Utrecht University

Department of Mathematics

Mathematical Sciences MSc thesis

Transfinite induction principles and iterated reflection principles

First examiner:

Dr Jaap van Oosten

Second examiner:

Dr Fan Yang

External advisor:

Prof Lev Beklemishev

Candidate:

Joonas Jürgen Kisel

December 11, 2024

Abstract

Two of the methods used to assign proof-theoretic ordinals to theories are using transfinite induction principles, inspired by Gentzen's consistency proof for PA via ε_0-induction, and approximating them by transfinite iterations of reflection principles, an approach first proposed by Turing and developed further by Feferman, Schmerl and Beklemishev. A number of well-known results by Feferman, Kreisel and Lévy, Schmerl, Sommer and Beklemishev relate transfinite induction to reflection. In this paper we study the relationships between partial transfinite induction principles obtained by restricting it to a class Σ_n or Π_n in the arithmetical hierarchy and iterated reflection principles. Our goal is to obtain a complete characterisation of transfinite induction principles in terms of reflection principles, closing the last remaining gap in the overall picture.

We first introduce our base theory of choice, the arithmetic EA of Kalmár elementary functions, and outline various details about its arithmetisation and conducting ordinal arithmetic (up to ε_0) within it. We then introduce partial transfinite induction and reflection principles and present some key facts established by Sommer, Schmerl and Beklemishev. Finally we state and prove our main result: the transfinite induction schema for Π_n formulae up to the ordinal ω^{ω^α} is equivalent to $\alpha + 1$ times iterated uniform reflection principle for Π_{n+2} formulae over EA.

Contents

1 Introduction **3**

1.1 Outline 4

1.2 Acknowledgements 5

2 Preliminaries **6**

2.1 Elementary arithmetic 6

2.2 Extensions and provability 12

2.3 Induction principles and partial truth predicates 15

2.4 Representing ordinals in EA 18

3 Transfinite induction principles **21**

3.1 Transfinite induction principles 21

4 Reflection principles **27**

4.1 Reflection principles 27

4.2 Transfinite progressions of reflection principles 31

5 $\mathsf{TI}[\omega_2(\alpha), \Pi_n]$ in terms of $\mathsf{RFN}^{\beta}_{\Pi_m}(\mathsf{EA})$ **34**

5.1 $\dashv$ entailment 34

5.2 $\vdash$ entailment 35

6 Summary **42**

Bibliography **44**

1. Introduction

A central concept in quantifying proof-theoretic strength of different theories is that of proof-theoretic ordinals. However, there are various natural ways in which theories can be assigned ordinals, each with their own advantages and shortcomings. For instance, they may be dependent on the coding of the ordinals or fail to distinguish between theories that differ by a single axiom of low quantifier complexity (e.g. PA and $\mathsf{PA} + \mathsf{Con}(\mathsf{PA})$, i.e. Peano arithmetic with and without the corresponding consistency assertion). This thesis will focus on two methods in particular.

Firstly, we have the approach via *transfinite induction*, i.e. of assigning to a theory the lowest ordinal where the theory's consistency can be proved by induction up to that ordinal. This was primarily motivated by Gentzen demonstrating in [4] that the corresponding ordinal for PA is ε_0. However, a naïve generalisation of this runs into multiple issues, such as dependency on the ordinal notation system and the provability predicate chosen, as well as the fact that the transfinite induction itself will have higher complexity than the consistency assertion for whose proof it is being employed. The difference in quantifier complexity leads to some gaps in the ordinal assignments that one obtains via transfinite induction schemata, even if one restricts the quantifier complexity of the induction formulas.

Secondly, we have the approach via *iterated reflection*. By Gödel's incompleteness theorems, iterating soundness assertions produces a strictly increasing progression of theories. Given a base theory, such progressions can be used to approximate stronger theories, a method pioneered by Turing in [12] and Feferman in [3], developed further by Schmerl in [10] and Beklemishev in [1].

Transfinite progressions of reflection principles provide a more refined method of measuring proof-theoretic strength of theories. In particular, transfinite induction principles can be expressed in terms of iterated reflection principles, but not always vice versa (for example, the theory $\mathsf{PA} + \mathsf{Con}(\mathsf{PA})$ is not equivalent to a natural restricted version of transfinite induction schema). The exact relation between restricted transfinite induction and iterated reflection schemata have been studied in the above mentioned literature. For example, Kreisel and Lévy have shown in [5] that α times iterated uniform reflection schema over PA is equivalent to transfinite induction schema up to ordinals below ε_α. Other basic relationships have been established by Schmerl, Sommer and Beklemishev; however, the general picture remained incomplete.

More specifically, Sommer established that transfinite induction schemata of the form $\mathsf{TI}[\alpha, \Pi_n]$ or $\mathsf{TI}[\alpha, \Sigma_n]$, where n>=0 and alpha is a natural ordinal notation $<_{\varepsilon_0}$, are equivalent to $\mathsf{TI}[\omega_2(\beta), \Pi_m]$, for some β, m. However, up until now we did not know a characterisation (axiomatisation) of theories $\mathsf{TI}[\omega_2(\beta), \Pi_m]$ in terms of iterated reflection principles for arbitrary β and m. The answer to this question is given in the present work.

1.1 Outline

In Chapter 2 we will outline the properties of our base theory EA, the arithmetic of Kalmár-elementary functions, and establish the notation and properties of various predicates and encodings, including an example of an encoding of the ordinals below ε_0 such that ordinal arithmetic can be performed by Kalmár-elementary functions.

In Chapter 3 we will explore the partial transfinite induction principles, determine the pairwise distinct ones and establish some key facts, as in [11]. Similarly, in Chapter 4 is dedicated to introducing partial reflection principles and establishing basic results about them and transfinite iterations thereof, as in [1] and [2].

In Chapter 5 we state and prove the advertised result: we close a remaining gap in the picture by showing that the transfinite induction schemata $\mathsf{TI}[\omega_2(\alpha), \Pi_n]$ are equivalent to $\alpha + 1$ times iterated uniform reflection principles for Π_{n+2} formulae (over EA), for canonical ordinal notations up to ε_0. This confirms the relationship conjectured by Beklemishev in correspondence with A. Freund. Thereby we classify all partial transfinite induction principles (restricted to Σ_n or Π_n for some $n \geq 1$) in terms of iterated partial reflection principles. Our methods and results combine those of Schmerl, Sommer and Beklemishev. Finally, in Chapter 6 we will summarise the results and pose further questions of interest.

1.2 Acknowledgements

I would like to thank Jaap van Oosten for agreeing to supervise my thesis and helping deal with organisational issues; Fan Yang for agreeing to be second reader; Lev Beklemishev for proposing the topic, relaying the conjecture proved in Chapter 4 and offering plenty of advice throughout; and all three for their patience throughout. I would also like to thank Rosalie Iemhoff for initially agreeing to be second reader and showing continued interest in the thesis; and friends both online and offline, as well as my family, for their unwavering support.

2. Preliminaries

2.1 Elementary arithmetic

We shall be operating in first-order logic, using the language of arithmetic (consisting of the constants 0 and 1, the relation $<$, the unary function exp representing the function $x \mapsto 2^x$, and the binary functions $+$ and $\times$).

We will begin by setting forth a certain notion of complexity for formulae in this language. Observe that the fundamental reason why propositional logic is complete whereas first-order logic is not is that the truth valuation on the quantifiers $\forall$ and $\exists$ translates to an infinite search. However, an exception occurs when all quantifiers occur in subformulae of the form $\forall x(x < t \rightarrow \varphi(x))$ or $\exists x(x < t \wedge \varphi(x))$ where in each case, the term t does not depend on the quantified variable x: in such cases, given any set of inputs, the terms adopt a finite set of values and the evaluation of the truth of these subformulae becomes equivalent to a finite case check. We refer to the quantifiers at the start of these subformulae as *bounded* and adopt the shorthands $(\forall x < t)\varphi(x)$ and $(\exists x < t)\varphi(x)$, respectively; similarly for $\leq$ in place of $<$.

Definition 2.1.1 (Bounded formulae)**.** *A formula where all the quantifiers are bounded is called a* bounded formula.

This leads us to define our notion of complexity:

Definition 2.1.2 (Arithmetical complexity)**.** *The set of all bounded formulae (in a language with exp) is labelled $\Delta_0(exp)$. We then define the following:*

- $\Sigma_0 = \Pi_0 = \Delta_0(exp)$;
- $\Sigma_{n+1} = \{\exists x\varphi(x, y) | \varphi(x, y) \in \Pi_n\}$;
- $\Pi_{n+1} = \{\forall x\varphi(x, y) | \varphi(x, y) \in \Sigma_n\}$.

Intuitively speaking, n is an upper bound for the number of nested iterations over $\mathbb{N}$ required to evaluate the truth of the formula.

Remark 2.1.3 (Arithmetical hierarchy)**.** *For all n, up to deductive equivalence, we have $\Sigma_n \subsetneq \Sigma_{n+1}, \Sigma_n \subsetneq \Pi_{n+1}$ and $\Pi_n \subsetneq \Sigma_{n+1}, \Pi_n \subsetneq \Pi_{n+1}$.*

Definition 2.1.4. Elementary (function) arithmetic *(*EA*, also labelled as* $\mathsf{I}\Delta_0(exp)$*) is the first-order theory with its signature as specified above and its axioms as follows:*

1. $\neg(x + 1 = 0)$;
2. $x + 1 = y + 1 \to x = y$;
3. $y = 0 \lor \exists x(x + 1 = y)$;
4. *the axioms for addition:* $x + 0 = x$, $x + (y + 1) = (x + y) + 1$;
5. *the axioms for multiplication:* $x \cdot 0 = 0$, $x \cdot (y + 1) = x \cdot y + x$;
6. *the axioms for exponentiation:* $exp(0) = 1$, $exp(x + 1) = 2 \cdot exp(x)$;
7. *the induction schema for bounded formulae:*

$$\varphi(0, y) \land \forall x(\varphi(x, y) \to \varphi(x + 1, y)) \to \forall x \varphi(x, y) \tag{2.1}$$

   for φ bounded.

We refer to bounded formulae in the language of EA *as* elementary formulae*.*

Remark 2.1.5. *We recognise axioms 1-5 as the* Robinson arithmetic*, a theory famously strong enough for Gödel's second incompleteness theorem to be applicable. Since* EA *is a consistent extension of it, the theorem also applies to it; this is also the case for all extensions of* EA *we will consider.*

The name "elementary arithmetic" derives from the fact that the class of provably total functions of EA is the class of *Kalmár elementary functions* which is the smallest class $\mathcal{E}$ containing the unary zero function, $+$, $\times$, exp, the indicator function for $<$ and the projection functions $\pi^n_k(x_1, \ldots, x_n) := x_k$ which is closed under composition and bounded minimisation:

Definition 2.1.6 (Bounded minimisation)**.** *For $f : \mathbb{N}^2 \to \mathbb{N}$, labelling*

$A_f(x,y) := \{z \in \mathbb{N} | z \leq y \wedge f(x,z) = 0\}$, *the* bounded minimisation *of* f *is given by*

$$\mu z \leq y(f(x,z) = 0) := \begin{cases} \min A_f(x,y) & \text{if } A_f(x,y) \neq \emptyset \\ y+1 & \text{otherwise} \end{cases} \tag{2.2}$$

Definition 2.1.7 (Multiexponential functions)**.** *For each* $n \in \mathbb{N}$*, we denote the* n*-fold composition of* $exp(x)$ *as* 2^x_n *(with* $2^x_0 := x$*) and refer to them as the* multiexponential functions*. Since* $exp \in \mathcal{E}$ *and* $\mathcal{E}$ *is closed under composition (and since* 2^x_0 *coincides with* $\pi^1_1(x)$*), the multiexponential functions are all Kalmár elementary. When a function in* $x_1, \ldots, x_n$ *is bounded above by* $2^{x_1+\ldots+x_n}_k$ *for some value of* k*, we say it is* multiexponentially bounded*.*

Remark 2.1.8. *It can be verified (e.g. by structural induction on the definition of* $\mathcal{E}$*) that every Kalmár elementary function in* x *is multiexponentially bounded.*

Given this, there is a natural choice for a metatheory for results which are not provable in EA alone:

Definition 2.1.9. *The first-order theory* EA^+ *is obtained by extending* EA *with an axiom asserting the totality of the function* $supexp(n,x) := 2^x_n$ *(as a function of* n *and* x*).*

Remark 2.1.10. EA^+ *is strong enough to prove cut-elimination for predicate logic; however, we will later see that most of the extensions we consider contain* EA^+*, thus making it a weaker choice of metatheory than one might expect.*

We will now outline some details about the arithmetisation of the syntax of first-order theories. As arithmetical formulae can be naturally identified with finite sequences of natural numbers by assigning a number to each symbol of our language, we will need a way to encode such sequences with a single natural number and invert this encoding in terms of elementary functions. Thus we define the following:

Definition 2.1.11 (Cantor coding)**.** *The* Cantor coding *for pairs of natural* *numbers is given by*

$$j(x,y) := \frac{(x+y)(x+y+1)}{2} + y \tag{2.3}$$

This is a bijection between $\mathbb{N} \times \mathbb{N}$ *and* $\mathbb{N}$. *We also define the inverses*

$$\begin{aligned} j_1(z) &:= \mu x \leq z(\exists y \leq z(z = j(x,y))) \\ j_2(z) &:= \mu y \leq z(\exists x \leq z(z = j(x,y))) \end{aligned} \tag{2.4}$$

Note that $j, j_1, j_2 \in \mathcal{E}$.

