A new method for establishing conservativity of classical systems over their intuitionistic version

Thierry Coquand and Martin Hofmann

Dedicated to Roger Hindley for his 60th Birthday

We use a syntactical notion of Kripke models to obtain interpretations of subsystems of arithmetic in their intuitionistic counterparts. This yields in particular a new proof of Buss' result that the Skolem functions of Bounded Arithmetic are polynomial time computable.

Introduction

One goal of Hilbert's programme (Hilbert 1971) was to make sense of classical reasoning, which makes use of the law of excluded-middle, from an intuitionistic viewpoint. If so understood, it has been answered positively for the case of Peano arithmetic PA by the usual Gödel-Gentzen-Bernays negative translation of this system in its intuitionistic version HA (Heyting arithmetic). This translation is a powerful and purely syntactical method for reducing a classical system to its intuitionistic version. We remark that this result motivated the distinction between *intuitionism* and *finitism*, which is usually equated with primitive recursive arithmetic PRA (Troelstra & van Dalen 1988). It follows from the work of Gödel that there is no interpretation of PA or HA in PRA; however Parsons' result, discussed below, shows that there are a priori non finitary systems that can be interpreted in PRA.

Unfortunately, the negative translation does not always work.

The most prominent example is the system $PA^\omega + AC$, where PA^ω is an extension of PA in which one can quantify over functions, functionals, ... and AC is the Axiom of Choice

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

In this case, the negative interpretation of AC which has the form

$$(\forall x)\neg\neg(\exists y)A^*(x,y) \to \neg\neg(\exists f)(\forall x)A^*(x,f(x))$$

cannot be proved in the intuitionistic version $HA^\omega + AC$. It can be shown however in this case that the proof-theoretic strength of the classical version $PA^\omega + AC$ is strictly greater than the one of its intuitionistic version $HA^\omega + AC$. This excludes a priori any purely syntactical translation of the classical system in its intuitionistic version.

There are other cases, where it is known by other methods (such as ordinal analysis) that the two versions have the same strength, but where the negative translation nevertheless does not seem to work. An example of this latter case is the system $I\Sigma_1^0 + EM$ which is a subsystem of PA in which induction is restricted to Σ_1^0 (existential) statements.

The negative translation of an existential statement $(\exists x)\varphi$ for atomic φ is easily seen to be equivalent to $\neg\neg(\exists x)\varphi$ which is not, in general, an existential statement. But it seems that we need induction over such statements in order to interpret the negative translation of $I\Sigma_1^0$!

In this case, it is natural to ask if a there is not a purely syntactical translation of the classical system in its intuitionistic version, which would avoid the use of an ordinal analysis.

In this paper we present such a method.

Applied to $I\Sigma_1^0 + EM$, it shows directly the conservativity of this system over its intuitionistic version for Π_2^0 statements. Since it is a direct consequence of modified realisability that $I\Sigma_1^0$ is conservative over primitive recursive arithmetic for Π_2^0 statements, this can be seen as yet a new proof of Parsons' result (Parsons 1970): $I\Sigma_1^0 + EM$ is conservative over primitive recursive arithmetic for Π_2^0 statements. This result is important since primitive recursive arithmetic is often thought to represent exactly finitism (as a restricted form of intuitionism). This conservativity establishes that one can make sense in a finitary way of classical logic with Σ_1^0 induction, schema which requires *a priori* some appeal to infinity (a set defined by a Σ_1^0 formula is in general not decidable).

The method we present can be extended to an interpretation of systems with König's lemma and the axiom of choice. We only sketch this extension and instead apply in detail the method to CPV^ω—a higher-order extension of Buss' Bounded Arithmetic introduced by Cook and Urquhart (Cook & Urquhart 1993). We obtain a new and relatively simple proof that the first-order functions definable therein are of polynomial complexity.

Our method is reminiscent of an argument due to Buchholz (Buchholz 1977) used by him in order to show that a positive inductive definition (ID_1) can be translated into a strictly positive one.

1. Peano arithmetic with Σ_1-induction

In this section we present a method to interpret a system equivalent to $I\Sigma_1^0 + EM$, i.e., a fragment of Peano arithmetic with induction restricted to Σ_1^0-formulas, to its intuitionistic version. We emphasise that the result in itself is not new; a proof using cut elimination can be found in (Parsons 1970). A proof using Gödel's *Dialectica* interpretation can be found in (Parsons 1972; Jervell 1998). The volume (Aczel *et al.* 1992) contains still two different proofs (Wainer & Wallen 1992; Pohlers 1992). Indeed, it was the analogy between the proof in (Wainer & Wallen 1992) and the Ω-rule of Buchholz (Buchholz 1981; Buchholz 1977) which suggested the present method.

