NEW DEVELOPMENTS IN PROOFS AND COMPUTATIONS

HELMUT SCHWICHTENBERG

It is a tempting idea to use formal existence proofs as a means to precisely and verifiably express algorithmic ideas. This is clearly possible for "constructive" proofs, which are informally understood via the Brouwer-Heyting-Kolmogorov interpretation (BHK-interpretation for short). This interpretation of intuitionistic (and minimal) logic explains what it means to prove a logically compound statement in terms of what it means to prove its components; the explanations use the notions of *construction* and *constructive proof* as unexplained primitive notions. For prime formulas the notion of proof is supposed to be given. The clauses of the BHK-interpretation are:

- p proves $A \wedge B$ if and only if p is a pair $\langle p_0, p_1 \rangle$ and p_0 proves A, p_1 proves B;
- p proves $A \to B$ if and only if p is a construction transforming any proof q of A into a proof $p(q)$ of B;
- $\bot$ is a proposition without proof;
- p proves $\forall_{x \in D} A(x)$ if and only if p is a construction such that for all $d \in D$, $p(d)$ proves $A(d)$;
- p proves $\exists_{x \in D} A(x)$ if and only if p is of the form $\langle d, q \rangle$ with d an element of D, and q a proof of $A(d)$.

The problem with the BHK-interpretation is its reliance on the unexplained concepts of construction and constructive proof. Gödel (1958) tried to replace the notion of constructive proof by something more definite, less abstract, his principal candidate being a notion of "computable functional of finite type" which is to be accepted as sufficiently well understood to justify the axioms and rules of his system T, an essentially logic-free theory of functionals of finite type. One only needs to know that certain basic functionals are computable (including primitive recursion operators in finite types), and that the computable functionals are closed under composition.

The general framework for proof interpretations as we understand it is to assign to every formula A a new one $\exists_x A_1(x)$ with $A_1(x)$ $\exists$-free. Then from a derivation $M \colon A$ we want to extract a "realizing" term r such that $A_1(r)$ can be proved. The intention here is that its meaning should in some sense be related to the meaning of the original formula A. The well-known (modified) realizability interpretation and Gödel's Dialectica interpretation both fall under this scheme (cf. Oliva (2006)). However, Gödel explicitely states in (1958, p.286) that his Dialectica interpretation is *not* the one intended by BHK-interpretation.

One might think that from the informal idea of a particular constructive proof it should be clear what its algorithmic content is. This, however, is not always true. An example is Tait's proof of the existence of normal forms

for the simply typed λ-calculus, which uses so-called computability predicates. Somewhat unexpectedly, it turns out that its computational content is the normalization-by-evaluation algorithm. This has first been observed by Berger (1993), and formally treated (including machine extraction of programs) in Berger et al. (2006).

An even greater challenge is the task of finding computational content in proofs of *classical* existence theorems, of the form $\neg\forall_y\neg A_0(y)$ with $A_0(y)$ quantifier-free; we use the shorthand $\tilde{\exists}_y A_0(y)$ for such formulas. It is well-known that we need to require that the kernel $A_0(y)$ is quantifier-free. Then the whole proof can be seen as deriving falsity from the (false) assumption $\forall_y\neg A_0(y)$. Now consider the long normal form of this proof. In this long normal form, each instance of the false assumption $\forall_y\neg A_0(y)$ must be applied to a closed term r_i of type $\mathbf{N}$, and for at least one of those r_i the kernel $\neg A_0(r_i)$ must be false and hence $A_0(r_i)$ true. This "direct method" has been described in Schwichtenberg (1993); in Berger and Schwichtenberg (1995) it has been shown that it gives the same results as the so-called A-translation of Friedman (1978) (and moreover, that we have the same algorithm in both cases). A refined form of the A-translation has been introduced in Berger et al. (2002), and further studied and applied in Berger et al. (2001); Seisenberger (2003).

An alternative to extract computational content from proofs of classical existence theorems is Gödel's Dialectica interpretation (1958), which is what we want to concentrate on in the present paper. Gödel assigned to every formula A a new one $\exists_{\vec{x}}\forall_{\vec{y}} A_D(\vec{x},\vec{y})$ with $A_D(\vec{x},\vec{y})$ quantifier-free. Here $\vec{x}$, $\vec{y}$ are lists of variables of finite types; the use of higher types is necessary even when the original formula A was first-order. He did this in such a way that whenever a proof of A say in constructive arithmetic was given, one could produce closed terms $\vec{r}$ such that the quantifier-free formula $A_D(\vec{r},\vec{y})$ is provable in T.

In (1958) Gödel referred to a Hilbert-style proof calculus. However, since the realizers will be formed in a λ-calculus formulation of system T, Gödel's interpretation becomes a lot more perspicious when it is done for a natural deduction calculus. Such a natural deduction based treatment of the Dialectica interpretation has been given by Jørgensen (2001) and Hernest (2006). Both authors use a formulation of natural deduction where open assumptions are viewed as *formulas*, and consequently the necessity of contractions arises when an application of the implication introduction rule $\to^+$ discharges more than one assumption formula. However, it seems to be more in the spirit of the Curry-Howard correspondence (formulas correspond to types, and proofs to terms) to view assumptions as *assumption variables*. This is particularly important when – say in an implementation – one wants to assign object terms ("realizers", in Gödel's T) to proof terms. To see the point, notice that a proof term M may have many occurrences of a free assumption variable u^A. The associated realizer $[\![M]\!]$ then needs to contain an object variable $x_u^{\tau(A)}$ uniquely associated with u^A, again with many occurrences. To organize this in an appropriate way it seems mandatory to be able to refer to an assumption A by means of its "label" u. The present exposition differs from previous ones mainly in this respect.

The rest of the paper is rather technical. We give a detailed natural deduction based proof of the soundness theorem for the Dialectica interpretation, and also extend it to the Dialectica interpretation with majorants (or "monotone" Dialectica interpretation), introduced by Kohlenbach (1992, 1996).

The main motivation for this work has been the desire to have a clean and explicit natural deduction based proof of the soundness theorem, for the exact Dialectica interpretation as well as for its variant with majorants, in such a way that this proof can be used as a template for an implementation. For the very same reason we have added a simplified and implementation-friendly proof of the fact – first observed by Kohlenbach (1992) – that WKL can be formulated as a $\forall\exists_{\leq}\forall$-axiom, and hence is covered by the Dialectica interpretation with majorants. However, it remains to be seen to what extent such an implementation will succeed in producing informative and usable realizers. A promising first step in this direction has been done by Hernest (2006); particularly interesting is his successful integration of the non-computational ("uniform") quantifiers of Berger (1993, 2005).

We begin in Sec.1 with a description of the arithmetic HA^ω in finite types that we consider. Sec.2 contains a proof of the Soundness Theorem for Gödel's Dialectica interpretation, and Sec.3 gives the majorant-based version of it. The final subsection contains a proof that WKL can be formulated as a $\forall\exists_{\leq}\forall$-axiom.

1. Arithmetic in Finite Types

1.1. **Types.** Our type system is defined by two type forming operations: arrow types $\rho \to \sigma$ and the formation of inductively generated types $\mu_{\vec{\alpha}}\vec{\kappa}$, where $\vec{\alpha} = (\alpha_j)_{j=1,\dots,N}$ is a list of distinct "type variables", and $\vec{\kappa} = (\kappa_i)_{i=1,\dots,k}$ is a list of "constructor types", whose argument types contain $\alpha_1, \dots, \alpha_N$ in strictly positive positions only.

For instance, $\mu_\alpha(\alpha, \alpha \to \alpha)$ is the type of natural numbers; here the list $(\alpha, \alpha \to \alpha)$ stands for two generation principles: α for "there is a natural number" (the 0), and $\alpha \to \alpha$ for "for every natural number there is a next one" (its successor).

Definition. Let $\vec{\alpha} = (\alpha_j)_{j=1,\dots,N}$ be a list of distinct type variables. *Types* $\rho, \sigma, \tau, \mu \in \mathrm{Ty}$ and *constructor types* $\kappa \in \mathrm{KT}_{\vec{\alpha}}$ are defined inductively:

$$\frac{\rho, \sigma \in \mathrm{Ty}}{\rho \to \sigma \in \mathrm{Ty}}, \quad \frac{\vec{\rho}, \vec{\sigma}_1, \dots, \vec{\sigma}_n \in \mathrm{Ty}}{\vec{\rho} \to (\vec{\sigma}_1 \to \alpha_{j_1}) \to \ldots \to (\vec{\sigma}_n \to \alpha_{j_n}) \to \alpha_j \in \mathrm{KT}_{\vec{\alpha}}}\ (n \geq 0),$$
$$\frac{\vec{\kappa} \in \mathrm{KT}_{\vec{\alpha}},\ \forall_{0<j\leq N}\exists_{j_1,\dots,j_n<j}\kappa_j = \vec{\rho} \to (\vec{\sigma}_1 \to \alpha_{j_1}) \to \ldots \to (\vec{\sigma}_n \to \alpha_{j_n}) \to \alpha_j}{(\mu_{\vec{\alpha}}(\kappa_1, \dots, \kappa_k))_j \in \mathrm{Ty}},$$

with $1 \leq N \leq k$; we call $\kappa_1, \dots, \kappa_N$ *nullary constructor types*.

Here $\vec{\rho} \to \sigma$ means $\rho_1 \to \ldots \to \rho_m \to \sigma$, associated to the right. We reserve μ for types of the form $(\mu_{\vec{\alpha}}(\kappa_1, \dots, \kappa_k))_j$. The *parameter types* of μ are the members of all $\vec{\rho}$ appearing in its constructor types $\kappa_1, \dots, \kappa_k$.

In the present paper it suffices to only consider the μ-types

$$\begin{aligned} &\mathbf{U} := \mu_\alpha \alpha, && \\ &\mathbf{B} := \mu_\alpha(\alpha, \alpha), && \mathrm{bin} := \mu_\alpha(\alpha, \alpha \to \alpha, \alpha \to \alpha), \\ &\mathbf{N} := \mu_\alpha(\alpha, \alpha \to \alpha), && \rho \wedge \sigma := \mu_\alpha(\rho \to \sigma \to \alpha). \end{aligned}$$

A type is *finitary* if it is a μ-type with all its parameter types $\vec{\rho}$ finitary, and all its constructor types are of the form $\vec{\rho} \to \alpha_{j_1} \to \ldots \to \alpha_{j_n} \to \alpha_j$, so the $\vec{\sigma}_1, \ldots, \vec{\sigma}_n$ in the general definition are all empty. For example, $\mathbf{U}$, $\mathbf{B}$, $\mathbf{N}$, bin are finitary, and $\rho \wedge \sigma$ is finitary provided its parameter types are.

