Proof Interpretations

Ulrich Kohlenbach

Ulrich Kohlenbach

kohlenb@brics.dk

BRICS[1]
Department of Computer Science
University of Aarhus
Ny Munkegade
DK-8000 Aarhus C, Denmark

June 1998

[1]**B**asic **R**esearch **I**n **C**omputer **S**cience,
Centre of the Danish National Research Foundation.

Preface

These lecture notes are a polished version of notes from a BRICS PhD course which I gave in the spring term 1998.

Their purpose is to give an introduction to two major proof theoretic techniques: functional interpretation and (modified) realizability. We focus on the possible use of these methods to extract programs, bounds and other effective data from given proofs.

Both methods are developed in the framework of intuitionistic arithmetic in higher types.

We also discuss applications to systems based on classical logic. We show that the combination of functional interpretation with the so-called negative translation, which allows to embed various classical theories into their intuitionistic counterparts, can be used to unwind non-constructive proofs.

Instead of combining functional interpretation with negative translation one can also use in some circumstances a combination of modified realizability with negative translation if one inserts the so-called A-translation (due to H. Friedman) as an intermediate step.

Acknowledment: I am grateful to the participants of my BRICS PhD course for helpful and clarifying comments.

Contents

Chapter 1

Introduction: Unwinding proofs

Proof interpretations of the kind we are going to study in these lectures are tools to extract constructive (computational) data from given proofs by recursion on the proof.
Such data quite often cannot directly be read off from a proof but are hidden behind the use of quantifiers.
G. Kreisel was the first to formulate the program of unwinding proofs under the general question:

'What more do we know if we have proved a theorem by restricted means than if we merely know that it is true?'

What do we mean by 'constructive data'?
E.g.

1) Realizing terms from a proof of an existential theorem $A \equiv \exists x\, B(x)$ (closed).

   A weaker requirement is to construct a list of terms $t_1, \ldots, t_n$ which are candidates for A, i.e. such that $B(t_1) \vee \ldots \vee B(t_n)$ holds.

   More general: If $A \equiv \forall x \exists y\, B(x,y)$, then one can ask for an algorithm p such that $\forall x\, B(x, p(x))$ holds (or – weaker– for a bounding function b such that $\forall x \exists y \leq b(x)\, B(x,y)$, if e.g. y ranges over the natural numbers).

2) weakening of the assumptions used in the proof: e.g. replacing general assumptions by specific instances of them.

What type of information one can expect (in general) depends of course on the structure of the theorem A to be proved and the principles used in its proof.
A first VERY ROUGH division of the structure of a sentence[1] A can be made according to the quantifier complexity of A:

1) A purely universal, i.e. $A \equiv \forall x A_0(x)$, where A_0 is quantifier-free.[2]
Such sentences A, sometimes called complete, don't ask for any witnessing data. So the problem of extracting data is empty here.

2) A purely existential, i.e. $A \equiv \exists x\, A_0(x)$. We treat this as a special case of

3) $A \equiv \forall x \exists y\, A_0(x, y)$. Lets consider the case where $x, y \in \mathbb{N}$ and $A_0 \in \mathcal{L}(\mathrm{PA})$ (here PA denotes first-order Peano arithmetic which we assume to contain all primitive recursive functions). A_0 is decidable (Exercise: $A_0(\underline{x}) \in \mathcal{L}(\mathrm{PA})$, then one can construct a primitive recursive function term t such that PA $\vdash \forall \underline{x}(t\underline{x} = 0 \leftrightarrow A_0(\underline{x})))$ and therefore defines a partial recursive function f, namely

$$f(x) := \begin{cases} \min y[A_0(x, y)], \text{ if } \exists y\, A_0(x, y) \\ \text{undefined, otherwise.} \end{cases}$$

A just says that f is total recursive.

Questions: How to extract a non-trivial program for f (different from simple unbounded search) from a proof of A? What is the complexity and the rate of growth of f if A is proved in a certain theory $\mathcal{T}$?

Theorems expressing that a set $\{y \in \mathbb{N} : A(y)\} \subseteq \mathbb{N}$ is infinite have the form $\forall x \in \mathbb{N} \exists y \geq x\, A(y)$. Quite often A can be expressed in a quantifier-free way

[1] As usual a sentence is a closed formula.

[2] From now on $A_0, B_0, C_0, \ldots$ always denote quantifier-free formulas. Instead of a single variable we may have (here and in the following) also a tuple $\underline{x} = x_1, \ldots, x_n$ of variables.

A_0 in PA, so that this falls under the general form $\forall x \exists y\, B_0(x,y)$, where $B_0(x,y) :\equiv (y \geq x \wedge A_0(y))$.

As an example consider the following

Proposition 1.1 *There are infinitely many prime numbers.*

The predicate $P(x) :\equiv$ 'x is a prime number' can be expressed in a quantifier-free way as a primitive recursive predicate (see e.g. [27],[68]).

Proof 1 (Euclid): Define $a := 1 + \prod\limits_{\substack{p \leq x \\ p\, prime}} p$. a cannot be divided by any prime number $p \leq x$. By the decomposition of every number into prime factors it follows that a contains a prime factor $q \leq a$ with $q > x$. $\square$

From this proof one immediately gets the bound $g(x) := 1 + x!\, (\geq 1 + \prod\limits_{\substack{p \leq x \\ p\, prime}} p)$.

By the Stirling formula we obtain
$g(x) \sim 1 + (2\pi x)^{\frac{1}{2}} (\frac{x}{e})^x = 1 + \sqrt{2\pi} \cdot e^{x \log x - x + \frac{1}{2} \log x}$ and hence $g(x) \leq e^{x \log x}$ for sufficiently large x.

Proof 2 (Euler): Suppose that there are only finitely many prime numbers $p_0, \ldots, p_r$ (listed in increasing order). One has

$$\begin{aligned} &\sum_{0 \leq \alpha_0, \ldots, \alpha_r \leq n} \frac{1}{p_0^{\alpha_0} \cdot \ldots \cdot p_r^{\alpha_r}} = \left(\sum_{i=0}^{n} \frac{1}{p_0^i}\right) \cdot \ldots \cdot \left(\sum_{i=0}^{n} \frac{1}{p_r^i}\right) \\ &< \frac{1}{1 - \frac{1}{p_0}} \cdot \ldots \cdot \frac{1}{1 - \frac{1}{p_r}} = \frac{p_0}{p_0 - 1} \cdot \ldots \cdot \frac{p_r}{p_r - 1} \\ &\leq \frac{2}{1} \cdot \frac{3}{2} \cdot \frac{4}{3} \cdot \ldots \cdot \frac{p_r}{p_r - 1} = p_r \end{aligned}$$

(note that this holds for all $n \in \mathbb{N}$).
It follows (using the decomposition into prime numbers) that for all $n \in \mathbb{N}$

$$\sum_{i=1}^{n} \frac{1}{i} \leq p_r.$$

But this contradicts the fact that $\sum\limits_{i=1}^{\infty} \frac{1}{i} = \infty$. $\square$

Quantitative analysis of Euler's proof:

We need a quantitative version of '$\sum\limits_{i=1}^{n} \frac{1}{i} \stackrel{n\to\infty}{\to} \infty$', more precisely we need a bound on $\exists n(\sum\limits_{i=1}^{n} \frac{1}{i} > p_r)$. It is known that $\sum\limits_{i=1}^{n} \frac{1}{i} - ln(n) \searrow C$, where $C \approx 0.5772...$ is the so-called Euler-Mascheroni constant. Hence for $n_r := \lceil e^{p_r - C} \rceil$ we have $\sum\limits_{i=1}^{n_r} \frac{1}{i} > p_r$ (and this is essentially optimal). From the proof above it follows that for all $n \in \mathbb{N}$

$$\sum_{0 \le \alpha_0,\dots,\alpha_r \le n} \frac{1}{p_0^{\alpha_0} \cdot \ldots \cdot p_r^{\alpha_r}} \le p_r.$$

Hence there must be an i ($1 \le i \le n_r$) which contains a prime factor p with $p_r < p \le i \le n_r$. So put together

$$\exists p(p \ prime \ \wedge p_r < p \le \lceil e^{p_r - C} \rceil).$$

Applying this argument to all prime numbers $p_0 < \ldots < p_{r_x} \le x$ we obtain

$$\forall x \exists p(p \ prime \ \wedge x < p \le \lceil e^{x - C} \rceil).$$

So we can take $g(x) := \lceil e^{x-C} \rceil$ (or an appropriate upper bound of this to make it computable).

Conclusion: Euler slightly better than Euclid!

Proof 3: Let $p_1, \ldots, p_j$ be the first j primes and define $N(x) := \{n \le x : n$ is not divisible by any prime $p > p_j\}$. We can express $n \in N(x)$ in the form $n = n_1^2 m$ where m is 'squarefree', i.e. is not divisible by a square of any prime.

We have $m = p_1^{b_1} \cdot p_2^{b_2} \cdot \ldots \cdot p_j^{b_j}$, where $b_i \in \{0, 1\}$. There are 2^j possible exponents and consequently at most 2^j different values of m. Also, because of $n_1 \le \sqrt{n} \le \sqrt{x}$, there are not more than $\sqrt{x}$ different values of n_1. Hence $|N(x)| \le 2^j \sqrt{x}$. Now if there were only finitely many primes $p_1, \ldots, p_j$, then $|N(x)| = x$ for every x and so $2^j \sqrt{x} \ge x$ for all x which is a contradiction.

From this proof one gets a bound as follows: Let $p_1, \ldots, p_j$ be the first j primes. Define $x := (2^j)^2 + 1 = 2^{2j} + 1$. Then $2^j\sqrt{x} < x$. Hence $\exists n \leq x(n$ is divisible by some prime $p > p_j)$ and so $\exists p(p$ prime $\wedge\, p_j < p \leq 2^{2j} + 1)$.
So we get a bound $g(j) := 2^{2j} + 1$ which is exponential in j (and no longer in $x \geq p_j$) which (for large enough j) is a significant improvement (e.g. from this bound one easily gets the lower bound $\frac{\log x}{2 \log 2}$ (for $x \geq 1$) for the Euler π-function $\pi(x) := |\{p : p \text{ prime } \wedge p \leq x\}|$ whereas the bound from Euclid's proof only yields $\log\log x$ (for $x \geq 2$) as a lower bound, see [23] for details).

For still another proof (in fact a variant of proof 3) see the exercise 1.

Discussion:

1) All three proofs provide more information than the mere fact that 'there are infinitely many primes' is true. By making their quantitative content explicit one can compare them with respect to their numerical quality.

2) The unwindings of the proofs 1)-3) were straightforward and didn't require any tools from logic as guiding principles. However there are more complicated proofs where the use of proof-theoretic tools turned out to be decisive in practice(see e.g. [14],[53],[31],[32]). The final verification of the data extracted will always be again an ordinary mathematical proof (obtained by a proof-theoretic transformation of the original proof) which does not rely on any logical meta-theorems (in contrast to the verification of the general procedure of transformation). This differs from many model theoretic applications to mathematics where the provability or the truth in some model of the conclusion is established without exhibiting a proof which doesn't rely on model theoretic theorems.

3) Already the a-priori information, provided by a general meta-theorem, that e.g. a certain computable bound must be extractable from a given proof which is formalizable in a certain system $\mathcal{T}$ can be an important step in actually finding such a bound even if the latter is carried

out by ad hoc methods and doesn't follow closely any proof-theoretic procedure.

Remark 1.2 *If A does not have the form $\forall x \exists y\, A_0(x,y)$ right away it may have so after some logical transformations, e.g.*

$$A :\equiv (\exists x \forall y\, A_0(x,y) \to \forall u \exists v\, B_0(u,v))$$

is logically equivalent to the prenex normal form

$$A^{pr} :\equiv \forall u, x \exists v, y (A_0(x,y) \to B_0(u,v))$$

so that the reasoning above applies to the A^{pr}.

4) $A \equiv \exists x \forall y\, A_0(x,y)$: From a proof of A (even in PL) one cannot (in general) obtain a realization $\forall y\, A_0(t,y)$ nor a list of candidates such that $\bigvee_{i=1}^{n} \forall y\, A_0(t_i,y)$ ($t, t_1, \ldots, t_n$ not containing y) holds:

Proposition 1.3 *There exists a Σ^0_2-sentence $A \equiv \exists x \forall y\, A_0(x,y) \in \mathcal{L}(\mathrm{PA})$ in the language of Peano arithmetic PA such that there is* **no** *list of closed terms $t_1, \ldots, t_k \in \mathcal{L}(PA)$ such that*

$$\mathrm{PA} \vdash \bigvee_{i=1}^{k} \forall y\, A_0(t_i,y).$$

Proof: Take $Px :\equiv Prov_{PA}(x, \overline{0} = \overline{1})$ and $A_0(x,y) :\equiv Px \vee \neg Py$ (here '$Prov_{PA}(x, \overline{0} = \overline{1})$' expresses primitive recursively 'x is the Gödel number of a PA-proof of $0 = 1$' (see e.g. [27]). Suppose there are closed terms $t_1, \ldots, t_k$ such that

$$(1)\ \mathrm{PA}\ \vdash \bigvee_{i=1}^{k} \forall y\, A_0(t_i,y).$$

Within PA each t_i can be computed to a numeral $\overline{n}_i$:

$$(2)\ \mathrm{PA}\ \vdash t_i = \overline{n}_i \text{ for } 1 \le i \le k.$$

By (1) and (2) we have

$$(3)\ \text{PA} \vdash \bigvee_{i=1}^{k} \forall y\, A_0(\overline{n}_i, y).$$

By the consistency of PA we know that

$$(4)\ \mathbb{N} \models \bigwedge_{i=1}^{k} \neg P\overline{n}_i.$$

Hence by the numeralwise representability of primitive recursive predicates in PA we have

$$(5)\ \text{PA} \vdash \bigwedge_{i=1}^{k} \neg P\overline{n}_i.$$

But (3) and (5) imply

$$(6)\ \text{PA} \vdash \forall y \neg Prov_{PA}(y, \overline{0} = \overline{1}),$$

which contradicts Gödel's second incompleteness theorem. □

However, although PA is not able to verify $\bigvee_{i=1}^{k} \forall y\, A_0(t_i, y)$ for some terms t_i **we** can (assuming the consistency of PA). In fact for **any** term t, e.g. for 0, we know that $A_0(t)$ is true in $\mathbb{N}$ simply because $\mathbb{N} \models \forall y \neg Prov_{PA}(y, \overline{0} = \overline{1})$.

But there are other examples where –in general– even this is not possible, e.g. take

$$A_e :\equiv \exists x \forall y (T(\overline{e}, \overline{e}, x) \vee \neg T(\overline{e}, \overline{e}, y)),$$

where T is the (primitive recursive) Kleene-T-predicate, i.e. $Txyz :\equiv$ 'the Turing machine with Gödel number x applied to the input y terminates with a computation whose Gödel number is z' (see e.g. [68]).
In general we are not able to determine closed terms $t_1, \ldots, t_k$ such that

$$\mathbb{N} \models \bigvee_{i=1}^{k} \forall y (T(\overline{e}, \overline{e}, t_i) \vee \neg T(\overline{e}, \overline{e}, y)),$$

since this would allow us to decide whether $\exists x T(\overline{e}, \overline{e}, x)$ or not (simply check whether $\bigvee_{i=1}^{k} T(\overline{e}, \overline{e}, t_i)$ is true or not).

In fact for

$$A :\equiv \forall x \exists y \forall z (Txxy \vee \neg Txxz)$$

A is provable in PA (using only the logical axioms and rules), but there is no computable bound g on '$\exists y$', i.e. no computable g such that

$$\forall x \exists y \leq gx \forall z (Txxy \vee \neg Txxz)$$

since this would make the (special) halting problem $\{x \in \mathbb{N} : \exists y \in \mathbb{N}(Txxy)\}$ decidable by the then computable function

$$fx := \begin{cases} 0, & \text{if } \exists y \leq gx(Txxy) \\ 1, & \text{otherwise.} \end{cases}$$

Two examples of non-constructive proofs from number theory:

Proposition 1.4 $\exists a, b \in \mathbb{R}\ \underbrace{(a, b \textit{ irrational } \wedge a^b \textit{ rational})}_{\textit{(sightly more complex than } \Pi^0_1\textit{)}}$.

Proof: Case 1: $\sqrt{2}^{\sqrt{2}}$ is rational. Put $a := b := \sqrt{2}$.
Case 2: $\sqrt{2}^{\sqrt{2}}$ irrational. Put $a := \sqrt{2}^{\sqrt{2}}, b := \sqrt{2}$. $\square$

From this proof we get two candidates for (a, b), namely $(\sqrt{2}, \sqrt{2})$ and $(\sqrt{2}^{\sqrt{2}}, \sqrt{2})$ but no decision which one satisfies the proposition.

Remark 1.5 *From a deep result of Gelfand and Schneider, stating that if a, b are algebraic, $a \neq 0, 1$ and b irrational, then a^b is transcendental, it follows that $\sqrt{2}^{\sqrt{2}}$ is transcendental and therefore irrational. So it is the pair $(\sqrt{2}^{\sqrt{2}}, \sqrt{2})$ which satisfies the proposition.*

Here is an example (communicated by H. Friedman) of a simple non-constructive PA-proof in number theory of a disjunction where none of the two disjuncts is known to be true up to now:

Proposition 1.6 PA $\vdash$ ($e - \pi$ is irrational) or ($e + \pi$ is irrational).

Proof: One easily formalizes the proof of the irrationality of e as given e.g. in [23] in PA. If both $e - \pi$ and $e + \pi$ were rational, then also their sum $2e$ and therefore e would be rational which is a contradiction. $\square$

We have seen that already for Σ^0_2, Π^0_3-sentences A it is not possible in general to compute witnesses resp. bounds. However one can obtain such witness candidates and bounds (and even realizing function(al)s) for a weakened version of A, namely its so-called Herbrand normal form A^H:

Definition 1.7 $A \equiv (\forall y_0)\exists x_1 \forall y_1 \ldots \exists x_n \forall y_n A_0(y_0, x_1, y_1, \ldots, x_n, y_n)$. *Then the Herbrand normal form of A is defined as*

$$A^H :\equiv \exists x_1, \ldots, x_n A_0(y_0, x_1, f_1 x_1, \ldots, x_n, f_n x_1 \ldots x_n),$$

where $f_1, \ldots, f_n$ are new function symbols, called index functions.