The advantage of proofs using functional interpretations like the present one or the one involving the *Dialectica* interpretation is that the extracted functional witness for a Π_2^0 is obtained directly by a structural induction on proofs and does not involve an exhaustive search component as witnesses obtained via cut elimination do. One might thus hope that our method (as well as the one based on the *Dialectica* interpretation) might have applications to program extraction from classical proofs. Our method is rather different from the *Dialectica* interpretation even at the level of the extracted programs. For example, the *Dialectica* interpretation translates an instance of the contraction rule as a case distinction whereas under our interpretation it is interpreted as a duplication of a variable like in realisability.

Whether or not this results in a gain of efficiency in the extracted programs remains of course to be found out by carrying out practical examples.

1.1. *The system PRA^ω*

We consider a variation of the system HA^ω (Troelstra & van Dalen 1988) called PRA^ω where

— The induction scheme is restricted to formulae of the form $A(x) \stackrel{\text{def}}{=} \exists y.g(x,y)=0$ (these are called purely existential or Σ^0_1-formulae.)

— The target type of primitive recursive definitions is restricted to ground type N.

We also include the axiom $x=0 \vee \neg x=0$ expressing decidability of atomic formulas. It can be seen that in the presence of this axiom quantifier-free formulas are equivalent to atomic formulas. On the other hand, with induction over Σ^0_1 formulas with quantifier-free kernel this would be provable.

It would be possible to use a restricted version of this system with variables ranging only over natural numbers and functions, like the system of elementary analysis EL_1 used in (Troelstra 1974).

By modified realisability it can be seen that if $PRA^\omega \vdash (\forall x)(\exists y)\ A(x,y)$ then there exists $f : N \to N$ primitive recursive such that $A(n, f(n))$ for all n.

The modified realisability we use is completely standard; see for instance (Troelstra & van Dalen 1988). It will be spelt out below for the similar system IPV^ω.

Our task is to interpret $PRA^\omega + EM$ in PRA^ω. As mentioned above, the negative translation in itself does not work because the translation does not preserve Σ^0_1-formulas. However it does provide a syntactical translation of $PRA^\omega + EM$ into the system $PRA^\omega + MP$ where MP is *Markov's principle* (Troelstra & van Dalen 1988) which is the schema

$$\neg\neg(\exists x)\varphi \to (\exists x)\varphi$$

for each purely existential formula $(\exists x)\varphi$. What we present now, is an interpretation of $PRA^\omega + MP$ into PRA^ω. The idea is to use a Kripke model, internally definable in PRA^ω, which will be a model of $PRA^\omega + MP$.

1.2. *Kripke semantics for Markov's principle*

Let $s, t : N \to N$ be (not necessarily closed terms). We write st for the term $\lambda x\colon N.s(x)t(x)$ (point-wise multiplication) and $T(s)$ for the formula $\exists x.s(x)=0$. The worlds of our Kripke model are terms $t : N \to N$ thought of as codes for Σ^0_1 formulae via $T(-)$. Accordingly, the partial order $s \preceq t$ is defined to be $T(t) \to T(s)$. Notice that for this partial order, there is a greatest element $\lambda x\colon N.1$ and st is a greatest lower bound of s and t, i.e., $T(st) \leftrightarrow T(s) \vee T(t)$. Thus, we have a meet-semi lattice. (In the case of ID_1 the meet-semi lattice can be formed by taking sequents which contain the inductively defined predicate only in positive positions; this is implicit in the reduction of $ID^i_\nu + EM$ to ID^i_ν, for the simplest case $\nu = 1$, in (Buchholz 1981), p. 224–227.)

