Partial conservativity of $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$ over Heyting arithmetic via realizability

Mattias Granberg Olsson, University of Gothenburg

(Joint work in progress with Graham Leigh)
Logic Coloquium 2019, Prague
2019-08-15

Setting

$\mathcal{L}_0 = \mathcal{L}_{\mathrm{PRA}}$.
P new (unary) relation symbol and $\mathcal{L}_P =$ "$\mathcal{L}_0 + P$".
For each $\Phi(P; x) \in \mathcal{L}_P$ (*operator form*) let I_Φ be a new unary relation symbol.
Consider the *fix-point axiom*

$$I_\Phi(x) \leftrightarrow \Phi(I_\Phi; x). \tag{1}$$

The theory $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$ is $\mathrm{HA} + (1)$ for all $\Phi(P; x) \in \mathcal{L}$ where P occurs only (strictly) positively. The classical variant is $\widehat{\mathrm{ID}}_1$.
(Note that there is no assumption on minimality or similar, as opposed to the classical theory of *least* fixed points ID_1.)

Setting

$\mathcal{L}_0 = \mathcal{L}_{\mathrm{PRA}}$.
P new (unary) relation symbol and $\mathcal{L}_P =$ "$\mathcal{L}_0 + P$".
For each $\Phi(P; x) \in \mathcal{L}_P$ (*operator form*) let I_Φ be a new unary relation symbol.
Consider the *fix-point axiom*

$$I_\Phi(x) \leftrightarrow \Phi(I_\Phi; x). \tag{1}$$

The theory $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$ is HA + (1) for all $\Phi(P; x) \in \mathcal{L}$ where P occurs only (strictly) positively. The classical variant is $\widehat{\mathrm{ID}}_1$.
(Note that there is no assumption on minimality or similar, as opposed to the classical theory of *least* fixed points ID_1.)

Setting

$\mathcal{L}_0 = \mathcal{L}_{\mathrm{PRA}}$.
P new (unary) relation symbol and $\mathcal{L}_P =$ "$\mathcal{L}_0 + P$".
For each $\Phi(P; x) \in \mathcal{L}_P$ (*operator form*) let I_Φ be a new unary relation symbol.
Consider the *fix-point axiom*

$$I_\Phi(x) \leftrightarrow \Phi(I_\Phi; x). \tag{1}$$

The theory $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$ is HA + (1) for all $\Phi(P; x) \in \mathcal{L}$ where P occurs only (strictly) positively. The classical variant is $\widehat{\mathrm{ID}}_1$.
(Note that there is no assumption on minimality or similar, as opposed to the classical theory of *least* fixed points ID_1.)

Setting

$\mathcal{L}_0 = \mathcal{L}_{\mathrm{PRA}}$.
P new (unary) relation symbol and $\mathcal{L}_P =$ "$\mathcal{L}_0 + P$".
For each $\Phi(P; x) \in \mathcal{L}_P$ (*operator form*) let I_Φ be a new unary relation symbol.
Consider the *fix-point axiom*

$$I_\Phi(x) \leftrightarrow \Phi(I_\Phi; x). \tag{1}$$

The theory $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$ is HA + (1) for all $\Phi(P; x) \in \mathcal{L}$ where P occurs only (strictly) positively. The classical variant is $\widehat{\mathrm{ID}}_1$.
(Note that there is no assumption on minimality or similar, as opposed to the classical theory of *least* fixed points ID_1.)

Setting

$\mathcal{L}_0 = \mathcal{L}_{\mathrm{PRA}}$.
P new (unary) relation symbol and $\mathcal{L}_P =$ "$\mathcal{L}_0 + P$".
For each $\Phi(P; x) \in \mathcal{L}_P$ (*operator form*) let I_Φ be a new unary relation symbol.
Consider the *fix-point axiom*

$$I_\Phi(x) \leftrightarrow \Phi(I_\Phi; x). \tag{1}$$

The theory $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$ is $\mathrm{HA} + (1)$ for all $\Phi(P; x) \in \mathcal{L}$ where P occurs only (strictly) positively. The classical variant is $\widehat{\mathrm{ID}}_1$.
(Note that there is no assumption on minimality or similar, as opposed to the classical theory of *least* fixed points ID_1.)