1.2. **Constants.** For each of our base types we have *constructors* C^μ_i and *recursion operators* $\mathcal{R}^\tau_\mu$, as follows:

$$\begin{aligned} &\mathtt{tt}^{\mathbf{B}} := \mathrm{C}^{\mathbf{B}}_1, \quad \mathtt{ff}^{\mathbf{B}} := \mathrm{C}^{\mathbf{B}}_2, \\ &\mathcal{R}^\tau_{\mathbf{B}} \colon \mathbf{B} \to \tau \to \tau \to \tau, \\ &0^{\mathbf{N}} := \mathrm{C}^{\mathbf{N}}_1, \quad \mathrm{S}^{\mathbf{N}\to\mathbf{N}} := \mathrm{C}^{\mathbf{N}}_2, \\ &\mathcal{R}^\tau_{\mathbf{N}} \colon \mathbf{N} \to \tau \to (\mathbf{N} \to \tau \to \tau) \to \tau, \\ &1^{\mathrm{bin}} := \mathrm{C}^{\mathrm{bin}}_1, \quad \mathrm{S}^{\mathrm{bin}\to\mathrm{bin}}_0 := \mathrm{C}^{\mathrm{bin}}_2, \quad \mathrm{S}^{\mathrm{bin}\to\mathrm{bin}}_1 := \mathrm{C}^{\mathrm{bin}}_3, \\ &\mathcal{R}^\tau_{\mathrm{bin}} \colon \mathrm{bin} \to \tau \to (\mathrm{bin} \to \tau \to \tau) \to (\mathrm{bin} \to \tau \to \tau) \to \tau, \\ &\big(\wedge^+_{\rho\sigma}\big)^{\rho\to\sigma\to\rho\wedge\sigma} := \mathrm{C}^{\rho\wedge\sigma}_1, \\ &\mathcal{R}^\tau_{\rho\wedge\sigma} \colon \rho \wedge \sigma \to (\rho \to \sigma \to \tau) \to \tau. \end{aligned}$$

1.3. **Terms.** *Terms* are inductively defined from typed variables x^ρ and the constants, that is, constructors C^μ_i and recursion operators $\mathcal{R}^\tau_\mu$, by abstraction $(\lambda_{x^\rho} M^\sigma)^{\rho\to\sigma}$ and application $(M^{\rho\to\sigma} N^\rho)^\sigma$. It is well known that every such term has a uniquely determined long normal form w.r.t. β- and $\mathcal{R}$-conversions and η-expansions. We consider two terms to be *definitionally equal* if they have the same long normal form, and identify such terms.

Notice that in the more general setting of Schwichtenberg (2006), where we also allow constants defined by computation rules, definitional equality should mean that there is a purely equational proof of their equality based on β- and $\mathcal{R}$-conversions and η-expansions.

Notice also that the boolean "recursion" operator $\mathcal{R}^\tau_{\mathbf{B}}$ does not make any recursive calls. We denote $\mathcal{R}^\tau_{\mathbf{B}} trs$ by $[\mathbf{if}\ t\ \mathbf{then}\ r\ \mathbf{else}\ s]$ (which also indicates that this term should be evaluated "lazily").

Using the recursion operators we can define boolean-valued functions representing (decidable) equality $=_\mu \colon \mu \to \mu \to \mathbf{B}$ for finitary base types μ, for instance $\mathbf{N}$:

$$\begin{aligned} &(0 = 0) := \mathtt{tt}, && (\mathrm{S}(m) = 0) := \mathtt{ff}, \\ &(0 = \mathrm{S}(n)) := \mathtt{ff}, && (\mathrm{S}(m) = \mathrm{S}(n)) := (m = n). \end{aligned}$$

The *projections* of a pair to its components can be defined easily:

$$r0 := \mathcal{R}^\rho_{\rho\wedge\sigma} r^{\rho\wedge\sigma}(\lambda_{x^\rho, y^\sigma} x^\rho), \quad r1 := \mathcal{R}^\rho_{\rho\wedge\sigma} r^{\rho\wedge\sigma}(\lambda_{x^\rho, y^\sigma} y^\sigma).$$

We also define the *canonical inhabitant* ε^ρ of a type ρ:

$$\varepsilon^{\mu_j} := \mathrm{C}^{\vec{\mu}}_j \varepsilon^{\vec{\rho}}(\lambda_{\vec{x}_1} \varepsilon^{\mu_{j_1}}) \ldots (\lambda_{\vec{x}_n} \varepsilon^{\mu_{j_n}}), \quad \varepsilon^{\rho\to\sigma} := \lambda_x \varepsilon^\sigma.$$

There are many canonical isomorphisms between types; $(\rho \wedge \sigma \to \tau) \sim (\rho \to \sigma \to \tau)$ is an example. The isomorphism pairs can be constructed explicitly from the functions above.

1.4. **Formulas.** Atomic formulas are $\mathrm{atom}(r^{\mathbf{B}})$, indicating that the argument is true. We may also allow further predicate constants, for instance inductively defined ones, like Leibniz equality.

Notice that there is no need for (logical) falsity $\bot$, since we can take the atomic formula $F := \mathrm{atom}(\mathsf{ff})$ – called *arithmetical falsity* – built from the boolean constant ff instead.

The *formulas* of HA^ω are built from atomic ones by the connectives $\to$, $\forall$, $\exists$ and $\wedge$. We define *negation* $\neg A$ by $A \to F$.

1.5. **Proof terms.** We use Gentzen's natural deduction calculus for logical derivations consisting of the well-known rules $\to^+$, $\to^-$, $\forall^+$ and $\forall^-$. It will be convenient to write derivations as terms, where the derived formula is viewed as the type of the term. This representation is known under the name *Curry-Howard correspondence*.

We give an inductive definition of derivation terms in Table 1, where for clarity we have written the corresponding derivations to the left. For the universal quantifier $\forall$ there is an introduction rule $\forall^+ x$ and an elimination rule $\forall^-$, whose right premise is the term r to be substituted. The rule $\forall^+ x$ is subject to the following *(Eigen-) variable condition*: The derivation term M of the premise A should not contain any open assumption with x as a free variable.

1.6. **Axioms.** The logical axioms are the *truth axiom* $\mathrm{Ax}_{\mathsf{tt}} \colon \mathrm{atom}(\mathsf{tt})$, the introduction and elimination axioms $\exists^+$ and $\exists^-$ for existence and $\wedge^+$, $\wedge^-$ for conjunction:

$$\begin{aligned}
&\exists^+ \colon \forall_z(A \to \exists_z A),\\
&\exists^- \colon \exists_z A \to \forall_z(A \to B) \to B \quad (z \notin \mathrm{FV}(B)),\\
&\wedge^+ \colon A \to B \to A \wedge B,\\
&\wedge^- \colon A \wedge B \to (A \to B \to C) \to C,
\end{aligned}$$

and the induction axioms

$$\begin{aligned}
&\mathrm{Ind}_{p,A} \colon \forall_p\big(A(\mathsf{tt}) \to A(\mathsf{ff}) \to A(p^{\mathbf{B}})\big),\\
&\mathrm{Ind}_{n,A} \colon \forall_m\big(A(0) \to \forall_n(A(n) \to A(\mathrm{S}n)) \to A(m^{\mathbf{N}})\big),\\
&\mathrm{Ind}_{b,A} \colon \forall_b\big(A(1) \to \forall_b(A(b) \to A(\mathrm{S}_0 b)) \to \forall_b(A(b) \to A(\mathrm{S}_1 b)) \to A(b^{\mathrm{bin}})\big),\\
&\mathrm{Ind}_{x,A} \colon \forall_x\big(\forall_{y^\rho, z^\sigma} A(\langle y, z\rangle) \to A(x^{\rho\wedge\sigma})\big),
\end{aligned}$$

where $\langle y, z\rangle$ is shorthand for $\wedge^+ yz$. The final axiom expresses that every object of a pair type is a pair; it is sometimes called *pair elimination axiom*.

Using boolean induction $\mathrm{Ind}_{p,A}$ we can derive the arithmetical form of *ex-falso-quodlibet*, that is, $F \to \mathrm{atom}(p^{\mathbf{B}})$ (recall $F := \mathrm{atom}(\mathsf{ff})$), and then $F \to A$ for arbitrary formulas A. Similarly – again using the fact that we only have decidable atoms of the form $\mathrm{atom}(r^{\mathbf{B}})$ – we can prove *compatibility*

$$x_1 =_\mu x_2 \to A(x_1) \to A(x_2) \quad (\mu \text{ finitary base type}).$$

derivation	term
$u\colon A$	u^A
$\dfrac{\begin{array}{c}[u\colon A]\\ \mid M\\ B\end{array}}{A \to B}\,{\to}^+u$	$(\lambda_{u^A} M^B)^{A\to B}$
$\dfrac{\mid M \qquad\qquad \mid N}{A \to B \qquad\qquad A}\,{\to}^-$	$(M^{A\to B} N^A)^B$
$\dfrac{\mid M}{A}\,\forall^+x$ (with var.cond.)	$(\lambda_x M^A)^{\forall_x A}$ (with var.cond.)
$\dfrac{\mid M}{\forall_x A(x) \qquad r}\,\forall^-$	$(M^{\forall_x A(x)} r)^{A(r)}$

TABLE 1. Derivation terms for $\to$, $\forall$

Let HA^ω be the theory based on the axioms above including the induction axioms, and ML^ω be the (many-sorted) minimal logic, where the induction axioms are left out.

We define *pointwise equality* $=_\rho$, by induction on the type. $x =_\mu y$ for μ a finitary base type is already defined, and

$$(x =_{\rho\to\sigma} y) := \forall_z(xz =_\sigma yz),$$
$$(x =_{\rho\wedge\sigma} y) := (x0 =_\rho y0) \wedge (x1 =_\sigma y1).$$

The *extensionality axioms* are

$$y_1 =_\rho y_2 \to x^{\rho\to\sigma} y_1 =_\sigma x^{\rho\to\sigma} y_2.$$

We write E-HA^ω when the extensionality axioms are present.

In Troelstra (1973), Howard proved that already the first non trivial instance of the extensionality scheme

$$y_1 =_1 y_2 \to xy_1 =_{\mathbf{N}} xy_2$$

(with $1 := \mathbf{N} \to \mathbf{N}$) does not have a Dialectica realizer. In fact, he introduced the majorizing relation as a tool to prove this result. This is in

contrast to the realizability interpretation, where extensionality axioms are unproblematic, since they are $\exists$-free.

As a substitute for extensionality one may add the *weak extensionality rule*

$$\frac{A_0 \to r =_\rho s}{A_0 \to t(r) =_\sigma t(s)} \quad (A_0 \text{ quantifier-free})$$

to the formal system considered. This "rule" is special in the sense that its premise must have been derived *without open assumptions*. – Since the conclusion is (equivalent to) a purely universal formula, adding the weak extensionality rule does not change the behaviour of the formal system w.r.t. the Dialectica interpretation.

We write WE-HA$^\omega$ when the weak extensionality rule is present, but not the extensionality axioms.

We will also consider some more axiom schemes. The *axiom of choice* (AC) is the scheme

$$\forall_{x^\rho}\exists_{y^\sigma} A(x,y) \to \exists_{f^{\rho\to\sigma}}\forall_{x^\rho} A(x,f(x)). \tag{1}$$

Independence of premise (IP$_\forall$) is the scheme

$$(A \to \exists_{x^\rho} B) \to \exists_{x^\rho}(A \to B) \quad (x \notin \mathrm{FV}(A)) \tag{2}$$

with A of the form $\forall_{y^\sigma} A_0$, A_0 quantifier-free. Moreover, we need the (constructively doubtful) *Markov principle* (MP), for a higher type variable x^ρ and quantifier-free formulas A_0, B_0:

$$(\forall_{x^\rho} A_0 \to B_0) \to \exists_{x^\rho}(A_0 \to B_0) \quad (x^\rho \notin \mathrm{FV}(B_0)). \tag{3}$$

2. Gödel's Dialectica Interpretation

Gödel (1958) assigned to every formula A a new one $\exists_{\vec{x}}\forall_{\vec{y}} A_D(\vec{x},\vec{y})$ with $A_D(\vec{x},\vec{y})$ quantifier-free. Here $\vec{x}$, $\vec{y}$ are lists of variables of finite types; the use of higher types is necessary even when the original formula A was first-order. He did this in such a way that whenever a proof of A say in constructive arithmetic was given, one could produce closed terms $\vec{r}$ such that the quantifier-free formula $A_D(\vec{r},\vec{y})$ is provable in T. Rather than working with tupels of variables and terms, we prefer to work with product types, in order to simplify the implementation. So we assign to every formula A its Gödel translation $\exists_x \forall_y |A|^x_y$, with $|A|^x_y$ quantifier-free.