For $n \geq 1$, *we define the n-ary function* $j^n(x_1, \ldots, x_n)$ *as follows:*

$$\begin{aligned} j^1(x_1) &:= x_1; \\ j^{m+1}(x_1, \ldots, x_{m+1}) &:= j(j^m(x_1, \ldots, x_m), x_{m+1}) \end{aligned} \tag{2.5}$$

j^n *is an bijection* $\mathbb{N}^n \to \mathbb{N}$. *We then define the inverses* j_i^n *for* $1 \leq i \leq n$, *given by*

$$\begin{aligned} j_1^1(z) &:= z; \\ j_i^{m+1}(z) &:= \begin{cases} j_i^m(j_1(z)) & \text{for } 1 \leq i \leq m; \\ j_2(z) & \text{for } i = m+1 \end{cases} \end{aligned} \tag{2.6}$$

We observe that j^n, j_i^n *are also elementary for all* $1 \leq i \leq n$.

Definition 2.1.12 (Code of a sequence)**.** *We denote the set of finite sequences of a set* A *as* A^*. *We fix a natural number* N_{seq}. *For* $[x_1, \ldots, x_n] \in \mathbb{N}^*$, *its code is* *given by the following:*

$$\begin{aligned} \langle\rangle &:= j(N_{seq}, 0); \\ \langle x_1, \ldots, x_n \rangle &:= j(N_{seq}, j(n-1, j^n(x_1, \ldots, x_n)) + 1) \end{aligned} \tag{2.7}$$

Due to j and j^n being elementary bijections, $\langle \cdot \rangle$ is an elementary function $\mathbb{N}^* \to \mathbb{N}$; moreover, x is the code of a sequence iff $j_1(x) = N_{seq}$. This means that the formula $\mathsf{Seq}(x)$, denoting "x is the code of a sequence", is elementary.

Let the function $\mathsf{lh}(x)$ be given by the following:

$$\mathsf{lh}(x) = \begin{cases} x+1 & \text{if } j_1(x) \neq N_{seq} \\ 0 & \text{if } x = j(N_{seq}, 0) \\ j_1(j_2(x)-1)+1 & \text{otherwise} \end{cases} \tag{2.8}$$

This is elementary and satisfies $\mathsf{lh}(\langle x_1, \ldots, x_n \rangle) = n$.

We further define the function $(x)_i$ of x and i as follows:

$$(x)_i := \begin{cases} j_i^{\mathsf{lh}(x)}(j_2(j_2(x)-1)), & \text{if } 1 \leq i \leq \mathsf{lh}(x) \\ 0, & \text{else} \end{cases} \tag{2.9}$$

We observe that $(x)_i$ is elementary and, for all $1 \leq i \leq n$, we have $(\langle x_1, \ldots, x_n \rangle)_i = x_i$.

Remark 2.1.13. *We observe that this justifies us not considering subformulae of the form $\forall x_1 \forall x_2 \ldots \forall x_n \varphi(x_1, \ldots, x_n)$ and $\exists x_1 \exists x_2 \ldots \exists x_n \varphi(x_1, \ldots, x_n)$ for $n \geq 2$ in 2.1.2: in both cases, we can instead quantify over $\langle x_1, \ldots, x_n \rangle$ without raising the complexity. Similarly, we can assume all formulae to have at most one free variable, as we can elementarily convert between (the Gödel numbers of) $\varphi(x_1, \ldots, x_n)$ and $\psi(x) := \varphi((x)_1, \ldots, (x)_n)$.*

Based on the coding of sequences of natural numbers, we will now fix a standard *Gödel encoding* of expressions in the language of EA, denoting the code of an expression σ (also known as its *Gödel number*) as $\ulcorner\sigma\urcorner$. Similarly to sequences, we will fix a natural number N_{expr} and assume without loss of generality that the Gödel number of an expression is of the form $j(N_{expr}, x)$ for some x. This will also have several syntactic properties be defined by elementary formulae.

Proposition 2.1.14. *Under a suitably chosen Gödel encoding, the following formulae are elementary:*

- $\mathsf{Term}(x), \mathsf{AtForm}(x), \mathsf{Form}(x), \mathsf{Sent}(x)$, *standing for "x is the Gödel number of a term/atomic formula/formula/sentence", respectively;*
- $\Pi_n(x), \Sigma_n(x)$, *standing for "x is the Gödel number of a* Π_n/Σ_n *formula", respectively;*
- $\mathsf{Ax}_{\mathsf{EA}}(x)$, *standing for "x is the Gödel number of an axiom of* EA*";*
- $\mathsf{MP}(x, y, z)$, *standing for "z is the Gödel number of a formula that follows from the formulae with Gödel numbers x and y by modus ponens";*
- $\mathsf{Gen}(x, y)$, *standing for "y is the Gödel number of a formula that follows from the formula with Gödel number x by generalisation".*

This follows from these formulae being primitive recursive in PRA, by observing that all instances of primitive recursion are multiexponentially bounded and can thus be replaced by bounded minimisation.

To cap off this section, we will establish two more pieces of notation:

Definition 2.1.15 (Numerals)**.** *The term* $(\ldots((0+1)+1)\ldots+1)$ *where* 1 *occurs n times is called the* numeral *of n and denoted* $\overline{n}$. *Observe that there is an elementary formula* $\mathsf{Num}(x)$, *standing for "x is the code of a numeral", and an elementary function* $\mathsf{num}(x)$ *which maps x to* $\ulcorner\overline{x}\urcorner$.

Definition 2.1.16. *We denote the formula whose Gödel encoding is e as* φ_e.

2.2 Extensions and provability

For Gödel's incompleteness theorems to apply to extensions of EA, there needs to be a recursively enumerable formula for the Gödel numbers of its axioms. However, it will be more convenient for us to require such a formula to be elementary, motivating the following definition:

Definition 2.2.1. *We say a theory T containing* EA *is* elementarily presented *if there is an elementary formula $\mathsf{Ax}_T(x)$ which is true iff x is the Gödel number of an axiom of T.*

From here on, unless otherwise stated, we will assume T to be elementarily presented and to contain EA. We denote the theory whose axioms are given by $\mathsf{Ax}_T(x) \vee \mathsf{Ax}_U(x)$ as $T + U$ and, for a formula $\varphi(y)$, we denote the theory whose axioms are given by $\mathsf{Ax}_T(x) \vee x = \ulcorner \varphi(y) \urcorner$ as $T + \varphi(y)$.

Definition 2.2.2 (Deductive equivalence)**.** *Theories T and U are said to be* deductively equivalent*, denoted $T \equiv U$, if for all formulae φ we have $T \vdash \varphi \iff U \vdash \varphi$. We say they are $\mathcal{C}$-*equivalent*, denoted $T \equiv_{\mathcal{C}} U$, if this condition is satisfied for all formulae $\varphi \in \mathcal{C}$.*

Recall that Hilbert's notion of a derivation is a sequence of formulae where each formula is either an axiom of the theory, the result of modus ponens applied to two of the previously listed formulae or the result of generalisation applied to one of the previously listed formulae, with the last line being the derived formula. Taking the Gödel number of a proof to be the code of the sequence of the Gödel numbers of the formulae, we can construct a *proof predicate* for a theory T:

$$\begin{aligned}\mathsf{Prf}_T(x,y) := \mathsf{Seq}(y) \wedge (y)_{\mathsf{lh}(y)} = x \wedge \\ \forall i \leq \mathsf{lh}(y)[i \geq 1 \rightarrow &(\mathsf{Ax}_T((y)_i) \vee \\ &\exists j,k < i(\mathsf{MP}((y)_j,(y)_k,(y)_i)) \vee \\ &\exists j < i(\mathsf{Gen}((y)_j,(y)_i)))]\end{aligned} \tag{2.10}$$

This represents y being the code of a proof of x. Note that since T is elementarily presented, $\mathsf{Prf}_T(x, y)$ is an elementary formula.

This in turn allows us to define a *provability predicate*:

Definition 2.2.3 (Provability predicate)**.** *We denote* $\exists y\ \mathsf{Prf}_T(x, y)$ *as* $\Box_T(x)$*. We will also adopt the shorthand* $\Box_T\varphi$ *for* $\Box_T(\overline{\ulcorner\varphi\urcorner})$*.*

We will now establish a few well-known results regarding $\Box_T$.

Lemma 2.2.4 (Deduction theorem)**.** *For any formulae* φ, ψ*,*

$$\mathsf{EA} \vdash \Box_{S+\varphi}\psi \to \Box_S(\varphi \to \psi) \tag{2.11}$$

Lemma 2.2.5 (Löb's conditions)**.** *For any sentences* φ, ψ*,*

LC1. $T \vdash \varphi \Rightarrow \mathsf{EA} \vdash \Box_T\varphi$

LC2. $\mathsf{EA} \vdash \Box_T(\varphi \to \psi) \to (\Box_T\varphi \to \Box_T\psi)$

LC3. $\mathsf{EA} \vdash \Box_T\varphi \to \Box_T\Box_T\varphi$

Proposition 2.2.6 (Fixed-point lemma)**.** *1. For any formula* $\varphi(x)$ *there is a formula* ψ *depending solely on the free variables of* φ *other than* x *such that*

$$\mathsf{EA} \vdash \psi \leftrightarrow \varphi(\ulcorner\psi\urcorner) \tag{2.12}$$

2. For any formula $\varphi(x_1, \ldots, x_n)$ *there is a formula* $\psi(x_1, \ldots, x_n)$ *such that*

$$\mathsf{EA} \vdash \psi(x_1, \ldots, x_n) \leftrightarrow \varphi(\ulcorner\psi(\dot{x}_1, \ldots, \dot{x}_n)\urcorner) \tag{2.13}$$

Theorem 2.2.7 (Löb's theorem)**.** *For any sentence* φ*,*

$$T \vdash \Box_T\varphi \to \varphi \iff T \vdash \varphi \tag{2.14}$$

Proof. The right-to-left implication is clear. For the converse, by 2.2.6 there is a sentence ψ such that $\mathsf{EA} \vdash \psi \leftrightarrow (\Box_T \psi \to \varphi)$. Using 2.2.5, as well as the fact that T contains EA, we can derive φ in T as follows:

1. $\Box_T(\psi \to (\Box_T \psi \to \varphi))$ (LC1)
2. $\Box_T \psi \to \Box_T(\Box_T \psi \to \varphi))$ (LC2)
3. $\Box_T \psi \to (\Box_T \Box_T \psi \to \Box_T \varphi)$ (LC2)
4. $\Box_T \psi \to \Box_T \varphi$ (LC3, modus ponens)
5. $\Box_T \psi \to \varphi$ (modus ponens on 4 and $T \vdash \Box_T \varphi \to \varphi$)
6. ψ (construction of ψ)
7. $\Box_T \psi$ (LC1)
8. φ (modus ponens on 5 and 7)

□

Löb's conditions generalise to formulae. However, observe that by the generalisation rule and the $\forall$-elimination axiom, $\varphi(x)$ is provable if and only if $\forall x \varphi(x)$ is provable, so naïvely replacing all instances of φ, ψ by $\varphi(x), \psi(x)$ gives rise to equivalent conditions. The proper generalisation is a strictly stronger result; to state it, we will first need to introduce some notation:

Definition 2.2.8. *We denote by* $\mathsf{sub}(x, y, z)$ *the function whose inputs are the Gödel code x of an expression, the Gödel code y of a variable and the Gödel code z of a term, and whose output is the Gödel code of the expression obtained by replacing all instances of the variable encoded by y in the expression encoded by x with the term encoded by z. As there are at most x symbols in the expression encoded by x, there are at most x instances of the variable in the expression, so this can be seen to be an elementary function.*

We now define $\ulcorner \varphi(\dot{x}_1, \ldots, \dot{x}_n) \urcorner$ *as satisfying the following:*

$$\begin{aligned} \ulcorner \varphi(\dot{x}_1) \urcorner &= \mathsf{sub}(\ulcorner \varphi(x_1) \urcorner, \ulcorner x_1 \urcorner, \mathsf{num}(x_1)); \\ \ulcorner \varphi(\dot{x}_1, \ldots, \dot{x}_n) \urcorner &= \mathsf{sub}(\ulcorner \varphi(\dot{x}_1, \ldots, \dot{x}_{n-1}, x_n) \urcorner, \ulcorner x_n \urcorner, \mathsf{num}(x_n)) \end{aligned} \tag{2.15}$$

Note that $\ulcorner\varphi(x_1,\ldots,x_n)\urcorner$ *is a constant whereas* $\ulcorner\varphi(\dot{x}_1,\ldots,\dot{x}_n)\urcorner$ *depends on* $x_1,\ldots,x_n$.

We can now state Löb's conditions for formulae:

Proposition 2.2.9. *For any formulae* $\varphi(x),\psi(x)$,

LC1f. $T\vdash\varphi(x)\Rightarrow \mathsf{EA}\vdash\Box_T\varphi(\dot{x})$

LC2f. $\mathsf{EA}\vdash\Box_T(\varphi(\dot{x})\to\psi(\dot{x}))\to(\Box_T\varphi(\dot{x})\to\Box_T\psi(\dot{x}))$

LC3f. $\mathsf{EA}\vdash\Box_T\varphi(\dot{x})\to\Box_T\Box_T\varphi(\dot{x})$

Proposition 2.2.10. *For any formula* $\varphi(x)$, *we have*

$$\mathsf{EA}\vdash\Box_T(\forall x\varphi(x))\to\forall x\Box_T\varphi(\dot{x}) \tag{2.16}$$

We finish the section by stating another important result.

