Weak König's Lemma and Extensional Equality

Diplomarbeit

der philosophisch-naturwissenschaftlichen Fakultät

der Universität Bern

vorgelegt von

Theo Burri

2004

Leiter der Arbeit:

Prof. Dr. Gerhard Jäger

Institut für Informatik und angewandte Mathematik

Contents

Introduction

Motivation

In this thesis we investigate the proof-theoretic strength of primitive recursive arithmetic in all finite types PRA^ω, with either weak or full extensional interpretation of equality, and how it is affected by extending with Weak König's Lemma WKL and Uniform Weak König's Lemma UWKL respectively.

The principle WKL asserts that every infinite binary tree has an infinite path. It has turned out to be quite an important principle when it comes to formulating mathematically strong, but proof-theoretically weak, subsystems of analysis. Consider for example the second order theory WKL_0, which as a fragment of full second order arithmetic is strong enough to prove the Heine-Borel covering theorem or Gödel's completeness theorem of first order logic, but is proof-theoretically weak for it is Π^0_2-conservative over PRA. The latter result, which was first shown by H. Friedman using a model-theoretic argument, can also be obtained as a corollary by considering results that follow from the sequel, without appealing to such techniques.

We use as base theory PRA^ω and extend with certain additional rules and axioms. When working in this environment, one first has to settle on an interpretation of equality between objects of higher type. In this thesis we only focus on the extensional interpretation of equality: two functionals are equal if from the same input they give the same output. An intensional interpretation would be saying that two functionals are equal, if they are given by the same description. Even though extensionality is basically just a way of defining equality at higher types, its handling via rules and axioms can be done in many different ways. We restrict ourselves to two ways, by either adding to our base theory PRA^ω Spector's quantifier free rule of extensionality $\mathsf{QF\text{-}ER}$ or the full extensionality axiom (E). As a first result we get that

$$\mathsf{WE\text{-}PRA}^\omega \oplus \mathsf{QF\text{-}AC} \oplus \mathsf{WKL} \text{ is } \Pi^0_2\text{-conservative over } \mathsf{PRA},$$

where WE means that we have added $\mathsf{QF\text{-}ER}$ and $\mathsf{QF\text{-}AC}$ stands for *quantifier-free axiom of choice*. In this context, two questions arise: what happens if we replace $\mathsf{QF\text{-}ER}$ by (E) and what happens if we replace WKL by a uniform version UWKL, which states that there exists a functional Φ, which selects from a given infinite binary tree f a infinite path $\Phi(f)$.

The first question is of interest, because of the following reason. $\mathsf{QF\text{-}ER}$ was introduced by Spector to have a form of extensionality to which Gödels D-interpretation, in contrast to (E), directly applies. $\mathsf{QF\text{-}ER}$ is an intuitively weaker form of dealing with extensionality, one can therefore ask how much weaker it is, i.e. whether the results obtained in the weakly extensional context still hold in the fully extensional context or whether even stronger results can be obtained in the latter. As a result to that question, we obtain that

$$\mathsf{E\text{-}PRA}^\omega + \mathsf{QF\text{-}AC}^{1,0} + \mathsf{QF\text{-}AC}^{0,1} + \mathsf{WKL} \text{ is } \Pi^0_2\text{-conservative over } \mathsf{PRA}.$$

The second question arises in the context of so called explicit mathematics as developed by S. Feferman. UWKL has turned out to be a very natural formulation of WKL within that context. Even though UWKL seems to be stronger than WKL we get the following result:

$$\mathsf{WE\text{-}PRA}^\omega \oplus \mathsf{QF\text{-}AC} \oplus \mathsf{UWKL} \text{ is } \Pi^0_2\text{-conservative over } \mathsf{PRA}.$$

Thus, making these two replacements individually, we find that the resulting theories, although intuitively stronger, remain Π^0_2-conservative over PRA. The picture changes when adding both UWKL and (E). It is in combination where the resulting theory turns out to be actually stronger;

$$\mathsf{E\text{-}PRA}^\omega + \mathsf{QF\text{-}AC}^{1,0} + \mathsf{QF\text{-}AC}^{0,1} + \mathsf{UWKL} \text{ is conservative over } \mathsf{PA}.$$

This result is not so interesting though, since it actually only shows that the methods used to prove the former three theorems can somehow not be applied in this theory. More interesting is the fact that $\mathsf{E\text{-}PRA}^\omega + \mathsf{QF\text{-}AC}^{1,0} + \mathsf{QF\text{-}AC}^{0,1} + \mathsf{UWKL}$, in contrast to the other theories, not only contains PRA but PA.

All these results have already been proved and can be found in compressed form in the paper by Kohlenbach [13]. The goal of this thesis can therefore be seen in understanding this paper and the methods used to prove the theorems therein. A lot of help in doing so came from Avigad, Feferman [1] and Troelstra [15].

We conclude with a short overview of the content of the three chapters of this thesis.

1. Definition of the language and the axioms and rules of arithmetic in all finite types and introduction of certain restrictions and variants. Definition of negative translation and proofs of some of its properties.

2. Definition of the quantifier-free part of arithmetic in all finite types, introduction of Gödel's D-interpretation and a proof of the main property of the D-interpretation. Discussing extensions of the base theory.

3. Definition of WKL and UWKL, introduction of some additional tools (hereditary majorizability, elimination of extensionality) and proofs of the above stated results.

Acknowledgements

I am grateful to Prof. Dr. Gerhard Jäger, for introducing me to proof theory and for supervising my work on this thesis, and to PD Dr. Thomas Strahm, for his guidance and constant readiness in answering my questions. Even though the last year has been a labour-intensive and strenuous one for both of them, they have always had time to give me competent advice whenever I was lost in the depths of knowledge. I would also like to thank all the other members of the group, especially Mathis Kretz who has helped me understand Gödel's functional interpretation, and Dr. Geoffrey Ostrin for proof-reading and commenting not only on the mathematical contents. I also owe Manuel Niklaus a debt of gratitude for going through my work not with the eye of a mathematician but that of a linguistic expert. Last but not least I thank my parents, without their help and support, my education and therefore this thesis would not have been possible.

Bern, January 2004 Theo Burri

All the limitative Theorems of metamathematics and the theory of computation suggest that once the ability to represent your own structure has reached a certain critical point, that is the kiss of death: it guarantees that you can never represent yourself totally. Gödel's Incompleteness Theorem, Church's Undecidability Theorem, Turing's Halting Problem, Tarski's Truth Theorem— all have the flavour of some ancient fairy tale which warns you that 'To seek self-knowledge is to embark on a journey which . . . will always be incomplete, cannot be charted on a map, will never halt, cannot be described.'

Douglas R. Hofstadter, *"Gödel, Escher, Bach: An Eternal Golden Braid"*

Chapter 1

Arithmetic in all finite types

The aim of this chapter is to establish the logical and proof-theoretical framework, in which our main results will be carried out. In section 1.1 we define the extensional and weakly extensional versions of arithmetic in all finite types, $\mathsf{E\text{-}PA}^\omega$ and $\mathsf{WE\text{-}PA}^\omega$. These theories are basically extensions of first order Peano Arithmetic PA over many-sorted predicate logic. They only differ in their respective interpretation of equality between objects of higher type (order). Some properties regarding their handling of equality at higher types will be proved at the end of the section.

Being formulated within a much more expressive language, $\mathsf{E\text{-}PA}^\omega$ and $\mathsf{WE\text{-}PA}^\omega$ turn out to be far more powerful than PA. For our purposes they are too strong, therefore we introduce restricted variants, $\mathsf{E\text{-}PRA}^\omega$ and $\mathsf{WE\text{-}PRA}^\omega$, which are in some sense weaker than PA as will be discussed in section 1.2. This restriction mirrors the weakening of PA to PRA. The means we use to prove the conservation results in chapter 3 also require the intuitionistic variants of all these theories. They will be introduced and certain properties of them will be proved in section 1.3.

All the theories mentioned above are well documented in the literature. The majority from our foundations is taken from Troelstra [15], the notation and the formulation of certain special axioms and rules, such as the equality axioms or the quantifier-free extensionality rule, are taken from Feferman and Avigad [1] and Kohlenbach [13]. First order theories will not especially be discussed here. For an introduction see Jäger [8] and [9]. Everything about intuitionistic variants is based on Troelstra [15] and [16].

1.1 The theories E-PA$^\omega$ and WE-PA$^\omega$

1.1.1 The language $\mathcal{L}_0^\omega$

The main difference between first order and many-sorted predicate logic is that instead of function symbols, the latter contains variables of all sorts. So before we can state the language of arithmetic in all finite types $\mathcal{L}_0^\omega$, we must know what is meant by variables of different sort, and how they are to be understood intuitively. It is common in the literature to use the term *type* instead of *sort*.

Definition 1.1.1. The set of all finite types, $\mathbf{T}$, is defined inductively by the following two clauses:

1. 0 is a type.
2. If σ and τ are types then so is $(\sigma \to \tau)$.[1]

The intended interpretation is that objects of type 0 denote natural numbers and that objects of type $\sigma \to \tau$ denote functions, taking type σ objects to type τ objects. Objects of type $(\sigma \to \tau) \to \rho$ have functions as arguments and are usually called functionals. Brackets shall be associated to the right, so that

$$\tau_1 \to \tau_2 \to \ldots \to \tau_n$$

is an abbreviation for

$$\tau_1 \to (\tau_2 \to (\ldots \to \tau_n) \ldots).$$

Definition 1.1.2. For each type σ its type level, $\mathrm{lev}(\sigma)$, is the natural number defined by:

1. $\mathrm{lev}(0) := 0$.
2. $\mathrm{lev}(\sigma \to \tau) := \max\{\mathrm{lev}(\sigma) + 1, \mathrm{lev}(\tau)\}$.

By this convention each type is assigned a finite type level, hence the name *finite types*.

Definition 1.1.3. For each natural number n, we define its pure type, (n), as follows:

1. $(0) := 0$.
2. $(n + 1) := (n) \to 0$.

When there is no confusion, we will omit brackets around pure types.

[1] There are many alternative notations in use such as $(\sigma)\tau$, $(\tau)\sigma$, τ^σ or (τ, σ), the latter being the notation Gödel used in his original paper [5].

Remark 1.1.4. This definition of types allows us to find for each type σ a sequence of types $\tau_1, \ldots, \tau_n$, such that $\sigma = \tau_1 \to \ldots \to \tau_n \to 0$, and that $\mathrm{lev}(\sigma) = \max_{1 \leqslant i \leqslant n}(\mathrm{lev}(\tau_i)+1)$.

Definition 1.1.5. The language $\mathcal{L}_0^\omega$ of arithmetic in all finite types contains the following basic symbols:

1. For each type σ, there are countably many free variables $u^\sigma, v^\sigma, w^\sigma, \ldots$
2. For each type σ, there are countably many bound variables $x^\sigma, y^\sigma, z^\sigma, \ldots$
3. A constant 0 of type (0) and a constant Sc of type (1).
4. For each pair of types σ, τ, there is a constant $\mathrm{K}_{\sigma,\tau}$ of type $\sigma \to \tau \to \sigma$.
5. For each triple of types ρ, σ, τ, there is a constant $\mathrm{S}_{\rho,\sigma,\tau}$ of type $(\rho \to \sigma \to \tau) \to (\rho \to \sigma) \to (\rho \to \tau)$.
6. For each type σ, there is a constant R_σ of type $\sigma \to (0 \to \sigma \to \sigma) \to 0 \to \sigma$.
7. A symbol for equality at type 0, $=_0$.
8. Logical operators: $\wedge$ (and), $\vee$ (or), $\to$ (implies), $\forall$ (for all), $\exists$ (exists).
9. Brackets: (,).

Let $\mathcal{S}$ and $\mathbf{a}_1, \ldots, \mathbf{a}_n$ be finite sequences of aforementioned symbols or *words*, and let $u_1^{\sigma_1}, \ldots, u_n^{\sigma_n}$ be a sequence of pairwise different free variables, then

$$\mathcal{S}[\mathbf{a}_1, \ldots, \mathbf{a}_n / u_1^{\sigma_1}, \ldots, u_n^{\sigma_n}]$$

shall denote the word that results from $\mathcal{S}$ if we simultaneously replace each free variable $u_i^{\sigma_i}$ by $\mathbf{a}_i$. Sometimes we will just write $\mathcal{S}[\mathbf{a}_1, \ldots, \mathbf{a}_n]$. The notation $\mathcal{S}(\mathbf{t})$ refers to $\mathbf{t}$ which may occur at some place within $\mathcal{S}$.

Definition 1.1.6. $\mathcal{L}_0^\omega$-terms and their types are defined inductively by:

1. Each free variable and each constant of $\mathcal{L}_0^\omega$ is a $\mathcal{L}_0^\omega$-term of its own type.
2. If s is a $\mathcal{L}_0^\omega$-term of type σ and t a $\mathcal{L}_0^\omega$-term of type $\sigma \to \tau$ then $t(s)$[2] is a $\mathcal{L}_0^\omega$-term of type τ.

This definition should of course yield the following interpretation. 0 denotes the constant zero, $\mathrm{Sc}(t)$ denotes the successor of t and $t(s)$ denotes the result of applying the function(al) t to the argument s. The meaning of the application of $\mathrm{K}_{\sigma,\tau}$, $\mathrm{S}_{\rho,\sigma,\tau}$ and R_σ will become clear later. Instead of $\mathrm{Sc}(t)$ we will often just write t'.

[2] Often one finds the alternative notation ts. Even though it is shorter, I find it less intuitive.

Whenever the context allows, type subscripts of terms and variables will be suppressed. Multiple term application will often be abbreviated as follows: Instead of writing $t(s)(r)$ (which is to be read by associating to the left $(t(s))(r)$) we will just write $t(s,r)$. Similarly $t(r_1,\ldots,r_n)$ for any finite number of applications. Finite sequences of variables will be indicated in bold face, e.g. $\boldsymbol{x} = (x_1,\ldots,x_n)$. So $t(r_1,\ldots,r_n)$ can also just be written as $t(\boldsymbol{r})$. By remark 1.1.4 it should be clear, that for each term t, there is a sequence of variables $\boldsymbol{x}$, such that $t(\boldsymbol{x})$ has type 0.

Definition 1.1.7. $\mathcal{L}_0^\omega$-formulas are defined inductively by:

1. If t, s are $\mathcal{L}_0^\omega$-terms of type 0 then $t =_0 s$ is a $\mathcal{L}_0^\omega$-prime formula.
2. If A, B are $\mathcal{L}_0^\omega$-formulas then $(A \wedge B)$, $(A \vee B)$ and $(A \rightarrow B)$ are $\mathcal{L}_0^\omega$-formulas.
3. If A is a $\mathcal{L}_0^\omega$-formula, u^σ a free variable of $\mathcal{L}_0^\omega$ and x^σ a bound variable of $\mathcal{L}_0^\omega$ which does not occur in A then $\exists x A[x/u]$ and $\forall x A[x/u]$ are $\mathcal{L}_0^\omega$-formulas.

In the sequel, we just speak of terms and formulas instead of $\mathcal{L}_0^\omega$-terms and $\mathcal{L}_0^\omega$-formulas. Parantheses shall be treated as usual (brackets to the right, $\wedge$, $\vee$ before $\rightarrow$). Analogously to term application, we may abbreviate $\exists x_1 \exists x_2 \ldots \exists x_n A[x_1,\ldots,x_n]$ by $\exists x_1,\ldots,x_n A[x_1,\ldots,x_n]$ or even $\exists \boldsymbol{x} A[\boldsymbol{x}]$.

By $FV(A)$ ($FV(t)$) we denote the set of free variables that appear in a formula A (term t). If a formula (term) does not contain free variables it is called a closed formula (term).

Remark 1.1.8. We will use the following definitional abbreviations:

1. Equivalence: $A \leftrightarrow B :\equiv (A \rightarrow B) \wedge (B \rightarrow A)$.
2. Falsum: $\bot :\equiv (0 =_0 0')$.
3. Negation: $\neg A :\equiv (A \rightarrow \bot)$.
4. Numerals: 0; $1 :\equiv 0'$; $2 :\equiv 0''$; $\ldots$.
5. Higher type equality: $r^\sigma =_\sigma s^\sigma :\equiv \forall x_1^{\tau_1},\ldots,x_n^{\tau_n}(r(\boldsymbol{x}) =_0 s(\boldsymbol{x}))$ for $\sigma = \tau_1 \rightarrow \ldots \rightarrow \tau_n \rightarrow 0$.
6. Inequality: $r \neq_\sigma s :\equiv \neg(r =_\sigma s)$.

1.1.2 Deduction framework

By deduction framework we mean the set, or system, of axioms and rules. We formulate these in a Hilbert style calculus. The logical axioms and rules are those of first order classical predicate logic, but where the quantifier axioms and rules are extended to variables of all types. We will follow Gödel in the formulation of the propositional part for the following two reasons: Intuitionistic logic is easily obtained and Gödel's functional interpretation can be verified without much work.

The non-logical axioms are Peano's axioms, including induction, and the defining equations of K, S and R which allow us to generate Gödel's "primitive recursive functionals of finite type". We will prove that this class includes terms for all primitive recursive functions.

Definition 1.1.9. Logical axioms and rules:

Propositional axioms and rules:

P1) $A \vee A \to A, \quad A \to A \wedge A$

P2) $A \to A \vee B, \quad A \wedge B \to A$

P3) $A \vee B \to B \vee A, \quad A \wedge B \to B \wedge A$

P4) $\perp \to A$

P5) $A \vee \neg A$

P6) From $A \to B$ and A conclude B

P7) From $A \to B$ and $B \to C$ conclude $A \to C$

P8) From $A \to B$ conclude $C \vee A \to C \vee B$

P9) From $A \wedge B \to C$ conclude $A \to (B \to C)$

P10) From $A \to (B \to C)$ conclude $A \wedge B \to C$

Quantifier axioms and rules for arbitrary types σ:

Q1) $\forall x^\sigma A[x/u] \to A[t/u]$, with t of type σ

Q2) $A[t/u] \to \exists x^\sigma A[x/u]$, with t of type σ

Q3) From $A \to B(u)$ conclude $A \to \forall x^\sigma B[x/u]$ assuming u does not occur in A

Q4) From $A(u) \to B$ conclude $\exists x^\sigma A[x/u] \to B$ assuming u does not occur in B

For the two quantifier rules, the premise is not to depend on assumptions containing the free variable u.

Equality axioms:

Eq1) $u =_0 u$

Eq2) $u =_0 v \wedge A[u/w] \rightarrow A[v/w]$

It is clear, that from $u =_0 u$, $u =_\sigma u$ can be derived for all $\sigma \in \mathbf{T}$. We can also derive symmetry and transitivity for $=_0$ from Eq2. and then generalize it to all finite types.

Definition 1.1.10. Non-logical axioms and rules:

Peano axioms:

PA1) $u' \neq_0 0$

PA2) $u' =_0 v' \leftrightarrow u =_0 v$

Ind) From $A(0)$ and $A(u) \rightarrow A(u')$ conclude $A(u)$

For the induction rule, the same restriction applies as to the quantifier rules, i.e. u does not occur in assumptions on which $A(u) \rightarrow A(u')$ depends.

Defining axioms for K, S and R:

1. $\mathrm{K}(s,t) =_\sigma s$, for s of type σ and t of type τ
2. $\mathrm{S}(r,s,t) =_\tau r(t)(s(t))$ for r of type $\rho \rightarrow \sigma \rightarrow \tau$, s of type $\rho \rightarrow \sigma$ and t of type ρ
3. $\mathrm{R}(f,g,0) =_\sigma f$
   $\mathrm{R}(f,g,n') =_\sigma g(n, \mathrm{R}(f,g,n))$ for f of type σ and g of type $0 \rightarrow \sigma \rightarrow \sigma$.

From the defining axioms of the *typed combinators* K and S we can now introduce the notion of λ-abstraction.

Theorem 1.1.11 (Definition of the λ-operator). *To each term t and each free variable u we can associate another term $\lambda u.t$ such that:*

$$(\lambda u.t)(s) = t[s/u]$$

Proof. We define $\lambda u.t$ by induction on the complexity of t and show that this definition has the claimed property. For the base case we must consider the following two cases:

1. If $u \notin FV(t)$ then: $\lambda u.t :\equiv \mathrm{K}(t)$.
2. If t is u then: $\lambda u.t :\equiv \mathrm{S}(\mathrm{K}, \mathrm{K})$.

It is clear that the claim holds for these two cases. For the induction step, let t be $t_1(t_2)$. Then

$$\lambda u.t \text{ is } \mathrm{S}(\lambda u.t_1, \lambda u.t_2).$$

This definition yields $(\lambda u.t)(s) = \mathrm{S}(\lambda u.t_1, \lambda u.t_2, s) = \lambda u.t_1(s)(\lambda u.t_2(s)) = t_1[s/u](t_2[s/u]) = t_1(t_2)[s/u] = t[s/u]$ using the induction hypothesis. $\square$

Again we use an abbreviation for multiple λ-abstraction. $\lambda \boldsymbol{u}.t$ or $\lambda u_1 u_2 \ldots u_n.t$ shall be interpreted as $\lambda u_1.(\lambda u_2.(\ldots(\lambda u_n.t)\ldots))$.

1.1.3 (Weakly) extensional interpretation of equality

To complete our theory of full arithmetic in all finite types, we include the extensionality axiom.

$$(\mathsf{E}) :\equiv \forall x^\rho, y^\rho, z^{\rho\to\tau}(x =_\rho y \to z(x) =_\tau z(y))$$

The resulting theory will be called $\mathsf{E\text{-}PA}^\omega$.

The *weakly extensional* version, $\mathsf{WE\text{-}PA}^\omega$, is obtained by replacing the extensionality axiom by the quantifier-free rule of extensionality

$$\mathsf{QF\text{-}ER}: \text{ From } A \to s =_\rho t \text{ conclude } A \to r[s/u] =_\tau r[t/u],$$

where A is quantifier-free, s^ρ, t^ρ, $r[x^\rho]^\tau$ are arbitrary terms and ρ and τ are arbitrary types.

Having defined all the axioms and rules needed, we can now say what a derivation or a proof within a certain theory is.

Definition 1.1.12. A proof of a formula A within a theory Th is a sequence of formulas $A_1, \ldots, A_n$ such that $A_n = A$ and for each $i < n$, A_i is either an axiom of Th or is obtained by a rule of Th from formulas A_j with $j < i$. We usually write $\mathsf{Th} \vdash A$ or $\vdash_{\mathsf{Th}} A$. If A is derived from $\mathsf{Th} + B_1 + \cdots + B_n$ (which is to be read as Th extended by additional axioms $B_1, \ldots, B_n$) then we write $B_1 + \ldots + B_n \vdash_{\mathsf{Th}} A$ or $\mathsf{Th} + B_1 + \ldots + B_n \vdash A$. The name of the theory can be omitted if the situation allows so.

$\mathsf{WE\text{-}PA}^\omega$ is weaker in the sense that the extensionality axiom (E) can only be derived for type 0 arguments; for all other types it is underivable. This is a direct consequence of theorems 2.3.2 and 3.4.2. Due to this fact, the two theories also differ in other aspects.

Theorem 1.1.13. *In contrast to* $\mathsf{E\text{-}PA}^\omega$, *the deduction theorem*[3] *(Ded) in general does not hold in* $\mathsf{WE\text{-}PA}^\omega$.

Proof. That the deduction theroem holds, if only logical axioms and rules are considered is a standard textbook proof, and can be looked up for example in Troelstra [15] p. 12. Adding further axioms does not affect the deduction theorem. That it does hold in $\mathsf{E\text{-}PA}^\omega$ therefore only requires verification of the induction rule.

For this, we extend the proof in [15] by the following. Assume $B \to (A(u) \to A(u'))$ and $B \to A(0)$ have already been proved, and B does not contain u. $B \to (A(u) \to A(u'))$ can now be transformed to $(B \to A(u)) \to (B \to A(u'))$ using only propositional logic. An application of the induction rule now yields $B \to A(u)$.

For $\mathsf{WE\text{-}PA}^\omega$, then from

$$t =_\sigma s \vdash t =_\sigma s$$

$\mathsf{QF\text{-}ER}$ yields

$$t =_\sigma s \vdash \forall r(r(t) =_\tau r(s)).$$

If the deduction theorem holds, we would obtain

$$\vdash t =_\sigma s \to \forall r(r(t) =_\tau r(s)),$$

which is underivable if σ is not 0, as has been mentioned above. Thus the deduction theorem does not hold. $\square$

Remark 1.1.14. Assume $B \vdash A$ in $\mathsf{WE\text{-}PA}^\omega$, whereby for every application of $\mathsf{QF\text{-}ER}$ within the proof the premise does not depend on B, then the deduction theorem does indeed hold, i.e. $\vdash B \to A$ does hold in $\mathsf{WE\text{-}PA}^\omega$.

Whenever we want to make full use of the deduction theorem, adding new axioms to $\mathsf{WE\text{-}PA}^\omega$ will always be denoted by $\oplus$ instead of $+$, meaning, that the additional axioms may not be used in the proof of a premise of an application of $\mathsf{QF\text{-}ER}$.