2.1. **Positive and negative types.** To determine the types of x and y, we assign to every formula A objects $\tau^+(A)$, $\tau^-(A)$ (a type or the "nulltype" symbol ε). $\tau^+(A)$ is intended to be the type of a (Dialectica-)realizer to be extracted from a proof of A, and $\tau^-(A)$ the type of a challenge for the claim that this term realizes A. The definition can be conveniently written if we extend the use of $\rho \to \sigma$ and $\rho \wedge \sigma$ to the nulltype symbol ε:

$$\begin{aligned} (\rho \to \varepsilon) &:= \varepsilon, & (\rho \wedge \varepsilon) &:= \rho, \\ (\varepsilon \to \sigma) &:= \sigma, & (\varepsilon \wedge \sigma) &:= \sigma, \\ (\varepsilon \to \varepsilon) &:= \varepsilon, & (\varepsilon \wedge \varepsilon) &:= \varepsilon. \end{aligned}$$

With this understanding of $\rho \to \sigma$ and $\rho \wedge \sigma$ we can simply write

$$\begin{aligned}
\tau^+(P(\vec{s})) &:= \varepsilon, & \tau^-(P(\vec{s})) &:= \varepsilon,\\
\tau^+(A \wedge B) &:= \tau^+(A) \wedge \tau^+(B), & \tau^-(A \wedge B) &:= \tau^-(A) \wedge \tau^-(B),\\
\tau^+(\forall_{x^\rho} A) &:= \rho \to \tau^+(A), & \tau^-(\forall_{x^\rho} A) &:= \rho \wedge \tau^-(A),\\
\tau^+(\exists_{x^\rho} A) &:= \rho \wedge \tau^+(A), & \tau^-(\exists_{x^\rho} A) &:= \tau^-(A).
\end{aligned}$$

and for implication

$$\begin{aligned}
\tau^+(A \to B) &:= \big(\tau^+(A) \to \tau^+(B)\big) \wedge \big(\tau^+(A) \to \tau^-(B) \to \tau^-(A)\big),\\
\tau^-(A \to B) &:= \tau^+(A) \wedge \tau^-(B).
\end{aligned}$$

In case $\tau^+(A)$ ($\tau^-(A)$) is $\neq \varepsilon$ we say that A has *positive (negative) computational content*. For formulas without positive or without negative content one can give an easy characterization, involving the well-known notion of positive or negative occurrences of quantifiers in a formula:

$$\begin{aligned}
&\tau^+(A) = \varepsilon \leftrightarrow A \text{ has no positive } \exists \text{ and no negative } \forall,\\
&\tau^-(A) = \varepsilon \leftrightarrow A \text{ has no positive } \forall \text{ and no negative } \exists,\\
&\tau^+(A) = \tau^-(A) = \varepsilon \leftrightarrow A \text{ is quantifier-free.}
\end{aligned}$$

Examples. (a) For quantifier-free A_0, B_0,

$$\begin{aligned}
&\tau^+(\forall_{x^\rho} A_0) = \varepsilon, \quad \tau^-(\forall_{x^\rho} A_0) = \rho,\\
&\tau^+(\exists_{x^\rho} A_0) = \rho, \quad \tau^-(\exists_{x^\rho} A_0) = \varepsilon,\\
&\tau^+(\forall_{x^\rho} \exists_{y^\sigma} A_0) = (\rho \to \sigma), \quad \tau^-(\forall_{x^\rho} \exists_{y^\sigma} A_0) = \rho.
\end{aligned}$$

(b) For arbitrary A, B, writing $\tau^\pm A$ for $\tau^\pm(A)$

$$\begin{aligned}
\tau^+(\forall_{z^\rho}(A \to B)) &= \rho \to (\tau^+ A \to \tau^+ B) \wedge (\tau^+ A \to \tau^- B \to \tau^- A),\\
\tau^+(\exists_{z^\rho} A \to B) &= (\rho \wedge \tau^+ A \to \tau^+ B) \wedge (\rho \wedge \tau^+ A \to \tau^- B \to \tau^- A),\\
\tau^-(\forall_{z^\rho}(A \to B)) &= \rho \wedge (\tau^+ A \wedge \tau^- B),\\
\tau^-(\exists_{z^\rho} A \to B) &= (\rho \wedge \tau^+ A) \wedge \tau^- B.
\end{aligned}$$

It is interesting to note that for an existential formula with a quantifier-free kernel the positive and negative type is the same, irrespective of the choice of the existential quantifier, constructive or classical.

Lemma. *$\tau^\pm(\tilde{\exists}_x A_0) = \tau^\pm(\exists_x A_0)$ for A_0 quantifier-free. In more detail,*
(a) *$\tau^+(\tilde{\exists}_x A) = \tau^+(\exists_x A) = \rho \wedge \tau^+(A)$ provided $\tau^-(A) = \varepsilon$,*
(b) *$\tau^-(\tilde{\exists}_x A) = \tau^-(\exists_x A) = \tau^-(A)$ provided $\tau^+(A) = \varepsilon$.*

Proof. For an arbitrary formula A we have

$$\begin{aligned}
&\tau^+(\forall_{x^\rho}(A \to \bot) \to \bot)\\
&= \tau^+(\forall_{x^\rho}(A \to \bot)) \to \tau^-(\forall_{x^\rho}(A \to \bot))\\
&= (\rho \to \tau^+(A \to \bot)) \to (\rho \wedge \tau^-(A \to \bot))\\
&= (\rho \to \tau^+(A) \to \tau^-(A)) \to (\rho \wedge \tau^+(A)),
\end{aligned}$$

$$\tau^+(\exists_{x^\rho} A) = \rho \wedge \tau^+(A).$$

Both types are equal if $\tau^-(A) = \varepsilon$. Similarly

$$\tau^-(\forall_{x^\rho}(A{\to}\bot) \to \bot) = \tau^+(\forall_{x^\rho}(A{\to}\bot)) = \tau^+(A{\to}\bot) = \tau^+(A) \to \tau^-(A),$$
$$\tau^-(\exists_{x^\rho} A) = \tau^-(A).$$

Both types are $= \tau^-(A)$ if $\tau^+(A) = \varepsilon$. □

2.2. **Gödel translation.** For every formula A and terms r of type $\tau^+(A)$ and s of type $\tau^-(A)$ we define a new quantifier-free formula $|A|^r_s$ by induction on A.

$$\begin{aligned} |P(\vec{s})|^r_s &:= P(\vec{s}), & |A \wedge B|^r_s &:= |A|^{r0}_{s0} \wedge |B|^{r1}_{s1},\\ |\forall_x A(x)|^r_s &:= |A(s0)|^{r(s0)}_{s1}, & |A \to B|^r_s &:= |A|^{s0}_{r1(s0)(s1)} \to |B|^{r0(s0)}_{s1}.\\ |\exists_x A(x)|^r_s &:= |A(r0)|^{r1}_s, \end{aligned}$$

The formula $\exists_x \forall_y |A|^x_y$ is called the *Gödel translation* of A and is often denoted by A^D. Its quantifier-free kernel $|A|^x_y$ is called *Gödel kernel* of A; it is denoted by A_D.

For readability we sometimes write terms of a pair type in pair form:

$$\begin{aligned} |\forall_z A|^f_{z,y} &:= |A|^{fz}_y, & |A \wedge B|^{x,z}_{y,u} &:= |A|^x_y \wedge |B|^z_u,\\ |\exists_z A|^{z,x}_y &:= |A|^x_y, & |A \to B|^{f,g}_{x,u} &:= |A|^x_{gxu} \to |B|^{fx}_u. \end{aligned}$$

Examples. (a) For quantifier-free formulas A_0, B_0 with $x^\rho \notin \mathrm{FV}(B_0)$

$$\begin{aligned} \tau^+(\forall_{x^\rho} A_0 \to B_0) &= \tau^-(\forall_{x^\rho} A_0) = \rho, & \tau^-(\forall_{x^\rho} A_0 \to B_0) &= \varepsilon,\\ \tau^+(\exists_{x^\rho}(A_0 \to B_0)) &= \rho, & \tau^-(\exists_{x^\rho}(A_0 \to B_0)) &= \varepsilon. \end{aligned}$$

Then

$$\begin{aligned} |\forall_{x^\rho} A_0 \to B_0|^x_\varepsilon &= |\forall_{x^\rho} A_0|^\varepsilon_x \to |B_0|^\varepsilon_\varepsilon = A_0 \to B_0,\\ |\exists_{x^\rho}(A_0 \to B_0)|^x_\varepsilon &= A_0 \to B_0. \end{aligned}$$

(b) For A with $\tau^+(A) = \varepsilon$ and $z \notin \mathrm{FV}(A)$, and arbitrary B

$$\begin{aligned} \tau^+(A \to \exists_{z^\rho} B) &= (\rho \wedge \tau^+(B)) \wedge (\tau^+(B) \to \tau^-(A)),\\ \tau^+(\exists_{z^\rho}(A \to B)) &= \rho \wedge (\tau^+(B) \wedge (\tau^+(B) \to \tau^-(A))),\\ \\ \tau^-(A \to \exists_{z^\rho} B) &= \tau^-(B),\\ \tau^-(\exists_{z^\rho}(A \to B)) &= \tau^-(B). \end{aligned}$$

Then

$$\begin{aligned} |A \to \exists_{z^\rho} B|^{\langle z,y\rangle,g}_v &= |A|^\varepsilon_{gv} \to |\exists_{z^\rho} B|^{z,y}_v = |A|^\varepsilon_{gv} \to |B|^y_v,\\ |\exists_{z^\rho}(A \to B)|^{z,\langle y,g\rangle}_v &= |A \to B|^{y,g}_v = |A|^\varepsilon_{gv} \to |B|^y_v. \end{aligned}$$

(c) For arbitrary A, B

$$\begin{aligned} \tau^+(\forall_{x^\rho}\exists_{y^\sigma} A(x,y)) &= (\rho \to \sigma \wedge \tau^+(A)),\\ \tau^+(\exists_{f^{\rho\to\sigma}}\forall_{x^\rho} A(x, fx)) &= (\rho \to \sigma) \wedge (\rho \to \tau^+(A)),\\ \\ \tau^-(\forall_{x^\rho}\exists_{y^\sigma} A(x,y)) &= \rho \wedge \tau^-(A),\\ \tau^-(\exists_{f^{\rho\to\sigma}}\forall_{x^\rho} A(x, fx)) &= \rho \wedge \tau^-(A). \end{aligned}$$