To each formula A of PRA^ω we now associate another formula $f \Vdash A$ with one extra variable $f : N \to N$. The defining clauses, which follow the usual definition of Kripke semantics, are as follows

$$\begin{array}{lcl} f \Vdash A & \stackrel{\text{def}}{=} & A \vee T(f) \quad \text{when } A \text{ is atomic} \\ f \Vdash A \wedge B & \stackrel{\text{def}}{=} & f \Vdash A \wedge f \Vdash B \\ f \Vdash A \vee B & \stackrel{\text{def}}{=} & f \Vdash A \vee f \Vdash B \\ f \Vdash \neg A & \stackrel{\text{def}}{=} & (\forall g)[g \Vdash A \to T(fg)] \\ f \Vdash A \to B & \stackrel{\text{def}}{=} & (\forall g)[g \Vdash A \to fg \Vdash B] \\ f \Vdash (\forall x)A & \stackrel{\text{def}}{=} & (\forall x) f \Vdash A \\ f \Vdash (\exists x)A & \stackrel{\text{def}}{=} & (\exists x) f \Vdash A \end{array}$$

Notice that we could write alternatively for the clauses of implication and negation

$$\begin{array}{lcl} f \Vdash \neg A & \stackrel{\text{def}}{=} & (\forall g \preceq f)[g \Vdash A \to T(g)] \\ f \Vdash A \to B & \stackrel{\text{def}}{=} & (\forall g \preceq f)[g \Vdash A \to g \Vdash B] \end{array}$$

and get an equivalent definition.

We remark that in a first version of this work, we used a composition of the present translation with the negative translation; we get then a direct interpretation of the law of excluded middle and the truth-values get a structure similar to the one described in (Sambin 1996).

Lemma 1.1. If A is provable in PRA^ω without using induction then $f \Vdash A$ is provable in PRA^ω.

Proof. Immediate by induction on derivations in PRA^ω. □

Lemma 1.2. If A is Σ^0_1 then $f \Vdash A$ is equivalent (in PRA^ω) to $A \vee T(f)$.

Proof. In this case A is of the form $(\exists n)[t(n) = 0]$ and so $f \Vdash A$ is $(\exists n)[t(n){=}0 \vee T(f)]$ which is equivalent to $A \vee T(f)$. □

Lemma 1.3. If A is Σ^0_1 then $f \Vdash \neg A$ is equivalent (in PRA^ω) to $A \to T(f)$.

Proof. $f \Vdash \neg A$ is $(\forall g)[g \Vdash A \to T(fg)]$. By lemma 1.2 $g \Vdash A$ is equivalent to $A \vee T(g)$ and hence $h \Vdash \neg A$ is equivalent to $A \to T(f)$. □

Lemma 1.4. If A is Σ^0_1 then $f \Vdash \neg\neg A$ is equivalent (in PRA^ω) to $A \vee T(f)$.

Proof. It is clear that $f \Vdash A$, which is equivalent to $A \vee T(f)$ by lemma 1.2, implies $f \Vdash \neg\neg A$. In the other direction, $f \Vdash \neg\neg A$ is $(\forall g)[g \Vdash \neg A \to T(fg)]$ while $g \Vdash \neg A$ is equivalent to $A \to T(g)$ by lemma 1.3. If we take g such that A is equivalent to $T(g)$ we get that $f \Vdash \neg\neg A$ implies $T(fg)$ that is equivalent to $A \vee T(f)$ as desired. □

Corollary 1.5. (Markov's Principle) If A is Σ^0_1 then $f \Vdash \neg\neg A \to A$.

Lemma 1.6. If $A(x)$ is Σ^0_1 then $f \Vdash A(0) \to (\forall x)[A(x) \to A(S\,x)] \to A(x)$ is provable in PRA^ω.

Proof. Direct by lemma 1.2. □

All this gives a proof of

Theorem 1.7. If A is provable is $PRA^\omega + MP$ then $f \Vdash A$ is provable in PRA^ω for *any* f.

There are two possible ways of seeing that a Σ^0_1 formula A provable in $PRA^\omega + MP$ is provable in PRA^ω using this theorem: we take $f = \lambda n.1$ or f such that $T(f)$ is equivalent to A. In both cases, $f \Vdash A$ is equivalent to A (in PRA^ω) and hence by the theorem, A is provable in PRA^ω. Using a negative translation, it follows that if a Π^0_2-formula A is provable in $PRA^\omega + EM$ then it is provable in PRA^ω.

We can see that our method is quite similar to Friedman's translation (Troelstra & van Dalen 1988); the important difference being that the disjunctively added formula is now a parameter which is "variable" and gets changed during the translation.