Setting

$\mathcal{L}_0 = \mathcal{L}_{\mathrm{PRA}}$.
P new (unary) relation symbol and $\mathcal{L}_P =$ "$\mathcal{L}_0 + P$".
For each $\Phi(P; x) \in \mathcal{L}_P$ (*operator form*) let I_Φ be a new unary relation symbol.
Consider the *fix-point axiom*

$$I_\Phi(x) \leftrightarrow \Phi(I_\Phi; x). \tag{1}$$

The theory $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$ is HA + (1) for all $\Phi(P; x) \in \mathcal{L}$ where P occurs only (strictly) positively. The classical variant is $\widehat{\mathrm{ID}}_1$.
(Note that there is no assumption on minimality or similar, as opposed to the classical theory of *least* fixed points ID_1.)

Setting

$\mathcal{L}_0 = \mathcal{L}_{\mathrm{PRA}}$.
P new (unary) relation symbol and $\mathcal{L}_P =$ "$\mathcal{L}_0 + P$".
For each $\Phi(P; x) \in \mathcal{L}_P$ (*operator form*) let I_Φ be a new unary relation symbol.
Consider the *fix-point axiom*

$$I_\Phi(x) \leftrightarrow \Phi(I_\Phi; x). \tag{1}$$

The theory $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$ is HA + (1) for all $\Phi(P; x) \in \mathcal{L}$ where P occurs only (strictly) positively. The classical variant is $\widehat{\mathrm{ID}}_1$.
(Note that there is no assumption on minimality or similar, as opposed to the classical theory of *least* fixed points ID_1.)

Known results

I $\widehat{\mathrm{ID}}_1$ is not conservative over PA.

II $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$ is conservative over HA [Ara11].

See also [Buc97], [Ara98] and [RS02].

Known results

I $\widehat{\mathrm{ID}}_1$ is not conservative over PA.

II $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$ is conservative over HA [Ara11].

See also [Buc97], [Ara98] and [RS02].

Known results

I $\widehat{\mathrm{ID}}_1$ is not conservative over PA.

II $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$ is conservative over HA [Ara11].

See also [Buc97], [Ara98] and [RS02].

The goal (vagueness intended)

We seek a new proof of II based on the following ideas:

- Using a notion of realizability $\underline{\mathrm{r}}$ to "transform" $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$ into $\underline{\mathrm{r}}\,\widehat{\mathrm{ID}}_1^{\mathrm{i}} \subseteq \widehat{\mathrm{ID}}_1^{\mathrm{i}}$ (stating the axioms of $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$ are realized), "reducing complexity".
- Construct satisfaction predicates and use the Diagonal Lemma to construct fix points in HA.

This talk will focus on the second part. We will outline that $\widehat{\mathrm{ID}}_1^{\mathrm{i}}{}^*$, the theory of positive fixpoints of almost negative (no $\vee$, $\exists$ only applies to equations) operator forms, is conservative over HA.

The goal (vagueness intended)

We seek a new proof of II based on the following ideas:

- Using a notion of realizability $\underline{\mathsf{r}}$ to "transform" $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$ into $\underline{\mathsf{r}}\,\widehat{\mathrm{ID}}_1^{\mathrm{i}} \subseteq \widehat{\mathrm{ID}}_1^{\mathrm{i}}$ (stating the axioms of $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$ are realized), "reducing complexity".
- Construct satisfaction predicates and use the Diagonal Lemma to construct fix points in HA.

This talk will focus on the second part. We will outline that $\widehat{\mathrm{ID}}_1^{\mathrm{i}}{}^*$, the theory of positive fixpoints of almost negative (no $\vee$, $\exists$ only applies to equations) operator forms, is conservative over HA.

The goal (vagueness intended)

We seek a new proof of II based on the following ideas:

- Using a notion of realizability $\underline{\mathrm{r}}$ to "transform" $\widehat{\mathrm{ID}}^{\mathrm{i}}_1$ into $\underline{\mathrm{r}}\,\widehat{\mathrm{ID}}^{\mathrm{i}}_1 \subseteq \widehat{\mathrm{ID}}^{\mathrm{i}}_1$ (stating the axioms of $\widehat{\mathrm{ID}}^{\mathrm{i}}_1$ are realized), "reducing complexity".
- Construct satisfaction predicates and use the Diagonal Lemma to construct fix points in HA.

This talk will focus on the second part. We will outline that $\widehat{\mathrm{ID}}^{\mathrm{i}}_1{}^*$, the theory of positive fixpoints of almost negative (no $\vee$, $\exists$ only applies to equations) operator forms, is conservative over HA.