This fact also affects the generalization of the equality axiom $u =_0 v \wedge A[u/w] \to A[v/w]$ with the help of (E) and $\mathsf{QF\text{-}ER}$.

Theorem 1.1.15. *Adding* (E) *or* $\mathsf{QF\text{-}ER}$ *allows us to extend Eq2 in the following way:*

1. $\mathsf{E\text{-}PA}^\omega \vdash u =_\sigma v \wedge A[u/w] \to A[v/w]$.
2. $\mathsf{WE\text{-}PA}^\omega \vdash u =_\sigma v \wedge A[u/w] \Rightarrow \mathsf{WE\text{-}PA}^\omega \vdash A[v/w]$ *where A is quantifier-free.*

[3] The deduction theorem is to be understood as usual: If $B_1, \ldots, B_n \vdash A$ then $B_1, \ldots, B_{n-1} \vdash B_n \to A$

Proof. For 1. this is easily proved by induction on the complexity of A. For $A \equiv s = t$ we can derive $s[u] = s[v]$ and $t[u] = t[v]$ using $u = v$ and (E). Using commutativity and transitivity of equality and $s[u] = t[u]$ we can now prove $s[v] = t[v]$. For the induction step we assume that

$$u = v \wedge A[u] \to A[v] \text{ and } u = v \wedge B[u] \to B[v]$$

have already been proved.

For "$\to$" assume $u = v \wedge (A[u] \to B[u]) \wedge A[v]$ and derive $A[u]$ using $u = v$ and the induction hypothesis. We can now derive $B[u]$ from the assumption. Applying the induction hypothesis again yields $B[v]$. An application of P9 results in

$$u = v \wedge (A[u] \to B[u]) \to (A[v] \to B[v]).$$

For "$\neg$" this works exactly the same, since it is defined by an implication.

For "$\wedge$" it is straightforward using $(C_1 \wedge C_2) \to C_1$ and $C_1 \to (C_2 \to (C_1 \wedge C_2))$.

For "$\vee$" we use the propositional laws $C_1 \wedge (C_2 \vee C_3) \leftrightarrow (C_1 \wedge C_2) \vee (C_1 \wedge C_3)$ and $(C_1 \to C_3) \to (C_2 \to C_3) \to (C_1 \vee C_2 \to C_3)$.

For "$\forall$" it is straightforward using the quantifier rules and axioms and the induction hypothesis.

For "$\exists$" we use the induction hypothesis and the quantifier rules to obtain $\exists x(u = v \wedge A[u, x]) \to \exists x A[v, x]$. Using $u = v \wedge \exists x A[u, x] \to \exists x(u = v \wedge A[u, x])$ we obtain what we want.

The proof of 2. uses theorem 2.2.5 to replace $A[u]$ by $t_A[u] = 0$. Using $u =_\sigma v$ and $\mathsf{QF\text{-}ER}$ we get $t_A[u] = t_A[v]$. By transitivity and again theorem 2.2.5 we obtain $A[v]$. $\square$

1.2 Restricted theories and first order theories

Definition 1.2.1 (PA and PRA). PA is basically the first order part of either $\mathsf{WE\text{-}PA}^\omega$ or $\mathsf{E\text{-}PA}^\omega$. PRA is obtained from PA by restricting induction to quantifier-free formulas. For an exact definition see Jäger [8] pp. 34-38. Note that we use PA for $\mathsf{PA}[PR]$ from [8].

From our proof-theoretic viewpoint the theories $\mathsf{E\text{-}PA}^\omega$ and $\mathsf{WE\text{-}PA}^\omega$ are far too strong for our purposes. We are more interested in fragments of these theories. For this reason we will weaken our theories analogously to the way PA is weakened to obtain PRA.

Definition 1.2.2 ($\mathsf{E\text{-}PRA}^\omega$ and $\mathsf{WE\text{-}PRA}^\omega$). The theories $\mathsf{E\text{-}PRA}^\omega$ and $\mathsf{WE\text{-}PRA}^\omega$ are obtained from $\mathsf{E\text{-}PA}^\omega$ and $\mathsf{WE\text{-}PA}^\omega$ respectively, by restricting induction to quantifier-free formulas and replacing the recursors R_σ by $\widehat{\mathrm{R}}_\sigma$, which are of type $\sigma \to (0 \to 0 \to \sigma) \to 0 \to \sigma$, and have the definitional axioms:

1. $\widehat{\mathrm{R}}_\sigma(f, g, 0, \boldsymbol{b}) =_0 f(\boldsymbol{b})$.
2. $\widehat{\mathrm{R}}_\sigma(f, g, n', \boldsymbol{b}) =_0 g(n, \widehat{\mathrm{R}}_\sigma(f, g, n, \boldsymbol{b}), \boldsymbol{b})$.

1.3 Intuitionistic variants

1.3.1 Intuitionistic logic

Intuitionistic logic was introduced to have a formal tool to investigate constructive mathematics. It differs from classical logic in that it only allows constructive proofs. For this thesis it is enough to know that constructive proofs differ in two major points from classical ones, namely in the proofs of existential statements and disjunctions. Proofs of these kinds must always provide us with a witness or respectively a method for deciding which disjunct is true. This of course implies that the law of excluded middle cannot in general be true within intuitionistic logic since we do not have a method which decides for every formula A whether it or its negation is true. Thus, to obtain intuitionistic logic, one must at least omit the law of the excluded middle. In fact, omitting just P5, $A \vee \neg A$, from our formulation of classical logic (P1-P10, Q1-Q4), is the only change necessary to obtain intuitionistic logic.

Clearly, the resulting logic is weaker than the classical one, but a lot of classically valid statements are still derivable, the deduction theorem to name something useful.

Lemma 1.3.1. *The following schemata and rules are derivable using only intuitionistic logic:*

1. $A \to B$ *implies* $\neg B \to \neg A$ *(contraposition).*
2. $A \to \neg\neg A$.
3. $\neg\neg\neg A \leftrightarrow \neg A$.
4. $\neg\neg(A \to B)$ *implies* $(A \to \neg\neg B)$.
5. $\neg\neg(A \wedge B)$ *implies* $(\neg\neg A \wedge \neg\neg B)$.
6. $\neg\neg\forall x A(x)$ *implies* $\forall x \neg\neg A(x)$.

Proof.

1. From $A \to B$ and $\neg B$ one can derive $\neg A$ using P7. Applying Ded yields the result.
2. When combining P1, P2 and P7 we can prove $\neg A \to \neg A$. P10 yields $\neg A \wedge A \to \bot$. By using commutativity and P9 we get the result.

3. "$\leftarrow$" is given by 2. "$\rightarrow$" can be obtained by applying 1. to 2.

4. Assume A, $\neg B$ and $\neg\neg(A \to B)$. From the first two we can derive $\neg(A \to B)$ by assuming $A \to B$ and by applying P6 twice and Ded once. We now use P6 again to get $\perp$. Eliminating A and $\neg B$ from the pool of assumptions by Ded gives us $A \to \neg\neg B$.

5. This proof is similar to 4. We would first show that $\neg(A \wedge B)$ follows from $\neg A$. Combining this result with $\neg\neg(A \wedge B)$ again gives us $\perp$ to which we apply Ded to obtain $\neg\neg A$. The same is now done with B.

6. From $\neg A(u)$ we derive $\neg\forall x A(x)$. As in 5. we get $\neg\neg A(u)$. Now we just have to introduce $\forall$, by way of Q3.

$\square$

1.3.2 Negative translation and application

Definition 1.3.2. For all formulas A of many sorted predicate logic, the *negative translation* $(A)^N$ is defined inductively on the formula complexity as follows:

1. $A^N :\equiv \neg\neg A$ for prime formulas A; $\perp^N :\equiv \perp$
2. $(A \wedge B)^N :\equiv A^N \wedge B^N$
3. $(A \to B)^N :\equiv A^N \to B^N$
4. $(\forall x A)^N :\equiv \forall x A^N$
5. $(A \vee B)^N :\equiv \neg(\neg A^N \wedge \neg B^N)$
6. $(\exists x A)^N :\equiv \neg\forall x \neg A^N$

Lemma 1.3.3. *For all formulas A, $\neg\neg A^N \to A^N$ is intuitionistically provable.*

Proof. By induction on the formula complexity. For prime formulas the result holds using 1.3.1.2. and 3. For the induction step consider the following.

Assume $\neg\neg C^N :\equiv \neg\neg(A^N \wedge B^N)$. By 1.3.1.5. this reduces to $\neg\neg A^N \wedge \neg\neg B^N$, the rest follows by the induction hypothesis.

For $\neg\neg C^N :\equiv \neg\neg\forall x A^N$ we use 1.3.1.6. and again the induction hypothesis.

The remaining cases are either trivial or similar. $\square$

We can now prove the main theorem for the negative translation.

Theorem 1.3.4. *For all formulas A, A is classically provable, if and only if A^N is intuitionistically provable.*

Proof. The proof from right to left is trivial. The proof from left to right is done by induction on the length of the proof. For the base case we solely have to check those axioms that contain either $\vee$ or $\exists$, since the negative translation of all the remaining axioms are again instances of themselves.

$(\text{P1})^N = \neg(\neg A^N \wedge \neg A^N) \to A^N$. We combine $\neg\neg A^N \to A^N$, which holds by 1.3.3, with the contraposition of $\neg A^N \to \neg A^N \wedge \neg A^N$.

$(\text{P2})^N = A^N \to \neg(\neg A^N \wedge \neg B^N)$. This is proved from the contraposition of an instance of P2, namely $\neg A^N \wedge \neg B^N \to \neg A^N$, using 1.3.1.2. and P7.

$(\text{P3})^N = \neg(\neg A^N \wedge \neg B^N) \to \neg(\neg B^N \wedge \neg A^N)$. This is just the contraposition of the conjunctive version of P3.

$(\text{P5})^N = \neg(\neg A^N \wedge \neg\neg A^N)$. From $\neg A^N \wedge \neg\neg A^N$, $\perp$ can easily be derived, we then just have to apply Ded.

$(\text{Q2})^N = A^N[t] \to \neg\forall x \neg A^N[x]$. Similar to the above, we take the contraposition of an instance of Q1 and use 1.3.1.2. and P7.

For the induction step for every rule, we assume that the negative translation of the premises of the rule are intuitionistically provable and prove that the same holds for the conclusion. For P6, P7, P9, P10 and Q3 this is trivial. So that leaves two remaining cases.

P8). $A^N \to B^N$ holds by induction hypothesis. We use its contraposition and instances of P2 to get $\neg C^N \wedge \neg B^N \to \neg C^N \wedge \neg A^N$. The contraposition of this is already the negative translation of the conclusion.

Q4). $A^N(u) \to B^N$ holds by induction hypothesis. Applying Q3 to the contraposition yields $\neg B^N \to \forall x \neg A^N[x]$. Again taking the contraposition and using 1.3.3 results in $(\exists x A[x] \to B)^N$. $\square$

Remark 1.3.5. Theorem 1.3.4 even holds if we add the equality axioms. Verification of these only requires lemma 1.3.1.2 and 1.3.7.

This result can be generalized to $\mathsf{WE\text{-}PA}^\omega$, $\mathsf{E\text{-}PA}^\omega$ and both their restricted variants. Whenever we want to refer to the intuitionistic variant of a theory Th, we write Th_i. For the case PA_i^ω it is also customary to use the term HA^ω.

Theorem 1.3.6. *For all formulas A, $\mathsf{WE\text{-}PA}^\omega \vdash A$ if and only if $\mathsf{WE\text{-}PA}_i^\omega \vdash A^N$. The same holds for $\mathsf{E\text{-}PA}^\omega$, $\mathsf{WE\text{-}PRA}^\omega$ and $\mathsf{E\text{-}PRA}^\omega$.*

Proof. We only need to verify that $\mathsf{WE\text{-}PA}_i^\omega$ proves the negative translation of the non-logical axioms of $\mathsf{WE\text{-}PA}^\omega$ and extend the proof of the induction step above by including the induction rule.

The defining axioms for K, S and R, $u' \neq_0 0$ and the induction rule are trivial. The proof of (E), $\mathsf{QF\text{-}ER}$ and $u' =_0 v' \leftrightarrow u =_0 v$ require the following theorem, for which a proof can be found in Troelstra and van Dalen [16]. $\square$

Theorem 1.3.7. *For all quantifier-free formulas A, $\neg\neg A \to A$ and $A \vee \neg A$ are provable in all our intuitionistic theories.*

Chapter 2

Gödel's dialectica interpretation

Having introduced all the logical and non-logical axioms and rules and having them grouped together to form certain theories, we now want to introduce the main tools that will be used to prove the conservation results in chapter 3.

In section 2.1 we first define the class of primitive recursive functions. For our purposes, this class of number-theoretic functions can be seen as the interface between PRA and the quantifier-free theories T or $\widehat{\mathsf{T}}$ respectively. This connection will be established and explained in the section 2.2, as will be the theories T and $\widehat{\mathsf{T}}$ themselves.

The most important section of this chapter, however, is section 2.3. In this section we introduce Gödel's so called *Dialectica* interpretation, which yields an interpretation of intuitionistic arithmetic in a quantifier-free theory of functionals of finite type. Deviating a bit from the original definition, we will define it for intuitionistic arithmetic in all finite types from the start. It was first published in [5] in 1958, even though Gödel had already started working on it in the 1930s. His original intention, when defining it, was to have a tool that would allow him to reduce a consistency proof of intuitionistic arithmetic HA ($= \mathsf{PA}_i$) to a consistency proof of a quantifier-free theory of functionals of finite type T. Our use of the D-interpretation is more direct. We use it to reduce proofs of one theory to proofs of T or $\widehat{\mathsf{T}}$ respectively.

At the end of this chapter, we will show that certain results concerning the D-interpretation still hold if we add variants of certain well known principles to the theories involved, such as Independence of Premise IP or Markov's Principle MP, which are already used in the justification of the D-interpretation.

2.1 Primitive recursive functions

Even though the primitive recursive functions are a well known subclass of the number-theoretic functions, their definition shall be repeated here. Additionally a relevant property of PRA will be proven.

Definition 2.1.1. The class of the primitive recursive functions is defined inductively by the following clauses:

1. $\mathrm{S} : \mathbb{N} \to \mathbb{N}$ with $\mathrm{S}(x) = x + 1$ is primitive recursive.
2. For all n, m; $\mathrm{Cs}^n_m : \mathbb{N}^n \to \mathbb{N}$ with $\mathrm{Cs}^n_m(\boldsymbol{x}) = m$ is primitive recursive.
3. For all n, $i < n$; $\mathrm{Pr}^n_i : \mathbb{N}^n \to \mathbb{N}$ with $\mathrm{Pr}^n_i(\boldsymbol{x}) = x_i$ is primitive recursive.
4. If f is a m-ary and $g_0, \ldots, g_{m-1}$ are n-ary primitive recursive functions, then so is $\mathrm{Comp}^n(f, g_0, \ldots, g_{m-1}) : \mathbb{N}^n \to \mathbb{N}$ with $\mathrm{Comp}^n(f, g_0, \ldots, g_{m-1})(\boldsymbol{x}) = f(g_0(\boldsymbol{x}), \ldots, g_{m-1}(\boldsymbol{x}))$.
5. If f is a n-ary and g a $(n+2)$-ary primitive recursive function, then so is $\mathrm{Rec}^{n+1}(f, g)$ with
$$\mathrm{Rec}^{n+1}(f, g)(0, \boldsymbol{y}) = f(\boldsymbol{y})$$
$$\mathrm{Rec}^{n+1}(f, g)(S(x), \boldsymbol{y}) = g(\mathrm{Rec}^{n+1}(f, g)(x, \boldsymbol{y}), x, \boldsymbol{y})$$

We recall, that several well known functions such as addition, multiplication, cut-off subtraction ($x \dot{-} y$), absolute difference ($|x - y|$), signum, minimum and maximum are primitive recursive. This fact allows us to prove the following theorem.

Theorem 2.1.2. *For all quantifier-free formulas A of PRA, there is a term t_A of PRA containing the same free variables, s.t* $\mathsf{PRA} \vdash A \leftrightarrow t_A = 0$

Proof. Since PRA contains function symbols for every primitive recursive function plus their defining axioms, the theorem can easily be proved by induction on the formula complexity:

1. $A \equiv t = s$. Let $t_A = |t - s|$.
2. $A \equiv \neg B$. Let $t_A = 1 \dot{-} t_B$.
3. $A \equiv B \wedge C$. Let $t_A = \max(t_B, t_C)$.
4. $A \equiv B \vee C$. Let $t_A = \min(t_B, t_C)$.
5. $A \equiv B \to C$. Let $t_A = \min(1 \dot{-} t_B, t_C)$.

$\square$

2.2 Quantifier-free part of weakly-extensional arithmetic in all finite types

2.2.1 Gödel's T

The quantifier-free theory that is used for the D-interpretation is basically the quantifier-free part of $\mathsf{WE\text{-}PA}^\omega$. Some changes of course are needed. Not containing quantifiers, T neither contains bound variables. Therefore, x and u will always denote free variables.

Definition 2.2.1. Gödel's T is defined as follows:

1. The formulas of T are all the quantifier-free formulas of $\mathsf{WE\text{-}PA}^\omega$.
2. For equality at higher types we use the abbreviation $r^\sigma =_\sigma s^\sigma$ for $r(\boldsymbol{x}) =_0 s(\boldsymbol{x})$.
3. The rules and axioms of T are those of $\mathsf{WE\text{-}PA}^\omega$, whereby the quantifier rules and axioms are replaced by a rule allowing for the substitution of arbitrary terms for variables of the same type.

Since all of our conservation results are carried out in the restricted theory $\mathsf{WE\text{-}PRA}^\omega$, we need to define the theory that mirrors its quantifier-free part as well. This theory will be named $\widehat{\mathsf{T}}$, differing from T only in the definition of the recursors $\widehat{\mathrm{R}}$ and R respectively.

2.2.2 Translating terms of PRA to terms of T

Theorem 2.2.2. *Let t and s be terms of* PRA. *There exists a natural translation to terms t^T and s^T of* T, *such that if* $\mathsf{PRA} \vdash t = s$ *then* $\mathsf{T} \vdash t^\mathsf{T} = s^\mathsf{T}$.

Proof. The translation is done by induction on the complexity of t. The function symbols $\cdot, +$ of PRA can be neglected, since they can be replaced by their corresponding primitive recursive function symbols. The conservation result follows directly from the way the translation is defined.

1. $t \equiv 0$. Let $t^\mathsf{T} :\equiv 0$.
2. $t \equiv u$. Let $t^\mathsf{T} :\equiv u$.
3. $t \equiv \mathrm{S}(s)$. Let $t^\mathsf{T} :\equiv \mathrm{Sc}(s^\mathsf{T})$.
4. $t \equiv \mathrm{Cs}^n_m(s_1, \ldots, s_n)$. Let $t^\mathsf{T} :\equiv (\lambda \boldsymbol{u}.m)(s_1^\mathsf{T}, \ldots, s_n^\mathsf{T})$.
5. $t \equiv \mathrm{Pr}^n_i(s_1, \ldots, s_n)$. Let $t^\mathsf{T} :\equiv (\lambda \boldsymbol{u}.u_i)(s_1^\mathsf{T}, \ldots, s_n^\mathsf{T})$, where $\boldsymbol{u}$ must not occur in s_i^T.

6. $t \equiv \mathrm{Comp}^n(r, l_1, \ldots, l_m)(s_1, \ldots, s_n)$. Let $t^{\mathsf{T}} :\equiv (\lambda \boldsymbol{u}.r^{\mathsf{T}}(l_1^{\mathsf{T}}(\boldsymbol{u}), \ldots, l_m^{\mathsf{T}}(\boldsymbol{u})))(s_1^{\mathsf{T}}, \ldots, s_n^{\mathsf{T}})$. $\boldsymbol{u}$ must not occur in any s_i^{T}, l_i^{T} or r^{T}.

7. $t \equiv \mathrm{Rec}^{n+1}(r, l)(s_1, \ldots, s_{n+1})$.
   Let $t^{\mathsf{T}} :\equiv (\lambda \boldsymbol{u}.\mathrm{R}(r^{\mathsf{T}}(u_2, \ldots, u_{n+1}), \lambda v_1 v_2.l^{\mathsf{T}}(v_2, v_1, u_2, \ldots, u_{n+1}), u_1))(s_1^{\mathsf{T}}, \ldots, s_{n+1}^{\mathsf{T}})$
   $\boldsymbol{u}$, v_1 or v_2 must not occur in any s_i^{T}, l^{T} or r^{T}.

□

Corollary 2.2.3. *The same holds if we replace* T *by* $\widehat{\mathsf{T}}$. *We only have to replace the last line by:* $t^{\widehat{\mathsf{T}}} = (\lambda \boldsymbol{u}.\widehat{\mathrm{R}}(r^{\widehat{\mathsf{T}}}, \lambda \boldsymbol{v}.l^{\widehat{\mathsf{T}}}(v_2, v_1, v_3, \ldots, v_{n+2}), u_1, \ldots, u_{n+1}))(s_1^{\widehat{\mathsf{T}}}, \ldots, s_{n+1}^{\widehat{\mathsf{T}}})$

Corollary 2.2.4. T *and* $\widehat{\mathsf{T}}$ *contain terms for every primitive recursive function.*

Theorem 2.2.5. *1. For every formula* A *of* T *there is a term* t_A *of* T *containing the same free variables, such that*

$$\mathsf{T} \vdash A \leftrightarrow t_A = 0.$$

2. There is a functional Cond*, such that*

$$\mathsf{T} \vdash \mathrm{Cond}(w, u, v) = \begin{cases} u & w = 0 \\ v & \textit{otherwise} \end{cases}$$

Both statements also hold for $\widehat{\mathsf{T}}$.

Proof. 1. is an easy consequence of theorem 2.1.2 and corollary 2.2.4.

For 2. let $\mathrm{Cond} = \lambda wuv.\mathrm{R}(u, \lambda v_1 v_2.v, w)$. For $\widehat{\mathsf{T}}$ we need a slight change. $\mathrm{Cond} = \lambda wuv\boldsymbol{b}.\widehat{\mathrm{R}}(u, \lambda v_1 v_2.v, w, \boldsymbol{b})$. We must not forget that equality at higher types is actually a definitional abbreviation. □

2.2.3 Type-1 conservation of $\widehat{\mathsf{T}}$ over PRA

The main theorem of this section is in a way the converse of theorem 2.2.2. Obviously, we cannot fully convert this theorem, since there is no way of interpreting variables of higher type in PRA. Nevertheless, we can prove the following theorem.

Theorem 2.2.6. *1. The closed type 1 terms of* $\widehat{\mathsf{T}}$ *denote primitive recursive functions.*

2. There exists a natural translation of terms t *of* $\widehat{\mathsf{T}}$, *which are of type 0 and only contain free variables of type 0 to terms* t^{PRA} *of* PRA, *such that if* $\widehat{\mathsf{T}} \vdash t = s$ *then* $\mathsf{PRA} \vdash t^{\mathsf{PRA}} = s^{\mathsf{PRA}}$.

The proof of this theorem requires some preparatory work. We need to define a new class of functionals KL, whose elements F are applied to lists of arguments $\boldsymbol{b}$ of arbitrary type, such that $F(\boldsymbol{b})$ is of type 0 only. Furthermore, each F in KL shall be defined by a term t of $\widehat{\mathsf{T}}$ and each term t in $\widehat{\mathsf{T}}$ shall be represented by an F in KL either directly (if it is of type 0) or by abstraction on some of the variables of F (otherwise).

In the definition of KL, and in the proof of these facts, we will use the following additional notations. For (possibly empty) sequences of types $(\sigma_1, \ldots, \sigma_n)$ we write $\underline{\sigma}$ and by $(\underline{\sigma} \to 0)$ we mean 0 if $n = 0$ and $(\sigma_1 \to \ldots \to \sigma_n \to 0)$ otherwise. For sequences of variables $\boldsymbol{u}$ we will also use $\boldsymbol{u}^{\underline{\sigma}}$, the meaning of which should be clear.

Definition 2.2.7. The functionals F in KL are generated by the following schemata, in which n is a variable of type 0 and $\boldsymbol{b} = (b_1^{\tau_1}, \ldots, b_n^{\tau_n})$.

1. $F(\boldsymbol{b}) = 0$
2. $F(n, \boldsymbol{b}) = n$
3. $F(n, \boldsymbol{b}) = n'$
4. a) $F(a^\sigma, \boldsymbol{b}, \boldsymbol{c}) = a(\boldsymbol{b})$, where $\sigma = (\underline{\tau} \to 0)$, $\sigma \neq 0$
4. b) $F(a^\sigma, \boldsymbol{b}) = a(t_1, \ldots, t_k)$ where $\sigma = (\underline{\alpha} \to 0)$, $\sigma \neq 0$, $\alpha_i = (\underline{\rho}_i \to 0)$ (possibly 0) and $t_i = \lambda \boldsymbol{w}_i^{\underline{\rho}_i}.G_i(\boldsymbol{w}_i, a, \boldsymbol{b})$ for $i = 1, \ldots, k$
5. $F(\boldsymbol{b}) = G(H(\boldsymbol{b}), \boldsymbol{b})$ where the first argument of G is of type 0
6. $F(\boldsymbol{b}) = G(\boldsymbol{b}_\pi)$ where $\boldsymbol{b}_\pi$ is a permutation of $\boldsymbol{b}$ by π
7. $F(0, \boldsymbol{b}) = G(\boldsymbol{b})$, $F(n', \boldsymbol{b}) = H(n, \boldsymbol{b}, F(n, \boldsymbol{b}))$

These functionals have already been introduced by Kleene in [10] while studying higher type recursion. Unlike Kleene, who used the term *primitive recursive* for all functionals in KL, we restrict this term to the functions (functionals) as defined in 2.1.

