The omega-rule interpretation of transfinite provability logic

David Fernández-Duque[a], Joost J. Joosten[b]

[a] *Department of Mathematics*
Ghent University
Krijgslaan 281, B 9000 Gent, Belgium
`David.FernandezDuque@UGent.be`
[b]*Dept. Lògica, Història i Filosofia de la Ciència, Universitat de Barcelona, Montalegre*
6, 08001 Barcelona, Catalonia, Spain
`jjoosten@ub.edu`

Abstract

Given a computable ordinal Λ, the transfinite provability logic GLP_Λ has for each $\xi < \Lambda$ a modality $[\xi]$ intended to represent a provability predicate within a chain of increasing strength. One possibility is to read $[\xi]\phi$ as ϕ *is provable in T using ω-rules of depth at most ξ,* where T is a second-order theory extending ACA_0.

In this paper we will formalize such iterations of ω-rules in second-order arithmetic and show how it is a special case of what we call *uniform provability predicates.* Uniform provability predicates are similar to Ignatiev's strong provability predicates except that they can be iterated transfinitely. Finally, we show that GLP_Λ is sound and complete for any uniform provability predicate.

Keywords: provability logic, arithmetic interpretation, iterated provability

1. Introduction

One compelling and particularly successful interpretation of modal logic is to think of $\Box\phi$ as *the sentence ϕ is provable,* where *provability* is understood within a formal theory T capable of coding syntax. This was suggested by Gödel; indeed, if we use $\Diamond\phi$ as shorthand for $\neg\Box\neg\phi$, the Second Incompleteness Theorem could be written as $\Diamond\top \to \neg\Box\Diamond\top$. It took some time, however, for a complete set of axioms to be assembled, namely until Löb

showed $\Box(\Box\phi \to \phi) \to \Box\phi$ to be valid; the resulting modal logic is called GL (for Gödel-Löb). It took longer still for GL to be proven complete by Solovay [34].

Later, Japaridze [27] enriched the language of GL by adding a sequence of provability modalities $[n]$ for $n < \omega$. The modality $[0]$ is now used as before to state that ϕ is derivable within some fixed formal theory T, while higher modalities represent provability in stronger and stronger theories.

There are many arithmetic interpretations for Japaridze's logic, and one of them also stems from an idea of Gödel, who introduced the notion of a theory T being ω-*consistent*: T is ω-*consistent* whenever for any formula ϕ, if $T \vdash \phi(\overline{n})$ for all $n \in \mathbb{N}$, then $T \nvdash \exists x \neg \phi(x)$. Dually to this notion one can define ω-*provability*: ϕ is ω-provable in T whenever $T + \neg\phi$ is ω-inconsistent.

One may then interpret $[1]\phi$ as ϕ *is* ω-*provable;* a detailed discussion of this is given by Boolos [14]. The higher modalities may then be interpreted by using iterated ω-rules. This idea was already explored by Japaridze in [27] and gives an interpretation for which the polymodal logic GLP_ω is sound and complete; Ignatiev [26] and Beklemishev [9] later improved on this result.

The logic GLP_ω is much more powerful than GL, and indeed Beklemishev has shown how it can be used to perform an ordinal analysis of Peano Arithmetic and its natural subtheories, gauging their proof-strength with respect to Π_1^0 sentences [6]. Boolos also considers a stronger provability predicate in second-order arithmetic which allows unbounded uses of the ω-rule and proves that GLP_2 is sound and complete when the modality $[1]$ is interpreted in this way [13].

Our proof-theoretic interpretation is a straightforward extension of Japaridze's, based on second-order arithmetic. We read $[\alpha]_T\phi$ as *The sentence* ϕ *is derivable in T using ω-rules of nesting depth at most α;* we shall make this precise later. In order to do this we consider a computable well-ordering $\Lambda = \langle |\Lambda|, <_\Lambda \rangle$ on the natural numbers and consider the logic GLP_Λ, already studied by the authors and Beklemishev [7, 21], with the sole difference that we shall represent ordinals as natural numbers rather than appending them as external entities.

Our main result is that GLP_Λ is sound and complete for arithmetical interpretations on 'suitable' theories T; we will mainly work with extensions of ACA_0, but as we shall discuss later, it is possible to work over a weaker theory. It is our hope and expectation that this paper will further the applicability of the transfinite polymodal provability logics GLP_Λ to the above-mentioned program of Π_1^0 ordinal analysis. A discussion on how we think the paper

contributes to the program is included in Appendix B.

Plan of the paper. Section 2 gives a brief review of second-order arithmetic, used in Section 3 to formalize the notion of iterated ω-provability. Section 4 then discusses some of the basic properties of iterated provability classes that can be established in ACA_0. Iterated provability is defined using a Δ_1^1 formula, which we will present in its Π_1^1 form. Section 5 briefly discusses an alternative Σ_1^1 presentation.

Iterated ω-provability is most naturally interpreted in *introspective theories,* introduced in Section 6. Section 7 reviews the logics GLP_Λ and proves that they are sound for their ω-rule interpretation, and more generally for any Λ-*uniform provability predicate.*

In order to prove completeness, Section 8 gives a brief review of the modal logic J, which is used in the completeness proof provided in Sections 9, 10 and 11. Finally, Appendix A discusses possible variations on the notion of iterated ω-provability and Appendix B discusses the choice of our base theory.

2. Second-order arithmetic

In this section we briefly review the subsystems of second-order arithmetic we will use and establish some basic conventions.

2.1. Syntax

Aside from the modal language $\mathcal{L}_\Lambda$, we will work mainly in the language of monadic second-order arithmetic with signature

$$\{0, 1, +, \cdot, \mathsf{exp}, =, <, \in\}$$

so that we have symbols for addition, multiplication and exponentiation; note that $=$ denotes first-order equality, and we in general do not use an equality symbol between sets. As is customary, we use $\mathbf{\Delta}_0^0$ to denote the set of all formulas (possibly with set parameters) where all quantifiers are *bounded,* that is, of the form $\forall\, x{<}t(\vec{y})\ \phi$ or $\exists\, x{<}t(\vec{y})\ \phi$, where $t(\vec{y})$ is some term in our language and x is not among the variables $\vec{y}$. We simultaneously define $\mathbf{\Sigma}_0^0 = \mathbf{\Pi}_0^0 = \mathbf{\Delta}_0^0$ and $\mathbf{\Sigma}_{n+1}^0$ to be the set of all formulas of the form $\exists x_0 \ldots \exists x_m \phi$ with $\phi \in \mathbf{\Pi}_n^0$, and similarly $\mathbf{\Pi}_{n+1}^0$ to be the set of all formulas of the form $\forall x_0 \ldots \forall x_m \phi$ with $\phi \in \mathbf{\Sigma}_n^0$. We denote by $\mathbf{\Pi}_\omega^0$ the union of all $\mathbf{\Pi}_n^0$; these are the *arithmetic formulas.*

The classes $\mathbf{\Sigma}_n^1, \mathbf{\Pi}_n^1$ and $\mathbf{\Pi}_\omega^1$ are defined analogously but using second-order quantifiers and setting $\mathbf{\Sigma}_0^1 = \mathbf{\Pi}_0^1 = \mathbf{\Delta}_0^1 = \mathbf{\Pi}_\omega^0$. It is well-known that every second-order formula is equivalent on the standard model to another formula in one of the above forms. We use lightface to indicate that no set variables appear free ($\Pi_m^n, \Sigma_m^n, \Delta_m^n$). Formulas without free variables (either first or second-order) are *sentences*, and the set of all Π_ω^1 sentences will be denoted $\mathcal{S}_\omega^1$.

We may also wish to add certain formulas to the above-defined classes. If θ is any formula, we define $\mathbf{\Pi}_n^m(\theta), \mathbf{\Sigma}_n^m(\theta)$, etc. as before, except that we add any first-order substitution instance of θ as an atomic formula. If θ is of the form $x \in Y$, we may write $\Gamma(Y)$ instead of $\Gamma(\theta)$.

We fix some natural Gödel numbering, mapping a formula $\psi \in \mathbf{\Pi}_\omega^1$ to its corresponding Gödel number $\ulcorner\psi\urcorner$, and similarly for terms and sequences of formulas (used to represent derivations). A required property of our Gödel numbering is that the code of substrings never exceed the code of the entire string. Moreover, we require that various simple syntactical operations correspond to operations on the Gödel numbers that are elementarily definable in such a way that the basic properties are readily provable. By *elementary* we mean Δ_0^0-definable, recalling that we have included exp in our language. In particular, we fix some set of *numerals,* which are closed terms such that each natural number n is denoted by exactly one numeral written as $\overline{n}$, where the function $n \mapsto \ulcorner\overline{n}\urcorner$ is elementary. Since we will be working mainly inside theories of arithmetic, we will often identify ψ with $\ulcorner\psi\urcorner$.

To simplify notation we use pseudo-terms, where an expression $\varphi(t(\vec{x}))$ should be seen as shorthand for $\exists y{<}\, s(\vec{x})\ (\psi(\vec{x}, y) \wedge \varphi(y))$, with ψ a Δ_0^0 formula defining the graph of the intended interpretation of t and s a standard term bounding the values of $t(\vec{x})$. We mention explicitly the use of the following pseudo-terms:

1. A pseudo-term $\langle x, y \rangle$ which returns a code of the ordered pair formed by x and y; for simplicity of exposition, we will assume $\langle \cdot, \cdot \rangle$ to be a bijection such that always $x \leq \langle x, y \rangle$ and $y \leq \langle x, y \rangle$. We will overload the notation $\langle \ldots \rangle$ and also use it for codes of n-tuples, defined (for example) by recursively setting $\langle \rangle = \langle 0, 0 \rangle$ and

$$\langle x_0, \ldots, x_{n+1} \rangle = \langle \langle \langle x_0, \ldots, x_n \rangle, x_{n+1} \rangle, n + 1 \rangle.$$

2. A pseudo-term $x[y/z]$ which, when x codes a formula ϕ, y a variable v and z a term t, returns the code of the result of subsituting t for v in ϕ

while avoiding variable clashes. Otherwise, its value is unspecified, for example it could be the default $\ulcorner \bot \urcorner$. We shall often just write $\phi(t)$ for this term if the context allows us to.

3. A pseudo-term $x \to y$ which, when x, y are codes for ϕ, ψ, returns a code of $\phi \to \psi$, and similarly for other Boolean operators or quantifiers. The context should always clarify if we use the symbol $\to$ as a part of a term or as a logical connective.

4. A pseudo-term $\overline{x}$ mapping a natural number to the code of its numeral.

5. For every formula ϕ and variables $x_0, \ldots, x_m$, a pseudo-term $\phi(\dot{x}_0, \ldots, \dot{x}_m)$ which, given natural numbers $n_0, \ldots, n_m$, returns the code of the outcome of $\phi[\vec{x}/\vec{n}]$, i.e., the code of $\phi(\overline{n_0}, \ldots, \overline{n_m})$.

The only purpose of using these pseudo-terms is to shorten complex formulas for the sake of legibility. Note that the graphs, ranges and inverses of all these pseudo-terms are elementarily definable. We will assume some of their simple properties without further comment whenever their universal forms are provable over a weak enough base theory.

2.2. Elementarily representable theories

We will say a theory T is *elementarily representable* or simply *representable* if it contains classical predicate logic, is closed under Generalization and Modus Ponens, and there is a Δ_0^0 formula $\texttt{Axioms}_T(x)$ which holds if and only if x is the code of an axiom of T. Subsequently, $\texttt{Proof}_T(x, y)$ will denote the Δ_0^0 formula stating that x codes a derivation in T of a formula coded by y in the usual way. We will be assuming all theories in the text to be representable. Of course, using Craig's trick, any theory with a c.e. set of axioms is deductively equivalent to a representable one.

We may also assume without loss of generality that any derivation d is a derivation of a unique formula ϕ, for example by representing d as a finite sequence of formulas whose last element is ϕ (as is done, for example, in [24]). Also, we assume that every formula that is derivable has arbitrarily large derivations; this is generally true of standard proof systems, for example one may add many copies of an unused axiom or many redundant cuts.

Whenever it does not lead to confusion we will work directly with codes rather than syntactical objects; for example, if ϕ is a natural number (supposedly coding a formula) we use $\Box_T \phi$ as shorthand for $\exists y\, \texttt{Proof}_T(y, \overline{\phi})$. We

write $\Box_T\phi(\dot{x}_0, \ldots, \dot{x}_n)$ as short for $\exists\psi\ (\psi = \phi(\dot{x}_0, \ldots, \dot{x}_n) \wedge \Box_T\psi)$ and adhere to the same convention for other notions of provability.

2.3. Induction and comprehension

It is important in this paper to keep track of the second-order principles that are used; below we describe the most important ones. We use $<$ to denote the standard ordering on the natural numbers and Γ to denote a set of formulas:

$$\Gamma\text{-CA} \quad \exists X \forall x \big(x \in X \leftrightarrow \phi(x) \big),$$
$$\text{where } \phi \in \Gamma \text{ and } X \text{ does not appear free in } \phi;$$

$$\text{I}\Gamma \quad \phi(0) \wedge \forall x \big(\phi(x) \to \phi(x+1) \big) \to \forall x \phi(x),$$
$$\text{where } \phi \in \Gamma;$$

$$\text{Ind} \quad \forall x \big((\forall y {<} x\ \ y \in X) \to x \in X \big) \to \forall x\, (x {\in} X).$$

We assume that all second-order theories extend two-sorted, classical first-order logic, so that they include Modus Ponens, Generalization, etc., as well as Robinson Arithmetic extended with the axioms for exp, which we denote Q^+. However, we will only use a first-order equality symbol. The axioms of Robinson Arithmetic are essentially those of PA without induction; see e.g. [24] for a precise description, and [33] for explicit definitions of PA and its second order equivalent, ACA_0. *Elementary arithmetic* is the first-order theory[1] $\mathrm{EA} = Q^+ + \mathrm{I}\Delta_0^0$.

Although not exactly an instance of $\Gamma\text{-CA}$, the following axiom is closely related:

$$\boldsymbol{\Delta}_1^0\text{-CA} \quad \forall x \Big(\phi(x) \leftrightarrow \psi(x) \Big) \to \exists X \forall x \Big(x \in X \leftrightarrow \phi(x) \Big),$$

where $\phi \in \boldsymbol{\Pi}_1^0$, $\psi \in \boldsymbol{\Sigma}_1^0$ and X does not appear free in ϕ or ψ.

Another principle that will be relevant to us is *transfinite recursion,* but this is a bit more elaborate to describe. It will be convenient to establish a

---

[1]Elementary arithmetic is often presented in a language *without* exponentiation as $Q^+ + \mathrm{I}\Delta_0^0 + \mathsf{Exp}$, where Exp is an axiom asserting that the graph of the exponential function is total. We do not need this additional axiom because we assume that our language includes exponentiation; this difference is innocuous, however, since our presentation is conservative over the other.

few conventions for working with binary relations in second-order arithmetic. First, let us write f *is a function:*

$$\texttt{funct}(f) = \forall x \exists! y (\langle x, y \rangle \in f).$$

Here, $\exists!$ is the standard abbreviation for *there exists a unique.*

Also for simplicity, we may write $n \, R \, m$ if R represents a relation and $\langle n, m \rangle \in R$ as well as $n \, \cancel{R} \, m$ for $\neg(\langle n, m \rangle \in R)$, or $n = f(m)$ if $\langle m, n \rangle \in f$ and f is meant to be interpreted as a function. Rather than working with a second-order equality symbol (which would require an additional extensionality axiom), it suffices to define $X \equiv Y$ by $\forall x \, (x \in X \leftrightarrow x \in Y)$. We also define $X \subseteq Y$ in the standard way by $\forall x \, (x \in X \rightarrow x \in Y)$.

2.4. Computable well-orders

Computable ordinals can be coded in the language of second-order arithmetic. We represent a linear order Λ by a pair of formulas $\langle x < \Lambda, x <_\Lambda y \rangle$; the set of numbers such that $x < \Lambda$ is the *domain* of Λ and written $|\Lambda|$, so that we may simply write $\Lambda = \langle |\Lambda|, <_\Lambda \rangle$. Exceptionally, we will also simply consider a linear order as a set parameter that codes the domain $|\Lambda|$ and the pairs $\langle x, y \rangle$ that stand in the $<_\Lambda$ ordering, but we will make this explicit when it is the case.

We will need to express Λ *is a linear order* and Λ *is a well-order,* as follows:

$\texttt{linear}(\Lambda):$

$$\forall x{<}\Lambda \, \big(\neg(x <_\Lambda x) \wedge \forall y{<}\Lambda \, (x <_\Lambda y \vee y <_\Lambda x \vee y = x)\big)$$
$$\wedge \ \forall x, y, z{<}\Lambda \, (x <_\Lambda y \wedge y <_\Lambda z \rightarrow x <_\Lambda z);$$

$\texttt{wo}(\Lambda):$

$$\texttt{linear}(\Lambda) \ \wedge \ \forall X \subseteq |\Lambda| \, \Big(\exists x \, (x \in X)$$
$$\rightarrow \exists y{\in}X \, \forall z \, \big(z <_\Lambda y \rightarrow z \notin X\big)\Big).$$

We will use Greek letters for natural numbers when viewed as ordered under $<_\Lambda$. In informal discussions we may use natural numbers to represent finite ordinals, so that, for example, 0 is the least element under $<_\Lambda$, independently of whether it truly corresponds to the natural number zero. We should

remark, however, that even if zero is the least element under $\leq_\Lambda$, this fact is not necessarily provable in our formal theory T; that is, we may have that $T \nvdash \forall \xi\ (0 \leq_\Lambda \xi)$.

We say that Λ is *computable* if $x < \Lambda$ is a Δ_0^0 formula[2] and both $x <_\Lambda y$ and $x \nless_\Lambda y$ are Σ_1^0. If Λ is computable, we obtain

$$\xi <_\Lambda \zeta\ \rightarrow\ \Box_T(\overline{\xi} <_\Lambda \overline{\zeta})$$

by formalized Σ_1^0 completeness (see e.g. [14]):

Lemma 2.1. *Let T be some representable arithmetic theory with induction for all Δ_0^0 formulas (with exponential). Let σ be a Σ_1^0 formula with free variables $x_1, \ldots, x_n$ for some $n \in \omega$. Then,*

$$T \vdash \sigma(x_1, \ldots, x_n) \rightarrow \Box_T \sigma(\dot{x}_1, \ldots, \dot{x}_n).$$

In applications, one might wish to work with notations that have further desirable properties, for example by working with paths within Kleene's $\mathcal{O}$ or by explicitly requiring that $<_\Lambda$ satisfies various basic desirable properties in that it can be used to interpret some theory of natural well-orderings like *NWO* from [5]. Nevertheless, for our purposes we only need the following standing assumptions:

1. We assume that all our (well-)orders are computable and, moreover, that linearity is provable in any theory containing induction for all Δ_0^0 formulas.

2. For simplicity, we will henceforth assume that $|\Lambda| = \mathbb{N}$, and thus write, for example, $\forall \lambda\ \phi(\lambda)$ instead of $\forall \lambda {<} \Lambda\ \phi(\lambda)$.

However, it is important to note that all our results can easily be adjusted for any Δ_0^0-definable $|\Lambda| \subsetneq \mathbb{N}$. This is particularly relevant when we wish for Λ to be a finite well-order.

---

[2]We will assume that $x < \Lambda$ is Δ_0^0 for simplicity, but more generally, one could allow $x < \Lambda$ to be Δ_1^0. However, this would require a bit more care, as one would have to keep track of whether our formal theories can prove that it is Δ_1^0.

2.5. Transfinite induction and recursion

Transfinite recursion is the principle that sets may be defined by iterating a formula along a well-order. To formalize this, let us consider a set X whose elements are of the form $\langle \xi, x \rangle$. Write X_ξ for $\{x \mid \langle \xi, x \rangle \in X\}$ and $X_{<_\Lambda \xi}$ for $\{\langle \zeta, x \rangle \in X \mid \zeta <_\Lambda \xi\}$. Then, given a set of formulas Γ, we define

$$\texttt{TR-}\Gamma \quad \forall \Lambda \left(\texttt{wo}(\Lambda) \to \exists X \forall \xi \forall x \ \left(x \in X_\xi \leftrightarrow \phi(x, X_{<_\Lambda \xi}) \right) \right) \quad \text{for } \phi \in \Gamma$$

(note that Λ is a set-variable here). With this, we may define the following systems of arithmetic:

$$
\begin{array}{rclclclc}
\text{RCA}_0 & := & \text{Q}^+ & + & \text{I}\boldsymbol{\Sigma}_1^0 & + & \boldsymbol{\Delta}_1^0\text{-CA} \\
\text{ACA}_0 & := & \text{Q}^+ & + & \text{Ind} & + & \boldsymbol{\Pi}_\omega^0\text{-CA} \\
\text{ATR}_0 & := & \text{ACA}_0 & + & \text{Ind} & + & \text{TR-}\boldsymbol{\Pi}_\omega^0.
\end{array}
$$

We list these from weakest to strongest, but remark that in the realm of second-order arithmetics there are important systems much stronger than ATR_0. For convenience we will work mainly in ACA_0, but in Appendix B will discuss how our techniques could be pushed down even below RCA_0, at the cost of slightly stronger transfinite induction.