Proposition 2.2.11. *Any elementarily presented theory* T *containing* EA *is* provably Σ_1-complete, *i.e. for every true* Σ_1 *sentence* φ, EA *proves* $\varphi\to\Box_T\varphi$; *and for every true* Σ_1 *formula* $\varphi(x_1,\ldots,x_n)$, *with all of its free variables explicitly shown,* EA *proves* $\varphi(x_1,\ldots,x_n)\to\Box_T\varphi(\dot{x}_1,\ldots,\dot{x}_n)$.

2.3 Induction principles and partial truth predicates

Recall that EA includes the induction schema for $\Delta_0(exp)$ formulae. We can also define induction schemata for higher classes of the arithmetical hierarchy. More explicitly:

Definition 2.3.1. *The* induction principle *for a class* $\mathcal{C}$ *is the schema* $\mathsf{I}\mathcal{C}$ *given by*

$$\forall y(\varphi(0,y)\land\forall x(\varphi(x,y)\to\varphi(x+1,y))\to\forall x\varphi(x,y)) \tag{2.17}$$

for all $\varphi\in\mathcal{C}$.

Aside from $\mathcal{C}=\Delta_0(exp)$, we will consider the cases $\mathcal{C}=\Sigma_n,\Pi_n$ for $n\geq 1$.

Since we have elementary formulae $\Sigma_n(x)$ and $\Pi_n(x)$, we find that for all n, $\mathsf{Ax}_{\mathsf{I}\Sigma_n}(x)$ and $\mathsf{Ax}_{\mathsf{I}\Pi_n}(x)$ are elementary, i.e. $\mathsf{I}\Sigma_n$ and $\mathsf{I}\Pi_n$ are elementarily presented.

Proposition 2.3.2. *For all n, we have $\mathsf{I}\Sigma_n \equiv \mathsf{I}\Pi_n$.*

Proof. As $\Sigma_0 = \Pi_0 = \Delta_0(exp)$, this clearly holds for $n = 0$. For $n \geq 1$, let $\varphi(x) \in \Pi_n$. We shall show that the instance of $\mathsf{I}\Pi_n$ for φ can be deduced from $\mathsf{I}\Sigma_n$.

Define $\psi(x, y) := \neg\varphi(y \dot{-} x)$ with $\psi \in \Sigma_n$, where $\dot{-}$ denotes truncated subtraction: $y \dot{-} x$ evaluates to $y - x$ if $y \geq x$, and 0 otherwise. We note that the premise $\forall x(\varphi(x) \rightarrow \varphi(x+1))$ implies $\forall x(\psi(x, y) \rightarrow \psi(x+1, y))$, therefore by $\mathsf{I}\Sigma_n$ we have $\forall y(\psi(0, y) \rightarrow \psi(y, y))$, i.e. $\forall y(\varphi(0) \rightarrow \varphi(y))$, hence $\varphi(0) \wedge \forall x(\varphi(x) \rightarrow \varphi(x+1)) \rightarrow \forall x \varphi(x)$ as required.

This establishes $\Pi_n \subseteq \Sigma_n$; for the other inclusion, take $\varphi(x) \in \Sigma_n$ instead.

□

Recall that by the *Tarski undefinability theorem*, there is no formula $\mathsf{Tr}(x)$ such that, for all $m \in \mathbb{N}$, $\mathsf{Tr}(m)$ holds in $\mathbb{N}$ if and only if m is the Gödel code of a sentence φ that holds in $\mathbb{N}$.

However, note that in Δ_0 sentences (i.e. elementary sentences not involving the exp symbol), all terms are either fixed or have boundedly quantified variables as their variables. We can also structurally define an *evaluation function* $\mathsf{eval}(u, x)$ such that

$$\mathsf{EA} \vdash \mathsf{eval}(\ulcorner t \urcorner, \langle x_1, \ldots, x_n \rangle) = t(x_1, \ldots, x_n) \tag{2.18}$$

for any term $t(x_1, \ldots, x_n)$. This extends to $\Delta_0(exp)$ sentences though there the totality is only proved in EA^+. This shows that bounded quantifications become EA^+-provably equivalent to finite conjunctions and disjunctions. Therefore each $\Delta_0(exp)$ sentence is equivalent to a quantifier-free sentence and, by the completeness of propositional logic, there is an elementary formula $\mathsf{Tr}_0(x)$ such that, for all $m \in \mathbb{N}$, $\mathsf{Tr}_0(m)$ holds in $\mathbb{N}$ if and only if m is the Gödel code of a $\Delta_0(exp)$ sentence φ that holds in $\mathbb{N}$.

We can further construct *partial truth predicates* $\mathsf{Tr}_{\Pi_n}(x)$ and $\mathsf{Tr}_{\Sigma_n}(x)$ as follows:

$$\begin{aligned} \mathsf{Tr}_{\Pi_0}(x) = \mathsf{Tr}_{\Sigma_0}(x) &:= \mathsf{Tr}_0(x) \\ \mathsf{Tr}_{\Pi_{m+1}}(x) &:= \forall y \mathsf{Tr}_{\Sigma_m}(\ulcorner \varphi_x(\dot{y}) \urcorner) \\ \mathsf{Tr}_{\Sigma_{m+1}}(x) &:= \exists y \mathsf{Tr}_{\Pi_m}(\ulcorner \varphi_x(\dot{y}) \urcorner) \end{aligned} \tag{2.19}$$

Since $\mathsf{Tr}_0(x)$ is elementary, we see that $\mathsf{Tr}_{\mathcal{C}}(x)$ has arithmetical complexity $\mathcal{C}$.

The following lemma illustrates that the partial truth definitions indeed work as we would expect them to.

Lemma 2.3.3. *For $\mathcal{C} = \Sigma_n, \Pi_n$ and any $\varphi(x_1, \ldots, x_m) \in \mathcal{C}$, we have*

$$\mathsf{EA} \vdash \varphi(x_1, \ldots, x_m) \leftrightarrow \mathsf{Tr}_{\mathcal{C}}(\ulcorner \varphi(\dot{x}_1, \ldots, \dot{x}_m) \urcorner) \tag{2.20}$$

This allows us to define *partial satisfiability predicates*: we take $\mathsf{Sat}_{\mathcal{C}}(e, v, x)$ to denote "e is the Gödel code of a $\mathcal{C}$ formula φ with a single free variable with Gödel code v, and x is the Gödel number of a term such that $\varphi(x)$ holds in $\mathbb{N}$". This is equivalent to $\mathsf{Tr}_{\mathcal{C}}(\mathsf{sub}(e, v, \mathsf{num}(x)))$ so it is a $\mathcal{C}$ formula.

In particular, note that $\mathsf{Sat}_{\mathcal{C}}(e, v, v) = \mathsf{Tr}_{\mathcal{C}}(\ulcorner \varphi_e(\dot{v}) \urcorner)$ which, by the above lemma, is EA-provably equivalent to $\varphi_e(v)$. This lets us establish the following:

Corollary 2.3.4. $\mathsf{I}\Pi_n$ *and* $\mathsf{I}\Sigma_n$ *are finitely axiomatisable.*

Proof. We note that for $\varphi_e \in \mathcal{C}$, $\mathsf{Sat}_{\mathcal{C}}(e, v, v)$ and $\varphi_e(v)$ are EA-provably equivalent $\mathcal{C}$ formulae so their instances of $\mathsf{I}\mathcal{C}$ are EA-provably equivalent.

This means that $\mathsf{I}\Pi_n$ and $\mathsf{I}\Sigma_n$ are entailed by

$$\begin{aligned}
\Pi_n - \mathsf{Ind}(e) &:= \Pi_n(e) \to (\mathsf{Sat}(e, v, \mathsf{num}(0)) \\
&\quad \wedge \forall x(\mathsf{Sat}(e, v, \mathsf{num}(x) \to \mathsf{Sat}(e, v, \mathsf{num}(x+1)))) \\
&\quad \to \forall x \mathsf{Sat}(e, v, \mathsf{num}(x))) \\
\Sigma_n - \mathsf{Ind}(e) &:= \Sigma_n(e) \to (\mathsf{Sat}(e, v, \mathsf{num}(0)) \\
&\quad \wedge \forall x(\mathsf{Sat}(e, v, \mathsf{num}(x) \to \mathsf{Sat}(e, v, \mathsf{num}(x+1)))) \\
&\quad \to \forall x \mathsf{Sat}(e, v, \mathsf{num}(x)))
\end{aligned} \tag{2.21}$$

respectively. □

2.4 Representing ordinals in EA

In order to define transfinite induction principles (see 3.1), as well as transfinite progressions of reflection principles (see 4.2), the ordinal notation system used needs to meet certain requirements: namely, it needs to be an *elementary well-order* as described here:

Definition 2.4.1. *An* elementary linear order $(D, \prec)$ *is a pair of elementary formulae $x \in D$ and $x \prec y$ such that $\prec$ linearly orders D provably in* EA. *If $\prec$ is a well-order on D, we call it an* elementary well-order; *note that we do not require that the well-ordering be provable in* EA.

Note that since $x \in D$ is elementary, we can use variables ranging over D within EA.

Definition 2.4.2. *The function $\omega_n(\alpha)$ is defined for $n < \omega$ and an ordinal α by $\omega_0(\alpha) := \alpha, \omega_{n+1}(\alpha) := \omega^{\omega_n(\alpha)}$. The ordinal $\bigcup_{n<\omega} \omega_n(0)$ is denoted ε_0.*

For illustrative purposes, we will outline the construction of a particular elementary well-order, the *canonical ordinal notation system for ε_0*. This was first described by Gentzen in [4] in the context of Peano arithmetic and later adapted to weaker theories by various authors. Sommer in particular developed a formalisation in $\mathsf{I}\Delta_0$, a theory weaker than EA, in [11];

consequently the results in this section will be both easier to achieve and stronger than in Sommer's work.

Definition 2.4.3. *The* Smullyan base-n representation *of a natural number x is the sequence of digits $\overline{x}^n := [a_k, \ldots, a_0]$ such that $1 \leq a_i \leq n$ for all $i \in \{0, \ldots, k\}$ and*

$$x = a_k \cdot n^k + a_{k-1} \cdot n^{k-1} \ldots + a_0 \tag{2.22}$$

The sequence $\overline{x}^n$ is uniquely defined; in fact, for $n \geq 2$, the map $x \mapsto \overline{x}^n$ is a bijection $\mathbb{N} \to \{1, \ldots, n\}^*$. In particular, 0 maps to the empty sequence.

We take $*$ to denote concatenation of the elements of $\{1, \ldots, n\}^*$.

Definition 2.4.4. *The* Cantor normal form *of an ordinal $\alpha > 0$ is the unique representation*

$$\alpha := \omega^{\alpha_k} \cdot n_k + \omega^{\alpha_{k-1}} \cdot n_{k-1} + \ldots + \omega^{\alpha_1} \cdot n_1 \tag{2.23}$$

where $\alpha_k > \alpha_{k-1} > \ldots > \alpha_1$ and $0 < n_i < \omega$. For $\alpha = 0$ we take $k = 0$ so the representation consists of 0 terms.

Remark 2.4.5. *For all $\alpha < \varepsilon_0$, we have $\alpha > \alpha_k$ in the Cantor normal form of α.*

This means we can define an elementary coding for ordinals $\alpha = \omega^{\alpha_k} \cdot n_k + \omega^{\alpha_{k-1}} \cdot n_{k-1} + \ldots + \omega^{\alpha_1} \cdot n_1 < \varepsilon_0$ by first defining a function $f : \varepsilon_0 \to \mathbb{N}$ given by $f(0) = 0$ and

$$\overline{f(\alpha)}^4 = [4] * \overline{f(\alpha_k)}^4 * [3] * \overline{n_k}^2 * \ldots * [4] * \overline{f(\alpha_1)}^4 * [3] * \overline{n_1}^2 \tag{2.24}$$

We then fix a natural number N_{ord} distinct from N_{seq} and N_{expr} and define $\ulcorner \alpha \urcorner := j(N_{ord}, f(\alpha))$. This gives rise to an elementary formula $\mathsf{Ord}(x)$ which is true if and only if x is the code of an ordinal.

Proposition 2.4.6. *With this encoding of the ordinals below ε_0, the following* *functions are elementary:*

1. *the function nt that maps* $\ulcorner \alpha \urcorner$ *to the number of terms in the Cantor normal form of* α*;*
2. *the function t that, for* $1 \leq j \leq i \leq nt(\alpha)$*, maps* $(\ulcorner \alpha \urcorner, i, j)$ *to* $\ulcorner \omega^{\alpha_i} \cdot n_i + \ldots + \omega^{\alpha_j} \cdot n_j \urcorner$*;*
3. *the function e that, for* $1 \leq i \leq nt(\alpha)$*, maps* $(\ulcorner \alpha \urcorner, i)$ *to* $\ulcorner \alpha_i \urcorner$*;*
4. *the function c that, for* $1 \leq i \leq nt(\alpha)$*, maps* $(\ulcorner \alpha \urcorner, i)$ *to* n_i*.*

A number of basic results and properties of ordinal arithmetic can similarly be expressed in terms of elementary formulae; however, rather than listing and proving them here, we will be tacitly assuming that to be the case and, throughout the following chapters, point out whenever we make use of this assumption.

3. Transfinite induction principles

3.1 Transfinite induction principles

The results of this section are due to Sommer in [11] with some modifications to better suit the purposes of this thesis.