2. Bounded arithmetic

In his thesis (Buss 1986) Sam Buss has introduced a fragment of Peano-Arithmetic S^1_2 in which the induction scheme is replaced by the weaker scheme of so-called *NP-induction.*

If $A(x)$ is a formula with free variable x let us define

$$PIND(A(x)) \stackrel{\text{def}}{=} A(0) \to (\forall x.A(\lfloor x/2 \rfloor) \to A(x)) \to \forall x.A(x)$$

The formula $PIND(A(x))$ is called an instance of *polynomial induction.* An instance of *NP-induction* is a formula $PIND(A(x))$ where $A(x)$ is a bounded Σ^0_1-formula, i.e., one of the form

$$A(x) \stackrel{\text{def}}{=} \exists y \leq t(x).s(x,y){=}0$$

where s, t are terms and t does not contain y.

In order that enough bounded Σ^0_1-formulas exist one needs a fair amount of basic functions and quantifier free axioms. We will discuss this point below for a certain extension of bounded arithmetic.

We also remark that in Buss' formulation of S^1_2 the kernel $s(x,y) = 0$ is replaced by a more general concept (sharply bounded formula) which turns out to be equivalent in the more general system which we consider below.

The main result of (Buss 1986) is that the Skolem functions of Π^2_0-statements provable in S^1_2 are polynomial time computable ($PTIME$). Buss proves this result by assigning $PTIME$-functions to cut-free proofs. Cook and Urquhart (Cook & Urquhart 1993) give an alternative proof involving a *Dialectica* interpretation of S^1_2 in IPV^ω—a higher-order generalisation of *intuitionistic* bounded arithmetic. The intuitionistic system IPV^ω admits a straightforward realisability interpretation in PV^ω—a $PTIME$-variant of Gödel's T from which the desired result follows directly.

We will now show how our method gives an alternative proof of Buss' result. The discussion of advantages of functional interpretations compared to cut elimination from Section 1 applies to this case. Moreover, since our method is relatively simple it might be possible to apply it to the weak subsystems of bounded arithmetic studied by Johannsen (Johannsen 1996). He reports that he did not succeed in applying the Dialectica translation whereas Buss' method involving cut elimination does go through.

2.1. *The systems PV^ω and IPV^ω*

The system PV^ω is the simply-typed lambda calculus over one base type o (for natural numbers in binary notation) and constants with types as indicated.

1 The constant zero: $0 : o$.

2 The two successor functions: $s_0, s_1 : o \to o$.

3 Integer division by two ("mix-fix notation"): $\lfloor \frac{}{2} \rfloor$ or $\lfloor\ /2 \rfloor$.

4 The parity function $\mathit{Parity} : o \to o$.

5 The (infix) functions *chop*, *pad*, and *smash*: $\div, \boxplus, \# : o \to o \to o$

6 The ternary conditional $\mathit{Cond} : o \to o \to o \to o$.

7 The bounded recursor $\mathcal{R} : o \to (o \to o \to o) \to (o \to o) \to o \to o$

The intended interpretation of the constants is as follows: $0 = 0, s_0(x) = 2x, s_1(x) = 2x+1, \mathit{Parity}(2x) = 0, \mathit{Parity}(2x+1) = 1, x \div y = \lfloor x/2^{|y|} \rfloor, x \boxplus y = x \cdot 2^{|y|}, x \# y = 2^{|x| \cdot |y|}$ where $|x| = \lceil \log_2(x+1) \rceil$ is the length of x in binary notation. The meaning of the conditional is $\mathit{Cond}(0, y, z) = y$, $\mathit{Cond}(x+1, y, z) = z$. The meaning of the recursor $\mathcal{R}$ is given by

$$\mathcal{R}(g, h, k, x) = \mathit{Cond}(x, g, \mathit{Cond}(t \div k(x), t, k(x)))$$

where $t \stackrel{\text{def}}{=} h(x, \mathcal{R}(g, h, k, \lfloor \frac{x}{2} \rfloor))$. To understand this definition assume functions g, h, k of appropriate type and let f be the function defined by

$$\begin{aligned} f(0) &= g \\ f(x) &= h(x, f(\lfloor \tfrac{x}{2} \rfloor)) \text{ when } x > 0 \end{aligned}$$

We have $\mathcal{R}(g, h, k, x) \leq 2^{|k(x)|}$ when $x > 0$ and $\mathcal{R}(g, h, k, x) = f(x)$ provided that $f(x) \leq 2^{|k(x)|}$ thus in particular if $f(x) \leq k(x)$. Therefore, the recursor $\mathcal{R}$ admits the definition of functions by Cobham's scheme of *bounded recursion on notation* and it follows from Cobham's theorem (Cobham 1965) that all PTIME-functions can be defined in PV^ω.