Remark 1.8 *In theories with function variables and function quantifiers we take the Herbrand normal form of A to be*

$$A^H :\equiv \forall (y_0), f_1, \ldots, f_n \exists x_1, \ldots, x_n A_0(y_0, x_1, f_1 x_1, \ldots, x_n, f_n x_1 \ldots x_n).$$

A and A^H are equivalent with respect to logical validity, i.e.

$$\models A \Leftrightarrow \models A^H,$$

but are not logically equivalent since in general

$$\text{PL} \not\vdash A^H \to A$$

However the converse implication holds

$$\text{PL} \vdash A \to A^H.$$

Let PL^2 denote the extension of PL obtained by the addition of n-ary function variables (for every n) and function quantifiers.
Let furthermore AC denote the schema of choice

$$AC: \ \forall \underline{x} \exists y\, A(\underline{x}, y) \rightarrow \exists f \forall \underline{x}\, A(\underline{x}, f\underline{x}) \ (\underline{x} = x_1 \dots x_n),$$

then it is an easy exercise to show that

$$\mathrm{PL}^2 + AC \vdash A \leftrightarrow A^H.$$

We now consider again the sentence

$$A \equiv \forall x \exists y \forall z (Pxy \vee \neg Pxz),$$

where P is some predicate symbol. In contrast to A, the Herbrand normal form A^H of A

$$A^H \equiv \exists y (P(x,y) \vee \neg P(x, gy))$$

allows an interpretation in form of a list of candidates (uniformly in x, g) for '$\exists y$', namely (x, gx) and also (c, gc) for any constant c does the job since the disjunction

$$A^{H,D} :\equiv (P(x,c) \vee \neg P(x, gc)) \vee (P(x, gc) \vee \neg P(x, g(g(c))))$$

is a tautology.
A tautology remains a tautology if we replace all occurrences of a term s by a variable y: Replace gc by y and $g(g(c))$ by z. Then $A^{H,D}$ becomes

$$A^D :\equiv (P(x,c) \vee \neg P(x,y)) \vee (P(x,y) \vee \neg P(x,z)),$$

which still is a tautology. From A^D we can derive A by a so-called direct proof (which uses only appropriate quantifier introduction rules, the shift of quantifiers over $\vee$ and contraction):

$$P(x,c) \vee \neg P(x,y) \vee P(x,y) \vee \neg P(x,z)$$

$$\Downarrow (\forall\text{-introduction})$$

$$P(x,c) \vee \neg P(x,y) \vee \forall z(P(x,y) \vee \neg P(x,z))$$

$$\Downarrow (\exists\text{-introduction})$$

$$P(x,c) \vee \neg P(x,y) \vee \exists y \forall z(P(x,y) \vee \neg P(x,z))$$

$$\Downarrow (\forall\text{-introduction})$$

$$\forall y(P(x,c) \vee \neg P(x,y)) \vee \exists y \forall z(P(x,y) \vee \neg P(x,z))$$

$$\Downarrow (\exists\text{-introduction})$$

$$\exists u \forall y(P(x,c) \vee \neg P(x,y)) \vee \exists y \forall z(P(x,y) \vee \neg P(x,z))$$

$$\Downarrow (\text{ contraction})$$

$$\exists y \forall z(P(x,y) \vee \neg P(x,z))$$

$$\Downarrow (\forall\text{-introduction})$$

$$\forall x \exists y \forall z(P(x,y) \vee \neg P(x,z))$$

Definition 1.9 *A formula A in the language of first-order predicate logic with equality ($PL_=$) is called a quasi-tautology if it is a tautological consequence of instances of $=$-axioms.*

Theorem 1.10 (Herbrand's Theorem)
Let $A \equiv \exists x_1 \forall y_1 \ldots \exists x_n \forall y_n A_0(x_1, y_1, \ldots, x_n, y_n)$. Then the following holds: $PL \vdash A$ iff there are terms $t_{1,1}, \ldots, t_{1,k_1}, \ldots, t_{n,1}, \ldots, t_{n,k_n}$ (built up out of the constants and variables of A and the index functions used for the formation of A^H) such that

$$A^{H,D} :\equiv \bigvee_{j_1=1}^{k_1} \ldots \bigvee_{j_n=1}^{k_n} A_0(t_{1,j_1}, f_1(t_{1,j_1}), \ldots, t_{n,j_n}, f_n(t_{1,j_1}, \ldots, t_{n,j_n}))$$

is a tautology.

The terms $t_{i,j}$ can be extracted constructively from a given PL-proof of A and conversely one can construct a PL-proof for A out of a given tautology $A^{H,D}$. The theorem holds for $PL_=$ if 'tautology' is replaced by 'quasi-tautology'.

Proof: See e.g. [60]. $\Box$

The most difficult part of the proof of Herbrand's theorem is the construction of the Herbrand terms $t_{i,j}$. The reverse direction for PL follows similar to the special case treated above: the terms f_i-terms in $A^{H,D}$ are replaced by new variables (starting from terms of maximal size) yielding an index-function-free Herbrand disjunction A^D. From this A is derived by a direct proof. For PL$_=$ the reverse direction is more complicated to establish since also instances of equality axioms $\underline{x} = \underline{y} \to f_i\underline{x} = f_i\underline{y}$ are now allowed in the proof of $A^{H,D}$.

In applications, the Herbrand disjunction A^D without index function has been particular useful (see [47],[53]). Although it is quite complicated to write down the general form of such a disjunction it is easy for Π^0_3-sentences (which is sufficient for many applications in mathematics):

For sentences $A \equiv \forall x \exists y \forall z\, A_0(x, y, z)$, A^D can always be written in the form

$$A_0(x, t_1, b_1) \vee A_0(x, t_2, b_2) \vee \ldots \vee A_0(x, t_k, b_k),$$

where the b_i are new variables and t_i does not contain any b_j with $i \leq j$ (see [47]).

Herbrand's theorem immediately extends to so-called open theories, i.e. first-order theories $\mathcal{T}$ whose non-logical axioms $G_1, \ldots, G_n$ are all purely universal ($G_i \equiv \forall a_i G_0^i(a_i)$), if '(quasi-)tautology' is replaced by 'tautological consequence of instances of equality axioms and the non-logical axioms'.

Proof: Apply Herbrand's theorem for logic to

$$\tilde{A} :\equiv \exists x_1 \forall y_1 \ldots \exists x_n \forall y_n \exists a_1, \ldots, a_m (\bigwedge_{i=1}^{n} G_0^i(a_i) \to A_0(x_1, y_1, \ldots, x_n, y_n)).$$

$\Box$

Warning: For the extension of Herbrand's theorem to open theories $\mathcal{T}$ it is important that the index function used in defining A^H are new and do not occur in the non-logical axioms. In particular if we have a schema of purely universal axioms then in the statement of Herbrand's theorem this schema is always understood with respect to the original language (without the index functions). Otherwise the reverse direction in Herbrand's theorem in general would fail (see [29] for a discussion of this and related matters thereby pointing out errors in the literature).

In general Herbrand's theorem in the form stated above does not hold for theories which are not open, e.g. it fails for PA.
However there are ways to extend the general idea behind Herbrand's theorem to theories like PA and beyond: in these lectures we will discuss Gödel's functional interpretation and the so-called no-counterexample interpretation (due to G. Kreisel [42],[43]). We conclude the first lecture by motivating the latter:

Lets consider again the example

$$A \equiv \forall x \exists y \forall z (P(x,y) \vee \neg P(x,z)).$$

If P is formulated in some theory like PA with decidable prime formulas, e.g. if $P(x,y) \equiv T(x,x,y)$, then we can realize the Herbrand normal form A^H of A instead of using a disjunction also by a computable functional of type level 2 which is defined by cases:

$$\Phi(x,g) := \begin{cases} x & \text{if } \neg T(x,x,gx) \\ gx & \text{otherwise.} \end{cases}$$

From this definition it easily follows that

$$\forall x, g (T(x,x,\Phi xg) \vee \neg T(x,x,g(\Phi xg)).$$

If A is not provable in PL but e.g. in PA we no longer can expect that functionals as simple as Φ above will be sufficient. In addition to the use of definition by cases we also have to allow certain recursive definitions whose complexity depends on the strength of the theory in which A is proved. In these lectures we will characterize in the case of PA (and subsystems) what functionals are needed.

Definition 1.11 *Let* $A \equiv \exists x_1 \forall y_1 \ldots \exists x_n \forall y_n A_0(x_1, y_1, \ldots, x_n, y_n)$. *If a tuple of functionals* $\Phi_1, \ldots, \Phi_n$ *realizes the Herbrand normal form* A^H *of* A, *i.e. if*

$$\forall \underline{f} A_0(\Phi_1 \underline{f}, f_1(\Phi_1 \underline{f}), \ldots, \Phi_n \underline{f}, f_n(\Phi_1 \underline{f}, \ldots, \Phi_n \underline{f}))$$

is true (where $\underline{f} = f_1, \ldots, f_n$*), then we say that* $\underline{\Phi}(= \Phi_1, \ldots, \Phi_n)$ *satisfies the no-counterexample interpretation of* A *(short:* $\underline{\Phi}$ *n.c.i.* A*).*
If A *starts with a universal quantifier* $\forall y_0$ *then* y_0 *is considered as a 0-place index function and* Φ_i *now depends on* y_0 *and* $\underline{f}$.

Motivation for the name 'no-counterexample interpretation':

Let A be as above. Then $\neg A$ is equivalent to

$$\forall x_1 \exists y_1 \ldots \forall x_n \exists y_n \neg A_0(x_1, y_1, \ldots, x_n, y_n).$$

So a counterexample to A is given by functions $f_1, \ldots, f_n$ such that

$$(+)\ \forall \underline{x} \neg A_0(x_1, f_1(x_1), \ldots, x_n, f_n(x_1, \ldots, x_n))$$

holds. Hence functionals $\underline{\Phi}$ satisfying the n.c.i. of A produce a counterexample to (+) i.e. to the existence of counterexample functions $f_1, \ldots, f_n$.

Definition 1.12 *A functional* F *of type level* ≤ 2 *is called* **primitive recursive in the sense of Kleene** *if it can be defined by the following schemas (*$\underline{x} = x_0, \ldots, x_{p-1}$ *is a list of number variables and* $\underline{f} = f_0, \ldots, f_{q-1}$ *is a list of function variables):*

(i) (Identity) $F(\underline{x}, \underline{f}) = x_i$ *(for* $i < p$*),*

(ii) (Function application) $F(\underline{x}, \underline{f}) = f_i(x_{j_0}, \ldots, n_{j_{l-1}})$
(for $i < q$ *and* $j_0, \ldots, j_{l-1} < p$*),*

(iii) (Successor) $F(\underline{x}, \underline{f}) = x_i + 1$ *(for* $i < p$*),*

(iv) *(Substitution)*
$F(\underline{x}, \underline{f}) = G(H_0(\underline{x}, \underline{f}), \ldots, H_{l-1}(\underline{x}, \underline{f}), \lambda y.K_0(y, \underline{x}, \underline{f}), \ldots, \lambda y.K_{j-1}(y, \underline{x}, \underline{f}))$,

(v) *(Primitive recursion)*
$F(0, \underline{x}, \underline{f}) = G(\underline{x}, \underline{f}), \quad F(y+1, \underline{x}, \underline{f}) = H(F(y, \underline{x}, \underline{f}), y, \underline{x}, \underline{f})$.

Exercises:

1) Consider $\Psi(x) :=$
$|\{n \in \mathbb{N} : 1 \leq n \leq x \wedge n$ is not divisible by any square number $\neq 1\}|$.
Show that $\Psi(x) \geq x - \sum_{\substack{p \text{ prime} \\ p \leq x}} [\frac{x}{p^2}]$ and use this to show that there are infinitely many primes. Use this proof to obtain an upper bound $g(j)$ for the next prime p_{j+1} as in the 3. proof of this statement above. Can you improve the bound we obtained from the latter (see Hacks [22])?

2) Let $(a_n)_{n\in\mathbb{N}}$ be a sequence of rational numbers in $[0, 1]$ with $\forall n \in \mathbb{N}(a_{n+1} \leq a_n)$. Since rational numbers can be coded by natural numbers one can consider (a_n) as a number theoretic function. The order relation $\leq$ and the usual arithmetical operations between rational numbers are primitive recursive in their codes. Construct a primitive recursive functional Φ which satisfies (uniformly in (a_n)) the n.c.i. of the theorem

$$\forall x \in \mathbb{N} \exists y \in \mathbb{N} \forall z \in \mathbb{N} \, (z > y \to a_y - a_z \leq \frac{1}{x+1}).$$

3) Construct primitive recursive functionals $\underline{\Phi}$ which satisfy the n.c.i. of (some prenex normal form of) the second-order axiom of Σ^0_1-induction:

$$\forall f (\exists y (f0y = 0) \wedge \forall x (\exists y (fxy = 0) \to \exists y (fx'y = 0)) \to \forall x \exists y (fxy = 0))$$

uniformly as a functional in f and the index functions.

Suggested further reading

1) On the general program of unwinding proofs: [47],[48],[49].

2) On Herbrand's theorem: [8],[14],[29],[47],[60].

3) On the no-counterexample interpretation: [14],[40],[42],[43].

Chapter 2

Intuitionistic logic and arithmetic in all finite types

In the following we formulate an axiomatic system for intuitionistic first-order predicate logic IL. The particular axiomatization we choose is due to [17] and particular suited to carry out proof interpretations inductively over the proof tree.

Intuitionistic first-order predicate logic IL

I. The language $\mathcal{L}(\mathrm{IL})$ of IL:

   As logical constants we use $\wedge, \vee, \rightarrow, \perp$ (absurdity of 'falsum'), $\exists, \forall$. $\mathcal{L}(\mathrm{IL})$ contains variables $x, y, z, \ldots$ (which can be free or bound). Furthermore we have function constants $f_1, f_2, f_3, \ldots$ and predicate constants $P_1, P_2, P_3, \ldots$ (both n-ary for every n).

   Abbreviations:

   $\neg A :\equiv A \rightarrow \perp$, $A \leftrightarrow B :\equiv (A \rightarrow B) \wedge (B \rightarrow A)$.

II. **Axioms of IL:**

   (i) $A \vee A \rightarrow A$, $A \rightarrow A \wedge A$ (axioms of contraction)

   (ii) $A \rightarrow A \vee B$, $A \wedge B \rightarrow A$ (axioms of weakening)

(iii) $A \vee B \to B \vee A$, $A \wedge B \to B \wedge A$ (axioms of permutation)

(iv) $\perp \to A$ (ex falso quodlibet)

(v) $\forall x\, A(x) \to A(t)$, $A(t) \to \exists x\, A(x)$, where t is free for x in A (quantifier axioms).

1) **Rules of IL:**

(i)

$$\frac{A\,,\, A \to B}{B}, \quad \frac{A \to B\,,\, B \to C}{A \to C}$$

(modus ponens and syllogism)

(ii)

$$\frac{A \wedge B \to C}{A \to (B \to C)}, \quad \frac{A \to (B \to C)}{A \wedge B \to C}$$

(exportation and importation)

(iii)

$$\frac{A \to B}{C \vee A \to C \vee B} \text{ (expansion)}$$

(iv)

$$\frac{B \to A(x)}{B \to \forall x\, A(x)}, \quad \frac{A(x) \to B}{\exists x\, A(x) \to B}, \text{ where } x \text{ is not free in } B$$

(quantifier rules).

The Brouwer-Heyting-Kolmogorov ('BHK') proof interpretation of the intuitionistic logical constants[1]

This interpretation is an informal attempt to explain the meaning of the logical constants of IL in terms of proof constructions:[2]

[1] Our exposition makes use of [58].

[2] 'Proof' is understood here as 'verification by a construction' and not as a formal proof in some fixed deductive framework like HA below.

(i) There is no proof for $\bot$.

(ii) A proof of $A \wedge B$ is a pair (q, r) of proofs, where q is a proof of A and r is a proof of B.

(iii) A proof of $A \vee B$ is a pair (n, q) consisting of an integer n and a proof q which proves A if $n = 0$ and resp. B if $n = 1$.

(iv) A proof p of $A \to B$ is a construction which transforms any hypothetical proof q of A into a proof $p(q)$ of B.

(v) A proof of $\forall x\, A(x)$ is a construction which produces for every construction c_d of an element d of the domain a proof $p(c_d)$ of $A(d)$.

(vi) A proof p of $\exists x\, A(x)$ is a pair (c_d, q), where c_d is the construction of an element d of the domain and q is a proof of $A(d)$.

Exercise 2.1 *1) Convince yourself that the axioms and rules of IL are sound under this interpretation.*

2) Convince yourself that $\neg\neg A \to A$ in general is not valid under this interpretation.

Discussion: There is one problem with the BHK-interpretation: from a strictly constructive point of view one would like to have a constructive verification of 'p is a proof of A' in case this is true, i.e. one would like to recognize a proof if one sees it. For $(i), (ii), (iii), (vi)$ there is no problem with this requirement. But for the universal statements in $(iv), (v)$ one would need an additional clause as suggested by Kreisel in [46]:

(iv)' A proof p of $A \to B$ is a pair (r, q), where q is a construction which transforms any hypothetical proof s of A into a proof $q(s)$ of B and r is a proof which verifies that q is such a construction.

(v)' A proof p of $\forall x\, A(x)$ is a pair (r, q) where q is a construction which produces for every construction c_d of an element d of the domain a proof $q(c_d)$ of $A(d)$ and r is a proof of the fact that q is such a construction.

