume that for each $\beta \gneqq \alpha$, we have $\mathsf{HA} \vdash \beta \mathcal{R} L \to \varphi_{\beta}$. For each β with $\alpha \mathcal{R} \beta$, by definition of φ_{α}, we have $\beta \Vdash \varphi_{\alpha}$. Hence by the induction hypothesis and Lemma 5.21, $\mathsf{HA} \vdash \exists x F(x) = \beta \to \varphi_{\alpha}$. Then $\mathsf{HA} \vdash \bigvee_{\alpha \mathcal{R} \beta} \exists x F(x) = \beta \to \varphi_{\alpha}$, which implies $\mathsf{HA} \vdash \alpha \mathcal{R} L \to \varphi_{\alpha}$. ❑

As a direct consequence of Lemma 5.20 and Lemma 5.22 we have the following result.

Corollary 5.23. *For any $\alpha \in K_0$, $\mathsf{HA} \vdash \exists x F(x) = \alpha \to \Box \varphi_{\alpha}$.*

Lemma 5.24. *Let $\alpha \in K$ and $\mathbb{N} \models L = \alpha$. Then $\mathsf{HA} \vdash \alpha \prec L \to \alpha \mathcal{R} L$.*

Proof. Assume $\alpha \lneqq \beta$ and $\alpha \not\mathcal{R} \beta$. Lemma 5.22 implies that for each $\gamma \geq \alpha$, $\mathsf{HA} \vdash \gamma \mathcal{R} L \to \varphi_{\gamma}$. So Lemma 5.18 implies $\mathsf{HA} \vdash L = \alpha \to \Box^{+} L \neq \beta$. Then $\mathbb{N} \models L = \alpha \to \Box^{+} L \neq \beta$, which implies $\mathsf{PA} \vdash L \neq \beta$. Hence $\mathsf{PA} \vdash \alpha \prec L \to \alpha \mathcal{R} L$. Now by Π_2-conservativity of PA over HA, $\mathsf{HA} \vdash \alpha \prec L \to \alpha \mathcal{R} L$. ❑

Lemma 5.25. *Let $\alpha \in K$ and $\mathbb{N} \models L = \alpha$. Then $\mathsf{PA} \cup \{L = \alpha, \Box \varphi_{\alpha}\}$ is consistent.*

Proof. Suppose not, i.e., $\mathsf{PA} \vdash (L = \alpha \wedge \Box \varphi_{\alpha}) \to \bot$. Then $\mathsf{PA} \vdash \Box \varphi_{\alpha} \to L \neq \alpha$. By Σ_1-completeness of PA, we have $\mathsf{PA} \vdash \exists x F(x) = \alpha$ and then by Lemma 5.5, $\mathsf{PA} \vdash \Box \varphi_{\alpha} \to L \succ \alpha$. By Π_2-conservativity of PA over HA, $\mathsf{HA} \vdash \Box \varphi_{\alpha} \to L \succ \alpha$. Lemma 5.24 implies $\mathsf{HA} \vdash \Box \varphi_{\alpha} \to \alpha \mathcal{R} L$. Then by Lemma 5.22, $\mathsf{HA} \vdash \Box \varphi_{\alpha} \to \varphi_{\alpha}$. Then by Löb's theorem, $\mathsf{HA} \vdash \varphi_{\alpha}$. Hence $\mathsf{HA} \vdash \Box \varphi_{\alpha}$. That implies $PA \vdash L \neq \alpha$, a contradiction with $\mathbb{N} \models L = \alpha$. ❑

Theorem 5.26. $\mathbb{N} \models L = \alpha_0$.

Proof. Suppose not, i.e., $\mathbb{N} \models L \neq \alpha_0$. Then by Lemma 5.5 items 2 and 3, $\mathbb{N} \models L = \alpha$, for some $\alpha > \alpha_0$. By Lemma 5.19, $\mathbb{N} \models \exists x F(x) = \alpha \to \Box^{+} \neg (L = \alpha \wedge \Box \varphi_{\alpha})$. This implies $\mathsf{PA} \vdash \neg (L = \alpha \wedge \Box \varphi_{\alpha})$, a contradiction with Lemma 5.25. ❑

Corollary 5.27. *For any $\alpha \leq \beta \in K$ we have $(L = \alpha \wedge \varphi_{\alpha}) \rhd (L = \beta \wedge \varphi_{\beta})$.*

Proof. We should show that $\mathbb{N} \models (L = \alpha \wedge \varphi_{\alpha}) \rhd (L = \beta \wedge \varphi_{\beta})$. This is a direct consequence of Theorem 5.26, Lemma 5.7 and Lemma 5.6. ❑

5.5 Proof of the main theorem

In this subsection, we will prove Theorem 5.1.