The system IPV^ω is a many-sorted intuitionistic predicate calculus over PV^ω. Its non-logical axioms are the defining equations for the PV^ω-constants (for a precise definition see (Cook & Urquhart 1993)) and all formulas $\mathit{PIND}(\exists y \leq t(x).s(x, y){=}0)$ where t contains free variables of base type only. We also include the axiom $x{=}0 \vee \neg x{=}0$ expressing decidability of atomic formulas. Like in the case of PRA^ω it follows that quantifier-free formulas are equivalent to atomic formulas.

The bounded quantifier $\exists x \leq t.A$ is shorthand for $\exists x.\mathit{Lesseq}(x, t){=}0 \wedge A$ where Lesseq is a PV^ω term denoting comparison of integers.

2.2. *Realisability for IPV^ω*

The definition of realisability for IPV^ω is completely standard. To each IPV^ω-formula A we associate a formula $\vec{x} \circledR A$ by the following definition:

$$
\begin{array}{lcl}
() \circledR A & \stackrel{\text{def}}{=} & A, \text{ if } A \text{ is atomic} \\
\vec{x}, \vec{y} \circledR A \wedge B & \stackrel{\text{def}}{=} & \vec{x} \circledR A \wedge \vec{y} \circledR B \\
z, \vec{x}, \vec{y} \circledR A \vee B & \stackrel{\text{def}}{=} & (z{=}0 \wedge \vec{x} \circledR A) \vee (z{=}1 \wedge \vec{x} \circledR B) \\
\vec{y} \circledR A \to B & \stackrel{\text{def}}{=} & (\forall \vec{x}.\vec{x} \circledR A \to \vec{y}(\vec{x}) \circledR B) \\
\vec{x} \circledR \forall y.A & \stackrel{\text{def}}{=} & \forall y.\vec{x}(y) \circledR A \\
z, \vec{x} \circledR \exists y.A & \stackrel{\text{def}}{=} & \vec{x} \circledR A(z)
\end{array}
$$

Here, $y(\vec{x})$ means $yx_1x_2\dots x_m$ and $\vec{y}(\vec{x})$ means $(y_1\vec{x}, y_2\vec{x}, \dots, y_n\vec{x})$ when $\vec{x} = (x_1, \dots, x_m)$ and $\vec{y} = (y_1, \dots, y_n)$.

Let us say that a formula A is realizable if there exists a sequence of PV^ω-terms $\vec{t}$ with free variables among those of A such that $IPV^\omega \vdash \vec{t} \circledR A$.

Induction on derivations shows that whenever $IPV^\omega \vdash A$ then A is realizable. The key observation is that an instance of NP-induction can be realised using bounded recursion on notation. Cook and Urquhart add $A \to B$ as a conjunct to $\vec{x} \circledR A \to B$. This ensures that the converse also holds, i.e., realizable formulas are provable. We do not want this property as below in Prop. 2.5 we need to realise a formula of which we do not know whether it is provable.

2.3. *Interpretation of $IPV^\omega + EM$ in IPV^ω*

The following is a straightforward application of realisability and appears as Corollary 8.17 in (Cook & Urquhart 1993).

Proposition 2.1. If $IPV^\omega \vdash \exists y{:}\, o.s(\vec{x}, y){=}0$ then there exists a term $t(\vec{x}) : o$ such that $IPV^\omega \vdash s(\vec{x}, t(\vec{x})) = 0$.

The central result of (Cook & Urquhart 1993) is that a similar result holds for $IPV^\omega{+}EM$ and hence for Buss' bounded arithmetic S^1_2.

Theorem 2.2. Let $A(\vec{x}) \stackrel{\text{def}}{=} \exists y{:}\, o.s(\vec{x}, y){=}0$ be a Σ^0_1 formula. If $IPV^\omega + EM \vdash A(\vec{x})$ then we can find a PV^ω term $t(\vec{x})$ such that $IPV^\omega \vdash s(\vec{x}, t(\vec{x})) = 0$.

Cook and Urquhart prove this by first using the negative translation to interpret $IPV^\omega + EM$ in $IPV^\omega + MP$ (where MP stands for Markov's principle) and then interpreting $IPV^\omega + MP$ in IPV^ω using a variant of the Dialectica interpretation.