Then

$$|\forall_{x^\rho}\exists_{y^\sigma} A(x,y)|^{\lambda_x\langle fx,z\rangle}_{x,u} = |\exists_{y^\sigma} A(x,y)|^{fx,z}_u = |A(x,fx)|^z_u,$$
$$|\exists_{f^{\rho\to\sigma}}\forall_{x^\rho} A(x,fx)|^{f,\lambda_x z}_{x,u} = |\forall_{x^\rho} A(x,fx)|^{\lambda_x z}_{x,u} = |A(x,fx)|^z_u.$$

(d) For arbitrary A, writing $\tau^\pm A$ for $\tau^\pm(A)$

$$\tau^+(\forall_{z^\rho}(A\to\exists_{z^\rho}A)) = \rho\to(\tau^+A\to\rho\wedge\tau^+A)\wedge(\tau^+A\to\tau^-A\to\tau^-A),$$
$$\tau^-(\forall_{z^\rho}(A\to\exists_{z^\rho}A)) = \rho\wedge(\tau^+A\wedge\tau^-A).$$

Then

$$\begin{aligned}|\forall_{z^\rho}(A\to\exists_{z^\rho}A)|^{\lambda_z\langle\lambda_x\langle z,x\rangle,\lambda_{x,w}\,w\rangle}_{z,\langle x,w\rangle} &= |A\to\exists_{z^\rho}A|^{\lambda_x\langle z,x\rangle,\lambda_{x,w}\,w}_{x,w}\\ &= |A|^x_w\to|\exists_{z^\rho}A|^{z,x}_w\\ &= |A|^x_w\to|A|^x_w.\end{aligned}$$

2.3. **Characterization.** We consider the question when the Gödel translation of a formula A is equivalent to the formula itself.

Theorem (Characterization)**.**

$$\mathrm{AC}+\mathrm{IP}_\forall+\mathrm{MP}\vdash A\leftrightarrow\exists_x\forall_y|A|^x_y.$$

Proof. Induction on A; we only treat one case.

$$\begin{aligned}(A\to B)&\leftrightarrow(\exists_x\forall_y|A|^x_y\to\exists_v\forall_u|B|^v_u) && \text{by IH}\\ &\leftrightarrow\forall_x(\forall_y|A|^x_y\to\exists_v\forall_u|B|^v_u)\\ &\leftrightarrow\forall_x\exists_v(\forall_y|A|^x_y\to\forall_u|B|^v_u) && \text{by } (\mathrm{IP}_\forall)\\ &\leftrightarrow\forall_x\exists_v\forall_u(\forall_y|A|^x_y\to|B|^v_u)\\ &\leftrightarrow\forall_x\exists_v\forall_u\exists_y(|A|^x_y\to|B|^v_u) && \text{by (MP)}\\ &\leftrightarrow\exists_f\forall_x\forall_u\exists_y(|A|^x_y\to|B|^{fx}_u) && \text{by (AC)}\\ &\leftrightarrow\exists_{f,g}\forall_{x,u}(|A|^x_{gxu}\to|B|^{fx}_u) && \text{by (AC)}\\ &\leftrightarrow\exists_{f,g}\forall_{x,u}|A\to B|^{f,g}_{x,u}\end{aligned}$$

where the last step is by definition. □

Without the Markov principle one can still prove some relations between A and its Gödel translation. This, however, requires conditions $G^+(A)$, $G^-(A)$ on A, defined inductively by

$$\begin{aligned}G^\pm(P(\vec s)) &:= \top,\\ G^+(A\to B) &:= (\tau^-(A)=\varepsilon)\wedge G^-(A)\wedge G^+(B),\\ G^-(A\to B) &:= G^+(A)\wedge G^-(B),\\ G^\pm(A\wedge B) &:= G^\pm(A)\wedge G^\pm(B),\\ G^\pm(\forall_x A) &:= G^\pm(A),\quad G^\pm(\exists_x A) := G^\pm(A).\end{aligned}$$

Proposition.

(4) $\mathrm{AC}\vdash\exists_x\forall_y|A|^x_y\to A$ *if* $G^-(A)$,

(5) $\mathrm{AC}\vdash A\to\exists_x\forall_y|A|^x_y$ *if* $G^+(A)$.

Proof. Both directions are proved simultaneously, by induction on A. □

2.4. **Soundness.** We prove soundness of the Dialectica interpretation, for our natural deduction formulation of the underlying logic.

We first treat some axioms, and show that each of them has a "logical Dialectica realizer", that is, a term t such that $\forall_y |A|^t_y$ can be proved logically.

For $(\exists^+)$ this was proved in Example (d) of 2.2. Conjunction introduction $(\wedge^+)$ and elimination $(\wedge^-)$ have obvious Dialectica realizers.

The axioms $(\exists^-)$, (MP), $(\mathrm{IP}_\forall)$ and (AC) all have the form $C \to D$ where $\tau^+(C) \sim \tau^+(D)$ and $\tau^-(C) \sim \tau^-(D)$, with $\rho \sim \sigma$ indicating that ρ and σ are canonically isomorphic. This has been verfied

- for the existence elimination axiom – written in the equivalent form $\forall_{z^\rho}(A \to B) \to \exists_{z^\rho} A \to B$ – in Example (b) of 2.1;
- for (MP), $(\mathrm{IP}_\forall)$ and (AC) in Examples (a)-(c) of 2.2, respectively.

Such canonical isomorphisms can be expressed by λ-terms

$$f^+ \colon \tau^+(C) \to \tau^+(D), \qquad f^- \colon \tau^-(C) \to \tau^-(D),$$
$$g^+ \colon \tau^+(D) \to \tau^+(C), \qquad g^- \colon \tau^-(D) \to \tau^-(C).$$

(they have been written explicitely in Examples (a)-(c) of 2.2). It is easy to check that the Gödel translations $|C|^u_{g^- v}$ and $|D|^{f^+ u}_v$ are equal (modulo β-conversion). But then $\langle f^+, \lambda_u\, g^- \rangle$ is a Dialectica realizer for the axiom $C \to D$, because

$$|C \to D|^{f^+, \lambda_u\, g^-}_{u,v} = |C|^u_{g^- v} \to |D|^{f^+ u}_v.$$

Theorem (Soundness)**.** *Let M be a derivation*

$$\text{WE-HA}^\omega + \text{AC} + \text{IP}_\forall + \text{MP} + \text{Ax}_\forall \vdash A$$

from assumptions $u_i \colon C_i$ $(i = 1, \dots, n)$. Let x_i of type $\tau^+(C_i)$ be variables for realizers of the assumptions, and y be a variable of type $\tau^-(A)$ for a challenge of the goal. Then we can find terms $[\![M]\!]^+ =: t$ of type $\tau^+(A)$ with $y \notin \mathrm{FV}(t)$ and $[\![M]\!]^-_i =: r_i$ of type $\tau^-(C_i)$, and a derivation $\mu(M)$

$$\text{WE-HA}^\omega + \text{Ax}_\forall \vdash |A|^t_y$$

from assumptions $\bar{u}_i \colon |C_i|^{x_i}_{r_i}$.

Proof. Induction on M. We begin with the logical rules and leave treatment of the axioms for the end. The axioms $(\wedge^\pm)$, $(\exists^\pm)$, (MP), $(\mathrm{IP}_\forall)$ and (AC) have just been dealt with, so we will only need to consider induction, $\mathrm{Ax}_\forall$ and the weak extensionality rule.

Case $u \colon A$. Let x of type $\tau^+(A)$ be a variable for a realizer of the assumption u. Define $[\![u]\!]^+ := x$ and $[\![u]\!]^- := y$.

Case $\lambda_{u^A} M^B$. By IH we have a derivation of $|B|^t_z$ from $\bar{u} \colon |A|^x_r$ and $\bar{u}_i \colon |C_i|^{x_i}_{r_i}$, where $\bar{u} \colon |A|^x_r$ may be absent. Substitute $y0$ for x and $y1$ for z. By $(\to^+)$ we obtain $|A|^{y0}_{r[x,z:=y0,y1]} \to |B|^{t[x:=y0]}_{y1}$, which is (up to β-conversion)

$$|A \to B|^{\lambda_x t, \lambda_{x,z} r}_y, \quad \text{from} \quad \bar{u}'_i \colon |C_i|^{x_i}_{r_i[x,z:=y0,y1]}.$$

Here r is the canonical inhabitant of the type $\tau^-(A)$ in case $\bar{u} \colon |A|^x_r$ is absent. Hence we can define the required terms by (assuming that u^A is u_1)

$$[\![\lambda_u M]\!]^+ := (\lambda_x [\![M]\!]^+, \lambda_{x,z} [\![M]\!]^-_1),$$

$$[\![\lambda_u M]\!]^-_i := [\![M]\!]^-_{i+1}[x, z := y0, y1].$$

Case $M^{A\to B}N^A$. By IH we have a derivation of

$$|A \to B|^t_x = |A|^{x0}_{t1(x0)(x1)} \to |B|^{t0(x0)}_{x1} \quad \text{from } |C_i|^{x_i}_{p_i},\ |C_k|^{x_k}_{p_k}, \text{ and of}$$
$$|A|^s_z \quad \text{from } |C_j|^{x_j}_{q_j},\ |C_k|^{x_k}_{q_k}.$$

Substituting $\langle s, y\rangle$ for x in the first derivation and of $t1sy$ for z in the second derivation gives

$$|A|^s_{t1sy} \to |B|^{t0s}_y \quad \text{from } |C_i|^{x_i}_{p'_i},\ |C_k|^{x_k}_{p'_k}, \text{ and}$$
$$|A|^s_{t1sy} \quad \text{from } |C_j|^{x_j}_{q'_j},\ |C_k|^{x_k}_{q'_k}.$$

Now we contract $|C_k|^{x_k}_{p'_k}$ and $|C_k|^{x_k}_{q'_k}$: since $|C_k|^{x_k}_w$ is quantifier-free, there is a boolean-valued term r_{C_k} such that

$$|C_k|^{x_k}_w \leftrightarrow r_{C_k} w = \mathtt{tt}. \tag{6}$$

Hence with $r_k := [\mathbf{if}\ r_{C_k}p'_k\ \mathbf{then}\ q'_k\ \mathbf{else}\ p'_k]$ we can derive both $|C_k|^{x_k}_{p'_k}$ and $|C_k|^{x_k}_{q'_k}$ from $|C_k|^{x_k}_{r_k}$. The derivation proceeds by cases on the boolean term $r_{C_k}p'_k$. If it is true, then r_k converts into q'_k, and we only need to derive $|C_k|^{x_k}_{p'_k}$. But this follows by substituting p'_k for w in (6). If $r_{C_k}p'_k$ is false, then r_k converts into p'_k, and we only need to derive $|C_k|^{x_k}_{q'_k}$, from $|C_k|^{x_k}_{p'_k}$. But the latter implies $\mathtt{ff} = \mathtt{tt}$ (substitute again p'_k for w in (6)) and therefore every quantifier-free formula, in particular $|C_k|^{x_k}_{q'_k}$.

Using $(\to^-)$ we obtain

$$|B|^{t0s}_y \quad \text{from } |C_i|^{x_i}_{p'_i},\ |C_j|^{x_j}_{q'_j},\ |C_k|^{x_k}_{r_k}.$$

Let $[\![MN]\!]^+ := t0s$ and $[\![MN]\!]^-_i := p'_i$, $[\![MN]\!]^-_j := q'_j$, $[\![MN]\!]^-_k := r_k$.