The goal (vagueness intended)

We seek a new proof of II based on the following ideas:

- Using a notion of realizability $\underline{\mathrm{r}}$ to "transform" $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$ into $\underline{\mathrm{r}}\,\widehat{\mathrm{ID}}_1^{\mathrm{i}} \subseteq \widehat{\mathrm{ID}}_1^{\mathrm{i}}$ (stating the axioms of $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$ are realized), "reducing complexity".
- Construct satisfaction predicates and use the Diagonal Lemma to construct fix points in HA.

This talk will focus on the second part. We will outline that $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,*}$, the theory of positive fixpoints of almost negative (no $\vee$, $\exists$ only applies to equations) operator forms, is conservative over HA.

The goal (vagueness intended)

We seek a new proof of II based on the following ideas:

- Using a notion of realizability $\underline{\mathsf{r}}$ to "transform" $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$ into $\underline{\mathsf{r}}\,\widehat{\mathrm{ID}}_1^{\mathrm{i}} \subseteq \widehat{\mathrm{ID}}_1^{\mathrm{i}}$ (stating the axioms of $\widehat{\mathrm{ID}}_1^{\mathrm{i}}$ are realized), "reducing complexity".
- Construct satisfaction predicates and use the Diagonal Lemma to construct fix points in HA.

This talk will focus on the second part. We will outline that $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,*}$, the theory of positive fixpoints of almost negative (no $\vee$, $\exists$ only applies to equations) operator forms, is conservative over HA.

An almost negative hierarchy Θ_n

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

$\mathcal{L}$ expansion of $\mathcal{L}_0$ with new relations.
Θ_0 are the formulae of the form

$$\exists x(s = t) \wedge \bigwedge_k \mathrm{at}_k$$

Θ_{n+1} are the formulae of the form

$$\forall x \bigwedge_k (\Theta_n \to \Theta_0)_k$$

Θ_0^* is the closure of $\Theta_0 \cup \mathrm{at} \cup \{\exists x(s = t)\}$ under conjunction.
Θ_{n+1}^* is the closure of $\Theta_{n+1} \cup \Theta_n^*$ under conjunction, universal quantification and the rule: If $\varphi \in \Theta_{n+1}^*$ and $\psi \in \Theta_n^*$, then $\psi \to \varphi \in \Theta_{n+1}^*$.
Important: "$=$" only relation in scope of $\exists$!

An almost negative hierarchy Θ_n

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

$\mathcal{L}$ expansion of $\mathcal{L}_0$ with new relations.

Θ_0 are the formulae of the form

$$\exists x(s = t) \wedge \bigwedge_k \mathrm{at}_k$$

Θ_{n+1} are the formulae of the form

$$\forall x \bigwedge_k (\Theta_n \to \Theta_0)_k$$

Θ_0^* is the closure of $\Theta_0 \cup \mathrm{at} \cup \{\exists x(s = t)\}$ under conjunction.
Θ_{n+1}^* is the closure of $\Theta_{n+1} \cup \Theta_n^*$ under conjunction, universal quantification and the rule: If $\varphi \in \Theta_{n+1}^*$ and $\psi \in \Theta_n^*$, then $\psi \to \varphi \in \Theta_{n+1}^*$.

Important: "$=$" only relation in scope of $\exists$!

An almost negative hierarchy Θ_n

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

$\mathcal{L}$ expansion of $\mathcal{L}_0$ with new relations.
Θ_0 are the formulae of the form

$$\exists x(s = t) \wedge \bigwedge_k \mathrm{at}_k$$

Θ_{n+1} are the formulae of the form

$$\forall x \bigwedge_k (\Theta_n \rightarrow \Theta_0)_k$$

Θ_0^* is the closure of $\Theta_0 \cup \mathrm{at} \cup \{\exists x(s = t)\}$ under conjunction.
Θ_{n+1}^* is the closure of $\Theta_{n+1} \cup \Theta_n^*$ under conjunction, universal quantification and the rule: If $\varphi \in \Theta_{n+1}^*$ and $\psi \in \Theta_n^*$, then $\psi \rightarrow \varphi \in \Theta_{n+1}^*$.
Important: "$=$" only relation in scope of $\exists$!