Definition 3.1.1 (Transfinite induction). *For $\alpha < \varepsilon_0$ and a formula φ, the* transfinite induction formula *up to α for φ, labelled $\mathsf{TI}(\alpha, \varphi)$, is the universal closure of*

$$\forall\beta((\forall\gamma < \beta)\varphi(\gamma) \to \varphi(\beta)) \to (\forall\gamma < \alpha)\varphi(\gamma) \tag{3.1}$$

The transfinite induction principle *up to α over a class $\mathcal{C}$ is the schema $\mathsf{TI}[\alpha, \mathcal{C}]$ given by $\{\mathsf{TI}(\ulcorner\alpha\urcorner, \varphi) | \varphi \in \mathcal{C}\}$. Note that, unlike with $\mathsf{TI}(\alpha, \varphi)$, here we do not let α be a variable.*

The subformula $\forall\beta((\forall\gamma < \beta)\varphi(\gamma) \to \varphi(\beta))$ is abbreviated as $\mathsf{Prog}(\varphi)$ and read as "φ is progressive *w.r.t. γ".*

In the following, we will restrict ourselves to the cases $\mathcal{C} = \Sigma_n, \Pi_n$. We can construct elementary formulae $\mathsf{Ax}_{\mathsf{TI}(\mathcal{C})}(\alpha, x)$, to present the family of theories $\mathsf{TI}[\alpha, \mathcal{C}]$ for $\alpha < \varepsilon_0$.

Lemma 3.1.2. *The following are provable in EA:*

i) $(\forall\alpha < \omega)(\Box_{\mathsf{EA}}\mathsf{TI}(\dot{\alpha}, \varphi))$*, since the formulae can be proved via finitely many applications of modus ponens.*

ii) $\forall\alpha(\forall\beta < \alpha)(\mathsf{TI}(\alpha, \varphi) \to \mathsf{TI}(\beta, \varphi))$*, since EA proves the transitivity of the ordinal $<$.*

iii) $\Box_{\mathsf{EA}+\mathsf{TI}[\omega,\mathcal{C}]}\varphi \leftrightarrow \Box_{\mathsf{EA}+\mathsf{I}\mathcal{C}}\varphi$ *since the function mapping natural numbers to their corresponding ordinals and its inverse are elementary.*

Lemma 3.1.3. *For all α and for all $m < \omega$ we have* $\mathsf{TI}[\alpha, \mathcal{C}] \equiv \mathsf{TI}[\alpha + m, \mathcal{C}]$ *over* EA.

Proof. Since EA proves that $\forall \alpha, \beta(\alpha + \beta \geq \alpha)$, the $\dashv$ direction follows from part ii) of 3.1.2.

For the $\vdash$ direction, we will informally proceed by meta-induction over m; however, as the codes of the proofs for each value of m are multiexponentially bounded as a function of m, this can be formalised in EA. Since EA proves $\alpha + 0 = \alpha$, the base case $m = 0$ is immediate. For $\varphi \in \mathcal{C}$, given $\mathsf{TI}(\alpha, \varphi) \to \mathsf{TI}(\alpha + m, \varphi)$, i.e. $\mathsf{Prog}(\varphi) \wedge (\forall \beta < \alpha)\varphi(\beta) \to (\forall \beta < \alpha + m)\varphi(\beta)$, note that $\mathsf{Prog}(\varphi)$ in particular implies $(\forall \beta < \alpha + m)\varphi(\beta) \to \varphi(\alpha + m)$. By modus ponens, we have $\mathsf{Prog}(\varphi) \wedge (\forall \beta < \alpha)\varphi(\beta) \to \varphi(\alpha + m)$. Combining this with the inductive hypothesis, we have

$$\mathsf{EA} \vdash \mathsf{Prog}(\varphi) \wedge (\forall \beta < \alpha)\varphi(\beta) \to (\forall \beta < \alpha + m)\varphi(\beta) \wedge \varphi(\alpha + m) \tag{3.2}$$

Since EA also proves that $\beta < \gamma + 1 \leftrightarrow \beta < \gamma \vee \beta = \gamma$, we conclude that $\mathsf{EA} \vdash \mathsf{Prog}(\varphi) \wedge (\forall \beta < \alpha)\varphi(\beta) \to (\forall \beta < \alpha + m + 1)\varphi(\beta)$, i.e. $\mathsf{EA} \vdash \mathsf{TI}(\alpha, \varphi) \to \mathsf{TI}(\alpha + m + 1, \varphi)$ as required. □

The following few results will be working towards a full classification of the schemata $\mathsf{TI}[\alpha, \Sigma_n]$ and $\mathsf{TI}[\alpha, \Pi_n]$ up to deductive equivalence. In particular, as we will see in 3.1.6 and 3.1.11, all such principles are equivalent to $\mathsf{TI}[\omega_2(\beta), \Pi_m]$ for some β, m dependent on α, n.

Proposition 3.1.4. *For any ordinal α and $1 \leq m, n < \omega$, we have* $\mathsf{TI}[\alpha, \Pi_n] \equiv \mathsf{TI}[\alpha^m, \Pi_n]$ *over* EA.

Proof. By part i) of 3.1.2 this holds for all $\alpha < \omega$ so suppose w.l.o.g. that $\alpha \geq \omega$.

We know the $\dashv$ direction by part ii) of 3.1.2 and by EA proving $m \geq 1 \to \alpha^m \geq \alpha$. Supposing the Cantor normal form of α is $\omega^{\alpha_k} \cdot n_k + \omega^{\alpha_{k-1}} \cdot n_{k-1} + \ldots + \omega^{\alpha_1} \cdot n_1$, we note that $\omega^{\alpha_k \cdot m} \leq \alpha^m \leq \omega^{\alpha_k \cdot (m+1)}$ so we can suppose w.l.o.g. $\alpha = \omega^\beta$ for some $\beta \geq 1$.

We proceed by (meta-level) induction on m. Since $(\omega^\beta)^0 \leq \omega^\beta$, the base case $m = 0$ follows by part ii) of 3.1.2. For the successor step, we need $\mathsf{TI}[\omega^{\beta \cdot m}, \Pi_n] \vdash \mathsf{TI}[\omega^{\beta \cdot (m+1)}, \Pi_n]$.

For $\varphi(\gamma) \in \Pi_n$, let $\psi(\delta) := (\forall \gamma < \omega^{\beta \cdot m} \cdot \delta)\varphi(\gamma)$. We note that $\psi(\delta) \in \Pi_n$. We will now show that $\mathsf{Prog}(\varphi) \vdash \mathsf{Prog}(\psi)$:

- if $\delta = 0$, then $\psi(\delta)$ is vacuously true, hence $\mathsf{Prog}(\psi)$ is vacuously true;
- if $\delta = \varepsilon + 1$, then $(\forall \delta' < \delta)\psi(\delta')$ is equivalent to $(\forall \delta' \leq \varepsilon)\psi(\delta')$ which, by the definition of ψ, implies $(\forall \gamma < \omega^{\beta \cdot m} \cdot \varepsilon)\varphi(\gamma)$.

  By $\mathsf{Prog}(\varphi)$, we have $\forall \zeta((\forall \zeta' < \zeta)\varphi(\omega^{\beta \cdot m} \cdot \varepsilon + \zeta') \rightarrow \varphi(\omega^{\beta \cdot m} \cdot \varepsilon + \zeta))$, i.e. $\varphi(\omega^{\beta \cdot m} \cdot \varepsilon + \zeta')$ is progressive w.r.t. ζ'. Applying $\mathsf{TI}[\omega^{\beta \cdot m}, \Pi_n]$ then yields $(\forall \zeta < \omega^{\beta \cdot m})\varphi(\omega^{\beta \cdot m} \cdot \varepsilon + \zeta)$ which in turn implies $(\forall \gamma < \omega^{\beta \cdot m} \cdot \delta)\varphi(\gamma)$. Applying $\mathsf{Prog}(\varphi)$ then implies $\varphi(\omega^{\beta \cdot m} \cdot \delta) = \psi(\delta)$;
- if δ is a non-zero limit ordinal, then, as $\omega^{\beta \cdot m} \cdot \delta = \bigcup_{\varepsilon < \delta} \omega^{\beta \cdot m} \cdot \varepsilon$, the inductive hypothesis $(\forall \varepsilon < \delta)\psi(\varepsilon)$ directly yields $(\forall \gamma < \omega^{\beta \cdot m} \cdot \delta)\varphi(\gamma)$ which, similarly to the successor case, implies $\psi(\delta)$ from $\mathsf{Prog}(\varphi)$, hence establishing $\mathsf{Prog}(\psi)$.

Applying $\mathsf{TI}[\omega^{\beta \cdot m}, \Pi_n]$ to ψ then gives us $(\forall \delta < \omega^{\beta \cdot m})\psi(\delta)$ which implies $(\forall \gamma < \omega^{\beta \cdot (m+1)})\varphi(\gamma)$, proving the instance of $\mathsf{TI}[\omega^{\beta \cdot (m+1)}, \Pi_n]$ for φ as required. □

Remark 3.1.5. *The above proof relies on the fact $\varphi \in \Pi_n \Rightarrow \psi \in \Pi_n$; however, $\varphi \in \Sigma_n$ would not imply $\psi \in \Sigma_n$ so the proof does not work for Σ_n. Indeed, we will see in 3.1.10 that $\mathsf{TI}[\omega, \Sigma_n] \nvdash \mathsf{TI}[\omega^2, \Sigma_n]$ for $n \geq 1$.*

Corollary 3.1.6. *For all $\omega \leq \alpha < \varepsilon_0$ and all $\beta < \alpha^\omega$, we have $\mathsf{TI}[\alpha, \Pi_n] \vdash \mathsf{TI}[\beta, \Pi_n]$. In particular, $\mathsf{TI}[\alpha, \Pi_n]$ is equivalent to $\mathsf{TI}[\omega_2(\gamma), \Pi_n]$ for some $\gamma < \varepsilon_0$.*

Proof. Given such α, β, we can find $m < \omega$ such that $\beta < \alpha^m$. Then, by 3.1.4 we have $\mathsf{TI}[\alpha, \Pi_n] \vdash \mathsf{TI}[\alpha^m, \Pi_n]$ and by part ii) of 3.1.2 we have $\mathsf{TI}[\alpha^m, \Pi_n] \vdash \mathsf{TI}[\beta, \Pi_n]$.

For the second statement, we have provably in EA that $\omega_2(0) = \omega$ and $(\omega_2(\gamma))^\omega = \omega_2(\gamma+1)$. Observe that for all $\omega \leq \alpha < \varepsilon_0$, there is some $\gamma < \varepsilon_0$ such that $\omega_2(\gamma) \leq \alpha < \omega_2(\gamma+1)$: more explicitly, γ is the exponent of the leading term in the Cantor normal form of α_k, the exponent of the leading term in the Cantor normal form of α. Then, since $\omega \leq \omega_2(\gamma) < \varepsilon_0$ and $\alpha < (\omega_2(\gamma))^\omega$, we conclude by the first half that $\mathsf{TI}[\omega_2(\gamma), \Pi_n] \vdash \mathsf{TI}[\alpha, \Pi_n]$ while the other direction follows by part ii) of 3.1.2. □

Remark 3.1.7. *It is worth noting that for $\alpha < \beta$, we have $\mathsf{TI}[\omega_2(\alpha), \Pi_n] \nvdash \mathsf{TI}[\omega_2(\beta), \Pi_n]$. Sommer shows this in Chapter 5 of [11] by using the* fast-growing hierarchy, *a transfinite hierarchy of functions of increasing rates of growth, to demonstrate that the set of functions in the hierarchy provably total within $\mathsf{TI}[\omega_2(\alpha), \Pi_n]$ strictly increases as α increases.*

However, 5.0.2 produces an alternative proof of this: $\mathsf{TI}[\omega_2(\alpha), \Pi_n]$ and $\mathsf{TI}[\omega_2(\beta), \Pi_n]$ are equivalent to different members of the same hierarchy of reflection principles which must be distinct by 4.2.6.

Proposition 3.1.8. *For $0 < m < \omega$, we have $\mathsf{TI}[\omega \cdot m, \Sigma_n] \equiv \mathsf{TI}[\omega, \Pi_n]$.*

Proof. By 2.3.2 and parts ii) and iii) of 3.1.2, it suffices to show $\mathsf{TI}[\omega, \Sigma_n] \vdash \mathsf{TI}[\omega \cdot m, \Sigma_n]$.

We shall proceed by meta-induction on m. This clearly holds for the base case $m = 1$. Now supposing $\mathsf{TI}[\omega, \Sigma_n] \vdash \mathsf{TI}[\omega \cdot m, \Sigma_n]$, it suffices to establish $\mathsf{TI}[\omega \cdot m, \Sigma_n] \vdash \mathsf{TI}[\omega \cdot (m+1), \Sigma_n]$ for the inductive step.

For $\varphi \in \Sigma_n$, $\mathsf{TI}[\omega \cdot m, \Sigma_n]$ establishes $\mathsf{Prog}(\varphi) \to (\forall \gamma < \omega \cdot m)\varphi(\gamma)$. Defining $\psi(\delta) := \varphi(\omega \cdot m + \delta) \in \Sigma_n$, we note that $(\forall \gamma < \omega \cdot m)\varphi(\gamma)$ and $\mathsf{Prog}(\varphi)$ entail $\mathsf{Prog}(\psi)$, therefore $\mathsf{Prog}(\varphi) \to \mathsf{Prog}(\psi)$. By $\mathsf{TI}[\omega, \Sigma_n] \subseteq \mathsf{TI}[\omega \cdot m, \Sigma_n]$ we have $\mathsf{Prog}(\psi) \to (\forall \delta < \omega)\psi(\delta)$, so we have $\mathsf{Prog}(\varphi) \to (\forall \gamma < \omega \cdot m + \omega)\varphi(\gamma)$ which is the corresponding instance of $\mathsf{TI}[\omega \cdot (m+1), \Sigma_n]$. □

Proposition 3.1.9. *For all $\alpha \geq \omega$, we have $\mathsf{TI}[\omega \cdot \alpha, \Sigma_n] \equiv \mathsf{TI}[\alpha, \Pi_{n+1}]$ over* EA.