The system ATR_0 is relevant because we will define iterated provability by recursion over the well-order Λ. However, as we shall see, we require much less than the full power of arithmetic transfinite recursion for the present work.

In various proofs we wish to reason by transfinite induction. By $\texttt{TI}(\Lambda, \phi)$ we denote the transfinite induction axiom for ϕ along the ordering Λ given by

$$\texttt{TI}(\Lambda, \phi) \ := \forall \xi <\Lambda \left(\left(\forall \zeta <_\Lambda \xi \ \phi(\zeta) \right) \to \phi(\xi) \right) \to \forall \xi <\Lambda \ \phi(\xi),$$

and if Γ is a set of formulas, $\texttt{TI}(\Lambda, \Gamma)$ denotes the scheme $\{\texttt{TI}(\Lambda, \phi) \mid \phi \in \Gamma\}$. The following lemma tells us that comprehension gives us access to transfinite induction for formulas of the right complexity:

Lemma 2.2. *In any second order arithmetic theory T we can prove*

$$\texttt{wo}(\Lambda) \wedge \{\neg \phi\}\text{-CA} \to \texttt{TI}(\Lambda, \phi).$$

Proof. Reason in T and assume $\mathtt{wo}(\Lambda) \wedge \{\neg\phi\}\text{-}\mathtt{CA}$. We prove $\mathtt{TI}(\Lambda, \phi)$ by contraposition. Thus, suppose that $\exists\lambda\neg\phi(\lambda)$. As $\{\xi \mid \neg\phi(\xi)\}$ is a non-empty set, we can apply $\mathtt{wo}(\Lambda)$ to obtain the $<_\Lambda$-minimal such λ. Clearly for this minimal λ we do not have $\big(\forall\zeta<_\Lambda\lambda\ \phi(\zeta)\big) \to \phi(\lambda)$. $\qquad\square$

3. Nested ω-rules

In this section we shall formalize the notion of iterated ω-rules inside second-order arithmetic. An application of the ω-rule allows us to conclude $\vdash \sigma$ whenever for some formula ϕ we have both $\forall n\ \vdash \phi(\overline{n})$ and $\vdash \forall x\phi(x) \to \sigma$. In Boolos ([14]) it is noted that multiple *parallel* applications of the ω-rule do not add extra strength. For example, the rule that allows us to conclude σ from

$$
\begin{array}{rcl}
\forall n & \vdash & \psi(\overline{n}) \\
\forall m & \vdash & \forall x\psi(x) \to \phi(\overline{m}) \qquad\qquad (\dagger) \\
& \vdash & \forall x\,\phi(x)\ \to\ \sigma
\end{array}
$$

can actually be derived by a single application of the ω rule. It is the *nesting depth* that gives extra strength and not the number of applications, where 'nesting' refers to *lying on the same branch* in the derivation tree.

When we admit slightly less uniformity in this last rule ($\dagger$) by allowing ψ to depend on m, and changing the premises to $\forall m\,\forall n\ \vdash \psi_m(\overline{n})$ and $\forall m\ \vdash \forall x\psi_m(x) \to \phi(\overline{m})$ respectively, we get our notion of 2-provability as we shall define in more detail below.

More generally, we may iterate this process to generate a hierarchy of increasingly powerful notions of provability with ξ-fold nesting/iteration of the ω-rule, for a computable ordinal ξ.

We will use $[\lambda]_T^\Lambda\phi$ to denote our representation of *The formula ϕ is provable in T using one application of an ω-rule of depth λ (according to Λ).* The desired recursion for such a sequence of provability predicates is given by the

following equivalence.[3]

$$[\lambda]_T^\Lambda \phi \;\leftrightarrow\; \Big(\Box_T \phi \vee \exists \psi \, \exists \xi <_\Lambda \lambda \, \big(\forall n \, [\xi]_T^\Lambda \psi(\dot n) \;\wedge\; \Box_T(\forall x \psi(x) \to \phi))\big)\Big) \quad (1)$$

As a first step towards formalizing this, we will use a set X as an 'iterated provability class', or IPC for short. Its elements are codes of pairs $\langle \lambda, \phi \rangle$, with λ a code for an ordinal and ϕ a code for a formula; we use $[\lambda]_X \phi$ as shorthand for $\langle \lambda, \phi \rangle \in X$ and $\langle \lambda \rangle_X \phi$ for $\langle \lambda, \neg \phi \rangle \notin X$.

An IPC will collect exactly those pairs of numbers $\langle \lambda, \phi \rangle$ so that (1) is satisfied. As such, any IPC will depend on a parameter Λ whose intended interpretation is a well-ordering on the natural numbers. We will define a formula $\mathtt{IPC}_T^\Lambda(X)$ as a formalization of:

$[\lambda]_X \phi$ *if and only if*

1. $\Box_T \phi$, *or*

2. *there is a formula $\psi(x)$ and an ordinal $\xi <_\Lambda \lambda$ such that*

    (a) *for each $n < \omega$, $[\xi]_X \psi(\overline{n})$, and*

    (b) $\Box_T(\forall x \psi(x) \to \phi)$.

We read $\mathtt{IPC}_T^\Lambda(X)$ as "X is an iterated provability class". Let us enter into a bit more detail:

Definition 3.1. *Define* $\mathtt{Rule}_T^\Lambda(d, \xi, \lambda, \psi, \phi \mid X)$ *to be the formula[4]*

$$\xi <_\Lambda \lambda \wedge \forall n \, [\xi]_X \psi(\dot n) \wedge \mathtt{Proof}_T(d, \forall x \psi(x) \to \phi),$$

---

[3]There are other natural ways of defining this recursion, which may or may not give equivalent notions of iterated provability. One particularly appealing variant is to set

$$[\lambda]_T^\Lambda \phi \;\leftrightarrow\; \Big(\Box_T \phi \vee \exists \psi \, \exists \xi <_\Lambda \lambda \, \big(\forall n \, [\xi]_T^\Lambda \psi(\dot n) \;\wedge\; [\xi]_T^\Lambda(\forall x \psi(x) \to \phi))\big)\Big);$$

but surprisingly this *would* give an equivalent iterated provability predicate. In Appendix A, we shall discuss this and other possible alternatives in more detail.

[4]As a general convention, we will use a bar to separate parameters that are meant to be "quantified away"; in Definition 4.7 we will define a related formula $\mathtt{Rule}_T^\Lambda(d, \xi, \lambda, \psi, \phi)$ which is independent of X.

$\mathrm{Proof}_T^\Lambda(c, \lambda, \phi \mid X)$ *to be*

$$\exists d \exists \xi \exists \psi \left(c = \langle d, \xi, \psi \rangle \wedge \left(\mathrm{Proof}_T(d, \phi) \vee \mathrm{Rule}_T^\Lambda(d, \xi, \lambda, \psi, \phi \mid X) \right) \right),$$

and let $\mathrm{IPC}_T^\Lambda(X)$ *be the formula*

$$\forall \lambda \forall \phi \left(\langle \lambda, \phi \rangle \in X \leftrightarrow \exists c \, \mathrm{Proof}_T^\Lambda(c, \lambda, \phi \mid X) \right).$$

Then, $[\lambda]_T^\Lambda \phi$ *is the* $\mathbf{\Pi}_1^1$*-formula*[5] $\forall X (\mathrm{IPC}_T^\Lambda(X) \to [\lambda]_X \phi)$.

Since we assume that our ordering $<_\Lambda$ is provably transitive, the following lemma becomes a simple observation:

Lemma 3.2. *Let* Λ *be a computable well-order,* T *be a representable theory, and* U *extend* RCA_0. *Then, we have that*

1. $U \vdash \eta \leq_\Lambda \lambda \wedge \mathrm{Rule}_T^\Lambda(d, \xi, \eta, \psi, \phi \mid X) \ \to \ \mathrm{Rule}_T^\Lambda(d, \xi, \lambda, \psi, \phi \mid X)$;

2. $U \vdash \eta \leq_\Lambda \lambda \wedge \mathrm{Proof}_T^\Lambda(c, \eta, \phi \mid X) \ \to \ \mathrm{Proof}_T^\Lambda(c, \lambda, \phi \mid X)$.

We place no restrictions on ψ and ϕ; they may have free first- or second-order quantifiers or variables. Note that the formulas $[\lambda]_X \phi$ and $\langle \lambda \rangle_X \phi$ are independent of T and of Λ and are merely of complexity $\mathbf{\Delta}_0^0$. Note also that since we work with computable well-orders, for r.e. theories T we have that $\mathrm{Proof}_T^\Lambda(c, \lambda, \phi \mid X)$ is equivalent to a $\mathbf{\Sigma}_2^0$-formula in classical predicate logic, whence $\mathrm{IPC}_T^\Lambda(X)$ is equivalent to a $\mathbf{\Pi}_3^0$-formula.

The notion of λ-provability that we have defined naturally compares to the existing literature; in particular, for natural well-orderings Λ and under the assumption that iterated provability classes exist, we have for $n < \omega$ that $[n]_T^\Lambda$ coincides with the provability notions discussed in e.g. [27, 29, 30]. Full ω-logic (i.e., the closure of a theory T under unbounded applications of the ω-rule) has also been formalized in second-order arithmetic [1, 22], and leads to another interpretation of GL, studied in [13] and [14, Chapter 14]. This provability operator is a natural upper bound for all $[\lambda]_T^\Lambda$.

---

[5]The formula is $\mathbf{\Pi}_1^1$ if we consider Λ as a set parameter. In case when Λ is a decidable ordering the complexity is simply Π_1^1. Similar remarks hold for the complexity estimates that follow.

It will be convenient to mention two results that fall somewhat outside the scope of this paper, but nevertheless provide motivation for studying the ω-rule interpretation. Both are discussed, for example, in [31] (or see [23, 32]).

Theorem 3.3. *If Λ is any computable well-order, T is a sound, representable theory, λ is a natural number and $\phi \in \mathcal{S}^1_\omega$, then $[\lambda]^\Lambda_T \phi$ implies that $\mathbb{N} \models \phi$.*

Theorem 3.4. *If ϕ is any $\mathbf{\Pi}^1_1$-sentence such that $\mathbb{N} \models \phi$, then there exists a computable well-order Λ and a natural number λ such that $[\lambda]^\Lambda_T \phi$.*

The proof of Theorem 3.3 follows from a straightforward induction on λ using the fact that the ω-rule preserves truth; however, it cannot be formalized directly within second-order arithmetic as it would require an unrestricted truth predicate. Theorem 3.4 may be seen as a motivation to restrict our attention to computable well-orders. It should also be noted that additional machinery would have to be developed if one were to extend the current work to non-computable orders.

4. Iterated provability classes

Transfinite induction is required in order to prove most of the properties of our iterated provability classes. As an exception, we may infer a weak form of the K-axiom for modal logic (Axiom 2 from Definition 7.1) directly from the definitions:

Lemma 4.1. *Given theories U, T where U extends RCA_0 and T is representable, we have that*

1. $U \vdash \forall\phi \, \forall\psi \, \big(\Box_T(\phi \to \psi) \to ([\lambda]^\Lambda_T \phi \to [\lambda]^\Lambda_T \psi)\big)$;

2. $U \vdash \forall\phi \, \forall\psi \, \big([\lambda]^\Lambda_T(\phi \to \psi) \to (\Box_T \phi \to [\lambda]^\Lambda_T \psi)\big)$.

Proof. For the first claim, assume that X is an IPC and that $\Box_T(\phi \to \psi)$ and $[\lambda]^\Lambda_T \phi$ both hold; let us prove $[\lambda]_X \psi$. From $[\lambda]^\Lambda_T \phi$ and $\mathtt{IPC}^\Lambda_T(X)$ we obtain $[\lambda]_X \phi$, which means that either $\Box_T \phi$ or there are a formula θ and an ordinal $\xi <_\Lambda \lambda$ such that $\forall n [\xi]_X \theta(\dot{n})$ and $\Box_T(\forall x \theta(x) \to \phi)$ both hold.

If $\Box_T \phi$ is the case, then by Modus Ponens we have that $\Box_T \psi$ and thus $[\lambda]_X \psi$. Otherwise, from $\Box_T(\forall x \theta(x) \to \phi)$ we obtain $\Box_T(\forall x \theta(x) \to \psi)$ and thus by one ω-rule we also have $[\lambda]_X \psi$. In either case, $[\lambda]_X \psi$, and since X was an arbitrary IPC we conclude that $[\lambda]^\Lambda_T \psi$.

The proof of the second claim is similar. $\qquad\square$

13

In fact, even the full RCA_0 is not needed for this argument, and in Appendix B we will discuss weaker suitable base-theories. However, for the sake of simplicity we will not consider theories below RCA_0 at this point.

Since $[\lambda]_T^\Lambda \phi$ is given by a $\mathbf{\Pi}_1^1$-formula, it may at first sight seem that one cannot prove properties of our predicate within ACA_0 when this requires induction. Fortunately, this turns out not to be the case. This is based on the observation that *if* an IPC exists, then it must be unique. Let us abbreviate

$$\forall X \forall Y \Big(\phi(X) \wedge \phi(Y) \to X \equiv Y \Big)$$

by $\exists_{\leq 1} X \ \phi(X)$.

Lemma 4.2. *Let T be representable. Then, we have that*

$$\text{RCA}_0 + \text{wo}(\Lambda) \vdash \exists_{\leq 1} X \ \text{IPC}_T^\Lambda(X).$$

Proof. The proof proceeds by a straightforward transfinite induction over Λ, which is available in view of Lemma 2.2, as the formula $x \in X \leftrightarrow x \in Y$ is Δ_0^0; recall that by $\exists_{\leq 1} X \ \text{IPC}_T^\Lambda(X)$ we denote merely uniqueness, and not existence, of an IPC. $\square$

Let $\theta(X)$ be any $\mathbf{\Pi}_\omega^1$ formula. We denote by $\mathbf{\Pi}_\omega^1 \restriction \theta$ the fragment of $\mathbf{\Pi}_\omega^1$ where all second-order quantifiers are of the form $\forall X \ (\theta(X) \to \phi)$ or $\exists X \ (\theta(X) \wedge \phi)$. This is the fragment of $\mathbf{\Pi}_\omega^1$ *protected by θ*. Of particular interest is when θ defines a single set, or more generally, *at most* one set.

Theorem 4.3. *Given a formula $\theta(X) \in \mathbf{\Pi}_\omega^1$,*

$$\text{ACA}_0 \vdash \forall \Lambda \ \big(\exists_{\leq 1} X \ \theta(X) \wedge \text{wo}(\Lambda) \to \text{TI}(\Lambda, \mathbf{\Pi}_\omega^1 \restriction \theta) \big).$$

Proof. For simplicity, let us assume that all second-order quantifiers are of the form $\forall X$. Reasoning within ACA_0 and assuming $\text{wo}(\Lambda)$, we consider two cases.

First assume that $\forall X \neg \theta(X)$. Then we have that, for any formula ψ, $\forall X \ (\theta(X) \to \psi)$ is true. For a formula $\phi \in \mathbf{\Pi}_\omega^1 \restriction \theta$, define $\phi^\top$ to be the result of replacing every occurrence of $\forall X \ (\theta(X) \to \phi)$ by $\top$. Then clearly, $\text{ACA}_0 \vdash \forall X \neg \theta(X) \to (\phi \leftrightarrow \phi^\top)$, and in particular, $\text{TI}(\Lambda, \psi)$ is equivalent to $\text{TI}(\Lambda, \psi^\top)$. But the latter is an instance of induction for an arithmetic formula, which is derivable in ACA_0 under the assumption that $\text{wo}(\Lambda)$, using Lemma 2.2 (see also e.g. [33]).

14

Now assume that $\exists X \theta(X)$. Let X_0 be any set satisfying $\theta(X_0)$. Then, clearly, $\forall X \ (\theta(X) \to \psi(X))$ is equivalent to $\psi(X_0)$.[6] More generally, let $\check{\phi}$ be the result of replacing every subformula of the form $\forall X \ (\theta(X) \to \psi(X))$ by $\psi(X_0)$. Then, for every formula ϕ,

$$\mathrm{ACA}_0 \vdash \theta(X_0) \wedge \exists_{\leq 1} X \theta(X) \to (\phi \leftrightarrow \check{\phi}).$$

In particular, $\mathtt{TI}(\Lambda, \phi)$ is equivalent to $\mathtt{TI}(\Lambda, \check{\phi})$ under these assumptions; but $\check{\phi}$ is arithmetic, hence as before we see that $\mathtt{TI}(\Lambda, \check{\phi})$ is derivable in ACA_0. $\quad\square$

As an immediate corollary we obtain the following very useful result which will allow us to do a substantial deal of induction.

Corollary 4.4. *If T is any representable theory,*

$$\mathrm{ACA}_0 \vdash \forall \Lambda \ \Big(\mathtt{wo}(\Lambda) \to \mathtt{TI}\big(\Lambda, \mathbf{\Pi}^1_\omega \restriction \mathtt{IPC}^\Lambda_T(X)\big)\Big).$$

This will be very convenient in subsequent sections for reasoning about iterated provability, but it is not the only nice consequence of Lemma 4.2. For example, uniqueness allows us to speak of codes of derivations much as with ordinary provability. To this end, let us say a few words about the parameter c in $\mathtt{Proof}^\Lambda_T(c, \lambda, \phi \mid X)$. This parameter plays the role of a proof code, much as the parameter d does in $\mathtt{Proof}_T(d, \phi)$.

Definition 4.5. *A natural number c is a λ-proof of ϕ if it satisfies the formula $[c : \lambda]^\Lambda_T \phi$ defined by*

$$[c : \lambda]^\Lambda_T \phi = \forall X \big(\mathtt{IPC}^\Lambda_T(X) \to \mathtt{Proof}^\Lambda_T(c, \lambda, \phi \mid X)\big).$$

As usual, we define $\langle c : \lambda \rangle^\Lambda_T \phi = \neg[c : \lambda]^\Lambda_T \neg\phi$. The following shows that indeed a λ-proof plays a similar role as a standard derivation:

Lemma 4.6. *Given theories U, T where U extends ACA_0 and T is representable, we have that*

1. $U \vdash \forall\lambda\forall\phi \ \Big(\exists c \,[c : \lambda]^\Lambda_T \phi \to [\lambda]^\Lambda_T \phi\Big)$;

2. $U + \mathtt{wo}(\Lambda) \vdash \forall\lambda\forall\phi \ \Big(\exists c \,[c : \lambda]^\Lambda_T \phi \leftrightarrow [\lambda]^\Lambda_T \phi\Big)$.

---

[6]Here we remind the reader that we are working in a language without set-equality; otherwise, this step would require extensionality.

Proof. Reason within U. Let λ be a natural number and ϕ a formula. For the first item, assume that $[c : \lambda]_T^\Lambda \phi$. Then, if X is any IPC, $\mathtt{Proof}_T^\Lambda(c, \lambda, \phi \mid X)$ holds, so that by definition of an IPC, $[\lambda]_X \phi$. Since X was arbitrary, $[\lambda]_T^\Lambda \phi$.

For the second, it only remains to prove the right-to-left implication. Suppose that $[\lambda]_T^\Lambda \phi$ holds. Consider two cases. If $\neg \exists X \ \mathtt{IPC}_T^\Lambda(X)$, then we may choose $c = 0$ and observe that both sides of the implication are vacuously true.

Otherwise, $\exists X \ \mathtt{IPC}_T^\Lambda(X)$ holds. Let X be an IPC. By the assumption that $[\lambda]_T^\Lambda \phi$ is true, we have that $[\lambda]_X \phi$. By $\mathtt{IPC}_T^\Lambda(X)$, there must exist c such that $\mathtt{Proof}_T^\Lambda(c, \lambda, \phi \mid X)$ holds. Now, if Y is any IPC, by Lemma 4.2, $X \equiv Y$, so we also have $\mathtt{Proof}_T^\Lambda(c, \lambda, \phi \mid Y)$. Since Y was arbitrary, we conclude that $[c : \lambda]_T^\Lambda \phi$, as claimed. $\qquad\square$

We may similarly characterize proof codes in terms of provability.