Remark 2.2 *There are various ways to formalize the idea behind the BHK-interpreta-tion which give rise to various forms of so-called realizability interpretations. The first version of realizability, the so-called recursive realizability, was introduced by Kleene in [26]. In these lectures we will focus on a typed variant of Kleene's type-free interpretation which is called 'modified realizability' and is due to Kreisel [44],[45].*

Intuitionistic ('Heyting'-)arithmetic HA

$\mathcal{L}$(HA) contains the logical constants of $\mathcal{L}$(IL), number variables $x, y, z, \ldots$, a constant 0 (zero), a unary function constant S (successor), function constants for all primitive recursive functions (more precisely for all derivations of primitive recursive functions) and a single binary predicate =.

Axioms and rules of HA:

(i) axioms and rules of IL (based on $\mathcal{L}$(HA))

(ii) =-axioms: $x = x$, $x = y \to y = x$, $x = y \wedge y = z \to x = z$, $\underline{x} = \underline{y} \to f(\underline{x}) = f(\underline{y})$ for every n-ary function constant f $(\underline{x} = x_1, \ldots, x_n, \underline{y} = y_1, \ldots, y_n)$

(iii) successor axioms:

$$\begin{cases} Sx \neq 0, \\ Sx = Sy \to x = y \end{cases}$$

(iv) defining equations for the primitive recursive functions

(v) axiom schema of complete induction

$$\text{IA}: \; A(0) \wedge \forall x (A(x) \to A(Sx)) \to \forall x\, A(x)$$

for every formula $A \in \mathcal{L}$(HA)

Convention: We often write x' or $x + 1$ for Sx.

Remark 2.3 *In HA we may identify* $\perp$ *with* $0 = 1$.

Instead of the axiom schema IA we could have formulated HA equivalently using the rule of induction

$$\text{IR:}\ \frac{A(0)\ ,\ A(x) \to A(Sx)}{A(x)}.$$

Exercise 2.4 *Show that IA is derivable from IR. Compare the complexity of the induction formula* $\tilde{A}(x)$ *of the IR-instance needed to prove an IA-instance with induction formula* $A(x)$ *with that of* $A(x)$.

Extensional intuitionistic ('Heyting'-) arithmetic E-HA$^\omega$ in all finite types

The set $\mathbf{T}$ of all finite types is generated inductively by the clauses

$$(i)\ 0 \in \mathbf{T},\ (ii)\ \rho, \tau \in \mathbf{T} \Rightarrow \tau(\rho) \in \mathbf{T}.$$

The type 0 is the type natural numbers. Objects of type $\tau(\rho)$ are functions which map objects of type ρ to objects of type τ (some authors write $(\rho)\tau$ or $\rho \to \tau$ instead of $\tau(\rho)$.
We often omit brackets which are uniquely determined and write e.g. $0(00)$ instead of $0(0(0))$.

The set $\mathbf{P} \subset \mathbf{T}$ of pure types is defined by

$$(i)\ 0 \in \mathbf{P},\ (ii)\ \rho \in \mathbf{P} \Rightarrow 0(\rho) \in \mathbf{P}.$$

Pure types are often denoted by natural numbers: $0(n) := n + 1$ (e.g. $00 = 1, 0(00) = 2$).
The type level or degree $deg(\rho)$ of a type ρ is defined as

$$deg(0) := 0,\ deg(\tau(\rho)) := \max(deg(\tau), deg(\rho) + 1)$$

(note that for pure types ρ, $deg(\rho)$ is just the number which denotes ρ).

Objects of type ρ with $deg(\rho) > 1$ are usually called functionals.

The language $\mathcal{L}$(E-HA$^\omega$) of E-HA$^\omega$ is based on a many-sorted version IL$^\omega$ of IL which contains variables $x^\rho, y^\rho, z^\rho, \ldots$ and quantifiers $\forall x^\rho$, $\exists y^\rho$ for every type ρ. As constants E-HA$^\omega$ contains 0^0 (zero), S^{00} (successor), $\Pi^{\rho\tau\rho}_{\rho,\tau}$ (projector), $\Sigma_{\delta,\rho,\tau}$ (combinator of type $\tau\delta(\rho\delta)(\tau\rho\delta)$) and recursor constants R_ρ of type $\rho(\rho0\rho)\rho0$ for all $\delta, \rho, \tau \in \mathbf{T}$. Furthermore $\mathcal{L}$(E-HA$^\omega$) contains a binary predicate constant $=_0$ for equality between object of type 0.

Terms of E-HA$^\omega$ are built up by

(i) constants c^ρ and variables x^ρ of type ρ are terms of type ρ

(ii) if $t^{\tau\rho}$ is a terms of type $\tau\rho$ and s^ρ is a term of type ρ, then (st) is a term of type τ.

Formulas of E-HA$^\omega$ are built up by

(i) prime formulas (also called 'atomic formulas') $s =_0 t$ are formulas (where s^0, t^0 are terms of type 0)

(ii) if A, B are formulas, then also $A \wedge B$, $A \vee B$ and $A \to B$ are formulas

(iii) if $A(x^\rho)$ is a formula, then also $\forall x^\rho A(x)$ and $\exists x^\rho A(x)$ are formulas.

Abbreviations:

1) Higher type equations $s =_\rho t$ between terms s, t of type $\rho = 0(\rho_k)\ldots(\rho_1)$ (where $k \geq 1$) are abbreviations for

$$\forall y_1^{\rho_1}, \ldots, y_k^{\rho_k}(sy_1 \ldots y_k =_0 ty_1 \ldots y_k),$$

where $y_1, \ldots, y_k$ are variables which don't occur in s, t.

2) As before: $\neg A :\equiv A \to \bot$, where $\bot :\equiv (0 = 1)$.
$A \leftrightarrow B :\equiv (A \to B) \wedge (B \to A)$.

Axioms and rules of E-HA$^\omega$

(i) all axioms and rules of IL$^\omega$

(ii) equality axioms for $=_0$

(iii) higher type extensionality:

$$E_\rho : \ \forall z^\rho, x_1^{\rho_1}, y_1^{\rho_1}, \ldots, x_k^{\rho_k}, y_k^{\rho_k} (\bigwedge_{i=1}^{k} (x_i =_{\rho_i} y_i) \to z\underline{x} =_0 z\underline{y}),$$

where $\rho = 0\rho_k \ldots \rho_1$

(iv) successor axioms

(v) induction schema

$$\text{IA:} \ \ A(0) \wedge \forall x^0 (A(x) \to A(Sx)) \to \forall x^0 A(x),$$

where $A(x^0)$ is an arbitrary formula of E-HA$^\omega$

(vi) axioms for $\Pi_{\rho,\tau}, \Sigma_{\delta,\rho,\tau}$ and R_ρ:

$$\begin{array}{lcl} (\Pi) & : & \Pi_{\rho,\tau} x^\rho y^\tau =_\rho x^\rho, \\ (\Sigma) & : & \Sigma_{\delta,\rho,\tau} xyz =_\tau xz(yz) \quad (x^{\tau\rho\delta},\ y^{\rho\delta},\ z^\delta), \\ (R) & : & \begin{cases} R_\rho 0 yz =_\rho y \\ R_\rho (Sx^0) yz =_\rho z(R_\rho xyz)x \quad (y^\rho,\ z^{\rho 0 \rho}). \end{cases} \end{array}$$

Definition 2.5 *Later on we will need also a variant WE-HA$^\omega$ of E-HA$^\omega$, where the extensionality axioms E_ρ are weakened to a quantifier-free rule of extensionality*

$$\text{QF-ER:} \ \ \frac{A_0 \to s =_\rho t}{A_0 \to r[s] =_\tau r[t]},$$

where $s^\rho, t^\rho, r[x^\rho]^\tau$ are terms of WE-HA$^\omega$ ($\rho, \tau \in \mathbf{T}$ arbitrary).

Warning: WE-HA$^\omega$ does not satisfy the deduction theorem.

WE-HA$^\omega$ allows the definition of λ-abstraction in the following sense:

Lemma 2.6 For every term $t[x^\rho]^\tau$ one can construct in WE-HA$^\omega$ a term $\lambda x^\rho.t[x]$ of type $\tau\rho$ (with $\mathrm{FV}(\lambda x^\rho.t[x]) = \mathrm{FV}(t[x]) \setminus \{x\}$)[3] such that

$$\text{WE-HA}^\omega \vdash (\lambda x^\rho.t[x])(s^\rho) =_\tau t[s].$$

Proof: Define

$$\begin{aligned}
\lambda x.x &:= \Sigma\Pi\Pi,\\
\lambda x.t &:= \Pi t, \text{ if } x \notin \mathrm{FV}(t)\\
\lambda x.(ts) &:= \Sigma(\lambda x.t)(\lambda x.s), \text{ if } x \in \mathrm{FV}(ts)
\end{aligned}$$

(here Π, Σ of suitable types). $\Box$

Exercise:
It is known that the function $\alpha(x, y)$ defined by the equations

$$(*) \begin{cases} \alpha(0, y) = y' \\ \alpha(x', 0) = \alpha(x, 1) \\ \alpha(x', y') = \alpha(x, \alpha(x', y)) \end{cases}$$

is not primitive recursive (in the sense of Kleene). In fact α is a variant due to R. Peter of the well-known Ackermann function.

Show that α is definable in in WE-HA$^\omega$ by a closed term $t^{0(0)(0)}$ (i.e. WE-HA$^\omega$ proves the equations $(*)$ for t).

Suggested further reading:
See [64] for further information on the BHK-interpretation. For more information on (W)E-HA$^\omega$ and its variants see [65].

[3] $\mathrm{FV}(t)$ ($\mathrm{FV}(A)$) denotes the set of all free variables of t (A).

Chapter 3

Modified realizability

Definition 3.1 (modified realizability) *For each formula A of E-HA$^\omega$ we define a formula $\underline{x}$ mr A (in words: '$\underline{x}$ modified realizes A') of E-HA$^\omega$ whose free variables are contained in that of A and $\underline{x}$, where $\underline{x}$ is a – possibly empty – tuple of variables which do not occur free in A. The length of $\underline{x}$ and the types of these variables are determined by the* **logical** *structure of A, since the definition of $\underline{x}$ mr A proceeds by induction over the logical structure of A:*

(i) $\underline{x}$ *mr* $A :\equiv A$ *with the empty tuple* $\underline{x}$*, if A is a prime formula.*

(ii) $\underline{x}, \underline{y}$ *mr* $(A \wedge B) :\equiv \underline{x}$ *mr* $A \wedge \underline{y}$ *mr* B.

(iii) $z^0, \underline{x}, \underline{y}$ *mr* $(A \vee B) :\equiv [(z = 0 \to \underline{x}$ *mr* $A) \wedge (z \neq 0 \to \underline{y}$ *mr* $B)]$.

(iv) $\underline{y}$ *mr* $(A \to B) :\equiv \forall \underline{x}(\underline{x}$ *mr* $A \to \underline{y}\,\underline{x}$ *mr* $B)$.

(v) $\underline{x}$ *mr* $(\forall y^\rho A(y)) :\equiv \forall y^\rho(\underline{x}y$ *mr* $A(y))$.

(vi) $z^\rho, \underline{x}$ *mr* $(\exists y^\rho A(y)) :\equiv \underline{x}$ *mr* $A(z)$.

Definition 3.2 *1) A formula $A \in \mathcal{L}$(E-HA)$^\omega$ is called $\exists$-free if it is built up from prime formulas by means of $\wedge, \to$ and $\forall$ only.*

2) A formula $A \in \mathcal{L}$(E-HA)$^\omega$ is called negative if it is built up from negated prime formulas by means of $\wedge, \to$ and $\forall$ only.

Remark 3.3 *In* WE-HA$^\omega$ *all prime formulas* P *are decidable and therefore* $\neg\neg P \leftrightarrow P$ *is provable in* WE-HA$^\omega$. *Hence every* $\exists$*-free formula is equivalent to a negative formula in* WE-HA$^\omega$.

Remark 3.4 *1) For* $\exists$*-free formulas* A *we have* $(\underline{x}\ mr\ A) \equiv A$ *with* $\underline{x}$ *being the empty tuple.*

2) $(\underline{x}\ mr\ A)$ *is always an* $\exists$*-free formula.*

We will also need a variant 'modified realizability with truth' mrt of mr:

Definition 3.5 $\underline{x}\ mrt\ A$ *is defined analogously to* $\underline{x}\ mr\ A$ *except that clause* (iv) *is replaced by*

$$(iv)'\ \underline{y}\ mrt\ (A \to B) :\equiv \forall \underline{x}(\underline{x}\ mrt\ A \to \underline{y}\,\underline{x}\ mrt\ B) \wedge (A \to B).$$

The name 'modified realizability with truth' is motivated by the following

Lemma 3.6 WE-HA$^\omega \vdash (\underline{x}\ mrt\ A) \to A$, *for every formula* A.

Proof: Straightforward. $\Box$

Exercise 3.7 WE-HA$^\omega \vdash (\underline{x}\ mrt\ \neg A) \leftrightarrow \neg A$ *for every formula* A.

The schema of choice AC:= $\bigcup\limits_{\rho,\tau\in\mathbf{T}} \{\,\mathrm{AC}^{\rho,\tau}\}$ is given by

$$\mathrm{AC}^{\rho,\tau}:\ \forall x^\rho \exists y^\tau A(x,y) \to \exists Y^{\tau\rho}\forall x^\rho A(x, Yx),$$

where A is an arbitrary formula of E-HA$^\omega$.

The independence-of-premise-schema $\mathrm{IP}^\omega_{ef} := \bigcup\limits_{\rho\in\mathbf{T}}\{\mathrm{IP}^\rho_{ef}\}$ for $\exists$-free formulas is given by

$$\mathrm{IP}^\rho_{ef}:\ (A \to \exists x^\rho B(x)) \to \exists x^\rho (A \to B(x)),$$

where A is $\exists$-free and doesn't contain x free.

Theorem 3.8 (soundness for mr**)** *Let* A *be an arbitrary formula in* $\mathcal{L}$(E-HA$^\omega$)*. Then the following rule holds*

$$E\text{-}HA^\omega + AC + IP^\omega_{ef} \vdash A \;\Rightarrow\; E\text{-}HA^\omega \vdash \underline{t}\ mr\ A,$$

where $\underline{t}$ *is a suitable tuple of terms of* E-HA^ω *with* $FV(\underline{t}) \subseteq FV(A)$ *which can be extracted from a given proof of* A*.*

Proof: Induction on the length of the derivation of A:
1) Axioms. $A \vee A \to A$: If $z^0, \underline{x}, \underline{y}\ mr\ (A \vee A)$, then $\underline{t}z^0\underline{x}\,\underline{y}\ mr\ A$, where $\underline{t}$ such that $\underline{t}z^0\underline{x}\,\underline{y} := \begin{cases} \underline{x}, & \text{if } z = 0 \\ \underline{y}, & \text{if } z \neq 0 \end{cases}$
(t_i can easily be defined using R_0: exercise!). Hence $\underline{t}\ mr\ (A \vee A \to A)$.
$A \to A \wedge A$ is realized by $\lambda\underline{x}.(\underline{x}, \underline{x})$.
$A \to A \vee B$: Let $\underline{x}\ mr\ A$, then $(0, \underline{x}, \underline{\mathcal{O}})\ mr\ (A \vee B)$ and hence $\lambda\underline{x}.[0, \underline{x}, \underline{\mathcal{O}}]\ mr\ (A \to A \vee B)$ (here $\underline{\mathcal{O}}$ is a suitable tuple $\mathcal{O}_1^{\rho_1}, \ldots, \mathcal{O}_k^{\rho_k}$ with suitable types ρ_i so that $\underline{\mathcal{O}}\ mr\ B$ is syntactically correct[1]).
$A \wedge B \to A$: if $(\underline{x}, \underline{y})\ mr\ A \wedge B$, then $\underline{x}\ mr\ A$. Hence
$\lambda\underline{x}, \underline{y}.\underline{x}\ mr\ (A \wedge B \to A)$.
$\perp\ \to A$ is realized by $\underline{\mathcal{O}}\ mr(\perp\ \to A)$ such that $\underline{\mathcal{O}}\ mr\ A$ is syntactically correct.
$A \vee B \to B \vee A$ is realized by $\lambda z^0, \underline{y}, \underline{x}.[\overline{sg}(z), \underline{x}, \underline{y}]$, where

$$\overline{sg}(z) := \begin{cases} 0^0, & \text{if } z \neq 0 \\ 1^0, & \text{otherwise.} \end{cases}$$

$A \wedge B \to B \wedge A$ is realized by $\lambda\underline{y}, \underline{x}.\underline{x}, \underline{y}$.
$\forall x^\rho A(x) \to A(t^\rho)$: Let $\underline{y}\ mr\ \forall x\, A(x)$. Then $\underline{y}(t)\ mr\ A(t)$. Hence
$\lambda\underline{y}.\underline{y}(t)\ mr\ (\forall x\, A(x) \to A(t))$.
$A(t^\rho) \to \exists x^\rho A(x)$ is realized by $\lambda\underline{y}.[t, \underline{y}]$, where $\underline{y}$ is a tuple of variables such that $\underline{y}\ mr\ A(t)$ is well-formed.
2) Rules. $\frac{A,\, A\to B}{B}$: Assume $\underline{t}\ mr\ A$ and $\underline{s}\ mr\ (A \to B)$. Let $\underline{r}$ be the terms which result from $\underline{t}(\underline{s})$ be replacing all free variables $\underline{a}$ which occur in A but not in B by $\underline{\mathcal{O}}$. Then $\underline{r}\ mr\ B$.

[1]Here for $\rho = 0(\rho_k)\ldots(\rho_1)$, $\mathcal{O}^\rho := \lambda x_1^{\rho_1}, \ldots, x_k^{\rho_k}.0^0$.

