Fix-point theories in first order logic

Mattias Granberg Olsson

2019-02-21[*]

In this talk we will consider two theories of (non-iterated) fix-points over a base theory, in classical and intuitionistic first order logic respectively. Following the basic definitions I will introduce the problem of (my) interest, namely that of conservativity over the base theory. It is known that the intuitionistic variant (and stronger theories) is conservative over HA, while the classical one is not conservative over PA. After a brief discussion of the history of the positive result, I will present work in progress on a new idea for a proof of this fact, or fragments of it. As such, all purported proofs are sketches and many details remain to be checked.

1 Introduction

We formulate our theories in the language $\mathcal{L}$ of *primitive recursive arithmetic* (PRA) wherein every primitive recursive function symbol is included. The basic theory HA has as axioms $\mathrm{S}(0) = 0 \to \bot$, defining equations for all primitive recursive function symbols and induction for all formulae of the language. PA is HA + LEM.

Let P be a new relation symbol (of unspecified arity for now). $\mathcal{L}(P)$ is the language $\mathcal{L}$ expanded with this symbol. We define the set $\mathrm{POS}(P) \subseteq \mathcal{L}(P)$ of formulae with only strictly positive occurrences of P inductively as follows:

1. Atomic formulae of $\mathcal{L}(P)$ are in $\mathrm{POS}(P)$.
2. If $\varphi, \psi \in \mathrm{POS}(P)$ then $\varphi \mathbin{\wedge\!\!\!\!\vee} \psi \in \mathrm{POS}(P)$.
3. If $\varphi \in \mathcal{L}$ and $\psi \in \mathrm{POS}(P)$ then $\varphi \to \psi \in \mathrm{POS}(P)$.
4. If $\varphi \in \mathrm{POS}(P)$ then $\exists\!\!\!\forall x \varphi \in POS(P)$.

Thus $\varphi \in \mathrm{POS}(P)$ iff P does not occur in an antecedent of an implication in φ. For the ensuing discussion we will also need the formula classes $\mathrm{POS}^0(P)$ and $\mathrm{POS}^+(P)$, which are defined by replacing 3 with

[*]Talk at the Initial Types Club at Chalmers, Göteborg.

3^0. If $\varphi \in \mathcal{L}$ is atomic and $\psi \in \mathrm{POS}^0(P)$ then $\varphi \to \psi \in \mathrm{POS}^0(P)$.

and omitting it entirely, respectively. Clearly $\mathrm{POS}^+(P) \subset \mathrm{POS}^0(P) \subset \mathrm{POS}(P)$.

Now suppose P is a unary predicate symbol. For every $\varphi(v_0) \in \mathrm{POS}(P)$ with exactly the variable v_0 free, let I_φ be a new unary predicate symbol and Ax_φ be the sentence

$$\forall v\, \mathrm{I}_\varphi(v) \leftrightarrow \varphi(\mathrm{I}_\varphi; v) \tag{†}$$

where the latter means $\varphi(v)$ with P replaced by I_φ. This is the *fix-point axiom* for φ.

The theories we consider are $\widehat{\mathrm{ID}}_1$ which is PA extended with the above fix-point axioms for all $\varphi(v_0) \in \mathrm{POS}(P)$ and induction for the entire language, and $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$ which is HA with the corresponding extension.

2 The Problem of Conservativity

An important (for my field) subtheory of $\widehat{\mathrm{ID}}_1$ is KF, the *Kripke-Feferman theory of truth*, which is PA expanded with the new unary predicate T, including induction for all of this language, extended by the following axioms[1]