Definition 4.7. *Define* $\mathtt{Rule}_T^\Lambda(d, \xi, \lambda, \psi, \phi)$ *by*

$$\xi <_\Lambda \lambda \wedge \forall n [\xi]_T^\Lambda \psi(\dot{n}) \wedge \mathtt{Proof}_T(d, \forall x \psi(x) \to \phi).$$

Lemma 4.8. *Given theories* U, T *where* U *extends* ACA_0 *and* T *is representable, we have that*

1. $U \vdash \forall c \forall \lambda \forall \phi \ \Big($

    $\exists d, \xi, \psi \ \Big(c = \langle d, \xi, \psi \rangle \wedge \big(\mathtt{Proof}_T(d, \phi) \vee \mathtt{Rule}_T^\Lambda(d, \xi, \lambda, \psi, \phi) \big) \Big)$

    $\to [c : \lambda]_T^\Lambda \phi \Big),$

2. $U + \mathtt{wo}(\Lambda) + \exists X \ \mathtt{IPC}_T^\Lambda(X) \vdash \forall c \forall \lambda \forall \phi \ \Big([c : \lambda]_T^\Lambda \phi$

    $\leftrightarrow \exists d \exists \xi \exists \psi \ \Big(c = \langle d, \xi, \psi \rangle \wedge \big(\mathtt{Proof}_T(d, \phi) \vee \mathtt{Rule}_T^\Lambda(d, \xi, \lambda, \psi, \phi) \big) \Big) \Big).$

Proof. The proof of the first item is simple and may be carried out entirely within two-sorted predicate logic.

For the remaining implication of the second item, let X_0 be an IPC. By the assumption that $[c : \lambda]_T^\Lambda \phi$ we obtain

$$\exists d \exists \xi \exists \psi \Big(c = \langle d, \xi, \psi \rangle \wedge \big(\mathtt{Proof}_T(d, \phi) \vee \mathtt{Rule}_T^\Lambda(d, \xi, \lambda, \psi, \phi | X_0) \big) \Big).$$

16

Recalling that

$$\texttt{Rule}_T^\Lambda(d, \xi, \lambda, \psi, \phi | X_0) = \xi <_\Lambda \lambda \wedge \forall n[\xi]_{X_0}\psi(\dot{n}) \wedge \texttt{Proof}_T(d, \forall x\psi(x) \to \phi),$$

we use Lemma 4.2 to see that $\forall n[\xi]_{X_0}\psi(\dot{n})$ implies $\forall n[\xi]_T^\Lambda\psi(\dot{n})$, thus obtaining

$$\texttt{Rule}_T^\Lambda(d, \xi, \lambda, \psi, \phi | X_0) \to \texttt{Rule}_T^\Lambda(d, \xi, \lambda, \psi, \phi).$$

The claim follows. □

Note that $\texttt{Rule}_T^\Lambda$ may refer to Definition 3.1 as well as to Definition 4.7 and the two will be distinguished by whether or not a set-parameter is displayed.

5. The existential interpretation

Observe that we defined $[\lambda]_T^\Lambda\phi$ by quantifying universally over all IPCs, but given the uniqueness of an IPC, we could instead have used an existential quantifier. While the two versions are provably equivalent over a strong enough theory, this is not the case for weaker theories. Because of this, it will sometimes be convenient to pass from one interpretation to the other in some proofs, most notably that of Lemma 7.6. The existential interpretation of $[\lambda]$ is as follows.

Definition 5.1. *Given a formula ϕ, a linear order Λ and $\lambda < \Lambda$, we define*

$$\widehat{[\lambda]}_T^\Lambda\phi = \exists X \left(\texttt{IPC}_T^\Lambda(X) \wedge [\lambda]_X\phi \right),$$

and $\widehat{\langle\lambda\rangle}_T^\Lambda\phi = \neg\widehat{[\lambda]}_T^\Lambda\neg\phi$.

As it turns out, $[\lambda]_T^\Lambda\phi$ and $\widehat{[\lambda]}_T^\Lambda\phi$ are equivalent when Λ is a well-order, although to prove this equivalence we need transfinite induction and the existence of an IPC.

Lemma 5.2. *If T is any representable theory and U extends ACA_0,*

1. *$U + \texttt{wo}(\Lambda) \vdash \widehat{[\lambda]}_T^\Lambda\phi \to [\lambda]_T^\Lambda\phi$;*

2. *$U + \exists X\texttt{IPC}_T^\Lambda(X) \vdash [\lambda]_T^\Lambda\phi \to \widehat{[\lambda]}_T^\Lambda\phi$.*

Proof. The first claim is an immediate consequence of Lemma 4.2, which states that there can be at most one IPC, and the second is straightforward and may be carried out within two-sorted predicate logic. □

One may similarly define a variant of $[c : \lambda]_T^\Lambda \phi$:

Definition 5.3. *Given a formula ϕ, an order Λ and $\lambda < \Lambda$, define*

$$\widehat{[c : \lambda]}_T^\Lambda \phi = \exists X \left(\mathtt{IPC}_T^\Lambda(X) \wedge \mathtt{Proof}_T^\Lambda(c, \lambda, \phi \mid X) \right)$$

and $\widehat{\langle c : \lambda \rangle}_T^\Lambda \phi = \neg \widehat{[c : \lambda]}_T^\Lambda \neg \phi$.

The analogue of Lemma 5.2 holds for $\widehat{[c : \lambda]}_T^\Lambda \phi$ and $[c : \lambda]_T^\Lambda \phi$:

Lemma 5.4. *If T is any representable theory, U extends ACA_0, and Λ a well-order, then*

1. *$U + \mathtt{wo}(\Lambda) \vdash \forall c\, \forall \lambda \left(\widehat{[c : \lambda]}_T^\Lambda \phi \rightarrow [c : \lambda]_T^\Lambda \phi \right);$*

2. *$U + \exists X\, \mathtt{IPC}_T^\Lambda(X) \vdash \forall c\, \forall \lambda \left([c : \lambda]_T^\Lambda \phi \rightarrow \widehat{[c : \lambda]}_T^\Lambda \phi \right).$*

Note that $\widehat{[\lambda]}_T^\Lambda \phi$ is Σ_1^1 rather than Π_1^1, and this will allow us to give a simplified analogue of Lemma 4.6:

Lemma 5.5. *Given theories U, T where U extends ACA_0 and T is representable, we have that*

$$U \vdash \forall \lambda \forall \phi \left(\exists c\, \widehat{[c : \lambda]}_T^\Lambda \phi \leftrightarrow \widehat{[\lambda]}_T^\Lambda \phi \right).$$

Proof. Reason within U. Let λ be a natural number and ϕ a formula, and assume that $\exists c \widehat{[c : \lambda]}_T^\Lambda \phi$. Then, there are an IPC X and a natural number c such that $\mathtt{Proof}_T^\Lambda(c, \lambda, \phi \mid X)$ holds, which implies that $[\lambda]_X \phi$. Using X as a witness, we conclude that $\widehat{[\lambda]}_T^\Lambda \phi$.

The other direction is similar and we skip it. $\qquad \square$

Lemma 5.6. *Let U, T be theories where U extends ACA_0 and T is representable. Define a formula $\rho = \rho(d, \xi, \lambda, \psi, \phi)$ by*

$$\xi <_\Lambda \lambda \wedge \forall n \widehat{[\xi]}_T^\Lambda \psi(\dot{n}) \wedge \mathtt{Proof}_T(d, \forall x \psi(x) \rightarrow \phi).$$

Then, it is provable in U that for all λ, ϕ, c, d, ξ and ψ,

$$\left(c = \langle d, \xi, \psi \rangle \wedge \widehat{[c : \lambda]}_T^\Lambda \phi \right) \rightarrow \left(\mathtt{Proof}_T(d, \phi) \vee \rho \right).$$

18

Proof. We omit the proof, which is straightforward and may be carried out in two-sorted predicate logic. Observe that ρ is the analogue of the formula $\mathtt{Rule}_T^\Lambda(d, \xi, \lambda, \psi, \phi)$, albeit replacing $[\xi]_T^\Lambda \psi(\dot{n})$ by $\widehat{[\xi]}_T^\Lambda \psi(\dot{n})$. $\qquad\square$

Thus as we have seen, $[\lambda]_T^\Lambda \phi$ and $\widehat{[\lambda]}_T^\Lambda \phi$ are in principle equivalent, but we need to be able to establish $\mathtt{wo}(\Lambda)$ and the existence of an IPC to prove this. Because of this, it will often be convenient to work with theories that include this assumption, and we will turn our attention to such theories in the next section.

6. Introspective theories

Our notion of ξ-provability, $[\xi]_T^\Lambda$, is a very weak one as it quantifies universally over all iterated provability classes X and it may be the case that there are no such classes. Dually, the notion of consistency $\langle \xi \rangle_T^\Lambda$ is very strong as it in particular asserts the existence of an IPC. In particular, we always provably have

$$\square_T \phi \to [0]_T^\Lambda \phi. \tag{2}$$

However, we cannot always prove $[0]_T^\Lambda \phi \to \square_T \phi$. Nevertheless, the two notions of provability coincide under the assumption that an iterated provability class exists:

Lemma 6.1. *Given a representable theory T and a theory U that extends* ACA_0, *we have that*

1. *$U \vdash \forall \phi \, \forall \lambda \, (\square_T \phi \;\to\; [\lambda]_T^\Lambda \phi)$;*

2. *$U + \exists X \, \mathtt{IPC}_T^\Lambda(X) \vdash \forall \phi \, \forall \mu \, \left(\forall \xi \, (\mu \leq_\Lambda \xi) \to (\square_T \phi \;\leftrightarrow\; [\mu]_T^\Lambda \phi) \right)$.*

Proof. The proofs are straightforward and are left to the reader. $\qquad\square$

Corollary 6.2. *Let U extend ACA_0 and T be representable. If $T \vdash \phi$, then $U \vdash \forall \lambda \, [\lambda]_T^\Lambda \phi$.*

Proof. If $T \vdash \phi$, then $\square_T \phi$ is a true Σ_1^0 sentence, whence $U \vdash \square_T \phi$, and thus $U \vdash \forall \lambda \, [\lambda]_T^\Lambda \phi$ by Lemma 6.1.1. $\qquad\square$

For a theory T to be able to reason about non-trivial facts of iterated provability at all, it is often necessary for it to at least "believe" that such a notion exists. For strong theories this is not an issue, but there is no reason

to assume that ACA_0 or any weaker theory is capable of proving that we have IPCs for non-trivial order types. Hence we shall pay attention to those theories that do prove this, and we shall call them *introspective theories.*

Definition 6.3 (Introspective theory). *Given a linear order Λ, we say that an arithmetic theory T is Λ-introspective if $T \vdash \exists X \ \mathtt{IPC}_T^\Lambda(X)$.*

We defined iterated provability classes by transfinite recursion, and as such it should be no surprise that ATR_0 is Λ-introspective for provable well-orders Λ:

Lemma 6.4. *Given an elementarily representable theory T,*

$$ATR_0 \vdash \mathtt{wo}(\Lambda) \to \exists X \ \mathtt{IPC}_T^\Lambda(X).$$

In particular, ATR_0 is Λ-introspective whenever $ATR_0 \vdash \mathtt{wo}(\Lambda)$.

However, we wish to work over much weaker theories than ATR_0, which may not always be Λ-introspective. Our strategy will be to consider some sort of an "introspective closure", but we do not wish for it to become much stronger than the original theory. Fortunately, this is not too difficult to achieve.

Definition 6.5. *We define the Λ-introspective closure of T as the theory $\overline{T}$ given by $T + \exists X \ \mathtt{IPC}_T^\Lambda(X)$.*

Below, we use the term "Gödelian" somewhat informally as being susceptible to Gödel's second incompleteness theorem; for example, it could be taken to mean sound, representable and extending RCA_0.

Lemma 6.6. *T is equiconsistent in elementary arithmetic with $\overline{T}$, provided T is Gödelian.*

Proof. Clearly the consistency of $\overline{T}$ implies the consistency of T. For the other direction we use that if T is Gödelian, then T is equiconsistent with $T' := T + \Box_T \bot$. We claim that $T' \vdash \exists X \ \mathtt{IPC}_T^\Lambda(X)$ so that $T' \supseteq \overline{T}$ whence

$$\begin{aligned}
\mathtt{Con}(T) &\Rightarrow \mathtt{Con}(T') \\
&\Rightarrow \mathtt{Con}(\overline{T}).
\end{aligned}$$

Indeed, reasoning within T, if T were inconsistent, then $\Box_T \phi$ for every formula ϕ. It follows that if X is an iterated provability class, then $[\lambda]_X \phi$ for all λ and ϕ; hence the trivial set consisting of all pairs $\langle \lambda, \phi \rangle$ is an iterated

20

provability class, and by Δ_0^0 comprehension, it forms a set. Note that this reasoning can be readily formalized in elementary arithmetic. $\qquad\square$

There is still a danger of sliding down a slippery-slope, where $\overline{T}$ is itself not Λ-introspective, thus needing to generate a sequence of theories where each theory is "introspective over the previous". Fortunately, this is not the case. In order to show this we need a technical lemma which will tell us, in particular, that introspection is preserved under taking finite extensions.

Definition 6.7. *Let X be a set of pairs of formulas and ordinals and θ a $\mathbf{\Pi}_\omega^1$-formula. We define the set X given θ, which we denote by $[\![X|\theta]\!]$, by*

$$\langle \lambda, \phi \rangle \in [\![X|\theta]\!] \ :\Longleftrightarrow\ \langle \lambda, \theta \to \phi \rangle \in X.$$

Lemma 6.8. *Let U be a theory containing $\mathbf{\Delta}_0^0$ comprehension, let T be representable and θ be any sentence. Then*

$$U \vdash \mathtt{IPC}_T^\Lambda(Y) \to \mathtt{IPC}_{T+\theta}^\Lambda([\![Y|\theta]\!]).$$

Proof. We reason in U and assume $\mathtt{IPC}_T^\Lambda(Y)$. By $\mathbf{\Delta}_0^0$ comprehension we see that $[\![Y|\theta]\!]$ is a set. We need to show that $[\lambda]_{[\![Y|\theta]\!]}\phi$ holds if and only if either $\Box_{T+\theta}\phi$ holds or there are ψ and $\xi <_\Lambda \lambda$ such that $\forall n[\xi]_{[\![Y|\theta]\!]}\psi(\dot{n})$ and $\Box_{T+\theta}(\forall x\psi(x) \to \phi)$.

If $[\lambda]_{[\![Y|\theta]\!]}\phi$ holds then $[\lambda]_Y(\theta \to \phi)$, which means that either $\Box_T(\theta \to \phi)$ and thus $\Box_{T+\theta}\phi$, or else for some ψ and $\xi <_\Lambda \lambda$ we have $\forall n[\xi]_Y\psi(\dot{n})$ and $\Box_T\big(\forall x\psi(x) \to (\theta \to \phi)\big)$. But using Lemma 4.1 we obtain $\forall n[\xi]_Y(\theta \to \psi(\dot{n}))$ and thus $\forall n[\xi]_{[\![Y|\theta]\!]}\psi(\dot{n})$. Meanwhile, using the tautology

$$\big(\forall x\psi(x) \to (\theta \to \phi)\big) \ \leftrightarrow\ \big(\theta \to (\forall x\psi(x) \to \phi)\big)$$

we obtain $\Box_{T+\theta}(\forall x\psi(x) \to \phi)$, as needed.

For the other direction, assume first that $\Box_{T+\theta}\phi$. Then by the deduction theorem we have $\Box_T(\theta \to \phi)$, and hence $[\lambda]_Y(\theta \to \phi)$, yielding $[\lambda]_{[\![Y|\theta]\!]}\phi$. Otherwise, for some ψ and $\xi <_\Lambda \lambda$ we have $\forall n[\xi]_{[\![Y|\theta]\!]}\psi(\dot{n})$ and $\Box_{T+\theta}(\forall x\psi(x) \to \phi)$. The first gives us $\forall n[\xi]_Y\big(\theta \to \psi(\dot{n})\big)$. Meanwhile, by the Deduction Theorem we obtain $\Box_T\big(\theta \to (\forall x\psi(x) \to \phi)\big)$, and using the tautology

$$\big(\theta \to (\forall x\psi(x) \to \phi)\big) \ \leftrightarrow\ \big(\forall x\,(\theta \to \psi(x)) \to (\theta \to \phi)\big)$$

this implies that $\Box_T\big(\forall x\,(\theta \to \psi(x)) \to (\theta \to \phi)\big)$, so that $[\lambda]_Y(\theta \to \phi)$ and therefore $[\lambda]_{[\![Y|\theta]\!]}\phi$, as needed. $\qquad\square$

Note that in general we may not expect for U to prove $\exists X\ \mathrm{IPC}^\Lambda_{T+\theta}(X) \to \exists X\ \mathrm{IPC}^\Lambda_T(X)$ since for $\theta = \bot$ the antecedent always holds (recall the proof of Lemma 6.6). This will be relevant in the proof of the following result, which is a version of the Deduction Theorem for iterated ω-rules.

Corollary 6.9. *Let U be a theory containing* ACA_0 *and let T be representable. We have that*

1. *$U \vdash [\lambda]^\Lambda_{T+\theta}\phi \to [\lambda]^\Lambda_T(\theta \to \phi)$*

2. *$U + \mathtt{wo}(\Lambda) + \exists X\mathrm{IPC}^\Lambda_T(X) \vdash [\lambda]^\Lambda_{T+\theta}\phi \leftrightarrow [\lambda]^\Lambda_T(\theta \to \phi)$.*

Proof. Reason within U. For the first claim, assume $[\lambda]^\Lambda_{T+\theta}\phi$ and let X satisfy $\mathrm{IPC}^\Lambda_T(X)$. Then, $Y = [\![X|\theta]\!]$ is an IPC for $T + \theta$, hence we have $[\lambda]_Y\phi$. But by definition, this means that $[\lambda]_X(\theta \to \phi)$, and since X was arbitrary, $[\lambda]^\Lambda_T(\theta \to \phi)$.

For the other direction we also need $\exists X\mathrm{IPC}^\Lambda_T(X)$. Assume $[\lambda]^\Lambda_T(\theta \to \phi)$ and let X be an IPC for T. Then, $Y = [\![X|\theta]\!]$ is an IPC for $T + \theta$, and by Lemma 4.2 it is the *only* IPC. As we have by definition that $[\lambda]_Y\phi$, it follows that $[\lambda]^\Lambda_{T+\theta}\phi$, as needed. $\square$

As another direct consequence of Lemma 6.8 we see that the Λ-introspective closure of a theory is indeed itself Λ-introspective.

Lemma 6.10. *If T is representable and contains $\mathbf{\Delta}^0_0$ comprehension then $\overline{T}$ is Λ-introspective.*

Proof. By the above Lemma 6.8, if Y is an iterated provability class for T, then $[\![Y|\exists X\mathrm{IPC}^\Lambda_T(X)]\!]$ is an iterated provability class for $\overline{T}$. Moreover, by $\mathbf{\Delta}^0_0$ comprehension, it forms a set according to $\overline{T}$. $\square$

We conclude that working with Λ-introspective theories is not too restrictive, as far as consistency strength is concerned:

Corollary 6.11. *Given a linear order Λ, any Gödelian theory T is equiconsistent over any theory that extends elementary arithmetic and contains $\mathbf{\Delta}^0_0$-comprehension with a Λ-introspective theory $\overline{T}$.*

A problem that $\overline{T}$ has is that it introduces a Σ^1_1 axiom, possibly increasing its logical complexity. Although not needed for this text, let us briefly discuss a way to avoid this, using *explicitly Λ-introspective* theories:

Definition 6.12. *Given a formal theory T, we define T^P over our second order language of arithmetic augmented with P which is a new set-constant and $T^P = T + \mathtt{IPC}_T^\Lambda(P)$.*

Much as with $\overline{T}$, T^P is Λ-introspective, provided T contains $\boldsymbol{\Delta}_0^0(P)$ comprehension (that is, P is allowed within comprehension instances). In fact, we can do a bit better in this case. Recall that the $\Delta_0^0(P)$ formulas are those bounded formulas where the only set parameter which may appear is P.

Lemma 6.13. *If T is Gödelian and contains $\Delta_0^0(P)$ comprehension and Λ is any linear order, then T^P is Λ-introspective.*

The proof proceeds as before and we omit it. Comprehension for formulas without second-order parameters is very convenient in that it does not "blow up", as it cannot be iterated; for example, $\boldsymbol{\Pi}_1^0$ comprehension *with* set parameters is equivalent to full arithmetic comprehension, but Π_1^0 comprehension is not.

Now that we have shown that Λ-introspective theories are not such a bad thing to work with, we will employ them freely in the next sections. Introspective theories are capable of reasoning about their own iterated provability; for example, we may prove the desired recursion as stated in (1), as can be seen by putting $U = T$ in the next lemma.