With the general method of constructing Kripke models for HA, invented by Smoryński [Smo73b], interpretability of theories containing PA plays an important role in constructing Kripke models of HA.

Definition 5.28. *A triple $\mathcal{I} := (K, <, T)$ is called an I-frame iff it has the following properties:*

- *$(K, <)$ is a finite tree,*
- *T is a function from K to arithmetical r.e. consistent theories containing* PA*,*

- *if $\beta < \gamma$, then T_β interprets T_γ ($T_\beta \rhd T_\gamma$).*

Theorem 5.29. *For every I-frame $\mathcal{I} := (K, <, T)$ there exists a first-order Kripke model $\mathcal{K} = (K, <, \mathfrak{M})$ such that $\mathcal{K} \Vdash \mathsf{HA}$ and moreover $\mathfrak{M}(\alpha) \models T_\alpha$, for any $\alpha \in K$. Note that both of the I-frame and Kripke model are sharing the same frame $(K, <)$.*

Proof. See [Smo73b, page 372-7]. For more detailed proof of a generalization of this theorem, see [AM14, Theorem 4.8]. ❑

Lemma 5.30. *For any $\beta \in K_0$, let $T_\beta := \mathsf{PA} \cup \{L = \beta\}$ and define $\mathcal{I} := (K_0, <_0, T)$. Then*

1. *$\mathcal{I}$ is an I-frame.*
2. *There exists a first-order Kripke model $\mathcal{K}_1 := (K_0, <_0, \mathfrak{M})$ of HA such that for all $\beta \in K_0$ and $B \in \mathsf{Sub}(\Gamma)$, $\mathcal{K}_0, \beta \Vdash B$ iff $\mathcal{K}_1, \beta \Vdash \sigma_{\mathsf{HA}}(B)$.*

Proof. 1. Corollary 5.27 implies that for each $\alpha \leq \beta$, $T_\alpha \rhd T_\beta$. Moreover, Lemma 5.25 implies that T_{α_0} is consistent. These finish the requirements of $\mathcal{I}$ to be an I-frame.

2. By Theorem 5.29, we can find a first-order Kripke model $\mathcal{K}_1 := (K_0, <_0, \mathfrak{M})$, such that for all $\beta \in K_0$, $\mathfrak{M}(\beta) \models T_\beta$ and $\mathcal{K}_1 \Vdash \mathsf{HA}$. Now we prove, by induction on the complexity of $B \in \mathsf{Sub}(\Gamma)$, that for all $\beta \in K_0$, $\mathcal{K}_0, \beta \Vdash B$ iff $\mathcal{K}_1, \beta \Vdash \sigma_{\mathsf{HA}}(B)$.

   - B is atomic. Then by definition, $\sigma_{\mathsf{HA}}(B) := \bigvee_{\gamma \Vdash B} L \succeq \gamma$. For left to right direction, let $\beta \Vdash B$. By second part of Lemma 5.5, $T_\beta \vdash \bigvee_{\gamma \Vdash B} L \succeq \gamma$ and hence $\mathfrak{M}(\beta) \models \bigvee_{\gamma \Vdash B} L \succeq \gamma$. Since $\bigvee_{\gamma \Vdash B} L \succeq \gamma$ is Σ_1-formula, by Lemma 3.8, we have $\beta \Vdash \bigvee_{\gamma \Vdash B} L \succeq \gamma$. For the other way around, let $\beta \nVdash B$. Then Lemma 5.5 item 2 implies $T_\beta \vdash \neg L \succeq \gamma$, for all $\gamma \Vdash B$. This implies that $\mathfrak{M}(\beta) \not\models L \succeq \gamma$, which by use of Lemma 3.8 implies $\beta \nVdash L \succeq \gamma$. So $\beta \nVdash \bigvee_{\gamma \Vdash B} L \succeq \gamma$.
   - B is conjunction, disjunction or implication. Result follows easily by induction hypothesis and inductive definition of $\Vdash$.
   - $B = \Box C$. Let $\beta \Vdash B$. Then by definition of φ_β and Corollary 5.23, $T_\beta \vdash \sigma_{\mathsf{HA}}(\Box C)$, and so $\mathfrak{M}(\beta) \models \sigma_{\mathsf{HA}}(\Box C)$. By Lemma 3.8, $\beta \Vdash \sigma_{\mathsf{HA}}(\Box C)$. For the other way around, suppose $\beta \nVdash \Box C$. Then Theorem 5.15 implies $T_\beta \vdash \neg\sigma_{\mathsf{HA}}(\Box C)$ and hence $\mathfrak{M}(\beta) \not\models \sigma_{\mathsf{HA}}(\Box C)$. Then Lemma 3.8 implies $\beta \nVdash \sigma_{\mathsf{HA}}(\Box C)$.