Proof. For the $\dashv$ direction, note that by 3.1.4, $\alpha \geq \omega$ and part iii) of 3.1.2, we have $\mathsf{TI}[\alpha, \Pi_{n+1}] \vdash \mathsf{TI}[\alpha^2, \Pi_{n+1}] \vdash \mathsf{TI}[\omega \cdot \alpha, \Pi_{n+1}]$ which establishes the required result as $\Sigma_n \subseteq \Pi_{n+1}$.

For the $\vdash$ direction, we write $\varphi \in \Pi_{n+1}$ as $\forall x \psi(x, \gamma)$ with $\psi \in \Sigma_n$ and, observing that $x < \omega, \gamma < \alpha \iff \omega \cdot \gamma + x < \omega \cdot \alpha$, define $\theta(\omega \cdot \gamma + x) := \psi(x, \gamma)$.

By 2.4.6, given that γ has the Cantor normal form $\gamma = \omega^{\gamma_k} \cdot n_k + \ldots + \omega^{\gamma_1} \cdot n_1$, the functions $e(\ulcorner \gamma \urcorner, 1) = \ulcorner \gamma_1 \urcorner$ and $c(\ulcorner \gamma \urcorner, 1) = n_1$ are elementary, so is the function

$$g_0(\gamma) := \begin{cases} n_1 & \text{if } \gamma_1 = 0 \\ 0 & \text{otherwise} \end{cases} \tag{3.3}$$

As EA also proves the totality and uniqueness of truncated ordinal right subtraction, the function $g_1(\gamma) := \gamma \dot{-} g_0(\gamma)$ (i.e. one that maps γ to the ordinal of the same Cantor normal form except without the ω^0 term, if it had one) is also elementary.

In particular, by bounded minimisation we can find the smallest j in $1 \leq j \leq k$ such that $\gamma_j \geq \omega$. This in turn means we can define the function

$$g_2(\gamma) := \omega^{\gamma_k} \cdot n_k + \ldots + \omega^{\gamma_j} \cdot n_j + \omega^{\gamma_{j-1} \dot{-} 1} \cdot n_{j-1} + \ldots + \omega^{\gamma_1 \dot{-} 1} \cdot n_1 \tag{3.4}$$

It follows that $g_2(\gamma)$ is elementary. We also observe that $\gamma = \omega \cdot g_2(g_1(\gamma)) + g_0(\gamma)$; in fact, we have

$$\mathsf{EA} \vdash \forall \gamma \forall x (g_0(\omega \cdot \gamma + x) = x \wedge g_2(g_1(\omega \cdot \gamma + x)) = \gamma) \tag{3.5}$$

Moreover, we have

$$\mathsf{EA} \vdash \alpha < \beta \rightarrow \forall x (\omega \cdot \alpha + x < \omega \cdot \beta) \tag{3.6}$$

This means that we can formalise our definition of $\theta(\gamma)$ as $\psi(g_0(\gamma), g_2(g_1(\gamma)))$.

It remains to show that $\mathsf{EA} \vdash \mathsf{Prog}(\varphi) \to \mathsf{Prog}(\theta)$, as then we can apply $\mathsf{TI}(\omega \cdot \alpha, \theta)$ and deduce $\mathsf{TI}(\alpha, \varphi)$.

Suppose $\mathsf{Prog}(\varphi)$ and $(\forall \gamma < \beta)\theta(\gamma)$, we want to deduce $\theta(\beta)$. By 3.6, we have for all $\gamma < g_2(g_1(\beta))$ and for all $x < \omega$ that $\omega \cdot \gamma + x < \omega \cdot g_2(g_1(\beta)) \leq \beta$, thus we have $(\forall \gamma < g_2(g_1(\beta)))\forall x\ \psi(g_0(\omega \cdot \gamma + x), g_2(g_1(\omega \cdot \gamma + x)))$. By 3.5, this implies $(\forall \gamma < g_2(g_1(\beta)))\forall x \psi(x, \gamma)$, i.e. $(\forall \gamma < g_2(g_1(\beta)))\varphi(\gamma)$. By $\mathsf{Prog}(\varphi)$, this in turn implies $\varphi(g_2(g_1(\beta)))$, i.e. $\forall x \psi(x, g_2(g_1(\beta)))$ whence in particular $\psi(g_0(\beta), g_2(g_1(\beta)))$ which in turn establishes $\theta(\beta)$, as required. □

Remark 3.1.10. *We see that* $\mathsf{TI}[\omega, \Sigma_n] \equiv \mathsf{I}\Sigma_n$ *but by 3.1.9 we have* $\mathsf{TI}[\omega^2, \Sigma_n] \equiv \mathsf{TI}[\omega, \Pi_{n+1}] \equiv \mathsf{I}\Pi_{n+1}$ *with* $\mathsf{I}\Sigma_n \subsetneq \mathsf{I}\Pi_{n+1}$*, as established by Parsons in [9].*

Corollary 3.1.11. *For all* $\omega \leq \alpha < \varepsilon_0$*,* $\mathsf{TI}[\alpha, \Sigma_n]$ *is equivalent to* $\mathsf{TI}[\omega_2(\beta), \Pi_m]$ *for some* $m < \omega, \beta < \varepsilon_0$*.*

Proof. By 3.1.3 and $\omega \cdot (\gamma + 1) = \omega \cdot \gamma + \omega$, every $\mathsf{TI}[\alpha, \Sigma_n]$ is equivalent to $\mathsf{TI}[\omega \cdot \gamma, \Sigma_n]$ for some $\gamma < \varepsilon_0$. Combining 3.1.8 and 3.1.9, it is equivalent to some $\mathsf{TI}[\delta, \Pi_m]$ which by 3.1.6 is equivalent to some $\mathsf{TI}[\omega_2(\beta), \Pi_m]$ as required. □

Remark 3.1.12. *Similarly to our proof of 2.3.4, we observe that for each* α *and each* n *we can define predicates* $\mathsf{TI}_{\alpha,n}(x)$ *denoting "*x *is the Gödel code of an instance of* $\mathsf{TI}[\omega_2(\alpha), \Pi_n]$*". Since these instances are* Π_{n+2} *sentences, we conclude that* $\mathsf{TI}[\omega_2(\alpha), \Pi_n]$ *is axiomatised by the formula*

$$\mathsf{ti}_n(\alpha) := \forall e, v(\mathsf{TI}_{\alpha,n}(e) \to \mathsf{Sat}_{\Pi_{n+2}}(e, \mathsf{num}(v), \mathsf{num}(v))) \tag{3.7}$$

4. Reflection principles

4.1 Reflection principles

The results of this chapter are summarised, along with many others, in [1] and [2] by Beklemishev.

As in Chapter 3, unless otherwise stated, T is assumed to be an elementarily presented theory containing EA. We introduce the notation $\Diamond_T(\overline{\ulcorner\varphi\urcorner})$ (abbreviated as $\Diamond_T\varphi$), meaning $\neg\Box_T\neg\varphi$. Note that this is equivalent to $\mathsf{Con}(T+\varphi)$.

Definition 4.1.1. *The* local reflection principle *for a theory T is the schema* $\mathsf{Rfn}(T)$ *consisting of*

$$\Box_T\varphi \to \varphi \tag{4.1}$$

for all sentences φ. Analogously to the generalisation of Löb's conditions in 2.2.9, the uniform reflection principle *is the schema* $\mathsf{RFN}(T)$ *consisting of*

$$\forall x(\Box_T\varphi(\dot{x}) \to \varphi(x)) \tag{4.2}$$

for all formulae $\varphi(x)$. The partial reflection principles*, denoted* $\mathsf{Rfn}_{\mathcal{C}}(T)$ *and* $\mathsf{RFN}_{\mathcal{C}}(T)$*, respectively, are obtained by restricting the sentences/formulae φ to the class $\mathcal{C}$; we shall mainly consider the cases $\mathcal{C} = \Sigma_n, \Pi_n$.*

Proposition 4.1.2. $\mathsf{Rfn}_{\Pi_1}(T) \equiv \mathsf{RFN}_{\Pi_1}(T) \equiv \mathsf{Con}(T)$ *over* EA.

Proof. Let $\varphi(x) \in \Pi_1$. By 2.2.11, we have

$$\begin{aligned} \vdash \Box_T\varphi(\dot{x}) \wedge \neg\varphi(x) &\rightarrow \Box_T\neg\varphi(\dot{x}) \\ &\rightarrow \Box_T(\varphi(\dot{x}) \wedge \neg\varphi(\dot{x})) \\ &\rightarrow \Box_T\bot \end{aligned} \tag{4.3}$$

By contraposition, $\neg\Box_T\bot \rightarrow (\Box_T\varphi(\dot{x}) \rightarrow \varphi(x))$, i.e. $\mathsf{Con}(T) \vdash \mathsf{RFN}_{\Pi_1}(T)$. Moreover, $\mathsf{RFN}_{\Pi_1}(T) \vdash \mathsf{Rfn}_{\Pi_1}(T)$ is clear by the definitions, so it remains to establish $\mathsf{Rfn}_{\Pi_1}(T) \vdash \mathsf{Con}(T)$. For that, note that we can think of $\mathsf{Con}(T)$ as $\Box_T\bot \rightarrow \bot$, an instance of $\mathsf{Rfn}_{\Pi_1}(T)$. □

Proposition 4.1.3. *For each $n \geq 1$, we have* $\mathsf{RFN}_{\Sigma_n}(T) \equiv \mathsf{RFN}_{\Pi_{n+1}}(T)$.

Proof. The $\dashv$ entailment is clear since $\Sigma_n \subseteq \Pi_{n+1}$. For the $\vdash$ direction, take $\varphi(x) \in \Pi_{n+1}$ and rewrite it as $\forall y\psi(x,y)$ for $\psi \in \Sigma_n$. By 2.2.10 and the definition of $\mathsf{RFN}_{\Sigma_n}(T)$, we have

$$\begin{aligned} \mathsf{EA} + \mathsf{RFN}_{\Sigma_n}(T) \vdash \Box_T\forall y\psi(\dot{x},y) &\rightarrow \forall y\Box_T\psi(\dot{x},\dot{y}) \\ &\rightarrow \forall y\psi(x,y) \end{aligned} \tag{4.4}$$

i.e. $\mathsf{EA} + \mathsf{RFN}_{\Sigma_n}(T)$ proves the instance of $\mathsf{RFN}_{\Pi_{n+1}}(T)$ for φ as required.

□

We now offer some alternate characterisations of $\mathsf{RFN}_{\mathcal{C}}(T)$. The first one shows it to be finitely axiomatisable.

Lemma 4.1.4. *Over* EA, $\mathsf{RFN}_{\mathcal{C}}(T)$ *is equivalent to*

$$\forall x(\Box_T\mathsf{Tr}_{\mathcal{C}}(\dot{x}) \rightarrow \mathsf{Tr}_{\mathcal{C}}(x)) \tag{4.5}$$

Proof. As $\mathsf{Tr}_{\mathcal{C}}(x)$ has arithmetical complexity $\mathcal{C}$, Equation 4.5 is an instance of $\mathsf{RFN}_{\mathcal{C}}(T)$ so it is clearly entailed by it. For the converse, we have the following:

$$
\begin{aligned}
&\mathsf{EA} \vdash \forall x_1, \ldots, x_n (\varphi(x_1, \ldots, x_n) \leftrightarrow \mathsf{Tr}_{\mathcal{C}}(\ulcorner \varphi(\dot{x}_1, \ldots, \dot{x}_n) \urcorner)) \\
&\quad \text{[generalisation applied to 2.3.3]} \\
&\ T \vdash \forall x_1, \ldots, x_n (\varphi(x_1, \ldots, x_n) \leftrightarrow \mathsf{Tr}_{\mathcal{C}}(\ulcorner \varphi(\dot{x}_1, \ldots, \dot{x}_n) \urcorner)) \\
&\quad \text{[since } T \text{ contains } \mathsf{EA} \text{ by assumption]} \\
&\mathsf{EA} \vdash \Box_T \forall x_1, \ldots, x_n (\varphi(x_1, \ldots, x_n) \leftrightarrow \mathsf{Tr}_{\mathcal{C}}(\ulcorner \varphi(\dot{x}_1, \ldots, \dot{x}_n) \urcorner)) \\
&\quad \text{[by Löb's 1st condition]} \\
&\mathsf{EA} \vdash \forall x_1, \ldots, x_n \Box_T (\varphi(\dot{x}_1, \ldots, \dot{x}_n) \leftrightarrow \mathsf{Tr}_{\mathcal{C}}(\ulcorner \varphi(\dot{x}_1, \ldots, \dot{x}_n) \urcorner)) \\
&\quad \text{[by 2.2.10]} \\
&\mathsf{EA} \vdash \Box_T (\varphi(\dot{x}_1, \ldots, \dot{x}_n) \leftrightarrow \mathsf{Tr}_{\mathcal{C}}(\ulcorner \varphi(\dot{x}_1, \ldots, \dot{x}_n) \urcorner)) \\
&\quad \text{[by the axiom } \forall x \varphi(x) \to \varphi(x)\text{]} \\
&\mathsf{EA} \vdash \Box_T \varphi(\dot{x}_1, \ldots, \dot{x}_n) \leftrightarrow \Box_T (\mathsf{Tr}_{\mathcal{C}}(\ulcorner \varphi(\dot{x}_1, \ldots, \dot{x}_n) \urcorner)) \\
&\quad \text{[by Löb's 2nd condition and modus ponens]}
\end{aligned}
\tag{4.6}
$$

We thus have

$$
\begin{aligned}
\mathsf{EA} \vdash \Box_T \varphi(\dot{x}_1, \ldots, \dot{x}_n) &\to \Box_T (\mathsf{Tr}_{\mathcal{C}}(\ulcorner \varphi(\dot{x}_1, \ldots, \dot{x}_n) \urcorner)) \\
&\to \mathsf{Tr}_{\mathcal{C}}(\ulcorner \varphi(\dot{x}_1, \ldots, \dot{x}_n) \urcorner) \\
&\to \varphi(x_1, \ldots, x_n)
\end{aligned}
\tag{4.7}
$$

where the second implication follows from applying Equation 4.5 to $x = \ulcorner \varphi(\dot{x}_1, \ldots, \dot{x}_n) \urcorner$. As this is the instance of $\mathsf{RFN}_{\mathcal{C}}(T)$ for $\varphi(x_1, \ldots, x_n) \in \mathcal{C}$, we are done. $\square$

By the following result due to Leivant in [6] and Ono in [7], for $T = \mathsf{EA}$ the finite axiomatisability is also proved by the following argument: for $n \geq 2$, $\mathsf{RFN}_{\Sigma_n}(\mathsf{EA})$ is equivalent to a partial induction principle which we established to be finitely axiomatisable in 2.3.4.