We will now give an interpretation of $IPV^\omega + MP$ in IPV^ω using our method.

To every formula A of IPV^ω and fresh variable $f : o \to o$ we can associate a formula $f \Vdash A$ of IPV^ω by the clauses given above in Section 1.2. The proof of Lemma 1.4 then goes through without changes and we can conclude that if A is a (not necessarily bounded) Σ^0_1-formula then $f \Vdash \neg\neg A$ is equivalent to $A \vee T(f)$ where as before $T(f) \stackrel{\text{def}}{=} \exists x.f(x){=}0$.

In order to establish the analogue of Lemma 1.6, i.e., that $f \Vdash A$ for all instances of NP-induction A we need to slightly strengthen the induction scheme of IPV^ω.

Let IPV^ω_+ stand for IPV^ω extended by all formulas $PIND(A(x) \vee B)$ where A is bounded Σ^0_1 and B is an arbitrary formula not containing x. This extension has been introduced by Buss in (Buss 1990). It is shown there that it is complete for a certain class of Kripke-models.

The following is a direct adaptation of Lemma 1.6.

Lemma 2.3. For every bounded Σ^0_1 formula A the formula $f \Vdash PIND(A)$ is provable in IPV^ω_+.

This gives a proof of:

Theorem 2.4. If A is provable in $IPV^\omega + MP$ then $f \Vdash A$ is provable in IPV^ω_+.

In order to deduce the main result that Skolem functions of $IPV^\omega + MP$ are $PTIME$ we must extend our polynomial realisability to IPV^ω_+.

Proposition 2.5. Let $A(x) \stackrel{\text{def}}{=} \exists y \leq t(x).s(x,y)$ and B be arbitrary so that x does not occur in B. The formula $PIND(A(x) \vee B)$ is realizable.

Proof. The formula $PIND(A(x) \vee B)$ is equivalent to

$$(A(0) \vee B) \to (\forall x.\forall z{\leq}t(\lfloor \tfrac{x}{2} \rfloor).s(\lfloor \tfrac{x}{2} \rfloor, z){=}0 \to (\exists y{\leq}t(x).s(x,y){=}0) \vee B) \to \forall x.A(x) \vee B$$

To realise the latter formula, assume variables $u, g_0, \vec{g}_1, v, h_0, \vec{h}_1$ such that (in the sense of IPV^ω-assumptions)

$$\begin{array}{l} u{=}0 \to g_0 \leq t(0) \wedge s(0, g_0){=}0 \\ u{=}1 \to (\vec{g}_1 \circledR B) \\ z{\leq}t(\lfloor \tfrac{x}{2} \rfloor) \wedge s(\lfloor \tfrac{x}{2} \rfloor, z){=}0 \wedge v(x,z){=}0 \to h_0(x,z){\leq}t(x) \wedge s(x, h_0(x,z)){=}0 \\ z \leq t(\lfloor \tfrac{x}{2} \rfloor) \wedge s(\lfloor \tfrac{x}{2} \rfloor, z){=}0 \wedge v(x,z){=}1 \to (\vec{h}_1(x,z) \circledR B) \end{array}$$

that is to say, we assume that

$$\begin{array}{l} u, g_0, \vec{g}_1 \circledR A(0) \vee B \\ v, h_0, \vec{h}_1 \circledR \forall x.\forall z{\leq}t(\lfloor \tfrac{x}{2} \rfloor).s(\lfloor \tfrac{x}{2} \rfloor, z){=}0 \to A(x) \vee B \end{array}$$

We must construct functions $w, f_0, \vec{f}_1$ such that (in IPV^ω under the above assumptions)

$$\begin{array}{l} w(x){=}0 \to f_0(x){\leq}t(x) \wedge s(x, f_0(x)){=}0 \\ w(x){=}1 \to (\vec{f}_1(x) \circledR B) \end{array}$$

i.e., $w, f_0, \vec{f}_1 \circledR A(x) \vee B$.