Case $\lambda_x M^{A(x)}$. By IH we have a derivation of $|A(x)|^t_z$ from $\bar{u}_i\colon |C_i|^{x_i}_{r_i}$. Substitute $y0$ for x and $y1$ for z. We obtain $|A(y0)|^{t[x:=y0]}_{y1}$, which is (up to β-conversion)

$$|\forall_x A(x)|^{\lambda_x t}_y, \quad \text{from} \quad \bar{u}'_i\colon |C_i|^{x_i}_{r_i[x,z:=y0,y1]}.$$

Hence we can define the required terms by

$$[\![\lambda_x M]\!]^+ := \lambda_x [\![M]\!]^+,$$
$$[\![\lambda_x M]\!]^-_i := [\![M]\!]^-_i[x, z := y0, y1].$$

Case $M^{\forall_x A(x)}s$. By IH we have a derivation of $|\forall_x A(x)|^t_z = |A(z0)|^{t(z0)}_{z1}$ from $|C_i|^{x_i}_{r_i}$. Substituting $\langle s, y\rangle$ for z gives

$$|A(s)|^{ts}_y \quad \text{from } |C_i|^{x_i}_{r_i[z:=\langle s,y\rangle]}.$$

Let $[\![Ms]\!]^+ := ts$ and $[\![Ms]\!]^-_i := r_i[z := \langle s, y\rangle]$.

We now come to induction, $\mathrm{Ax}_\forall$ and the weak extensionality rule. For induction, consider for instance the algebra of natural numbers, given by constructors 0 and S. The induction schema then reads

$$\forall_n\big(A(0) \to \forall_m(A(m) \to A(m+1)) \to A(n)\big). \tag{7}$$

Let $B(n) := A(0) \to \forall_m(A(m) \to A(m+1)) \to A(n)$. Clearly we can derive $B(0)$ and $B(n) \to B(n+1)$. By those parts of the proof of the Soundness Theorem that we have dealt with already, we obtain realizing terms s and t, r and derivations of $|B(0)|^s_y$ and of $|B(n) \to B(n+1)|^{t,r}_{x,u}$, hence of

$$\begin{aligned}
&|B(n)|^x_{rxu} \to |B(n+1)|^{tx}_u \\
&\forall_y |B(n)|^x_y \to |B(n+1)|^{tx}_u \\
&\forall_y |B(n)|^x_y \to \forall_y |B(n+1)|^{tx}_y.
\end{aligned}$$

So if we define $g(0) := s$ and $g(n+1) := t(g(n))$, then we have proved by induction that $\forall_y |B(n)|^{g(n)}_y$, hence that $\exists_g \forall_y |\forall_n B(n)|^g_y$.

Now consider a purely universal formula $B = \forall_x A_0$, with A_0 quantifier-free. Then $\tau^+(B) = \varepsilon$, and moreover $|B|^\varepsilon_y = A_0$. Hence such axioms are interpreted by themselves. The weak extensionality rule can be dealt with in the same way. □

2.5. **Practical aspects of constructing Dialectica realizers.** In the proof of the Soundness Theorem above, at two points we have made (implicit) use of Dialectica realizers for logically derivable formulas:

- In the treatment of $\exists^-$, the equivalence of $\exists_{z^\rho} A \to \forall_{z^\rho}(A \to B) \to B$ with $\forall_{z^\rho}(A \to B) \to \exists_{z^\rho} A \to B$, and
- for induction, that we can derive $B(0)$ and $B(n) \to B(n+1)$, for $B(n) := A(0) \to \forall_m(A(m) \to A(m+1)) \to A(n)$.

Although these logical derivations are very easy, the fact that the formulas involved contain nested implications makes their Dialectica realizers complex. This shows up drastically in an implementation of the Dialectica interpretation. Two such implementations are presently available, both in the proof assistant and program extraction system Minlog[1]: one by Hernest (2006), and another one by the author, following the present paper.

Much more perspicious Dialectica realizers are obtained if one replaces the existence elimination and induction *axioms* by their equivalent *rule* formulations. Technically in our natural deduction setting with derivation terms this means that the derivation constants $\exists^-$ and Ind appear with sufficiently many arguments. Clearly this can always be assumed (use η-expansion). Then Dialectica realizers are constructed as follows.

Case $\mathrm{Ind}_{n,A} m M_0^{A(0)} M_1^{\forall_n(A(n)\to A(n+1))}$ By IH we have derivations of

$$\begin{aligned}
&|\forall_n(A(n) \to A(n+1))|^t_{n,f,y} = \\
&|A(n) \to A(n+1)|^{tn}_{f,y} = \\
&|A(n)|^f_{tn1fy} \to |A(n+1)|^{tn0f}_y \quad \text{from } |C_i|^{x_i}_{r_{i1}(n,f,y)}
\end{aligned}$$

and of

$$|A(0)|^{t_0}_{x_0} \quad \text{from } |C_i|^{x_i}_{r_{i0}(x_0)}.$$

[1] See `http://www.minlog-system.de`

i ranges over all assumption variables in $\mathrm{Ind}_{n,A}mM_0M_1$ (if necessary choose canonical terms r_{i0} and r_{i1}). It suffices to construct terms (involving recursion operators) $\tilde{t}$, $\tilde{r}_i$ with free variables among $\vec{x}$ such that

$$\forall_{m,y}\big((|C_i|^{x_i}_{\tilde{r}_i my})_i \to |A(m)|^{\tilde{t}m}_y\big). \tag{8}$$

For then we can define $[\![\mathrm{Ind}_{n,A}mM_0M_1]\!]^+ := \tilde{t}m$ and $[\![\mathrm{Ind}_{n,A}mM_0M_1]\!]^-_i := \tilde{r}_i my$. The recursion equations for $\tilde{t}$ are

$$\tilde{t}0 = t_0, \quad \tilde{t}(n+1) = tn0(\tilde{t}n)$$

and for $\tilde{r}_i$

$$\tilde{r}_i 0y = r_{i0}, \quad \tilde{r}_i(n+1)y = \begin{cases} r_{i1}(n,\tilde{t}n,y) =: s & \text{if } \neg|C_i|^{x_i}_s, \\ \tilde{r}_i n(tn1(\tilde{t}n)y) & \text{otherwise.} \end{cases}$$

$\tilde{t}$, $\tilde{r}_i$ can be written explicitely with recursion operators:

$$\begin{aligned} \tilde{t}m &= \mathcal{R}mt_0(\lambda_n(tn0)), \\ \tilde{r}_i m &= \mathcal{R}m(\lambda_y r_{i0})\big(\lambda_{n,p,y}[\mathbf{if}\ r_{C_i}s\ \mathbf{then}\ p(tn1(\tilde{t}n)y)\ \mathbf{else}\ s]\big) \end{aligned}$$

with s as above. It remains to prove (8). We only consider the successor case. Assume $|C_i|^{x_i}_{\tilde{r}_i(n+1)y}$ for all i. We must show $|A(n+1)|^{\tilde{t}(n+1)}_y$. If $\neg|C_i|^{x_i}_s$ for some i, then by definition $\tilde{r}_i(n+1)y = s$ and we have $|C_i|^{x_i}_s$, a contradiction. Hence $|C_i|^{x_i}_s$ for all i, and therefore $\tilde{r}_i(n+1)y = \tilde{r}_i n(tn1(\tilde{t}n)y)$. The IH (8) with $y := tn1(\tilde{t}n)y$ gives $|A(n)|^{\tilde{t}n}_{tn1(\tilde{t}n)y}$. Recall that the global IH (for the step derivation) gives with $f := \tilde{t}n$

$$(|C_i|^{x_i}_s)_i \to |A(n)|^{\tilde{t}n}_{tn1(\tilde{t}n)y} \to |A(n+1)|^{tn0(\tilde{t}n)}_y$$

and we are done.

Case $\exists^-_{x,A,B}M^{\exists_x A}N^{\forall_x(A\to B)}$. We proceed similar to the treatment of $(\to^-)$ above. By IH we have a derivation of

$$\begin{aligned} |\forall_x(A(x)\to B)|^t_x &= |A(x0)\to B|^{t(x0)}_{x1} \\ &= |A(x0)|^{x10}_{t(x0)1(x10)(x11)} \to |B|^{t(x0)0(x10)}_{x11} \end{aligned}$$

from $|C_i|^{x_i}_{p_i}$, $|C_k|^{x_k}_{p_k}$, and of

$$|\exists_x A(x)|^s_z = |A(s0)|^{s1}_z \quad \text{from } |C_j|^{x_j}_{q_j},\ |C_k|^{x_k}_{q_k}.$$

Substituting $\langle s0,\langle s1,y\rangle\rangle$ for x in the first derivation and of $t(s0)1(s1)y$ for z in the second derivation gives

$$\begin{aligned} &|A(s0)|^{s1}_{t(s0)1(s1)y} \to |B|^{t(s0)0(s1)}_y && \text{from } |C_i|^{x_i}_{p'_i},\ |C_k|^{x_k}_{p'_k}, \text{ and} \\ &|A(s0)|^{s1}_{t(s0)1(s1)y} && \text{from } |C_j|^{x_j}_{q'_j},\ |C_k|^{x_k}_{q'_k}. \end{aligned}$$

Now we contract $|C_k|^{x_k}_{p'_k}$ and $|C_k|^{x_k}_{q'_k}$ as in case $(\to^-)$ above; with $r_k := [\mathbf{if}\ r_{C_k}p'_k\ \mathbf{then}\ q'_k\ \mathbf{else}\ p'_k]$ we can derive both $|C_k|^{x_k}_{p'_k}$ and $|C_k|^{x_k}_{q'_k}$ from $|C_k|^{x_k}_{r_k}$. Using $(\to^-)$ we obtain

$$|B|^{t(s0)0(s1)}_y \quad \text{from } |C_i|^{x_i}_{p'_i},\ |C_j|^{x_j}_{q'_j},\ |C_k|^{x_k}_{r_k}.$$

So $[\![\exists^- MN]\!]^+ := t(s0)0(s1)$ and

$$[\![\exists^- MN]\!]^-_i := p'_i, \quad [\![\exists^- MN]\!]^-_j := q'_j, \quad [\![\exists^- MN]\!]^-_k := r_k.$$

2.6. **A unified treatment of modified realizability and the Dialectica interpretation.** Following Oliva (2006), we show that modified realizability can be treated in such a way that similarities with the Dialectica interpretation become visible. To this end, one needs to change the definitions of $\tau^+(A)$ and $\tau^-(A)$ and also of the Gödel translation $|A|^x_y$ in the implicational case, as follows.

$$\begin{aligned} &\tau^+_{\mathrm{mr}}(A \to B) := \tau^+_{\mathrm{mr}}(A) \to \tau^+_{\mathrm{mr}}(B), \\ &\tau^-_{\mathrm{mr}}(A \to B) := \tau^+_{\mathrm{mr}}(A) \wedge \tau^-_{\mathrm{mr}}(B), \end{aligned} \qquad \|A \to B\|^f_{x,u} := \forall_y \|A\|^x_y \to \|B\|^{fx}_u.$$

Notice that the (changed) Gödel translation $\|A\|^x_y$ is not quantifier-free any more, but only $\exists$-free. – Then the standard definition of modified realizability mr (cf. Troelstra (1973)) can be expressed in terms of the (new) $\|A\|^x_y$:

$$\vdash r \ \mathrm{mr}\ A \leftrightarrow \forall_y \|A\|^r_y.$$

This is proved by induction on A. For prime formulas the claim is obvious. *Case* $A \to B$, with $\tau^+_{\mathrm{mr}}(A) \neq \varepsilon$, $\tau^-_{\mathrm{mr}}(A) \neq \varepsilon$.

$$\begin{aligned} r \ \mathrm{mr}\ (A \to B) &\leftrightarrow \forall_x(x \ \mathrm{mr}\ A \to rx \ \mathrm{mr}\ B) && \text{by definition} \\ &\leftrightarrow \forall_x(\forall_y \|A\|^x_y \to \forall_u \|B\|^{rx}_u) && \text{by IH} \\ &\leftrightarrow \forall_{x,u}(\forall_y \|A\|^x_y \to \|B\|^{rx}_u) \\ &= \forall_{x,u} \|A \to B\|^r_{x,u} && \text{by definition.} \end{aligned}$$

The other cases are similar (even easier).