Theorem 4.1.5 (Leivant's theorem)**.** *For each $n \geq 1$, we have* $\mathsf{EA} + \mathsf{RFN}_{\Sigma_{n+1}}(\mathsf{EA}) \equiv$

$\mathsf{I}\Sigma_n$.

This yields the classical result by Kreisel and Lévy established in [5]:

Corollary 4.1.6 (Kreisel, Lévy)**.** $\mathsf{PA} \equiv \mathsf{EA} + \mathsf{RFN}(\mathsf{EA})$.

We now give the other characterisation.

Lemma 4.1.7. *For each $n \geq 1$ and each elementarily presented theory S containing* EA*,* $\mathsf{RFN}_{\Pi_{n+2}}(S)$ *is equivalent to the schema* $\forall x(\pi(x) \rightarrow \Diamond_S \pi(\dot{x}))$ *for all* $\pi \in \Pi_{n+1}$.

Proof. By 4.1.3, contraposition and substituting $\pi := \neg\varphi$, we have

$$\begin{aligned} \mathsf{RFN}_{\Pi_{n+2}}(S) &\equiv \mathsf{RFN}_{\Sigma_{n+1}}(S) \\ &\equiv \{\forall x(\Box_S \varphi(\dot{x}) \rightarrow \varphi(x)) | \varphi \in \Sigma_{n+1}\} \\ &\equiv \{\forall x(\neg\varphi(x) \rightarrow \neg\Box_S \varphi(\dot{x})) | \varphi \in \Sigma_{n+1}\} \\ &\equiv \{\forall x(\pi(x) \rightarrow \neg\Box_S \neg\pi(\dot{x})) | \pi \in \Pi_{n+1}\} \\ &\equiv \{\forall x(\pi(x) \rightarrow \Diamond_S \pi(\dot{x})) | \pi \in \Pi_{n+1}\} \end{aligned} \tag{4.8}$$

□

By considering the totality statement for the superexponential function $supexp(n, x)$, one can obtain this result which demonstrates that most of the theories we are going to consider in fact contain EA^+ which we introduced in 2.1.9:

Proposition 4.1.8. $\mathsf{RFN}_{\Pi_2}(\mathsf{EA}) \vdash \mathsf{EA}^+$

We present two results which will be of use in the next chapter:

Proposition 4.1.9. *For U, V elementarily presented theories containing* EA*, we have*

$$U \equiv_{\Pi_n} V \Rightarrow \mathsf{EA}^+ \vdash \mathsf{RFN}_{\Pi_n}(U) \equiv \mathsf{RFN}_{\Pi_n}(V) \tag{4.9}$$

More explicitly, this is provable in EA^+*, i.e.*

$$\begin{aligned}\mathsf{EA}^+ \vdash \forall y(\Pi_n(y) \to (\Box_U(y) \leftrightarrow \Box_V(y))) \\ \to \forall y(\Box_{\mathsf{RFN}_{\Pi_n}(U)}(y) \leftrightarrow \Box_{\mathsf{RFN}_{\Pi_n}(V)}(y))\end{aligned} \tag{4.10}$$

Proof. It suffices to show that, assuming $\forall y(\Pi_n(y) \to (\Box_U(y) \leftrightarrow \Box_V(y)))$, EA^+ proves $\mathsf{RFN}_{\Pi_n}(U) \vdash \mathsf{RFN}_{\Pi_n}(V)$ (the other implication follows by the symmetry of the argument). For $\varphi(x) \in \Pi_n$, if $\Box_V\varphi(\dot{x})$, then by $\forall y(\Pi_n(y) \to (\Box_U(y) \leftrightarrow \Box_V(y)))$ we have $\Box_U\varphi(\dot{x})$. By $\mathsf{RFN}_{\Pi_n}(U)$ we have $\forall x(\Box_U\varphi(\dot{x}) \to \varphi(x))$, so we obtain $\forall x(\Box_V\varphi(\dot{x}) \to \varphi(x))$, i.e. the corresponding instance of $\mathsf{RFN}_{\Pi_n}(V)$. $\Box$

Lemma 4.1.10. $\mathsf{EA} \vdash \forall \pi \in \Pi_m((\mathsf{RFN}_{\Sigma_m}(T) \wedge \mathsf{Tr}_{\Pi_m}(\pi)) \leftrightarrow \mathsf{RFN}_{\Sigma_m}(T+\pi))$

Proof. We only need to prove the $\to$ implication. Let $\sigma(y) \in \Sigma_m$. We reason in EA: by a formalisation of the deduction theorem, $\Box_{T+\pi}\sigma(\dot{y})$ implies $\Box_T(\pi \to \sigma(\dot{y}))$. By 2.3.3, this in turn implies $\Box_T(\mathsf{Tr}_{\Pi_m}(\dot{\pi}) \to \sigma(\dot{y}))$. Modulo logical equivalence, $\mathsf{Tr}_{\Pi_m}(\pi) \to \sigma(y)$ is a Σ_m formula. Hence, by applying $\mathsf{RFN}_{\Sigma_m}(T)$ to $\Box_T(\mathsf{Tr}_{\Pi_m}(\dot{\pi}) \to \sigma(\dot{y}))$, we obtain $\mathsf{Tr}_{\Pi_m}(\pi) \to \sigma(y)$. Since we assume $\mathsf{Tr}_{\Pi_m}(\pi)$, this yields $\sigma(y)$, hence establishing the corresponding instance of $\mathsf{RFN}_{\Sigma_m}(T+\pi)$. $\Box$

It is also worth noting that one of the earliest results regarding transfinite induction principles established, by Kreisel and Lévy in [5], relates it to a reflection principle:

Theorem 4.1.11. $\mathsf{TI}[\varepsilon_0] := \bigcup_{n<\omega} \mathsf{TI}[\varepsilon_0, \Pi_n] \equiv \mathsf{PA} + \mathsf{RFN}(\mathsf{PA})$

In particular, this classical result implies that $\mathsf{TI}[\varepsilon_0]$ cannot be axiomatised by formulae of bounded complexity over PA.

4.2 Transfinite progressions of reflection principles

We let $(D, \prec)$ be an elementary well-order as defined in 2.4.1. Starting from a given theory T, we consider a progression of theories satisfying

$\mathsf{RFN}^{\alpha}_{\mathcal{C}}(T) := T + \{\mathsf{RFN}_{\mathcal{C}}(\mathsf{RFN}^{\beta}_{\mathcal{C}}(T)) | \beta \prec \alpha\}$ (this in itself does not amount to a formal definition, that is established by 4.2.1). Note in particular that $\mathsf{RFN}^{0}_{\mathcal{C}}(T) = T$.

It is worth noting that this is an *elementarily presented family*, meaning there is an elementary formula $\mathsf{Ax}_T(\alpha, x)$ which expresses "x is the code of an axiom in $\mathsf{RFN}^{\alpha}_{\mathcal{C}}(T)$". We already established in 4.1.4 that instances of $\mathsf{RFN}_{\mathcal{C}}$ can be elementarily expressed: more explicitly, we will define an elementary formula $\mathsf{Rcode}(e, x)$ for a theory U denoting "e is the code of the Σ_1 formula $\varphi(v) := \Box_U v$ and x is the code of an instance of $\mathsf{RFN}_{\mathcal{C}}(U)$". We also define the Σ_1 predicate $\Box_T(\alpha, x)$ meaning "$\mathsf{RFN}^{\alpha}_{\mathcal{C}}(T)$ proves the formula coded by x". Then the following results can be established:

Lemma 4.2.1. *For any elementary linear order* $(D, \prec)$ *and theory* T*, there is an elementary formula* $\mathsf{Ax}_T(\alpha, x)$ *such that*

$$\mathsf{EA} \vdash \mathsf{Ax}_T(\alpha, x) \leftrightarrow (\mathsf{Ax}_T(x) \vee \exists \beta \prec \alpha(\mathsf{Rcode}(\ulcorner \Box_T(\dot{\beta}, v) \urcorner, x))) \tag{4.11}$$

A formula satisfying this property is called an explicit numeration.

In particular, the new provability predicate is monotone with respect to $\prec$:

Lemma 4.2.2. $\mathsf{EA} \vdash \alpha \prec \beta \to (\Box_T(\alpha, x) \to \Box_T(\beta, x))$.

The proof of 4.2.1 uses the fixed-point lemma 2.2.6 which gives rise to the concern that the formula $\mathsf{Ax}_T(\alpha, x)$ might depend on the construction of the fixed point. The following result, however, shows that this is not the case, as the progression obtained depends solely on the initial theory and the elementary linear order:

Lemma 4.2.3. *For* U, V *elementarily presented theories containing* EA, $(D, \prec)$ *an elementary linear order, we have*

$$\mathsf{EA} \vdash \forall x(\Box_U(x) \leftrightarrow \Box_V(x)) \Rightarrow \mathsf{EA} \vdash \forall \alpha \forall x(\Box_U(\alpha, x) \leftrightarrow \Box_V(\alpha, x)) \tag{4.12}$$

The proof relies on a technique known as *reflexive induction* or *reflexive progressivity* which can be thought of as induction which uses the provability, rather than the truth, of the previous cases as the inductive hypothesis.

Theorem 4.2.4 (Reflexive induction)**.** *For any elementary linear ordering $(D, \prec)$, T is closed under the following rule:*

$$\forall\alpha(\Box_T(\forall\beta \prec \dot{\alpha}\ A(\beta)) \to A(\alpha)) \vdash \forall\alpha A(\alpha) \tag{4.13}$$

We refer to the subformula $\Box_T(\forall\beta \prec \dot{\alpha}\ A(\beta))$ as the (reflexive) induction hypothesis*.*

Proof. We note that $T \vdash \Box_T\forall\alpha A(\alpha) \to \forall\alpha(\Box_T(\forall\beta \prec \dot{\alpha}\ A(\beta)))$ which, given the assumption $T \vdash \forall\alpha(\Box_T(\forall\beta \prec \dot{\alpha}\ A(\beta)) \to A(\alpha))$, implies $T \vdash \Box_T\forall\alpha A(\alpha) \to \forall\alpha A(\alpha)$. We conclude by 2.2.7. □

For further details and proofs of Lemmata 4.2.1 and 4.2.2 for the case $\mathcal{C} = \Pi_1$, see Chapter 2 of [1]. To make the previous results more explicit, we can establish the following:

Lemma 4.2.5. *The theories $\mathsf{RFN}^{\alpha+1}_{\Pi_n}(\mathsf{EA})$ are finitely axiomatisable; in particular, $\mathsf{RFN}^{\alpha+1}_{\Pi_n}(\mathsf{EA})$ is equivalent to an elementary formula $\mathsf{re}_n(\alpha)$ satisfying the following:*

$$\begin{aligned} \mathsf{re}_n(0) &\leftrightarrow \forall y(\Pi_n(y) \wedge \Box_T y \to \mathsf{Tr}_{\Pi_n}(y)) \\ \mathsf{re}_n(\beta+1) &\leftrightarrow \forall y(\Pi_n(y) \wedge \Box_T(\mathsf{re}_n(\dot{\beta}) \to y) \to \mathsf{Tr}_{\Pi_n}(y)) \\ \mathsf{re}_n(\lambda) &\leftrightarrow \forall\alpha < \lambda(\mathsf{re}_n(\alpha)) \textit{ for } \lambda \textit{ a non-zero limit ordinal.} \end{aligned} \tag{4.14}$$

Remark 4.2.6. *Since by Gödel's and Rosser's results, none of the theories in the progression $\mathsf{RFN}^{\alpha}_{\mathcal{C}}(T)$ can prove their own consistency while each iterand proves the previous one's consistency (assuming $\mathcal{C} = \Sigma_n, \Pi_n$ for $n \geq 1$), the progression must be strictly increasing.*

5. $\mathsf{TI}[\omega_2(\alpha), \Pi_n]$ in terms of $\mathsf{RFN}^{\beta}_{\Pi_m}(\mathsf{EA})$

In this chapter we will establish an expression for the transfinite induction principles $\mathsf{TI}[\omega_2(\alpha), \Pi_n]$ in terms of the reflection principles $\mathsf{RFN}^{\beta}_{\Pi_m}(\mathsf{EA})$ where m will be a function of n and β will be a function of α.