❑

Proof. (of Theorem 5.1) Let σ be the substitution that we have defined at the beginning of this section and $\mathcal{K}_1$ be as we have by Lemma 5.30. Then the assertion of Lemma 5.30 implies that for any $A \in \mathsf{Sub}(\Gamma)$ we have:

$$\mathcal{K}_0, \alpha \Vdash A \quad \text{iff} \quad \mathcal{K}_1, \alpha \Vdash \sigma_{\mathsf{HA}}(A)$$

❑

6 The Σ_1-provability logic of HA

In this section, we will show that $\mathcal{PL}_\sigma(\mathsf{HA}) = \mathsf{iH}_\sigma$ (see Definition 2.2). Moreover, we will show that iH_σ is decidable. As a by-product of Theorem 5.1, we show that $\mathsf{HA} + \Box\bot$ has de Jongh property. Before we continue with the statement and proof of soundness and completeness theorems, let us apply our techniques presented in this paper to Examples 1.2 and 1.3 in section 1.

Example 6.1. Let $A = \Box(p \vee q) \to (\Box p \vee \Box q)$. We will refute the modal proposition A from the provability logic (and Σ_1-provability logic) of HA.

First of all note that the Kripke model $\mathcal{K}_0$ from Example 1.2 is a counter-model for A. Then by Theorem 5.1, there exists some first-order Kripke model $\mathcal{K}_1 \Vdash \mathsf{HA}$ and some Σ_1-substitution σ

such that $\mathcal{K}_1 \nVdash \sigma_{_{\mathsf{HA}}}(A)$. Hence we have $\mathsf{HA} \nvdash \sigma_{_{\mathsf{HA}}}(A)$, in other words, if we define $B := \sigma(p)$ and $C := \sigma(q)$, then we have $\mathsf{HA} \nvdash \Box(B \vee C) \to (\Box B \vee \Box C)$.

Example 6.2. In this example, we show that how to refute $A = \neg\neg\Box(\neg\neg p \to p) \to \Box(\neg\neg p \to p)$ from the provability logic of HA and also from the Σ_1-provability logic of HA. First we compute the TNNIL-approximation of A, i.e. A^+. By TNNIL-algorithm (Section 4.1.3), $A^+ = \Box(p \vee \neg p) \vee \neg\Box(p \vee \neg p)$. Then $\mathcal{K}_0$ from Example 1.3 is a countermodel for A^+. Then by Theorem 5.1, there exists some first-order Kripke model $\mathcal{K}_1 \Vdash \mathsf{HA}$ and some Σ_1-substitution σ such that $\mathcal{K}_1 \nVdash \sigma_{_{\mathsf{HA}}}(A^+)$. Hence $\mathsf{HA} \nvdash \sigma_{_{\mathsf{HA}}}(A^+)$. Since $\mathsf{HA} \vdash \sigma_{_{\mathsf{HA}}}(\Box A \leftrightarrow \Box A^+)$ (Corollary 4.8), we can deduce $\mathsf{HA} \nvdash \sigma_{_{\mathsf{HA}}}(A)$.

Soundness

Theorem 6.3. iH_σ *is sound for* Σ_1*-arithmetical interpretations in* HA*, i.e.* $\mathsf{iH}_\sigma \subseteq \mathcal{PL}_\sigma(\mathsf{HA})$.

Proof. We must show that Σ_1-interpretations of all axioms of iH_σ hold in HA. For a proof that Σ_1-interpretations of axioms $\Box A \to \Box B$ with $A \blacktriangleright B$, hold in HA, see [Vis02], Theorem 10.2. The other axioms are well-known or obvious, except for the Extended Leivant's principle, which holds by Theorem 3.19. ❑

The following corollary, shows that LC captures the TNNIL part of the theory iH_σ. In other words, as far as we are interested in TNNIL propositions, we can work with the rather simple theory LC instead of iH_σ.

Corollary 6.4. *For any* TNNIL *modal proposition* A*,* $\mathsf{LC} \vdash A$ *if and only if* $\mathsf{iH}_\sigma \vdash \mathsf{A}$.