$\frac{A\to B,\, B\to C}{A\to C}$: $\underline{s}$ mr $(A \to B)$, $\underline{t}$ mr $(B \to C)$. If $\underline{x}$ mr A, then $\underline{s}\,\underline{x}$ mr B and hence $\underline{t}(\underline{s}\,\underline{x})$ mr C. Thus $\lambda\underline{x}.\underline{t}(\underline{s}\,\underline{x})$ mr $(A \to C)$ (if necessary replace free variables which don't occur in $A \to C$ by $\mathcal{O}$).
$\frac{A\wedge B\to C}{A\to(B\to C)}$ and $\frac{A\to(B\to C)}{A\wedge B\to C}$ are trivially satisfied: use the terms from the premise for the conclusion.
$\frac{A\to B}{C\vee A\to C\vee B}$: Assume $\underline{t}$ mr $(A \to B)$ and $z^0, \underline{x}, \underline{y}$ mr $(C \vee A)$. Then either $z = 0$, $\underline{x}$ mr C or $z \neq 0$, $\underline{y}$ mr A. In the second case we have $\underline{t}\,\underline{y}$ mr B. Hence $\lambda z^0, \underline{x}, \underline{y}.[z^0, \underline{x}, \underline{t}\,\underline{y}]$ mr $(C \vee A \to C \vee B)$.
$\frac{B\to A(x^\rho)}{B\to\forall x^\rho A(x)}$: Assume $\underline{t}[x]$ mr $(B \to A(x))$ and $\underline{z}$ mr B.
Then $\lambda x.(\underline{t}[x]\underline{z})$ mr $\forall x\, A(x)$ and therefore $\lambda\underline{z}, x.(\underline{t}[x]\underline{z})$ mr $(B \to \forall x\, A(x))$.
$\frac{A(x^\rho)\to B}{\exists x^\rho A(x)\to B}$: Assume $\underline{t}[x]$ mr $(A(x) \to B)$ and $x, \underline{z}$ mr $\exists x\, A(x)$. Then $\underline{z}$ mr $A(x)$ and therefore $\underline{t}[x]\underline{z}$ mr B. Thus $\lambda x, \underline{z}.(\underline{t}[x]\underline{z})$ mr $(\exists x\, A(x) \to B)$.
3) Axioms for $=_0, S, \Pi, \Sigma, R$ **and** E_ρ**:** These axioms are all $\exists$-free and therefore realized by themselves.
4) The induction schema: Let $\underline{x}$ mr $A(0)$ and $\underline{y}$ mr $\forall z^0(A(z) \to A(z{+}1))$. Define $\underline{t}$ by simultaneous primitive recursion in higher types (which can be reduced to ordinary primitive recursion in higher types, exercise!) such that

$$\begin{cases} \underline{t}\,\underline{x}\,\underline{y}0 = \underline{x} \\ \underline{t}\,\underline{x}\,\underline{y}(z+1) = \underline{y}z(\underline{t}\,\underline{x}\,\underline{y}z). \end{cases}$$

By induction on z^0 one shows that $\underline{t}\,\underline{x}\,\underline{y}z$ mr $A(z)$ and hence $\underline{t}\,\underline{x}\,\underline{y}$ mr $\forall z\, A(z)$.
5) The interpretations for AC and IP^ω_{ef} are trivial (note that AC and IP^ω_{ef} are not needed to verify their mr-interpretation). □

Definition 3.9 ([65]) *The subset* Γ_1 *of formulas* $\in \mathcal{L}(E\text{-}HA^\omega)$ *is defined inductively by*

1) Prime formulas are in Γ_1.[2]

2) $A, B \in \Gamma_1 \Rightarrow A \wedge B, A \vee B, \forall x\, A(x), \exists x\, A(x) \in \Gamma_1$.

[2]Note that in our theories quantifier–free formulas can be written as prime formulas $s =_0 t$.

3) If A is $\exists$-free and $B \in \Gamma_1$, then $(\exists \underline{x} A \to B) \in \Gamma_1$.

Lemma 3.10 *For $A \in \Gamma_1$ we have*

$$\text{E-HA}^\omega \vdash (\underline{x}\ mr\ A) \to A.$$

Proof: Straightforward induction on the generation of Γ_1. $\square$

Corollary 3.11 *E-HA$^\omega$+ AC + IP$^\omega_{ef}$ is conservative over E-HA$^\omega$ with respect to formulas $A \in \Gamma_1$.*
In particular E-HA$^\omega$+ AC + IP$^\omega_{ef}$ is consistent relative to E-HA$^\omega$ since $(0 = 1) \in \Gamma_1$.

Remark 3.12 *One can show by much more complicated methods that E-HA$^\omega$+ AC is conservative over HA. For a 'neutral' version of E-HA$^\omega$ without extensionality this is due to [19],[20],[21] (see also [54],[57]). The extension to E-HA$^\omega$ is due to [2].*
In contrast to corollaries 3.11 and 3.14 (below), this result does not relativize to subsystems with restricted induction (see [39]).

Theorem 3.13 (soundness for mrt) *Let $H^\omega :=$ E-HA$^\omega \overset{+}{-}$ AC $\overset{+}{-}$ IP$^\omega_{ef}$ and A be an arbitrary formula in $\mathcal{L}(\text{E-HA}^\omega)$. The following rule holds*

$$H^\omega \vdash A \ \Rightarrow \ \ H^\omega \vdash \underline{t}\ mrt\ A,$$

where $\underline{t}$ is a suitable tuple of terms of E-HA$^\omega$ with $FV(\underline{t}) \subseteq FV(A)$ which can be extracted from a given proof of A.

Proof: The treatment of the logical axioms is analogous to the one for the mr-interpretation in the proof of of theorem 3.8. The same applies for the modus ponens. For the remaining rules the new second clause in the mrt-interpretation of the conclusion follows from the corresponding second clause(s) of the premise(s) using the same rule. Only in the case of the $\to$-introduction rule one has to be a bit more careful because of the nested implications:
By induction hypothesis we have terms $\underline{s}$ such that

$$\underline{s}\ mrt\ (A \wedge B \to C).$$

Hence

$$\forall \underline{x}, \underline{y}(\underline{x}\ mrt\ A \wedge \underline{y}\ mrt\ B \rightarrow \underline{s}\, \underline{x}\, \underline{y}\ C) \wedge (A \wedge B \rightarrow C),$$

which is equivalent to

$$\forall \underline{x}(\underline{x}\ mrt\ A \rightarrow \forall \underline{y}(\underline{y}\ mrt\ B \rightarrow \underline{s}\, \underline{x}\, \underline{y}\ mrt\ C)) \wedge (A \rightarrow (B \rightarrow C)).$$

By lemma 3.6 we have $(\underline{x}\ mrt\ A) \rightarrow A$. Hence

$$\forall \underline{x}(\underline{x}\ mrt\ A \rightarrow \forall \underline{y}(\underline{y}\ mrt\ B \rightarrow \underline{s}\, \underline{x}\, \underline{y}\ mrt\ C) \wedge (B \rightarrow C)) \wedge (A \rightarrow (B \rightarrow C))$$

and hence

$$\forall \underline{x}(\underline{x}\ mrt\ A \rightarrow \underline{s}\, \underline{x}\ mrt\ (B \rightarrow C)) \wedge (A \rightarrow (B \rightarrow C)),$$

i.e.

$$\underline{s}\ mrt\ A \rightarrow (B \rightarrow C).$$

We leave it as an exercise to the reader to adopt the mr-interpretation of the non-logical axioms and rules from the proof of theorem 3.8 to the mrt-interpretation. $\square$

Corollary 3.14 *Let $H^\omega := E\text{-}HA^\omega \overset{+}{-} AC \overset{+}{-} IP^\omega_{ef}$. Then the following rules hold:*

1)

$$H^\omega \vdash A \vee B \;\Rightarrow\; H^\omega \vdash A \text{ or } H^\omega \vdash B,$$

for closed formulas $A \vee B$ (disjunction property DP*)*

2)

$$H^\omega \vdash \exists x^\rho A(x) \;\Rightarrow\; H^\omega \vdash A(t),$$

for a suitable term t^ρ of H^ω with $FV(t) \subseteq FV(A) \setminus \{x^\rho\}$ (the special case of this property for closed formulas $\exists x^\rho A(x)$ is called existence property EP*)*

3)

$$H^\omega \vdash \forall x^\rho \exists y^\tau A(x,y) \;\Rightarrow\; H^\omega \vdash \exists Y^{\tau\rho} \forall x^\rho A(x, Yx)$$

(closure of H^ω under the rule of choice ACR*).*

4)

$$H^\omega \vdash (A \to \exists x^\rho B(x)) \;\Rightarrow\; H^\omega \vdash \exists x^\rho (A \to B(x^\rho)),$$

where A is $\exists$-free and doesn't contain x free (closure of H^ω under the rule of independence of premise for $\exists$-free formulas IPR^ω_{ef}).

Proof: 1) Suppose that $\mathrm{H}^\omega \vdash A \vee B$ for a closed formula $A \vee B$. By theorem 3.13 one finds closed terms $t^0, \underline{s}, \underline{r}$ such that

$$\mathrm{H}^\omega \vdash (t =_0 0 \to \underline{s} \; mrt \; A) \wedge (t \neq 0 \to \underline{r} \; mrt \; B).$$

In E-HA$^\omega$ the closed number term t^0 can be reduced (computed) to a numeral $\overline{n}$ and so

$$\mathrm{H}^\omega \vdash t =_0 \overline{n}.$$

The conclusion now follows from the fact that

$$\mathrm{H}^\omega \vdash \overline{n} = 0 \text{ or } \mathrm{H}^\omega \vdash \overline{n} \neq 0$$

and lemma 3.6.
2) By theorem 3.13 the assumptions yields terms $t^\rho, \underline{s}$ with $FV(t, \underline{s}) \subseteq FV(A) \setminus \{x\}$ such that

$$\mathrm{H}^\omega \vdash \underline{s} \; mrt \; A(t).$$

The claim now follows using lemma 3.6.
3) By 2) applied to the open formula $\exists y^\tau A(x^\rho, y^\tau)$ we get a term $t[x^\rho]^\tau$ such that

$$\mathrm{H}^\omega \vdash A(x, t[x]).$$

The conclusion follows by taking $Y := \lambda x^\rho . t[x^\rho]$.
4) Theorem 3.13 applied to

$$\mathrm{H}^\omega \vdash A \to \exists x^\rho B(x)$$

yields terms $t^\rho, \underline{s}$ such that (using that $(\underline{x}\ mrt\ A) \leftrightarrow A$ with the empty tuple $\underline{x}$ for $\exists$-free formulas A)

$$\mathrm{H}^\omega \vdash A \to \underline{s}\ mrt\ B(t)$$

and hence (by lemma 3.6)

$$\mathrm{H}^\omega \vdash A \to B(t)$$

and so

$$\mathrm{H}^\omega \vdash \exists x(A \to B(x)),$$

where x is not free in A. □

Proposition 3.15 *For all formulas A of* E-HA$^\omega$ *one has*

$$\textit{E-HA}^\omega + \ AC + \ IP^\omega_{ef} \vdash (\underline{x}\ mr\ A) \leftrightarrow A.$$

This also holds for mrt instead of mr.

Proof: Straightforward induction on the logical structure of A. □

Exercise:

The so-called Markov Principle in all finite types is the schema

$$\mathrm{MP}^\omega: \ \neg\neg\exists \underline{x}\, A_0(\underline{x}) \to \exists \underline{x}\, A_0(\underline{x}),$$

where A_0 is an arbitrary quantifier-free formula of WE-HA$^\omega$ and $\underline{x}$ is a tuple of variables of arbitrary types.

Show, using modified realizability, that already for $\underline{x} = x^0$, MP$^\omega$ is not derivable in E-HA$^\omega$+AC+IP$_{ef}$.

Suggested further reading: [65] (chapter III, section 4) provides a very concise and detailed treatment of modified realizability interpretation. For a general survey on various forms of realizability interpretations see [67]. For an application of modified realizability to the extraction of a program from a specific proof see [3].

Chapter 4

Majorizability and the fan rule

In this lecture we investigate an interesting structural property of the functionals which are definable by the closed terms of E-HA$^\omega$, namely their majorizability. We will indicate the far-reaching use one can make out of this observation by showing the closure of E-HA$^\omega \underline{+}$AC $\underline{+}$ IP$_{ef}$ under the so-called fan rule:

Definition 4.1 *For arbitrary $\rho \in \mathbf{T}$ we define the relation $x_1 \geq_\rho x_2$ between functionals x_1, x_2 of type ρ by induction on ρ:*

$$\begin{cases} x_1 \geq_0 x_2 :\equiv x_1 \geq x_2 \ (\textit{for the usual primitive recursive relation } \geq) \\ x_1 \geq_{\tau\rho} x_2 :\equiv \forall y^\rho (x_1 y \geq_\tau x_2 y). \end{cases}$$

Lemma 4.2 *Let $\rho = \tau\rho_k \ldots \rho_1$. Then*
WE-HA$^\omega \vdash x_1 \geq_\rho x_2 \leftrightarrow \forall y_1^{\rho_1}, \ldots, y_k^{\rho_k}(x_1\underline{y} \geq_\tau x_2\underline{y})$.

Definition 4.3 (W.A. Howard [24]) *We define the relation x^* maj$_\rho$ x (x^* majorizes x) between functionals of type ρ by induction on ρ:*

$$\begin{cases} x^* \ maj_0 \ x :\equiv x^* \geq_0 x, \\ x^* \ maj_{\tau\rho} \ x :\equiv \forall y^*, y(y^* \ maj_\rho \ y \to x^* y^* \ maj_\tau \ xy). \end{cases}$$

Lemma 4.4 *WE-HA$^\omega$ proves:*

(i) $\tilde{x}^* =_\rho x^* \wedge \tilde{x} =_\rho x \wedge x^* \; maj_\rho \; x \to \tilde{x}^* \; maj_\rho \; \tilde{x}$.

(ii) $x^* \; maj_\rho \; x \wedge x \geq_\rho y \to x^* \; maj_\rho \; y$.

(iii) *For* $\rho = \tau\rho_k \ldots \rho_1$:

$$x^* \; maj_\rho \; x \leftrightarrow \forall y_1^*, y_1, \ldots, y_k^*, y_k (\bigwedge_{i=1}^{k} (y_i^* \; maj_{\rho_i} \; y_i) \to x^* \underline{y}^* \; maj_\tau \; x\underline{y}).$$

Proof: Induction on the type respectively on k. $\Box$

Definition 4.5 *Define* $\varphi^{1(1)}$ *by recursion (using only* R_0*) such that*

$$\varphi(x^1, 0) =_0 x0$$
$$\varphi(x, z+1) =_0 \max\nolimits_0(\varphi(x,z), x(z+1)),$$

where $\max_0$ *is the usual (primitive recursively definable) maximum between natural numbers.*
We write $x^M := \lambda z^0.\varphi(x,z)$ *(note that* $x^M(z) = \max\limits_{i \leq z}(x(i))$*).*
This definition easily extends to finite types by λ*-abstraction: For* x *of type* $\rho 0$ *with* $\rho = 0\rho_k \ldots \rho_1$ *we define* $x^M := \lambda z, \underline{v}.\varphi(\lambda z.xz\underline{v}, z)$*, where* $\underline{v} = v_1^{\rho_1}, \ldots, v_k^{\rho_k}$.

One easily proves the following

Lemma 4.6
WE-HA$^\omega$ $\vdash \forall x^{\rho 0}(x^M 0 =_\rho x0 \wedge x^M(z+1) =_\rho \max_\rho(x^M z, x(z+1)))$,
where $\max_{\tau\rho}(x_1, x_2) :\equiv \lambda y^\rho. \max_\tau(x_1 y, x_2 y)$ *for complex types.*

Remark 4.7 *Using recursion of type* ρ *one can define* x^M *directly by iteration of* $\max_\rho$*. However our a bit more complicated approach shows that actually* R_0 *is sufficient.*

Lemma 4.8 *WE-HA$^\omega$* $\vdash \forall x^{\rho 0}, \tilde{x}^{\rho 0}(\forall n^0(\tilde{x}n \; maj_\rho \; xn) \to \tilde{x}^M \; maj_{\rho 0} x)$.

Proof: Let $\rho = 0\rho_k \ldots \rho_1$ and $\underline{v} = v_1^{\rho_1}, \ldots, v_k^{\rho_k}$. One easily shows by (quantifier-free) induction on n that

$$\forall n^0 (\forall m \leq n (\tilde{x}^M n \underline{v} \geq_0 \tilde{x} m \underline{v})).$$

Together with the assumption that $\forall n (\tilde{x} n \ maj_\rho \ xn)$ this yields

$$\forall n, m, \underline{v}^*, \underline{v} (n \geq_0 m \wedge \underline{v}^* \ maj \ \underline{v} \rightarrow \tilde{x}^M n \underline{v}^* \geq_0 x m \underline{v})$$

and hence $\tilde{x}^M \ maj_\rho \ x$. □

Corollary 4.9 *WE-HA*$^\omega \vdash \forall x^1 (x^M \ maj_1 \ x)$.

Proposition 4.10 (W.A. Howard [24]) *For each closed term t^ρ of WE-HA$^\omega$ one can construct a closed term t^{*^ρ} of WE-HA$^\omega$ such that*

$$\text{WE-HA}^\omega \vdash t^* \ maj_\rho \ t.$$

Proof: Induction on the structure of t:
Constants c: $0^0 \ maj_0 \ 0^0$, $S \ maj_1 \ S$. Using lemma 4.4(iii) we also have $\Pi_{\rho,\tau} \ maj \ \Pi_{\rho,\tau}$ and $\Sigma_{\delta,\rho,\tau} \ maj \ \Sigma_{\delta,\rho,\tau}$.
By induction on x^0 one shows (using again lemma 4.4(iii))

$$\forall x^0 (R_\rho x \ maj \ R_\rho x).$$

Hence by lemma 4.8

$$R^* := R_\rho^M \ maj \ R_\rho.$$

So for every constant c of WE-HA$^\omega$ we have a closed term t^* such that

$$\text{WE-HA}^\omega \vdash t^* \ maj \ c.$$

The proposition now follows from that fact that $t^* \ maj_{\tau\rho} \ t \wedge s^* \ maj_\rho \ s$ implies $t^* s^* \ maj_\tau \ ts$. □

Theorem 4.11 ([28]) *H^ω:=E-HA$^\omega\,\underline{+}$ AC $\underline{+}$ IP$_{ef}$. Let s be a closed term, $A(x,y,z)$ a formula containing only x,y,z as free variables and* $\deg(\tau)\leq 2$. *Then the following rule holds:*

$$\begin{cases} H^\omega \vdash \forall x^1 \forall y \leq_\rho sx \exists z^\tau A(x,y,z) \ \Rightarrow \\ H^\omega \vdash \forall x^1 \forall y \leq_\rho sx \exists z \leq_\tau tx\, A(x,y,z), \end{cases}$$

where t is a suitable closed term which can be extracted from a given proof of the assumption.