Lemma 6.14. *Let U be a theory that extends ACA_0 and T be representable. Then, we have that*

1. *U proves that*

$$\left(\Box_T \phi \vee \exists \psi \, \exists \xi <_\Lambda \lambda \, \left(\forall n \, [\xi]_T^\Lambda \psi(\dot{n}) \, \wedge \, \Box_T(\forall x \psi(x) \to \phi) \right) \right) \; \to \; [\lambda]_T^\Lambda \phi;$$

2. *$U + \mathtt{wo}(\Lambda) + \exists X \, \mathtt{IPC}_T^\Lambda(X)$ proves that*

$$\left(\Box_T \phi \vee \exists \psi \, \exists \xi <_\Lambda \lambda \, \left(\forall n \, [\xi]_T^\Lambda \psi(\dot{n}) \, \wedge \, \Box_T(\forall x \psi(x) \to \phi) \right) \right) \; \leftrightarrow \; [\lambda]_T^\Lambda \phi.$$

Proof. In the first item, we reason in U and need to prove $[\lambda]_T^\Lambda \phi$ under the assumption of the antecedent. To this end, we assume $\mathtt{IPC}_T^\Lambda(X)$ and show $[\lambda]_X \phi$. However, this follows directly from the definition of X being an IPC since we can in particular replace $\forall n \, [\xi]_T^\Lambda \psi(\dot{n})$ in the antecedent by $\forall n \, [\xi]_X \psi(\dot{n})$.

23

For the remaining implication, we assume that $[\lambda]_T^\Lambda \phi$ and use the assumption that $\exists X \, \mathrm{IPC}_T^\Lambda(X)$ to fix an IPC X. Then, from $[\lambda]_T^\Lambda \phi$ we in particular obtain $[\lambda]_X \phi$, which gives us either $\Box_T \phi$ or $\forall n [\xi]_X \psi(\dot{n}) \wedge \Box_T(\forall x \psi(x) \to \phi)$ for some formula ψ and some $\xi <_\Lambda \lambda$. But by Lemma 4.2, $\forall n [\xi]_X \psi(\dot{n})$ implies that $\forall n [\xi]_T^\Lambda \psi(\dot{n})$. Thus we have that

$$\Box_T \phi \vee \exists \psi \, \exists \xi {<_\Lambda} \lambda \, \big(\forall n \; [\xi]_T^\Lambda \psi(\dot{n}) \; \wedge \; \Box_T(\forall x \psi(x) \to \phi) \big),$$

as needed. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

7. Soundness

Given a linear order Λ with $|\Lambda| \subseteq \mathbb{N}$, one may define a logic GLP_Λ with modalities $[\lambda]$ for each $\lambda < \Lambda$, generalizing Japaridze's GLP_ω as was first done in [7]. For this, fix a countable set $\mathbb{P}$ of propositional variables. The formulas of the language $\mathcal{L}_\Lambda$ are built from $\bot$ and variables $p \in \mathbb{P}$ using Boolean connectives $\neg, \wedge$ and a modality $[\xi]$ for each $\xi < \Lambda$. We use $\langle \xi \rangle$ as shorthand for $\neg [\xi] \neg$.

Definition 7.1. *The logic GLP_Λ is given by the following rules and axioms:*

1. *all propositional tautologies,*

2. $[\xi](\phi \to \psi) \to ([\xi]\phi \to [\xi]\psi)$ *for all $\xi < \Lambda$,*

3. $[\xi]([\xi]\phi \to \phi) \to [\xi]\phi$ *for all $\xi < \Lambda$,*

4. $\langle \zeta \rangle \phi \to \langle \xi \rangle \phi$ *for $\xi <_\Lambda \zeta < \Lambda$,*

5. $\langle \xi \rangle \phi \to [\zeta] \langle \xi \rangle \phi$ *for $\xi <_\Lambda \zeta < \Lambda$,*

6. *Modus Ponens, Substitution and Necessitation:* $\dfrac{\phi}{[\xi]\phi}$ *for $\xi < \Lambda$.*

We will normally be interested in the case where Λ is a well-order. In particular, under this notation, the standard unimodal GL becomes GLP_1. Note that that $\langle \xi \rangle \top$ is consistent with GLP_Λ for all $\xi < \Lambda$, although it is not derivable [10].

In this section we shall see that indeed GLP_Λ is sound for our ω-rule interpretation. While proving soundness, we shall keep a clear distinction between the theory (denoted T below) and the *meta-theory* (denoted U below), since

24

they play essentially different roles and hence have different requirements. Also, separating these theories is needed if the logics GLP_Λ are to be employed for Π^0_1 ordinal analysis, as discussed in Appendix B.

We begin the soundness proof by showing the monotonicity of our operators:

Lemma 7.2. *Given theories U, T where U extends ACA_0 and T is representable, we have that*

$$U \vdash \forall \phi \, \forall \xi \, \forall \zeta \, \big((\xi \leq_\Lambda \zeta) \to ([\xi]^\Lambda_T \phi \to [\zeta]^\Lambda_T \phi) \big).$$

Proof. This is immediate from Lemma 3.2 after unpacking the definition of $[\xi]^\Lambda_T \phi$. $\qquad\qquad\square$

To prove that the remaining axioms of GLP_Λ hold we will need transfinite induction over Λ. To this end we define, given a second-order theory T, a new theory T^Λ by

$$T^\Lambda := T + \mathtt{wo}(\Lambda).$$

For ease of exposition we will reason within theories that extend ACA_0, although in Appendix B we shall discuss this choice. Since introspection is closed under taking finite extensions, both $\overline{T}^\Lambda$ and $\overline{T^\Lambda}$ are Λ-introspective (though not necessarily equivalent); for all our arguments below it is irrelevant which one we use.

Lemma 7.3. *Given theories U, T where U extends ACA_0 and T is representable, then*

$$U^\Lambda \vdash \forall \lambda \, \forall \phi_1 \forall \phi_2 \, \big([\lambda]^\Lambda_T (\phi_1 \to \phi_2) \to ([\lambda]^\Lambda_T \phi_1 \to [\lambda]^\Lambda_T \phi_2) \big).$$

Proof. We reason within U and proceed by transfinite induction on λ. Observe that this induction is available over ACA_0 (and hence over U) given Corollary 4.4.

Without loss of generality we may assume the existence of an IPC – since otherwise the consequent is trivially true– and with this have access to Lemma 6.14.2. So, we assume that $[\lambda]^\Lambda_T \phi_1 \wedge [\lambda]^\Lambda_T (\phi_1 \to \phi_2)$.

If either $\Box_T \phi_1$ or $\Box_T (\phi_1 \to \phi_2)$ holds we may use Lemma 4.1. Otherwise, $\neg \Box_T \phi_1$ and $\neg \Box_T (\phi_1 \to \phi_2)$ hold. Then, there must be ψ_1, ψ_2 such that

1. for each $i = 1, 2$ there is $\xi_i <_\Lambda \lambda$ such that for all $n < \omega$, $[\xi_i]^\Lambda_T \psi_i(\overline{n})$,

2. $\Box_T (\forall x \psi_1(x) \to \phi_1)$,

3. $\Box_T(\forall x \psi_2(x) \to (\phi_1 \to \phi_2))$.

By first-order logic we see that

$$\Box_T(\forall x(\psi_1(x) \wedge \psi_2(x)) \to \phi_2). \tag{3}$$

Let $\xi = \max\{\xi_1, \xi_2\}$. By induction on $\xi <_\Lambda \lambda$ and several uses of Modus Ponens inside $[\xi]_T^\Lambda$ we obtain for each n that $[\xi]_T^\Lambda(\psi_1(\overline{n}) \wedge \psi_2(\overline{n}))$. This shows in combination with (3) that $[\lambda]_T^\Lambda \phi_2$ and by transfinite induction on λ, we conclude the claim. $\qquad \square$

Close inspection of the proof of Lemma 7.3 reveals that, by Lemma 2.2, we actually only need $\mathbf{\Sigma}_1^0$ comprehension (with set parameters) to have access to the transfinite induction. And actually the set-parameter is only needed to plug in an iterated provability class. Thus, if our language had a constant P for such a provability class, then $\Sigma_1^0(P)$ comprehension (without additional set-parameters) would suffice. The same holds for the other proofs in this section and we shall no longer explicitly observe this everywhere. In Appendix B we shall see that the amount of comprehension required can often be lowered farther still.

So far we have shown that some of the axioms of GLP_Λ are sound for our omega-rule interpretation; Löb's axiom and the "negative introspection" axiom remain to be checked. For the former, the following lemma will be quite useful.

Lemma 7.4. *Extend* GL *with a new operator* $\blacksquare$ *and the following axioms for all formulas* ϕ, *and* ψ:

1. $\vdash \Box\phi \to \blacksquare\phi$,

2. $\vdash \blacksquare(\phi \to \psi) \to (\blacksquare\phi \to \blacksquare\psi)$ *and,*

3. $\vdash \blacksquare\phi \to \blacksquare\blacksquare\phi$,

and call the resulting system $\mathsf{GL}^\blacksquare$.

Then for all ϕ,

$$\mathsf{GL}^\blacksquare \vdash \blacksquare(\blacksquare\phi \to \phi) \to \blacksquare\phi.$$

Proof. It is well-known that GL is equivalent to $\mathsf{K4}$ plus the Löb Rule [14]:

$$\frac{\Box\phi \to \phi}{\phi}.$$

26

Thus it suffices to check that this rule holds for $\blacksquare$. But indeed, assume that $\mathsf{GL}^{\blacksquare} \vdash \blacksquare\phi \to \phi$. Then, using $\Box\phi \to \blacksquare\phi$ we obtain $\Box\phi \to \phi$, and by Löb's rule (for $\Box$) we see that $\mathsf{GL}^{\blacksquare} \vdash \phi$, as desired. $\qquad\square$

Thus the following result implies that $[\lambda]_T^\Lambda$ is Löbian for all λ:

Lemma 7.5. *Given an order Λ and theories U, T extending ACA_0 such that T is representable, we have that*

1. $U^\Lambda \vdash \forall c \forall \lambda \forall \phi \left([c : \lambda]_T^\Lambda \phi \to [\lambda]_T^\Lambda [\dot{c} : \dot{\lambda}]_T^\Lambda \dot{\phi} \right);$

2. $U^\Lambda \vdash \forall \phi \forall \lambda \left([\lambda]_T^\Lambda \phi \to [\lambda]_T^\Lambda [\dot{\lambda}]_T^\Lambda \dot{\phi} \right).$

Proof. Reason within U^Λ. We will prove both claims simultaneously by induction on λ; as before, Corollary 4.4 allows us to carry out this induction within ACA_0. Without loss of generality we may assume that there is an IPC for if not, the lemma trivially holds.

We begin by proving the first claim. Suppose that ϕ is a formula and c, λ satisfy $[c : \lambda]_T^\Lambda \phi$, so that by Lemma 4.8, $c = \langle d, \xi, \psi \rangle$ and either $\mathtt{Proof}_T(d, \phi)$ or $\mathtt{Rule}_T^\Lambda(d, \xi, \lambda, \psi, \phi)$ holds. By provable Σ_1^0-completeness, we have that $\Box_T \big(\overline{c} = \langle \overline{d}, \overline{\xi}, \overline{\psi} \rangle \big)$. Similarly, if $\mathtt{Proof}_T(d, \phi)$ holds then $\Box_T \mathtt{Proof}_T(\overline{d}, \overline{\phi})$, and thus in this case we obtain $[\lambda]_T^\Lambda [\overline{c} : \overline{\lambda}]_T^\Lambda \overline{\phi}$.

On the other hand, if $\mathtt{Rule}_T^\Lambda(d, \xi, \lambda, \psi, \phi)$ holds, then $\xi <_\Lambda \lambda$ and, once again by Σ_1^0-completeness, $[\lambda]_T^\Lambda(\overline{\xi} <_\Lambda \overline{\lambda})$. Moreover, we have that $\forall n [\xi]_T^\Lambda \psi(\dot{n})$ holds, so that applying the second claim inductively to $\xi <_\Lambda \lambda$, $\forall n [\xi]_T^\Lambda [\overline{\xi}]_T^\Lambda \psi(\dot{n})$ holds as well. By one ω-rule this gives us $[\lambda]_T^\Lambda \forall n [\overline{\xi}]_T^\Lambda \psi(\dot{n})$. Meanwhile, once again by Σ_1^0-completeness we have $\Box_T \mathtt{Proof}_T \big(\overline{d}, \forall x \overline{\psi}(x) \to \overline{\phi} \big)$, and hence $[\lambda]_T^\Lambda \mathtt{Rule}_T^\Lambda(\overline{d}, \overline{\xi}, \overline{\lambda}, \overline{\psi}, \overline{\phi})$. Once again by Lemma 4.8.1 (within T),

$$\Box_T \big(\overline{c} = \langle \overline{d}, \overline{\xi}, \overline{\psi} \rangle \wedge \mathtt{Rule}_T^\Lambda(\overline{d}, \overline{\xi}, \overline{\lambda}, \overline{\psi}, \overline{\phi}) \to [\overline{c} : \overline{\lambda}]_T^\Lambda \overline{\phi} \big),$$

and we may use Lemma 4.1 to conclude that $[\lambda]_T^\Lambda [\overline{c} : \overline{\lambda}]_T^\Lambda \overline{\phi}$.

For the second claim, if $[\lambda]_T^\Lambda \phi$ holds, by Lemma 4.6.2 there is c such that $[c : \lambda]_T^\Lambda \phi$ holds, hence by the previous item, $[\lambda]_T^\Lambda [\overline{c} : \overline{\lambda}]_T^\Lambda \overline{\phi}$. Using Lemma 4.6.1 within T, this implies that $[\lambda]_T^\Lambda [\overline{\lambda}]_T^\Lambda \overline{\phi}$. $\qquad\square$

It remains to check that the axiom $\langle \xi \rangle \phi \to [\lambda] \langle \xi \rangle \phi$ for $\xi <_\Lambda \lambda$ is valid. Here we will need for the first and only time the assumption that T is Λ-introspective.

Lemma 7.6. *If U, T are theories extending ACA_0, Λ is computable and T is representable and Λ-introspective, then*

1. $U^\Lambda \vdash \forall \lambda \forall \phi \forall c \left(\langle c : \lambda \rangle^\Lambda_T \phi \to [\lambda]^\Lambda_T \langle \dot{c} : \dot{\lambda} \rangle^\Lambda_T \dot\phi \right)$;

2. $U^\Lambda \vdash \forall \lambda \, \forall \phi \, \forall \xi <_\Lambda \lambda \left(\langle \xi \rangle^\Lambda_T \phi \to [\lambda]^\Lambda_T \langle \dot\xi \rangle^\Lambda_T \dot\phi \right)$.

Proof. Reason within U^Λ. Instead of the original claims, we will prove that, for all λ and ϕ,

$$\forall c \left(\langle c : \lambda \rangle^\Lambda_T \phi \to [\lambda]^\Lambda_T \widehat{\langle \dot{c} : \dot\lambda \rangle^\Lambda_T \dot\phi} \right); \tag{4}$$

$$\forall \xi <_\Lambda \lambda \left(\langle \xi \rangle^\Lambda_T \phi \to [\lambda]^\Lambda_T \widehat{\langle \dot\xi \rangle^\Lambda_T \dot\phi} \right). \tag{5}$$

The lemma immediately follows, since T is Λ-introspective, so that by Lemma 5.4.2 we have that

$$\Box_T \left(\widehat{\langle c : \lambda \rangle^\Lambda_T \phi} \to \langle c : \lambda \rangle^\Lambda_T \phi \right),$$

and similarly by Lemma 5.2.2, $\Box_T \, \forall c \, \forall \lambda \left(\widehat{\langle \lambda \rangle^\Lambda_T \phi} \to \langle \lambda \rangle^\Lambda_T \phi \right)$.

We proceed to prove (4) and (5) simultaneously by transfinite induction on λ. We begin by proving (5).

Assume that $\xi <_\Lambda \lambda$ is such that $\langle \xi \rangle^\Lambda_T \phi$ holds. By Lemma 4.6.2, this is provably equivalent in U^Λ to $\forall c \, \langle c : \xi \rangle^\Lambda_T \phi$. By the induction hypothesis applied to (4), this implies that $\forall c \, [\xi]^\Lambda_T \widehat{\langle \dot{c} : \overline{\xi} \rangle^\Lambda_T \overline\phi}$. By one application of the ω-rule, this yields $[\lambda]^\Lambda_T \forall c \, \widehat{\langle c : \overline\xi \rangle^\Lambda_T \overline\phi}$, and by Lemma 5.5 this gives us $[\lambda]^\Lambda_T \widehat{\langle \overline\xi \rangle^\Lambda_T \overline\phi}$, as needed.

To prove (4), fix an arbitrary c. Since the pairing function is bijective, we may uniquely write $c = \langle d, \xi, \psi \rangle$. By provable Σ^0_1-completeness, this gives us $[\lambda]^\Lambda_T \left(\overline{c} = \langle \overline{d}, \overline\xi, \overline\psi \rangle \right)$.

If $\langle c : \lambda \rangle^\Lambda_T \phi$ holds, by Lemma 4.8.1 we have that both $\neg \mathtt{Proof}_T(d, \neg\phi)$ and $\neg \mathtt{Rule}^\Lambda_T(d, \xi, \lambda, \psi, \neg\phi)$ hold. By Σ^0_1-completeness and Lemma 6.1.1 we also have $[\lambda]^\Lambda_T \neg \mathtt{Proof}_T(\overline{d}, \neg\overline\phi)$.

Since $\neg \mathtt{Rule}^\Lambda_T(d, \xi, \lambda, \psi, \neg\phi)$ is true, this means that at least one of $\xi \geq_\Lambda \lambda$, $\neg \mathtt{Proof}_T(d, \forall x \psi(x) \to \neg\phi)$, or $\exists n \, \langle \xi \rangle^\Lambda_T \neg \psi(\dot{n})$ holds. If

$$\xi \geq_\Lambda \lambda \vee \neg \mathtt{Proof}_T(d, \forall x \psi(x) \to \neg\phi)$$

28

is the case, then as this is a Σ_1^0-formula we also have

$$[\lambda]_T^\Lambda\big(\overline{\xi} \geq_\Lambda \overline{\lambda} \vee \neg\mathtt{Proof}_T(\overline{d}, \forall x\overline{\psi}(x) \to \neg\overline{\phi})\big).$$

If instead we have that $\exists n\langle\xi\rangle_T^\Lambda\neg\psi(\dot{n})$, choose such an n. By (5), we have that $[\lambda]_T^\Lambda\widehat{\langle\overline{\xi}\rangle}_T^\Lambda\neg\overline{\psi}(\overline{n})$.

Setting

$$\rho = \xi <_\Lambda \lambda \wedge \exists n\widehat{[\overline{\xi}]}_T^\Lambda\psi(\dot{n}) \wedge \mathtt{Proof}_T\big(d, \forall x\psi(x) \to \neg\phi\big),$$

we see that in both cases $[\lambda]_T^\Lambda\neg\rho$ holds. Since we also had $[\lambda]_T^\Lambda\neg\mathtt{Proof}_T(\overline{d}, \neg\overline{\phi})$, we obtain

$$[\lambda]_T^\Lambda\big(\overline{c} = \langle\overline{d}, \overline{\xi}, \overline{\psi}\rangle \wedge \neg\mathtt{Proof}_T(\overline{d}, \neg\overline{\phi}) \wedge \neg\rho\big),$$

so that by the contrapositive of Lemma 5.6 we have that $[\lambda]_T^\Lambda\widehat{\langle\overline{c} : \overline{\lambda}\rangle}_T^\Lambda\overline{\phi}$, as desired. $\qquad\square$

We have essentially proven that GLP_Λ is sound for its omega-rule interpretation. However, it will be convenient to cast this soundness result in a more general setting, so we introduce the notion of *uniform provability predicates* which will be central in our completeness proof further on.

Definition 7.7. *Let T be representable and Λ a linear order. Given a formula $\pi(c, \lambda, \phi)$, we introduce the notation $[c : \lambda]_\pi\phi = \pi(c, \lambda, \phi)$, as well as $[\lambda]_\pi\phi = \exists c[c : \lambda]_\pi\phi$.*

A Λ-uniform proof predicate over T is a formula $\pi(c, \lambda, \phi)$ (with all free variables shown) satisfying

1. *$T \vdash \mathtt{I}\Sigma_1^0(\pi)$;*

2. *$T \vdash \forall\lambda\forall\phi\ (\Box_T\phi \to [\lambda]_\pi\phi)$;*

3. *$T \vdash \forall\lambda\forall\phi\forall\psi\ \Big([\lambda]_\pi(\psi \to \phi) \wedge [\lambda]_\pi\psi \to [\lambda]_\pi\phi\Big)$;*

4. *$T \vdash \forall c\forall\lambda\forall\xi\leq_\Lambda\lambda\forall\phi\ \Big([c : \xi]_\pi\phi \to [c : \lambda]_\pi\phi\Big)$;*

5. *$T \vdash \forall c\forall\lambda\forall\phi\ \Big([c : \lambda]_\pi\phi \to [\lambda]_\pi[\dot{c} : \dot{\lambda}]_\pi\dot{\phi}\Big)$;*

6. *$T \vdash \forall c\forall\lambda\forall\phi\ \Big(\langle c : \lambda\rangle_\pi\phi \to [\lambda]_\pi\langle\dot{c} : \dot{\lambda}\rangle_\pi\dot{\phi}\Big)$;*

29

7. $T \vdash \forall \lambda \, \forall \xi <_\Lambda \lambda \, \forall \phi \, \left(\langle \xi \rangle_\pi \phi \to [\lambda]_\pi \langle \dot{\xi} \rangle_\pi \dot{\phi} \right).$

We say π is sound[7] if, moreover, $\mathbb{N} \models \forall \lambda \forall \phi \, ([\lambda]_\pi \phi \to \phi)$.