Proof. The deduction from left to right is by Theorem 4.24 and the fact that $\mathsf{iH}_\sigma \vdash \mathsf{LLe}^+$, which holds by definition of iH_σ. For the right to left direction, assume that $\mathsf{LC} \nvdash A$. Then by Theorem 4.26, there exists some Kripke model $\mathcal{K}_0$ such that $\mathcal{K}_0 \nVdash A$. Then by Theorem 5.1, we can find some Σ_1-substitution σ and some first-order Kripke model $\mathcal{K}_1$, such that $\mathcal{K}_1 \Vdash \mathsf{HA}$ and $\mathcal{K}_1 \nVdash \sigma_{_{\mathsf{HA}}}(A)$. Hence $\mathsf{HA} \nvdash \sigma_{_{\mathsf{HA}}}(A)$. Now Theorem 6.3 implies that $\mathsf{iH}_\sigma \nvdash \mathsf{A}$, as desired. ❑

Completeness

Theorem 6.5. iH_σ *is complete for* Σ_1*-arithmetical interpretations in* HA*, i.e.*

$$\mathcal{PL}_\sigma(\mathsf{HA}) \subseteq \mathsf{iH}_\sigma$$

Proof. Let $\mathsf{iH}_\sigma \nvdash \mathsf{A}$. Then by Corollary 4.19, $\mathsf{iH}_\sigma \nvdash \mathsf{A}^-$. Then by Theorem 4.5 item 1, $\mathsf{iH}_\sigma \nvdash (\mathsf{A}^-)^*$, which implies $\mathsf{iH}_\sigma \nvdash \mathsf{A}^+$. Hence $\mathsf{LLe}^+ \nvdash A^+$ and by Theorem 4.24, $\mathsf{LC} \nvdash A^+$. Then by Theorem 4.26, there exists some Kripke model $\mathcal{K}_0$ such that $\mathcal{K}_0 \nVdash A^+$. Then by Theorem 5.1, we can find some Σ_1-substitution σ and some first-order Kripke model $\mathcal{K}_1$, such that $\mathcal{K} \Vdash \mathsf{HA}$ and $\mathcal{K} \nVdash \sigma_{_{\mathsf{HA}}}(A^+)$. Hence $\mathsf{HA} \nvdash \sigma_{_{\mathsf{HA}}}(A^+)$. Now Corollary 4.8 item 1 implies $\mathsf{HA} \nvdash \sigma_{_{\mathsf{HA}}}(A)$, as desired. ❑

Although the axioms of the theory iH_σ sounds very complicated, however we have the following surpring result.

Theorem 6.6. *The* Σ_1*-provability logic of* HA *(*iH_σ*) is decidable.*

Proof. Let A be a given modal proposition. We explain how to decide $\mathsf{iH}_\sigma \vdash \mathsf{A}$ or $\mathsf{iH}_\sigma \nvdash \mathsf{A}$. First by TNNIL algorithm, compute A^+. Then by Corollary 4.27, we can decide $\mathsf{LC} \vdash A^+$. If $\mathsf{LC} \vdash A^+$, we say 'yes" to $\mathsf{iH}_\sigma \vdash \mathsf{A}$, and otherwise we say "no" to $\mathsf{iH}_\sigma \vdash \mathsf{A}$. Corollary 6.4 guarantees validity of the algorithm.

❑

Theorem 6.7. $\mathsf{HA} + \Box\bot$ *has the de Jongh property, i.e. for all non-modal proposition A, $\mathsf{IPC} \vdash A$ iff for all arithmetical substitution σ, $\mathsf{HA} + \Box\bot \vdash \sigma(A)$.*

Proof. If $\mathsf{IPC} \vdash A$, we apparently have $\mathsf{HA} \vdash \sigma(A)$, for all σ, and hence $\mathsf{HA} + \Box\bot \vdash \sigma(A)$.

For the other way around, let $\mathsf{IPC} \nvdash A$. Hence by Theorem 4.5 item 1, $\mathsf{IPC} \nvdash A^*$. Then by Theorem 4.28, $\mathsf{LC} \nvdash \Box\bot \to A^*$. This implies that $\mathsf{LLe}^+ \nvdash \Box\bot \to A^*$. Hence by Theorem 5.1, there exists some substitution σ such that $\mathsf{HA} + \Box\bot \nvdash \sigma(A)$, as desired. ❑

Acknowledgement. We would like to thank Albert Visser for reading of the first draft of this paper and his very helpful comments, remarks and corrections. Some results of of this paper is obtained during the second author's visit to the Department of Philosophy in Utrecht University in May-June 2011. The second author is thankful to Albert Visser who patiently answered to his questions and having very fruitful conversations. The second author is also very thankful to Jeroen Goudsmit for his helps, hospitality and conversations.