We will use the PV^ω-recursor to define a function $F : o \to o$ meeting the following specification:

1 if $F(x)$ is even then

$$\lfloor F(x)/2 \rfloor \circledR A(x)$$

2 if $F(x)$ is odd then

$$u{=}1 \ \vee \ (y_0{\leq}t(\lfloor \tfrac{x_0}{2} \rfloor) \wedge v(x_0, y_0){=}1 \wedge s(\lfloor \tfrac{x_0}{2} \rfloor, y_0){=}0$$

where

$$\begin{array}{rcl} x_0 & = & \lfloor F(x)/2 \rfloor \\ y_0 & = & \lfloor F(\lfloor x_0/2 \rfloor)/2 \rfloor \end{array}$$

Notice that in the second case $\vec{h}_1(x_0, y_0) \circledR B$ by the assumption on $\vec{h}_1$. The reason for not making this consequence part of the specification is that it need not be a Σ^1_b-formula hence cannot necessarily be established by NP-induction. The official specification of F, however, is equivalent to an atomic formula.

Once we have such function F we can define the desired functions $w, f_0, \vec{f}_1$ by

$$\begin{array}{l} w(x) = \mathit{Parity}(F(x)) \\ f_0(x) = \lfloor \frac{F(x)}{2} \rfloor \\ \vec{f}_1(x) = \mathit{Cond}(u, \vec{g}_1, \vec{h}_1(x_0, y_0) \\ \quad \text{where } x_0 = \lfloor F(x)/2 \rfloor \text{ and } y_0 = \lfloor F(\lfloor x_0/2 \rfloor)/2 \rfloor \end{array}$$

The following is a sugared definition of F by recursion on notation:

$$\begin{array}{l} F(0) = \textbf{if } u{=}0 \textbf{ then } s_0(g_0) \textbf{ else } s_1(0) \\ F(x) = \textbf{if } \mathit{Parity}(F(\lfloor x/2 \rfloor)){=}0 \\ \quad \textbf{then let } z = \lfloor F(\lfloor x/2 \rfloor)/2 \rfloor \textbf{ in} \\ \quad\quad \textbf{if } v(x,z){=}0 \\ \quad\quad\quad \textbf{then } s_0(h_0(x,z)) \\ \quad\quad\quad \textbf{else } s_1(x) \\ \quad \textbf{else } F(\lfloor x/2 \rfloor) \end{array}$$

where the second case applies when $x > 0$.

Induction on x (using the defining property of h_0) now shows that if $F(x)$ is even then $F(x) \le s_0(t(x)) \wedge s(x, \lfloor \frac{F(x)}{2} \rfloor) = 0$. If $F(x)$ is odd then clearly $F(x) \le s_1(x)$. Therefore, the above can be transformed into a legal PV^ω-definition with $k(x) := \max(s_0(t(x)), s_1(x))$.

By formalised NP-induction on x it is now possible to show that the resulting function F satisfies its defining equations (because the bound k is "valid") and the above specification. The required properties of the derived functions $w, f_0, \vec{f}_1$ are then direct. $\square$

Putting things together yields a proof of Theorem 2.2.

3. Conclusions

We have presented a new method for establishing conservativity of excluded middle with respect to Π^0_2-formulas in situations where the usual negative translation is not applicable. The method thus provides an alternative to techniques using cut elimination or *Dialectica*-interpretation which have been used before. We believe that the new method is simpler than these previous methods and that it should give rise to a more efficient program extraction procedure than those. Of course, some concrete examples would have to be carried out to substantiate this claim. Independently, Avigad has applied this method to a system of polynomial strength (Avigad 1998) and to the system with Σ^1_1 axiom of choice. Avigad's result on bounded arithmetic is slightly weaker than ours since it only gives conservativity for bounded Π^0_2-formulas. Accordingly, the detour via IPV^ω_+ is not needed in his proof.

This method has also been used in the work of Burr (Burr 1998) on fragments of Heyting arithmetic.

Some extensions of the proof for PRA^ω are possible. The negative translation provides us with an interpretation of $PRA^\omega + EM + \Sigma^0_1\text{-}AC$ in $PRA^\omega + EM + \Sigma^0_1\text{-}AC$. It is readily seen that $\Sigma^0_1\text{-}AC$ is valid under our Kripke semantics which is why we can interpret the latter system in $PRA^\omega + \Sigma^0_1\text{-}AC$ which is Π^0_2-conservative over PRA^ω by realisability.

Let EL'_1 be a the system EL_1 of (Troelstra 1974) with induction restricted to Σ^0_1 formula. We can add the fan theorem (FAN) to $EL'_1 + AC$. Then we get an interpretation of $EL'_1 + EM + \Sigma^0_1\text{-}AC + WKL$ (weak König's lemma) in $EL'_1 + FAN + AC$. The usual method of elimination of choice sequence shows that this system is conservative over PRA for Π^0_2 sentences, reading (Troelstra 1974) with EL'_1 instead of EL_1.