Remark 4.12 *Note that in the previous theorem the bound tx on '$\exists z$' does not depend on y.*

Corollary 4.13 (Fan Rule [66]) *Let A be a formula of E-HA$^\omega$ containing only free variables of type levels ≤ 1. Then for H^ω as above the following rule holds*

$$\begin{cases} H^\omega \vdash \forall x \leq_1 y \exists n^0 A(x,n) \ \Rightarrow \\ H^\omega \vdash \exists m^0 \forall x \leq_1 y \exists n \leq_0 m\, A(x,n). \end{cases}$$

Proof of the theorem 4.11: Suppose that

$$\mathrm{H}^\omega \vdash \forall x^1 \forall y \leq_\rho sx \exists z^\tau A(x,y,z).$$

Then by corollary 3.14,2)-4) one can extract a closed term t such that

$$\mathrm{H}^\omega \vdash \forall x^1 \forall y \leq_\rho sx\, A(x,y,txy).$$

By proposition 4.10 there are closed terms s^*, t^* such that

$$\text{E-HA}^\omega \vdash s^*\ maj\ s \wedge t^*\ maj\ t.$$

By lemma 4.4 we have in E-HA$^\omega$:

$$\forall x^1 (s^* x^M\ maj_\rho\ sx)$$

and therefore

$$\forall x^1 \forall y \leq_\rho sx (s^* x^M\ maj_\rho\ y).$$

Hence

$$\forall x^1 \forall y \leq_\rho sx(t^*(x^M, s^*x^M)\ maj_\tau\ txy).$$

For simplicity lets now consider only the case $\tau = 2$:

$$\forall x^1 \forall y \leq_\rho sx \forall z^1((t^*(x^M, s^*x^M))z^M \geq_0 txyz).$$

Hence

$$\forall x^1 \forall y \leq_\rho sx(\tilde{t}x \geq_2 txy),$$

where $\tilde{t} := \lambda x, z.[(t^*(x^M, s^*x^M))(z^M)]$. Thus $\tilde{t}$ satisfies the claim of the theorem. $\Box$

For further applications of modified realizability combined with majorization see [38].

Exercises:

Define the type-structure $\mathcal{S}^\omega$ of **all** set-theoretic functionals as follows:

$$S_0 := \mathbb{N}$$
$$S_{\tau\rho} := \{\text{ all set-theoretic functionals } \varphi : S_\rho \to S_\tau \}$$
$$\mathcal{S}^\omega := \bigcup_{\rho \in \mathbf{T}} S_\rho.$$

It is clear that $\mathcal{S}^\omega$ is a model of E-HA$^\omega$. We can relativize 'maj' to $\mathcal{S}^\omega$. In the following we refer to this majorizability in $\mathcal{S}^\omega$.

1) Construct a functional $\varphi^2 \in S_2$ such that no functional $\varphi^* \in S_2$ exists with $\varphi^*\ maj\ \varphi$ (not even on primitive recursive arguments).

2) Show that if $\varphi \in S_2$ is continuous (in the sense of the Baire space), then it can be majorized by a suitable $\varphi^* \in S_2$.

Chapter 5

Gödel's functional ('Dialectica-')interpretation

The Gödel functional interpretation, introduced in [17], assigns to each formula $A(\underline{a})$ of WE-HA$^\omega$ a formula $A^D \equiv \exists \underline{x} \forall \underline{y}\, A_D(\underline{x}, \underline{y}, \underline{a})$, where A_D is quantifier-free (and hence decidable) and $\underline{x}, \underline{y}$ are tuples of variables of finite type.

In contrast to the no-counterexample interpretation, which we briefly discussed in the first lecture, the functional interpretation does not require A to be in prenex normal form and therefore is applicable in an intuitionistic context like WE-HA$^\omega$ where not every formula is provably equivalent to a prenex one.

In the next lecture we will introduce a translation of the classical variant WE-PA$^\omega$ of WE-HA$^\omega$ (i.e. WE-HA$^\omega$ plus the tertium-non-datur schema $A \vee \neg A$) into WE-HA$^\omega$, the so-called negative translation $A \mapsto A'$ due to [16]. We will see that the composition of $'$ and D, $A \mapsto (A')^D$ provides a very subtle constructive interpretation of A which faithfully reflects the proof-theoretic and computational strength of A (in contrast to the no-counterexample interpretation of (a prenex normal form of) A which in general is a much weaker interpretation and can be established as a particular corollary of the functional interpretation). The price to be paid for this is the necessity to use functionals of arbitrary finite types already for $A \in \mathcal{L}(\text{PA})$. Moreover the functional interpretation is much more involved than the modified real-

izability interpretation but has the crucial benefit that it trivially interprets the Markov principle M^ω which has no constructive modified realizability interpretation as we have seen. It is this fact which makes the composition of negative translation and functional interpretation a powerful tool of extractive proof theory for classical non-constructive proofs. Note that in contrast to this the combination of negative translation and modified realizability interpretation $\underline{x}\ mr\ A'$ would be useless since A' is an $\exists$-free formula and therefore $(\underline{x}\ mr\ A') \equiv A'$ where $\underline{x}$ is the empty tuple.[1]
Functional interpretation has the same nice behaviour with respect to the logical deduction rules as the modified realizability interpretation.

Motivation of the functional interpretation:

The definition of A^D (like $\underline{x}\ mr\ A$) proceeds by induction on the logical structure of A (i.e. the length and the types of $\underline{x}, \underline{y}$ only depend on the logical structure of A).
The most interesting and difficult case again is the implication whose treatment we are going to motivate now:
Suppose we have already defined the functional interpretations
$A^D \equiv \exists\underline{x}\forall\underline{y}\, A_D(\underline{x},\underline{y})$ and $B^D \equiv \exists\underline{u}\forall\underline{v}\, B_D(\underline{u},\underline{v})$. We are trying to define $(A \to B)^D$:
First consider

$$(A^D \to B^D) \equiv (\exists\underline{x}\forall\underline{y}\, A_D(\underline{x},\underline{y}) \to \exists\underline{u}\forall\underline{v}\, B_D(\underline{u},\underline{v})).$$

Our strategy to obtain from this a formula of the form $\exists\underline{a}\forall\underline{b}(A \to B)_D$ (with $(A \to B)_D$ quantifier-free) is to transform $(A^D \to B^D)$ into prenex normal form and then to apply the axiom of choice AC.
It is an easy exercise to verify that there are four different prenex normal forms of $A^D \to B^D$. We try to choose the most constructive (or rather: the least non-constructive) one:

[1] One can however use modified realizability in connection with the negative translation if one applies the so-called Friedman-Dragalin A-translation as an intermediate step. See lecture 7 below.

For the first step we have two possibilities:

$$\begin{aligned}&\exists \underline{x}\forall \underline{y}\, A_D(\underline{x},\underline{y}) \to \exists \underline{u}\forall \underline{v}\, B_D(\underline{u},\underline{v})\\ &\mapsto \begin{cases} (1)\ \forall \underline{x}(\forall \underline{y}\, A_D(\underline{x},\underline{y}) \to \exists \underline{u}\forall \underline{v}\, B_D(\underline{u},\underline{v})) \\ (2)\ \exists \underline{u}(\exists \underline{x}\forall \underline{y}\, A_D(\underline{x},\underline{y}) \to \forall \underline{v}\, B_D(\underline{u},\underline{v})). \end{cases}\end{aligned}$$

Here the choice is obvious: the passage to (1) is intuitionistically valid, whereas the passage to (2) not even holds in IL^ω+ IP^ω_{ef}+ M^ω.
From (1) there are two ways to proceed further:

$$(1) \mapsto \begin{cases} (1.1)\, \forall \underline{x}\exists \underline{u}(\forall \underline{y}\, A_D(\underline{x},\underline{y}) \to \forall \underline{v}\, B_D(\underline{u},\underline{v})) \\ (1.2)\, \forall \underline{x}\exists \underline{y}(A_D(\underline{x},\underline{y}) \to \exists \underline{u}\forall \underline{v}\, B_D(\underline{u},\underline{v})). \end{cases}$$

This time the choice is more difficult since both implications (1) $\to$ (1.1) and (1) $\to$ (1.2) are not provable in IL^ω. So we have to compromise our goal to use only strictly constructive transformation steps. However the first implication only requires a weak form of IP^ω_{ef} (for purely universal formulas A) in addition to IL^ω which has some constructive justification by the results of lecture 3. So lets choose (1.1).
From there we have two possibilities to finish our prenexation:

$$(1.1) \mapsto \begin{cases} (1.1.1)\, \forall \underline{x}\exists \underline{u}\forall \underline{v}\exists \underline{y}(A_D(\underline{x},\underline{y}) \to B_D(\underline{u},\underline{v})) \\ (1.1.2)\, \forall \underline{x}\exists \underline{u}\exists \underline{y}\forall \underline{v}(A_D(\underline{x},\underline{y}) \to B_D(\underline{u},\underline{v})). \end{cases}$$

Again the choice is not obvious: both implications are not provable in IL^ω.
Lets consider (1.1) $\to$ (1.1.1) first: The first step to

$$\forall \underline{x}\exists \underline{u}\forall \underline{v}(\forall \underline{y}\, A_D(\underline{x},\underline{y}) \to B_D(\underline{u},\underline{v}))$$

is perfectly valid from an intuitionistic point of view. However from there we – intuitionistically – only get

$$\forall \underline{x}\exists \underline{u}\forall \underline{v}\neg\neg\exists \underline{y}(A_D(\underline{x},\underline{y}) \to B_D(\underline{u},\underline{v}))$$

and so need the Markov-principle M^ω to obtain (1.1.1).
For (1.1) $\to$ (1.1.2) the first step to

$$\forall \underline{x} \exists \underline{u} \exists \underline{y} (A_D(\underline{x}, \underline{y}) \to \forall \underline{v}\, B_D(\underline{u}, \underline{v}))$$

is not intuitionistically valid but again only the passage to the weaker

$$\forall \underline{x} \exists \underline{u} \neg\neg \exists \underline{y} (A_D(\underline{x}, \underline{y}) \to \forall \underline{v}\, B_D(\underline{u}, \underline{v})).$$

However this time not even M^ω suffices to get rid of $\neg\neg$ since '$A_D(\underline{x}, \underline{y}) \to \forall \underline{v}\, B_D(\underline{u}, \underline{v})$' is not quantifier-free.
So the implication '(1.1) $\to$ (1.1.1)' is less non-constructive than '(1.1) $\to$ (1.1.2)'. Hence we now 'officially' choose (1.1.1) as our prenex normal form of $A^D \to B^D$. Applying AC to (1.1.1) we finally obtain

$$(A \to B)^D :\equiv \exists \underline{U}, \underline{Y} \forall \underline{x}, \underline{v} (\underbrace{A_D(\underline{x}, \underline{Y}\, \underline{x}\, \underline{v}) \to B_D(\underline{U}\, \underline{x}, \underline{v})}_{(A \to B)_D :\equiv}).$$

Despite of the fact that we had to make various compromises to end up with $(A \to B)^D$, this interpretation works while any of the remaining three prenex normal forms of $A^D \to B^D$ would result in a definition of $(A \to B)^D$ which even for $B :\equiv A$ in general would fail to have a constructive (computable) realization (exercise).

Definition 5.1 (Gödel [17]) *To every formula A of WE-HA$^\omega$ we assign a translation $A^D \equiv \exists \underline{x} \forall \underline{y}\, A_D(\underline{x}, \underline{y})$ in the same language. The free variables of A^D are that of A. The types and length of $\underline{x}, \underline{y}$ depend only on the logical structure of A. A_D is a quantifier-free formula.*

(i) $A^D :\equiv A_D :\equiv A$ *for prime formulas A.*

(ii) $(A \wedge B)^D :\equiv \exists \underline{x}, \underline{u} \forall \underline{y}, \underline{v} [A \wedge B]_D$
$:\equiv \exists \underline{x}, \underline{u} \forall \underline{y}, \underline{v} [A_D(\underline{x}, \underline{y}) \wedge B_D(\underline{u}, \underline{v})]$,

(iii) $(A \vee B)^D :\equiv \exists z^0, \underline{x}, \underline{u} \forall \underline{y}, \underline{v} [A \vee B]_D$
$:\equiv \exists z^0, \underline{x}, \underline{u} \forall \underline{y}, \underline{v} [(z = 0 \to A_D(\underline{x}, \underline{y})) \wedge (z \neq 0 \to B_D(\underline{u}, \underline{v}))]$,

(iv) $(\exists z^\rho A(z))^D :\equiv \exists z, \underline{x} \forall \underline{y} (\exists z\, A(z))_D :\equiv \exists z, \underline{x} \forall \underline{y}\, A_D(\underline{x}, \underline{y}, z)$,

(v) $(\forall z^\rho A(z))^D :\equiv \exists \underline{X} \forall z, \underline{y} (\forall z\, A(z))_D :\equiv \exists \underline{X} \forall z, \underline{y}\, A_D(\underline{X} z, \underline{y}, z)$,

(vi) $(A \to B)^D :\equiv \exists \underline{U}, \underline{Y} \forall \underline{x}, \underline{v} (A_D(\underline{x}, \underline{Y}\,\underline{x}\,\underline{v}) \to B_D(\underline{U}\,\underline{x}, \underline{v}))$,
$(A \to B)_D :\equiv (A_D(\underline{x}, \underline{Y}\,\underline{x}\,\underline{v}) \to B_D(\underline{U}\,\underline{x}, \underline{v}))$.

Remark 5.2 *As a consequence of the treatment of implication we obtain*

(i) $(\neg A)^D \equiv \exists \underline{Y} \forall \underline{x} \neg A_D(\underline{x}, \underline{Y}\,\underline{x})$,

(ii) $(\neg\neg A)^D \equiv \exists \underline{X} \forall \underline{Y} \neg\neg A_D(\underline{X}\,\underline{Y}, \underline{Y}(\underline{X}\,\underline{Y})) \leftrightarrow \exists \underline{X} \forall \underline{Y} A_D(\underline{X}\,\underline{Y}, \underline{Y}(\underline{X}\,\underline{Y}))$, *where the equivalence is provable in WE-HA*$^\omega$.

Definition 5.3 *The independence-of-premise schema IP*$^\omega_\forall$ *for universal premises is the union (for all types) of*

$$IP^\rho_\forall :\ (\forall \underline{x} A_0(\underline{x}) \to \exists y^\rho B(y)) \to \exists y^\rho (\forall \underline{x} A_0(\underline{x}) \to B(y)),$$

where $\underline{y}$ *not free in* $\forall \underline{x}\, A_0(\underline{x})$.

Theorem 5.4 (soundness of functional interpretation [17],[65])

$$\begin{cases} \textit{WE-HA}^\omega + AC + IP^\omega_\forall + M^\omega \vdash A(\underline{a}), \quad \textit{then} \\ \textit{WE-HA}^\omega \vdash \forall \underline{y}\, A_D(\underline{t}\,\underline{a}, \underline{y}, \underline{a}), \end{cases}$$

where $\underline{t}$ *is a suitable tuple of closed terms of WE-HA*$^\omega$ *which can be extracted from a given proof of the assumption.*

Proof: As in the proof of the soundness theorem for modified realizability we proceed by induction on the length of the derivation.

1) Logical axioms and rules: We will discuss only two axioms and two rules to give an idea of the general proof (for full details see [65],[52]).

$\underline{A \to A \wedge A}$:

$(A \to A \wedge A)^D \equiv$

$(\exists \underline{x} \forall \underline{y} A_D(\underline{x}, \underline{y}, \underline{a}) \to \exists \underline{x}', \underline{x}'' \forall \underline{y}', \underline{y}''(A_D(\underline{x}', \underline{y}', \underline{a}) \wedge A_D(\underline{x}'', \underline{y}'', \underline{a})))^D$

$\exists \underline{Y}, \underline{X}', \underline{X}'' \forall \underline{x}, \underline{y}', \underline{y}''(A_D(\underline{x}, \underline{Y}\, \underline{x}\, \underline{y}' \underline{y}'', \underline{a}) \to A_D(\underline{X}' \underline{x}, \underline{y}', \underline{a}) \wedge A_D(\underline{X}'' \underline{x}, \underline{y}'', \underline{a}))$.

Hence

$$\underline{t}_{X'} := \underline{t}_{X''} := \lambda \underline{a}, \underline{x}.\underline{x}$$
$$\underline{t}_Y \underline{a}\, \underline{x}\, \underline{y}' \underline{y}'' := \begin{cases} \underline{y}', & \text{if } t_{A_D} \underline{x}\, \underline{y}' \underline{a} \neq 0 \\ \underline{y}'', & \text{if } t_{A_D} \underline{x}\, \underline{y}' \underline{a} = 0, \end{cases}$$

satisfy the functional interpretation of $A \to A \wedge A$ (here t_{A_D} is a closed term of WE-HA$^\omega$ such that WE-HA$^\omega \vdash t_{A_D} \underline{x}\, \underline{y}\, \underline{a} =_0 0 \leftrightarrow A_D(\underline{x}, \underline{y}, \underline{a})$).

$\underline{A \to A \vee B}$:

$(A \to A \vee B)^D \equiv (\exists \underline{x} \forall \underline{y} A_D(\underline{x}, \underline{y}, \underline{a}) \to$

$\exists z^0, \underline{x}', \underline{u} \forall \underline{y}', \underline{v}((z = 0 \to A_D(\underline{x}', \underline{y}', \underline{a})) \wedge (z \neq 0 \to B_D(\underline{u}, \underline{v}, \underline{a}'))))^D$

$\equiv \exists \underline{Y}, Z, \underline{X}', \underline{U} \forall \underline{x}, \underline{y}', \underline{v}(A_D(\underline{x}, \underline{Y}\, \underline{x}\, \underline{y}' \underline{v}, \underline{a}) \to$

$((Z\underline{x} = 0 \to A_D(\underline{X}' \underline{x}, \underline{y}', \underline{a})) \wedge (Z\underline{x} \neq 0 \to B_D(\underline{U}\, \underline{x}, \underline{v}, \underline{a}'))))$.