2.7. **Extraction.** As a consequence of the Soundness and Characterization Theorems we obtain

Theorem (Extraction)**.** *Assume*

$$\mathrm{WE\text{-}HA}^\omega + \mathrm{AC} + \mathrm{IP}_\forall + \mathrm{MP} + \mathrm{Ax}_\forall \vdash \forall_x \exists_y A(x,y)$$

with A arbitrary. Then we can find a closed HA^ω*-term t such that*

$$\mathrm{WE\text{-}HA}^\omega + \mathrm{AC} + \mathrm{IP}_\forall + \mathrm{MP} + \mathrm{Ax}_\forall \vdash \forall_x A(x, tx).$$

Moreover, in case the condition $G^-(A)$ is satisfied we even have

$$\mathrm{WE\text{-}HA}^\omega + \mathrm{AC} + \mathrm{Ax}_\forall \vdash \forall_x A(x, tx).$$

Proof. Recall that

$$|\forall_x \exists_y A(x,y)|^{\lambda_x \langle fx, gx\rangle}_{x,b} = |\exists_y A(x,y)|^{fx,gx}_b = |A(x,fx)|^{gx}_b.$$

By the Soundness Theorem we obtain closed terms t, s such that

$$\mathrm{WE\text{-}HA}^\omega + \mathrm{Ax}_\forall \vdash \forall_{x,b} |A(x,tx)|^{sx}_b$$

and hence

$$\mathrm{WE\text{-}HA}^\omega + \mathrm{Ax}_\forall \vdash \forall_x \exists_a \forall_b |A(x,tx)|^a_b.$$

By the Characterization Theorem we have

$$\mathrm{AC} + \mathrm{IP}_\forall + \mathrm{MP} \vdash \exists_a \forall_b \, |A(x,tx)|^a_b \to A(x,tx).$$

By (4), ($\mathrm{IP}_\forall$) and (MP) are not needed here provided the condition $G^-(A)$ is satisfied. Therefore the claim follows. □

Theorem (Extraction from classical proofs)**.** *Assume*

$$\text{WE-HA}^\omega + \text{AC} + \text{IP}_\forall + \text{MP} + \text{Ax}_\forall \vdash \forall_x \tilde{\exists}_y A_0(x,y),$$

$A_0(x,y)$ a quantifier-free formula with at most the displayed variables free. Then we can find a closed HA^ω*-term t such that*

$$\text{WE-HA}^\omega + \text{Ax}_\forall \vdash \forall_x A_0(x,tx).$$

Proof. This follows from the Soundness Theorem in 2.4 and

$$|\forall_x \tilde{\exists}_y A_0(x,y)|^t_x = |\tilde{\exists}_y A_0(x,y)|^{tx}_\varepsilon = \neg\neg A_0(x,tx). \qquad \square$$

3. Gödel's Dialectica Interpretation With Majorants

Generally, the Dialectica interpretation has a strong tendency to produce complex extracted terms, as opposed to the realizability interpretation. This is partially due to contraction (necessary in the $\to^-$-rule). Therefore it is advisable (even more so than for the realizability interpretation) to

- consider derivations from lemmata (whose proofs are not analyzed), and
- try to simplify extracted terms by only aiming at majorants.

This has led Kohlenbach (1992, 1996) to develop his "monotone Dialectica interpretation", where one only looks for bounds of realizers rather than exact realizers.

An essential point observed by Kohlenbach (1996) is that when one restricts attention to bounds rather than exact realizers, then one can conveniently deal with additional assumptions $\text{Ax}_{\forall\exists_\le\forall}$ of the form

$$\forall_{x^\rho}\exists_{y\le_\sigma rx}\forall_{z^\tau} A_0(x,y,z) \qquad (A_0 \text{ quantifier-free}),$$

with r a closed term of type $\rho\to\sigma$. We then need to consider strenghtened versions $\text{Ax}'_{\forall\exists_\le\forall}$ of these assumptions as well:

$$\exists_{Y\le_{\rho\to\sigma} r}\forall_{x^\rho,z^\tau} A_0(x,Yx,z).$$

Note that with (AC) one can prove the strenghtened version from the original one.

3.1. **Majorization.** We define *pointwise majorization* $\ge_\rho$, by induction on the type. $x\ge_\mu y$ for μ s finitary base type is already defined, and

$$\begin{aligned}(x \ge_{\rho\to\sigma} y) &:= \forall_z(xz \ge_\sigma yz),\\ (x \ge_{\rho\wedge\sigma} y) &:= (x0 \ge_\rho y0)\wedge(x1\ge_\rho y1),\end{aligned}$$

For simplicity we treat the majorization relation of Howard (1973) just for types built from the base type $\mathbf{N}$ by $\rho\to\sigma$. We extend $\ge_{\mathbf{N}}$ to higher types, in a *pointwise* fashion (as we did for $=_\mu$ in 1.6 above)

$$(x_1 \ge_{\rho\to\sigma} x_2) := \forall_y(x_1y \ge_\sigma x_2y).$$

Following Howard (1973), we define a relation $x^*\ \text{maj}_\rho\ x$ (x^* *hereditarily majorizes* x) for $x^*, x\in G_\rho$, by induction on the type ρ:

$$\begin{aligned}(x^*\ \text{maj}_\mu\ x) &:= (x^*\ge_\mu x),\\ (x^*\ \text{maj}_{\rho\to\sigma}\ x) &:= \forall_{y^*,y}(y^*\ \text{maj}_\rho\ y \to x^*y^*\ \text{maj}_\sigma\ xy).\end{aligned}$$

Lemma.

(a) $\vdash x^* =_\rho \tilde{x}^* \to x =_\rho \tilde{x} \to x^* \operatorname{maj}_\rho x \to \tilde{x}^* \operatorname{maj}_\rho \tilde{x}$.
(b) $\vdash x^* \operatorname{maj}_\rho x \to x \geq_\rho \tilde{x} \to x^* \operatorname{maj}_\rho \tilde{x}$.

Proof. Induction on ρ. We argue informally, and only treat (b). *Case* $\rho \to \sigma$. Assume $y^* \operatorname{maj}_\rho y$. Then $x^*y^* \operatorname{maj}_\sigma xy$ and $xy \geq_\sigma \tilde{x}y$, hence by IH $x^*y^* \operatorname{maj}_\sigma \tilde{x}y$. □

3.2. **Majorization of closed HA^ω-terms.** Let 1 denote the type $\mathbf{N} \to \mathbf{N}$. Clearly, for every monotone function D of type 1 we have $D \operatorname{maj} D$. Moreover, $\mathcal{R}^\tau_\mu$ is hereditarily majorizable:

Lemma (Majorization)**.** (a) *Define* $M\colon (\mu \to \tau) \to \mu \to \tau$ *with* $\tau = \vec{\rho} \to \mu'$ *by*

$$M f n \vec{x} := \max_{i \leq n} f i \vec{x}.$$

Then $\mathrm{HA}^\omega \vdash \forall_n \bar{f} n \operatorname{maj} f n \to M\bar{f} \operatorname{maj} f$.
(b) $\mathrm{HA}^\omega \vdash f^*, g^* \operatorname{maj} f, g \to \mathcal{R}_\mu f^* g^* n \operatorname{maj} \mathcal{R}_\mu f g n$.
(c) *Define* $\mathcal{R}^*_\mu f g := M(\mathcal{R}_\mu f g)$. *Then* $\mathrm{HA}^\omega \vdash \mathcal{R}^*_\mu \operatorname{maj} \mathcal{R}_\mu$.

Proof. We argue informally.
(a) Let $n^* \geq n$ and $\vec{x}^* \operatorname{maj} \vec{x}$; we must show $M\bar{f}n^*\vec{x}^* \geq f n \vec{x}$.

$$M\bar{f}n^*\vec{x}^* = \max_{i \leq n^*} \bar{f} i \vec{x}^* \geq \bar{f} n \vec{x}^* \geq f n \vec{x}.$$

(b) Induction on n; for simplicity we assume $\mu = \mathbf{N}$. For 0 the claim is obvious, and in the step we have by IH $\mathcal{R} f^* g^*(\mathrm{S}n) := g^* n (\mathcal{R} f^* g^* n) \operatorname{maj} g n (\mathcal{R} f g n) := \mathcal{R} f g (\mathrm{S}n)$, where $:=$ is definitional equality.
(c) Let $f^*, g^* \operatorname{maj} f, g$. We must show $M(\mathcal{R} f^* g^*) \operatorname{maj} \mathcal{R} f g$. By (a) it suffices to prove $\forall_n \mathcal{R} f^* g^* n \operatorname{maj} \mathcal{R} f g n$. But this holds by (b). □

The following theorem is due to Howard (1973).

Theorem. *Let* $r(\vec{x})$ *be a* HA^ω*-term with free variables among* $\vec{x}$*. Assume that* $\mathrm{HA}^\omega \vdash c^* \operatorname{maj} c$ *for all constants* c *in* r*. Let* r^* *be* r *with all constants* c *replaced by* c^**. Then* $\mathrm{HA}^\omega \vdash \vec{x}^* \operatorname{maj} \vec{x} \to r^*(\vec{x}^*) \operatorname{maj} r(\vec{x})$.

Proof. Induction on r. *Case* $\lambda_y r(y, \vec{x})$. We argue informally. Assume $\vec{x}^* \operatorname{maj} \vec{x}$. We must show $y^* \operatorname{maj} y \to (\lambda_y r^*(y, \vec{x}^*)) y^* \operatorname{maj} (\lambda_y r(y, \vec{x})) y$. So assume $y^* \operatorname{maj} y$. Then by IH $r^*(y^*, \vec{x}^*) \operatorname{maj} r(y, \vec{x})$, which is our claim. □

Hence every closed term r of HA^ω is hereditarily majorizable. In fact, we have constructed a closed term r^* of HA^ω such that $r^* \operatorname{maj} r$.