Lemma 2.2.8. *For these newly defined functionals, we establish the following facts:*

1. *The functionals of KL are closed under substitution: If F and G are functionals of KL and a^τ a variable of F with $\tau = (\underline{\sigma} \to 0)$, $\tau \neq 0$, then the functional that results if we substitute $\lambda \boldsymbol{u}^{\underline{\sigma}}.H(\boldsymbol{u}, \ldots)$ for a is also in KL.*
2. *For each F in KL we can find a term t of $\widehat{\mathsf{T}}$, which defines F. Meaning, that whenever $F(\boldsymbol{b}) = a$, then also $t(\boldsymbol{b}) = a$.*
3. *If t is a term of $\widehat{\mathsf{T}}$ with free variables $\boldsymbol{b}$, and t is of type $(\underline{\sigma} \to 0)$ (possibly 0), then we can find a functional $F(\boldsymbol{u}, \boldsymbol{b})$ of KL such that $t(\boldsymbol{u}) = F(\boldsymbol{u}, \boldsymbol{b})$ for all $\boldsymbol{u}, \boldsymbol{b}$.*
4. *If $F(\boldsymbol{b})$ has all its variables $b_1, \ldots, b_n$ of type 0, then F is primitive recursive.*

Proof.

1. This corresponds to the Full Substitution Theorem of Kleene [10] section 1.6 and will not be proved here. It is proved by induction on the level of τ and therein by induction on the generation of F in KL.

2. This is done by induction on the generation of F, where the first 4 schemata are to be considered for the base case and the rest for the induction step.

   1. $t = \lambda \boldsymbol{b}.0$
   2. $t = \lambda n\boldsymbol{b}.n$
   3. $t = \lambda n\boldsymbol{b}.Sc(n)$

   4.a. $t = \lambda a\boldsymbol{b}\boldsymbol{c}.a(\boldsymbol{b})$

   4.b. $t = \lambda a\boldsymbol{b}.a(r_1, \dots, r_k)$ where $r_i = \lambda \boldsymbol{w}_i^{\rho}.t_i(\boldsymbol{w}_i, a, \boldsymbol{b})$ and t_i is the term which defines G_i by induction hypothesis.

   5. $t = \lambda \boldsymbol{b}.r(s(\boldsymbol{b}), \boldsymbol{b})$ where r and s define G and H.
   6. $t = \lambda \boldsymbol{b}.r(b_{\pi(1)}, \dots, b_{\pi(n)})$ where r defines G.
   7. $t = \lambda k\boldsymbol{b}.\widehat{\mathrm{R}}(r, s, k, \boldsymbol{b})$ where r and s define G and H.

3. By induction on the complexity of t. For the base case we consider the following 7 cases.

   1. $t = 0$. Then $F \overset{1.}{=} 0$
   2. $t = u$. Then $F(u) \overset{2.}{=} u$
   3. $t = \mathrm{Sc}$. Then $F(u) \overset{3.}{=} \mathrm{Sc}(u)$
   4. $t = b$. Then $F(\boldsymbol{u}, b) \overset{6.}{=} F(b, \boldsymbol{u}) \overset{4.a.}{=} b(\boldsymbol{u}) = t(\boldsymbol{u})$
   5. $t = \mathrm{K}$. Then $F(\boldsymbol{u}) \overset{6.}{=} F(u_1, u_3, \dots, u_k, u_2) \overset{4.a.}{=} u_1(u_3, \dots, u_k) = t(\boldsymbol{u})$
   6. $t = \mathrm{S}$. Then $F(\boldsymbol{u}) \overset{4.b.}{=} u_1(t_3, t_2, t_4, \dots, t_k) = u_1(u_3, u_2(u_3), u_4, \dots, u_k) = t(\boldsymbol{u})$ Where $t_i = \lambda \boldsymbol{w}.G_i(\boldsymbol{w}, \boldsymbol{u}) \overset{6.\&4.a.}{=} \lambda \boldsymbol{w}.u_i(\boldsymbol{w}) = u_i$ for $i > 2$. The case $i = 2$ is done analogously.
   7. $t = \mathrm{R}$. Then $F(\boldsymbol{u}) \overset{6.}{=} F(u_3, u_1, u_2, u_4, \dots, u_k)$. With
   $F(0, u_1, u_2, u_4, \dots, u_k) \overset{7.}{=} G(u_1, u_2, u_4, \dots, u_k) =$
   $\overset{6.\&4.a.}{=} u_1(u_4, \dots, u_k) = t(u_1, u_2, 0, u_4, \dots, u_k)$
   $F(u_3', u_1, u_2, u_4, \dots, u_k) \overset{7.}{=} H(u_3, u_2, u_2, u_4, \dots, u_k, F(u_3, u_2, u_2, u_4, \dots, u_k)) =$
   $\overset{6.\&4.a.}{=} u_2(u_3, F(u_3, u_2, u_2, u_4, \dots, u_k), u_4, \dots, u_k) = t(u_1, u_2, u_3', u_4, \dots, u_k)$

For the induction step assume $t_1(\boldsymbol{u}) = F_1(\boldsymbol{u}, \boldsymbol{b}_1)$ and $t_2(\boldsymbol{v}) = F_2(\boldsymbol{v}, \boldsymbol{b}_2)$. For $t = t_1(t_2)$ we take

$$F(\boldsymbol{u}, \boldsymbol{b}_1, \boldsymbol{b}_2) = F_1(\lambda \boldsymbol{v}.F_2(\boldsymbol{v}, \boldsymbol{b}_2), u_2, \ldots, u_k, \boldsymbol{b}_1)$$

which is a functional of KL by 1.

4. If all variables of $F(\boldsymbol{b})$ are of type 0, we readily see, that the schemata 4.a. and 4.b. cannot be used in the generation of F. If we now look at the remaining schemata we see that they define primitive recursive functions.

$\square$

We can now prove theorem 2.2.6

Proof. 1. This follows directly from 2.2.8, parts 3. and 4.

2. The translation of t is done in two steps. We first build a term t^{KL} and from there a term t^{PRA} using 3. and 4. of the above lemma. For the proofs, we do practically the same. We use 2.2.8.3. to transform proofs of $\widehat{\mathsf{T}}$ to proofs from schemata 1-7. With 2.2.8.4. we can transform these proofs to proofs of PRA. $\square$

2.3 The D-Interpretation

The definition, justification and proof of the main theorem of the D-interpretation follow Avigad and Feferman [1] and Troelstra [15]. From now on, we will somewhat depart from our original use of notation. The difference of free and bound variables will no longer be emphasized by using different letters. Additionally, we will also use capital letters to denote variables, usually when we want to denote functional variables with explicit arguments, as in $X(y, z)$.

2.3.1 Definition and justification

The *dialectica* (or D-) *interpretation* assigns to each formula A of $\mathsf{WE}\text{-}\mathsf{PA}^\omega_i$ a formula A^D of the form $\exists \boldsymbol{x} \forall \boldsymbol{y} A_D$, where A_D is a quantifier-free formula of T, containing as free variables those free in A together with the sequences $\boldsymbol{x}$ and $\boldsymbol{y}$.

Definition 2.3.1. The associations $(*)^D$ and $(*)_D$ are defined inductively as follows.

For A a prime formula, $\boldsymbol{x}$ and $\boldsymbol{y}$ are both empty and $A^D = A_D = A$.

For the induction clause assume

$$A^D = \exists \boldsymbol{x} \forall \boldsymbol{y} A_D \text{ and } B^D = \exists \boldsymbol{u} \forall \boldsymbol{v} B_D.$$

1. $(A \wedge B)^D = \exists \boldsymbol{x}, \boldsymbol{u} \forall \boldsymbol{y}, \boldsymbol{v}(A_D \wedge B_D)$.
2. $(A \vee B)^D = \exists z, \boldsymbol{x}, \boldsymbol{u} \forall \boldsymbol{y}, \boldsymbol{v}((z = 0 \wedge A_D) \vee (z \neq 0 \wedge B_D))$.
3. $(\forall z A(z))^D = \exists \boldsymbol{X} \forall z, \boldsymbol{y} A_D(\boldsymbol{X}(z), \boldsymbol{y}, z)$.
4. $(\exists z A(z))^D = \exists z, \boldsymbol{x} \forall \boldsymbol{y} A_D(\boldsymbol{x}, \boldsymbol{y}, z)$.
5. $(A \to B)^D = \exists \boldsymbol{U}, \boldsymbol{Y} \forall \boldsymbol{x}, \boldsymbol{v}(A_D(\boldsymbol{x}, \boldsymbol{Y}(\boldsymbol{x}, \boldsymbol{v})) \to B_D(\boldsymbol{U}(\boldsymbol{x}), \boldsymbol{v}))$.

For negation we obtain the following using clause 5.

$$\begin{aligned}(\neg A)^D = (A \to \bot)^D = (\exists \boldsymbol{x} \forall \boldsymbol{y} A_D(\boldsymbol{x}, \boldsymbol{y}) \to \bot)^D = \\ = \exists \boldsymbol{Y} \forall \boldsymbol{x}(A_D(\boldsymbol{x}, \boldsymbol{Y}(\boldsymbol{x})) \to \bot) = \exists \boldsymbol{Y} \forall \boldsymbol{x} \neg A_D(\boldsymbol{x}, \boldsymbol{Y}(\boldsymbol{x})).\end{aligned}$$

The definition of A^D for A prime and the definition of $(A \wedge B)^D$ and of $(\exists z A(z))^D$ needs no comment. They are all justified from a constructive as well as classical point of view. $(A \vee B)^D$ becomes clear, if you consider Gödel's original intention. The definition of $(\forall z A(z))^D$ is justified by applications of the axiom of choice AC,

$$\forall x \exists y A(x, y) \to \exists Y \forall x A(x, Y(x)),$$

which is accepted by many constructivists. For the definition of $(A \to B)^D$ one uses AC and the following 4 equivalences:

1. $(\forall x A(x) \to B) \leftrightarrow \exists x(A(x) \to B)$, x not appearing free in B.
2. $(A \to \exists x B(x)) \leftrightarrow \exists x(A \to B(x))$, x not appearing free in A.
3. $(\exists x A(x) \to B) \leftrightarrow \forall x(A(x) \to B)$, x not appearing free in B.
4. $(A \to \forall x B(x)) \leftrightarrow \forall x(A \to B(x))$, x not appearing free in A.

Equivalences 3 and 4 are intuitionistically valid, while 1 and 2 are only classically valid. The direction from left to right of equivalence 2 is usually called *independence of premise* IP. The reason why it is intuitionistically problematic, is that a constructive reading of the hypothesis says that the choice of a witness for x in $\exists x B$ depends on the proof of A, while the conclusion tells us that this witness can be chosen independently from a proof of A. Equivalence 1 can be justified by a generalization of Markov's principle $\neg\neg\exists x B \to \exists x B$

MP_+ $\quad \neg \forall x A \to \exists x \neg A$

if the law of excluded middle holds for the conclusion, but as we will see, this will always be the case. If we consider B to be true, 1 is justified and anything can be taken to witness the existential-quantifier. If B is false $(\forall x A \to B)$ can be transformed to $(\neg\forall x A)$ to which MP_+ is applied With P4 the final result can be obtained. The other direction can be proved intuitionistically without MP_+. MP_+ is intuitionistically problematic, because there is no evident way to choose constructively a witness x to $\neg A$ from a proof that $\forall x A$ leads to a contradiction.

Clearly, to obtain the definition of $(A \to B)^D$, these 4 equivalences must be applied in a specific order, namely 3,2,4,1. The reason why this particular order has been chosen is explained in [15].

The three principles AC, IP and MP_+ will be discussed more thoroughly in section 2.3.3.

2.3.2 Verifying the axioms of arithmetic

The main result concerning the D-interpretation is the following theorem.

Theorem 2.3.2. *If* $\mathsf{WE\text{-}PA}_i^\omega \vdash A(\boldsymbol{z})$ *(A containing at most $\boldsymbol{z}$ free) and* $(A(\boldsymbol{z}))^D = \exists \boldsymbol{x} \forall \boldsymbol{y} A_D(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ *then, there is a sequence of closed terms $\boldsymbol{t}$, such that* $\mathsf{T} \vdash A_D(\boldsymbol{t}, \boldsymbol{y}, \boldsymbol{z})$.

This result can be rephrased by saying, that $\mathsf{WE\text{-}PA}_i^\omega$ is D-interpreted in T.

Proof. This proof is carried out by induction on the length of the proof of A in $\mathsf{WE\text{-}PA}_i^\omega$. One only has to verify, that the claim holds true when A is an axiom of $\mathsf{WE\text{-}PA}_i^\omega$ and that it is maintained under the rules of inference. For most cases additional free variables $\boldsymbol{z}$ need not be considered.

Propositional axioms and rules:

P1).

$$\boxed{(A \vee A \to A)^D =}$$
$$\big(\exists z, \boldsymbol{x}_1, \boldsymbol{x}_2 \forall \boldsymbol{y}_1, \boldsymbol{y}_2 [(z = 0 \wedge A_D(\boldsymbol{x}_1, \boldsymbol{y}_1)) \vee (z \neq 0 \wedge A_D(\boldsymbol{x}_2, \boldsymbol{y}_2))] \to \exists \boldsymbol{x} \forall \boldsymbol{y} A_D(\boldsymbol{x}, \boldsymbol{y})\big)^D =$$
$$\exists \boldsymbol{X}, \boldsymbol{Y}_1, \boldsymbol{Y}_2 \forall z, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y} \big([z = 0 \wedge A_D(\boldsymbol{x}_1, \boldsymbol{Y}_1(z, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}))] \vee$$
$$\vee [z \neq 0 \wedge A_D(\boldsymbol{x}_2, \boldsymbol{Y}_2(z, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}))] \to A_D(\boldsymbol{X}(z, \boldsymbol{x}_1, \boldsymbol{x}_2), \boldsymbol{y})\big)$$

For $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ we take sequences of terms $\boldsymbol{t}_1 = \boldsymbol{t}_2 = \lambda z \boldsymbol{x}_1 \boldsymbol{x}_2 \boldsymbol{y}.\boldsymbol{y}$. And for $\boldsymbol{X}$ take $\boldsymbol{t} = \mathbf{Cond}$ from theorem 2.2.5. Clearly the following holds

$$\mathsf{T} \vdash [(z = 0 \wedge A_D(\boldsymbol{x}_1, \boldsymbol{t}_1(z, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}))) \vee (z \neq 0 \wedge A_D(\boldsymbol{x}_2, \boldsymbol{t}_2(z, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})))] \to A_D(\boldsymbol{t}(z, \boldsymbol{x}_1, \boldsymbol{x}_2), \boldsymbol{y})$$

for if we evaluate and replace respective terms, this reads as

$$\mathsf{T} \vdash [(z = 0 \wedge A_D(\boldsymbol{x}_1, \boldsymbol{y})) \vee (z \neq 0 \wedge A_D(\boldsymbol{x}_2, \boldsymbol{y}))] \to A_D(\mathbf{Cond}(z, \boldsymbol{x}_1, \boldsymbol{x}_2), \boldsymbol{y}).$$

$\boxed{(A \to A \wedge A)^D =}$

$$\big(\exists \boldsymbol{x} \forall \boldsymbol{y} A_D(\boldsymbol{x}, \boldsymbol{y}) \to \exists \boldsymbol{x}_1, \boldsymbol{x}_2 \forall \boldsymbol{y}_1, \boldsymbol{y}_2 (A_D(\boldsymbol{x}_1, \boldsymbol{y}_1) \wedge A_D(\boldsymbol{x}_2, \boldsymbol{y}_2))\big)^D =$$
$$\exists \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y} \forall \boldsymbol{x}, \boldsymbol{y}_1, \boldsymbol{y}_2 \big(A_D(\boldsymbol{x}, \boldsymbol{Y}(\boldsymbol{x}, \boldsymbol{y}_1, \boldsymbol{y}_2)) \to A_D(\boldsymbol{X}_1(\boldsymbol{x}), \boldsymbol{y}_1) \wedge A_D(\boldsymbol{X}_2(\boldsymbol{x}), \boldsymbol{y}_2)\big)$$

For $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ just take $\boldsymbol{t}_1 = \boldsymbol{t}_2 = \lambda \boldsymbol{x}.\boldsymbol{x}$. Now let t_{A_D} be as in theorem 2.2.5, then we can take $\boldsymbol{t} = \lambda \boldsymbol{x}\boldsymbol{y}_1\boldsymbol{y}_2.\mathbf{Cond}(t_{A_D}(\boldsymbol{x}, \boldsymbol{y}_1), \boldsymbol{y}_2, \boldsymbol{y}_1)$ for $\boldsymbol{Y}$. Again, the following can be shown

$$\mathsf{T} \vdash A_D(\boldsymbol{x}, \boldsymbol{t}(\boldsymbol{x}, \boldsymbol{y}_1, \boldsymbol{y}_2)) \to A_D(\boldsymbol{t}_1(\boldsymbol{x}), \boldsymbol{y}_1) \wedge A_D(\boldsymbol{t}_2(\boldsymbol{x}), \boldsymbol{y}_2).$$

P2).

$\boxed{(A \to A \vee B)^D =}$

$$\big(\exists \boldsymbol{x} \forall \boldsymbol{y} A_D(\boldsymbol{x}, \boldsymbol{y}) \to \exists z, \boldsymbol{x}_1, \boldsymbol{x}_2 \forall \boldsymbol{y}_1, \boldsymbol{y}_2 [(z = 0 \wedge A_D(\boldsymbol{x}_1, \boldsymbol{y}_1)) \vee (z \neq 0 \wedge B_D(\boldsymbol{x}_2, \boldsymbol{y}_2))]\big)^D =$$
$$\exists Z, \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{Y} \forall \boldsymbol{y}_1, \boldsymbol{y}_2, \boldsymbol{x} \big(A_D(\boldsymbol{x}, \boldsymbol{Y}(\boldsymbol{x}, \boldsymbol{y}_1, \boldsymbol{y}_2)) \to [(Z(\boldsymbol{x}) = 0 \wedge A_D(\boldsymbol{X}_1(\boldsymbol{x}), \boldsymbol{y}_1)) \vee$$
$$\vee (Z(\boldsymbol{x}) \neq 0 \wedge B_D(\boldsymbol{X}_2(\boldsymbol{x}), \boldsymbol{y}_2))]\big)$$

For Z take $t = \lambda \boldsymbol{x}.0$, for $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ take $\boldsymbol{t}_1 = \boldsymbol{t}_2 = \lambda \boldsymbol{x}.\boldsymbol{x}$, and for $\boldsymbol{Y}(\boldsymbol{x}, \boldsymbol{y}_1, \boldsymbol{y}_2)$ take $\boldsymbol{t} = \lambda \boldsymbol{x}\boldsymbol{y}_1\boldsymbol{y}_2.\boldsymbol{y}_1$. We then get

$$\mathsf{T} \vdash A_D(\boldsymbol{x}, \boldsymbol{t}(\boldsymbol{x}, \boldsymbol{y}_1, \boldsymbol{y}_2)) \to [(t(\boldsymbol{x}) = 0 \wedge A_D(\boldsymbol{t}_1(\boldsymbol{x}), \boldsymbol{y}_1)) \vee (t(\boldsymbol{x}) \neq 0 \wedge B_D(\boldsymbol{t}_2(\boldsymbol{x}), \boldsymbol{y}_2))].$$

$\boxed{(A \wedge B \to A)^D =}$

$$\big(\exists \boldsymbol{x}_1, \boldsymbol{x}_2 \forall \boldsymbol{y}_1, \boldsymbol{y}_2 (A_D(\boldsymbol{x}_1, \boldsymbol{y}_1) \wedge B_D(\boldsymbol{x}_2, \boldsymbol{y}_2)) \to \exists \boldsymbol{x} \forall \boldsymbol{y} A_D(\boldsymbol{x}, \boldsymbol{y})\big)^D =$$
$$\exists \boldsymbol{X}, \boldsymbol{Y}_1, \boldsymbol{Y}_2 \forall \boldsymbol{y}, \boldsymbol{x}_1, \boldsymbol{x}_2 \big(A_D(\boldsymbol{x}_1, \boldsymbol{Y}_1(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})) \wedge B_D(\boldsymbol{x}_2, \boldsymbol{Y}_2(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})) \to A_D(\boldsymbol{X}(\boldsymbol{x}_1, \boldsymbol{x}_2), \boldsymbol{y})\big)$$

For $\boldsymbol{X}$ take $\boldsymbol{t} = \lambda \boldsymbol{x}_1\boldsymbol{x}_2.\boldsymbol{x}_1$ and for $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ take $\boldsymbol{t}_1 = \boldsymbol{t}_2 = \lambda \boldsymbol{x}_1\boldsymbol{x}_2\boldsymbol{y}.\boldsymbol{y}$. So trivially

$$\mathsf{T} \vdash A_D(\boldsymbol{x}_1, \boldsymbol{t}_1(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})) \wedge B_D(\boldsymbol{x}_2, \boldsymbol{t}_2(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y})) \to A_D(\boldsymbol{t}(\boldsymbol{x}_1, \boldsymbol{x}_2), \boldsymbol{y}).$$

P3).

$\boxed{(A \vee B \to B \vee A)^D =}$

$$\big(\exists z_1, \boldsymbol{x}_1, \boldsymbol{x}_2 \forall \boldsymbol{y}_1, \boldsymbol{y}_2 [(z_1 = 0 \wedge A_D(\boldsymbol{x}_1, \boldsymbol{y}_1)) \vee (z_1 \neq 0 \wedge B_D(\boldsymbol{x}_2, \boldsymbol{y}_2))] \to$$
$$\to \exists z_2, \boldsymbol{x}_3, \boldsymbol{x}_4 \forall \boldsymbol{y}_3, \boldsymbol{y}_4 [(z_2 = 0 \wedge B_D(\boldsymbol{x}_3, \boldsymbol{y}_3)) \vee (z_2 \neq 0 \wedge A_D(\boldsymbol{x}_4, \boldsymbol{y}_4))]\big)^D =$$
$$\exists Z_2, \boldsymbol{X}_3, \boldsymbol{X}_4, \boldsymbol{Y}_1, \boldsymbol{Y}_2 \forall z_1, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_3, \boldsymbol{y}_4$$
$$\big([(z_1 = 0 \wedge A_D(\boldsymbol{x}_1, \boldsymbol{Y}_1(z_1, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_3, \boldsymbol{y}_4))) \vee (z_1 \neq 0 \wedge B_D(\boldsymbol{x}_2, \boldsymbol{Y}_2(z_1, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_3, \boldsymbol{y}_4)))] \to$$
$$\to [(Z_2(z_1, \boldsymbol{x}_1, \boldsymbol{x}_2) = 0 \wedge B_D(\boldsymbol{X}_3(z_1, \boldsymbol{x}_1, \boldsymbol{x}_2), \boldsymbol{y}_3)) \vee (Z_2(z_1, \boldsymbol{x}_1, \boldsymbol{x}_2) \neq 0 \wedge A_D(\boldsymbol{X}_4(z_1, \boldsymbol{x}_1, \boldsymbol{x}_2), \boldsymbol{y}_4))]\big)$$

Take the following terms:

For $\boldsymbol{Y}_1$ take $\boldsymbol{t}_1 = \lambda z_1 \boldsymbol{x}_1 \boldsymbol{x}_2 \boldsymbol{y}_3 \boldsymbol{y}_4 . \boldsymbol{y}_4$.
For $\boldsymbol{Y}_2$ take $\boldsymbol{t}_2 = \lambda z_1 \boldsymbol{x}_1 \boldsymbol{x}_2 \boldsymbol{y}_3 \boldsymbol{y}_4 . \boldsymbol{y}_3$.
For $\boldsymbol{X}_3$ take $\boldsymbol{t}_3 = \lambda z_1 \boldsymbol{x}_1 \boldsymbol{x}_2 . \boldsymbol{x}_2$.
For $\boldsymbol{X}_4$ take $\boldsymbol{t}_4 = \lambda z_1 \boldsymbol{x}_1 \boldsymbol{x}_2 . \boldsymbol{x}_1$.
For Z_2 take $t = \lambda z_1 \boldsymbol{x}_1 \boldsymbol{x}_2 . (1 \dot{-} z_1)$.