An almost negative hierarchy Θ_n

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

$\mathcal{L}$ expansion of $\mathcal{L}_0$ with new relations.
Θ_0 are the formulae of the form

$$\exists x(s = t) \wedge \bigwedge_k \mathsf{at}_k$$

Θ_{n+1} are the formulae of the form

$$\forall x \bigwedge_k (\Theta_n \to \Theta_0)_k$$

Θ_0^* is the closure of $\Theta_0 \cup \mathsf{at} \cup \{\exists x(s = t)\}$ under conjunction.
Θ_{n+1}^* is the closure of $\Theta_{n+1} \cup \Theta_n^*$ under conjunction, universal quantification and the rule: If $\varphi \in \Theta_{n+1}^*$ and $\psi \in \Theta_n^*$, then $\psi \to \varphi \in \Theta_{n+1}^*$.

Important: "$=$" only relation in scope of $\exists$!

An almost negative hierarchy Θ_n

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

$\mathcal{L}$ expansion of $\mathcal{L}_0$ with new relations.
Θ_0 are the formulae of the form

$$\exists x(s = t) \wedge \bigwedge_k \mathrm{at}_k$$

Θ_{n+1} are the formulae of the form

$$\forall x \bigwedge_k (\Theta_n \to \Theta_0)_k$$

Θ_0^* is the closure of $\Theta_0 \cup \mathrm{at} \cup \{\exists x(s = t)\}$ under conjunction.
Θ_{n+1}^* is the closure of $\Theta_{n+1} \cup \Theta_n^*$ under conjunction, universal quantification and the rule: If $\varphi \in \Theta_{n+1}^*$ and $\psi \in \Theta_n^*$, then $\psi \to \varphi \in \Theta_{n+1}^*$.

Important: "$=$" only relation in scope of $\exists$!

An almost negative hierarchy Θ_n

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

$\mathcal{L}$ expansion of $\mathcal{L}_0$ with new relations.
Θ_0 are the formulae of the form

$$\exists x(s = t) \wedge \bigwedge_k \mathrm{at}_k$$

Θ_{n+1} are the formulae of the form

$$\forall x \bigwedge_k (\Theta_n \to \Theta_0)_k$$

Θ_0^* is the closure of $\Theta_0 \cup \mathrm{at} \cup \{\exists x(s = t)\}$ under conjunction.
Θ_{n+1}^* is the closure of $\Theta_{n+1} \cup \Theta_n^*$ under conjunction, universal quantification and the rule: If $\varphi \in \Theta_{n+1}^*$ and $\psi \in \Theta_n^*$, then $\psi \to \varphi \in \Theta_{n+1}^*$.

Important: "$=$" only relation in scope of $\exists$!

An almost negative hierarchy Θ_n

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

$\mathcal{L}$ expansion of $\mathcal{L}_0$ with new relations.
Θ_0 are the formulae of the form

$$\exists x(s = t) \wedge \bigwedge_k \mathrm{at}_k$$

Θ_{n+1} are the formulae of the form

$$\forall x \bigwedge_k (\Theta_n \to \Theta_0)_k$$

Θ_0^* is the closure of $\Theta_0 \cup \mathrm{at} \cup \{\exists x(s = t)\}$ under conjunction.
Θ_{n+1}^* is the closure of $\Theta_{n+1} \cup \Theta_n^*$ under conjunction, universal quantification and the rule: If $\varphi \in \Theta_{n+1}^*$ and $\psi \in \Theta_n^*$, then $\psi \to \varphi \in \Theta_{n+1}^*$.

Important: "$=$" only relation in scope of $\exists$!

An almost negative hierarchy Θ_n

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

$\mathcal{L}$ expansion of $\mathcal{L}_0$ with new relations.
Θ_0 are the formulae of the form

$$\exists x(s = t) \wedge \bigwedge_k \mathsf{at}_k$$

Θ_{n+1} are the formulae of the form

$$\forall x \bigwedge_k (\Theta_n \to \Theta_0)_k$$

Θ_0^* is the closure of $\Theta_0 \cup \mathsf{at} \cup \{\exists x(s = t)\}$ under conjunction.
Θ_{n+1}^* is the closure of $\Theta_{n+1} \cup \Theta_n^*$ under conjunction, universal quantification and the rule: If $\varphi \in \Theta_{n+1}^*$ and $\psi \in \Theta_n^*$, then $\psi \to \varphi \in \Theta_{n+1}^*$.
Important: "$=$" only relation in scope of $\exists$!