We already have the necessary results to derive such an expression for $\alpha = 0$. Note that $\mathsf{TI}[\omega_2(0), \Pi_n] = \mathsf{TI}[\omega, \Pi_n]$ which by part iii) of 3.1.2 is equivalent to $\mathsf{I}\Pi_n$ over EA. By 2.3.2, this is equivalent to $\mathsf{I}\Sigma_n$. This means that the desired result is obtained by combining 4.1.5 and 4.1.3.

Corollary 5.0.1. *For each $n \geq 1$, $\mathsf{TI}[\omega, \Pi_n] \equiv \mathsf{RFN}_{\Pi_{n+2}}(\mathsf{EA})$ over EA.*

The following generalisation was first conjectured in the correspondence between Beklemishev and Freund.

Theorem 5.0.2. *For each $n \geq 1, \alpha \geq 0$, $\mathsf{TI}[\omega_2(\alpha), \Pi_n] \equiv \mathsf{RFN}^{\alpha+1}_{\Pi_{n+2}}(\mathsf{EA})$ over EA.*

5.1 $\dashv$ entailment

One of the entailments for 5.0.2 was proved by Schmerl in [10], with PRA instead of EA as the base theory.

Theorem 5.1.1. *For all $\alpha \geq 0, n \geq 1$, we have $\mathsf{RFN}^{\alpha+1}_{\Pi_{n+2}}(\mathsf{EA}) \vdash \mathsf{TI}[\omega_2(\alpha), \Pi_n]$.*

Proof. By 4.2.4, it suffices to show $\forall\alpha(\Box_{\mathsf{EA}}(\forall\beta < \dot{\alpha}(\mathsf{re}_{n+2}(\beta) \to \mathsf{ti}_n(\omega_2(\beta)))) \to (\mathsf{re}_{n+2}(\alpha) \to \mathsf{ti}_n(\omega_2(\alpha))))$, so suppose that the reflexive inductive hypothesis $\Box_{\mathsf{EA}}(\forall\beta < \dot{\alpha}(\mathsf{re}_{n+2}(\beta) \to \mathsf{ti}_n(\omega_2(\beta))))$ holds. We will present the remainder of the argument informally.

Note that $\mathsf{re}_{n+2}(\alpha) \equiv \mathsf{RFN}_{\Pi_{n+2}}(\bigcup_{\beta<\alpha} \mathsf{re}_{n+2}(\dot{\beta})) \vdash \mathsf{RFN}_{\Pi_{n+2}}(\bigcup_{\beta<\alpha} \mathsf{ti}_n(\omega_2(\dot{\beta})))$ by the inductive hypothesis. Observe that $S_\alpha := \bigcup_{\beta<\alpha} \mathsf{ti}_n(\omega_2(\dot{\beta}))$ is an elementarily presented family with $\mathsf{Ax}_S(\alpha, x) = \mathsf{Ax}_{\mathsf{EA}}(x) \vee (\exists\beta < \alpha)(x = \ulcorner \mathsf{ti}_n(\omega_2(\dot{\beta}))\urcorner)$.

Formalising the proof of 3.1.6, we have $(\forall\gamma < \omega_2(\alpha))\Box_{S_\alpha}\mathsf{ti}_n(\dot\gamma)$ (since for each $\gamma < \omega_2(\alpha)$, we can elementarily find $\beta < \alpha$ such that $\omega_2(\beta) \leq \gamma < \omega_2(\beta+1)$); therefore $S_\alpha \equiv \bigcup_{\gamma<\omega_2(\alpha)} \mathsf{ti}_n(\dot\gamma)$.

It remains to show $\mathsf{RFN}_{\Pi_{n+2}}(S_\alpha) \vdash \mathsf{TI}[\omega_2(\alpha), \Pi_n]$; we will establish this by showing it implies each instance. Let $\varphi \in \Pi_n$ be arbitrary. Since $\mathsf{Prog}(\varphi) \in \Pi_{n+1}$, we have $\mathsf{TI}(\gamma,\varphi) \in \Pi_{n+2}$. This shows that $\mathsf{RFN}_{\Pi_{n+2}}(S_\alpha) \vdash \forall\gamma(\Box_{S_\alpha}(\mathsf{TI}(\dot\gamma,\varphi)) \rightarrow \mathsf{TI}(\gamma,\varphi))$, so it remains to establish $\mathsf{RFN}_{\Pi_{n+2}}(S_\alpha) \vdash \Box_{S_\alpha}(\mathsf{TI}(\omega_2(\dot\alpha),\varphi))$. By Löb's first condition, it will suffice to show $S_\alpha \vdash \mathsf{Prog}(\varphi) \rightarrow (\forall\gamma < \omega_2(\dot\alpha))\varphi(\gamma)$. This, however, follows from the equivalence $S_\alpha \equiv \bigcup_{\gamma<\omega_2(\alpha)} \mathsf{ti}_n(\dot\gamma)$.

□

Schmerl also offers a partial converse:

Theorem 5.1.2. *For $\alpha \geq \omega$ and $n \geq 1$, we have* $\mathsf{TI}[\alpha,\Pi_n] \vdash \mathsf{RFN}^{\alpha+1}_{\Pi_n}(\mathsf{EA})$.

However, this is considerably weaker and is not strict even in the base case $\alpha = \omega$. In the following section, we will establish a full converse.

5.2 $\vdash$ entailment

By 4.1.7, it suffices to show the following for the $\vdash$ entailment of 5.0.2:

Theorem 5.2.1. $\mathsf{TI}[\omega_2(\alpha),\Pi_n] \vdash \forall z(\pi(z) \rightarrow \Diamond_{\mathsf{RFN}^{\alpha}_{\Pi_{n+2}}(\mathsf{EA})}\pi(\dot z))$ *for all* $\pi(z) \in \Pi_{n+1}$ *over* EA.

We shall first prove a similar result for $\pi \in \Pi_{n+1}$ restricted to sentences:

Theorem 5.2.2. *For all sentences $\pi \in \Pi_{n+1}$, we have* $\mathsf{TI}[\omega_2(\alpha),\Pi_n] \vdash \pi \rightarrow \mathsf{RFN}_{\Pi_n}(\mathsf{RFN}^{\alpha}_{\Pi_{n+2}}(\mathsf{EA}) + \pi)$.

We begin with the following lemma:

Lemma 5.2.3. *For each ordinal λ and each elementarily presented theory T containing* EA, *we have* $\mathsf{TI}[\lambda,\Pi_n] + \mathsf{RFN}_{\Sigma_n}(T) \vdash \mathsf{RFN}^{\lambda+1}_{\Pi_n}(T)$.

Proof. A formalisation of the following meta-inductive proof for the claim "for T a Σ_n-sound elementarily presented theory containing EA, $\mathsf{RFN}^{\alpha}_{\Pi_n}(T)$ is Π_n-sound for all ordinals α": we have provably in EA that $\mathsf{RFN}^{\alpha}_{\Pi_n}(T) \equiv T + \{\mathsf{RFN}^{\beta+1}_{\Pi_n}(T) | \beta < \alpha\}$ and by the inductive hypothesis, each $\mathsf{RFN}^{\beta}_{\Pi_n}(T)$

is Π_n-sound, hence the axioms $\mathsf{RFN}^{\beta+1}_{\Pi_n}(T)$ are true Π_n sentences. Supposing $T + \{\mathsf{RFN}^{\beta+1}_{\Pi_n}(T) | \beta < \alpha\}$ proves a false Σ_n sentence ψ, the proof would use a finite conjunction φ of the Π_n sentences $\mathsf{RFN}^{\beta_i+1}_{\Pi_n}(T))$ for $\beta_1, \ldots, \beta_k < \alpha$. By the deduction theorem, we would then have $T \vdash \varphi \to \psi$, a false Σ_n sentence which contradicts T being Σ_n-sound. Thus the right-hand side is Σ_n-sound, hence Π_n-sound. $\square$

Taking $\lambda = \omega_2(\alpha)$ and $T = \mathsf{EA} + \pi$ for $\pi \in \Pi_{n+1}$, we note by 4.1.10 that

$$\begin{aligned} \mathsf{RFN}_{\Sigma_{n+1}}(T) &\equiv \mathsf{RFN}_{\Sigma_{n+1}}(\mathsf{EA}) + \pi \\ &\equiv \mathsf{I}\Sigma_n + \pi \\ &\equiv \mathsf{I}\Pi_n + \pi \\ &\equiv \mathsf{TI}[\omega, \Pi_n] + \pi \end{aligned} \tag{5.1}$$

Since $\mathsf{TI}[\omega_2(\alpha), \Pi_n] \vdash \mathsf{TI}[\omega, \Pi_n]$, we find that $\mathsf{TI}[\omega_2(\alpha), \Pi_n] + \pi \vdash \mathsf{RFN}_{\Sigma_{n+1}}(T)$, hence $\mathsf{TI}[\omega_2(\alpha), \Pi_n] + \pi \vdash \mathsf{RFN}_{\Sigma_n}(T)$ and thus, by 5.2.3,

$$\mathsf{TI}[\omega_2(\alpha), \Pi_n] + \pi \vdash \mathsf{RFN}^{\omega_2(\alpha)+1}_{\Pi_n}(T) \tag{5.2}$$

Recall the definition of $\mathcal{C}$-equivalence in 2.2.2 and that formally, $T \equiv_{\mathcal{C}} U$ is expressed as $\forall \varphi \in \mathcal{C}(\Box_T \varphi \leftrightarrow \Box_U \varphi)$.

The final major link is provided by a result known as *Schmerl's formula*. While originally established in [10] for $T = \mathsf{PRA}$, the generalisation below is due to [1].

Theorem 5.2.4 (Schmerl's formula)**.** *For all $n \geq 1$ and all Π_{n+1}-axiomatised elementarily presented theories T extending* EA*, the following is provable in* EA^+*:*

$$\forall \alpha \geq 1 (\mathsf{RFN}^{\alpha}_{\Pi_{n+1}}(T) \equiv_{\Pi_n} \mathsf{RFN}^{\omega_1(\alpha)}_{\Pi_n}(T)) \tag{5.3}$$

Corollary 5.2.5. *For all $n \geq 1, \alpha \geq 1$ and all Π_{n+1}-axiomatised elementarily presented theories T extending* EA*, we have provably in* EA^+ *that* $\mathsf{RFN}_{\Pi_n}(\mathsf{RFN}^{\alpha}_{\Pi_{n+2}}(T)) \equiv \mathsf{RFN}^{\omega_2(\alpha)+1}_{\Pi_n}(T)$.

Proof. As $\Pi_{n+1} \subseteq \Pi_{n+2}$, T is also Π_{n+2}-axiomatised. Hence, by 5.2.4, we have

$$\forall \alpha \geq 1(\mathsf{RFN}^{\alpha}_{\Pi_{n+2}}(T) \equiv_{\Pi_{n+1}} \mathsf{RFN}^{\omega_1(\alpha)}_{\Pi_{n+1}}(T)) \tag{5.4}$$

Making use of the fact that $\omega_1(\omega_1(\alpha)) = \omega_2(\alpha)$, 5.2.4 also gives us

$$\forall \alpha \geq 1(\mathsf{RFN}^{\omega_1(\alpha)}_{\Pi_{n+1}}(T) \equiv_{\Pi_n} \mathsf{RFN}^{\omega_2(\alpha)}_{\Pi_n}(T)) \tag{5.5}$$

Since Π_{n+1}-equivalence implies Π_n-equivalence, we deduce

$$\forall \alpha \geq 1(\mathsf{RFN}^{\alpha}_{\Pi_{n+2}}(T) \equiv_{\Pi_n} \mathsf{RFN}^{\omega_2(\alpha)}_{\Pi_n}(T)) \tag{5.6}$$

By 4.1.9, we then have $\mathsf{RFN}_{\Pi_n}(\mathsf{RFN}^{\alpha}_{\Pi_{n+2}}(T)) \equiv \mathsf{RFN}^{\omega_2(\alpha)+1}_{\Pi_n}(T)$ as required. □

Now we can conclude the proof of 5.2.2.

Proof. By Equation 5.2, $\mathsf{TI}[\omega_2(\alpha), \Pi_n] + \pi \vdash \mathsf{RFN}^{\omega_2(\alpha)+1}_{\Pi_n}(\mathsf{EA} + \pi)$ for all $\pi \in \Pi_{n+1}$.