$$\forall s,t(\mathrm{T}(s \ulcorner{=}\urcorner t) \leftrightarrow \mathrm{val}(s) = \mathrm{val}(t)) \tag{1}$$
$$\forall s,t(\mathrm{T}(\ulcorner\neg\urcorner(s\ulcorner{=}\urcorner t)) \leftrightarrow \mathrm{val}(s) \neq \mathrm{val}(t)) \tag{2}$$
$$\forall f(\mathrm{T}(\ulcorner\neg\neg\urcorner f) \leftrightarrow \mathrm{T}(f)) \tag{3}$$
$$\forall f,g(\mathrm{T}(f\ulcorner\wedge\urcorner g) \leftrightarrow \mathrm{T}(f) \wedge \mathrm{T}(g)) \tag{4}$$
$$\forall f,g(\mathrm{T}(\ulcorner\neg\urcorner(f\ulcorner\wedge\urcorner g) {=}{\leftrightarrow} \mathrm{T}(\ulcorner\neg\urcorner f) \vee \mathrm{T}(\ulcorner\neg\urcorner g)) \tag{5}$$
$$\forall f,g(\mathrm{T}(f\ulcorner\vee\urcorner g) \leftrightarrow \mathrm{T}(f) \vee \mathrm{T}(g)) \tag{6}$$
$$\forall f,g(\mathrm{T}(\ulcorner\neg\urcorner(f\ulcorner\vee\urcorner g)) \leftrightarrow \mathrm{T}(\ulcorner\neg\urcorner f) \wedge \mathrm{T}(\ulcorner\neg\urcorner g)) \tag{7}$$
$$\forall x,f(\mathrm{T}(\ulcorner\forall\urcorner xf) \leftrightarrow \forall v\, \mathrm{T}(f[x/\dot{v}])) \tag{8}$$
$$\forall x,f(\mathrm{T}(\ulcorner\neg\forall\urcorner xf) \leftrightarrow \exists v\, \mathrm{T}(\ulcorner\neg\urcorner f[x/\dot{v}])) \tag{9}$$
$$\forall x,f(\mathrm{T}(\ulcorner\exists\urcorner xf) \leftrightarrow \exists v\, \mathrm{T}(f[x/\dot{v}])) \tag{10}$$
$$\forall x,f(\mathrm{T}(\ulcorner\neg\exists\urcorner xf) \leftrightarrow \forall v\, \mathrm{T}(\ulcorner\neg\urcorner f[x/\dot{v}])) \tag{11}$$
$$\forall f(\mathrm{T}(\ulcorner\mathrm{T}\urcorner(f)) \leftrightarrow \mathrm{T}(\mathrm{val}(f))) \tag{12}$$
$$\forall f(\mathrm{T}(\ulcorner\neg\,\mathrm{T}\urcorner(f)) \leftrightarrow \mathrm{T}(\ulcorner\neg\urcorner\,\mathrm{val}(f)) \vee \neg\,\mathrm{Sent}_{\mathrm{T}}(\mathrm{val}(f))) \tag{13}$$

(NB: slightly nonstandard notation). The symbols in raised corners are the (codes of the) formal symbols inside, which PA can "reason about". Likewise, val is the formal valuation function on formal terms, $\dot{}$ is the formal numeral function, $[x/y]$ denotes formal substitution of y for x and $\mathrm{Sent}_{\mathrm{T}}$ is the formal $\mathcal{L}(\mathrm{T})$-sentence predicate. Note that the right hand sides of the axioms are strictly positive in T as defined above. Thus these axioms can be combined into a single sentence which is an instance of (†) (with T for I).

[1] In my talk I actually presented only half of these axioms. This is the correct axiomatisation of KF. This doesn't affect the other properties I claimed KF to have, e.g. the ability to prove the consistency of PA.

For a truth theory like KF, the question of conservativity is that of how much "strength" is added by the truth predicate T (it has for instance at times been argued that the "correct" truth predicate should be described by a conservative theory). It is known that KF is not conservative over PA, since it proves the consistency of PA ([4, pp. 106, 159 & 217]). Hence the larger theory $\widehat{\mathrm{ID}}_1$ does and is as well.

2.1 A partial conservativity result

The following is a preliminary conservativity result for a fragment of $\widehat{\mathrm{ID}}_1$.

Proposition 1. *Let* $\widehat{\mathrm{ID}}_1^{\Pi_1}$ *be* $\mathrm{PA} + \{\mathrm{Ax}_\varphi \mid \varphi(v_0) \in \mathrm{POS}(P) \cap \Pi_1\}$*. Then* $\widehat{\mathrm{ID}}_1^{\Pi_1}$ *is conservative over (in fact relatively interpretable in)* PA.