Then,

$$\mathsf{T} \vdash [(z_1 = 0 \wedge A_D(\boldsymbol{x}_1, \boldsymbol{t}_1(z_1, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_3, \boldsymbol{y}_4))) \vee (z_1 \neq 0 \wedge B_D(\boldsymbol{x}_2, \boldsymbol{t}_2(z_1, \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_3, \boldsymbol{y}_4)))] \to$$
$$\to [(t(z_1, \boldsymbol{x}_1, \boldsymbol{x}_2) = 0 \wedge B_D(\boldsymbol{t}_3(z_1, \boldsymbol{x}_1, \boldsymbol{x}_2), \boldsymbol{y}_3)) \vee (Z_2(z_1, \boldsymbol{x}_1, \boldsymbol{x}_2) \neq 0 \wedge A_D(\boldsymbol{t}_4(z_1, \boldsymbol{x}_1, \boldsymbol{x}_2), \boldsymbol{y}_4))]$$

Replacing and evaluating terms, the above statement results in the more readable form:

$$\mathsf{T} \vdash (z_1 = 0 \wedge A_D(\boldsymbol{x}_1, \boldsymbol{y}_4)) \vee (z_1 \neq 0 \wedge B_D(\boldsymbol{x}_2, \boldsymbol{y}_3)) \to$$
$$\to (1 \dot{-} z_1 = 0 \wedge B_D(\boldsymbol{x}_2, \boldsymbol{y}_3)) \vee (1 \dot{-} z_1 \neq 0 \wedge A_D(\boldsymbol{x}_1, \boldsymbol{y}_4)).$$

$\boxed{(A \wedge B \to B \wedge A)^D =}$

$$\big(\exists \boldsymbol{x}_1, \boldsymbol{x}_2 \forall \boldsymbol{y}_1, \boldsymbol{y}_2 (A_D(\boldsymbol{x}_1, \boldsymbol{y}_1) \wedge B_D(\boldsymbol{x}_2, \boldsymbol{y}_2)) \to \exists \boldsymbol{x}_3, \boldsymbol{x}_4 \forall \boldsymbol{y}_3, \boldsymbol{y}_4 (B_D(\boldsymbol{x}_3, \boldsymbol{y}_3) \wedge A_D(\boldsymbol{x}_4, \boldsymbol{y}_4))\big)^D =$$
$$\exists \boldsymbol{Y}_1, \boldsymbol{Y}_2, \boldsymbol{X}_3, \boldsymbol{X}_4 \forall \boldsymbol{y}_3, \boldsymbol{y}_4, \boldsymbol{x}_1, \boldsymbol{x}_2 \big(A_D(\boldsymbol{x}_1, \boldsymbol{Y}_1(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_3, \boldsymbol{y}_4)) \wedge B_D(\boldsymbol{x}_2, \boldsymbol{Y}_2(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_3, \boldsymbol{y}_4))) \to$$
$$\to B_D(\boldsymbol{X}_3(\boldsymbol{x}_1, \boldsymbol{x}_2), \boldsymbol{y}_3) \wedge A_D(\boldsymbol{X}_4(\boldsymbol{x}_1, \boldsymbol{x}_2), \boldsymbol{y}_4)\big)$$

Take the following terms:

For $\boldsymbol{Y}_1$ take $\boldsymbol{t}_1 = \lambda z_1 \boldsymbol{x}_1 \boldsymbol{x}_2 \boldsymbol{y}_3 \boldsymbol{y}_4 . \boldsymbol{y}_4$.
For $\boldsymbol{Y}_2$ take $\boldsymbol{t}_2 = \lambda z_1 \boldsymbol{x}_1 \boldsymbol{x}_2 \boldsymbol{y}_3 \boldsymbol{y}_4 . \boldsymbol{y}_3$.
For $\boldsymbol{X}_3$ take $\boldsymbol{t}_3 = \lambda z_1 \boldsymbol{x}_1 \boldsymbol{x}_2 . \boldsymbol{x}_2$.
For $\boldsymbol{X}_4$ take $\boldsymbol{t}_4 = \lambda z_1 \boldsymbol{x}_1 \boldsymbol{x}_2 . \boldsymbol{x}_1$.

The rest works exactly as above.

P4).

$$(\perp \to A)^D = \exists \boldsymbol{x} \forall \boldsymbol{y} (\perp \to A_D(\boldsymbol{x}, \boldsymbol{y}))$$

For $\boldsymbol{x}$ just take any closed term of suitable type.

P6). Assume

$\mathsf{T} \vdash A_D(\boldsymbol{t}_1, \boldsymbol{y}_1)$,

$\mathsf{T} \vdash A_D(\boldsymbol{x}_1, \boldsymbol{t}_2(\boldsymbol{x}_1, \boldsymbol{y}_2)) \rightarrow B_D(\boldsymbol{t}_3(\boldsymbol{x}_1), \boldsymbol{y}_2)$.

We have to find terms $\boldsymbol{t}_4$ such that

$\mathsf{T} \vdash B_D(\boldsymbol{t}_4, \boldsymbol{y}_2)$.

Since T allows substitution, we replace $\boldsymbol{x}_1$ with $\boldsymbol{t}_1$ and $\boldsymbol{y}_1$ with $\boldsymbol{t}_2(\boldsymbol{t}_1, \boldsymbol{y}_2)$ to obtain

$\mathsf{T} \vdash A_D(\boldsymbol{t}_1, \boldsymbol{t}_2(\boldsymbol{t}_1, \boldsymbol{y}_2))$ and

$\mathsf{T} \vdash A_D(\boldsymbol{t}_1, \boldsymbol{t}_2(\boldsymbol{t}_1, \boldsymbol{y}_2)) \rightarrow B_D(\boldsymbol{t}_3(\boldsymbol{t}_1), \boldsymbol{y}_2)$.

Applying P6 yields

$\mathsf{T} \vdash B_D(\boldsymbol{t}_3(\boldsymbol{t}_1), \boldsymbol{y}_2)$.

So we take $\boldsymbol{t}_3(\boldsymbol{t}_1)$ for $\boldsymbol{t}_4$.

P7). Assume

$\mathsf{T} \vdash A_D(\boldsymbol{x}_1, \boldsymbol{t}_1(\boldsymbol{x}_1, \boldsymbol{y}_2)) \rightarrow B_D(\boldsymbol{t}_2(\boldsymbol{x}_1), \boldsymbol{y}_2)$ and

$\mathsf{T} \vdash B_D(\boldsymbol{x}_3, \boldsymbol{t}_3(\boldsymbol{x}_3, \boldsymbol{y}_4)) \rightarrow C_D(\boldsymbol{t}_4(\boldsymbol{x}_3), \boldsymbol{y}_4)$.

We need terms $\boldsymbol{t}_5$ and $\boldsymbol{t}_6$ such that

$\mathsf{T} \vdash A_D(\boldsymbol{x}_1, \boldsymbol{t}_5(\boldsymbol{x}_1, \boldsymbol{y}_4)) \rightarrow C_D(\boldsymbol{t}_6(\boldsymbol{x}_1), \boldsymbol{y}_4)$.

Again we use substitution and replace $\boldsymbol{t}_2(\boldsymbol{x}_1)$ for $\boldsymbol{x}_3$ and $\boldsymbol{t}_3(\boldsymbol{t}_2(\boldsymbol{x}_1), \boldsymbol{y}_4)$ for $\boldsymbol{y}_2$ and get

$\mathsf{T} \vdash A_D(\boldsymbol{x}_1, \boldsymbol{t}_1(\boldsymbol{x}_1, \boldsymbol{t}_3(\boldsymbol{t}_2(\boldsymbol{x}_1), \boldsymbol{y}_4))) \rightarrow B_D(\boldsymbol{t}_2(\boldsymbol{x}_1), \boldsymbol{t}_3(\boldsymbol{t}_2(\boldsymbol{x}_1), \boldsymbol{y}_4))$ and

$\mathsf{T} \vdash B_D(\boldsymbol{t}_2(\boldsymbol{x}_1), \boldsymbol{t}_3(\boldsymbol{t}_2(\boldsymbol{x}_1), \boldsymbol{y}_4)) \rightarrow C_D(\boldsymbol{t}_4(\boldsymbol{t}_2(\boldsymbol{x}_1)), \boldsymbol{y}_4)$.

By P7 we obtain

$\mathsf{T} \vdash A_D(\boldsymbol{x}_1, \boldsymbol{t}_1(\boldsymbol{x}_1, \boldsymbol{t}_3(\boldsymbol{t}_2(\boldsymbol{x}_1), \boldsymbol{y}_4))) \rightarrow C_D(\boldsymbol{t}_4(\boldsymbol{t}_2(\boldsymbol{x}_1)), \boldsymbol{y}_4)$.

So we take $\lambda \boldsymbol{x}_1 \boldsymbol{y}_4 . \boldsymbol{t}_1(\boldsymbol{x}_1, \boldsymbol{t}_3(\boldsymbol{t}_2(\boldsymbol{x}_1), \boldsymbol{y}_4))$ for $\boldsymbol{t}_5$ and $\lambda \boldsymbol{x}_1 . \boldsymbol{t}_4(\boldsymbol{t}_2(\boldsymbol{x}_1))$ for $\boldsymbol{t}_6$.

P8). Assume

$\mathsf{T} \vdash A_D(\boldsymbol{x}_1, \boldsymbol{t}_1(\boldsymbol{x}_1, \boldsymbol{y}_2)) \rightarrow B_D(\boldsymbol{t}_2(\boldsymbol{x}_1), \boldsymbol{y}_2)$.

We must construct terms t, $\boldsymbol{t}_3$, $\boldsymbol{t}_4$, $\boldsymbol{t}_5$ and $\boldsymbol{t}_6$, such that

$$\mathsf{T} \vdash [(z_1 = 0 \wedge A_D(\boldsymbol{x}_1, \boldsymbol{t}_3(z_1, \boldsymbol{x}_1, \boldsymbol{x}_3, \boldsymbol{y}_2, \boldsymbol{y}_4))) \vee (z_1 \neq 0 \wedge C_D(\boldsymbol{x}_3, \boldsymbol{t}_4(z_1, \boldsymbol{x}_1, \boldsymbol{x}_3, \boldsymbol{y}_2, \boldsymbol{y}_4)))] \rightarrow$$
$$\rightarrow [(t(z_1, \boldsymbol{x}_1, \boldsymbol{x}_3) = 0 \wedge B_D(\boldsymbol{t}_5(z_1, \boldsymbol{x}_1, \boldsymbol{x}_3), \boldsymbol{y}_2)) \vee (t(z_1, \boldsymbol{x}_1, \boldsymbol{x}_3) \neq 0 \wedge C_D(\boldsymbol{t}_6(z_1, \boldsymbol{x}_1, \boldsymbol{x}_3), \boldsymbol{y}_4))].$$

Starting from our assumption, we can easily prove

$\mathsf{T} \vdash (z_1 = 0 \wedge A_D(\boldsymbol{x}_1, \boldsymbol{t}_1(\boldsymbol{x}_1, \boldsymbol{y}_2))) \rightarrow (z_1 = 0 \wedge B_D(\boldsymbol{t}_2(\boldsymbol{x}_1), \boldsymbol{y}_2))$.

Applying P8 we finally get

$$\mathsf{T} \vdash (z_1 = 0 \wedge A_D(\boldsymbol{x}_1, \boldsymbol{t}_1(\boldsymbol{x}_1, \boldsymbol{y}_2))) \vee (z_1 \neq 0 \wedge C_D(\boldsymbol{x}_3, \boldsymbol{y}_4)) \rightarrow$$
$$\rightarrow (z_1 = 0 \wedge B_D(\boldsymbol{t}_2(\boldsymbol{x}_1), \boldsymbol{y}_2)) \vee (z_1 \neq 0 \wedge C_D(\boldsymbol{x}_3, \boldsymbol{y}_4)).$$

So we take the following terms

For t take $\lambda z_1 \boldsymbol{x}_1 \boldsymbol{x}_3 . z_1$.
For $\boldsymbol{t}_3$ take $\lambda z_1 \boldsymbol{x}_1 \boldsymbol{x}_3 \boldsymbol{y}_2 \boldsymbol{y}_4 . \boldsymbol{t}_1(\boldsymbol{x}_1, \boldsymbol{y}_2)$.
For $\boldsymbol{t}_4$ take $\lambda z_1 \boldsymbol{x}_1 \boldsymbol{x}_3 \boldsymbol{y}_2 \boldsymbol{y}_4 . \boldsymbol{y}_4$.
For $\boldsymbol{t}_5$ take $\lambda z_1 \boldsymbol{x}_1 \boldsymbol{x}_3 . \boldsymbol{t}_2(\boldsymbol{x}_1)$.
For $\boldsymbol{t}_6$ take $\lambda z_1 \boldsymbol{x}_1 \boldsymbol{x}_3 . \boldsymbol{x}_3$.

P9). Assume

$$\mathsf{T} \vdash A_D(\boldsymbol{x}_1, \boldsymbol{t}_1(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_3)) \rightarrow [B_D(\boldsymbol{x}_2, \boldsymbol{t}_2(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_3)) \rightarrow C_D(\boldsymbol{t}_3(\boldsymbol{x}_1, \boldsymbol{x}_2), \boldsymbol{y}_3)].$$

Terms $\boldsymbol{t}_4$, $\boldsymbol{t}_5$ and $\boldsymbol{t}_6$ are needed, such that

$$\mathsf{T} \vdash [A_D(\boldsymbol{x}_1, \boldsymbol{t}_4(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_3)) \wedge B_D(\boldsymbol{x}_2, \boldsymbol{t}_5(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{y}_3))] \rightarrow C_D(\boldsymbol{t}_6(\boldsymbol{x}_1, \boldsymbol{x}_2), \boldsymbol{y}_3).$$

Obviously, we just have to apply P9 and take $\boldsymbol{t}_{i-3}$ for $\boldsymbol{t}_i$ $(i \in \{4, 5, 6\})$.

P10). This works completely analogously to P9.

Quantifier axioms and rules

Q1).

$$(\forall z A \rightarrow A[t/z])^D =$$
$$\big(\exists \boldsymbol{X}_1 \forall \boldsymbol{y}_1, z A_D(\boldsymbol{X}_1(z), \boldsymbol{y}_1, z) \rightarrow \exists \boldsymbol{x}_2, \forall \boldsymbol{y}_2 A_D(\boldsymbol{x}_2, \boldsymbol{y}_2, t)\big)^D$$
$$\exists \boldsymbol{X}_2, \boldsymbol{Y}_1, Z \forall \boldsymbol{X}_1, \boldsymbol{y}_2 \big(A_D(\boldsymbol{X}_1(Z(\boldsymbol{X}_1, \boldsymbol{y}_2)), \boldsymbol{Y}_1(\boldsymbol{X}_1, \boldsymbol{y}_2), Z(\boldsymbol{X}_1, \boldsymbol{y}_2)) \rightarrow A_D(\boldsymbol{X}_2(\boldsymbol{X}_1), \boldsymbol{y}_2, t)\big)$$

For $\boldsymbol{Y}_1$, $\boldsymbol{X}_2$ and Z take terms $\boldsymbol{t}_1 = \lambda \boldsymbol{X}_1 \boldsymbol{y}_1 . \boldsymbol{y}_2$, $\boldsymbol{t}_2 = \lambda \boldsymbol{X}_1 . \boldsymbol{X}_1(t)$ and $t_3 = \lambda \boldsymbol{X}_1 \boldsymbol{y}_1 . t$. The following now obviously holds

$$\mathsf{T} \vdash A_D(\boldsymbol{X}_1(t_3(\boldsymbol{X}_1, \boldsymbol{y}_1)), \boldsymbol{t}_1(\boldsymbol{X}_1, \boldsymbol{y}_1), t_3(\boldsymbol{X}_1, \boldsymbol{y}_1)) \rightarrow A_D(\boldsymbol{t}_2(\boldsymbol{x}_1), \boldsymbol{y}_2, t).$$

Q2).

$$(A[t/z] \rightarrow \exists z A)^D =$$
$$\big(\exists \boldsymbol{x}_1 \forall \boldsymbol{y}_1 A_D(\boldsymbol{x}_1, \boldsymbol{y}_1, t) \rightarrow \exists z, \boldsymbol{x}_2 \forall \boldsymbol{y}_2 A_D(\boldsymbol{x}_2, \boldsymbol{y}_2, z)\big)^D =$$
$$\exists \boldsymbol{X}_2, \boldsymbol{Y}_1, Z \forall \boldsymbol{x}_1, \boldsymbol{y}_2 \big(A_D(\boldsymbol{x}_1, \boldsymbol{Y}_1(\boldsymbol{x}_1, \boldsymbol{y}_2), t) \rightarrow A_D(\boldsymbol{X}_2(\boldsymbol{x}_1), \boldsymbol{y}_2, Z(\boldsymbol{x}_1))\big)$$

For $\boldsymbol{Y}_1$, $\boldsymbol{X}_2$ and Z take terms $\boldsymbol{t}_1 = \lambda \boldsymbol{x}_1 \boldsymbol{y}_2 . \boldsymbol{y}_2$, $\boldsymbol{t}_2 = \lambda \boldsymbol{x}_1 . \boldsymbol{x}_1$ and $t_3 = \lambda \boldsymbol{x}_1 . t$. Again the rest is more or less trivial

$$\mathsf{T} \vdash A_D(\boldsymbol{x}_1, \boldsymbol{t}_1(\boldsymbol{x}_1, \boldsymbol{y}_2), t) \to A_D(\boldsymbol{t}_2(\boldsymbol{x}_1), \boldsymbol{y}_2, t_3(\boldsymbol{x}_1)).$$

For the last two rules, we need to consider the additional free variable z as well.

Q3). Assume

$$\mathsf{T} \vdash A_D(\boldsymbol{x}_1, \boldsymbol{t}_1(z, \boldsymbol{x}_1, \boldsymbol{y}_2)) \to B_D(\boldsymbol{t}_2(z, \boldsymbol{x}_1), \boldsymbol{y}_2, z).$$

We need to find terms $\boldsymbol{t}_3$ and $\boldsymbol{t}_4$, such that

$$\mathsf{T} \vdash A_D(\boldsymbol{x}_1, \boldsymbol{t}_3(\boldsymbol{x}_1, \boldsymbol{y}_2, z)) \to B_D(\boldsymbol{t}_4(\boldsymbol{x}_1, z), \boldsymbol{y}_2, z).$$

Obviously we just take $\lambda \boldsymbol{x}_1 \boldsymbol{y}_2 z . \boldsymbol{t}_1(z, \boldsymbol{x}_1, \boldsymbol{y}_2)$ for $\boldsymbol{t}_3$ and $\lambda \boldsymbol{x}_1 z . \boldsymbol{t}_2(z, \boldsymbol{x}_1)$ for $\boldsymbol{t}_4$.

Q4). Assume

$$\mathsf{T} \vdash A_D(z, \boldsymbol{x}_1, \boldsymbol{t}_1(z, \boldsymbol{x}_1, \boldsymbol{y}_2)) \to B_D(\boldsymbol{t}_2(z, \boldsymbol{x}_1), \boldsymbol{y}_2).$$

We need to find terms $\boldsymbol{t}_3$ and $\boldsymbol{t}_4$, such that

$$\mathsf{T} \vdash A_D(z, \boldsymbol{x}_1, \boldsymbol{t}_3(z, \boldsymbol{x}_1, \boldsymbol{y}_2)) \to B_D(\boldsymbol{t}_4(z, \boldsymbol{x}_1), \boldsymbol{y}_2).$$

But this is completely trivial.

<u>Non-logical axioms and rules</u>

Most of the non-logical axioms are unproblematic, since they are either purely universal or do not include quantifiers at all. Only $\mathsf{QF\text{-}ER}$, Eq2 and Ind need comment.

$\mathsf{QF\text{-}ER}$). Assume

$$\mathsf{T} \vdash A_D(\boldsymbol{x}, \boldsymbol{t}_1(\boldsymbol{x}, \boldsymbol{z}_1)) \to s(\boldsymbol{z}_1) =_0 t(\boldsymbol{z}_1).$$

We need terms $\boldsymbol{t}_2$ such that

$$\mathsf{T} \vdash A_D(\boldsymbol{x}, \boldsymbol{t}_2(\boldsymbol{x}, \boldsymbol{z}_2)) \to r(s, \boldsymbol{z}_2) =_0 r(t, \boldsymbol{z}_2).$$

But since A is quantifier-free, quantifiers in A^D are only introduced via disjunctions. These additional variables can be eliminated by T, so that we obtain

$$\mathsf{T} \vdash A \to s(\boldsymbol{z}_1) =_0 t(\boldsymbol{z}_1).$$

We now apply $\mathsf{QF\text{-}ER}$

$$\mathsf{T} \vdash A \to r(s, \boldsymbol{z}_2) =_0 r(t, \boldsymbol{z}_2)$$

and then reintroduce the additional variables and terms to get what is required from above.

Eq2).

$$(u = v \wedge A[u/w] \rightarrow A[v/w])^D = \\ \exists \boldsymbol{X}_2, \boldsymbol{Y}_1 \forall \boldsymbol{x}_1, \boldsymbol{y}_2 (u = v \wedge A_D(\boldsymbol{x}_1, \boldsymbol{Y}_1(\boldsymbol{x}_1, \boldsymbol{y}_2), u) \rightarrow A_D(\boldsymbol{X}_2(\boldsymbol{x}_1), \boldsymbol{y}_2, v))$$

It is obvious, what terms are needed for $\boldsymbol{Y}_1$ and $\boldsymbol{X}_2$.

Ind). Assume

$\mathsf{T} \vdash A_D(0, \boldsymbol{t}_1, \boldsymbol{y})$ and

$\mathsf{T} \vdash A_D(u, \boldsymbol{x}_1, \boldsymbol{t}_2(u, \boldsymbol{x}_1, \boldsymbol{y}_2)) \rightarrow A_D(u', \boldsymbol{t}_3(u, \boldsymbol{x}_1), \boldsymbol{y}_2)$.

We need terms $\boldsymbol{t}_4$, such that

$\mathsf{T} \vdash A_D(u, \boldsymbol{t}_4(u), \boldsymbol{y})$.

Using the recursors R, we define the following sequence of terms $\boldsymbol{t}_4 = t_{4_1}, \ldots, t_{4_r}$, by

$$\begin{array}{rcl} t_{4_i}(0) & = & t_{1_i} \\ t_{4_i}(u') & = & t_{3_i}(u, \boldsymbol{t}_4(u)) \end{array} \quad \Rightarrow \quad t_{4_i}(u) = \mathrm{R}(t_{1_i}, \lambda u t_{4_i}.t_{3_i}(u, \boldsymbol{t}_4(u))), u)$$

By substitution into our assumptions

$\mathsf{T} \vdash A_D(0, \boldsymbol{t}_4(0), \boldsymbol{y})$ and

$\mathsf{T} \vdash A_D(u, \boldsymbol{t}_4(u), \boldsymbol{t}_2(u, \boldsymbol{t}_4(u), \boldsymbol{y}_2)) \rightarrow A_D(u', \boldsymbol{t}_3(u, \boldsymbol{t}_4(u)), \boldsymbol{y}_2)$,

which is

$\mathsf{T} \vdash A_D(u, \boldsymbol{t}_4(u), \boldsymbol{t}_2(u, \boldsymbol{t}_4(u), \boldsymbol{y}_2)) \rightarrow A_D(u', \boldsymbol{t}_4(u'), \boldsymbol{y}_2)$.

By the induction lemma 2.3.3, we finally get

$\mathsf{T} \vdash A_D(u, \boldsymbol{t}_4(u), \boldsymbol{y})$. □

Theorem 2.3.3 (Induction Lemma). *From assumptions $A(0, \boldsymbol{y})$ and $A(u, \boldsymbol{t}(u, \boldsymbol{y})) \rightarrow A(u', \boldsymbol{y})$, we can prove $A(u, \boldsymbol{y})$ in T.*

Proof. This proof will be informal and also requires the use of new symbols ($<, >, \leqslant, \geqslant$). These symbols are all introduced via primitive recursion, which makes all of their properties used in the proof provable in T. For more details see [15].