3.3. **Soundness with majorants.**

Theorem (Soundness with majorants)**.** *Let* M *be a derivation*

$$\mathrm{WE\text{-}HA}^\omega + \mathrm{AC} + \mathrm{IP}_- + \mathrm{MP} + \mathrm{Ax}_{\forall\exists_\leq\forall} \vdash A$$

from assumptions $u_i\colon C_i$ $(i = 1, \ldots, n)$*. Let* x_i *of type* $\tau^+(C_i)$ *be variables for realizers of the assumptions, and* y *of type* $\tau^-(A)$ *be a variable for a challenge of the goal. Let* $\vec{z}$ *of type* $\vec{\rho}$ *be the variables free in* M*. Then we can find closed terms* $[\![\lambda_{\vec{z},\vec{u}} M]\!]^{*+} =: T^*$ *of type* $\tau^+(C_1) \to \ldots \to \tau^+(C_n) \to \vec{\rho} \to \tau^+(A)$ *and* $[\![\lambda_{\vec{z},\vec{u}} M]\!]^{*-}_i =: R^*_i$ *of type* $\tau^+(C_1) \to \ldots \to \tau^+(C_n) \to \vec{\rho} \to \tau^-(A) \to \tau^-(C_i)$*, and a derivation* $\mu(M)$ *in*

$$\mathrm{WE\text{-}HA}^\omega + \mathrm{Ax}'_{\forall\exists_\leq\forall}$$

of the formula

$$\exists_{T,R_1,\dots,R_n}\big(T^*\ \mathrm{maj}\ T \land R_1^*\ \mathrm{maj}\ R_1 \land \cdots \land R_n^*\ \mathrm{maj}\ R_n \land \forall_{\vec{x},\vec{z},y}(|C_1|^{x_1}_{R_1\vec{x}\vec{z}y} \to \cdots \to |C_n|^{x_n}_{R_n\vec{x}\vec{z}y} \to |A|^{T\vec{x}\vec{z}}_y)\big).$$

Proof. Induction on M.

Case $u\colon A$. Let x of type $\tau^+(A)$ be a variable for a realizer of the assumption u. We need T^* and R^* such that

$$\exists_{T,R}\big(T^*\ \mathrm{maj}\ T \land R^*\ \mathrm{maj}\ R \land \forall_{x,y}(|A|^x_{Rxy} \to |A|^{Tx}_y)\big).$$

We can take $Tx := x$ and $Rxy := y$, which both majorize themselves.

Case $c\colon A$, c an axiom. Consider an axiom

$$\forall_{x^\rho}\exists_{y\le_\sigma rx}\forall_{z^\tau} A_0(x,y,z) \qquad (A_0 \text{ quantifier-free}),$$

with r a closed term of type $\rho\to\sigma$. We have to find a majorant of some T such that the following holds:

$$\begin{aligned}
&\forall_{x,z}|\forall_{x^\rho}\exists_{y\le_\sigma rx}\forall_{z^\tau} A_0(x,y,z)|^T_{x,z}\\
&\forall_{x,z}|\exists_{y\le_\sigma rx}\forall_{z^\tau} A_0(x,y,z)|^{Tx}_z\\
&\forall_{x,z}(Tx \le rx \land |\forall_{z^\tau} A_0(x,Tx,z)|_z)\\
&\forall_{x,z}(Tx \le rx \land A_0(x,Tx,z)).
\end{aligned}$$

We now use the corresponding axiom in $\mathrm{Ax}'_{\forall\exists_\le\forall}$:

$$\exists_{Y\le_{\rho\to\sigma} r}\forall_{x^\rho,z^\tau} A_0(x,Yx,z).$$

Pick this Y as the desired T. Then as a majorant for Y we can take a closed term r^* majorizing r.

For the other axioms we have already constructed a Dialectica realizer, and we can take an arbitrary majorant of it. However, we can also directly provide a majorant of some Dialectica realizer.

Case $\lambda_{u^A}M^B$. By IH we have a derivation of

$$\begin{aligned}
\exists_{T,R_1,\dots,R_n,R}\big(&T^*\ \mathrm{maj}\ T \land R_1^*\ \mathrm{maj}\ R_1 \land \cdots \land R_n^*\ \mathrm{maj}\ R_n \land R^*\ \mathrm{maj}\ R \land\\
&\forall_{x_1,\dots,x_n,x,z}(|C_1|^{x_1}_{R_1x_1\dots x_nxz} \to \cdots \to |C_n|^{x_n}_{R_nx_1\dots x_nxz} \to\\
&\quad |A|^x_{Rx_1\dots x_nxz} \to |B|^{Tx_1\dots x_nx}_z)\big).
\end{aligned}$$

We argue informally. Instantiating x with $y0$ and z with $y1$ gives

$$\begin{aligned}
&\forall_{x_1,\dots,x_n,y}(|C_1|^{x_1}_{R_1x_1\dots x_n(y0)(y1)} \to \cdots \to |C_n|^{x_n}_{R_nx_1\dots x_n(y0)(y1)} \to\\
&\quad |A|^{y0}_{Rx_1\dots x_n(y0)(y1)} \to |B|^{Tx_1\dots x_n(y0)}_z),
\end{aligned}$$

which is

$$\begin{aligned}
&\forall_{x_1,\dots,x_n,y}(|C_1|^{x_1}_{R_1x_1\dots x_n(y0)(y1)} \to \cdots \to |C_n|^{x_n}_{R_nx_1\dots x_n(y0)(y1)} \to\\
&\quad |A\to B|^{Tx_1\dots x_n,Rx_1\dots x_n}_y.
\end{aligned}$$

Therefore we can define the required $\tilde{T}^*, \tilde{R}_i^*$ by

$$\tilde{T}^*\vec{x} := \langle T^*\vec{x}, R^*\vec{x}\,\rangle, \quad \tilde{R}_i^*\vec{x}y := R_i^*\vec{x}(y0)(y1).$$

Case $M^{A\to B}N^A$. We argue informally. By IH we have

$|A\to B|^{T\vec{x}_i\vec{x}_k}_x = |A|^{x0}_{T\vec{x}_i\vec{x}_k1(x0)(x1)} \to |B|^{T\vec{x}_i\vec{x}_k0(x0)}_{x1}$ from $|C_i|^{x_i}_{P_i\vec{x}_i\vec{x}_kx}$, $|C_k|^{x_k}_{P_k\vec{x}_i\vec{x}_kx}$

$$|A|_z^{S\vec{x}_j\vec{x}_k} \quad \text{from } |C_j|^{x_j}_{Q_j\vec{x}_j\vec{x}_k z},\ |C_k|^{x_k}_{Q_k\vec{x}_j\vec{x}_k z}.$$

Instantiating x with $\langle S\vec{x}_j\vec{x}_k, y\rangle$ in the first and z with $T\vec{x}_i\vec{x}_k 1(S\vec{x}_j\vec{x}_k)y$ in the second derivation gives

$$\begin{aligned} &|A|^{S\vec{x}_j\vec{x}_k}_{T\vec{x}_i\vec{x}_k 1(S\vec{x}_j\vec{x}_k)y} \to |B|_y^{T\vec{x}_i\vec{x}_k 0(S\vec{x}_j\vec{x}_k)} && \text{from } |C_i|^{x_i}_{p_i'},\ |C_k|^{x_k}_{p_k'}, \text{ and}\\ &|A|^{S\vec{x}_j\vec{x}_k}_{T\vec{x}_i\vec{x}_k 1(S\vec{x}_j\vec{x}_k)y} && \text{from } |C_j|^{x_j}_{q_j'},\ |C_k|^{x_k}_{q_k'}, \end{aligned}$$

with

$$\begin{aligned} p_i' &:= P_i\vec{x}_i\vec{x}_k\langle S\vec{x}_j\vec{x}_k, y\rangle, & p_k' &:= P_k\vec{x}_i\vec{x}_k\langle S\vec{x}_j\vec{x}_k, y\rangle,\\ q_j' &:= Q_j\vec{x}_j\vec{x}_k(T\vec{x}_i\vec{x}_k 1(S\vec{x}_j\vec{x}_k)y), & q_k' &:= Q_k\vec{x}_j\vec{x}_k(T\vec{x}_i\vec{x}_k 1(S\vec{x}_j\vec{x}_k)y). \end{aligned}$$

Hence we can take

$$\begin{aligned} \tilde{T}^*\vec{x}_i\vec{x}_j\vec{x}_k &:= T^*\vec{x}_i\vec{x}_k 0(S^*\vec{x}_j\vec{x}_k),\\ R_i^*\vec{x}_i\vec{x}_j\vec{x}_k y &:= P_i^*\vec{x}_i\vec{x}_k\langle S^*\vec{x}_j\vec{x}_k, y\rangle,\\ R_j^*\vec{x}_i\vec{x}_j\vec{x}_k y &:= Q_j^*\vec{x}_j\vec{x}_k(T^*\vec{x}_i\vec{x}_k 1(S^*\vec{x}_j\vec{x}_k)y),\\ R_k^*\vec{x}_i\vec{x}_j\vec{x}_k y &:= \max(P_k^*\vec{x}_i\vec{x}_k\langle S^*\vec{x}_j\vec{x}_k, y\rangle, Q_k^*\vec{x}_j\vec{x}_k(T^*\vec{x}_i\vec{x}_k 1(S^*\vec{x}_j\vec{x}_k)y)). \end{aligned}$$

For the verifying derivation we again need to contract $|C_k|^{x_k}_{p_k'}$ and $|C_k|^{x_k}_{q_k'}$: since $|C_k|^{x_k}_w$ is quantifier-free, there is a boolean-valued term r_{C_k} such that

$$|C_k|^{x_k}_w \leftrightarrow r_{C_k}w = \mathtt{tt}.$$

Hence with $r_k := [\mathbf{if}\ r_{C_k}p_k'\ \mathbf{then}\ q_k'\ \mathbf{else}\ p_k']$ we can derive both $|C_k|^{x_k}_{p_k'}$ and $|C_k|^{x_k}_{q_k'}$ from $|C_k|^{x_k}_{r_k}$. Using $(\to^-)$ we obtain

$$|B|_y^{T\vec{x}_i\vec{x}_k 0(S\vec{x}_j\vec{x}_k)} \quad \text{from } |C_i|^{x_i}_{p_i'},\ |C_j|^{x_j}_{q_j'},\ |C_k|^{x_k}_{r_k}.$$

Case $\lambda_x M^{A(x)}$. By IH we have a derivation of $|A(x)|_z^{Tx_1\dots x_n x}$ from $|C_i|^{x_i}_{R_i x_1\dots x_n xz}$. Instantiating x with $y0$ and z with $y1$ gives $|A(y0)|_{y1}^{Tx_1\dots x_n(y0)}$, which is

$$|\forall_x A(x)|_y^{Tx_1\dots x_n}, \quad \text{from } |C_i|^{x_i}_{R_i x_1\dots x_n(y0)(y1)}.$$

Hence we can take

$$\begin{aligned} \tilde{T}^* x_1\dots x_n &:= T^* x_1\dots x_n,\\ \tilde{R}_i^* x_1\dots x_n y &:= R_i^* x_1\dots x_n(y0)(y1). \end{aligned}$$

Case $M^{\forall_x A(x)}s$. By IH we have a derivation of $|\forall_x A(x)|_z^{Tx_1\dots x_n}$, which is $|A(z0)|_{z1}^{Tx_1\dots x_n(z0)}$, from $|C_i|^{x_i}_{R_i x_1\dots x_n z}$. Instantiating z with $\langle s, y\rangle$ gives

$$|A(s)|_y^{Tx_1\dots x_n s} \quad \text{from } |C_i|^{x_i}_{R_i x_1\dots x_n\langle s,y\rangle}.$$

Assume for simplicity that s is closed. Then we can take

$$\begin{aligned} \tilde{T}^* x_1\dots x_n &:= T^* x_1\dots x_n s^*,\\ \tilde{R}_i^* x_1\dots x_n y &:= R_i^* x_1\dots x_n\langle s^*, y\rangle. \end{aligned}$$

□

3.4. **The weak Lemma of König as a $\forall\exists_{\leq}\forall$-Axiom.** We show that the "weak" (that is, binary) Lemma of König WKL can be brought into the form of an axiom in $\mathrm{Ax}_{\forall\exists_{\leq}\forall}$. This has been observed by Kohlenbach (1992). Here we give a somewhat simplified proof of this fact; it is based on ideas of Ishihara (2006).