Note that $\mathsf{EA} + \pi$ is Π_{n+1}-axiomatised by assumption. By 5.2.5, we have

$$\mathsf{TI}[\omega_2(\alpha), \Pi_n] + \pi \vdash \mathsf{RFN}_{\Pi_n}(\mathsf{RFN}^{\alpha}_{\Pi_{n+2}}(\mathsf{EA} + \pi)) \tag{5.7}$$

By the deduction theorem, this means

$$\mathsf{TI}[\omega_2(\alpha),\Pi_n] \vdash \pi \to \mathsf{RFN}_{\Pi_n}(\mathsf{RFN}^{\alpha}_{\Pi_{n+2}}(\mathsf{EA}+\pi)) \tag{5.8}$$

Since $\mathsf{RFN}^{\alpha}_{\Pi_{n+2}}(\mathsf{EA}+\pi)$ entails both $\mathsf{RFN}^{\alpha}_{\Pi_{n+2}}(\mathsf{EA})$ and π, we are done. □

In order to prove 5.2.1, we need to formalise the fact that π in the statement of 5.2.2 ought to be a formula with its Gödel number being quantified over. We define the following theories:

$$\begin{aligned} T_z &:= \mathsf{EA}+\pi(\overline{z}) \\ U^{\alpha}_z &:= \mathsf{RFN}^{\omega_2(\alpha)}_{\Pi_n}(T_z) \\ V^{\alpha}_z &:= \mathsf{RFN}^{\alpha}_{\Pi_{n+2}}(T_z) \end{aligned} \tag{5.9}$$

Formally (and more generally), for S a Π_{n+1}-axiomatised elementarily presented theory containing EA, $\pi \in \Pi_{n+1}$ and $S_z := S+\pi(\overline{z})$, we define the explicit numeration of $\mathsf{Ax}_{S_z}(x)$ as

$$\mathsf{Ax}_{S_z}(x) := Ax_S(x) \vee x = \ulcorner \pi(\dot{z}) \urcorner \tag{5.10}$$

Note that the predicate, as well as the corresponding provability predicate $\Box_{S_z}(x)$, has x and z as free variables. In particular, this allows us to construct these predicates for T_z which in turn leads to the corresponding ones for U^{α}_z and V^{α}_z with α, z, x as their free variables.

The general strategy will be to closely follow the proof of 5.2.2, except our theories now have a parameter z which we need to account for throughout. We start with the uniform analogue of 5.2.3:

Theorem 5.2.6. *For each ordinal λ, we have* $\mathsf{TI}[\lambda,\Pi_n] \vdash \forall z(\mathsf{RFN}_{\Sigma_n}(T_z) \to \mathsf{RFN}^{\lambda+1}_{\Pi_n}(T_z))$.

Proof. Similarly to 5.2.3, this is a formalisation of the meta-inductive proof for the claim "for each z such that T_z is Σ_n-sound, $\mathsf{RFN}^{\alpha}_{\Pi_n}(T_z)$ is Π_n-sound for all ordinals α"; neither the proof nor the formalisation requires any substantial changes. □

Lemma 5.2.7. *For all Π_{n+1} formulae $\pi(z)$, EA proves $\forall z((\mathsf{RFN}_{\Sigma_{n+1}}(\mathsf{EA}) \wedge \mathsf{Tr}_{\Pi_{n+1}}(\ulcorner \pi(\dot{z}) \urcorner)) \equiv \mathsf{RFN}_{\Sigma_{n+1}}(T_z))$.*

Proof. This follows from 4.1.10, with π replaced by the term $\ulcorner \pi(\dot{z}) \urcorner$. □

Corollary 5.2.8.

$$\mathsf{TI}[\omega_2(\alpha), \Pi_n] \vdash \forall z(\pi(z) \to \mathsf{RFN}_{\Pi_n}(U^{\alpha}_z)) \tag{5.11}$$

Proof. By definition, $\forall z(\mathsf{RFN}_{\Pi_n}(U^{\alpha}_z) \leftrightarrow \mathsf{RFN}^{\omega_2(\alpha)+1}_{\Pi_n}(T_z))$. By 5.2.6, we thus have $\mathsf{TI}[\omega_2(\alpha), \Pi_n] \vdash \forall z(\mathsf{RFN}_{\Sigma_n}(T_z) \to \mathsf{RFN}_{\Pi_n}(U^{\alpha}_z))$. Hence it remains to show that

$$\mathsf{TI}[\omega_2(\alpha), \Pi_n] \vdash \forall z(\pi(z) \to \mathsf{RFN}_{\Sigma_n}(T_z)) \tag{5.12}$$

Note that $\mathsf{TI}[\omega_2(\alpha), \Pi_n] \vdash \mathsf{TI}[\omega, \Pi_n] \equiv \mathsf{I}\Pi_n$. By 2.3.2 and 4.1.5, this entails $\mathsf{RFN}_{\Sigma_{n+1}}(\mathsf{EA})$. This means it suffices to show that

$$\mathsf{RFN}_{\Sigma_{n+1}}(\mathsf{EA}) \vdash \forall z(\pi(z) \to \mathsf{RFN}_{\Sigma_{n+1}}(T_z)) \tag{5.13}$$

This follows from 2.3.3 and 5.2.7. □

We generalise 5.2.4 as follows:

Theorem 5.2.9. *For $n \geq 1$, let S and S_z be as above. Then EA^+ proves the following:*

$$\forall z\ \forall \alpha \geq 1(\mathsf{RFN}^{\alpha}_{\Pi_{n+1}}(S_z) \equiv_{\Pi_n} \mathsf{RFN}^{\omega_1(\alpha)}_{\Pi_n}(S_z)) \tag{5.14}$$

We will not prove this theorem in detail. However, we remark that its statement differs from 5.2.4 only in formalisation: in 5.2.4 the theory T is arbitrary but fixed whereas in 5.2.9 we state the formalisation uniformly in z for each extension by a Π_{n+1} sentence of the form $\pi(\overline{z})$ for a fixed π. The formalisation consists of going through the proof of Schmerl's formula in [1] and verifying that all steps of the proof are formalisable in EA^+ uniformly in z. Although a full formalisation would be laborious, there are no substantial changes in the argument.

Corollary 5.2.10. *For all $n \geq 1$, we have provably in EA^+ that $\forall \alpha \geq 1\ \forall z(\mathsf{RFN}_{\Pi_n}(U^{\alpha}_z) \equiv \mathsf{RFN}_{\Pi_n}(V^{\alpha}_z))$.*

Proof. We proceed as in the proof of 5.2.5, using 5.2.9 in the case $S = \mathsf{EA}$ instead of 5.2.4. The steps that yielded Equations 5.4, 5.5 and 5.6 remain perfectly analogous, so we have

$$\forall \alpha \geq 1\ \forall z(U^{\alpha}_z \equiv_{\Pi_n} V^{\alpha}_z) \tag{5.15}$$

or, more formally,

$$\forall \alpha \geq 1\ \forall z \forall y(\Pi_n(y) \to (\Box_{U^{\alpha}_z} y \leftrightarrow \Box_{V^{\alpha}_z}(y))) \tag{5.16}$$

We give the final step of the argument in more detail, following the proof of 4.1.9.

For $\varphi(x) \in \Pi_n$ we reason in EA^+: suppose $\Box_{V^{\alpha}_z} \varphi(\dot{x})$. By applying Equation 5.16 to $y = \ulcorner \varphi(\dot{x}) \urcorner$, we have $\Box_{U^{\alpha}_z} \varphi(\dot{x})$. By $\mathsf{RFN}_{\Pi_n}(U^{\alpha}_z)$ we have $\forall x(\Box_{U^{\alpha}_z} \varphi(\dot{x}) \to \varphi(x))$ which then entails $\varphi(x)$. In summary, we have $\forall x(\Box_{V^{\alpha}_z} \varphi(\dot{x}) \to \varphi(x))$, an instance of $\mathsf{RFN}_{\Pi_n}(V^{\alpha}_z)$, derived from the corresponding instance of $\mathsf{RFN}_{\Pi_n}(U^{\alpha}_z)$. As the code of this derivation, in terms of $\ulcorner \varphi(x) \urcorner$, $\ulcorner \alpha \urcorner$ and z, can be bounded above by $supexp$, it follows that $\mathsf{EA}^+ \vdash \forall \alpha \geq 1\ \forall z(\mathsf{RFN}_{\Pi_n}(U^{\alpha}_z) \vdash \mathsf{RFN}_{\Pi_n}(V^{\alpha}_z))$. For the other implication, reverse the roles of U^{α}_z and V^{α}_z in the above argument.

□

We now conclude the proof of 5.2.1.

Proof. By 5.2.8, we have $\mathsf{TI}[\omega_2(\alpha), \Pi_n] \vdash \forall z(\pi(z) \rightarrow \mathsf{RFN}_{\Pi_n}(U_z^\alpha))$ for all $\pi \in \Pi_{n+1}$.

Note that $\mathsf{EA} + \pi(\overline{z})$ is Π_{n+1}-axiomatised by assumption. By 5.2.10, we have

$$\mathsf{TI}[\omega_2(\alpha), \Pi_n] \vdash \forall z(\pi(z) \rightarrow \mathsf{RFN}_{\Pi_n}(V_z^\alpha)) \tag{5.17}$$

The theory V_z^α EA-provably contains both $\mathsf{RFN}^\alpha_{\Pi_{n+2}}(\mathsf{EA})$ and $\pi(\overline{z})$. Hence, $\mathsf{RFN}_{\Pi_n}(V_z^\alpha)$ implies $\mathsf{Con}(V_z^\alpha)$ and $\mathsf{Con}(\mathsf{RFN}^\alpha_{\Pi_{n+2}}(\mathsf{EA}) + \pi(\overline{z}))$ which is provably equivalent to $\Diamond_{\mathsf{RFN}^\alpha_{\Pi_{n+2}}(\mathsf{EA})}\pi(\dot{z})$ as required. □

6. Summary

With 5.0.2 we have closed a gap in our knowledge by expressing partial transfinite induction principles in terms of transfinitely iterated partial reflection principles, and strengthening an implication established by Schmerl into an equivalence.

We hereby pose some further questions to be explored:

- Instead of transfinite induction principles, we could consider transfinite induction rules, given by $\mathsf{TI}^R[\alpha, \Pi_n] : \frac{\mathsf{Prog}(\varphi)}{\forall\beta{<}\alpha\ \varphi(\beta)}$ for $\varphi \in \Pi_n$. These are known to be weaker than the corresponding induction principles but it remains unknown whether a similar expression can be derived for those.
- Alternatively, one could look for analogous results in second-order arithmetic. A natural setting for this is ACA_0, a conservative extension of PA obtained by comprehension for arithmetical formulae. Some preliminary results and bounds are known; refer to [8] by Pakhomov and Walsh for an overview.

Bibliography

[1] Beklemishev, Lev. “Proof-theoretic analysis by iterated reflection”. In: *Archive for Mathematical Logic* 42 (Jan. 2003), pp. 515–552. DOI: 10.1007/s00153-002-0158-7.

[2] Beklemishev, Lev. “Reflection principles and provability algebras in formal arithmetic”. In: *Russian Mathematical Surveys* 60 (Apr. 2005). DOI: 10.1070/RM2005v060n02ABEH000823.

[3] Feferman, Solomon. “Transfinite recursive progressions of axiomatic theories”. In: *Journal of Symbolic Logic* 27.3 (1962), pp. 259–316. DOI: 10.2307/2964649.

[4] Gentzen, Gerhard. “Die Widerspruchsfreiheit der reinen Zahlentheorie”. In: *Mathematische Annalen* 112 (1936), pp. 493–565. DOI: 10.1007/BF01565428. URL: https://doi.org/10.1007/BF01565428.

[5] Kreisel, Georg and Lévy, Azriel. “Reflection principles and their use for establishing the complexity of axiomatic systems”. In: *Mathematical Logic Quarterly* 14.7-12 (1968), pp. 97–142. DOI: https://doi.org/10.1002/malq.19680140702. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/malq.19680140702. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/malq.19680140702.

[6] Leivant, Daniel. “The optimality of induction as an axiomatization of arithmetic”. In: *The Journal of Symbolic Logic* 48.1 (1983), pp. 182–184. ISSN: 00224812. URL: http://www.jstor.org/stable/2273332 (visited on 05/13/2024).

[7] Ono, Hiroakira. “Reflection principles in fragments of Peano arithmetic”. In: *Mathematical Logic Quarterly* 33.4 (1987), pp. 317–333. DOI: https://doi.org/10.1002/malq.19870330407. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/malq.19870330407. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/malq.19870330407.

[8] Pakhomov, Fedor and Walsh, James. “Reflection ranks and ordinal analysis”. In: *The Journal of Symbolic Logic* 86.4 (July 2020), pp. 1350–1384. ISSN: 1943-5886. DOI: 10.1017/jsl.2020.9. URL: http://dx.doi.org/10.1017/jsl.2020.9.

[9] Parsons, Charles. “On *n*-Quantifier Induction”. In: *The Journal of Symbolic Logic* 37.3 (1972), pp. 466–482. ISSN: 00224812. URL: http://www.jstor.org/stable/2272731 (visited on 08/10/2024).

[10] Schmerl, Ulf R. “A fine structure generated by reflection formulas over primitive recursive arithmetic”. In: Studies in Logic and the Foundations of Mathematics 97 (1979). Ed. by Boffa, Maurice,

Dalen, Dirkvan, and Mcaloon, Kenneth, pp. 335–350. ISSN: 0049-237X. DOI: https://doi.org/10.1016/S0049-237X(08)71633-1. URL: https://www.sciencedirect.com/science/article/pii/S0049237X08716331.

[11] Sommer, Richard. "Transfinite induction within Peano arithmetic". In: *Annals of Pure and Applied Logic* 76.3 (1995), pp. 231–289. ISSN: 0168-0072. DOI: https://doi.org/10.1016/0168-0072(95)00029-G. URL: https://www.sciencedirect.com/science/article/pii/016800729500029G.

[12] Turing, Alan M. "Systems of Logic Based on Ordinals". In: *Proceedings of the London Mathematical Society* s2-45.1 (1939), pp. 161–228. DOI: https://doi.org/10.1112/plms/s2-45.1.161. eprint: https://londmathsoc.onlinelibrary.wiley.com/doi/pdf/10.1112/plms/s2-45.1.161. URL: https://londmathsoc.onlinelibrary.wiley.com/doi/abs/10.1112/plms/s2-45.1.161.