Using the Recursors and the functionals $\boldsymbol{t}$, we define the following functionals $\boldsymbol{e}$

$$\begin{array}{rcl} \boldsymbol{e}(0, u, \boldsymbol{y}) & = & \boldsymbol{y} \\ \boldsymbol{e}(v', u, \boldsymbol{y}) & = & \boldsymbol{t}(u \dot{-} v', \boldsymbol{e}(v, u, \boldsymbol{y})) \end{array} \qquad e_i = \lambda v u \boldsymbol{y}.\mathrm{R}(y_i, \lambda a e_i.t_i(u \dot{-} a', \boldsymbol{e}(v, u, \boldsymbol{y})), v)$$

By the use of these functionals we obtain,

I) $\quad \mathsf{T} \vdash w < u \to [A(w, \boldsymbol{e}(u \dot{-} w, u, \boldsymbol{y})) \to A(w', \boldsymbol{e}(u \dot{-} w', u, \boldsymbol{y}))]$

Proof: Assume $w < u$. Then $u \dot{-} w = (u \dot{-} w')'$ and so

$$\begin{aligned} \boldsymbol{e}(u \dot{-} w, u, \boldsymbol{y}) &= \boldsymbol{e}((u \dot{-} w')', u, \boldsymbol{y}) \\ &= \boldsymbol{t}(u \dot{-} (u \dot{-} w')', \boldsymbol{e}(u \dot{-} w', u, \boldsymbol{y})) \\ &= \boldsymbol{t}(u \dot{-} (u \dot{-} w), \boldsymbol{e}(u \dot{-} w', u, \boldsymbol{y})) \\ &= \boldsymbol{t}(u, \boldsymbol{e}(u \dot{-} w', u, \boldsymbol{y})) \end{aligned}$$

Because of

$$\mathsf{T} \vdash A(u, \boldsymbol{t}(u, \boldsymbol{y})) \to A(u', \boldsymbol{y})$$

we finally get

$$\mathsf{T} \vdash A(w, \boldsymbol{t}(w, \boldsymbol{e}(u \dot{-} w', u, \boldsymbol{y}))) \to A(w', \boldsymbol{e}(u \dot{-} w', u, \boldsymbol{y}))$$

$$\mathsf{T} \vdash A(w, \boldsymbol{e}(u \dot{-} w, u, \boldsymbol{y})) \to A(w', \boldsymbol{e}(u \dot{-} w', u, \boldsymbol{y})).$$

The next step is to show the following

II) $\quad \mathsf{T} \vdash w \leqslant u \to A(w, \boldsymbol{e}(u \dot{-} w, u, \boldsymbol{y}))$

Proof: By induction on w

$w = 0$: By our first assumption $A(0, \boldsymbol{y})$ holds in T. And since $0 \leqslant u$ also holds in T, the claim is derivable for $w = 0$.

$w \to w'$: We first combine $w' \leqslant u \to w < u$ and I) to obtain

$$\mathsf{T} \vdash w' \leqslant u \to [A(w, \boldsymbol{e}(u \dot{-} w, u, \boldsymbol{y})) \to A(w', \boldsymbol{e}(u \dot{-} w', u, \boldsymbol{y}))]$$

and then again combine $w' \leqslant u \to w < u$ with the induction hypothesis to get

$$\mathsf{T} \vdash w' \leqslant u \to A(w, \boldsymbol{e}(u \dot{-} w, u, \boldsymbol{y})).$$

The rest is trivial.

For the last step, we now use II), setting w to be u

$$\mathsf{T} \vdash u \leqslant u \to A(u, \boldsymbol{e}(u \dot{-} u, u, \boldsymbol{y})).$$

But since $u \leqslant u$ always holds in T and $u \dot{-} u = 0$

$$\mathsf{T} \vdash A(u, \boldsymbol{e}(0, u, \boldsymbol{y})),$$

which is

$$\mathsf{T} \vdash A(u, \boldsymbol{y})$$

by definition of $\boldsymbol{e}$. $\square$

Lemma 2.3.4. *The law of the excluded middle (P5) is not D-interpretable.*

Proof. Assume that it were, then we would have to be able to find closed terms t_1, $\boldsymbol{t}_2$ and $\boldsymbol{t}_3$, such that

$$\mathsf{T} \vdash [t_1 = 0 \wedge A_D(\boldsymbol{t}_2, \boldsymbol{y})] \vee [t_1 \neq 0 \wedge \neg A_D(\boldsymbol{x}, \boldsymbol{t}_3(\boldsymbol{x}))].$$

Even though

$$\mathsf{T} \vdash A_D(\boldsymbol{t}_2, \boldsymbol{y}) \vee \neg A_D(\boldsymbol{x}, \boldsymbol{t}_3(\boldsymbol{x})),$$

for any closed terms $\boldsymbol{t}_2$ and $\boldsymbol{t}_3$ of suitable type, we can not find a term t_1, since it would have to decide which disjunct is true and this is in general not possible. $\square$

The D-interpretation is not only applicable to $\mathsf{WE\text{-}PA}_i^\omega$. It can obviously be applied to any theory which is defined over the language $\mathcal{L}_0^\omega$. But what about theorem 2.3.2? What happens if we take a restricted or classical variant of $\mathsf{WE\text{-}PA}_i^\omega$ or if we replace $\mathsf{QF\text{-}ER}$ by (E)?

For classical logic, we only have to combine theorem 1.3.6 and 2.3.2 to obtain.

Corollary 2.3.5. *If $\mathsf{WE\text{-}PA}^\omega \vdash A(\boldsymbol{z})$ (A containing at most $\boldsymbol{z}$ free), then there is a sequence of closed terms $\boldsymbol{t}$, such that $\mathsf{T} \vdash (A^N)_D(\boldsymbol{t}, \boldsymbol{y}, \boldsymbol{z})$. We say that $\mathsf{WE\text{-}PA}^\omega$ is ND-interpreted in T.*

For the restricted theories it suffices to replace T by $\widehat{\mathsf{T}}$.

Corollary 2.3.6. *$\mathsf{WE\text{-}PRA}_i^\omega$ and $\mathsf{WE\text{-}PRA}^\omega$ are (N)D-interpreted in $\widehat{\mathsf{T}}$.*

Proof. If we go through the proof of theorem 2.3.2, we see that the recursors R are only used in two cases, namely for the definition of Cond used in the interpretation of P1 and P2, and in the interpretation of induction. In the definition of Cond we already know that R can be replaced by $\widehat{\mathrm{R}}$. If we consider induction in $\mathsf{WE\text{-}PRA}_i^\omega$ we see that neither R nor $\widehat{\mathrm{R}}$ are used in the interpretation of it, since it is only allowed for quantifier-free formulas and can therefore be directly interpreted by the induction rule of $\widehat{\mathsf{T}}$. $\square$

An easy consequence of these theorems is that all these theories are conservative over T, respectively $\widehat{\mathsf{T}}$, for quantifier-free formulas. Fully extensional theories, such as $\mathsf{E\text{-}PA}^\omega$, are not D-interpretable, since there is no functional of this theory which satisfies the D-interpretation of (E). This was proved by Howard in [15]; his proof will be quoted in section 3.4.1.

2.3.3 Extending $\mathsf{WE\text{-}PA}_i^\omega$ and $\mathsf{WE\text{-}PA}^\omega$

We now want to see what happens if we add the principles AC, IP and MP_+ to the theories $\mathsf{WE\text{-}PA}_i^\omega$ and $\mathsf{WE\text{-}PA}^\omega$. Our first observation is that adding the latter two principles to $\mathsf{WE\text{-}PA}^\omega$ is redundant, since they can both be proved therein. The next, not so obvious, observation is that in the context in which these three principles are used, IP and MP_+ are actually only used in their following weaker forms

IP' $\quad (\forall \boldsymbol{x} A \to \exists y B) \to \exists y(\forall \boldsymbol{x} A \to B) \quad$ with A quantifier-free,

MP' $\quad \neg\forall x A \to \exists x \neg A \quad$ with A quantifier-free.

If we add the stronger versions, our theories might become too strong, an assumption which will turn out to be true when considering D-interpretability. From now on, we therefore only want to consider IP' and MP'. Already the next theorem, which should actually come as no surprise, shows that these weaker forms are strong enough for our purposes.

Theorem 2.3.7. *For all formulas A, $\mathsf{WE\text{-}PA}_i^\omega \oplus \mathsf{AC} \oplus \mathsf{IP}' \oplus \mathsf{MP}' \vdash A \leftrightarrow A^D$.*

Proof. We prove this by induction on the formula complexity. The base case is trivial. For the induction step, assume the assertion holds for $A \leftrightarrow A^D$ and $B \leftrightarrow B^D$. We readily see, that in this extended theory, $A \wedge B \leftrightarrow (A \wedge B)^D$ and $A \vee B \leftrightarrow (A \vee B)^D$ hold. The case for $\exists$ is trivial and for $\forall$ we just use AC. It only remains to consider implication.

By induction hypothesis, we can assume that $(A \to B) \leftrightarrow (A^D \to B^D)$ is provable. We now transform the right side of this equivalence step by step.

$$
\begin{array}{rlcc}
A^D \to B^D \equiv & \exists \boldsymbol{x}\forall \boldsymbol{y} A_D(\boldsymbol{x},\boldsymbol{y}) \to \exists \boldsymbol{u}\forall \boldsymbol{v} B_D(\boldsymbol{u},\boldsymbol{v}) & \leftrightarrow & \text{(i)} \\
& \forall \boldsymbol{x}(\forall \boldsymbol{y} A_D(\boldsymbol{x},\boldsymbol{y}) \to \exists \boldsymbol{u}\forall \boldsymbol{v} B_D(\boldsymbol{u},\boldsymbol{v})) & \leftrightarrow & \text{(ii)} \\
& \forall \boldsymbol{x}\exists \boldsymbol{u}(\forall \boldsymbol{y} A_D(\boldsymbol{x},\boldsymbol{y}) \to \forall \boldsymbol{v} B_D(\boldsymbol{u},\boldsymbol{v})) & \leftrightarrow & \text{(iii)} \\
& \forall \boldsymbol{x}\exists \boldsymbol{u}\forall \boldsymbol{v}(\forall \boldsymbol{y} A_D(\boldsymbol{x},\boldsymbol{y}) \to B_D(\boldsymbol{u},\boldsymbol{v})) & \leftrightarrow & \text{(iv)} \\
& \forall \boldsymbol{x}\exists \boldsymbol{u}\forall \boldsymbol{v}\exists \boldsymbol{y}(A_D(\boldsymbol{x},\boldsymbol{y}) \to B_D(\boldsymbol{u},\boldsymbol{v})) & \leftrightarrow & \text{(v)} \\
& \forall \boldsymbol{x}\exists \boldsymbol{u},\boldsymbol{Y}_1\forall \boldsymbol{v}(A_D(\boldsymbol{x},\boldsymbol{Y}_1(\boldsymbol{v})) \to B_D(\boldsymbol{u},\boldsymbol{v})) & \leftrightarrow & \text{(vi)} \\
& \exists \boldsymbol{U},\boldsymbol{Y}\forall \boldsymbol{x},\boldsymbol{v}(A_D(\boldsymbol{x},\boldsymbol{Y}(\boldsymbol{x},\boldsymbol{v})) \to B_D(\boldsymbol{U}(\boldsymbol{x}),\boldsymbol{v})) & \multicolumn{2}{l}{\equiv (A \to B)^D}
\end{array}
$$

All equivalences (i)-(vi) are provable in $\mathsf{WE\text{-}PA}_i^\omega \oplus \mathsf{AC} \oplus \mathsf{IP}' \oplus \mathsf{MP}'$. For (v) and (vi) we use AC, (ii) and (iv) are provable using IP' and MP' and the rest are provable in intuitionistic logic. We recall, that for using MP' in (iv), the conclusion must be decidable, but since in this case it is quantifier-free, it is also decidable. All this has already been discussed in the definition of the D-interpretation. $\square$

Corollary 2.3.8. *For all formulas A,* $\mathsf{WE\text{-}PRA}_i^\omega \oplus \mathsf{AC} \oplus \mathsf{IP}' \oplus \mathsf{MP}' \vdash A \leftrightarrow A^D$

Proof. The proof works exactly as above, since the relative weakness of this theory does in no way affect the proof of theorem 2.3.7 □

With regard to the next chapter, the principles AC, IP' and MP' allow us to prove another important theorem.

Theorem 2.3.9. *The theory* $\mathsf{WE\text{-}PA}_i^\omega \oplus \mathsf{AC} \oplus \mathsf{IP}' \oplus \mathsf{MP}'$ *and its restricted variant are D-interpreted in* T *and* $\widehat{\mathsf{T}}$ *respectively.*

Proof. Obviously, we only have to extend the proof of theorem 2.3.2 by the following three cases:

AC).

$$(\forall x \exists y A(x,y))^D = (\forall x \exists y \exists \boldsymbol{u} \forall \boldsymbol{v} A_D(x,y,\boldsymbol{u},\boldsymbol{v}))^D = \exists Y, \boldsymbol{U} \forall x, v A_D(x, Y(x), \boldsymbol{U}(x), \boldsymbol{v})$$

And

$$(\exists Y \forall x A(x, Y(x)))^D = (\exists Y \forall x \exists \boldsymbol{u} \forall \boldsymbol{v} A_D(x, Y(x), \boldsymbol{u}, \boldsymbol{v}))^D = \exists Y, \boldsymbol{U} \forall x, v A_D(x, Y(x), \boldsymbol{U}(x), \boldsymbol{v})$$

We see, that the D-interpretation of the hypothesis and the conclusion of an instance of AC are identical. Hence interpreting an instance of AC reduces to interpreting an instance of $B \to B$, which of course can easily be done in both T and $\widehat{\mathsf{T}}$.

For IP' and MP' this works exactly the same.

IP').

$$(\forall \boldsymbol{x} A \to \exists y B)^D = (\forall \boldsymbol{x} A_D(\boldsymbol{x}) \to \exists y, \boldsymbol{u} \forall \boldsymbol{v} B_D(y, \boldsymbol{u}, \boldsymbol{v}))^D = \\ \exists \boldsymbol{X}, y, \boldsymbol{u} \forall \boldsymbol{v} (A_D(\boldsymbol{X}(\boldsymbol{v})) \to B_D(y, \boldsymbol{u}, \boldsymbol{v}))$$

And

$$(\exists y (\forall \boldsymbol{x} A \to B))^D = (\exists y (\forall \boldsymbol{x} A_D(\boldsymbol{x}) \to \exists \boldsymbol{u} \forall \boldsymbol{v} B_D(y, \boldsymbol{u}, \boldsymbol{v})))^D = \\ \exists y, \boldsymbol{X}, \boldsymbol{u} \forall \boldsymbol{v} (A_D(\boldsymbol{X}(\boldsymbol{v})) \to B_D(y, \boldsymbol{u}, \boldsymbol{v}))$$

MP').

$$(\neg \forall x_1 A \to \exists x_2 \neg A)^D = (\neg \forall x_1 A_D(x_1) \to \exists x_2 \neg A_D(x_2))^D = (\exists x_1 \neg A_D(x_1) \to \exists x_2 \neg A_D(x_2))^D$$

□

This theorem would not hold, if we were to consider IP and MP_+. If we look at the classical variants of these theories we see that in combination with theorem 1.3.6, a slight variation of theorem 2.3.7 also holds for the theory $\mathsf{WE\text{-}PA}^\omega \oplus \mathsf{AC}$, namely

$$\mathsf{WE\text{-}PA}^\omega \oplus \mathsf{AC} \vdash A \leftrightarrow (A^N)^D,$$

and its restricted variant. The principles IP' and MP' need not be added anymore, since they follow from classical logic. A problem though arises if we consider theorem 2.3.9. The difficulty is that AC is not in general N-interpretable. For the negative translation of $\forall x \exists y A(x, y) \to \exists Y \forall x A(x, Y(x))$ is

$$\forall x \neg \forall y \neg A^N(x, y) \to \neg \forall Y \neg \exists x A^N(x, Y(x))$$

or equivalently,

$$\forall x \neg\neg \exists y A^N(x, y) \to \neg\neg \exists Y \forall x A^N(x, Y(x))$$

which cannot in general be proved in $\mathsf{WE\text{-}PA}^\omega_i \oplus \mathsf{AC} \oplus \mathsf{IP}' \oplus \mathsf{MP}'$. But again we can prove a slight variation.

Definition 2.3.10. $\mathsf{QF\text{-}AC}$ is AC restricted to quantifier-free formulas A.

$$\mathsf{QF\text{-}AC} \qquad \forall \boldsymbol{x} \exists \boldsymbol{y} A(\boldsymbol{x}, \boldsymbol{y}) \to \exists \boldsymbol{Y} \forall \boldsymbol{x} A(\boldsymbol{x}, \boldsymbol{Y}(\boldsymbol{x}))$$

Theorem 2.3.11. *The theory* $\mathsf{WE\text{-}PA}^\omega \oplus \mathsf{QF\text{-}AC}$ *and its restricted variant are ND-interpreted in* T *and* $\widehat{\mathsf{T}}$ *respectively.*

Proof. First we need to show that $\mathsf{WE\text{-}PA}^\omega \oplus \mathsf{QF\text{-}AC}$ is N-interpreted in $\mathsf{WE\text{-}PA}^\omega_i \oplus \mathsf{AC} \oplus \mathsf{IP}' \oplus \mathsf{MP}'$. Considering theorem 1.3.6, this only requires verification of

$$\mathsf{WE\text{-}PA}^\omega_i \oplus \mathsf{AC} \oplus \mathsf{IP}' \oplus \mathsf{MP}' \vdash (\mathsf{QF\text{-}AC})^N.$$

From $\forall x \neg \forall y \neg A^N(x, y)$ we can prove $\forall x \exists y \neg\neg A^N(x, y)$ using MP'. Since A is quantifier-free we get $\forall x \exists y A^N(x, y)$. Then from AC we infer $\exists Y \forall x A^N(x, Y(x))$ which implies intuitionistically its double negation $\neg\neg \exists Y \forall x A^N(x, Y(x))$, which finally is equivalent to $\neg \forall Y \neg \forall x A^N(x, Y(x))$.

The rest follows from theorem 2.3.9 $\square$

But now we seem to have a small discrepancy between this theorem and the fact that $\mathsf{WE\text{-}PA}^\omega \oplus \mathsf{AC} \vdash A \leftrightarrow (A^N)^D$. Luckily we can actually get the following proper analogue to theorem 2.3.7.

Theorem 2.3.12. *For all formulas A,* $\mathsf{WE\text{-}PA}^\omega \oplus \mathsf{QF\text{-}AC} \vdash A \leftrightarrow (A^N)^D$

Proof. Obviously $\mathsf{WE\text{-}PA}^\omega \oplus \mathsf{QF\text{-}AC} \vdash A \leftrightarrow A^N$. So it remains to be shown that $\mathsf{WE\text{-}PA}^\omega \oplus \mathsf{QF\text{-}AC} \vdash A^N \leftrightarrow (A^N)^D$. But this is the same as $\mathsf{WE\text{-}PA}^\omega \oplus \mathsf{QF\text{-}AC} \vdash A \leftrightarrow A^D$ for formulas A not containing disjunctions and existential-quantifiers.

By induction on the complexity of A we show that A^D has the form $\exists \boldsymbol{x} \forall \boldsymbol{y} A^*(\boldsymbol{x}(\boldsymbol{y}), \boldsymbol{y})$ with $A^*(\boldsymbol{x}(\boldsymbol{y}), \boldsymbol{y}) = A_D(\boldsymbol{x}, \boldsymbol{y})$ and simultaneously that $\mathsf{WE\text{-}PA}^\omega \oplus \mathsf{QF\text{-}AC} \vdash A \leftrightarrow A^D$.

For A prime both assertions are obvious. For the induction step let $A^D = \exists \boldsymbol{x} \forall \boldsymbol{y} A^*(\boldsymbol{x}(\boldsymbol{y}), \boldsymbol{y})$ and $B^D = \exists \boldsymbol{u} \forall \boldsymbol{v} B^*(\boldsymbol{u}(\boldsymbol{v}), \boldsymbol{v})$.

$$\begin{aligned}(\forall z A)^D = \exists \boldsymbol{x} \forall z, \boldsymbol{y} A^*(\boldsymbol{x}(z, \boldsymbol{y}), z, \boldsymbol{y}) &\overset{\mathsf{QF\text{-}AC}}{\leftrightarrow} \forall z, \boldsymbol{y} \exists \boldsymbol{x} A^*(\boldsymbol{x}, z, \boldsymbol{y}) \leftrightarrow \\ &\overset{\mathsf{QF\text{-}AC}}{\leftrightarrow} \forall z \exists \boldsymbol{x} \forall \boldsymbol{y} A^*(\boldsymbol{x}(\boldsymbol{y}), z, \boldsymbol{y}) = \forall z A^D \overset{\text{i.h.}}{\leftrightarrow} \forall z A\end{aligned}$$

$$\begin{aligned}(A \wedge B)^D = \exists \boldsymbol{x}, \boldsymbol{u} \forall \boldsymbol{y}, \boldsymbol{v} (A^*(\boldsymbol{x}(\boldsymbol{y}), \boldsymbol{y}) \wedge B^*(\boldsymbol{u}(\boldsymbol{v}), \boldsymbol{v})) \leftrightarrow \\ \leftrightarrow \exists \boldsymbol{x} \forall \boldsymbol{y} A^*(\boldsymbol{x}(\boldsymbol{y}), \boldsymbol{y}) \wedge \exists \boldsymbol{u} \forall \boldsymbol{v} B^*(\boldsymbol{u}(\boldsymbol{v}), \boldsymbol{v}) = A^D \wedge B^D \overset{\text{i.h.}}{\leftrightarrow} A \wedge B\end{aligned}$$

$$\begin{aligned}&(A \to B)^D = \exists \boldsymbol{y}, \boldsymbol{u} \forall \boldsymbol{x}, \boldsymbol{v} \big[A^*(\boldsymbol{x}(\boldsymbol{y}(\boldsymbol{x}, \boldsymbol{v})), \boldsymbol{y}(\boldsymbol{x}, \boldsymbol{v})) \to B^*(\boldsymbol{u}(\boldsymbol{x}, \boldsymbol{v}), \boldsymbol{v})\big] \leftrightarrow \\ &\overset{\mathsf{QF\text{-}AC}}{\leftrightarrow} \forall \boldsymbol{x}, \boldsymbol{v} \exists \boldsymbol{y}, \boldsymbol{u} \big[A^*(\boldsymbol{x}(\boldsymbol{y}), \boldsymbol{y}) \to B^*(\boldsymbol{u}, \boldsymbol{v})\big] \leftrightarrow \forall \boldsymbol{x}, \boldsymbol{v} \big[\forall \boldsymbol{y} A^*(\boldsymbol{x}(\boldsymbol{y}), \boldsymbol{y}) \to \exists \boldsymbol{u} B^*(\boldsymbol{u}, \boldsymbol{v})\big] \leftrightarrow \\ &\big[\exists \boldsymbol{x} \forall \boldsymbol{y} A^*(\boldsymbol{x}(\boldsymbol{y}), \boldsymbol{y}) \to \forall \boldsymbol{v} \exists \boldsymbol{u} B^*(\boldsymbol{u}, \boldsymbol{v})\big] \overset{\mathsf{QF\text{-}AC}}{\leftrightarrow} \big[\exists \boldsymbol{x} \forall \boldsymbol{y} A^*(\boldsymbol{x}(\boldsymbol{y}), \boldsymbol{y}) \to \exists \boldsymbol{u} \forall \boldsymbol{v} B^*(\boldsymbol{u}(\boldsymbol{v}), \boldsymbol{v})\big] = \\ &\qquad = (A^D \to B^D) \overset{\text{i.h.}}{\leftrightarrow} (A \to B)\end{aligned}$$

□

This proof raises the question, whether theorem 2.3.7 would also work with $\mathsf{QF\text{-}AC}$ instead of AC. But if we were to extend the above proof by the case for the existential-quantifier we would face the following difficulty.

$$\exists z A \overset{\text{i.h.}}{\leftrightarrow} \exists z A^D = \exists z, \boldsymbol{x} \forall \boldsymbol{y} A^*(\boldsymbol{x}(\boldsymbol{y}), \boldsymbol{y}, z) \not\leftrightarrow \exists z, \boldsymbol{x} \forall \boldsymbol{y} A^*(\boldsymbol{x}(\boldsymbol{y}), \boldsymbol{y}, z(\boldsymbol{y})) = (\exists z A)^D$$

For convenience, we will use the following abbreviations:

$\mathsf{WE\text{-}PA}_i^{\omega+} := \mathsf{WE\text{-}PA}_i^\omega \oplus \mathsf{AC} \oplus \mathsf{IP}' \oplus \mathsf{MP}'$
$\mathsf{WE\text{-}PRA}_i^{\omega+} := \mathsf{WE\text{-}PRA}_i^\omega \oplus \mathsf{AC} \oplus \mathsf{IP}' \oplus \mathsf{MP}'$

Chapter 3

Conservation and subtheory results for WKL

In this chapter we now want to look at the proof-theoretic strength of certain theories. In order to do so we first introduce two notions: one theory being a *subtheory* of another; and one theory being *conservative* over another. We then look at $\mathsf{WE\text{-}PRA}^\omega \oplus \mathsf{QF\text{-}AC}$ and show that it contains PRA as a subtheory, and that it is conservative over PRA for a certain class of formulas. This case will in a way serve as the basis for the following extensions of the theory. Namely, we then want to see what happens if we add WKL or UWKL to this theory or the fully extensional version of this theory. We will see that for most of these cases the picture does not change although their proofs will require further definitions and lemmas. Only in the last case, when adding UWKL to $\mathsf{E\text{-}PRA}^\omega + \mathsf{QF\text{-}AC}^{1,0} + \mathsf{QF\text{-}AC}^{0,1}$, will we get stronger results.

All of these results are based upon Kohlenbach [13],[11], and Avigad, Feferman [1]. Elimination of extensionality is taken from Luckhardt [14], while non-interpretability of extensionality comes from Howard [7].

3.1 Preliminaries

3.1.1 Subtheories and conservation

A theory Th_1 is said to be a subtheory of a theory Th_2, $\mathsf{Th}_1 \subseteq \mathsf{Th}_2$, if there is a natural translation of formulas A of Th_1 to formulas A^{Th_2} of Th_2 such that $\mathsf{Th}_1 \vdash A$ implies $\mathsf{Th}_2 \vdash A^{\mathsf{Th}_2}$.