Proofsketch. We show for each $\varphi(v_0) \in \mathrm{POS}(P) \cap \Pi_1$ there is a $\psi(v_0) \in \mathcal{L}_{\mathrm{PA}}$ such that

$$\mathrm{PA} \vdash \forall x \psi(x) \leftrightarrow \varphi(\psi; x).$$

Let Sat_{Π_1} be a Π_1-satisfaction predicate for Π_1-formulae, i.e. a Π_1-formula s.th.

$$\mathrm{PA} \vdash \mathrm{Sat}_{\Pi_1}(\langle \bar{x} \rangle, \ulcorner \vartheta \urcorner) \leftrightarrow \vartheta(\bar{x})$$

for all Π_1-formulae ϑ. Given φ consider $\tilde{\varphi} = \varphi(\mathrm{Sat}'_{\Pi_1}; v_0)$ where $\mathrm{Sat}'_{\Pi_1}(x)$ is $\mathrm{Sat}_{\Pi_1}(\langle x \rangle, v_1)$ (v_1 is treated as a constant in the substitution, so $\tilde{\varphi}$ has two free variables, v_1 from Sat'_{Π_1} and v_0 from φ). By the diagonal lemma

$$\mathrm{PA} \vdash \psi(x) \leftrightarrow \tilde{\varphi}(x, \ulcorner \psi \urcorner)$$

for some $\psi(v_0) \in \mathcal{L}_{\mathrm{PA}}$. Since $\tilde{\varphi} \in \Pi_1(\mathrm{PA})$ (P occurs only positively), $\psi \in \Pi_1(\mathrm{PA})$ and

$$\mathrm{PA} \vdash \psi(x) \leftrightarrow \tilde{\varphi}(x, \ulcorner \psi \urcorner) \leftrightarrow \varphi(\mathrm{Sat}'_{\Pi_1}[v_1/\ulcorner \psi \urcorner]; x) \leftrightarrow \varphi(\psi; x).$$

□

In particular, $\mathrm{KF} \restriction \Pi_1$ should be conservative over PA.

3 The Intuitionistic Case

The intuitionistic theory $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$, on the other hand, is conservative over HA.

3.1 Some History

To the best of my knowledge, the first result along these lines was by Buchholz ([3]), who proved that the fragment of $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$ axiomatised by (†) only for $\varphi \in \mathrm{POS}^+(P)$ is conservative over HA for *essentially* $\exists$*-free* formulae; these contain no disjunctions and $\exists$ only in front of atomic formulae (i.e. term equations).[2] This was subsequently improved by Arai ([1], [2]) and Rüede and Strahm ([5]) (approximately) as follows:

[2](Proper) $\exists$-free formulae contains neither $\vee$ nor $\exists$.

1997 Buchholz: Conservativity for ess. $\exists$-free formulae with fix-points for $\mathrm{POS}^+(P)$.

1998 Arai: Conservativity for all formulae with (finitely iterated) fix-points for $\mathrm{POS}^+(P)$.

2002 Rüede and Strahm: Conservativity for $\exists$-free and Π_2 formulae with (finitely iterated) fix-points for $\mathrm{POS}(P)$.

2011 Arai: Conservativity for all formulae with (finitely iterated) fix-points for $\mathrm{POS}(P)$.

3.2 Adapting the idea in 2.1 to the intuitionistic case

We will now sketch an idea for a proof of the following extension of Buchholz' result. Let $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,0}$ be the fragment of $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$ with fix-point axioms only for formulae in $\mathrm{POS}^0(P)$.

Theorem 1. *$\widehat{\mathrm{ID}}_1^{\mathrm{i}\,0}$ is conservative over HA with respect to essentially $\exists$-free formulae.*

We use a realizability interpretation to "reduce" the number of existential quantifiers in the axioms. We use the "standard" Kleene realizability by numbers in HA (see e.g. [6]):

$$\begin{aligned}
x\,\underline{\mathrm{r}}(t = s) &\text{ is } t = s \\
x\,\underline{\mathrm{r}}(\varphi \wedge \psi) &\text{ is } \rho_1(x)\,\underline{\mathrm{r}}\,\varphi \wedge \rho_2(x)\,\underline{\mathrm{r}}\,\psi \\
x\,\underline{\mathrm{r}}(\varphi \vee \psi) &\text{ is } (\rho_1(x) = 0 \to \rho_2(x)\,\underline{\mathrm{r}}\,\varphi) \wedge (\rho_1(x) \neq 0 \to \rho_2(x)\,\underline{\mathrm{r}}\,\psi) \\
x\,\underline{\mathrm{r}}(\varphi \to \psi) &\text{ is } \forall y\,\underline{\mathrm{r}}\,\varphi(\exists u T(x,y,u) \wedge \forall v(T(x,y,v) \to U(v)\,\underline{\mathrm{r}}\,\psi)) && [u, v \text{ fresh}] \\
x\,\underline{\mathrm{r}}(\forall y \varphi) &\text{ is } \forall y(\exists u T(x,y,u) \wedge \forall v(T(x,y,v) \to U(v)\,\underline{\mathrm{r}}\,\varphi)) && [u, v \text{ fresh}] \\
x\,\underline{\mathrm{r}}(\exists y \varphi) &\text{ is } \rho_2(x)\,\underline{\mathrm{r}}\,\varphi[y/\rho_1(x)]
\end{aligned}$$

Here T and U are Kleene's predicate and result-extracting function and ρ_1, ρ_2 are the first and second component functions. Note that the only $\exists$ left are in front of atoms.