Hence $\underline{t_Y} := \lambda \underline{\tilde{a}}, \underline{x}, \underline{y}', \underline{v}.\underline{y}'$, $t_Z := \lambda \underline{\tilde{a}}, \underline{x}.0^0$, $\underline{t_{X'}} := \lambda \underline{\tilde{a}}, \underline{x}.\underline{x}$, $\underline{t_U} := \lambda \underline{\tilde{a}}, \underline{x}.\underline{\mathcal{O}}$, where $\underline{\tilde{a}} = \{\underline{a}\} \cup \{\underline{a}'\}$, satisfy the functional interpretation of $A \to A \vee B$.

The modus ponens rule: Assume

$$(1)\ \forall \underline{y} A_D(\underline{t}_1 \underline{a}, \underline{y}, \underline{a})$$

and

$$(2)\ \forall \underline{x}, \underline{v}(A_D(\underline{x}, \underline{t}_2 \underline{\tilde{a}}\, \underline{x}\, \underline{v}, \underline{a}) \to B_D(\underline{t}_3 \underline{\tilde{a}}\, \underline{x}, \underline{v}, \underline{a}')),$$

where again $\underline{\tilde{a}} = \{\underline{a}\} \cup \{\underline{a}'\}$.
We have to construct $\underline{t}_4$ such that

$$\forall \underline{v}\, B_D(\underline{t}_4 \underline{a}', \underline{v}, \underline{a}').$$

Apply (1) to $\underline{y} := \underline{t}_2 \underline{\tilde{a}}\, \underline{x}\, \underline{v}$, then

$$(3)\ \forall \underline{x}, \underline{v} A_D(\underline{t}_1 \underline{a}, \underline{t}_2 \underline{\tilde{a}}\, \underline{x}\, \underline{v}, \underline{a}).$$

Apply (2) and (3) to $\underline{x} := \underline{t}_1\underline{a}$. Then

$$(4)\ \forall \underline{v}(A_D(\underline{t}_1\underline{a}, \underline{t}_2(\underline{\tilde{a}}, \underline{t}_1\underline{a}, \underline{v}), \underline{a}) \to B_D(\underline{t}_3(\underline{\tilde{a}}, \underline{t}_1\underline{a}), \underline{v}, \underline{a}'))$$

and

$$(5)\ \forall \underline{v}\, A_D(\underline{t}_1\underline{a}, \underline{t}_2(\underline{\tilde{a}}, \underline{t}_1\underline{a}, \underline{v}), \underline{a}).$$

Hence

$$(6)\ \forall \underline{v} B_D(\underline{t}_3(\underline{\tilde{a}}, \underline{t}_1\underline{a}), \underline{v}, \underline{a}').$$

Let $\underline{t}[\underline{a}']$ be the result of replacing all variables a_i in $\underline{t}_3(\underline{\tilde{a}}, \underline{t}_1\underline{a})$ which do not occur in $\underline{a}'$ by $\mathcal{O}$ of appropriate type. Then $\underline{t}_4 := \lambda\underline{a}'.\underline{t}[\underline{a}']$ does the job.

The rule $\frac{A\to B\,,\,B\to C}{A\to C}$: For notational simplicity we omit the free parameters this time.
Assume

$$(1)\ \forall \underline{x}, \underline{v}(A_D(\underline{x}, \underline{t}_1\underline{x}\,\underline{v}) \to B_D(\underline{t}_2\underline{x}, \underline{v}))$$

and

$$(2)\ \forall \underline{u}, \underline{w}(B_D(\underline{u}, \underline{t}_3\underline{u}\,\underline{w}) \to C_D(\underline{t}_4\underline{u}, \underline{w})).$$

We have to construct $\underline{t}_5, \underline{t}_6$ such that

$$\forall \underline{x}, \underline{w}(A_D(\underline{x}, \underline{t}_5\underline{x}\,\underline{w}) \to C_D(\underline{t}_6\underline{x}, \underline{w})).$$

Apply (1) to $\underline{v} = \underline{t}_3(\underline{t}_2\underline{x}, \underline{w})$ and (2) to $\underline{u} = \underline{t}_2\underline{x}$. Then

$$(3)\ \forall \underline{x}, \underline{w}(A_D(\underline{x}, \underline{t}_1(\underline{x}, \underline{t}_3(\underline{t}_2\underline{x}, \underline{w}))) \to C_D(\underline{t}_4(\underline{t}_2\underline{x}), \underline{w})).$$

Hence

$$\underline{t}_5 := \lambda\underline{x}, \underline{w}.\underline{t}_1(\underline{x}, \underline{t}_3(\underline{t}_2\underline{x}, \underline{w})),\ \ \underline{t}_6 := \lambda\underline{x}.\underline{t}_4(\underline{t}_2\underline{x})$$

do the job.

2) Axioms for $=_0, S, \Pi, \Sigma, R$**:** These purely universal axioms are identical with their own functional interpretation.

3) The quantifier-free extensionality rule QF-ER: Both the premise and the conclusion are purely universal and so are identical to their functional interpretation.[2]

[2] We may assume that A_0 in QF-ER does not contain any $\vee$, since $A_0(\underline{x})$ can be written as $t\underline{x} =_0 0$ in WE-HA$^\omega$.

4) The schema of induction: It is easier to use the equivalent induction rule: Let $B(y^0)^D \equiv \exists \underline{u} \forall \underline{v} B_D(\underline{u}, \underline{v}, y, \underline{a})$ and assume that we have already proved

$$\begin{cases} \forall \underline{v}\, B_D(\underline{t}_1 \underline{a}, \underline{v}, 0, \underline{a}) \text{ and} \\ \forall \underline{u}, \underline{w}(B_D(\underline{u}, \underline{t}_2 y\, \underline{a}\, \underline{u}\, \underline{w}, y, \underline{a}) \to B_D(\underline{t}_3 y\, \underline{a}\, \underline{u}, \underline{w}, y+1, \underline{a})). \end{cases}$$

Define $\underline{t}$ by simultaneous primitive recursion in higher types such that

$$\begin{cases} \underline{t}(\underline{a}, 0) = \underline{t}_1 \underline{a} \\ \underline{t}(\underline{a}, y+1) = \underline{t}_3(y, \underline{a}, \underline{t}(\underline{a}, y)). \end{cases}$$

Then

$$\begin{cases} \forall \underline{v} B_D(\underline{t}(\underline{a}, 0), \underline{v}, 0, \underline{a}) \text{ and} \\ \forall \underline{w}(B_D(\underline{t}(\underline{a}, y), \underline{t}_2(y, \underline{a}, \underline{t}(\underline{a}, y), \underline{w}), y, \underline{a}) \to B_D(\underline{t}(\underline{a}, y+1), \underline{w}, y+1, \underline{a})) \end{cases}$$

and therefore

$$\begin{cases} \forall \underline{v} B_D(\underline{t}(\underline{a}, 0), \underline{v}, 0, \underline{a}) \text{ and} \\ \forall \underline{v} B_D(\underline{t}(\underline{a}, y), \underline{v}, y, \underline{a}) \to \forall \underline{v} B_D(\underline{t}(\underline{a}, y+1), \underline{v}, y+1, \underline{a}). \end{cases}$$

Hence by the induction rule we obtain

$$\forall \underline{v} B_D(\underline{t}(\underline{a}, y), \underline{v}, y, \underline{a}).$$

5) The functional interpretations of AC, M^ω and $\mathrm{IP}^\omega_\forall$ only need simple λ-terms and can be verified already in WE-HA^ω without the use of AC, M^ω and $\mathrm{IP}^\omega_\forall$. $\square$

Warning: The soundness theorem does not hold for E-HA^ω(see [24]).

Remark 5.5 *Gödel actually established the conclusion of the soundness theorem in (an intensional variant of) a quantifier-free fragment* qf-(WE-HA$^\omega$) *of* WE-HA$^\omega$ *which results if quantifiers are omitted from the language, the axiom schema of induction is replaced by a quantifier-free rule of induction*

$$\text{QF-IR}: \quad \frac{A_0(0)\,,\, A_0(x^0) \to A_0(x+1)}{A_0(x)}$$

and a substitution rule

$$Sub: \quad \frac{A(x^\rho)}{A(t^\rho)}$$

(replacing $\forall$-elimination) is added.
To verify $A_D(\underline{t}\,\underline{a}, \underline{y}, \underline{a})$ in qf-(WE-HA$^\omega$) *requires a somewhat more complicated treatment of induction (see e.g.* [65]*).*

Definition 5.6 ([65]) *The subset Γ_2 of formulas $\in \mathcal{L}$(WE-HA$^\omega$) is defined inductively by*

1) *Prime formulas are in Γ_2.*

2) $A, B \in \Gamma_2 \Rightarrow A \wedge B, A \vee B, \forall x\, A(x), \exists x\, A(x) \in \Gamma_2$.

3) *If A is purely universal and $B \in \Gamma_2$, then $(\exists \underline{x} A \to B) \in \Gamma_2$.*

Lemma 5.7 *For $A \in \Gamma_2$ one has WE-HA$^\omega \vdash A^D \to A$.*

Proof: Easy induction on the logical structure of A. □

Corollary 5.8 *WE-HA$^\omega$+ AC + IP$^\omega_\forall$+ M$^\omega$ is conservative over WE-HA$^\omega$ with respect to formulas $A \in \Gamma_2$.*

Proof: The corollary follows from theorem 5.4 and lemma 5.7. □

Proposition 5.9 *For all formulas A of WE-HA$^\omega$ one has*

$$\textit{WE-HA}^\omega + AC + IP^\omega_\forall + M^\omega \vdash A \leftrightarrow A^D.$$

Proof: Easy induction on the logical structure of A. $\square$

Corollary 5.10 *WE-HA$^\omega$+ AC + IP$^\omega_\forall$+ M$^\omega$ has the disjunction property* DP*, the existence property* EP *and is closed under the rules of choice ACR and of independence-of-premise for purely universal formulas IPR$^\omega_\forall$.*

Proof: The corollary follows similarly to corollary 3.14 but with theorem 5.4 and proposition 5.9 instead of modified realizability with truth. $\square$

An application of functional interpretation and majorization

Definition 5.11 (hereditarily extensional equality [65])

$$\begin{cases} x_1 \approx_0 x_2 :\equiv (x_1 =_0 x_2), \\ x_1 \approx_{\tau\rho} x_2 :\equiv \forall y_1^\rho, y_2^\rho (y_1 \approx_\rho y_2 \to x_1 y_1 \approx_\tau x_2 y_2). \end{cases}$$

Lemma 5.12 *WE-HA$^\omega$* $\vdash x_1 =_\rho \tilde{x}_1 \wedge x_2 =_\rho \tilde{x}_2 \wedge x_1 \approx_\rho x_2 \to \tilde{x}_1 \approx_\rho \tilde{x}_2$.

Proof: Induction on ρ. $\square$

Proposition 5.13 *Let t^ρ be a closed term of WE-HA$^\omega$. Then*

$$\textit{WE-HA}^\omega \vdash t \approx_\rho t.$$

Proof: Induction on the structure of t.
(i) Constants: One easily verifies that $0^0 \approx_0 0^0$, $S \approx_1 S$, $\Pi_{\rho,\tau} \approx \Pi_{\rho,\tau}$, $\Sigma_{\delta,\rho,\tau} \approx \Sigma_{\delta,\rho,\tau}$.
R_ρ: We show by induction on x^0 that $R_\rho x \approx R_\rho x$:
Suppose that $y_1 \approx y_2, z_1 \approx z_2$:
$R_\rho 0 y_1 z_1 = y_1 \approx y_2 = R_\rho 0 y_2 z_2 \Rightarrow R_\rho 0 y_1 z_2 \approx R_\rho 0 y_2 z_2$.
$R_\rho(x+1) y_1 z_1 = z_1(R_\rho x y_1 z_1)x \overset{I.H.}{\approx} z_2(R_\rho x y_2 z_2)x = R_\rho(x+1) y_2 z_2$
$\Rightarrow R_\rho(x+1) y_1 z_1 \approx R_\rho(x+1) y_2 z_2$.
Since $x_1 =_0 x_2 \leftrightarrow x_1 \approx_0 x_2$, we have by $=_0$-axioms

$$\forall x_1, x_2 (x_1 \approx_0 x_2 \to R_\rho x_1 \approx R_\rho x_2),$$

i.e. $R_\rho \approx R_\rho$.
(ii) $t \approx_{\tau\rho} t \wedge s \approx_\rho s \to ts \approx_\tau ts$. $\square$

Corollary 5.14 *Let $t^{1(1)}$ be a closed term of WE-HA$^\omega$. Then*

$$\text{WE-HA}^\omega \vdash \forall x^1, y^1(x =_1 y \to tx =_1 ty).$$

Proof: This follows from proposition 5.13 since
WE-HA$^\omega \vdash x =_1 y \leftrightarrow x \approx_1 y$. □

Proposition 5.15 *Let $t^{1(1)}$ be closed. Then $t^{1(1)}$ is uniformly continuous on each set $\{x : x \leq_1 y\}$ with a modulus of uniform continuity which is definable in WE-HA$^\omega$ (uniformly in y), i.e. there is a closed term $\tilde{t}^{0(1)(0)}$ of WE-HA$^\omega$:*

$$\text{WE-HA}^\omega \vdash \forall k^0 \forall x_1, x_2 \leq_1 y(\bigwedge_{i=0}^{\tilde{t}ky}(x_1 i =_0 x_2 i) \to \bigwedge_{j=0}^{k}(tx_1 j =_0 tx_2 j)).$$

Proof [28]: By the corollary above we have

$$\text{WE-HA}^\omega \vdash \forall x_1, x_2(\forall i(x_1 i =_0 x_2 i) \to \forall k \forall j \leq k(tx_1 j =_0 tx_2 j)).$$

Hence

$$\text{WE-HA}^\omega + M^\omega \vdash \forall k \forall x_1, x_2 \exists i(x_1 i =_0 x_2 i \to \bigwedge_{j=0}^{k}(tx_1 j =_0 tx_2 j)).$$

By theorem 5.4 there exists a term $\hat{t}$ such that

$$\text{WE-HA}^\omega \vdash \forall k \forall x_1, x_2(x_1(\hat{t}kx_1x_2) =_0 x_2(\hat{t}kx_1x_2) \to \bigwedge_{j=0}^{k}(tx_1 j =_0 tx_2 j)).$$

By proposition 4.10 there exists a closed term $\hat{t}^*$ such that

$$\text{WE-HA}^\omega \vdash \hat{t}^* \; maj \; \hat{t}$$

and hence by lemma 4.4

$$\text{WE-HA}^\omega \vdash \forall k \forall x_1, x_2 \leq_1 y(\hat{t}^* k y^M y^M \geq_0 \hat{t}kx_1x_2).$$

So $\tilde{t} := \lambda k^0, y^1.\hat{t}^* k y^M y^M$ does the job. □

Exercises:

1) Solve the functional interpretation of

$$\neg\neg(\exists x A_0(x) \vee \neg\exists x A_0(x))$$

by a closed term of WE-HA$^\omega$.

2) In addition to the prenex normal form we used in the definition of $(A \to B)^D$, there are three more prenex normal forms of $(A^D \to B^D)$ which give rise to corresponding functional interpretations $(A \to B)^i$ $(i = 1, 2, 3)$ of $A \to B$.

   For all three of them already $(A \to A)^i$ fails to have a computable solution for suitable A.

   Compute these interpretations $(A \to B)^i$ and give a counterexample to the computable solvability of $(A \to A)^i$ for at least one of them.

Suggested further reading: [17] (the original paper which introduced functional interpretation; an English translation with extended introductory notes by A.S. Troelstra can be found in [18]), [65] (chapter 3, section 5; this is a very concise and compact treatment of functional interpretation), [52] (covers in detail C. Spector's extension of functional interpretation to analysis by means of bar recursion), [1] (a very readable and comprehensive treatment of the whole subject). [5] contains an interesting discussion about the functional interpretation of '$\to$' from a constructive point of view. Applications of functional interpretation to systems of bounded arithmetic are given in [9]. For further applications of functional interpretation combined with majorization see [30],[33],[34],[35],[36],[37]. For applications of functional interpretation (and majorization) in the context of approximation theory see [31],[32].

Chapter 6

Negative translation and its use combined with functional interpretation

There are several interpretations – so-called 'negative' or 'double-negation' translations – of classical logic as well as many theories based on classical logic into their intuitionistic variant. All these translations $A \mapsto A'$ have in common that A' is (or is intuitionistically equivalent to) a negative formula. The first such translation is due to Gödel [16] (although G. Gentzen independently discovered a similar translation). There is some preceding work by Kolmogorov [41] and Glivenko [15]. Two further variants of Gödel's translation are due to Kuroda [50] and it his one of these which we will adopt here:

Definition 6.1 *Let A be a formula in a theory based on $\mathcal{L}(\mathrm{IL}^\omega)$. A' is defined as $A' :\equiv \neg\neg A^*$, where A^* is defined by induction on the logical structure of A:*

(i) $A^* :\equiv A$, *if A is a prime formula,*

(ii) $(A \Box B)^* :\equiv (A^* \Box B^*)$, *where* $\Box \in \{\wedge, \vee, \rightarrow\}$,

(iii) $(\exists x^\rho A(x))^* :\equiv \exists x^\rho (A(x))^*$,

(iv) $(\forall x^\rho A(x))^* :\equiv \forall x^\rho \neg\neg (A(x))^*$.