Theorem 3.1.1. $\mathsf{PA} \subseteq \mathsf{WE\text{-}PA}^\omega$ *and* $\mathsf{PRA} \subseteq \mathsf{WE\text{-}PRA}^\omega$.

Proof. Theorem 2.2.2 guarantees, that we can naturally translate terms of PA (PRA) to terms of $\mathsf{WE\text{-}PA}^\omega$ ($\mathsf{WE\text{-}PRA}^\omega$). The translation of formulas then is trivial. Since all logical axioms of PA (PRA) are also axioms of $\mathsf{WE\text{-}PA}^\omega$ ($\mathsf{WE\text{-}PRA}^\omega$), for the translation of proofs we only need to consider nonlogical axioms. For a list of the nonlogical axioms of PA (PRA) see [8]. Induction is unproblematic for either theory. The provability of the others is either guaranteed by theorem 2.2.2 or by the remark made in the proof of theorem 2.3.3. □

In this case, being a subtheory can also be looked at as being a lower bound. So PRA is a first order lower bound of $\mathsf{WE\text{-}PRA}^\omega$. It is actually the greatest lower bound, since any theory, which is stronger than PRA, as PA is for example, is not a subtheory of $\mathsf{WE\text{-}PRA}^\omega$ anymore.

Definition 3.1.2. Let $\mathsf{Th}_1 \subseteq \mathsf{Th}_2$ and let Γ be a subset of all formulas of Th_1. Th_2 is said to be conservative over Th_1 relative to Γ if for all $A \in \Gamma$

$$\mathsf{Th}_2 \vdash A^{\mathsf{Th}_2} \quad \Rightarrow \quad \mathsf{Th}_1 \vdash A$$

Theorem 3.1.3. $\mathsf{WE\text{-}PRA}^\omega$ *is conservative over* PRA *relative to all quantifier-free formulas.*

Proof. If

$$\mathsf{WE\text{-}PRA}^\omega \vdash A^{\mathsf{WE\text{-}PRA}^\omega}$$

for a quantifier-free formula A of PRA, theorem 2.2.5 assures that there is a term $t_{A^{\mathsf{WE\text{-}PRA}^\omega}}$ such that

$$\mathsf{WE\text{-}PRA}^\omega \vdash t_{A^{\mathsf{WE\text{-}PRA}^\omega}} = 0.$$

Using the fact, that $\mathsf{WE\text{-}PRA}^\omega$ is conservative over $\widehat{\mathsf{T}}$ for quantifier-free formulas and theorem 2.2.6 we obtain

$$\mathsf{PRA} \vdash t^{\mathsf{PRA}}_{A^{\mathsf{WE\text{-}PRA}^\omega}} = 0.$$

If we now combine the natural subtheory translation with theorem 2.1.2 we finally get

$$\mathsf{PRA} \vdash A.$$

□

When dealing with natural translations, the name of the theory will often be omitted, since it makes reading more difficult and should be clear from the context anyway. So instead of t^{Th} or A^{Th} we just write t and A.

If a subtheory can be looked at as a lower bound, then a theory over which another is conervative, can be looked at as an upper bound. In this case, PRA is a first order upper bound of $\mathsf{WE\text{-}PRA}^\omega$ with respect to quantifier-free formulas. It is actually the lowest upper bound, since the above stated theorem obviously also holds if we replace PRA by any theory which is stronger, for example PA.

3.1.2 Σ- and Π-formulas

Since conservations results are usually made about specific sets of formulas and in our case are stated with respect to first order theories, we need to have a more subtle classification of first order formulas than just quantifier-free ones and the rest. For that reason the so called Π^n_m- and Σ^n_m-formulas are defined.

For the following definition we always consider formulas of PA or PRA respectively. First we define the class of all Δ^0_0-formulas. Basically this is the class of all quantifier-free formulas. If our theory includes the relation $\leqslant$, we also allow bounded quantifiers. But since bounded quantification is primitive recursive it can be replaced by a term and so it does not really make a difference.

Definition 3.1.4. 1. A is a Σ^0_n-formula if it is of the form

$$\exists x_1 \forall x_2 \ldots Q_n x_n A^*(x_1, \ldots, x_n)$$

where $A^*(u_1, \ldots, u_n)$ is a Δ^0_0-formula and where Q is an all-quantifier if n is even and an existential-quantifier if n is odd.

2. A is a Π^0_n-formula if it is of the form

$$\forall x_1 \exists x_2 \ldots Q_n x_n A^*(x_1, \ldots, x_n)$$

where $A^*(u_1, \ldots, u_n)$ is a Δ^0_0-formula and where Q is an all-quantifier if n is odd and an existential-quantifier if n is even.

This definition can be generalized to formulas of PA^ωor PRA^ω respectively. We only need to introduce an additional class of formulas, the Δ^m_0-formulas. This class contains all formulas, in which all bound variables are of lower type level than m. In this case, the Δ^0_0-formulas would again be the quantifier-free ones, but with free variables of arbitrary types.

Definition 3.1.5. 3. A is a Σ^m_n-formula if it is of the form

$$\exists x_1 \forall x_2 \ldots Q_n x_n A^*(x_1, \ldots, x_n)$$

where the x_i's are of type m, $A^*(u_1, \ldots, u_n)$ is a Δ^m_0-formula and where Q is an all-quantifier if n is even and an existential-quantifier if n is odd.

4. A is a Π^m_n-formula if it is of the form

$$\forall x_1 \exists x_2 \ldots Q_n x_n A^*(x_1, \ldots, x_n)$$

where the x_i's are of type m, $A^*(u_1, \ldots, u_n)$ is a Δ^m_0-formula and where Q is an all-quantifier if n is odd and an existential-quantifier if n is even.

3.1.3 $\mathsf{WE\text{-}PRA}^\omega \oplus \mathsf{QF\text{-}AC}$

The theorem in this subsection is of interest for two reasons. The first point is that it uses almost all the means introduced and defined in the first two chapters and therefore shows that all things, even if at first sight they do not seem to be linked, all play together. The second reason is that it will be the foundation for all the conservation results in the next three sections.

Theorem 3.1.6. $\mathsf{WE\text{-}PRA}^\omega \oplus \mathsf{QF\text{-}AC}$ *is* Π^0_2*-conservative over* PRA

Proof. Assume

$$\mathsf{WE\text{-}PRA}^\omega \oplus \mathsf{QF\text{-}AC} \vdash \forall x \exists y A(x, y).$$

The first step is to apply negative translation to obtain $\mathsf{WE\text{-}PRA}^\omega_i \oplus \mathsf{QF\text{-}AC} \vdash \forall x \neg \forall y \neg A^N(x, y)$. We now extend the theory and get

$$\mathsf{WE\text{-}PRA}^{\omega+}_i \vdash \forall x \neg \forall y \neg A^N(x, y).$$

With the help of MP' together with the fact that A is quantifier-free we can eliminate the negations and have again a pure Π^0_2-formula.

$$\mathsf{WE\text{-}PRA}^{\omega+}_i \vdash \forall x \exists y A^N(x, y).$$

The next step is to apply the D-interpretation (theorem 2.3.9) which yields $\widehat{\mathsf{T}} \vdash A^N_D(x, t(x))$. Since A^N does not contain disjunctions nor quantifiers it is identical to A^N_D, the D can therefore be dropped. $\widehat{\mathsf{T}}$ is defined over classical logic, we therefore have $\widehat{\mathsf{T}} \vdash A \leftrightarrow A^N$. This leaves us with

$$\widehat{\mathsf{T}} \vdash A(x, t(x)).$$

Theorem 2.2.5 gives us a term t_A such that $\widehat{\mathsf{T}} \vdash t_A(x, t(x)) = 0$. All arguments of this term are of type 0 and so is its value. By theorem 2.2.6 we can therefore translate it into a term of PRA and get

$$\mathsf{PRA} \vdash t_A(x, t(x)) = 0$$

The last step is to reintroduce A to get $\mathsf{PRA} \vdash A(x, t(x))$ and then to introduce quantifiers to obtain the final result

$$\mathsf{PRA} \vdash \forall x \exists y A(x, y).$$

□

3.1.4 Weak König's lemma WKL and uniform weak König's lemma

WKL asserts that every infinite binary tree has an infinite path. UWKL goes one step further in saying that there is a functional which selects from every infinite binary tree an infinite path. In order to express these two principles we must introduce some new definitions and notions.

WE-PRA$^\omega$, to which we want to add WKL, does not include sets nor the $\in$-relation. In order to be able to talk about sets of natural numbers, we introduce characteristic functions. So whenever we write $x \in M$ or $x \notin M$ we actually mean $\mathrm{ch}_M(x) = 0$ or $\mathrm{ch}_M(x) = 1$. And if we say that a set is primitive recursive we mean that its characteristic function is so. For our next definition, we recall that finite sequences of natural numbers can primitive recursively be coded as natural numbers. This means that the function $\langle * \rangle : \mathbb{N}^k \to \mathbb{N}$ which codes sequences of length k, the function $lh(s) : \mathbb{N} \to \mathbb{N}$ which returns the length of the sequence coded by s, the function $\pi(s, i) : \mathbb{N}^2 \to \mathbb{N}$ which returns the ith element of the sequence coded by s, the function $* : \mathbb{N}^2 \to \mathbb{N}$, which concatenates two sequences, the unary predicate $Seq(s)$ which asserts that s is the code of a sequence and the binary predicate $t \subseteq s$ which asserts that t is an initial segment of s are all primitive recursive. For reference see [8] or [15].

Definition 3.1.7. 1. Let $\{0,1\}^k$ (resp. $\{0,1\}^{<\omega}$) denote the set of length-k (resp. finite) binary sequences. These sets are primitive recursive, as can be seen from the following:

$$s \in \{0,1\}^k \quad \Leftrightarrow \quad \mathrm{Seq}(s) \wedge lh(s) = k \wedge \forall i \leqslant (k \dot{-} 1)(\pi(s,i) \leqslant 1)$$

2. If b is a function (type $0 \to 0$), let $\bar{b}$ denote the initial segment function

$$\bar{b}(x) = \langle b(0), b(1), \ldots, b(x-1) \rangle.$$

3. We define the unary predicate BinFunc on functions b which asserts that b is a binary function

$$\mathrm{BinFunc}(b^{0\to 0}) \equiv \forall x (b(x) = 0 \vee b(x) = 1).$$

4. Let BinTree be the unary predicate on sets of natural numbers f, which asserts that f is a binary tree, that is, a set of binary sequences (coded as natural numbers) closed under initial segments;

$$\mathrm{BinTree}(f^{0\to 0}) \equiv \forall s \in f (s \in \{0,1\}^{<\omega} \wedge \forall t \subseteq s (t \in f)).$$

5. Last we define the binary predicate Bounded on binary trees f and natural numbers k, which asserts that the hight of the binary tree f is less than or equal to k;

$$\mathrm{Bounded}(f, k) \equiv \forall s \in \{0,1\}^k (s \notin f) \wedge \mathrm{BinTree}(f).$$

With the help of these definitions we can now write out weak König's lemma WKL and uniform weak König's lemma UWKL.

WKL $\quad \forall f(\mathrm{BinTree}(f) \wedge \forall k \neg \mathrm{Bounded}(f,k) \rightarrow \exists b(\mathrm{BinFunc}(b) \wedge \forall k\, \bar{b}(k) \in f))$

UWKL $\quad \exists \Phi \forall f(\mathrm{BinTree}(f) \wedge \forall k \neg \mathrm{Bounded}(f,k) \rightarrow \forall k(\overline{\Phi(f)}(k) \in f))$

Note that Φ is of type $(0 \rightarrow 0) \rightarrow (0 \rightarrow 0)$.

These are quite nice and intuitive formulations of WKL and UWKL. But there are two minor flaws. The first one is that it contains redundant parts. The predicate BinFunc is actually not needed. Since if there is a function b such that all its initial segments are in f it must automatically be a binary function. The second flaw is that the predicate BinTree requires a universal quantifier and is therefore not primitive recursive and trickier to work with. This can be avoided by applying the following trick. Given a function f we define the functional tree of type $(0 \rightarrow 0) \rightarrow (0 \rightarrow 0)$ so that for any s

$$s \in \mathrm{tree}(f) \leftrightarrow (s \in \{0,1\}^{<\omega} \wedge \forall t \subseteq s(t \in f)).$$

$\mathrm{tree}(f)$ or f^{tree} then becomes a binary tree and we can prove in WE-PRA$^\omega$

$$\mathrm{BinTree}(f^{\mathrm{tree}}) \wedge (\mathrm{BinTree}(f) \rightarrow f =_1 f^{\mathrm{tree}}).$$

Having fixed these minor flaws we can formulate WKL and UWKL by

WKL $\quad \forall f(\forall k \neg \mathrm{Bounded}(f^{\mathrm{tree}},k) \rightarrow \exists b \forall k\, \bar{b}(k) \in f^{\mathrm{tree}})$

UWKL $\quad \exists \Phi \forall f(\forall k \neg \mathrm{Bounded}(f^{\mathrm{tree}},k) \rightarrow \forall k(\overline{\Phi(f)}(k) \in f^{\mathrm{tree}}))$

Obviously, we also drop the predicate BinTree within the definition of Bounded.

3.2 The theory WE-PRA$^\omega$ $\oplus$ QF-AC $\oplus$ WKL

We have shown that PRA is a lower and an upper bound for WE-PRA$^\omega \oplus$QF-AC. Obviously it is still a lower bound if we add WKL. Actually it is still the greatest lower bound, but this will not be proved explicitly, here. What is not so obvious, is that it still is an upper bound, with respect to Π^0_2-formulas. The only thing we actually need to do is to eliminate WKL from the proof of an arbitrary Π^0_2-formula.

Even if we add UWKL instead of WKL the results stay the same, this though will be shown in the next section.

3.2.1 Hereditary majorizability

With primitive recursion we have a tool, which allows us to introduce the well known relation $\leqslant$ for type 0 terms via for example cutoff-subtraction

$$s \leqslant t :\equiv s \dot{-} t = 0$$

It has already been mentioned that bounded quantification is primitive recursive and that formulas which contain bounded quantifiers instead of unbounded ones are therefore easier to work with. If working in a theory which includes higher type variables, it is interesting to ask whether one can generalize $\leqslant$ to arbitrary types, such that properties of this generalized relation can be used to prove other things or to simplify conjectures. In [7] Howard has done exactly that. He introduced the notion of *hereditary majorizability* to prove that the extensionality axiom (E) is not D-interpretable. For now we want to use it to eliminate WKL from a proof of a Π^0_2-formula.

Definition 3.2.1. For each type σ, we define a relation $a \leqslant_* b$ for terms a and b of type σ as follows.

1. If $\sigma = 0$, $a \leqslant_* b$ is just $a \leqslant b$.
2. If $\sigma = (\tau \to \rho)$, then $a \leqslant_* b$ if and only if

$$\forall x, y(x \leqslant_* y \to a(x) \leqslant_* b(y))$$

   where the variables x and y are of type τ. When working in $\widehat{\mathsf{T}}$ we drop the quantifiers to adopt the definition.

The relation $a \leqslant_* b$ is read "a is hereditarily majorized by b."

We now want to show that all type 1 terms of $\widehat{\mathsf{T}}$ are majorizable and that all closed terms of any finite type of $\widehat{\mathsf{T}}$ are majorizable. The intuitive assumption, that $f \leqslant_* f$ is false, as can easily be verified when considering for example $f = \lambda x.(10 \dot{-} x)$.

Theorem 3.2.2. *1. Given a type 1 term f of $\widehat{\mathsf{T}}$, define*

$$f^*(x) := \max_{y \leqslant x} f(y).$$

Then $f \leqslant_* f^*$.

2. For every closed term F in $\widehat{\mathsf{T}}$ there is another closed term F^* such that $\widehat{\mathsf{T}}$ proves $F \leqslant_* F^*$.

Proof.

1. Follows directly from the definition of f^*.

2. By induction on the complexity of F. Since the type of F can always be written as $\tau_1 \to \ldots \to \tau_n \to 0$ we need to construct a term F^*, such that $\widehat{\mathsf{T}}$ proves

$$\boldsymbol{x}^{\underline{\tau}} \leqslant_* \boldsymbol{y}^{\underline{\tau}} \to F(\boldsymbol{x}) \leqslant_* F^*(\boldsymbol{y}).$$

   Which is equivalent to $F \leqslant_* F^*$.

   For the basis consider the five cases 0, Sc, K, S and $\widehat{\mathrm{R}}$. Obviously we can take 0^* and Sc^* to be 0 and Sc respectively. For K^* and S^* we can take K and S respectively, as can be easily verified.

   Consider K. $\widehat{\mathsf{T}}$ obviously proves

$$x_1 \leqslant_* y_1 \wedge \ldots \wedge x_n \leqslant_* y_n \to x_1(x_3, \ldots, x_n) \leqslant_* y_1(y_3, \ldots, y_n).$$

   Here, the right hand side is just $\mathrm{K}(x_1, x_2)(x_3, \ldots, x_n) \leqslant_* \mathrm{K}^*(y_1, y_2)(y_3, \ldots, y_n)$, which is $\mathrm{K}(\boldsymbol{x}) \leqslant_* \mathrm{K}^*(\boldsymbol{y})$.

   For S this works completely analogous. The only critical case is $\widehat{\mathrm{R}}$. In this case we first show that $\widehat{\mathsf{T}}$ proves

$$x_1 \leqslant_* y_1 \wedge x_2 \leqslant_* y_2 \wedge \boldsymbol{b} \leqslant_* \boldsymbol{c} \to \widehat{\mathrm{R}}(x_1, y_1, n, \boldsymbol{b}) \leqslant_* \widehat{\mathrm{R}}(x_2, y_2, n, \boldsymbol{c}).$$

   This can easily be shown by induction on n. We now take $\widehat{\mathrm{R}}^*$ to be

$$\widehat{\mathrm{R}}^* := \lambda x_1 y_1 n \boldsymbol{b}. \sum_{m \leqslant n} \widehat{\mathrm{R}}(x_1, y_1, m, \boldsymbol{b}).$$

   This is unproblematic, since bounded sums are primitive recursive and can therefore be replaced by terms of $\widehat{\mathsf{T}}$.

   For the induction step assume $F = G(H)$, $G \leqslant_* G^*$ and $H \leqslant_* H^*$. Clearly we can just take F^* to be $G^*(H^*)$.

□

The following technical lemma will be needed for the proof of theorem 3.2.4

Theorem 3.2.3. *1.* $\widehat{\mathsf{T}} \vdash \forall f(\mathrm{BinFunc}(f) \leftrightarrow f \leqslant_* \lambda x.1)$.

2. If the term B is of type $1 \to 1$, then $\widehat{\mathsf{T}}$ proves

$$\forall f \mathrm{BinFunc}(B(f)) \leftrightarrow B \leqslant_* \lambda f x.1.$$

Proof.

1. This follows from the fact, that $f \leqslant_* \lambda x.1$ actually stands for $\forall x, y(f(x) \leqslant_* \lambda x.1(y))$, which is $\forall x(f(x) \leqslant_* 1)$.

2. $B \leqslant_* \lambda f x.1$ is equivalent to

$$\forall f, g(f \leqslant_* g \to B(f) \leqslant_* \lambda x.1).$$

   By using 1. we obtain

$$\forall f, g(f \leqslant_* g \to \mathrm{BinFunc}(B(f))).$$

   Taking $g = f^*$ from the above theorem yields

$$\forall f(f \leqslant_* f^* \to \mathrm{BinFunc}(B(f))),$$

   where the antecedent is universally true and can therefore be eliminated. The other direction is trivial.

□

3.2.2 Conservation result

Theorem 3.2.4. $\mathsf{WE\text{-}PRA}^\omega \oplus \mathsf{QF\text{-}AC} \oplus \mathsf{WKL}$ *is* Π^0_2*-conservative over* PRA.

The proof of this theorem will be carried out in three steps. The first one is to reduce the proof to $\mathsf{WE\text{-}PRA}^{\omega+}_i \oplus \mathsf{WKL}$ via negative translation. The next step is to apply the deduction theorem to WKL and then to weaken it with the help of the D-interpretation. In the last step we will show, that this weak form of WKL can now be proved in $\mathsf{WE\text{-}PRA}^{\omega+}_i$ using the notion of hereditary majorizability. We have then reduced the proof to the proof of theorem 3.1.6. This 3 step programm requires the following 3 lemmas.

Lemma 3.2.5. $\mathsf{WE\text{-}PRA}^\omega \oplus \mathsf{QF\text{-}AC} \oplus \mathsf{WKL}$ *is N-interpreted in* $\mathsf{WE\text{-}PRA}^{\omega+}_i \oplus \mathsf{WKL}$.

Proof. We only need to verify that the claim holds for WKL.

$$(\mathsf{WKL})^N = \forall f(\forall k \neg \mathrm{Bounded}(f^{\mathrm{tree}}, k) \to \neg \forall b \neg \forall k\, \bar{b}(k) \in f^{\mathrm{tree}})$$

But since the two principles $\neg \forall x \neg A \leftrightarrow \neg\neg \exists x A$ and $A \to \neg\neg A$ are intuitionistically valid, $(\mathsf{WKL})^N$ follows from WKL. □

We now define a more convenient variant of WKL as follows:

$$\mathsf{WKL}' \qquad \forall f \exists b \forall k (\neg \mathrm{Bounded}(f^{\mathrm{tree}}, k) \to \bar{b}(k) \in f^{\mathrm{tree}}).$$

The idea is basically to take all quantifiers to the front so that we can later make use of lemma 3.2.7.

Lemma 3.2.6. $\mathsf{WKL}' \to \mathsf{WKL}$ *can be proved intuitionistically.*

Proof. The proof basically reduces to showing that

$$\exists x \forall y [A(y) \to B(x,y)] \to [\forall y A(y) \to \exists x \forall y B(x,y)]$$

can be proved using only the rules and axioms of intuitionistic many-sorted predicate logic. But this can be shown rather easy if we consider a natural deduction system as defined for example in [15]. □

Lemma 3.2.7. *Suppose*

$$\mathsf{WE\text{-}PRA}_i^{\omega+} \vdash \exists a \forall b, c B(a,b,c) \to \forall x \exists y A(x,y),$$

with A and B quantifier-free. Then there is a specific term $\tilde{c}(a,x)$ such that

$$\mathsf{WE\text{-}PRA}_i^{\omega+} \vdash \forall x \exists a \forall b B(a,b,\tilde{c}(a,x)) \to \forall x \exists y A(x,y).$$

Proof. We apply the D-interpretation with $\forall x$ deleted and obtain

$$\widehat{\mathsf{T}} \vdash B_D(a, \tilde{b}(a,x), \tilde{c}(a,x)) \to A_D(x, \tilde{y}(a)).$$

Since $\widehat{\mathsf{T}}$ is a subtheory of $\mathsf{WE\text{-}PRA}_i^{\omega+}$, we can also prove it in the latter theory. Additionally we introduce quantifiers, so that

$$\mathsf{WE\text{-}PRA}_i^{\omega+} \vdash \exists B, Y \forall a [B_D(a, B(a,x), \tilde{c}(a,x)) \to A_D(x, Y(a))].$$

Using the fact that $A \leftrightarrow A^D$ is provable over this theory we get

$$\mathsf{WE\text{-}PRA}_i^{\omega+} \vdash \exists a \forall b B(a,b,\tilde{c}(a,x)) \to \exists y A(x,y).$$

By binding and distributing x we get the final result. □

We now have enough tools at hand to proof theorem 3.2.4

Proof. [3.2.4] Assume

$$\mathsf{WE\text{-}PRA}^{\omega} \oplus \mathsf{QF\text{-}AC} \oplus \mathsf{WKL} \vdash \forall x \exists y A(x,y).$$

By applying lemma 3.2.5 and using MP' we obtain

$$\mathsf{WE\text{-}PRA}_i^{\omega+} \oplus \mathsf{WKL} \vdash \forall x \exists y A^N(x,y).$$

The next step is to combine the deduction theorem with lemma 3.2.6;

$$\mathsf{WE\text{-}PRA}_i^{\omega+} \vdash \mathsf{WKL}' \to \forall x \exists y A^N(x,y),$$

which is

$$\mathsf{WE\text{-}PRA}_i^{\omega+} \vdash \forall f \exists b \forall k (\neg \mathrm{Bounded}(f^{\mathrm{tree}}, k) \to \bar{b}(k) \in f^{\mathrm{tree}}) \to \forall x \exists y A^N(x,y).$$

Here we apply AC to the hypothesis. Actually we do not really need AC since this strengthening of the hypothesis can be proved without it;

$$\mathsf{WE\text{-}PRA}_i^{\omega+} \vdash \exists B \forall f, k (\neg \mathrm{Bounded}(f^{\mathrm{tree}}, k) \to \overline{B(f)}(k) \in f^{\mathrm{tree}}) \to \forall x \exists y A^N(x,y)$$

and finally, with lemma 3.2.7;

$$\mathsf{WE\text{-}PRA}_i^{\omega+} \vdash \exists B \forall f (\neg \mathrm{Bounded}(f^{\mathrm{tree}}, \tilde{k}(B,x)) \to \overline{B(f)}(\tilde{k}(B,x)) \in f^{\mathrm{tree}}) \to \forall x \exists y A^N(x,y).$$

We are now reduced to showing that $\mathsf{WE\text{-}PRA}_i^{\omega+}$ proves the antecedent of this implication. This is done by bringing in the notion of hereditary majorizability to bound the value of $\tilde{k}(B,x)$. By theorem 3.2.2 we find a term k^* such that $\tilde{k} \leqslant_* k^*$. By theorem 3.2.3 we get

$$\mathsf{WE\text{-}PRA}_i^{\omega+} \vdash \forall f \mathrm{BinFunc}(B(f)) \to (B \leqslant_* \lambda f x.1 \wedge x \leqslant_* x).$$

Which by using the definition of hereditary majorizability results in

$$\text{I)} \qquad \mathsf{WE\text{-}PRA}_i^{\omega+} \vdash \forall f \mathrm{BinFunc}(B(f)) \to \tilde{k}(B,x) \leqslant_* \overbrace{k^*(\lambda f x.1, x)}^{k_0(x)}.$$

We now need to find an appropriate B, that can finally witness the existential-quantifier.