A formula $\hat{\pi}$ is a Λ-uniform provability predicate over T if $T \vdash \hat{\pi} \leftrightarrow \exists c \, \pi$, where π is a Λ-uniform proof predicate.

Note that in general we do not require $[0]_\pi \phi \to \Box_T \phi$ to hold. Our ω-rule interpretation gives an example of a uniform provability predicate:

Lemma 7.8. *If Λ is a well-order and T is a sound, representable, Λ-introspective theory extending* ACA_0 *and such that $T \vdash \mathtt{wo}(\Lambda)$, then $[\lambda]_T^\Lambda \phi$ is a sound, Λ-uniform provability predicate with proof predicate $[c : \lambda]_T^\Lambda \phi$.*

Proof. By Lemma 4.6.2, we have that $[\lambda]_T^\Lambda \phi$ is provably equivalent to $\exists c \, [c : \lambda]_T^\Lambda \phi$, so it remains to check that $[c : \lambda]_T^\Lambda \phi$ is a sound, Λ-uniform proof predicate. Property 1 follows from Corollary 4.4, Property 2 is Lemma 6.1.1, Property 4 is Lemma 3.2, Property 3 is Lemma 7.3, Property 5 is Lemma 7.5.1 and Properties 6 and 7 are Lemma 7.6. Soundness is Theorem 3.3. □

Definition 7.9. *An* arithmetic interpretation *is a function[8] $f : \mathbb{P} \to \mathcal{S}_\omega^1$.*

If π is a Λ-uniform proof predicate over T, we denote by f_π the unique extension of f such that $f_\pi(p) = f(p)$ for every propositional variable p, $f_\pi(\bot) = \bot$, f_π commutes with Booleans and $f_\pi([\lambda]\phi) = [\bar{\lambda}]_\pi f_\pi(\phi)$.

Lemma 7.10. *If Λ is any computable linear order on the natural numbers, T is a representable theory extending* ACA_0, *π is a Λ-uniform proof predicate over T and $\mathsf{GLP}_\Lambda \vdash \phi$, then $T \vdash f_\pi(\phi)$ for every arithmetic interpretation f.*

Proof. By an easy induction on the length of a GLP_Λ-proof of ϕ, using the fact that each of the axioms is derivable. Necessitation uses Σ_1^0-completeness and Property 2. □

Theorem 7.11 (Soundness)**.** *Let Λ be any computable well-order on the natural numbers, T a representable, Λ-introspective theory extending* ACA_0 *such that $T \vdash \mathtt{wo}(\Lambda)$ and $f_T^\Lambda = f_{\mathtt{Proof}_T^\Lambda}$. Then, if $\mathsf{GLP}_\Lambda \vdash \phi$ it follows that $T \vdash f_T^\Lambda(\phi)$ for every arithmetic interpretation f.*

Proof. Immediate from Lemmas 7.8 and 7.10. □

---

[7]Observe that for π to be sound, we must have that T itself was already sound.

[8]Recall that $\mathbb{P}$ is the set of propositional variables and $\mathcal{S}_\omega^1$ the set of Π_ω^1 sentences.

Now that we have proven that GLP_Λ is sound, our main objective will be to prove the converse of Theorem 7.11, that is, the completeness of GLP_Λ. For this, let us first review the modal logic J.

8. The logic J

Kripke semantics are often a useful tool for studying modal logics, and the logic GLP_1 is sound and complete for the class of finite, well-founded trees. However, the logic GLP_Λ is not sound and complete for any class of Kripke frames whenever $\Lambda > 1$. In order to remedy this situation, Beklemishev, improving on an idea of Ignatiev [26], has proposed the weaker logic J in [8]. This logic J is very similar to GLP_ω, except that we replace Axiom 4 of GLP_ω by the two axioms

6. $[n]\phi \to [m][n]\phi$, for $n \leq m$ and

7. $[n]\phi \to [n][m]\phi$, for $n < m$.

We will denote by J_N the fragment of J which only contains modalities less than N.

For our purposes, a *Kripke frame* is a structure $\mathfrak{F} = \langle W, \langle R_n \rangle_{n<N} \rangle$, where W is a set, N a natural number and $\langle R_n \rangle_{n<N}$ a family of binary relations on W. A *valuation* on $\mathfrak{F}$ is a function $\llbracket \cdot \rrbracket : \mathcal{L}_N \to \mathcal{P}(W)$ such that

$$
\begin{array}{rclcrcl}
\llbracket \bot \rrbracket & = & \varnothing & \qquad & \llbracket \neg\phi \rrbracket & = & W \setminus \llbracket \phi \rrbracket \\
\llbracket \phi \wedge \psi \rrbracket & = & \llbracket \phi \rrbracket \cap \llbracket \psi \rrbracket & \qquad & \llbracket \langle n \rangle \phi \rrbracket & = & R_n^{-1} \llbracket \phi \rrbracket,
\end{array}
$$

where for $X \subseteq W$, we define $R_n^{-1}X = \{w \in W : \exists v \in X \ w R_n v\}$.

A *Kripke model* is a Kripke frame equipped with a valuation $\llbracket \cdot \rrbracket$; note that propositional variables may be assigned arbitrary subsets of W. Clearly, a valuation is uniquely determined once we have fixed its values for the propositional variables. As usual, ϕ is *satisfied* on $\langle \mathfrak{F}, \llbracket \cdot \rrbracket \rangle$ if $\llbracket \phi \rrbracket \neq \varnothing$, and *valid* on $\langle \mathfrak{F}, \llbracket \cdot \rrbracket \rangle$ if $\llbracket \phi \rrbracket = W$. Often we will write $\mathfrak{F}, x \Vdash \psi$ instead of $x \in \llbracket \psi \rrbracket$, and write $\mathfrak{F} \Vdash \psi$ if ψ is valid on $\mathfrak{F}$. It is well-known that $[n]$ obeys GL whenever R_n^{-1} is well-founded and transitive, in which case we write R_n^{-1} as $<_n$.[9]

---

[9]Note that this strays from convention in that $v <_n w$ means that v is n-accessible from w, and not vice-versa. There are several reasons why we prefer this presentation; for example, some of the most natural models of GLP_Λ come from ordinals, where the "possible worlds" from ξ are all less than ξ [11, 18, 20, 25].

Unlike GLP_N, the logics J_N are proven in [8, 10] to be sound and complete for a class of Kripke frames, defined as follows:

Definition 8.1. *A J_N-frame is a structure $\mathfrak{W} = \langle W, \langle >_n \rangle_{n < N} \rangle$, where W is a finite set and*

1. *the relations $<_n$ are transitive and well-founded,*

2. *if $n < m$ and $w <_m v$ then $<_n(w) = <_n(v)$ (where $<_n(w) = \{u : u <_n w\}$), and*

3. *if $n < m$ then $w <_m v <_n u$ implies that $w <_n u$.*

It will also be convenient to define some auxiliary relations on the J_N-frame $\mathfrak{W}$. Say that:

- $w \ll_n v$ if for some $m \geq n$, $w <_m v$. We will use $\lll_n$ to denote the reflexive closure of $\ll_n$.

- $w \lll_n v$ if there is $u \in W$ such that $w \ll_n u \gg_{n+1} v$.

We will also use $\approx_n$ to denote the symmetric, reflexive, transitive closure of $\ll_n$, and $[w]_n$ to denote the equivalence class of w under $\approx_n$. Write $[w]_{n+1} <_n [v]_{n+1}$ if there exist $w' \in [w]_{n+1}$, $v' \in [v]_{n+1}$ such that $w' <_n v'$. Next, let us state some useful properties of J-frames:

Lemma 8.2. *Let $\mathfrak{W} = \langle W, \langle >_n \rangle_{n < N} \rangle$ be any J_N-frame. Then,*

1. *If $v <_m w$ and $v <_n w$ then $m = n$.*

2. *If $u <_m v <_n w$ then $u <_{\min\{m,n\}} w$.*

3. *The relation $\ll_n$ is transitive and well-founded.*

4. *If $w \approx_m v$ and $n < m$, then $<_n(w) = <_n(v)$.*

Proof. For the first item, assume that $v <_m w$ and $v <_n w$. Towards a contradiction, assume that $m \neq n$; without loss of generality, we may assume $n < m$. Then, by the J-frame condition 2, $<_n(w) = <_n(v)$, which means in particular that $v <_n v$. But this contradicts the well-foundedness of $<_n$, according to the J-frame condition 1.

For the second, if $n < m$, by the J-frame condition 3, we have that $u <_n w$. If $m = n$, then $u <_m w$ by transitivity (J-frame condition 1). Finally, if
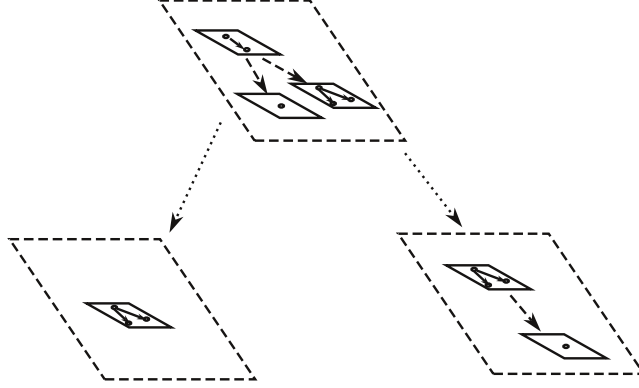
32

Figure 1: A stratified J_3-frame; dotted arrows represent $<_0$, dashed arrows $<_1$ and solid arrows $<_2$; accodringly, dashed boxes represent $\approx_1$ and solid ones $\approx_2$. Note that per our convention, a dotted arrow from w to v means that $v <_0 w$. The frame may be seen as a partial ordering $<_0$ whose nodes are the $\approx_1$-equivalence classes. Each of these classes is, in turn, a tree-like J-frame but now only with relations $<_1, <_2$, and each of whose nodes, in turn, is a J-frame on $<_2$. It is helpful to think of the equivalence classes $[w]_n$ as lower-dimensional "sheets", as depicted in the figure.

$m < n$, then by the J-frame condition 2, we have that $<_m(v) = <_m(w)$, and in particular $u <_m w$.

For the third, we begin by establishing transitivity. Assume that $u \ll_n v \ll_n w$, so that there are $m, k \geq n$ with $u <_m v <_k w$. Then, by the previous item, $u <_{\min\{m,k\}} w$, and hence $u \ll_n w$. For well-foundedness, observe that, since W is assumed finite and $\ll_n$ is transitive, it suffices to check that $\ll_n$ is irreflexive. But, if $w \ll_n w$, this means that $w <_m w$ for some $m \geq n$, contradicting the well-foundedness of $<_m$.

The fourth claim follows by an easy induction using condition 2. $\qquad\square$

A J_N-frame $\mathfrak{W}$ is said to be *stratified* if whenever $[w]_{n+1} <_n [v]_{n+1}$, it follows that $w <_n v$. This means that $w <_n v$ and $w <_m u$ implies $u <_n v$ whenever $m > n$. Strictly speaking, stratification is not needed for any of our proofs, but stratified frames are particularly easy to visualize (see Figure 1). We say $\mathfrak{W}$ is *rooted* if there is a world 0 such that for all $w \in W$, $w \lessapprox_0 0$, in which case 0 is its *root*.

With this we may state the following completeness result:

Theorem 8.3. *Given $\phi \in \mathcal{L}_N$, ϕ is J_N-consistent if and only if ϕ can be satisfied at the root of a stratified J_N-frame.*

This result is proven for the full logic J in [8] and for the fragments J_N in [10], where it is shown that, in fact, J is conservative over J_N. Thus if we can reduce GLP_ω to J, we will be able to work with finite well-behaved Kripke models. For this, given a formula ϕ whose maximum modality is N, define

$$M(\phi) = \bigwedge_{\substack{[n]\psi \in \mathrm{sub}(\phi) \\ n < m \leq N}} \big([n]\psi \to [m]\psi\big).$$

Then we set $M^+(\phi) = M(\phi) \wedge \bigwedge_{n \leq N}[n]M(\phi)$.

The formulas $M(\phi)$ make expressions of the form $[n]\psi$ become much stronger.

Lemma 8.4. *Let $\mathfrak{W} = \mathfrak{W} = \langle W, \langle >_n \rangle_{n<N} \, [\![\cdot]\!] \rangle$ be a J_N-model and $\phi \in \mathcal{L}_N$ be such that $\mathfrak{W} \Vdash M(\phi)$. Suppose that $[n]\psi \in \mathrm{sub}(\phi)$ and $w \in W$ are such that $\mathfrak{W}, w \Vdash [n]\psi$.*

Then, whenever $v \lll_n w$, it follows that $\mathfrak{W}, v \Vdash \psi$.

Proof. Suppose that $\mathfrak{W}, w \Vdash [n]\psi$ and $[n]\psi \in \mathrm{sub}(\phi)$. If $v \lll_n w$, then there is $u \in W$ such that $v \ll_n u \ggg_{n+1} w$; note that $u \approx_{n+1} w$, so by Lemma 8.2.4, $<_n(u) = <_n(w)$, from which it follows that $\mathfrak{W}, u \Vdash [n]\psi$. Moreover, $v <_m u$ for some m with $n \leq m < N$, so that from the assumption that $\mathfrak{W} \Vdash M(\phi)$ we obtain $\mathfrak{W} \Vdash [n]\psi \to [m]\psi$, hence $\mathfrak{W}, u \Vdash [m]\psi$. In particular, it follows that $\mathfrak{W}, v \Vdash \psi$, as claimed. $\qquad\square$

Moreover, the formulas $M^+(\phi)$ provide a reduction from GLP_ω to J, as proven in [8]:

Lemma 8.5. *For any formula $\phi \in \mathcal{L}_\Lambda$ with modalities less than N, $\mathsf{GLP}_\omega \vdash \phi$ if and only if $\mathsf{J}_N \vdash M^+(\phi) \to \phi$.*

Once again, the result is established for J in [8] and for J_N in [10]. Combining this with Theorem 8.3, we see that if $\phi \in \mathcal{L}_N$ is GLP_ω-consistent, then ϕ is satisfiable on some rooted J_N-model $\mathfrak{W}$. Moreover, since $\mathfrak{W}$ is rooted and the root satisfies $M^+(\phi)$, it is easy to see that $\mathfrak{W} \Vdash M(\phi)$. Nevertheless, we often want a model where $\Diamond\phi$, rather than ϕ, is satisfied. We can obtain such a model as follows.

34

Lemma 8.6. *Let* $\mathfrak{W} = \langle W, \langle >_n \rangle_{n<N}, [\![\cdot]\!] \rangle$ *be a* J_N*-model with* $0 \notin W$, *and consider a new structure*

$$\mathfrak{W}' = \langle W', \langle >'_n \rangle_{n<N}, [\![\cdot]\!]' \rangle,$$

where

(i) $W' = W \cup \{0\}$,

(ii) $>'_0 = >_0 \cup (\{0\} \times W)$ *and if* $n < N$, $>'_n = >_n$;

(iii) $[\![p]\!]' = [\![p]\!]$ *for any propositional variable* p.

Then, for any $\phi \in \mathcal{L}_N$:

1. $\mathfrak{W}'$ *is a rooted* J_N*-model with root* 0;

2. *if* $\mathfrak{W}$ *is stratified, then* $\mathfrak{W}'$ *is stratified;*

3. *if* $w \in W$ *then* $\mathfrak{W}', w \Vdash \phi$ *if and only if* $\mathfrak{W}, w \Vdash \phi$;

4. *if* $1 \in W$ *is such that* $\mathfrak{W}, 1 \Vdash \phi$, *then* $\mathfrak{W}', 0 \Vdash \Diamond\phi$, *and*

5. *if* $\mathfrak{W} \Vdash M(\phi)$, *then* $\mathfrak{W}' \Vdash M(\phi)$.

Proof. Items 1 and 2 are proven by checking each of the required conditions. For example, we show that if $u <'_m v <'_n w$ then $u <'_{\min\{m,n\}} w$. If u, v, w are all different from 0, then we already had $u <_m v <_n w$, so that $u <_{\min\{m,n\}} w$, and hence $u <'_{\min\{m,n\}} w$. Otherwise, it is easy to see from the definition that only w can be 0, and thus $0 = n = \min\{m, n\}$. But then, we have that $u <'_0 0$ by definition.

For item 3, we observe that W is *generated* in $\mathfrak{W}'$, i.e. if $w \in W$ and $w >_n v$ it follows that $v \in W$ as well. By standard modal theory [15] (or, alternately, by a straightforward structural induction on ϕ), we see that for any $w \in W$, $\mathfrak{W}, w \Vdash \phi$ if and only if $\mathfrak{W}', w \Vdash \phi$. Item 4 is then immediate, since in particular we obtain that $\mathfrak{W}', 1 \Vdash \phi$, hence $\mathfrak{W}, 0 \Vdash \phi$.

Finally, for item 5, assume that $\mathfrak{W} \Vdash M(\phi)$. By item 3, for all $w \in W$, $\mathfrak{W}', w \Vdash M(\phi)$. Hence it remains to check that $\mathfrak{W}', 0 \Vdash M(\phi)$. Observe that $\mathfrak{W}', 0 \Vdash [m]\psi$ whenever $m > 0$ *independently* of ψ, simply because there is no w such that $w <'_m 0$. Moreover, if $n < m$ we forcibly have $m > 0$, thus $\mathfrak{W}', 0 \Vdash [n]\psi \to [m]\psi$ for any ψ. Recalling the definition of $M(\phi)$, it readily follows that $\mathfrak{W}', 0 \Vdash M(\phi)$. □

Let us put all of these results together into the form that will be useful to us.

Lemma 8.7. *If $N < \omega$ and $\phi \in \mathcal{L}_N$ is GLP_ω-consistent, then there is a finite, stratified J_N-model $\mathfrak{W}$ with root 0 such that*

1. *$\mathfrak{W}, 0 \Vdash \Diamond\psi$, and*

2. *$\mathfrak{W} \Vdash M(\phi)$.*

Proof. If $\phi \in \mathcal{L}_N$ is GLP_ω-consistent, by Lemma 8.5, $M^+(\phi) \wedge \phi$ is J_N-consistent, so that by Theorem 8.3, $M^+(\phi) \wedge \phi$ is satisfied on the root of some stratified J_N-model $\mathfrak{W}$. Note that since the root of $\mathfrak{W}$ satisfies $M^+(\phi)$, it readily follows that $\mathfrak{W} \Vdash M(\phi)$. Thus using Lemma 8.6, we see that there is a finite, rooted, stratified J_N-model $\mathfrak{W}'$ with root 0 such that $\mathfrak{W}', 0 \Vdash \langle 0 \rangle \phi$ and $\mathfrak{W}' \Vdash M(\phi)$, as needed. $\square$

We shall use this result in the next section in order to prove arithmetical completeness by "piggybacking" on the completeness of J_N for finite frames.

9. On arithmetical completeness

In the following sections we want to prove that GLP_Λ is complete with respect to T for any sound uniform proof predicate π over T, and in particular for its ω-rule interpretation. This means that, given a GLP_Λ-consistent formula ϕ, there is an arithmetic interpretation f such that $\neg f_\pi(\phi)$ is not derivable in T (we will make this claim precise in Theorem 10.2).

There are many proofs of completeness of GL and GLP_ω, and it is possible to go back to an existing proof and adjust it to prove completeness in our setting. Because of this, we should say a few words about our choice of including a full proof in this paper. There are essentially two reasons.

The first is that, while our result follows to a certain degree from known *proofs,* it does not follow from known *results;* even then, there would be several technical issues in adjusting known arguments to our setting, as they make assumptions that are not available to us.

The second is that the argument we propose carries some simplifications over previous proofs that could also be applied to standard interpretations of GLP_ω, thus contributing to an ongoing effort to find simpler arguments for this celebrated completeness result.

There have been early completeness proofs of transfinite provability logics (see [3, 4], where the latter has completeness proofs with respect to progressions within or through Kleene's universal notation system $\mathcal{O}$), although the interpretations considered yield modal logics that are very different from GLP_Λ. For GLP_Λ itself there are at least seven arithmetical completeness results in the literature, all with $\Lambda \leq \omega$:

1. Solovay originally constructed a function h with domain ω of a self-referential nature and used statements about h to prove the completeness theorem for the unimodal GLP_1 in [34]. The proof used the recursion theorem.

2. Boolos introduced in [12] a modification of Solovay's original proof using the fixpoint theorem instead of the recursion theorem, where the function h is simulated via finite sequences that represent computations.