Definition 6.2 PL^ω ($\mathrm{WE\text{-}PA}^\omega$, $\mathrm{E\text{-}PA}^\omega$) *is the extension of* IL^ω ($\mathrm{WE\text{-}HA}^\omega$, $\mathrm{E\text{-}HA}^\omega$) *obtained by adding the tertium-non-datur schema* $A \vee \neg A$.

Proposition 6.3 *(i)* $\mathrm{PL}^\omega \vdash A \ \Rightarrow \mathrm{IL}^\omega \vdash A'$,

(ii) $\mathrm{(W)E\text{-}PA}^\omega \vdash A \ \Rightarrow \mathrm{(W)E\text{-}HA}^\omega \vdash A'$.

Proof: (i) Induction on the length of the derivation: We only treat the tertium-non-datur schema, the modus ponens and the $\forall$-introduction rule: $(A \vee \neg A)' \equiv \neg\neg(A^* \vee \neg A^*)$ which is provable in IL^ω (it is an easy exercise that $\neg\neg(A \vee \neg A)$ holds intuitionistically for arbitrary formulas A).
Consider $\frac{A,\, A \to B}{B}$: By induction hypothesis we have $\neg\neg A^*$ and $\neg\neg(A^* \to B^*)$. By intuitionistic logic we get $\neg\neg A^* \to \neg\neg B^*$ and hence by modus ponens $\neg\neg B^*$, i.e. B'.
Consider $\frac{A \to B(x)}{A \to \forall x B(x)}$: By induction hypothesis we have $\neg\neg(A^* \to (B(x))^*)$ and therefore (by intuitionistic logic) $A^* \to \neg\neg(B(x))^*$. By $\forall$-intoduction we obtain
$A^* \to \forall x \neg\neg(B(x))^*$, i.e. $(A \to \forall x B(x))^*$ and therefore $(A \to \forall x B(x))'$.
(ii) We only have to extend the proof of (i) by the treatment of the non-logical axioms and rules: The negative translation of the purely universal $=_0, S, \Pi, \Sigma, R$-axioms trivially follows from the axioms themselves (note that $\mathrm{WE\text{-}HA}^\omega \vdash \neg\neg A_0 \leftrightarrow A_0$ and so the translation of purely universal axioms is in fact equivalent (relative to $\mathrm{WE\text{-}HA}^\omega$) to the axioms since intuitionistically $\neg\neg\forall x \neg\neg A(x) \leftrightarrow \forall \neg\neg A(x)$). Similarly the negative translation of (E) (resp. of the premise and the conclusion of QF-ER) can be seen to be equivalent to (E) in $\mathrm{WE\text{-}HA}^\omega$.
The induction rule: By the induction hypothesis we have $\neg\neg(A(0))^*$ and $\neg\neg((A(x))^* \to (A(x+1))^*)$. Hence also $\neg\neg(A(x))^* \to \neg\neg(A(x+1))^*$. Thus by the induction rule we obtain $\neg\neg(A(x))^*$, i.e. $(A(x))'$. $\Box$

Definition 6.4 *The schema* QF-AC *of quantifier-free choice is the restriction of* AC *to quantifier-free formulas* $A_0 \equiv A$. *For convenience we formulate* *this schema for tuples*[1]*:*

$$\text{QF-AC}: \ \forall \underline{x} \exists \underline{y} A_0(\underline{x}, \underline{y}) \to \exists \underline{Y} \forall \underline{x} A_0(\underline{x}, \underline{Y}\,\underline{x}),$$

where A_0 *is quantifier-free.*

Proposition 6.5
WE-PA$^{\omega}$+ QF-AC $\vdash A \ \Rightarrow$ WE-HA$^{\omega}$+ QF-AC $+M^{\omega} \vdash A'$.

Proof: We only have to extend the proof of proposition 6.3 by showing that

$$\text{WE-HA}^{\omega} + \text{QF-AC} + M^{\omega} \vdash (\text{QF-AC})'.$$

We have in WE-HA$^{\omega}$

$$\Big(\forall \underline{x} \exists \underline{y} A_0(\underline{x}, \underline{y}) \to \exists \underline{Y} \forall \underline{x} A_0(\underline{x}, \underline{Y}\,\underline{x})\Big)' \leftrightarrow$$
$$\Big(\forall \underline{x} \neg\neg \exists \underline{y} A_0(\underline{x}, \underline{y}) \to \neg\neg \exists \underline{Y} \forall \underline{x} \neg\neg A_0(\underline{x}, \underline{Y}\,\underline{x})\Big),$$

which clearly is implied by

$$(*) \ \forall \underline{x} \neg\neg \exists \underline{y} A_0(\underline{x}, \underline{y}) \to \exists \underline{Y} \forall \underline{x} A_0(\underline{x}, \underline{Y}\,\underline{x}),$$

which in turn follows from M^{ω} and QF-AC. □

Theorem 6.6

$$\begin{cases} \textit{WE-PA}^{\omega} + \textit{ QF-AC } \vdash A(\underline{a}) \\ \Rightarrow \ \textit{one can extract closed terms } \underline{t} \textit{ of WE-HA}^{\omega} \textit{ such that} \\ \textit{WE-HA}^{\omega} \vdash \forall \underline{y} (A')_D(\underline{t}\,\underline{a}, \underline{y}, \underline{a}). \end{cases}$$

Proof: Combine proposition 6.5 with theorem 5.4. □

[1] Actually one can show in WE-HA$^{\omega}$ that finite tuples $\underline{x}$ of variables (of different types) can be coded together into a single variable x whose type depends on the types of $\underline{x}$ (see [65] for details on that).

Corollary 6.7 *Let $A_0(x,y)$ be a (quantifier-free) formula of WE-PA$^\omega$ which only contains x,y as free variables. Then the following rule holds:*

$$\begin{cases} \textit{WE-PA}^\omega\textit{+ QF-AC} \vdash \forall x^\rho \exists y^\tau A_0(x,y) \\ \Rightarrow \textit{ one can extract a closed term } t \textit{ of WE-HA}^\omega \textit{ such that} \\ \textit{WE-HA}^\omega \vdash \forall x A_0(x,tx). \end{cases}$$

Proof:

$$\text{WE-PA}^\omega\text{+QF-AC} \vdash \forall x \exists y A_0(x,y) \overset{prop.6.5}{\Rightarrow}$$
$$\text{WE-HA}^\omega\text{+QF-AC} + M^\omega \vdash \neg\neg\forall x\neg\neg\exists y A_0(x,y) \overset{M^\omega}{\Rightarrow}$$
$$\text{WE-HA}^\omega\text{+QF-AC} + M^\omega \vdash \forall x \exists y A_0(x,y) \overset{thm.5.4}{\Rightarrow}$$
$$\text{WE-HA}^\omega \vdash \forall x A_0(x,tx) \text{ for a suitable closed term } t$$

(note that A_0 can be treated as a prime formula in WE-HA$^\omega$). $\square$

Combined with the majorization technique from lecture 4 we obtain

Proposition 6.8 ([28]) *Let $A_0(x^1, y^\rho, z^\tau)$ be a (quantifier-free) formula of WE-PA$^\omega$ containing only x,y,z as free variables, $\tau \le 2$ and s be a closed term. Then*

$$\begin{cases} \textit{WE-PA}^\omega\textit{+ QF-AC} \vdash \forall x^1 \forall y \le_\rho sx \exists z^\tau A_0(x,y,z) \\ \Rightarrow \textit{ one can extract a closed term } t \textit{ of WE-HA}^\omega \textit{ such that} \\ \textit{WE-HA}^\omega \vdash \forall x^1 \forall y \le_\rho sx \exists z \le_\tau tx\, A_0(x,y,z). \end{cases}$$

As a further application of corollary 6.7 we obtain the no-counterexample interpretation of PA by terms of WE-HA$^\omega$:

Proposition 6.9 *Let $A \in \mathcal{L}(\mathrm{PA})$ be a prenex sentence. Then the following rule holds:*

$$\begin{cases} \mathrm{PA} \vdash A \\ \Rightarrow \textit{ one can extract closed terms } \underline{\Phi} \textit{ of WE-HA}^\omega \textit{ such that} \\ \textit{WE-HA} \vdash \underline{\Phi} \textit{ n.c.i. } A. \end{cases}$$

Proof: Apply corollary 6.7 to A^H. □

Remark 6.10 *Strictly speaking* PA *is not a subsystem of WE-PA$^\omega$ since we have included symbols for all primitive recursive functions as primitive notions in* PA *whereas they are defined notions in WE-PA$^\omega$. However* PA *is a subsystem of a corresponding definitorial extension of WE-PA$^\omega$ to which corollary 6.7 applies as well.*

The combination of negative translation and functional interpretation $(A')^D$ is much closer related to A than the no-counterexample interpretation is (for prenex arithmetical A), since the equivalence of A and $(A')^D$ can be proved using only quantifier-free choice (although in higher types) whereas the no-counterexample interpretation of A only implies A in the presence of (number-theoretic) choice for arithmetical formulas:

Proposition 6.11 *Let A be an arbitrary formula of WE-PA$^\omega$. We may assume that A does not contain $\vee$ (which can be defined in terms of $\wedge$ and $\neg$ in WE-PA$^\omega$). Then PL$^\omega$+ QF-AC $\vdash A \leftrightarrow (A')^D$.*

Proof: Exercise (Hint: Show that PL$^\omega$+ QF-AC $\vdash A \leftrightarrow A^D$ for $\exists$-free formulas A). □

Elimination of extensionality

Definition 6.12 *We define $E^\rho(x^\rho)$ and $x =^e_\rho y$ by induction on ρ :*

$$\begin{aligned}
&E^0(x) :\equiv (x =_0 x),\ x =^e_0 y :\equiv (x =_0 y),\\
&E^{\tau\rho}(x) :\equiv \forall y, z(y =^e_\rho z \to xy =^e_\tau xz),\\
&x =^e_{\tau\rho} y :\equiv \forall z^\rho(E^\rho(z) \to xz =^e_\tau yz) \wedge E^{\tau\rho}(x) \wedge E^{\tau\rho}(y).
\end{aligned}$$

Definition 6.13 *For every formula A of E-PA$^\omega$ we define a translation A_e by relativizing all quantifiers to hereditarily extensional functionals in the sense of E^ρ:*

(i) $A_e :\equiv A$, *if A is a prime formula,*

(ii) $(A\Box B)_e :\equiv (A_e \Box B_e)$, *where* $\Box \in \{\wedge, \vee, \to\}$,

(iii) $(\exists x^\rho A(x))_e :\equiv \exists x^\rho (E^\rho(x) \wedge A_e(x))$,

(iv) $(\forall x^\rho A(x))_e :\equiv \forall x^\rho (E^\rho(x) \to A_e(x))$.

Proposition 6.14

$$\begin{cases} \textit{E-PA}^\omega + \textit{QF-AC}^{0,1} + \textit{QF-AC}^{1,0} \vdash A(\underline{a}) \;\Rightarrow \\ \textit{WE-PA}^\omega + \textit{QF-AC}^{0,1} + \textit{QF-AC}^{1,0} \vdash E(\underline{a}) \to A_e(\underline{a}), \end{cases}$$

where $\underline{a}$ *are all the free variables of* A.

Proof: See [52]. □

Let $\widehat{\text{(W)E-PA}}^\omega\!\restriction$ ($\widehat{\text{(W)E-HA}}^\omega\!\restriction$) be the fragment of (W)E-PA$^\omega$ ((W)E-HA$^\omega$) where we only have the recursor R_0 for type-0-recursion and the induction schema is restricted to the schema of quantifier-free induction

$$\text{QF-IA} \;:\; A_0(0) \wedge \forall x^0(A_0(x) \to A_0(x+1)) \to \forall x^0 A_0(x).$$

Lemma 6.15 $\widehat{\textit{WE-PA}}^\omega\!\restriction + \textit{QF-AC}^{0,0} \vdash \Sigma^0_1$ -IA, *where*

$$\Sigma^0_1\text{-IA} :\; \exists y^0 A_0(0,y) \wedge \forall x^0(\exists y A_0(x,y) \to \exists y A_0(x+1,y)) \to \forall x \exists y A_0(x,y).$$

Proof: Assume $\exists y_0 A_0(0,y_0)$ and $\forall x, y_1 \exists y_2 (A_0(x,y_1) \to A_0(x+1,y_2))$. By QF-AC^{0,0} we get

$$\exists f \forall x, y_1 (A_0(x,y_1) \to A_0(x+1, fxy_1)).$$

Define

$$\begin{cases} \Phi(0,y,f) :=_0 y \\ \Phi(x+1,y,f) :=_0 f(x, \Phi(x,y,f)) \end{cases}$$

(note that this can by done by R_0). Then by QF-IA one easily shows that

$$\forall x A_0(x, \Phi(x, y_0, f))$$

for y_0 such that $A_0(0,y_0)$ and therefore $\forall x \exists y A_0(x,y)$. □

Proposition 6.16 ([11])

$$\begin{cases} \widehat{WE\text{-}PA}^{\omega}\!\restriction + QF\text{-}AC \vdash A(\underline{a}) \\ \Rightarrow \exists \textit{ closed terms } \underline{t} \textit{ of } \widehat{WE\text{-}HA}^{\omega}\!\restriction \textit{ such that} \\ \widehat{WE\text{-}HA}^{\omega}\!\restriction \vdash \forall \underline{y}(A')_D(\underline{t}\,\underline{a}, \underline{y}, \underline{a}) \end{cases}$$

(Here $\underline{a}$ are all of the free variables of $A(\underline{a})$.

Proof: The proof theorem 6.6 easily relativizes to $\widehat{\text{WE-PA}}^{\omega}\!\restriction$. □

One can show that the function(al)s of types ≤ 2 definable by closed terms in $\widehat{\text{WE-HA}}^{\omega}\!\restriction$ are just the usual primitive recursive ones in the sense of Kleene (see [11]). From the proof of this fact combined with the previous proposition plus elimination of extensionality one gets

Proposition 6.17 *Let $R(x, y)$ be a primitive recursive relation (in the sense of PRA). Then the following rule holds:*

$$\begin{cases} \widehat{E\text{-}PA}^{\omega}\!\restriction + QF\text{-}AC^{1,0} + QF\text{-}AC^{0,1} \vdash \forall x^0 \exists y^0 R(x, y) \\ \Rightarrow \exists \textit{ primitive recursive function } p \textit{ such that} \\ PRA \vdash R(x, px). \end{cases}$$

Let PA_1 be the restriction of PA to induction for Σ^0_1-formulas only.

Corollary 6.18 (Parsons,Mints,Takeuti,...) *PA_1 is Π^0_2-conservative over PRA.*

Extensions to fragments of analysis

An analytical principle which has received great attention during the last 20 years in proof theory is the binary ('weak') König's lemma WKL formalized by the following axiom

$$\begin{array}{l}\mathrm{WKL} :\equiv \\ \forall f^1(T(f) \wedge \forall x^0 \exists n^0(lth(n) =_0 x \wedge f(n) =_0 0) \to \exists b \leq_1 1 \forall x^0(f(\overline{b}x) =_0 0)),\end{array}$$

where

$$T(f) :\equiv \forall n^0, m^0(f(n*m) = 0 \to f(n) = 0) \wedge \forall n^0, x^0(f(n*\langle x\rangle) = 0 \to x \leq 1)$$

(here $lth, *, \overline{b}x, \langle\cdot\rangle$ refer to a standard primitive recursive coding of finite sequences of numbers).

The interest in WKL rests on the following facts:

1) Already relative to a second-order fragment RCA_0 of $\widehat{\mathrm{WE\text{-}PA}}^{\omega}\!\upharpoonright$+ QF-$\mathrm{AC}^{0,0}$, WKL proves substantial parts of (non-constructive) classical mathematics and in particular of analysis (see e.g. [12],[62],[63],[6],[7], [25],[59]).

2) Despite of the mathematical strength of WKL, it is weak from the proof-theoretic point of view, namely a classical result due to H. Friedman states that RCA_0+ WKL is Π^0_2-conservative over PRA.

One can use functional interpretation combined with majorization to prove various generalizations of Friedman's result, e.g.:

Theorem 6.19 ([30]) *Let $A_0(x, y, z)$ be a (quantifier-free) formula of $\widehat{E\text{-}PA}^{\omega}\!\upharpoonright$ containing only x^1, y^1, z^0 as free variables and let s be a closed term. Then the following rule holds:*

$$\begin{cases} \widehat{E\text{-}PA}^{\omega}\!\upharpoonright + \textit{QF-AC}^{1,0} + \textit{QF-AC}^{0,1} + \textit{WKL} \vdash \forall x^1 \forall y \leq_1 sx \exists z^0 A_0(x, y, z) \\ \Rightarrow \textit{ one can extract a primitive recursive (in Kleene's sense) } \Phi \textit{ such that} \\ \widehat{WE\text{-}HA}^{\omega}\!\upharpoonright \vdash \forall x^1 \forall y \leq_1 sx \exists z \leq_0 \Phi x\, A_0(x, y, z). \end{cases}$$

Corollary 6.20 ([30]) *$\widehat{E\text{-}PA}^{\omega}\restriction + QF\text{-}AC^{1,0} + QF\text{-}AC^{0,1} +$ WKL is Π^0_2-conservative over PRA.*

The proofs for these results go beyond the scope of our lectures. One problem to be dealt with is that the functional interpretation of the negative translation of WKL is not even solvable in E-PA$^{\omega}$.

Chapter 7

The Friedman A-translation

In [13] H. Friedman introduced a strikingly simple device to establish closure under the Markov rule in the form

$$T_i \vdash \neg\neg\exists x P(\underline{a}, x) \Rightarrow T_i \vdash \exists x P(\underline{a}, x)$$

(P is a prime formula)[1] which works for many intuitionistic theories T_i. As we have seen one can use functional interpretation to obtain closure under the Markov rule of those theories to which functional interpretation applies. The important feature of Friedman's so-called A-translation however is that it is much easier to apply and also works for some systems like intuitionistic Zermelo-Fraenkel set theory ZFI for which no functional interpretation has been developed yet.

Combined with the negative translation, Friedman's A-translation therefore can be used to show Π^0_2-conservation of many classical theories T over their intuitionistic counterpart T_i.