WKL says that every infinite binary tree has an infinite path. When we try to directly formalize it in our (functional) language, it does not quite have the required form, since the assumption that the given tree is infinite needs an additional $\forall$ in the premise. However, one can easily find an equivalent statement of the required form. To this end, we define the "infinite extension" of a given tree, and let WKL$'$ say that for every t, the infinite extension $I(\hat{t})$ of its "associated tree" $\hat{t}$ has an infinite path. It then is easy to see that WKL and WKL$'$ are equivalent.

Let us first introduce some basic definitions. Let $\mathbf{N}$ be the type of unary and bin the type of binary natural numbers. It is convenient here to view binary numbers as lists of booleans $\mathtt{tt}$, $\mathtt{ff}$, and to write these lists in reverse order, that is, add elements at the end. We fix the types of some variables and state their intended meaning:

a, b, c	of type bin	for nodes,
r, s, t	of type bin $\to \mathbf{B}$	for decidable sets of nodes,
f, g, h	of type $\mathbf{N} \to \mathbf{B}$	for paths,
n, m, k, i, j	of type $\mathbf{N}$	for natural numbers,
p, q	of type $\mathbf{B}$	for booleans.

Let $\mathrm{lh}(a)$ be the *length* of a (viewed as list of booleans). Let $\bar{a}(n)$ denote the initial segment of a of length n, if $n \leq \mathrm{lh}(a)$, and a otherwise. Similarly let $\bar{f}(n)$ denote the initial segment of f of length n, that is, the list $(f(0), f(1), \dots, f(n-1))$. Let $(a)_n$ denote the n-th element of a, if $n < \mathrm{lh}(a)$, and $\mathtt{tt}$ otherwise. f is a *path in* t if all its initial segments $\bar{f}(n)$ are in t. Call t *infinite* if for every n there is a node of length n in t. Call t a *tree* if it is downwards closed, i.e., $\forall_a \forall_{n \leq \mathrm{lh}(a)} (a \in t \to \bar{a}(n) \in t)$. So WKL says that

$$\begin{aligned}
&\forall_t \big(\forall_a \forall_{n \leq \mathrm{lh}(a)} (a \in t \to \bar{a}(n) \in t) \to && (t \text{ is a tree}) \\
&\quad \forall_n \exists_{a \in t} \mathrm{lh}(a) = n \to && (t \text{ is infinite}) \\
&\quad \exists_f \forall_n \bar{f}(n) \in t \big) && (t \text{ has an infinite path}),
\end{aligned}$$

which – because of the two premises saying that t is an infinite tree – is not of the required logical form.

To obtain an equivalent formulation in the required form, we introduce some further notions.

$$\begin{aligned}
&\hat{t} := \{\, a \mid \forall_{n < \mathrm{lh}(a)} \bar{a}(n) \in t \,\} && \text{the associated tree } \hat{t} \text{ for } t, \\
&b = a * \mathtt{tt}^{\mathrm{lh}(b) - \mathrm{lh}(a)} && b \text{ is the } \mathtt{tt}\text{-extension of } a, \\
&\forall_{c; \mathrm{lh}(c) = \mathrm{lh}(b)} c \notin \hat{t} && b \text{ is } t\text{-big};
\end{aligned}$$

here $*$ denotes concatenation of lists. Let $\min_{\mathrm{lex}}$ denote the minimum of a set of nodes w.r.t. the lexicographical ordering, and $\mathrm{maxlen}_{<n}(t)$ be the maximal length of all nodes of t of length $< n$. Then $\mathrm{ll}_n(t)$ is the leftmost largest node in t of length $< n$:

$$\mathrm{maxlen}_{<n}(t) := \max\{\, \mathrm{lh}(a) \mid a \in t \wedge \mathrm{lh}(a) < n \,\},$$

$$\mathrm{ll}_n(t) \qquad := \min_{\mathrm{lex}}\{\, c \in t \mid \mathrm{lh}(c) = \mathrm{maxlen}_{<n}(t)\,\}.$$

We can now define the infinite extension $I(t)$ of a tree t:

$$I(t) := \{\, b \mid b \in t \vee (b \text{ is } t\text{-big} \wedge b \text{ is the } \mathtt{tt}\text{-extension of } \mathrm{ll}_{\mathrm{lh}(b)})\,\}.$$

All these notions are definable in HA^ω. They clearly have the following properties:

$\hat{t}$ is a tree;

if t is a tree, then $\hat{t} = t$;

if t is a tree, then $I(t)$ is an infinite tree extending t;

if t is an infinite tree, then $I(t) = t$.

Then WKL is equivalent (provably in HA^ω) to

$$\mathrm{WKL}' := \forall_t \exists_f \forall_n \bar{f}(n) \in I(\hat{t}).$$

To see this, assume WKL, and let t be arbitrary. Then $I(\hat{t})$ is an infinite tree extending t. By WKL applied to $I(\hat{t})$, $\exists_f \forall_n \bar{f}(n) \in I(\hat{t})$. Conversely, let t be an infinite tree. Then $I(\hat{t}) = t$ and therefore $\exists_f \forall_n \bar{f}(n) \in t$.

Remark. From the results of Ishihara (1990) it is known WKL implies Brouwer's fan theorem. Moreover, a direct proof of this implication has been given by Ishihara in 2002 (published in (2006)). In Berger and Ishihara (2005), it is shown that a weakened form WKL! of WKL, where as an additional hypothesis it is required that in an effective sense infinite paths are unique, is equivalent to Fan. One direction (WKL! implies Fan) is essentially the proof by Ishihara (2006), enhanced by the additional requirement that the tree extension to be constructed satisfies the effective uniqueness condition (as in Berger and Ishihara (2005)). The main tool of this proof is the construction of $I(\hat{t})$ described above. The other direction (Fan implies WKL!) is far less directly proved in Berger and Ishihara (2005), where the emphasis rather was to provide a fair number of equivalents to Fan, and to do the proof economically by giving a circle of implications. A direct proof of the equivalence of Fan with WKL! is in Schwichtenberg (2005). The latter paper also reports on a formalization in the Minlog proof assistant, and gives rather short and perspicious realizing terms (w.r.t. modified realizability) machine-extracted from each of the two directions of this proof.

References

J. Berger and H. Ishihara. Brouwer's fan theorem and unique existence in constructive analysis. *Mathematical Logic Quarterly*, 51(4):360–364, 2005.

U. Berger. Program extraction from normalization proofs. In M. Bezem and J. Groote, editors, *Typed Lambda Calculi and Applications*, volume 664 of *LNCS*, pages 91–106. Springer Verlag, Berlin, Heidelberg, New York, 1993.

U. Berger. Uniform Heyting Arithmetic. *Annals Pure Applied Logic*, 133: 125–148, 2005.

U. Berger, S. Berghofer, P. Letouzey, and H. Schwichtenberg. Program extraction from normalization proofs. *Studia Logica*, 82:27–51, 2006.

U. Berger, W. Buchholz, and H. Schwichtenberg. Refined program extraction from classical proofs. *Annals of Pure and Applied Logic*, 114:3–25, 2002.

U. Berger and H. Schwichtenberg. Program development by proof transformation. In H. Schwichtenberg, editor, *Proof and Computation*, volume 139 of *Series F: Computer and Systems Sciences*, pages 1–45. NATO Advanced Study Institute, International Summer School held in Marktoberdorf, Germany, July 20 – August 1, 1993, Springer Verlag, Berlin, Heidelberg, New York, 1995.

U. Berger, H. Schwichtenberg, and M. Seisenberger. The Warshall Algorithm and Dickson's Lemma: Two Examples of Realistic Program Extraction. *Journal of Automated Reasoning*, 26:205–221, 2001.

H. Friedman. Classically and intuitionistically provably recursive functions. In D. Scott and G. Müller, editors, *Higher Set Theory*, volume 669 of *Lecture Notes in Mathematics*, pages 21–28. Springer Verlag, Berlin, Heidelberg, New York, 1978.

K. Gödel. Über eine bisher noch nicht benützte Erweiterung des finiten Standpunkts. *Dialectica*, 12:280–287, 1958.

M.-D. Hernest. *Feasible programs from (non-constructive) proofs by the light (monotone) Dialectica interpretation*. PhD thesis, Ecole Polytechnique Paris and LMU München, 2006.

W. A. Howard. Hereditarily majorizable functionals of finite type. In A. Troelstra, editor, *Mathematical Investigation of Intuitionistic Arithmetic and Analysis*, volume 344 of *Lecture Notes in Mathematics*, pages 454–461. Springer Verlag, Berlin, Heidelberg, New York, 1973.

H. Ishihara. An omniscience principle, the König lemma and the Hahn-Banach theorem. *Zeitschr. f. math. Logik und Grundlagen d. Math.*, 36: 237–240, 1990.

H. Ishihara. Weak König lemma implies Brouwer's fan theorem: a direct proof. *Notre Dame J. Formal Logic*, 47:249–252, 2006.

K. F. Jørgensen. Finite type arithmetic. Master's thesis, University of Roskilde, 2001.

U. Kohlenbach. Effective bounds from ineffective proofs in analysis: an application of functional interpretation and majorization. *The Journal of Symbolic Logic*, 57(4):1239–1273, 1992.

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

P. Oliva. Unifying functional interpretations. *Notre Dame J. Formal Logic*, 47:262–290, 2006.

H. Schwichtenberg. Proofs as programs. In P. Aczel, H. Simmons, and S. Wainer, editors, *Proof Theory. A selection of papers from the Leeds Proof Theory Programme 1990*, pages 81–113. Cambridge University Press, 1993.

H. Schwichtenberg. A direct proof of the equivalence between Brouwer's fan theorem and König's lemma with a uniqueness hypothesis. *Journal of Universal Computer Science*, 11(12):2086–2095, 2005. `http://www.jucs.`

org/jucs_11_12/a_direct_proof_of.

H. Schwichtenberg. Recursion on the partial continuous functionals. In C. Dimitracopoulos, L. Newelski, D. Normann, and J. Steel, editors, *Logic Colloquium '05*, volume 28 of *Lecture Notes in Logic*, pages 173–201. Association for Symbolic Logic, 2006.

M. Seisenberger. *On the Constructive Content of Proofs*. PhD thesis, Mathematisches Institut der Universität München, 2003.

A. S. Troelstra, editor. *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*, volume 344 of *Lecture Notes in Mathematics*. Springer Verlag, Berlin, Heidelberg, New York, 1973.