3. Japaridze proved in [27] the completeness of GLP_ω for a first-order version of the ω-rule interpretation we are presenting here.

4. De Jongh, Jumelet and Montagna gave in [17] a more elementary construction of the fixpoints using the simultaneous fixpoint theorem. Moreover, the reasoning about these simultaneous fixpoints could be largely formalized within a modal logic incorporating witness comparison statements (essentially Guaspari-Solovay's logic R^-).

5. Ignatiev generalized Japaridze's result in [26] to a large family of "strong provability predicates".

6. Boolos provides in [13] a completeness proof for GLP_2 where the modality [1] is interpreted in second-order arithmetic as *provable by any number of applications of the omega-rule.* Both this paper and Japaridze's [27] work with Solovay-style functions that are defined on a finite initial segment of ω.

7. Beklemishev gave in [9] a simplified argument using the logic J, which is very well-behaved. This proof considers a family of N Solovay functions h_n with domain ω, where N is the number of modal operators appearing in our "target formula" ϕ and where each $h_{n+1}(0)$ is defined to be the limit of the previous h_n function. In a sense, these h_n thus glue together to a global Solovay function on $\omega \cdot N$.

Despite Ignatiev's provability predicates being quite general, they do not include our interpretation, as for example it is assumed that they are of increasing syntactical complexity whereas our iterated provability classes are all given by a single Π^1_1 formula. Although we do not have a syntactical difference between $[\lambda]^\Lambda_T \phi$ and $[\mu]^\Lambda_T \phi$ for $\lambda < \mu$, it is clear that in extension the formulas do increase in complexity; that is, the complexity of the hyperarithmetical sets of formulas provable using the different predicates increases with the ordinals (see e.g. [2]).

Thus we will instead work with uniform provability predicates. For the proof to go smoothly, we need a technical assumption on such predicates:

Definition 9.1. *Let π be a Λ-uniform proof predicate over a theory T. We say that π is* normalized *if it is provable in T that for every λ we have that every λ-derivable formula has infinitely many λ-derivations and, whenever $[c : \lambda]_\pi \phi$ and $[c : \lambda]_\pi \psi$, it follows that $\phi = \psi$; in other words, every derivation must be a derivation of a single formula.*

Observe that if π is a Λ-uniform proof predicate such that $[c : \xi]_\pi \phi$ and $[c : \lambda]_\pi \psi$, it also follows that $\phi = \psi$. For, without loss of generality, we may assume that $\xi \leq_\Lambda \lambda$, and thus by Definition 7.7.4, $[c : \lambda]_\pi \phi$ as well, so that $\psi = \phi$.

As mentioned in Section 2, there are natural ways of obtaining normalized proof predicates, for example letting c be a sequence of formulas which are taken to prove only the last formula and allowing for redundant steps in proofs. But we may also enforce this condition in an ad-hoc way:

Lemma 9.2. *Given a Λ-uniform proof predicate π for a theory T, there is a normalized proof predicate π' such that $\mathrm{RCA}_0 \vdash \forall\lambda\forall\phi([\lambda]_\pi \phi \leftrightarrow [\lambda]_{\pi'} \phi)$.*

Proof. Define $\pi'(c, \lambda, \phi)$ if and only if there are c' and m such that $c = \langle c', \phi, m\rangle$ and $\pi(c', \lambda, \phi)$. Note that the parameter m works as "padding" to lengthen proofs. $\square$

Thus we may freely restrict our attention to normalized proof predicates.

10. Arithmetical completeness of $\mathbf{GLP_\Lambda}$

The completeness proof we present here is the first that considers well-orders beyond ω (excluding [3, 4], which as we mentioned apply to different modal logics) and combines ideas from [17] and [9] by considering finite paths

over a polymodal J-frame. We do so by introducing an additional trick, which is to work with all modalities simultaneously, where our path makes a λ_n-step whenever appropriate. Readers familiar with known proofs might find it surprising that this is not problematic, but indeed it isn't and otherwise the argument proceeds as in other settings. As always, we will mimic a Kripke structure using arithmetic formulas and define our arithmetic interpretation based on them.

Since GLP_Λ is Kripke-incomplete, we will resort to J-models instead. These models are related to GLP_ω as described in Section 8. The step from GLP_Λ to GLP_ω is provided by the following easy lemma which is also given in [10].

Lemma 10.1. *Let ϕ be a GLP_Λ formula and $\lambda_0 <_\Lambda \ldots <_\Lambda \lambda_{N-1}$ be all the elements of Λ appearing in ϕ. By ϕ^{c} we denote the* condensation *of ϕ that arises by simultaneously replacing each occurrence of $[\lambda_i]$ by $[i]$. Then, we have that*

$$\mathsf{GLP}_\omega \vdash \phi^{\mathsf{c}} \implies \mathsf{GLP}_\Lambda \vdash \phi.$$

Proof. Given a GLP_ω-derivation d of ϕ^{c} we may replace every occurrence of $[n]$ in d by $[\lambda_n]$, thus obtaining a derivation of ϕ. $\qquad\square$

With this lemma at hand we may give an outline of the proof of our completess theorem, which reads as follows.

Theorem 10.2. *If Λ is a computable linear order, T is any sound, representable theory extending RCA_0, π is a sound, normalized, Λ-uniform proof predicate over T and ϕ is any $\mathcal{L}_\Lambda$-formula, $\mathsf{GLP}_\Lambda \vdash \phi$ if and only if, for every arithmetic interpretation f, $T \vdash f_\pi(\phi)$.*

Proof plan. One direction is soundness and has already been established in Lemma 7.10.

For the other, if ϕ is GLP_Λ-consistent then by Lemma 10.1, ϕ^{c} is GLP_ω-consistent. Let N be the number of modalities appearing in ϕ. By Lemma 8.7, there is a finite J_N-model

$$\mathfrak{W} = \langle W, \langle >_n \rangle_{n<N}, [\![\cdot]\!] \rangle$$

with root 0 such that $\mathfrak{W}, 0 \Vdash \langle 0 \rangle \phi^{\mathsf{c}}$ (so that $\mathfrak{W}, 1 \Vdash \phi^{\mathsf{c}}$ for some world $1 \neq 0$) and $\mathfrak{W} \Vdash M(\phi^{\mathsf{c}})$. We will assign to each $w \in W$ an arithmetic sentence σ_w so that the formulas $\boldsymbol{\sigma} = \langle \sigma_w \rangle_{w \in W}$ are a "snapshot" of $\mathfrak{W}$. We will make this

precise in Definition 10.4, but let us outline the essential properties that we need from $\boldsymbol{\sigma}$.

First, we need for the arithmetic interpretation f that sends a propositional variable p to $f(p) := \bigvee_{w \in [\![p]\!]} \sigma_w$ to have the properties that

$$\begin{aligned}
\mathfrak{W}, w \Vdash \psi^{\mathsf{c}} &\iff T \vdash \sigma_w \to f_\pi(\psi) \\
\mathfrak{W}, w \not\Vdash \psi^{\mathsf{c}} &\iff T \vdash \sigma_w \to \neg f_\pi(\psi)
\end{aligned} \tag{6}$$

for each $w \in W \setminus \{0\}$ and each subformula ψ of ϕ. In particular, we have that $T \vdash \sigma_1 \to \neg f_\pi(\phi)$ from which we obtain

$$T \vdash \Diamond_T \sigma_1 \to \neg \Box_T f_\pi(\phi). \tag{7}$$

Our desired result will follow if the formulas $\boldsymbol{\sigma}$ satisfy two more properties: the first is that
$$T \vdash \sigma_0 \to \Diamond_T \sigma_1,$$
and the second, that $\mathbb{N} \models \sigma_0$. By the assumption that T is sound we conclude that $\mathbb{N} \models \neg \Box_T f_\pi(\phi)$. Hence, $f_\pi(\phi)$ is not provable in T which is what was to be shown. $\qquad \square$

Before we proceed to give the details needed to complete the proof we state as an easy consequence of our arithmetic completeness theorem the following corollary, which was also proven by purely modal means in [10].

Corollary 10.3. *Given a computable well-order Λ and an $\mathcal{L}_\Lambda$-formula ϕ we have that*
$$\mathsf{GLP}_\Lambda \vdash \phi \iff \mathsf{GLP}_\omega \vdash \phi^{\mathsf{c}}.$$

Proof. One direction is Lemma 10.1. For the other direction, suppose $\mathsf{GLP}_\omega \not\vdash \phi^{\mathsf{c}}$. By Theorem 10.2, we find an arithmetic interpretation f so that $\overline{\mathrm{ACA}_0}^\Lambda \not\vdash f_T^\Lambda(\phi)$. By the soundness theorem (Theorem 7.11), we conclude that $\mathsf{GLP}_\Lambda \not\vdash \phi$. $\qquad \square$

Before entering into further detail, we first say what it means for a collection of sentences $\boldsymbol{\sigma} = \langle \sigma_w \rangle_{w \in W}$ to be a snapshot of a Kripke structure inside a theory. Generally speaking, this means that each world w will be associated with an arithmetic sentence σ_w so that this sentence carries all the necessary information in terms of accessible worlds. Below, recall that all J_N-frames are assumed finite.

Definition 10.4. *Given a computable linear order Λ, a sequence*

$$\boldsymbol{\lambda} = \lambda_0 <_\Lambda \lambda_1 <_\Lambda \ldots <_\Lambda \lambda_{N-1},$$

a J_N-frame $\mathfrak{W} = \langle W, \langle >_n \rangle_{n<N} \rangle$ with root 0, and a Λ-uniform proof predicate π for some formal theory T, a family of sentences $\boldsymbol{\sigma} = \langle \sigma_w \rangle_{w \in W}$ is a $\boldsymbol{\lambda}$-snapshot of $\mathfrak{W}$ over π if

1. $T \vdash \displaystyle\bigwedge_{w \neq v \in W} \neg(\sigma_w \wedge \sigma_v)$;

2. $T + \sigma_w \vdash \langle \overline{\lambda_n} \rangle_\pi \sigma_v$ *for all $w \in W$, $n < N$ and $v <_n w$;*

3. *for all $n < N$ and for each world $w \neq 0$,*

$$T + \sigma_w \vdash [\overline{\lambda_n}]_\pi \bigvee_{v \lll_n w} \sigma_v;$$

4. $\mathbb{N} \models \sigma_0$.

If $\mathfrak{W}$, $\boldsymbol{\sigma}$, $\boldsymbol{\lambda}$, T are as above we will write $\boldsymbol{\sigma} : \mathfrak{W} \overset{\boldsymbol{\lambda}}{\hookrightarrow} \pi$.

The following result is proven in [9].

Lemma 10.5. *Suppose that $\boldsymbol{\sigma} : \mathfrak{W} \overset{\boldsymbol{\lambda}}{\hookrightarrow} \pi$, ϕ is an $\mathcal{L}_\Lambda$-formula with modalities amongst $\boldsymbol{\lambda}$ such that $\mathfrak{W} \Vdash M(\phi^{\mathsf{c}})$, and $f(p) := \bigvee_{w \in \llbracket p \rrbracket} \sigma_w$.*
Then, for all $0 \neq w \in W$ and every subformula ψ of ϕ,

1. *if $w \in \llbracket \psi^{\mathsf{c}} \rrbracket$ then $T + \sigma_w \vdash f_\pi(\psi)$;*

2. *if $w \notin \llbracket \psi^{\mathsf{c}} \rrbracket$ then $T + \sigma_w \vdash \neg f_\pi(\psi)$.*

Proof. By induction on the complexity of ψ. The cases for propositional variables and Booleans are straightforward, so we focus on $\psi = [\lambda_n]\theta$.

First we prove item 2. Suppose that $w \notin \llbracket ([n]\theta)^{\mathsf{c}} \rrbracket$. Then, there is $v \in W$ such that $v <_n w$ and $v \notin \llbracket \theta^{\mathsf{c}} \rrbracket$. By the induction hypothesis, $T + \sigma_v \vdash \neg f_\pi(\theta)$, hence by contrapositive, $T \vdash f_\pi(\theta) \to \neg \sigma_v$. By Definition 7.7.3, we have that $T \vdash [\overline{\lambda_n}]_\pi f_\pi(\theta) \to [\overline{\lambda_n}]_\pi \neg \sigma_v$; once again by contrapositive, $T \vdash \langle \overline{\lambda_n} \rangle_\pi \sigma_v \to \neg[\overline{\lambda_n}]_\pi f_\pi(\theta)$. It follows by Definition 10.4.2 that $T + \sigma_w \vdash \neg[\overline{\lambda_n}]_\pi f_\pi(\theta)$, and the latter is equal to $\neg f_\pi(\psi)$.

Item 1 is similar, but we use Lemma 8.4. If $w \in [\![\,([n]\theta)^{\mathsf{c}}\,]\!]$, then we also have that $v \in [\![\,\theta^{\mathsf{c}}\,]\!]$ for all $v \lll_n w$. By the induction hypothesis we have that $T + \sigma_v \vdash f_\pi(\theta)$ for all $v \lll_n w$, and using Definition 10.4.3 we easily obtain $T + \sigma_w \vdash f_\pi(\psi)$. $\hfill\square$

In the next section, we shall mainly see how to produce snapshots of a given Kripke model $\mathfrak{W}$ over a uniform proof predicate π. We define the corresponding sentences σ_w for $w \in W$ in a standard way as limit statements about certain Solovay functions.

11. Solovay sequences and snapshots of Kripke models

In our presentation, the J_N-frame $\mathfrak{W} = \langle W, \langle >_n \rangle_{n<N} \rangle$ will be externally given and fixed beforehand. When reasoning within a formal theory, $x <_n y$ will be understood as shorthand for $\bigvee_{v <_n w}(x = \overline{v} \wedge y = \overline{w})$, and similarly for the relations $\approx_n, \lll_n$, etc. We will simulate quantifiers over W with (finite) conjunctions and disjunctions. We will usually write these conjunctions or disjunctions over W explicitly, but when this becomes impractical, a quantifier over W should be understood as an abbreviation (e.g. $\exists! w \in W\ \theta(\overline{w})$ should be understood to be described using conjunctions and disjunctions). An alternative presentation (which we will not pursue) is to use quantifiers, which would first require describing the frame $\mathfrak{W}$ within T.

A central notion to our proof is what we call a *Solovay sequence* or *path;* these sequences are given by a recursion based on a uniform provability predicate which depends on a parameter ϕ. Later we will choose an appropriate value of ϕ via a fixpoint construction. We shall use the following notation: $\mathtt{Seq}(x)$ is a Δ_0^0 formula stating that x codes a finite sequence of numbers, $\mathtt{last}(x)$ is a term that picks out the last element of x if it is non-empty (we may set $\mathtt{last}(\langle\rangle) = 0$), $x \sqsubseteq y$ is a Δ_0^0 formula that states that the sequence x is an initial segment of y, $|x|$ gives the length of x and x_y is a term which picks the y-coordinate of x. As in previous sections, it is not necessary to have these terms available in our language, as we can define their graphs and replace them by pseudo-terms, but we shall write them as such for simplicity of exposition.

Recall that throughout this paper we assume that we work with well-behaved implementations of our syntactical operations. In particular, we assume that all the basic properties will be provable in EA.

We will now define a formula $\mathtt{Lim}$ stating that the paths satsifying ϕ "converge" to w:

Definition 11.1. *Let ϕ be a formula with a single free number variable. Define*

$$\texttt{Lim}_\phi(w) \; := \; \exists s \left(\texttt{Seq}(s) \wedge \phi(s) \wedge \forall s' \sqsupseteq s \left(\phi(s') \to \texttt{last}(s') = w \right) \right).$$

We shall use these formulas to define our recursive paths.

Definition 11.2. *Using the notation of Definition 7.7, for a formula ϕ with a single free number variable, we define a formula $\texttt{Sol}(s|\phi)$ by*

$$
\begin{aligned}
\texttt{Sol}(s|\phi) := \; &\texttt{Seq}(s) \wedge s_0 = 0 \\
&\wedge \forall x{<}|s|{-}1 \bigwedge_{w \in W} \Bigg(s_x = \overline{w} \\
&\quad \to \Bigg(\Bigg(\Big(\bigwedge_{n<N} \bigwedge_{v<_n w} \langle x : \lambda_n \rangle_\pi \texttt{Lim}_\phi(\overline{v}) \Big) \to s_{x+1} = \overline{w} \Bigg) \\
&\qquad \wedge \bigwedge_{n<N} \bigwedge_{v<_n w} \Big(\big([x : \lambda_n]_\pi \neg\texttt{Lim}_\phi(\overline{v}) \big) \to s_{x+1} = \overline{v} \Big) \Bigg) \Bigg).
\end{aligned}
$$

With this we can say what it means to be a Solovay path.

Definition 11.3 (Solovay path)**.** *Using the fixpoint theorem on the parameter ϕ in $\texttt{Sol}(s|\phi)$, we define $\texttt{Sol}(s)$ to be a formula satisfying*

$$\mathrm{RCA}_0 \vdash \forall x \Big(\texttt{Sol}(x) \leftrightarrow \texttt{Sol}(x | \ulcorner \texttt{Sol}(s) \urcorner) \Big).$$

We say s is a Solovay path *if $\texttt{Sol}(s)$ holds. Further, we say w is a* Solovay value *at i if there is a Solovay path s with $s_i = w$, and formalize this by*

$$\texttt{sol}(i) = w \; := \; \exists s \left(\texttt{Sol}(s) \wedge |s| > i \wedge s_i = w \right).$$

Finally, say that w is a limit Solovay value *if every Solovay path that is long enough has w as its last value, formalized by*

$$\texttt{Lim}(w) \; := \; \texttt{Lim}_{\texttt{Sol}}(w).$$

Below we list some derivable properties of Solovay paths. Intuitively, item 1 states that Solovay paths are linearly ordered by $\sqsubseteq$, item 2 that there

are arbitrarily long Solovay paths, item 3 that Solovay values are uniquely determined, item 4 that Solovay paths are $\lll_0$-decreasing and, more generally, Item 5 states that Solovay paths are $\lll_n$-decreasing if whenever $m < n$ and $u <_m v$, it is λ_m-consistent that u is the Solovay limit.

Lemma 11.4. *Under the assumptions of Theorem 10.2, it is derivable in T that*

1. *$\forall s \forall s' (\texttt{Sol}(s) \wedge \texttt{Sol}(s') \to s \sqsubseteq s' \vee s' \sqsubseteq s)$;*

2. *$\forall i \exists s \ (\ \texttt{Sol}(s) \wedge |s| > i)$;*

3. *$\forall i \exists! w \in W \ \ \texttt{sol}(i) = \overline{w}$;*

4. *$\bigwedge_{w,v \in W} \forall i \, \forall j < i \ \left(\texttt{sol}(j) = \overline{v} \wedge \texttt{sol}(i) = \overline{w} \to \overline{w} \lll_0 \overline{v} \right)$;*

5. *$\bigwedge_{w,v \in W} \forall i \, \forall j < i \ \Big(\left(\texttt{sol}(j) = \overline{v} \wedge \texttt{sol}(i) = \overline{w} \right.$*

$$\wedge \bigwedge_{m<n} \bigwedge_{u <_m v} \langle \lambda_m \rangle_\pi \, \texttt{Lim}(\overline{u}) \,) \ \to \overline{w} \lll_n \overline{v} \Big).$$

6. *$\bigwedge_{w \approx_n v} \forall i \, \forall j < i \ \left(\left(\texttt{sol}(j) = \overline{v} \wedge \texttt{sol}(i) = \overline{w} \right) \to \overline{w} \lll_n \overline{v} \right).$*

Proof.

1. Clearly it suffices to prove that

$$\left(\texttt{Sol}(s) \wedge \texttt{Sol}(s') \wedge i < |s| \wedge i < |s'| \right) \to s_i = s'_i,$$

for then if s, s' are any two paths and, say, $|s| \leq |s'|$, it follows that $s_i = s'_i$ for all $i < |s|$ and thus $s \sqsubseteq s'$. Moreover, this formula is $\Delta^0_0(\pi)$ and hence, in view of Definition 7.7.1, we may proceed by induction on i.

The base case is trivial since $s_0 = s'_0 = 0$. For the inductive step, we assume $w = s_i = s'_i$. Then, we must have that either $[i : \lambda_n]_\pi \neg \texttt{Lim}(\overline{v})$ holds for some n, $v <_n w$, or it does not. If it does, then the value of v is uniquely determined (as i is the code of a derivation of at most one formula) and thus $s_{i+1} = s'_{i+1} = v$. Otherwise, $\bigwedge_n \bigwedge_{v <_n w} \langle x : \lambda_n \rangle_\pi \texttt{Lim}(\overline{v})$ holds and $s_{i+1} = s'_{i+1} = w$.

2. The proof follows the above structure; here we observe that if s is a Solovay path, we may always add one extra element to s depending on which condition is met.