Define:

$$\begin{aligned} l &:= \max_{\alpha \leqslant k_0(x)} \exists s \leqslant \mathrm{AllOne}(\alpha)(lh(s) = \alpha \wedge s \in f^{\mathrm{tree}}) \\ s_0 &:= \min_{\alpha \leqslant \mathrm{AllOne(l)}} (lh(\alpha) = l \wedge \alpha \in f^{\mathrm{tree}}) \end{aligned}$$

Where $\mathrm{AllOne}(x)$ is the code of the sequence which consists of ones only and has length x. All parts of these definitions are primitive recursive and can therefore be represented by terms of $\mathsf{WE\text{-}PRA}_i^{\omega+}$. We now define the following binary function

$$B'(x,f)(y) := \begin{cases} \pi(s_0, y) & \text{if } y < l \\ 0 & \text{otherwise.} \end{cases}$$

$B'(x, f)$ therefore consists of s_0 followed by zeros. We could also choose the following definition

$$B'(x, f)(y) := \pi(s_0, y) \cdot \mathrm{sgn}(lh(s_0) + 1 \dot{-} y).$$

With these definitions we get, that $\mathsf{WE\text{-}PRA}_i^{\omega+}$ proves

$$\text{II)} \qquad l \leqslant k_0(x) \wedge \neg\mathrm{Bounded}(f^{\mathrm{tree}}, l) \to \overline{B'(x, f)}(l) \in f^{\mathrm{tree}}.$$

With the definition $B(x) := \lambda f.B'(x, f)$, we get that $\mathsf{WE\text{-}PRA}_i^{\omega+}$ proves

$$\forall f \mathrm{BinFunc}(B(x, f))$$

and so, using I),

$$\forall x(\tilde{k}(B(x), x) \leqslant k_0(x)).$$

If we now combine this with II), we then have

$$\forall f(\neg\mathrm{Bounded}(f^{\mathrm{tree}}, \tilde{k}(B(x), x)) \to \overline{B(x, f)}(\tilde{k}(B(x), x)) \in f^{\mathrm{tree}}).$$

Or by introducing the existential-quantifier

$$\mathsf{WE\text{-}PRA}_i^{\omega+} \vdash \exists B \forall f(\neg\mathrm{Bounded}(f^{\mathrm{tree}}, \tilde{k}(B, x)) \to \overline{B(f)}(\tilde{k}(B, x)) \in f^{\mathrm{tree}}).$$

We are finally left with

$$\mathsf{WE\text{-}PRA}_i^{\omega+} \vdash \forall x \exists y A^N(x, y).$$

The rest of the proof reduces therefore to the proof of theorem 3.1.6. □

3.3 The theory $\mathsf{WE\text{-}PRA}^\omega \oplus \mathsf{QF\text{-}AC} \oplus \mathsf{UWKL}$

Theorem 3.3.1. $\mathsf{WE\text{-}PRA}^\omega \oplus \mathsf{QF\text{-}AC} \oplus \mathsf{UWKL}$ *is* Π_2^0*-conservative over* PRA*.*

Proof. Eliminating UWKL and therefore reducing the proof to theorem 3.2.4 is quite simple. The only two things we have to show, are that UWKL is N-interpretable and that UWKL follows from WKL′ in $\mathsf{WE\text{-}PRA}_i^{\omega+}$.

The first step is easy since $(\mathsf{UWKL})^N = \neg\neg\mathsf{UWKL}$ and therefore trivially follows from UWKL in $\mathsf{WE\text{-}PRA}_i^{\omega+}$.

The second step is almost as easy if we look at WKL′

$$\mathsf{WKL}' \qquad \forall f \exists b \forall k(\neg\mathrm{Bounded}(f^{\mathrm{tree}}, k) \to \bar{b}(k) \in f^{\mathrm{tree}})$$

If we now apply AC we get

$$\exists B \forall f, k(\neg\mathrm{Bounded}(f^{\mathrm{tree}}, k) \to \overline{B(f)}(k) \in f^{\mathrm{tree}})$$

This obviously implies UWKL.

So if we have a proof of a Π^0_2-formula in $\mathsf{WE\text{-}PRA}^\omega \oplus \mathsf{QF\text{-}AC} \oplus \mathsf{UWKL}$ we first apply the N-interpretation then the deduction theorem and finally replace UWKL in the antecedent by WKL'. The rest follows from the proof of theorem 3.2.4. □

3.4 The theory $\mathsf{E\text{-}PRA}^\omega + \mathsf{QF\text{-}AC}^{1,0} + \mathsf{QF\text{-}AC}^{0,1} + \mathsf{WKL}$

The next question we ask, is what happens now if we add WKL to the fully extensional theory $\mathsf{E\text{-}PRA}^\omega + \mathsf{QF\text{-}AC}^{1,0} + \mathsf{QF\text{-}AC}^{0,1}$. Again PRA is still a lower bound. It is also still an upper bound. The proof of the latter though requires now a bit more work.

3.4.1 Non-interpretability of extensionality

If we look at the extensionality axiom

$$(\mathsf{E}) \qquad \forall x, y, z(x = y \to z(x) = z(y))$$

We readily see that it is N-interpretable. So it seems that for the conservation result, the same path as above could be chosen. But, as has been mentioned before and shall be shown now, this will fail, since (E) is not D-interpretable.

The proof will make use of the following lemma.

Lemma 3.4.1. *Suppose a closed term t of $\widehat{\mathsf{T}}$ of type $\sigma_1 \to \ldots \to \sigma_n \to 0$. Let k be a fixed numeral and let M^k_σ denote the following set of functionals*

$$x^\sigma \in M^k_\sigma \quad \Leftrightarrow \quad \forall \boldsymbol{z}(x(\boldsymbol{z}) \leqslant k).$$

Then

$$\mathsf{WE\text{-}PRA}^{\omega+}_i \vdash \exists m(\forall x_1 \in M^k_{\sigma_1}) \ldots (\forall x_n \in M^k_{\sigma_n})(t(x_1, \ldots, x_n) \leqslant m).$$

Proof. By theorem 3.2.2.2 there is a term t^* such that $t \leqslant_* t^*$. Now let $g^*_\sigma = \lambda \boldsymbol{z}.k$ (choose $\boldsymbol{z}$ so that g^*_σ is of type σ). By definition of hereditary majorizability we get

$$\mathsf{WE\text{-}PRA}^{\omega+}_i \vdash (\forall x \in M^k_\sigma)(x \leqslant_* g^*_\sigma).$$

Hence

$$\mathsf{WE\text{-}PRA}^{\omega+}_i \vdash (\forall x_1 \in M^k_{\sigma_1}) \ldots (\forall x_n \in M^k_{\sigma_n})(t(x_1, \ldots, x_n) \leqslant_* t^*(g^*_{\sigma_1}, \ldots, g^*_{\sigma_n})).$$

Thus $t^*(g^*_{\sigma_1}, \ldots, g^*_{\sigma_n})$ is the required numeral m. □

Theorem 3.4.2. *For the simplest non-trivial case of* (E)*,*

$$(\mathsf{E})_2 = \forall x^1, y^1, z^2(\forall u^0[x(u) = y(u)] \to z(x) = z(y)),$$

with

$$(\mathsf{E})_2^D = \exists U \forall x^1, y^1, z^2(x(U(x,y,z)) = y(U(x,y,z)) \to z(x) = z(y))$$

already, there is no closed term t such that

$$\widehat{\mathsf{T}} \vdash x(t(x,y,z)) = y(t(x,y,z)) \to z(x) = z(y)$$

or equivalently

$$\widehat{\mathsf{T}} \vdash z(x) \neq z(y) \to x(t(x,y,z)) \neq y(t(x,y,z)).$$

Proof. Assume that there is such a term t, so that we have

$$\mathsf{WE\text{-}PRA}_i^{\omega+} \vdash z(x) \neq z(y) \to x(t(x,y,z)) \neq y(t(x,y,z)).$$

It is easy to define primitive recursive functionals $\lambda x.z_n$ and $\lambda u.x_n$ such that

$$z_n(x) = \begin{cases} 1 & \text{if } (\forall u < n)(x(u) = 0) \wedge x(n) = 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad x_n(u) = \begin{cases} 0 & \text{if } u < n \\ 1 & \text{if } u \geqslant n \end{cases}$$

Using these functionals we get

$$\mathsf{WE\text{-}PRA}_i^{\omega+} \vdash z_n(x_n) \neq z_n(\lambda u.0) \to x_n(t(x_n, \lambda u.0, z_n)) \neq \lambda u.0(t(x_n, \lambda u.0, z_n)),$$

which reduces to

$$\mathsf{WE\text{-}PRA}_i^{\omega+} \vdash z_n(x_n) \neq 0 \to x_n(t(x_n, \lambda u.0, z_n)) \neq 0.$$

But since $z_n(x_n) = 1$ we obtain $x_n(t(x_n, \lambda u.0, z_n)) \neq 0$. Looking at the definition of x_n, we see that therefore $t(x_n, \lambda u.0, z_n) \geqslant n$. This clearly contradicts the above mentioned lemma, since n is arbitrary, $\lambda u.0, x_n \in M^1_{(1)}$, $z_n \in M^1_{(2)}$ and

$$\mathsf{WE\text{-}PRA}_i^{\omega+} \vdash \exists m(\forall x \in M^1_{(1)})(\forall y \in M^1_{(1)})(\forall z \in M^1_{(2)})(t(x,y,z) \leqslant m).$$

Whereas the assumption, that $(\mathsf{E})_2$ is D-interpretable, leads to

$$\mathsf{WE\text{-}PRA}_i^{\omega+} \vdash \forall m(\exists x \in M^1_{(1)})(\exists y \in M^1_{(1)})(\exists z \in M^1_{(2)})(t(x,y,z) \geqslant m).$$

$(\mathsf{E})_2$ Is therefore not D-interpretable. $\square$

3.4.2 Elimination of extensionality

Since we know now that (E) is not D-interpretable, we must find another way to get rid of it. The method best suited for this is the method introduced by Luckhardt in [14].

Definition 3.4.3. For all types σ, we define the functionals E_σ and the relation $\equiv_\sigma$ inductively as follows

1. $E_0(n) \leftrightarrow (n = n)$ $\qquad (n \equiv_0 m) \leftrightarrow (n = m)$
2. $E_\tau(f) \leftrightarrow \forall \boldsymbol{x}^{\underline{\sigma}}, \boldsymbol{y}^{\underline{\sigma}}(\boldsymbol{x} \equiv_{\underline{\sigma}} \boldsymbol{y} \to f(\boldsymbol{x}) =_0 f(\boldsymbol{y}))$
   $f \equiv_\tau g \leftrightarrow \forall \boldsymbol{z}^{\underline{\sigma}}(E_{\underline{\sigma}}(\boldsymbol{z}) \to f(\boldsymbol{z}) =_0 g(\boldsymbol{z})) \wedge E_\tau(f) \wedge E_\tau(g)$

We use the intuitive abbreviations $E_{\underline{\sigma}}(\boldsymbol{u})$ for $E_{\sigma_1}(u_1) \wedge \ldots \wedge E_{\sigma_n}(u_n)$ and $\boldsymbol{u} \equiv_{\underline{\sigma}} \boldsymbol{v}$ for $u_1 \equiv_{\sigma_1} v_1 \wedge \ldots \wedge u_n \equiv_{\sigma_n} v_n$. Type subscripts are dropped, if no ambiguities arise.

Remark 3.4.4. Considering these functionals, we get the following direct consequences of the above definitions.

1. $\forall x^0 E_0(x)$ obviously holds. Using the second axiom of equality we can show that $\forall a, b(a \equiv_0 b \to f(a) =_0 f(b))$, which is equivalent to $\forall f^1 E_1(f)$, also holds for any f.
2. $\equiv_\sigma$ is symmetric and transitive.
3. $r \equiv s \leftrightarrow \forall \boldsymbol{x}, \boldsymbol{y}(\boldsymbol{x} \equiv \boldsymbol{y} \to r(\boldsymbol{x}) =_0 s(\boldsymbol{y}))$
4. $r \equiv r \leftrightarrow E(r)$
5. $E(x) \wedge E(y) \to (x = y \leftrightarrow x \equiv y)$

Definition 3.4.5. For all formulas A of many-sorted predicate logic, we define the formula A^E inductively as follows:

1. $A^E :\equiv A$ for prime formulas A.
2. $(A \circ B)^E :\equiv A^E \circ B^E$ for $\circ \in \{\wedge, \vee, \to\}$.
3. $(\forall x^\sigma A(x))^E :\equiv \forall x(E_\sigma(x) \to A^E(x))$.
4. $(\exists x^\sigma A(x))^E :\equiv \exists x(E_\sigma(x) \wedge A^E(x))$.

A direct consequence of this in combination with 1. of the above remark, is that if A only contains bound variables of type 0 or 1, $A \leftrightarrow A^E$ holds.

For the main theorem we need the following lemma.

Lemma 3.4.6. *The following claims all hold in* $\mathsf{WE\text{-}PRA}^\omega$.

1. $E(\mathrm{K})$, $E(\mathrm{S})$, $E(\widehat{\mathrm{R}})$ *for all types.*
2. $E(\boldsymbol{u}) \to E(t)$, *where the sequence* $\boldsymbol{u}$ *contains the variables of* t.
3. $E(t)$ *for all closed terms* t

Proof.

1. Take $E(\mathrm{K})$:
$$E(\mathrm{K}) \leftrightarrow \forall \boldsymbol{x}, \boldsymbol{y}(\boldsymbol{x} \equiv \boldsymbol{y} \to \mathrm{K}(\boldsymbol{x}) =_0 \mathrm{K}(\boldsymbol{y})).$$
From remark 3.4.4.3 we get $u_1 \equiv v_1 \to \forall \boldsymbol{x}, \boldsymbol{y}(\boldsymbol{x} \equiv \boldsymbol{y} \to u_1(\boldsymbol{x}) =_0 v_1(\boldsymbol{y}))$ which can also be read as
$$u_1 \equiv v_1 \to \forall \boldsymbol{x}, \boldsymbol{y}(\boldsymbol{x} \equiv \boldsymbol{y} \to \mathrm{K}(u_1, u_2)(\boldsymbol{x}) =_0 \mathrm{K}(v_1, v_2)(\boldsymbol{y})),$$
which implies
$$u_1 \equiv v_1 \wedge u_2 \equiv v_2 \wedge \boldsymbol{x} \equiv \boldsymbol{y} \to \mathrm{K}(u_1, u_2)(\boldsymbol{x}) =_0 \mathrm{K}(v_1, v_2)(\boldsymbol{y})).$$
This obviously implies $E(\mathrm{K})$. For S and $\widehat{\mathrm{R}}$ this works similarly.

2. By induction on the complexity of t. For the basis consider the following two cases. t is either a functional constant or $t = u_i$, both of which are trivial. For the induction step let $t[\boldsymbol{u}] = r[\boldsymbol{u}](s[\boldsymbol{u}])$. By induction hypothesis
$$E(\boldsymbol{u}) \to E(r[\boldsymbol{u}]) \text{ and } E(\boldsymbol{u}) \to E(s[\boldsymbol{u}])$$
By definition, the first one implies
$$E(\boldsymbol{u}) \to \forall \boldsymbol{x}, \boldsymbol{y}(\boldsymbol{x} \equiv \boldsymbol{y} \to r[\boldsymbol{u}](\boldsymbol{x}) =_0 r[\boldsymbol{u}](\boldsymbol{y})),$$
or
$$E(\boldsymbol{u}) \to \forall \boldsymbol{x}', \boldsymbol{y}'(s[\boldsymbol{u}] \equiv s[\boldsymbol{u}] \wedge \boldsymbol{x}' \equiv \boldsymbol{y}' \to r[\boldsymbol{u}](s[\boldsymbol{u}], \boldsymbol{x}') =_0 r[\boldsymbol{u}](s[\boldsymbol{u}], \boldsymbol{y}')).$$
Using remark 3.4.4.4, the second one implies $E(\boldsymbol{u}) \to s[\boldsymbol{u}] \equiv s[\boldsymbol{u}]$. When combining the two, they yield
$$E(\boldsymbol{u}) \to \forall \boldsymbol{x}', \boldsymbol{y}'(\boldsymbol{x}' \equiv \boldsymbol{y}' \to r[\boldsymbol{u}](s[\boldsymbol{u}], \boldsymbol{x}') =_0 r[\boldsymbol{u}](s[\boldsymbol{u}], \boldsymbol{y}')),$$
which is $E(r[\boldsymbol{u}](s[\boldsymbol{u}]))$.

3. Direct consequence from 2. □

Theorem 3.4.7. *Let A be a formula of the language $\mathcal{L}_0^\omega$, and let $\boldsymbol{u}$ be the set of all free variables of A. The following holds.*

If

$$\mathsf{E\text{-}PRA}^\omega + \mathsf{QF\text{-}AC}^{1,0} + \mathsf{QF\text{-}AC}^{0,1} + \mathsf{WKL} \vdash A,$$

then

$$\mathsf{WE\text{-}PRA}^\omega \oplus \mathsf{QF\text{-}AC} \oplus \mathsf{WKL} \vdash E_{\underline{\sigma}}(\boldsymbol{u}) \to A^E.$$

Proof. This is shown by induction on the length of the proof. We have to show that it holds for all axioms and that it is preserved under application of a rule. Since the $(*)^E$ translation does not change the propositional structure, the theorem directly follows for all propositional axioms P1-P5, the equality axioms and the Peano axioms. As for the definitional axioms for the functional constants, lets take for example K. $\mathrm{K}(s,t) = s$ actually reads as $\forall \boldsymbol{z}(\mathrm{K}(s,t)(\boldsymbol{z}) = s(\boldsymbol{z}))$, we can therefore easily derive $\forall \boldsymbol{z}(E(\boldsymbol{z}) \to \mathrm{K}(s,t)(\boldsymbol{z}) = s(\boldsymbol{z}))$, which is nothing else than $(\mathrm{K}(s,t) = s)^E$. Putting $E(\boldsymbol{u})$ in front is now trivial.

As for the rules, we see that the propositional rules with only one premise are also trivial, since no new free variables are introduced.

So we only have to check P6, P7, Q1-Q4, Ind, $\mathsf{QF\text{-}AC}^{1,0}$, $\mathsf{QF\text{-}AC}^{0,1}$, (E) and WKL.

P6). Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be the free variables of A and B respectively. By induction hypothesis, the following hold.

$$E(\boldsymbol{u}) \to A^E \text{ and } E(\boldsymbol{u}) \wedge E(\boldsymbol{v}) \to (A^E \to B^E).$$

These can be combined using only propositional logic to obtain

$$E(\boldsymbol{u}) \wedge E(\boldsymbol{v}) \to B^E.$$

We now reduce the sequence $\boldsymbol{u}$ to a sequence $\boldsymbol{u}'$ by losing all variables which already appear in $\boldsymbol{v}$, and then replace the remaining ones with closed terms $\boldsymbol{F}$ and get $E(\boldsymbol{F}) \wedge E(\boldsymbol{v}) \to B^E$. By lemma 3.4.6.3, this now implies $E(\boldsymbol{v}) \to B^E$.

P7). Let $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w}$ be the free variables of A, B and C respectively. By induction hypothesis

$$E(\boldsymbol{u}) \wedge E(\boldsymbol{v}) \to (A^E \to B^E) \text{ and } E(\boldsymbol{v}) \wedge E(\boldsymbol{w}) \to (B^E \to C^E)$$

hold. These imply

$$E(\boldsymbol{u}) \wedge E(\boldsymbol{v}) \wedge E(\boldsymbol{w}) \to (A^E \to B^E) \text{ and } E(\boldsymbol{u}) \wedge E(\boldsymbol{v}) \wedge E(\boldsymbol{w}) \to (B^E \to C^E),$$

which imply by propositional logic

$$E(\boldsymbol{u}) \wedge E(\boldsymbol{v}) \wedge E(\boldsymbol{w}) \to (A^E \to C^E).$$

The rest is done analogously to P6.

Q1)+Q2). Let $\boldsymbol{u}$ be the free variables of $A[t]$. Using $E(\boldsymbol{u}) \to E(t)$ from lemma 3.4.6 and the tautology $E(\boldsymbol{u}) \to (A^E[t] \to A^E[t])$ we can derive

$$E(\boldsymbol{u}) \to (E(t) \to A^E[t]) \to A^E[t].$$

Using Q1 we get

$$E(\boldsymbol{u}) \to \forall x(E(x) \to A^E[x]) \to A^E[t],$$

which is $E(\boldsymbol{u}) \to (\forall x A[x] \to A[t])^E$

Using the same formulas as above we can also derive

$$E(\boldsymbol{u}) \to (A^E[t] \to E(t) \wedge A^E[t]).$$

we now use Q2 to obtain $E(\boldsymbol{u}) \to (A[t] \to \exists x A[x])^E$.

Q3). Let $\boldsymbol{u}$ be the free variables of $A \to \forall x B[x]$. By induction hypothesis

$$E(\boldsymbol{u}) \wedge E(v) \to (A^E \to B^E[v])$$

holds. This implies

$$E(\boldsymbol{u}) \to (A^E \to (E(v) \to B^E[v])).$$

Since v does not appear neither in A nor in $\boldsymbol{u}$, an application of Q3 yields

$$E(\boldsymbol{u}) \to (A^E \to \forall x(E(x) \to B^E[x])),$$

which is $E(\boldsymbol{u}) \to (A \to \forall x B[x])^E$.

Q4). Let $\boldsymbol{u}$ be the free variables of $\exists x A[x] \to B$. By induction hypothesis

$$E(\boldsymbol{u}) \wedge E(v) \to (A^E[v] \to B^E)$$

holds. This implies

$$E(\boldsymbol{u}) \to (E(v) \wedge A^E[v] \to B^E).$$

Since v does not appear neither in B nor in $\boldsymbol{u}$, an application of Q4 yields

$$E(\boldsymbol{u}) \to (\exists x(E(x) \wedge A^E[x]) \to B^E),$$

which is $E(\boldsymbol{u}) \to (\exists x A[x] \to B)^E$.

Ind). Let $\boldsymbol{u}$ be the free variables of $A(0)$. Since A is quantifier-free, they are all of type 0. By induction hypothesis

$$E(\boldsymbol{u}) \to A^E(0) \text{ and } E(\boldsymbol{u}) \wedge E(n) \to (A^E(n) \to A^E(n')).$$

Since $E(n)$ holds, we can eliminate it from the second hypothesis and then derive

$$(E(\boldsymbol{u}) \wedge A^E(n)) \to (E(\boldsymbol{u}) \wedge A^E(n')).$$

To this we can now apply the induction rule and then reintroduce $E(n)$ to get

$$E(\boldsymbol{u}) \wedge E(n) \to A^E(n).$$

WKL). Since WKL only contains bound variables of type $\leqslant 1$, $\mathsf{WKL} \leftrightarrow \mathsf{WKL}^E$ and therefore WKL^E holds. Since WKL does not contain free variables, we are done.

(E)). From the definition of $E(z)$ we get

$$E(z) \to \forall x, y(x \equiv y \to \forall \boldsymbol{v}, \boldsymbol{w}(\boldsymbol{v} \equiv \boldsymbol{w} \to z(x, \boldsymbol{v}) =_0 z(y, \boldsymbol{w}))),$$

which we first transform to

$$E(x) \wedge E(y) \wedge E(z) \to \forall x, y(x \equiv y \to \forall \boldsymbol{v}(\boldsymbol{v} \equiv \boldsymbol{v} \to z(x, \boldsymbol{v}) =_0 z(y, \boldsymbol{v})))$$

and then to

$$\forall x, y, z(E(x) \wedge E(y) \wedge E(z) \to (x \equiv y \to \forall \boldsymbol{v}(E(\boldsymbol{v}) \to z(x, \boldsymbol{v}) =_0 z(y, \boldsymbol{v})))).$$

Now we take the definition of $z(x) \equiv z(y)$

$$z(x) \equiv z(y) \leftrightarrow \forall \boldsymbol{v}(E(\boldsymbol{v}) \to z(x, \boldsymbol{v}) =_0 z(y, \boldsymbol{v})) \wedge E(z(x)) \wedge E(z(y)),$$

which implies

$$E(x) \wedge E(y) \wedge E(z) \wedge \forall \boldsymbol{v}(E(\boldsymbol{v}) \to z(x, \boldsymbol{v}) =_0 z(y, \boldsymbol{v})) \to z(x) \equiv z(y).$$

The combination of the two now yields

$$\forall x, y, z(E(x) \wedge E(y) \wedge E(z) \to (x \equiv y \to z(x) \equiv z(y))).$$

We now use remark 3.4.4.5 to obtain the final

$$\forall x, y, z(E(x) \wedge E(y) \wedge E(z) \to ((x = y)^E \to (z(x) = z(y))^E)).$$

QF-AC^{1,0}, QF-AC^{0,1}). We only consider the second case, the first is a bit more complicated but can be looked up in [14].