As a corollary of this one gets that T has the same provably recursive functions as T_i.

By combining negative translation and A-translation with modified realizability one obtains an alternative method (to the use of negative translation and functional interpretation) for unwinding proofs of Π^0_2-sentences in e.g. PA. Note that the direct combination of negative translation and modified realizability without the intermediate step of the A-translation would be

[1] In theories in which every quantifier-free formula $A_0(\underline{x})$ can be written as a prime formula $t_{A_0}(\underline{x}) = 0$ this implies the usual form of the Markov rule.

useless since the modified realizability interpretation is trivial for negative formulas which result under negative translation.

However one should mention also some serious limitations of the approach based on the A-translation:

1) the A-translation is not sound for QF-ER and so doesn't apply to our systems WE-HA$^\omega$ and $\widehat{\text{WE-HA}}^\omega\!\upharpoonright$ while functional interpretation does.

2) A-translation only shows closure under the Markov **rule** but doesn't establish conservation results for the Markov **principle** with respect to general classes of formulas (which functional interpretation does). In particular it is not sound for the negative translation of QF-AC (which follows from QF-AC only in the presence of the Markov principle) and therefore cannot be used to show that e.g. the provably recursive functions of WE-PA$^\omega$+ QF-AC are just the ones definable by closed terms of WE-HA$^\omega$ even if we omit the extensionality rule QF-ER.

3) The combination of negative translation, A-translation and modified realizability is not known to be faithful for subsystems PA$_n$ of PA with restricted induction (respectively for corresponding finite type extensions of PA$_n$) whereas negative translation combined with functional interpretation is (see [56] for the latter).

Remark 7.1 *The A-translation was independently also discovered by A. Dragalin in [10].*

In this lecture we will establish the A-translation only for HA since this suffices to illustrate the general pattern. For extensions to other systems like ZFI the reader should consult Friedman's original paper [13].

Definition 7.2 ([13]) *Let $A \in \mathcal{L}(\text{HA})$ be a formula of* HA. *For every formula $F \in \mathcal{L}(\text{HA})$ (such that A doesn't contain free variables which are bound in F) we define the A-translation F^A of F as follows: F^A results if all prime formulas P in F are replaced by $P \vee A$.*

Proposition 7.3 $HA \vdash F \Rightarrow HA \vdash F^A$.

Proof: Easy induction on the length of the derivation. □

Corollary 7.4 $HA \vdash \forall x \neg\neg \exists y A_0(x,y) \Rightarrow HA \vdash \forall x \exists y A_0(x,y)$.

Proof: In HA, $A_0(x,y)$ can be written as a prime formula $t_{A_0}(x,y) = 0$. Hence (since $\neg G$ is an abbreviation for $G \to 0 = 1$)

$$\text{HA } \vdash \forall x \neg\neg \exists y A_0(x,y)$$

implies

$$\text{HA } \vdash (\exists y(t_{A_0}(x,y) = 0) \to 0 = 1) \to 0 = 1.$$

By the A-translation for $A :\equiv \exists y(t_{A_0}(x,y) = 0)$ we get

$$\text{HA } \vdash ((\exists y(t_{A_0}(x,y) = 0) \to 0 = 1) \to 0 = 1)^{\exists y(t_{A_0}(x,y)=0)},$$

i.e. HA proves

$$(\exists y(t_{A_0}(x,y) = 0 \vee \exists y(t_{A_0}(x,y) = 0)) \to \exists y(t_{A_0}(x,y) = 0)) \to \exists y(t_{A_0}(x,y) = 0),$$

and hence HA proves

$$((\exists y t_{A_0}(x,y) = 0 \vee \exists y(t_{A_0}(x,y) = 0)) \to \exists y(t_{A_0}(x,y) = 0)) \to \exists y(t_{A_0}(x,y) = 0).$$

Since $G \vee G \to G$ holds by intuitionistic logic, we get

$$\text{HA } \vdash \exists y(t_{A_0}(x,y) = 0)$$

and hence

$$\text{HA } \vdash \forall x \exists y A_0(x,y).$$

□

For more information on the A-translation see [51]. Applications of the combination of negative translation, A-translation and realizability can be found in [4] and [55].

Chapter 8

Final comments

In the previous lectures we have studied various proof interpretations and indicated their use as tools for unwinding the computational content of proofs (both constructive as well as non-constructive ones).
A common feature of all these interpretations is that they translate a system $\mathcal{T}$ into another system $\mathcal{S}$ by assigning to every formula A of the former a formula A^* of the latter such that the implication

$$\mathcal{T} \vdash A \;\Rightarrow\; \mathcal{S} \vdash A^*$$

holds. Moreover the proof of A^* in $\mathcal{S}$ can be obtained by a simple recursion over a given proof of A in $\mathcal{T}$ since the interpretations respect the logical deduction rules (**locality** or **modularity** of proof interpretations).

As a consequence of this such proof interpretations preserve to a certain extent the structure of the original proof and the resulting $\mathcal{S}$-proof of A^* will not be much longer than the original proof of A in $\mathcal{T}$. This is in sharp contrast to structural proof transformations like cut-elimination or normalization which in general cause a non-elementary recursive blow-up of the original proof. Of course at a few places (proposition 6.17, corollary 6.20) we had to normalize the term extracted by the proof interpretation which again is of non-elementary complexity. However as we have seen one can also make substantial use of terms involving higher types by exploiting the mathematical structure of the functionals denoted by these terms without having to normalize them (see e.g. theorem 4.11, corollary 4.13, proposition 5.15 and

– if we allow Φ to be an arbitrary closed term of $\widehat{\text{WE-HA}}^{\omega}\!\restriction$– also theorem 6.19)! So proof interpretations of the sort we investigated in the present lectures allow to separate those aspects of unwinding proofs which can be carried out locally by recursion over the proof from those which involve a global rebuilding of a proof or a term like normalization.

Another important consequence of the modularity of proof interpretations is that they can be easily extended to systems $\tilde{\mathcal{T}} \supset \mathcal{T}$ obtained by adding further non-logical axioms Γ to $\mathcal{T}$. If the interpretation Γ^* of Γ is provable in $\mathcal{S}$ (resp. in some extension $\tilde{\mathcal{S}}$ of $\mathcal{S}$), then the given interpretation immediately extends to an interpretation of $\tilde{\mathcal{T}}$ in $\mathcal{S}$ (resp. in $\tilde{\mathcal{S}}$). So it suffices to consider the new axioms.

As a simple example for such an extension consider e.g. the following: both functional interpretation and negative translation are trivial for purely universal sentences $\Gamma := \forall x A_0(x)$. Because of this the proofs of e.g. theorem 6.6 and corollary 6.7 immediately extend to WE-PA$^\omega + \Gamma$, WE-HA$^\omega + \Gamma$ instead of WE-PA$^\omega$, WE-HA$^\omega$ (for Γ in the language of WE-PA$^\omega$).
As a corollary we obtain that the addition of ($\mathcal{S}^\omega$-true) universal axioms to WE-PA$^\omega$+ QF-AC doesn't change the provably recursive functionals of the system. This observation –which has been stressed in the context of first-order arithmetic by G. Kreisel– can be extended also to more general classes of formulas (see e.g. [30]).

Finally, proof interpretations can easily be combined with each other: e.g in chapter 6 we used a combination of three different interpretations: elimination of extensionality, negative translation and functional interpretation.

Bibliography

[1] Avigad, J., Feferman, S., Gödel's functional ('Dialectica') interpretation. To appear in: Handbook of Proof Theory (edited by S. Buss).

[2] Beeson, M., Goodman's theorem and beyond. Pacific J. Math. **84**, pp. 1-16 (1979).

[3] Berger, U., Program extraction from normalization proofs. In: M. Bezem et al.(eds.), TLCA'93, pp. 91-106. Springer LNCS **664** (1993).

[4] Berger,U., Schwichtenberg, H., Program extraction from classical proofs. In: Leivant, D. (ed.), Logic and Computational Complexity Workshop LCC'94. Springer LNCS **960**, pp. 77-97 (1995).

[5] Bishop, E., Mathematics as a numerical language. In: Kino, Myhill, Vesley (eds.), Intuitionism and Proof Theory, pp. 53-71, North-Holland, Amsterdam (1970).

[6] Brown, D.K.– Simpson, S.G., Which set existence axioms are needed to prove the separable Hahn–Banach theorem? Ann. Pure Appl. Logic **31**, pp. 123–144 (1986).

[7] Brown, D., Giusto, M., Simpson, S., Vitali's theorem and WWKL. Preprint (1997).

[8] Buss, S.R., On Herbrand's Theorem. In: Logic and Computational Complexity (Leivant, D., ed.), pp. 195-209. Springer LNCS **960** (1995).

[9] Cook, S., Urquart, A., Functional interpretations of feasibly constructive arithmetic. Ann. Pure Applied Logic **63**, pp. 103-200 (1993).

[10] Dragalin, A.G., New kinds of realizability and the Markov rule. Dokl. Akad. Nauk. SSSR **251**, pp. 534-537 (1980) (Russian). English translation: Soviet Math. Dokl. **21**, pp. 461-464 (1980).

[11] Feferman, S., Theories of finite type related to mathematical practice. In: Barwise, J. (ed.), Handbook of Mathematical Logic, pp. 913-972, North-Holland, Amsterdam (1977).

[12] Friedman, H., Systems of second-order arithmetic with restricted induction (abstract). J. Symbolic Logic **41**, pp. 558-559 (1976)

[13] Friedman, H., Classical and intuitionistically provably recursive functions. In: Müller,G.H., Scott, D.S. (eds.), Higher Set Theory, pp. 21-27. Springer LNM **669** (1978).

[14] Girard, J.-Y., Proof Theory and Logical Complexity Vol.I. Studies in Proof Theory. Bibliopolis (Napoli) and Elsevier Science Publishers (Amsterdam) 1987.

[15] Glivenko, V.I., Sur quelques points de la logique de M. Brouwer. Bull. Soc. Math. Belg. **15**, pp. 183-188 (1929).

[16] Gödel, K., Zur intuitionistischen Arithmetik und Zahlentheorie. Ergebnisse eines Mathematischen Kolloquiums, vol. 4 pp. 34–38 (1933).

[17] Gödel, K., Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes. Dialectica **12**, pp. 280–287 (1958).

[18] Gödel, K., Collected Work, Vol.2, S. Feferman et al. eds. Oxford University Press. New York

[19] Goodman, N., The faithfulness of the interpretation of arithmetic in the theory of constructions. J. Symbolic Logic **38**, pp.453-459 (1973).

[20] Goodman, N., The theory of the Gödel functionals. J. Symbolic Logic **41**, pp. 574-582 (1976).

[21] Goodman, N., Relativized realizability in intuitionistic arithmetic of all finite types. J. Symbolic Logic **43**, pp. 23-44 (1978).

[22] Hacks, J., Einige Anwendungen der Funktion $[x]$. Acta Math. **14**, 329-336 (1890).

[23] Hardy, G.H., Wright, E.M., An introduction to the theory of numbers. 5th edition, Oxford Science Publications (1979).

[24] Howard, W.A., Hereditarily majorizable functionals of finite type. In: Troelstra (ed.), Metamathematical investigation of intuitionistic arithmetic and analysis, pp. 454-461. Springer LNM 344 (1973).

[25] Humphreys, A.J., Simpson, S., Separation and weak König's lemma. Preprint (1997).

[26] Kleene, S.C., On the interpretation of intuitionistic number theory. J. Symbolic Logic **10**, pp. 109-124 (1945).

[27] Kleene, S.C., Introduction to Metamathematics. North-Holland (Amsterdam), Noordhoff (Groningen), Van Nostrand (New-York) 1952.

[28] Kohlenbach, U., Pointwise hereditary majorization and some applications. Arch. Math. Logic **31**, pp. 227-241 (1992).

[29] Kohlenbach, U., Remarks on Herbrand normal forms and Herbrand realizations. Arch. Math. Logic **31**, pp. 305-317 (1992).

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

[31] Kohlenbach, U., Effective moduli from ineffective uniqueness proofs. An unwinding of de La Vallée Poussin's proof for Chebycheff approximation. Ann. Pure Appl. Logic **64**, pp. 27–94 (1993).

[32] Kohlenbach, U., New effective moduli of uniqueness and uniform a–priori estimates for constants of strong unicity by logical analysis of known proofs in best approximation theory. Numer. Funct. Anal. and Optimiz. **14**, pp. 581–606 (1993).

[33] Kohlenbach, U., Analysing proofs in analysis. In: W. Hodges, M. Hyland, C. Steinhorn, J. Truss, editors, *Logic: from Foundations to Applications. European Logic Colloquium* (Keele, 1993), pp. 225–260, Oxford University Press (1996).

[34] Kohlenbach, U., Mathematically strong subsystems of analysis with low rate of growth of provably recursive functionals. Arch. Math. Logic **36**, pp. 31–71 (1996).

[35] Kohlenbach, U., Elimination of Skolem functions for monotone formulas in analysis. To appear in: Archive for Mathematical Logic.

[36] Kohlenbach, U., The use of a logical principle of uniform boundedness in analysis. To appear in: Proc. 'Logic in Florence 1995'.

[37] Kohlenbach, U., Arithmetizing proofs in analysis. To appear in: Proc. Logic Colloquium 96 (San Sebastian).

[38] Kohlenbach, U., Relative constructivity. To appear in: J. Symbolic Logic.

[39] Kohlenbach, U., A note on Goodman's theorem. To appear in: Studia Logica.

[40] Kohlenbach, U., On the no-counterexample interpretation. BRICS Report Series RS-97-42, 26pp., to appear in: J. Symbolic Logic.

[41] Kolmogorov, A.N., On the principle of the excluded middle (Russian). Mat. Sb. **32**, pp.646-667 (1925).

[42] Kreisel, G., On the interpretation of non-finitist proofs, part I. J. Symbolic Logic **16**, pp.241-267 (1951).

[43] Kreisel, G., On the interpretation of non-finitist proofs, part II: Interpretation of number theory, applications. J. Symbolic Logic **17**, pp. 43-58 (1952).

[44] Kreisel, G., Interpretation of analysis by means of constructive functionals of finite types. In: Constructivity in Mathematics, Heyting, A. (ed.). North-Holland (Amsterdam), pp. 101-128 (1959).

[45] Kreisel, G., On weak completeness of intuitionistic predicate logic. J. Symbolic Logic **27**, pp.139–158 (1962).

[46] Kreisel, G., Foundations of intuitionistic logic. In: Proc. Logic Methodology and Philosophy of Science, Nagel, E., Suppes, P., Tarski, A. (eds.). Stanford University Press (Stanford), pp. 198-210 (1962).

[47] Kreisel, G., Finiteness theorems in arithmetic: an application of Herbrand's theorem for Σ_2-formulas. Proc. of the Herbrand symposium (Marseille, 1981), North-Holland (Amsterdam), pp. 39-55 (1982).

[48] Kreisel, G., Proof Theory and Synthesis of Programs: Potentials and Limitations. Eurocal '85 (Linz 1985), Springer LNCS **203**, pp. 136-150 (1986).

[49] Kreisel, G., Macintyre, A., Constructive logic versus algebraization I. Proc. L.E.J. Brouwer Centenary Symposium (Noordwijkerhout 1981), North-Holland (Amsterdam), pp. 217-260 (1982).

[50] Kuroda, S., Intuitionistische Untersuchungen der formalistischen Logik. Nagoya Math. **3**, pp. 35-47 (1951).

[51] Leivant, D., Syntactic translations and provably recursive functions. J. Symbolic Logic **50**, pp. 252-258 (1985).

[52] Luckhardt, H., Extensional Gödel Functional Interpretation. Springer Lecture Notes in Mathematics **306**, 1973.

[53] Luckhardt, H., Herbrand-Analysen zweier Beweise des Satzes von Roth: Polynomiale Anzahlschranken. J. Symbolic Logic **54**, pp. 234-263 (1989).

[54] Mints, G.E., Finite investigations of transfinite derivations. J. Soviet Math. **10**, pp. 548-596 (1978).

[55] Murthy, C., Extracting Constructive Content from Classical Proofs. PhD thesis, Cornell University (1990).

[56] Parsons, C., On n-quantifier induction. J. Symbolic Logic **37**, pp. 466-482 (1972).

[57] Renardel de Lavalette, G.R., Extended bar induction in applicative theories. Ann. Pure Applied Logic **50**, 139-189 (1990).

[58] Ruitenberg, W., Basic logic and Fregean set theory. In: Barendregt, Bezem Klop (eds.), Dirk van Dalen Festschrift. Quaestiones Infinitae Vol. V. Utrecht University 1993.

[59] Shioji, N.–Tanaka, K., Fixed point theory in weak second–order arithmetic. Ann. Pure Appl. Logic **47**,pp. 167–188 (1990).

[60] Shoenfield, J.S., Mathematical Logic. Addison-Wesley Publishing Company (Reading, Massachusetts) 1967.

[61] Sieg, W., Fragments of arithmetic. Ann. Pure Appl. Logic **28**,pp. 33–71 (1985).

[62] Simpson, S.G., Which set existence axioms are needed to prove the Cauchy/Peano theorem for ordinary differential equations. J. Symbolic Logic **49**,pp. 783–801 (1984).

[63] Simpson, S.G., Reverse Mathematics. Proc. Symposia Pure Math. **42**, pp. 461–471, AMS, Providence (1985).

[64] Troelstra, A.S., Principles of Intuitionism. Springer Lecture Notes in Mathematics Vol. 95, Berlin - Heidelberg - New York, Springer 1969.

[65] Troelstra, A.S. (ed.) Metamathematical investigation of intuitionistic arithmetic and analysis. Springer Lecture Notes in Mathematics **344** (1973).

[66] Troelstra, A.S., Some models for intuitionistic finite type arithmetic with fan functional. J. Symbolic Logic **39**, pp. 194-202 (1977).

[67] Troelstra, A.S., Realizability. To appear in: Handbook of Proof Theory (edited by S. Buss).

[68] Troelstra, A.S., van Dalen, D., Constructivism in Mathematics, vol.1. North-Holland (Amsterdam) 1988.