3. This is immediate from items 1 and 2.

4. By the recursive definition of a Solovay path, if s is a Solovay path with $k+1 < |s|$, it is always the case that $s_{k+1} \ll_0 s_k$. Since $\ll_0$ is transitive, this implies inductively that $s_i \ll_0 s_j$ whenever $j < i$, and this induction can be easily formalized in T.

5. This is a simple generalization of the previous item. Let s be a Solovay path such that $j < |s|$. The assumption that $\bigwedge_{m<n} \bigwedge_{u<_m v} \langle \lambda_m \rangle_\pi \mathtt{Lim}(\overline{u})$ implies that any possible first step in s away from s_j is to a world $s_k \ll_n s_j$. Moreover, the same holds for possible further steps since $\ll_n$ is transitive (Lemma 8.2.3), and since $u <_m s_k \ll_n s_j$ implies $u <_m s_j$ (Lemma 8.2.4). Again, by induction we obtain that $s_i \ll_n s_j$ whenever $j < i$.

6. This follows a similar induction as the previous item. Assume that $s_j \approx_n s_i$, and towards a contradiction suppose that $s_i \ll_n s_j$ fails. By the transitivity of $\ll_n$, this can only be if $s_{k+1} <_m s_k$ for some $k \in [j, i)$ and some $m < n$. Moreover, we may assume that k is minimal, so that $s_k \ll_n s_j$, and thus $s_k \approx_n s_i$. By Lemma 8.2.4, $s_{k+1} <_m s_i$ and thus $s_{k+1} \ll_0 s_i$, but by item 4, we also have that $s_i \ll_0 s_{k+1}$, contradicting the irreflexivity of $\ll_0$. We conclude that there can be no such k, and hence $s_i \ll_n s_j$. $\square$

It should be remarked that ACA_0 proves that all pairs $\langle i, w \rangle$ such that $\mathtt{sol}(i) = w$ form a set, provided π is comprehensible. However, in this section we do not assume that T extends ACA_0, so we will work directly with the Σ^0_1 formula $\mathtt{sol}(i) = w$. It will be useful to note that its complexity can be lowered somewhat.

Corollary 11.5. *Under the assumptions of Theorem 10.2, it is provable in T that*

$$\bigwedge_{w \in W} \left(\mathtt{sol}(i) = \overline{w} \ \leftrightarrow \ \forall s \left(\mathtt{Sol}(s) \wedge |s| > i \to s_i = \overline{w} \right) \right).$$

Proof. Immediate from Lemma 11.4.2. $\square$

Thus, $\mathtt{sol}(i) = w$ is provably $\Delta^0_1(\pi)$. In fact, we can do better than this: since we only need to exhibit a path s such that $\mathtt{Sol}(s) \wedge |s| > i$ holds, we can assume, under standard assumptions on the coding of sequences, that such a sequence is bounded elementarily in $i + |W|$. Thus, the quantification over s can be bounded in $\mathtt{sol}(i) = w$, and we can write it as a $\Delta^0_0(\pi)$ formula. However, the $\Delta^0_1(\pi)$ presentation will suffice for our purposes.

Lemma 11.6. *Under the assumptions of Theorem 10.2, it is derivable in T that*

1. *A Solovay limit exists:*
$$\bigvee_{w \in W} \mathtt{Lim}(\overline{w}).$$

2. *More specifically, that if $w \in W$ is a Solovay value, then there is $v \lll_0 w$ such that v is a Solovay limit:*
$$\bigwedge_{w \in W} \forall i \left(\mathtt{sol}(i) = \overline{w} \to \bigvee_{v \lll_0 w} \mathtt{Lim}(\overline{v}) \right).$$

Proof. The first item follows from the second, if we observe that $\mathtt{sol}(0) = 0$ trivially holds since $\langle 0 \rangle$ is a Solovay path (of length one). Thus we focus on the second.

Here, we proceed by external induction on $\lll_0$; to be precise, let $w \in W$ and suppose that
$$T \vdash \forall i \left(\mathtt{sol}(i) = \overline{u} \to \bigvee_{v \lll_0 u} \mathtt{Lim}(\overline{v}) \right)$$
whenever $u \lll_0 w$.

Suppose that w is a Solovay value, so that for some i we have that $\mathtt{sol}(i) = w$, and let s be a Solovay path with $|s| > i$ and $s_i = w$. Observe that, by Lemma 11.4.4, we have that whenever $j > i$ and $\mathtt{sol}(j) = v$ holds, then $v \lll_0 w$.

Here we consider two cases. First assume that there are $s' \sqsupseteq s$, $j > i$ and $u \neq w$ such that $s'_j = u$; since $u \neq w$ we have that $u \lll_0 w$, and by our induction hypothesis $\bigvee_{v \lll_0 u} \mathtt{Lim}(\overline{v})$, which implies $\bigvee_{v \lll_0 w} \mathtt{Lim}(\overline{v})$.

Now assume that there is no such s'. But this means that whenever $s' \sqsupseteq s$ we must have $\mathtt{last}(s') = w$, i.e. $\mathtt{Lim}(\overline{w})$ holds, and once again this implies $\bigvee_{v \lll_0 w} \mathtt{Lim}(\overline{v})$, as desired. $\square$

Lemma 11.7. *Under the assumptions of Theorem 10.2, if $w \in W$ and $n < N$, then:*

1. $T \vdash \displaystyle\bigwedge_{w \in W} \forall k \left(\mathtt{sol}(k) = \overline{w} \to [\overline{\lambda}_n]_\pi \mathtt{sol}(\dot{k}) \approx_{n+1} \overline{w} \right);$

2. *if $m \leq n$ and $v >_m w$,*

$$T \vdash \bigwedge_{w,v \in W} \forall k \Big(\mathtt{sol}(k) = \overline{v} \wedge \mathtt{sol}(k+1) = \overline{w} \to [\overline{\lambda}_n]_\pi \, \mathtt{sol}(\dot{k}+1) = \overline{w} \Big).$$

Proof. Reasoning within T, we will prove claim 1 by induction on k; notice that $[\lambda]_\pi \phi$ is $\Sigma_1^0(\pi)$ and $\mathtt{sol}(x) = y$ is $\Delta_1^0(\pi)$, and hence the induction formula is $\Sigma_1^0(\pi)$. The case $k = 0$ is trivial, as $s = \langle 0 \rangle$ is a Solovay path with $s_0 = 0$. For the inductive step, assume that the claim is true for k and let s be a Solovay path with $|s| > k+1$ and $s_{k+1} = w$. Meanwhile, for $v = s_k$, by our induction hypothesis

$$[\lambda_n]_\pi \forall x (\mathtt{Sol}(x) \wedge |x| > \overline{k} \to x_{\overline{k}} \approx_{n+1} \overline{v}). \tag{8}$$

Consider two cases.

Case 1. Suppose that $s_{k+1} <_m s_k$ for some $m \leq n$. Note that this m is unique since the accessibility relations are disjoint, so we must have that $[k : \lambda_m]_\pi \neg \mathtt{Lim}(\overline{w})$ holds, and by Properties 4 and 5 of Definition 7.7, we also have $[\lambda_n]_\pi [\overline{k} : \overline{\lambda}_m]_\pi \neg \mathtt{Lim}(\overline{w})$.

Then, $w <_m v$ implies that $w <_m u$ whenever $u \approx_{n+1} v$ by the J-frame Condition 2 and the definition of a Solovay path. Thus, reasoning within $[\lambda_n]_\pi$, since $\bigvee_{u \approx_{n+1} v} x_k = \overline{u}$ holds for any sufficiently long Solovay path x, we must have that $\overline{w} <_m x_{\overline{k}}$, so that

$$[\lambda_n]_\pi \forall x \Big(\mathtt{Sol}(x) \wedge |x| > \overline{k} + 1 \to x_{\overline{k}+1} = \overline{w} \Big).$$

The claim immediately follows from the facts that $w \approx_{n+1} w$ and that x was arbitrary.

Case 2. Suppose that for no $m \leq n$ do we have that $s_{k+1} <_m s_k$. Recall that $v = s_k$, so that according to the Solovay recursion, $\langle k : \lambda_m \rangle_\pi \mathtt{Lim}(\overline{u})$ holds whenever $u <_m v$.

Now, reason within $[\lambda_n]_\pi$, and let x be such that $\mathtt{Sol}(x) \wedge |x| > i$. Let $m \leq n$ and $u \in W$. If $u \not<_m v$, from $x_k \approx_{n+1} \overline{v}$ we obtain $\overline{u} \not<_m x_k$ as in case 1, while if $u <_m v$ we see using the assumption that π is a uniform provability predicate (Definition 7.7.6) that $\langle \overline{k} : \overline{\lambda}_m \rangle_\pi \mathtt{Lim}(\overline{u})$ is derivable. It follows by the Solovay recursion that $x_{k+1} = x_k$ or $x_{k+1} <_j x_k$ for some $j > n$, so that in either case $x_{k+1} \lll_{n+1} x_k$.

47

Thus,

$$x_{k+1} \lll_{n+1} x_k \overset{\text{IH}}{\approx}_{n+1} \overline{s_k} \ggg_{n+1} \overline{s_{k+1}}.$$

This implies that $x_{k+1} \approx_{n+1} \overline{s_{k+1}}$. Since $s_{k+1} = w$ and this reasoning was realized within $[\lambda_n]_\pi$, we conclude that

$$[\lambda_n]_\pi \forall x \Big(\mathtt{Sol}(x) \wedge |x| > \overline{k} + 1 \to x_{\overline{k}+1} \approx_{n+1} \overline{w} \Big),$$

from which the first claim clearly follows.

Claim 2 is a direct consequence of claim 1. Suppose that $m \le n$, $v >_m w$, $\mathtt{sol}(k) = \overline{v}$ and $\mathtt{sol}(k+1) = \overline{w}$. Then, by Claim 1, $[\lambda_n]_\pi \mathtt{sol}(\overline{k}) \approx_{n+1} \overline{v}$, and by Properties 4 and 5 of Definition 7.7, $[\lambda_n]_\pi [\overline{k} : \overline{\lambda_m}]_\pi \neg \mathtt{Lim}(\overline{w})$.

Let x be a Solovay path with $|x| > k+1$. Recall that we have set $w = s_{k+1}$. Then, $w <_m v$ together with $x_k \approx_{n+1} v$ imply that $w <_m x_k$ by Lemma 8.2.4, and thus using the definition of a Solovay path, $x_{k+1} = w$. Formalizing this within $[\lambda_n]_\pi$ we obtain

$$[\lambda_n]_\pi \forall x \Big(\mathtt{Sol}(x) \wedge |x| > \overline{k} + 1 \to x_{\overline{k}+1} = \overline{w} \Big).$$

By Lemma 11.4.2, there provably exists a Solovay path x with $|x| > k + 1$, so that $[\lambda_n]_\pi \mathtt{sol}(\overline{k} + 1) = \overline{w}$, as claimed. $\qquad \square$

From here on, it remains to show that the formulas $\mathtt{Lim}(\overline{w})$ give a snapshot of our Kripke model.

Lemma 11.8. *Under the assumptions of Theorem 10.2, if $n < N$ and $v <_n w$, then*

$$T \quad + \quad \mathtt{Lim}(\overline{w}) \quad \vdash \quad \langle \overline{\lambda_n} \rangle_\pi \ \mathtt{Lim}(\overline{v}).$$

Proof. Reason within $T + \mathtt{Lim}(\overline{w})$ and, towards a contradiction, suppose that $[\lambda_n]_\pi \neg \mathtt{Lim}(\overline{v})$, so that there exists some i which satisfies $[i : \lambda_n]_\pi \neg \mathtt{Lim}(\overline{v})$. Now, by assumption, $\mathtt{Lim}(w)$ holds, and hence we may choose s such that $\mathtt{last}(s') = w$ for all Solovay paths $s' \sqsupseteq s$. Since π is normalized, every derivable formula has arbitrarily large derivations, and thus we may pick $j > |s|$ such that $[j : \lambda_n]_\pi \neg \mathtt{Lim}(\overline{v})$ holds and, in view of Lemma 11.4.2, a Solovay path s' with $|s'| > j + 1$. Then, $s'_j = s'_{j+1} = w$, but by the Solovay recursion we should have $s'_{j+1} = v$, a contradiction. $\qquad \square$

Lemma 11.9. *Under the assumptions of Theorem 10.2, if $w \neq 0$ and $n < N$, then*

$$T \; + \; \mathtt{Lim}(\overline{w}) \;\; \vdash \;\; [\overline{\lambda}_n]_\pi \bigvee_{v \lll_n w} \mathtt{Lim}(\overline{v}).$$

Proof. We reason in $T + \mathtt{Lim}(\overline{w})$. Let s be a Solovay path with $w = \mathtt{last}(s)$. Let $k_* < |s|$ be the greatest value such that $s_{k_*} \not\approx_{n+1} w$; note that k_* is well-defined since $s_0 = 0 \not\approx_{n+1} w$ and $s_k = w$ for k large enough. Note also that $s_{k_*+1} <_{m_*} s_{k_*}$ for some $m_* \leq n$ (otherwise $s_{k_*+1} \approx_{n+1} s_{k_*}$ and k_* would not be maximal), and that by Lemma 11.4.6, $w \lll_{n+1} s_{k_*+1}$.

By Lemma 11.7.2,

$$[\lambda_n]_\pi \forall x \left(\mathtt{Sol}(x) \wedge |x| > \overline{k}_* + 1 \;\to\; x_{\overline{k}_*+1} = \overline{s_{k_*+1}} \right). \tag{9}$$

Let $m < n$ and $v <_m s_{k_*+1}$. Since $s_{k_*+1} \approx_{n+1} w$, we also have $v <_m w$, so that $\langle \lambda_m \rangle_\pi \mathtt{Lim}(\overline{v})$ holds by Lemma 11.8. By Definition 7.7.7 we also have

$$[\lambda_n]_\pi \bigwedge_{m<n} \bigwedge_{v<_m s_{k_*+1}} \langle \overline{\lambda}_m \rangle_\pi \mathtt{Lim}(\overline{v}).$$

By Lemma 11.4.5 this implies that

$$[\lambda_n]_\pi \forall x \forall j \left(\overline{k}_* + 1 < j \wedge \mathtt{Sol}(x) \wedge |x| > j \to x_j \lll_n x_{\overline{k}_*+1} \right). \tag{10}$$

Putting (9) and (10) together, along with the fact that if $\mathtt{Sol}(x)$ and $|x| > k_* + 1$ then $\mathtt{last}(x) \lll_n s_{k_*+1} \ggg_{n+1} w$, we see that

$$[\lambda_n]_\pi \; \forall x \left(\mathtt{Sol}(x) \wedge |x| > \overline{k}_* + 1 \to (\mathtt{last}(x) \lll_n \overline{w} \vee \mathtt{last}(x) = \overline{s_{k_*+1}}) \right).$$

We now claim that this implies

$$[\lambda_n]_\pi \left(\mathtt{Lim}(\overline{s_{k_*+1}}) \vee \bigvee_{v \lll_n w} \mathtt{Lim}(\overline{v}) \right). \tag{11}$$

By Lemma 11.6.1, $\mathtt{Lim}(v_*)$ holds for some v_*, so we can pick a Solovay path r whose last element is v_* and such that $r' \sqsupseteq r$ implies that $\mathtt{last}(r') = v_*$ for any Solovay path r'. In particular we can choose r' such that $|r'| > k_* + 1$ using Lemma 11.4.2, so that either $\mathtt{last}(r') \lll_n w$ or $\mathtt{last}(r') = s_{k_*+1}$. It follows that $v_* \lll_n w$ or $v_* = s_{k_*+1}$, and formalizing this within T we obtain (11).

It remains to show that $[\lambda_n]_\pi \neg \mathtt{Lim}(\overline{s_{k_*+1}})$. But in view of the Solovay recursion we must have that $[k_* : \lambda_{m_*}]_\pi \neg \mathtt{Lim}(\overline{s_{k_*+1}})$. In view of Property 4 of Definition 7.7 and the fact that $\lambda_{m_*} \leq_\Lambda \lambda_n$ this implies that $[\lambda_n]_\pi \neg \mathtt{Lim}(\overline{s_{k_*+1}})$, as required. $\square$

We are now ready to prove that the formulas $\texttt{Lim}(\overline{w})$ provide a snapshot of our J-model.

Lemma 11.10. *Under the assumptions of Theorem 10.2, define*

$$\overrightarrow{\texttt{Lim}} = \langle \texttt{Lim}(\overline{w}) : w \in W \rangle .$$

Then, $\overrightarrow{\texttt{Lim}}$ *is a* $\boldsymbol{\lambda}$*-snapshot of* $\mathfrak{W}$ *over* π*, that is,* $\overrightarrow{\texttt{Lim}} : \mathfrak{W} \overset{\lambda}{\hookrightarrow} \pi$.

Proof. We must check each of the conditions of Definition 10.4.

1. For the first, reasoning within T, suppose that $\texttt{Lim}(\overline{w})$ and $\texttt{Lim}(\overline{v})$ hold. In view of $\texttt{Lim}(\overline{w})$, pick a Solovay path s such that any extension of s has last element w and similarly s' such that any extension of s' has last element v. By Lemma 11.4.1, either $s \sqsupseteq s'$ or $s' \sqsupseteq s$; in either case, it follows by $\texttt{Lim}(\overline{w}) \wedge \texttt{Lim}(\overline{v})$ that $w = v$.

2-3. The second condition is Lemma 11.8 and the third, Lemma 11.9.

4. For the fourth, we must use the fact that $[\lambda]_\pi$ is sound for all λ. Towards a contradiction, suppose that $\texttt{Lim}(\overline{w})$ holds for some $w \neq 0$ and let s be a Solovay path with $\texttt{last}(s) = w$. Then, there exist k such that $s_k \neq w$ but $s_{k+1} = w$. This implies that $[k : \lambda_n]_\pi \neg \texttt{Lim}(\overline{w})$ for some $n < N$. But by soundness we conclude that $\neg \texttt{Lim}(\overline{w})$, a contradiction. $\square$

We may finally prove our main completeness result.

Proof of Theorem 10.2. We have already seen that GLP_Λ is sound for its arithmetic interpretation. For the other direction, if ϕ is consistent over GLP_Λ, then by Lemma 8.5, ϕ^c is consistent over GLP_ω and thus by Lemma 8.7, there is a finite J_N-model $\mathfrak{W}$ with root 0 such that $\mathfrak{W}, 0 \Vdash \Diamond \phi^\mathsf{c}$ and $\mathfrak{W} \Vdash M(\phi^\mathsf{c})$, so that $\mathfrak{W}, 1 \Vdash \phi^\mathsf{c}$ for some world $1 \neq 0$. Let $\boldsymbol{\lambda}$ be the modalities appearing in ϕ. By Lemma 11.10,

$$\overrightarrow{\texttt{Lim}} : \mathfrak{W} \overset{\lambda}{\hookrightarrow} \pi,$$

so that by Lemma 10.5.1, $T + \texttt{Lim}(\overline{1}) \vdash f_\pi(\phi)$. Hence, by Items 2 and 4 of Definition 10.4, $\mathbb{N} \models \Diamond_T f_\pi(\phi)$, i.e. $f_\pi(\phi)$ is consistent with T. $\square$

Appendix A. Alternative provability predicates

We have already discussed how the provability predicate $[\lambda]_T^\Lambda$ may be replaced by its 'existential' counterpart $\widehat{[\lambda]}_T^\Lambda$. In this section we shall briefly consider some more possible variants of our provability predicates. We do so in an informal setting and in particular shall refer to defining recursions rather than formalizations in second-order arithmetic. Moreover, we shall not be too concerned about the amount of transfinite induction needed in the arguments. Also, we shall assume that Λ is sufficiently well-behaved by, for example, demanding it satisfies a theory of natural of well-orderings within T, such as *NWO* from [5].

Appendix A.1. A wider notion of iterated provability

We could consider an apparently slightly more liberal notion of α-provability – let us write $[\alpha]_T^{\Lambda,w}$ – defined by the following recursion:

$$[\alpha]_T^{\Lambda,w}\phi \ :\Leftrightarrow\ \Box_T\phi \vee \exists\psi\,\exists\,\beta <_\Lambda \alpha(\forall n\,[\beta]_T^{\Lambda,w}\psi(\overline{n}) \wedge [\beta]_T^{\Lambda,w}(\forall x\,\psi(x) \to \phi)).$$

However, it is easy to see by transfinite induction that $[\alpha]_T^{\Lambda,w}\phi \ \Leftrightarrow\ [\alpha]_T^\Lambda\phi$. The $\Leftarrow$ direction is obvious. For the other direction we assume that we can formalize the notion of $[\alpha]_T^{\Lambda,w}$ just like $[\alpha]_T^\Lambda$ and prove all the necessary lemmata like monotonicity, distribution axioms, etc. Suppose that

$$\forall n\,[\beta]_T^{\Lambda,w}\psi(\overline{n}) \wedge [\beta]_T^{\Lambda,w}(\forall x\,\psi(x) \to \phi)$$

for some formula ψ and ordinal $\beta <_\Lambda \alpha$. Then, clearly also

$$\forall n\,[\beta]_T^{\Lambda,w}\Big(\psi(\overline{n}) \wedge (\forall x\,\psi(x) \to \phi)\Big).$$

But, as

$$[0]_T^\Lambda\Big(\forall x\,(\psi(x) \wedge (\forall x\,\psi(x) \to \phi)) \to \phi\Big),$$

we get

$$\forall n\,[\beta]_T^{\Lambda,w}\Big(\psi(\overline{n}) \wedge (\forall x\,\psi(x) \to \phi)\Big) \wedge \Box_T\Big(\forall x\,(\psi(x) \wedge (\forall x\,\psi(x) \to \phi)) \to \phi\Big),$$

which by the induction hypothesis for β implies $[\alpha]_T^\Lambda\phi$.