Let R be a new binary predicate, the intended meaning of which is realizability of P; that is we stipulate that

$$x\,\underline{\mathrm{r}}\,P(y) \text{ is } R(x,y)$$

in HA expanded to $\mathcal{L}(R)$. By straightforward induction we get

Lemma 1. *If $\varphi(P;\bar{y}) \in \mathrm{POS}^0(P)$, then $x\,\underline{\mathrm{r}}\,\varphi(P;\bar{y}) \in \mathrm{POS}^0(R)$ and is essentially $\exists$-free.*

Definition 1 (Slightly nonstandard notation)**.** Let Σ_1 be the set of all formulae of the form $\exists y \psi$ for ψ quantifier free, and let Π_2 be the set of all formulae of the form $\forall x \exists y \psi$ with ψ quantifier free. Let $\Sigma_1(\mathrm{HA})$ be the least set containing Σ_1 which is closed under conjunction, existential quantification and implications by atomic antecedents. Let $\Pi_2(\mathrm{HA})$ be the least set containing Π_2 and closed under conjunction, universal quantification and implications with $\Sigma_1(\mathrm{HA})$-antecedents.

Lemma 2. *There is a primitive recursive function (symbol) which (provably in HA) converts any $\Pi_2(\mathrm{HA})$-formula to an equivalent (over HA) formula in Π_2.*

Proofsketch. We show existence via straightforward induction, the only notable case being $\to$. The proof makes it probable that the transformation is primitive recursive. $\square$

Lemma 3. *If $\varphi \in \mathrm{POS}^0(P)$ is essentially $\exists$-free and $\vartheta \in \Pi_2$ has two free variables, then $\varphi(\vartheta; \bar{z}) \in \Pi_2(\mathrm{HA})$.*

Proofsketch. Induction. $\square$

The idea now is to use a (rather specific) Π_2-formulation Sat of the classical satisfaction predicate for Π_2-formulae, and show that it is a satisfaction predicate for Π_2-formulae in HA. Then we follow the idea of the classical case and define the interpretation $\iota(x, y) \in \mathcal{L}$ of $x \, \underline{\mathrm{r}} \, \mathrm{I}_\varphi(y)$ via the Diagonal lemma:

$$\mathrm{HA} \vdash \iota(x, y) \leftrightarrow \tilde{\varphi}(\mathrm{Sat}(\langle \cdot, \cdot \rangle, \ulcorner \iota \urcorner); x, y)$$

where $\tilde{\varphi}(R; x, y)$ is $x \, \underline{\mathrm{r}} \, \varphi(y)$.

This is part of showing that

$$\mathrm{HA} \vdash \exists x \, \underline{\mathrm{r}} \, \mathrm{Ax}$$

for every axiom Ax of $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,0}$. The rest is showing that realizability is closed under (enough) intuitionistic logic. Finally

$$\mathrm{HA} \vdash \exists x \, \underline{\mathrm{r}} \, \varphi \to \varphi$$

for essentially $\exists$-free φ (see [6]), yielding Theorem 1.

References

[1] Toshiyasu Arai. Some results on cut-elimination, provable well-orderings, induction and reflection. *Annals of Pure and Applied Logic*, 95:93–184, 1998.

[2] Toshiyasu Arai. Quick cut-elimination for strictly positive cuts. *Annals of Pure and Applied Logic*, 162:807–815, 2011.

[3] Wilfried Buchholz. An intuitionistic fixed point theory. *Archive for Mathematical Logic*, 37:21–27, 1997.

[4] Volker Halbach. *Axiomatic Theories of Truth.* Cambridge University Press, 2011.

[5] Christian Rüede and Thomas Strahm. Intuitionistic fixed point theories for strictly positive operators. *Mathematical Logic Quaterly*, 48:195–202, 2002.

[6] Anne Sjerp Troelstra and Dirk van Dalen. *Constructivism in Mathematics: An Introduction, vol. 1*, volume 121 of *Studies in Logic and the Foundations of Mathematics.* Elsevier Science Publishers B.V., 1988.