These results may be interesting for comparing some recent intuitionistic proofs of theorems such as Hahn-Banach, Heine-Borel (Cederquist 1997), existence of prime ideals in their localic formulation using formal topology and their corresponding proofs in reverse mathematics where these theorems are proved classically in systems similar to $EL'_1 + EM + WKL + \Sigma^0_1\text{-}AC$. For an extension of these results to PRA^ω and sharper formulations, using a functional interpretation, see (Kohlenbach 1992). Whether our method gives an alternative proof of his result remains unexplored.

References

J. Avigad. Interpreting classical theories in constructive ones. Preprint, available from http://macduff.andrew.cmu.edu/avigad/. 1998.

W. Burr. Fragments of Heyting-Arithmetic to appear in Journal of Symbolic Logic, 1999.

W. Buchholz, S. Feferman, W. Pohlers and W. Sieg. *Iterated Inductive Definitions and Subsystems of Analysis: Recent Proof-Theoretical Studies.* Lecture Notes in Mathematics 897, Springer-Verlag, 1981.

W. Buchholz. Some proof-theoretical results in the theories ID^c_ν, ID^i_ν. Preprint, march 1977.

S. R. Buss. On model theory for intuitionistic bounded arithmetic with applications to independence results. In S. R. Buss and P. J. Scott, editors, *Feasible Mathematics*, pages 27–47. Birkhäuser, 1990.

S. R. Buss. *Bounded Arithmetic.* Bibliopolis, 1986.

J. Cederquist. Formal Topology in Type Theory. Ph.D. thesis, Chalmers University, 1997.

A. Cobham. The intrinsic computational difficulty of functions. In Y. Bar-Hillel, editor, *Logic, Methodology, and Philosophy of Science II*, pages 24–30. Springer Verlag, 1965.

S. Cook and A. Urquhart. Functional interpretations of feasibly constructive arithmetic. *Annals of Pure and Applied Logic*, 63:103–200, 1993.

J. van Heijenoort. *From Frege to Gödel.* J. van Heijenoort, editor, Harvard University Press, MA, 1971.

D. Hilbert. The foundations of mathematics. In From Frege to Gödel, J. van Heijenoort, editor, Harvard University Press, MA, 1971, p. 465-479.

H. R. Jervell. A course in proof theory. *Manuscript.* Currently (April 1998) available under www.uio.no/~herman/ptheory.ps.

J. Johannsen. *Schwache Fragmente der Arithmetik und Schwellwertschaltkreise beschränkter Tiefe.* PhD thesis, Universität Erlangen, 1996.

P. Johnstone. *Stone Spaces.* Cambridge University Press, 1982.

U. Kohlenbach. Effective Bounds from Ineffective Proofs in Analysis: an Application of Functional Interpretation and Majorization. Journal of Symbolic Logic, 1992, p. 1239-1273.

P. Aczel, H. Simmons, and S. Wainer, eds. *Proof Theory.* (Leeds, 1990), Cambridge Univ. Press, Cambridge, 1992.

C. Parsons. On a number theoretic choice schema and its relation to induction. In Intuitionism and Proof Theory, Kino, Myhill and Vesley eds., North-Holland, 1970, p. 459-473.

C. Parsons. On n-quantifier induction. Journal of Symbolic Logic vol. 37, 1972, p. 466-482.

W. Pohlers. A short course in ordinal analysis. in Proof theory (Leeds, 1990), 27–78, Cambridge Univ. Press, Cambridge, 1992.

G. Sambin. (1996) A new and elementary method to represent every complete Boolean algebra. in Logic and algebra (Pontignano, 1994), 655–665, Lecture Notes in Pure and Appl. Math., 180, Edited by Aldo Ursini and Paolo Agliano, Dekker, New York.

W. Tait. Normal Derivability in Classical Logic. in Lecture Notes in Mathematics 72, Springer-Verlag, J. Barwise ed., pp. 204-236. 1968

A. Troelstra. Note on the Fan Theorem. Journal of Symbolic Logic, Volume 39, 1974, p. 584-596.

A. Troelstra and D. van Dalen. *Constructivism in Mathematics, vol. II.* North-Holland, 1988.

S. Wainer and L. Wallen. Basic proof theory. in Proof theory (Leeds, 1990), 1–26, Cambridge Univ. Press, Cambridge, 1992.