Appendix A.2. Diagonializing at limits

Note that in our definition of $[\alpha]_T^\Lambda$ there is still some uniformity present in that we choose one particular $\beta <_\Lambda \alpha$ with $[\beta]_T^\Lambda \psi(\overline{n})$ for all numbers n. We can make this β also dependent on n. In a sense, this boils down to diagonalizing at limit ordinals. Thus, we define our notion of $[\alpha]_T^{\Lambda,d}$ as follows:

$$[\alpha]_T^{\Lambda,d}\phi \quad :\Leftrightarrow \quad \Box_T\phi \ \lor \ \exists\psi\left(\forall n\, \exists\beta_n<_\Lambda\alpha \ [\beta_n]_T^{\Lambda,d}\psi(\overline{n}) \ \land \ \Box_T(\forall x\, \psi(x) \to \phi)\right).$$

By an argument similar as before, we see that, also in this notion we can replace the $\Box(\forall x\, \psi(x) \to \phi)$ by $\exists\gamma<_\Lambda\alpha \ [\gamma]_T^{\Lambda,d}(\forall x\, \psi(x) \to \phi)$ without gaining any strength. This new notion of provability is related to $[\alpha]_T^\Lambda$ in a simple fashion as is expressed in Lemma Appendix A.2 below. Again, we assume that we can formalize the notion $[\alpha]_T^{\Lambda,d}$ in a suitable way so that the basic properties are provable. We first state a simple but useful observation.

Lemma Appendix A.1.

1. $[\alpha + 1]_T^\Lambda\phi \ \Leftrightarrow \ \exists\psi\left(\forall n \ [\alpha]_T^\Lambda\psi(\overline{n}) \ \land \ \Box(\forall x\psi(x) \to \phi)\right)$

2. $[\alpha + 1]_T^{\Lambda,d}\phi \ \Leftrightarrow \ \exists\psi\left(\forall n \ [\alpha]_T^{\Lambda,d}\psi(\overline{n}) \ \land \ \Box(\forall x\psi(x) \to \phi)\right).$

Proof. This follows directly from the definition and monotonicity. $\qquad\square$

Lemma Appendix A.2.

1. $[n]_T^\Lambda\phi \ \Leftrightarrow [n]_T^{\Lambda,d}\phi$ *for* $n \in \omega$;

2. $[\alpha + 1]_T^\Lambda\phi \ \Leftrightarrow [\alpha]_T^{\Lambda,d}\phi$ *for* $\alpha \geq_\Lambda \omega$, *and*

3. $[\lambda]_T^\Lambda\phi \ \Leftrightarrow \exists\gamma < \lambda \ [\gamma]_T^{\Lambda,d}\phi$ *for limit* λ.

Proof. The proofs proceed by induction on n and α, respectively, and we omit them. $\qquad\square$

As can be seen, there is a fair amount of freedom in defining transfinite iterations of the ω-rule. We chose the current paper's presentation both for the sake of simplicity and because a more refined hierarchy is in general terms more convenient; after all, it is easy to remove intermediate operators later if they are not needed. We also suspect it will be the appropriate notion useful later for a Π_1^0-ordinal analysis of second-order arithmetics, a goal which now seems within our reach.

Appendix A.3. Restricting iterated provability classes

While we have treated the parameter Λ as constant, we may also vary it in our provability predicates. This can be especially useful when our base theory does not establish $\exists X\, \mathtt{IPC}_T^\Lambda(X)$ but it does prove $\exists X\, \mathtt{IPC}_T^\Xi(X)$ whenever Ξ is a proper initial segment of Λ.

In such a case, we may define a 'restricted' version of our uniform provability predicate as follows. First, for $\lambda < \Lambda$ define $\Lambda \restriction \lambda$ by restricting Λ to $\{\xi : \xi <_\Lambda \lambda\}$. Then, set

$$[\lambda]_T^{\Lambda,r}\phi = [\lambda]_T^{\Lambda\restriction\lambda+1}\phi.$$

We have chosen to work with $[\lambda]_T^\Lambda$ rather than $[\lambda]_T^{\Lambda,r}$ for the sake of simplicity, but our results and techniques may readily be modified to work directly with $[\lambda]_T^{\Lambda,r}$ instead. This may be useful, for example, in defining a predicate chain of length Γ_0 in ATR_0, which is Λ-introspective for all $\Lambda < \Gamma_0$ but cannot prove that Γ_0 is well-ordered.

Appendix A.4. Provability by ω-trees

Finally, we mention an entirely different approach can be taken to iterated provability, as found in e.g. [1, 22]. There, derivability with the ω-rule is formalized by the existence of an (infinite) derivation tree. An *ω-tree* is a tree S such that each node has at most countably many daughters. A *labeled ω-tree* is a pair $\langle S, L \rangle$ such that S is an ω-tree and $L\colon S \to \mathbb{N}$.

Given a theory T, a *preproof for T* is a labeled ω-tree $\langle S, L\rangle$ such that for every $s \in S$, $L(s)$ is a formula in the language of T, and s is of one of two types:

(a) s is a leaf of S and $\Box_T L(s)$ holds, or

(b) there is a formula $\psi(x)$ such that one daughter of s is labeled by $\forall x \left(\psi(x) \to L(s)\right)$, while for each $n \in \mathbb{N}$, s has a daughter labeled by $\psi(\overline{n})$.

If S is well-founded, we will say that $\langle S, L\rangle$ is an *ω-proof*. The *depth* of $\langle S, L\rangle$ is the depth of S as a tree (with the leaves at depth 0).

Then, we can define $[\lambda]_T^{\Lambda,t}\phi$ as a formalization of *There exists an ω-proof of ϕ of depth at most λ.* Here we will not go into further detail, and instead

refer the reader to [1, 22].[10] Nevertheless, in [19] we compare the operators $[\lambda]_T^{\Lambda,t}$ and $[\lambda]_T^{\Lambda}$, in particular showing that provability in the two senses is equivalent over ATR_0, although over weaker theories this is generally not the case.

Appendix B. An afterword on the choice of base theory

In this paper we have shown that the logic GLP_Λ is sound and complete for the interpretation where each modality $[\xi]$ is interpreted as "provable in ACA_0 using at most ξ nested applications of the omega-rule". The main application we have in mind with this result is to provide Π_1^0 ordinal analyses of theories much stronger than PA in the style of Beklemishev [6]. However, for the mere soundness of the logic there were quite some strong principles needed: the existence of an iterated provability class, plus a certain amount of transfinite induction. We shall discuss here how these principles fit into the intended application to ordinal analysis.

A consistency proof for a theory S in a finitary base theory U plus $\mathtt{TI}(\Omega, \Pi_1^0)$ (with Ω a large enough well-order) is closely related to the Π_1^0 ordinal analysis of S and we shall focus our discussion on such a consistency proof.

Beklemishev has shown how such a consistency proof for Peano Arithmetic can be performed almost entirely within the modal logic GLP_ω. There are three theories involved in such an analysis. First, there is a strong target theory S whose consistency one wishes to prove (PA in Beklemishev's analysis). Second, there is a finitary base theory U in which the consistency proof will be formalized, together with transfinite induction up to an appropriately large computable ordinal. In Beklemishev's analysis, $U = \mathrm{EA}^+$, i.e., EA together with an axiom asserting that the superexponential is total. Finally, there is a theory T which works as a 'consistency unit', in this case EA. If we let $\langle n \rangle_{\mathrm{EA}}^{\mathtt{True}} \phi$ denote 'ϕ is derivable in T from a true Π_n formula', Beklemishev then uses the fact that $\mathrm{PA} \equiv \mathrm{EA} + \{\langle n \rangle_{\mathrm{EA}}^{\mathtt{True}} \top\}_{n<\omega}$ to realize a consistency proof for PA using EA^+ and transfinite induction up to ε_0 [6]. Note that in this analysis, $\Lambda = \omega$ and $\Omega = \varepsilon_0$, and in general one expects $\Lambda \leq \Omega$.

---

[10]Note, however, that our 'accounting' is somewhat different since these authors count the applications of all deduction rules when computing the height of a derivation, whereas we only count the nesting depth of ω-rules.

Thus, it is important that all the modal reasoning is available in the base theory U. This translates to the requirement that soundness of the modal logic should be provable in the base theory, and for this purpose it was important to separate the consistency unit T from the meta-theory U throughout this paper. With this in mind, we could have added standing assumptions on U and T throughout; however, we have chosen to state the assumptions on U and T in a case-by-case fashion since over weaker systems one might get logics different from GLP_Λ; the corresponding logics might be of interest, and it would be good to know how much of the reasoning in this paper carries over and how much does not.

Consistency proofs within a finitary base theory U plus $\mathtt{TI}(\Omega, \Pi_1^0)$ can be seen as partial realizations of Hilbert's program in the sense that over finitsitic mathematics one can prove the consistency of a strong theory with just one additional non-finitist ingredient. As such, one could say that S is safeguarded by this method. If one accepts this method of safeguarding it makes sense to use S itself as new base theory to safeguard even stronger theories by the same method. It is in this perspective that having ACA_0 as our base theory is not a bad thing since ACA_0 has already been safeguarded over EA using some amount of transfinite induction. However, there are reasons why it might be desirable to dispense with such an intermediate step.

In [28] it is noted that it is soundness rather than completeness of GLP_Λ that is involved in a consistency proof. But also in the soundness proof presented in this paper, we needed transfinite induction as well as resorting to the Λ-introspective closure of a theory. As we have seen in Corollary 6.11, as concerns consistency strength, it does not matter whether we consider either a theory or its Λ-introspective closure, as both theories are provably equiconsistent. Thus the key issue is the amount of transfinite induction needed in our soundness proof for GLP_Λ.

First of all, let us note that our soundness proof for GLP_Λ uses[11] at most $\mathtt{TI}(\Lambda, \boldsymbol{\Pi}_1^0)$. In a sense, this is not bad at all, because it is exactly this ingredient (in parameter-free form) that is added to our base theory U to perform a consistency proof for our target theory. So, by adding this amount of

---

[11]The referee most kindly pointed out that it would be desirable to use transfinite induction for initial segments of Λ in the the soundness proof of GLP_Λ, rather than all of Λ. Indeed, this is sufficient to prove each instance of the axioms rather than their universal closure as we have done in this paper.

transfinite induction (with parameters) to U, we get access to exactly the soundness of GLP_Λ needed to perform this consistency proof, although our base theory was taken to be ACA_0, not EA^+. This choice of U has been mainly to simplify our exposition, but the needed amount of arithmetic can be pushed down a lot farther.

In Beklemishev's analysis, the 'base theory' U (EA^+) is an extension of the 'consistency unit' T (EA). However, rather than an actual extension, it is only necessary that U prove the consistency of T. In this sense, working with T or with its introspective closure does not make much difference, since T is provably equiconsistent to $T + \Box_T\bot$ (see Lemma 6.6).

Finally, we observed that to prove $\mathtt{TI}(\Lambda, \boldsymbol{\Pi}_1^0)$ over $\mathrm{EA}^\Lambda + \mathtt{Ind}$ we need $\boldsymbol{\Sigma}_1^0$ comprehension, but $\boldsymbol{\Sigma}_1^0$ comprehension with second order parameters already proves ACA_0.

However, we can do better still in the sense that we need less comprehension by allowing slightly stronger well-ordering assumptions, as we shall see in the next lemma. Let us fix some computable well-order Λ, and let $<$ denote the usual ordering on the natural numbers. Using a bijective pairing function we define a new computable well-order $\omega\Lambda$ with $|\omega\Lambda| = \mathbb{N} \times |\Lambda|$ and

$$\langle n, \xi\rangle <_{\omega\Lambda} \langle m, \zeta\rangle := \xi <_\Lambda \zeta \vee (\xi = \zeta \wedge n < m).$$

Clearly, $\omega\Lambda$ is a new well-order. We then have:

Lemma Appendix B.1. *For any computable well-order Λ,*

$$\boldsymbol{\Delta}_0^0\text{-}\mathtt{CA} + \mathtt{wo}(\omega\Lambda) \vdash \mathtt{TI}(\Lambda, \boldsymbol{\Pi}_1^0).$$

Proof. By the usual argument we see that $\boldsymbol{\Delta}_0^0\text{-}\mathtt{CA} + \mathtt{wo}(\omega \cdot \Lambda) \vdash \mathtt{TI}(\omega\Lambda, \boldsymbol{\Delta}_0^0)$. Thus, we shall prove $\mathtt{TI}(\Lambda, \boldsymbol{\Pi}_1^0)$ using $\mathtt{TI}(\omega\Lambda, \boldsymbol{\Delta}_0^0)$. Let $\phi(z, x)$ be some $\boldsymbol{\Delta}_0^0$ formula possibly with further parameters, and assume

$$\forall x \ (\forall y <_\Lambda x \ \forall z \phi(z, y) \rightarrow \forall z \phi(z, x)). \tag{B.1}$$

If we assume that $\forall y <_{\omega\Lambda} x \ \phi(y_0, y_1)$, by the way we defined the relation $<_{\omega\Lambda}$, we get $\forall y <_\Lambda x_1 \forall z \ \phi(z, y)$. Thus, using (B.1) we get $\forall z \ \phi(z, x_1)$, so certainly $\phi(x_0, x_1)$. Now, by $\mathtt{TI}(\omega\Lambda, \boldsymbol{\Delta}_0^0)$ we obtain $\forall x \forall z \phi(z, x)$. $\qquad\square$

Note that $\omega\Lambda$ is not much larger than Λ. In particular, if the last term in Cantor normal form of Λ is at least ω^ω, we get that $\omega\Lambda = \Lambda$. Thus, for natural proof-theoretical ordinals we have this equation whence we get

the extra induction for free. Consequently, the full ACA_0 is not needed for proving soundness of GLP_Λ and performing a $\mathbf{\Pi}^0_1$ ordinal analysis as sketched above; it may be entirely carried out in the theory ECA_0 given by

$$\mathrm{ECA}_0 = \mathbf{\Delta}^0_0\text{-}\mathtt{CA} + \mathtt{Ind}$$

(here it is important to recall that we assumed exponentiation was in our language, which is needed for Σ^0_1-completeness in Lemma 2.1). This theory is conservative over Elementary Arithmetic [16], and thus perfect candidates for our finitary base theory and consistency unit are $U = \mathrm{ECA}^+_0$ (i.e., ECA_0 with superexponential) and $T = \overline{\mathrm{ECA}}_0$, that is, its introspective closure. As an alternative, one may take more familiar, but somewhat stronger, theories, such as $U = \mathrm{RCA}_0$ and $T = \overline{\mathrm{RCA}^*_0}$, where RCA^*_0 is a weaker version of RCA_0 [33].

Acknowledgments

References

[1] Arai, T., 1998. Some results on cut-elimination, provable well-orderings, induction and reflection. Annals of Pure and Applied Logic 95 (1), 93 – 184.

[2] Ash, C. J., Knight, J., 2000. Computable Structures and the Hyperarithmetical Hierarchy. Elsevier.

[3] Beklemishev, L. D., 1991. Provability logics for natural Turing progressions of arithmetical theories. Studia Logica 50 (1), 109–128.

[4] Beklemishev, L. D., 1992. Independent numerations of theories and recursive progressions. Sibirskii Matematichskii Zhurnal 33 (5), 22–46, (In Russian. English translation in *Siberian Math. Journal*, 33 (1992)).

[5] Beklemishev, L. D., 1995. Iterated local reflection versus iterated consistency. Annals of Pure and Applied Logic 75, 25–48.

[6] Beklemishev, L. D., 2004. Provability algebras and proof-theoretic ordinals, I. Annals of Pure and Applied Logic 128, 103–124.

[7] Beklemishev, L. D., 2005. Veblen hierarchy in the context of provability algebras. In: Hájek, P., Valdés-Villanueva, L., Westerståhl, D. (Eds.), Logic, Methodology and Philosophy of Science, Proceedings of the Twelfth International Congress. Kings College Publications, pp. 65–78.

[8] Beklemishev, L. D., 2010. Kripke semantics for provability logic GLP. Annals of Pure and Applied Logic 161 (6), 756–774.

[9] Beklemishev, L. D., 2011. A simplified proof of the arithmetical completeness theorem for the provability logic GLP. Trudy Matematicheskogo Instituta imeni V.A. Steklova 274 (3), 32–40, in Russian. English translation in *Proceedings of the Steklov Institute of Mathematics*, 274(3):25–33, 2011.

[10] Beklemishev, L. D., Fernández-Duque, D., Joosten, J. J., 2014. On provability logics with linearly ordered modalities. Studia Logica 102 (3), 541–566.

[11] Beklemishev, L. D., Gabelaia, D., 2013. Topological completeness of the provability logic GLP. Annals of Pure and Applied Logic 164 (12), 1201–1223.

[12] Boolos, G., 1979. The Unprovability of Consistency: An Essay in Modal Logic. Cambridge University Press, Cambridge.

[13] Boolos, G., 1993. The analytical completeness of Dzhaparidze's polymodal logics. Annals of Pure and Applied Logic 61 (1-2), 95–111.

[14] Boolos, G., 1993. The Logic of Provability. Cambridge University Press, Cambridge.

[15] Chagrov, A., Zakharyaschev, M., 1997. Modal Logic. Vol. 35 of Oxford logic guides. Oxford University Press.

[16] Cordón Franco, A., Fernández-Duque, D., Joosten, J. J., Lara Martín, F., 2017. Predicativity through transfinite reflection. Journal of Symbolic Logic.

[17] de Jongh, D., Jumelet, M., Montagna, F., 1991. On the proof of Solovay's theorem. Studia Logica 50, 51–69.

[18] Fernández-Duque, D., 2014. The polytopologies of transfinite provability logic. Archive for Mathematical Logic 53 (3-4), 385–431.

[19] Fernández-Duque, D., 2015. Impredicative consistency and reflection. ArXiv e-prints.

[20] Fernández-Duque, D., Joosten, J. J., 2012. Kripke models of transfinite provability logic. In: Advances in Modal Logic. Vol. 9. College Publications, pp. 185–199.

[21] Fernández-Duque, D., Joosten, J. J., 2013. Models of transfinite provability logics. Journal of Symbolic Logic 78 (2), 543–561.

[22] Girard, J.-Y., 1987. Proof theory and logical complexity. Vol. 1. Studies in proof theory. Bibliopolis, Napoli.

[23] Grzegorczyk, A., Mostowski, A., Ryll-Nardzewski, C., 1958. The classical and the omega-complete arithmetic. Journal of Symbolic Logic 23 (2), 188–206.

[24] Hájek, P., Pudlák, P., 1993. Metamathematics of First Order Arithmetic. Springer-Verlag, Berlin, Heidelberg, New York.

[25] Icard III, T. F., 2011. A topological study of the closed fragment of GLP. Journal of Logic and Computation 21, 683–696.

[26] Ignatiev, K. N., 1993. On strong provability predicates and the associated modal logics. The Journal of Symbolic Logic 58, 249–290.

[27] Japaridze, G., 1988. The polymodal provability logic. In: Intensional logics and logical structure of theories: material from the Fourth Soviet-Finnish Symposium on Logic. Metsniereba, Telavi, in Russian.

[28] Joosten, J. J., 2013. Π_1^0-ordinal analysis beyond first-order arithmetic. Mathematical Communications 18, 109–121.

[29] Kotlarski, H., 1986. Bounded induction and satisfaction classes. Zeitschrift für mathematische Logik und Grundlagen der Mathematik 32 (31-34), 531–544.
URL http://dx.doi.org/10.1002/malq.19860323107

[30] Michalski, G., 1992. On the iterated ω-rule. Zeitschrift für mathematische Logik und Grundlagen der Mathematik 38, 203–208.

[31] Pohlers, W., 2009. Proof Theory, The First Step into Impredicativity. Springer-Verlag, Berlin Heidelberg.

[32] Sacks, G. E., 1990. Higher Recursion Theory. Springer-Verlag, Berlin.

[33] Simpson, S. G., 2009. Subsystems of Second Order Arithmetic, 2nd Edition. Cambridge University Press, New York.

[34] Solovay, R. M., 1976. Provability interpretations of modal logic. Israel Journal of Mathematics 25, 287–304.