$(\mathsf{QF\text{-}AC}^{0,1})^E = \forall x^0(E(x) \to \exists y^1(E(y) \wedge A^E(x,y))) \to \exists Y(E(Y) \wedge \forall x(E(x) \to A^E(x, Y(x))))$

By QF-AC^{0,1}

$$\forall x \exists y(E(y) \wedge A^E(x,y)) \to \exists Y \forall x(E(Y(x)) \wedge A^E(x, Y(x)))$$

holds, which, by using $\forall x^0 E(Y(x)) \to E(Y)$, implies

$$\forall x \exists y(E(y) \wedge A^E(x,y)) \to \exists Y(E(Y) \wedge \forall x A^E(x, Y(x))).$$

The rest is trivial. □

In the above proof, we did not actually use the full force of QF-AC within the weakly extensional theory. The theorem would still hold if we replaced it by QF-AC^{1,0} and QF-AC^{0,1}.

3.4.3 Conservation result

Theorem 3.4.8. E-PRA$^\omega$ + QF-AC^{1,0} + QF-AC^{0,1} + WKL *is* Π^0_2*-conservative over* PRA.

Proof. Being able to eliminate extensionality, this proof now becomes rather easy. Consider

$$\mathsf{E\text{-}PRA}^\omega + \mathsf{QF\text{-}AC}^{1,0} + \mathsf{QF\text{-}AC}^{0,1} + \mathsf{WKL} \vdash \forall x \exists y A(x,y).$$

By theorem 3.4.7 this implies

$$\mathsf{WE\text{-}PRA}^\omega \oplus \mathsf{QF\text{-}AC} \oplus \mathsf{WKL} \vdash E(\boldsymbol{u}) \to (\forall x \exists y A(x,y))^E.$$

Now, we first eliminate the antecedent by using the fact that all the variables $\boldsymbol{u}$ are of type 0 and that therefore $E(\boldsymbol{u})$ holds. Secondly we use the fact that x and y are of type 0 and that therefore $\forall x \exists y A(x,y) \leftrightarrow (\forall x \exists y A(x,y))^E$ holds. This leaves us with

$$\mathsf{WE\text{-}PRA}^\omega \oplus \mathsf{QF\text{-}AC} \oplus \mathsf{WKL} \vdash \forall x \exists y A(x,y)$$

and the proof reduces to the proof of theorem 3.2.4. □

3.5 The theory E-PRA$^\omega$ + QF-AC^{1,0} + QF-AC^{0,1} + UWKL

The most interesting case now is this one. We still know that PRA is a lower bound, but it is no longer the lowest upper bound. We can now proof that PA is also a subtheory. As for the upper bound result, it can be proved that PA is one with respect to all first order formulas. This is then obviously also true for the former three theories.

3.5.1 Subtheory result

To prove that PA is a subtheory, we first show that UWKL is discontinuous. We then use an argument, which is known as 'Grilliot's trick' [6] to show that UWKL implies the existence of φ, the functional which represents numerical quantification. This functional can then be used to extend theorem 2.1.2.

Definition 3.5.1. A type 2 functional f of $\mathsf{E\text{-}PRA}^\omega$ is *effectively discontinuous* if there exists a sequence of closed type 1 terms $(g_i)_{i\in\mathbb{N}}$ (i.e. $\lambda ix.g(i,x)$ is a closed type $0 \to 1$ term) and a closed type 1 term g, such that $g = \lim_{i\to\infty} g_i$ but $f(g) \neq \lim_{i\to\infty} f(g_i)$ (if $\lim f(g_i)$ should exists).

$g = \lim_{i\to\infty} g_i$ is to be read as

$$\forall n^0 \exists i_n^0 (\forall m < n \forall i > i_n (g(m) = g_i(m))).$$

Theorem 3.5.2. $\mathsf{E\text{-}PRA}^\omega$ *proves that any* Φ *satisfying* UWKL *is effectively discontinuous. To be exact, we would have to say, that* $\lambda f.\Phi(f,0)$ *is effectively discontinuous.*

Proof. We want to show that

$$\mathsf{E\text{-}PRA}^\omega \vdash \left\{ \begin{array}{l} \forall \Phi^{1\to1}[\forall f^1\{\forall k \neg \mathrm{Bounded}(f^{\mathrm{tree}},k) \to \forall k(\overline{\Phi(f)}(k) \in f^{\mathrm{tree}})\} \to \\ \to \exists g_{(\cdot)}^{0\to1}, g^1\{\forall k \neg \mathrm{Bounded}(g^{\mathrm{tree}},k) \wedge \forall i,k \neg \mathrm{Bounded}(g_i^{\mathrm{tree}},k) \wedge \\ \wedge \forall i \forall j \geqslant i(g_j(i) = g(i)) \wedge \forall i,j(\Phi(g_i,0) = \Phi(g_j,0) \neq \Phi(g,0))\}] \end{array} \right.$$

First we define g primitive recursive, such that

$$g(k) = \left\{ \begin{array}{lll} 0 & , & \forall m < lh(k)(\pi(k,m) = 0) \vee \forall m < lh(k)(\pi(k,m) = 1) \\ 1 & , & \text{otherwise.} \end{array} \right.$$

g represents a binary tree with two infinite paths, one consisting solely of 0's the other solely of 1's. Obviously we have $\mathsf{E\text{-}PRA}^\omega \vdash \forall k \neg \mathrm{Bounded}(g^{\mathrm{tree}},k)$.

We now define the terms g_i according to what the value of $\Phi(g,0)$ is.

<u>Case 1</u>: $\Phi(g,0) = 0$. We define a primitive recursive function $\lambda ik.g_i(k)$ such that

$$g_i(k) = \left\{ \begin{array}{lll} 0 & , & (lh(k) \leqslant i \wedge \forall m < lh(k)(\pi(k,m) = 0)) \vee (\forall m < lh(k)(\pi(k,m) = 1)) \\ 1 & , & \text{otherwise.} \end{array} \right.$$

We see that g_i differs from g only in one aspect, namely that its left branch has been cut at level i. Again we readily see that $\mathsf{E\text{-}PRA}^\omega \vdash \forall k \neg \mathrm{Bounded}(g_i^{\mathrm{tree}},k)$. Also do we get $\mathsf{E\text{-}PRA}^\omega \vdash \forall k \forall l \geqslant lh(k)(g_l(k) = g(k))$. Since lh has the property that $lh(k) < k$, we get

$$\mathsf{E\text{-}PRA}^\omega \vdash \forall k \forall l \geqslant k(g_l(k) = g(k)).$$

The last thing we have to do is to look at the value of $\Phi(g_i, 0)$. But since the only infinite path in g_i is the one consisting only of 1's, it follows that

$$\forall i(\Phi(g_i, 0) = 1).$$

<u>Case 2</u>: $\Phi(g, 0) = 1$. We only make a slight change in the definition of the term $\lambda ik.g_i(k)$.

$$g_i(k) = \begin{cases} 0 & , \quad (lh(k) \leqslant i \wedge \forall m < lh(k)(\pi(k,m) = 1)) \vee (\forall m < lh(k)(\pi(k,m) = 0)) \\ 1 & , \quad \text{otherwise.} \end{cases}$$

The rest works as in case 1.

To finish the proof of the discontinuity of Φ, we name the two sequences of terms according to the case they were defined in, namely $g_{1,i}$ and $g_{2,i}$, and then combine them to a new sequence g_i, by

$$g_i(k) = (1 \dot{-} \Phi(g, 0)) \cdot g_{1,i}(k) + \Phi(g, 0) \cdot g_{2,i}(k).$$

□

Theorem 3.5.3. $\mathsf{E\text{-}PRA}^\omega \vdash \mathsf{UWKL} \leftrightarrow \exists \varphi^2 \forall f^1(\varphi(f) = 0 \leftrightarrow \exists x^0(f(x) = 0))$.

Proof. 1) '$\rightarrow$' We use the fact that any Φ satisfying UWKL is discontinuous, or rather the sequences $g_{k,i}$ defined in the proof of this fact, to define φ primitive recursively in Φ, such that $(\varphi) :\equiv \forall f(\varphi(f) = 0 \leftrightarrow \exists x(f(x) = 0))$ holds provably in $\mathsf{E\text{-}PRA}^\omega$. It is this step, which is known as 'Grilliot's trick'.

We again define φ according to the value of $\Phi(g, 0)$. Assume $\Phi(g, 0) = 0$. The first step is to define a term $t_1^{1\rightarrow 1}$ such that we have

$$t_1(h, i) = \begin{cases} g_{1,j}(i) & , \quad \text{for the least } j < i \text{ such that } h(j) > 0, \text{ if such a } j \text{ exists} \\ g_{1,i}(i) & , \quad \text{otherwise.} \end{cases}$$

This term t_1 can be constructed in $\mathsf{E\text{-}PRA}^\omega$ since the bounded minimum operator is primitive recursive. Using the two facts $\forall i \forall j \geqslant i(g_{1,j}(i) = g_{1,i}(i))$ and $\forall i(g_{1,i}(i) = g(i))$ about the term $\lambda ik.g_i(k)$, we obtain the two following implications.

$$\exists j(h(j) > 0) \rightarrow t_1(h) =_1 g_{1,j} \text{ for the least such } j, \text{ and } \forall j(h(j) = 0) \rightarrow t_1(h) =_1 g.$$

We now apply the extensionality axiom for type 2 functionals to the second implication and get

$$\forall j(h(j) = 0) \rightarrow \Phi(t_1(h), 0) =_0 \Phi(g, 0).$$

This is the place where the extensionality axiom now makes the difference between the weakly and the fully extensional theory.

If we now look at the first of the two above mentioned implications, we see that it is equivalent to

$$\forall j(t_1(h) \neq g_{1,j}) \to \forall j(h(j) = 0),$$

which implies

$$t_1(h) =_1 g \to \forall j(h(j) = 0).$$

As above we now introduce Φ to this formula.

$$\Phi(t_1(h), 0) = \Phi(g, 0) \to \forall j(h(j) = 0).$$

If we now combine the two implications containing Φ we obtain

$$\forall j(h(j) = 0) \leftrightarrow \Phi(t_1(h), 0) =_0 \Phi(g, 0).$$

This of course can also be written as

$$\forall j(h(j) = 0) \leftrightarrow |\Phi(t_1(h), 0) - \Phi(g, 0)| = 0$$

or equivalently, as

$$\forall j(1 \dot{-} h(j) = 0) \leftrightarrow |\Phi(t_1(\lambda x.(1 \dot{-} h(x))), 0) - \Phi(g, 0)| = 0.$$

By taking the contrapositive and changing the existential-quantifier to an all-quantifier we obtain

$$\exists j(h(j) = 0) \leftrightarrow |\Phi(t_1(\lambda x.(1 \dot{-} h(x))), 0) - \Phi(g, 0)| \neq 0,$$

or

$$\exists j(h(j) = 0) \leftrightarrow 1 \dot{-} |\Phi(t_1(\lambda x.(1 \dot{-} h(x))), 0) - \Phi(g, 0)| = 0.$$

So finally we define φ by

$$\varphi := \lambda h.(1 \dot{-} |\Phi(t_1(\lambda x.(1 \dot{-} h(x))), 0) - \Phi(g, 0)|).$$

We must not forget, that we have defined φ assuming case 1, but we also need a term t_2 according to case 2, so that we can combine them to obtain a final version of φ. This can be reached by constructing a closed term χ, such that

$$\mathsf{E\text{-}PRA}^\omega \vdash \forall x((x = 0 \to \chi(x) = t_1) \wedge (x \neq 1 \to \chi(x) = t_2)).$$

We can now define the final version of φ by

$$\varphi := \lambda h.(1 \dot{-} |\Phi(\chi(\Phi(g, 0))(\lambda x.(1 \dot{-} h(x))), 0) - \Phi(g, 0)|).$$

2) '$\leftarrow$' We define Φ primitive recursive in φ as follows.

$$\Phi(f,0) = \begin{cases} 0 & , \quad \exists m^0 \forall s^0 \leqslant \text{AllOne}(m)(\langle 1\rangle * s \notin f^{\text{tree}}) \\ 1 & , \quad \text{otherwise.} \end{cases}$$

$$\Phi(f,k') = \begin{cases} 0 & , \quad \exists m^0 \forall s^0 \leqslant \text{AllOne}(m)(\overline{\Phi(f,k')} * \langle 1\rangle * s \notin f^{\text{tree}}) \\ 1 & , \quad \text{otherwise.} \end{cases}$$

In this form Φ is obviously not yet a primitive recursive term in φ, all the problematic parts though can successively be removed. The first step is to replace the $\notin$-relation symbol and the all-quantifier. $\notin$ is just replaced by its definition, while the all-quantifier, since it is bounded, can be replaced by a bounded sum, which in turn can be represented by a primitive recursive term.

$$\Phi(f,0) = \begin{cases} 0 & , \quad \exists m^0 [\sum\limits_{i=0}^{\text{AllOne}(m)} (1 \dot{-} f^{\text{tree}}(\langle 1\rangle * s)) = 0] \\ 1 & , \quad \text{otherwise.} \end{cases}$$

$$\Phi(f,k') = \begin{cases} 0 & , \quad \exists m^0 [\sum\limits_{i=0}^{\text{AllOne}(m)} (1 \dot{-} f^{\text{tree}}(\overline{\Phi(f,k')} * \langle 1\rangle * s)) = 0] \\ 1 & , \quad \text{otherwise.} \end{cases}$$

The last step now is trivial, since we have φ at hand.

$$\Phi(f,0) = \begin{cases} 0 & , \quad \varphi(\lambda m.[\sum\limits_{i=0}^{\text{AllOne}(m)} (1 \dot{-} f^{\text{tree}}(\langle 1\rangle * s))]) = 0 \\ 1 & , \quad \text{otherwise.} \end{cases}$$

$$\Phi(f,k') = \begin{cases} 0 & , \quad \varphi(\lambda m.[\sum\limits_{i=0}^{\text{AllOne}(m)} (1 \dot{-} f^{\text{tree}}(\overline{\Phi(f,k')} * \langle 1\rangle * s))]) = 0 \\ 1 & , \quad \text{otherwise.} \end{cases}$$

$\square$

As a consequence of this proof we see, that UWKL taken for itself is weaker than (φ), meaning that UWKL implies (φ) only in combination with (E), whereas UWKL can be defined using only (φ).

Lemma 3.5.4. *Let A be a formula of* PA*, then there is a term t_A of* E-PRA$^\omega$ + UWKL*, such that*

$$\text{E-PRA}^\omega + \text{UWKL} \vdash A \leftrightarrow t_A = 0$$

Proof. By induction on the formula complexity. The base case and the cases for the logical connectives have already been proved in theorem 2.1.2. What remains are the two quantifier cases:

$\exists$). $A \equiv \exists x B(x)$. By induction hypothesis $B(x) \leftrightarrow t_B(x) = 0$. From this we get $\exists x B(x) \leftrightarrow \exists x(t_B(x) = 0)$, and using φ, $\exists x(t_B(x) = 0) \leftrightarrow \varphi(t_B) = 0$. So we put $t_A \equiv \varphi(t_B)$.

$\forall$). $A \equiv \forall x B(x)$. From the induction hypothesis we again get $\forall x B(x) \leftrightarrow \forall x(t_B(x) = 0)$, and finally

$$\begin{aligned}
\forall x t_B(x) = 0 &\leftrightarrow \neg\exists x \neg(t_B(x) = 0) \\
&\leftrightarrow \neg\exists x(t_B(x) \neq 0) \\
&\leftrightarrow \neg\exists x(1 \dot{-} t_B(x) = 0) \\
&\leftrightarrow \neg\varphi(\lambda x.(1 \dot{-} t_B(x))) = 0 \\
&\leftrightarrow 1 \dot{-} \varphi(\lambda x.(1 \dot{-} t_B(x))) = 0
\end{aligned}$$

So $t_A \equiv \varphi(\lambda x.(1 \dot{-} t_B(x)))$. □

Theorem 3.5.5. PA *is a subtheory of* $\mathsf{E\text{-}PRA}^\omega + \mathsf{QF\text{-}AC}^{1,0} + \mathsf{QF\text{-}AC}^{0,1} + \mathsf{UWKL}$.

Proof. We only need to show that $\mathsf{E\text{-}PRA}^\omega + \mathsf{QF\text{-}AC}^{1,0} + \mathsf{QF\text{-}AC}^{0,1} + \mathsf{UWKL}$ proves induction for every formula A of PA. But this follows directly from the above lemma.

Assume you have $A(0) \wedge A(n) \to A(n')$ provable in $\mathsf{E\text{-}PRA}^\omega + \mathsf{QF\text{-}AC}^{1,0} + \mathsf{QF\text{-}AC}^{0,1} + \mathsf{UWKL}$ for any formula A of PA. By the above lemma we get $t_A(0) = 0 \wedge t_A(n) = 0 \to t_A(n') = 0$ to which we can apply quantifier free induction rule to obtain $t_A(n) = 0$, which finally implies $A(n)$. □

3.5.2 Conservation result

To prove theorem 3.4.8 we used elimination of extensionality to reduce the proof to the one of theorem 3.2.4. Obviously one would try to use the same method with the theory $\mathsf{E\text{-}PRA}^\omega + \mathsf{QF\text{-}AC}^{1,0} + \mathsf{QF\text{-}AC}^{0,1} + \mathsf{UWKL}$. Unfortunately this method will not work, because theorem 3.4.7 does not work with UWKL instead of WKL.

Consider $(\mathsf{UWKL})^E$

$$(\mathsf{UWKL})^E = \exists \Phi\Big(E(\Phi) \wedge \forall f\Big(E(f) \to \big(\forall k\big(E(k) \to \neg\mathrm{Bounded}(f^{\mathrm{tree}}, k)\big) \to$$
$$\to \forall k\big(E(k) \to (\overline{\Phi(f)}(k) \in f^{\mathrm{tree}})\big)\big)\Big)\Big)$$

Since $E(x)$ holds if x is of type 0 or 1, this can be reduced to

$$\exists \Phi(E(\Phi) \wedge \forall f(\mathrm{BinTree}(f) \wedge \forall k \neg\mathrm{Bounded}(f, k) \to \forall k(\overline{\Phi(f)}(k) \in f)))$$

The question now is whether this follows from UWKL in $\mathsf{WE\text{-}PRA}^\omega \oplus \mathsf{QF\text{-}AC} \oplus \mathsf{UWKL}$. Unfortunately it does not. The problem is that $E_{1\to 1}(F)$ cannot be shown without (E).

$$E_{1\to 1}(F) \leftrightarrow \forall x^1, y^1, v^0, w^0(x \equiv_1 y \wedge v \equiv_0 w \to F(x,v) =_0 F(y,w))$$

Since $v \equiv_0 w \leftrightarrow v =_0 w$ and

$$\begin{aligned} x \equiv_1 y &\leftrightarrow \forall z^0(E(z) \to x(z) =_0 y(z)) \wedge E(x) \wedge E(y) \\ &\leftrightarrow \forall z(x(z) =_0 y(z)) \\ &\leftrightarrow x =_1 y \end{aligned}$$

we can reduce the above statement to

$$E_{1\to 1}(F) \leftrightarrow \forall x^1, y^1, v^0, w^0(x =_1 y \wedge v =_0 w \to F(x,v) =_0 F(y,w)),$$

where the right side is (E) and can therefore not be proved in $\mathsf{WE\text{-}PRA}^\omega \oplus \mathsf{QF\text{-}AC} \oplus \mathsf{UWKL}$.

The route taken to prove the conservation result in the above three theories can in this case not be taken, we can though prove another somewhat weaker theorem, which of course holds for the other theories, too.

Theorem 3.5.6. $\mathsf{E\text{-}PRA}^\omega + \mathsf{QF\text{-}AC}^{1,0} + \mathsf{QF\text{-}AC}^{0,1} + \mathsf{UWKL}$ *is conservative over* PA.

The proof of this theorem shall only be sketched, since its writing out in detail would require to much additional material.

Proof. Let A be a formula of PA. The first step is to reduce

$$\mathsf{E\text{-}PRA}^\omega + \mathsf{QF\text{-}AC}^{1,0} + \mathsf{QF\text{-}AC}^{0,1} + \mathsf{UWKL} \vdash A$$

to

$$\mathsf{E\text{-}PRA}^\omega + \mathsf{QF\text{-}AC}^{1,0} + \mathsf{QF\text{-}AC}^{0,1} + (\mu) \vdash A$$

where $(\mu) := \exists \mu^2 \forall f^1(\exists x^0(f(x) = 0) \to f(\mu(f)) = 0)$. This is done by showing, that

$$\mathsf{E\text{-}PRA}^\omega + \mathsf{QF\text{-}AC}^{1,0} + \mathsf{QF\text{-}AC}^{0,1} \vdash \mathsf{UWKL} \leftrightarrow \exists \mu^2 \forall f^1(\exists x^0(f(x) = 0).$$

But (μ) obviously implies (φ), and (φ) implies (μ) by applying $\mathsf{QF\text{-}AC}^{1,0}$ to $\forall f \exists x(\varphi(f) = 0 \to f(x) = 0)$. The rest then follows from theorem 3.5.3.

The next step is to eliminate extensionality, which unlike UWKL, works with (μ).

$$(\mu)^E = \exists \mu^2(E(\mu) \wedge \forall f^1(\exists x^0(f(x) = 0) \to f(\mu(f)) = 0)$$

Considering $E(\mu)$, we see that it is equivalent to $\forall x, y(x =_1 y \to \mu(x) =_0 \mu(y))$ which is derivable from QF-ER.

We now have

$$\mathsf{WE\text{-}PRA}^\omega \oplus \mathsf{QF\text{-}AC} \oplus (\mu) \vdash A.$$

The fact that this can now be reduced to $\mathsf{PA} \vdash A$ requires knowledge about arithmetical hierarchies and term modelling in PA. We do not want to introduce these notions here, but a full proof can be found in [4] or [1]. $\square$

Bibliography

[1] Avigad J., Feferman S., *Gödel's Functional ("Dialectica") Interpretation*, In: [3], pp. 337-405 (1998). 5, 7, 26, 41, 66

[2] Barwise, J. (editor), *Handbook of mathematical logic*, North Holland (1977) 67

[3] Buss S.R. (editor), *Handbook of Proof Theory. Studies in Logic and the Foundation of Mathematics*, Elsevier Science B.V., Amsterdam (1998). 67

[4] Feferman, S., *Theories of finite type related to mathematical practice*, In: [2], pp. 913-972 (1977) 66

[5] Gödel, K., *Über eine bisher noch nicht benüzte Erweiterung des finiten Standpunktes*, Dialectica **12**, pp. 280-287 (1958) 8, 20

[6] Grilliot, T. J., *On effectively discontinuous type-2 objects*, J. Symbolic Logic **36**, pp. 245-248 (1971) 61

[7] Howard, W.A., *Hereditarliy majorizable functionals of finite type*, In: [15], pp. 454-461 (1973) 41, 47

[8] Jäger, G., *Beweistheorie*, Lecture Notes (1998) 7, 15, 42, 45

[9] Jäger, G., *Einführung in die theoretische Informatik*, Lecture Notes (2003) 7

[10] Kleene, S.C., *Recursive functionals and quantifiers of finite types, I*, Trans. of the A.M.S. **91**, pp. 1-52 (1959) 24, 25

[11] Kohlenbach U., *Effective bounds from ineffective proofs in analysis: an application of functional interpretation and majorization*, J. Symbolic Logic **57**, pp. 1239-1273 (1992) 41

[12] Kohlenbach U., *A note on Spector's quantifier-free rule of extensionality*, Arch. Math. Logic **40**, pp. 89-92 (2001)

[13] Kohlenbach U., *On uniform weak König's lemma*, Ann. Pure Applied Logic **114**, pp. 103-119 (2002) 5, 7, 41

[14] Luckhardt, H., *Extensional Gödel Functional Interpretation*, Springer Lecture Notes in Mathematics **306** (1973) 41, 55, 60

[15] Troelstra, A.S. (editor), *Metamathematical investigation of intuitionistic arithmetic and analysis*, Springer Lecture Notes in Mathematics **344** (1973) 5, 7, 14, 26, 28, 34, 36, 45, 50, 67

[16] Troelstra, A.S., van Dalen, D., *Constructivism in Mathematics: An Introduction*, vol. 1, North-Holland, Amsterdam (1988) 7, 19