References

[AB04] S. Artemov and L. Beklemishev, *Provability logic*, in Handbook of Philosophical Logic (D. Gabbay and F. Guenthner, eds.), vol. 13, Springer, 2nd ed., 2004, pp. 189–360.

[AM14] M. Ardeshir and S. M. Mojtahedi, *The de Jongh property for Basic Arithmetic*, Archive for Mathematical Logic **53** (2014), no. 7-8, 881–895.

[AM15] ______, *Reduction of provability logics to Σ_1-provability logics*, Logic Journal of IGPL **23** (2015), no. 5, 842–847.

[Ber90] A. Berarducci, *The Interpretability Logic of Peano Arithmetic*, Journal of Symbolic Logic **55** (1990), no. 3, 1059–1089.

[Boo95] G. Boolos, *The logic of provability*, Cambridge University Press, 1995.

[BV06] L. Beklemishev and A. Visser, *Problems in the logic of provability*, Mathematical problems from applied logic. I, Int. Math. Ser. (N. Y.), vol. 4, Springer, New York, 2006, pp. 77–136.

[dJVV11] D. de Jongh, R. Verbrugge, and A. Visser, *Intermediate logics and the de jongh property*, Archive for Mathematical Logic **50** (2011), 197–213.

[Fef60] S. Feferman, *Arithmetization of metamathematics in a general setting*, Fundamenta Mathematicae **49** (1960), no. 1, 35–92 (eng).

[Fri75] H. Friedman, *The disjunction property implies the numerical existence property*, Proc. Nat. Acad. Sci. U.S.A. **72** (1975), no. 8, 2877–2878. MR 0379141 (52 #47)

[Göd31] K. Gödel, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I*, Monatsh. Math. Phys. **38** (1931), no. 1, 173–198.

[Göd33] ______, *Eine interpretation des intuitionistischen aussagenkalkuls*, Ergebnisse eines mathematischen Kolloquiums **4** (1933), 39–40, English translation in: S. Feferman etal., editors, Kurt Gödel Collected Works, Vol. 1, pages 301-303. Oxford University Press, 1995.

[Iem01] R. Iemhoff, *Provability logic and admissible rules*, Ph.D. thesis, University of Amsterdam, 2001.

[Lei75] D. Leivant, *Absoluteness in intuitionistic logic*, Ph.D. thesis, University of Amsterdam, 1975.

[Myh73] J. Myhill, *A note on indicator-functions*, Proceedings of the AMS **39** (1973), 181–183.

[Smo73a] C. Smoryński, *Invesigations of intuitionistic formal systems by means of Kripke models*, Ph.D. thesis, University of Chicago, 1973.

[Smo73b] C. A. Smoryński, *Applications of Kripke models*, Metamathematical investigation of intuitionistic arithmetic and analysis, Springer, Berlin, 1973, pp. 324–391. Lecture Notes in Mathematics, Vol. 344.

[Smo85] C. Smoryński, *Self-reference and modal logic*, Universitext, Springer-Verlag, New York, 1985.

[Sol76] R. M. Solovay, *Provability interpretations of modal logic*, Israel J. Math. **25** (1976), no. 3-4, 287–304.

[TvD88] A. S. Troelstra and D. van Dalen, *Constructivism in mathematics. Vol. I*, Studies in Logic and the Foundations of Mathematics, vol. 121, North-Holland Publishing Co., Amsterdam, 1988, An introduction.

[Vis81] A. Visser, *Aspects of diagonalization and provability*, Ph.D. thesis, Utrecht University, 1981.

[Vis82] ______, *On the completeness principle: a study of provability in Heyting's arithmetic and extensions*, Ann. Math. Logic **22** (1982), no. 3, 263–295.

[Vis98] ______, *An overview of interpretability logic*, Advances in modal logic, Vol. 1 (Berlin, 1996), CSLI Lecture Notes, vol. 87, CSLI Publ., Stanford, CA, 1998, pp. 307–359.

[Vis02] ______, *Substitutions of Σ_1^0 sentences: explorations between intuitionistic propositional logic and intuitionistic arithmetic*, Ann. Pure Appl. Logic **114** (2002), no. 1-3, 227–271, Commemorative Symposium Dedicated to Anne S. Troelstra (Noordwijkerhout, 1999).

[VvBdJRdL95] A. Visser, J. van Benthem, D. de Jongh, and G. R. R. de Lavalette, NNIL, *a study in intuitionistic propositional logic*, Modal logic and process algebra (Amsterdam, 1994), CSLI Lecture Notes, vol. 53, CSLI Publ., Stanford, CA, 1995, pp. 289–326.