An almost negative hierarchy Θ_n

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

$\mathcal{L}$ expansion of $\mathcal{L}_0$ with new relations.
Θ_0 are the formulae of the form

$$\exists x(s = t) \wedge \bigwedge_k \mathrm{at}_k$$

Θ_{n+1} are the formulae of the form

$$\forall x \bigwedge_k (\Theta_n \to \Theta_0)_k$$

Θ_0^* is the closure of $\Theta_0 \cup \mathrm{at} \cup \{\exists x(s = t)\}$ under conjunction.
Θ_{n+1}^* is the closure of $\Theta_{n+1} \cup \Theta_n^*$ under conjunction, universal quantification and the rule: If $\varphi \in \Theta_{n+1}^*$ and $\psi \in \Theta_n^*$, then $\psi \to \varphi \in \Theta_{n+1}^*$.
Important: "$=$" only relation in scope of $\exists$!

An almost negative hierarchy Θ_n

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

$\mathcal{L}$ expansion of $\mathcal{L}_0$ with new relations.
Θ_0 are the formulae of the form

$$\exists x(s = t) \wedge \bigwedge_k \mathrm{at}_k$$

Θ_{n+1} are the formulae of the form

$$\forall x \bigwedge_k (\Theta_n \to \Theta_0)_k$$

Θ_0^* is the closure of $\Theta_0 \cup \mathrm{at} \cup \{\exists x(s = t)\}$ under conjunction.
Θ_{n+1}^* is the closure of $\Theta_{n+1} \cup \Theta_n^*$ under conjunction, universal quantification and the rule: If $\varphi \in \Theta_{n+1}^*$ and $\psi \in \Theta_n^*$, then $\psi \to \varphi \in \Theta_{n+1}^*$.
Important: "$=$" only relation in scope of $\exists$!

An almost negative hierarchy Θ_n

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Lemma (1. Θ_n)

1. Θ_n^* *exhaust the almost negative formulae:* $\bigcup_{n\in\mathbb{N}} \Theta_n^* = \mathrm{AN}$.
2. $n > 0$*:* Θ_n *are provably equivalent in* PA *to* Π_{n+1} *and vice versa. Similarly,* Θ_0 *are* PA*-equivalent to* Σ_1*.*
3. *There is a prim. rec.* θ_n *transforming* Θ_n^* *into* Θ_n*, preserving* HA*-equivalence (we will suppress this).*
4. *If* $\varphi(P;\bar{x}) \in \Theta_n^*$ *is positive in* P *and* $\vartheta \in \Theta_n^*$*, then* $\varphi(\vartheta;\bar{x}) \in \Theta_n^*$*.*

Conjecture (2. Diagonal Lemma)

Θ_n^* *is "stable" under the diagonal lemma: if* $\varphi \in \Theta_n^*$ *then the diag. lma. gives* $\psi \in \Theta_n^*$ *with*

$$\mathrm{HA} \vdash \psi \leftrightarrow \varphi(\ulcorner \psi \urcorner).$$

An almost negative hierarchy Θ_n

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Lemma (1. Θ_n)

1. *Θ_n^* exhaust the almost negative formulae:* $\bigcup_{n\in\mathbb{N}} \Theta_n^* = \mathsf{AN}$.
2. *$n > 0$: Θ_n are provably equivalent in PA to Π_{n+1} and vice versa. Similarly, Θ_0 are PA-equivalent to Σ_1.*
3. *There is a prim. rec. θ_n transforming Θ_n^* into Θ_n, preserving HA-equivalence (we will suppress this).*
4. *If $\varphi(P;\bar{x}) \in \Theta_n^*$ is positive in P and $\vartheta \in \Theta_n^*$, then $\varphi(\vartheta;\bar{x}) \in \Theta_n^*$.*

Conjecture (2. Diagonal Lemma)

Θ_n^* is "stable" under the diagonal lemma: if $\varphi \in \Theta_n^*$ then the diag. lma. gives $\psi \in \Theta_n^*$ with

$$\mathrm{HA} \vdash \psi \leftrightarrow \varphi(\ulcorner \psi \urcorner).$$

An almost negative hierarchy Θ_n

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Lemma (1. Θ_n)

1. *Θ_n^* exhaust the almost negative formulae:* $\bigcup_{n\in\mathbb{N}} \Theta_n^* = \mathsf{AN}$.
2. *$n > 0$: Θ_n are provably equivalent in PA to Π_{n+1} and vice versa. Similarly, Θ_0 are PA-equivalent to Σ_1.*
3. *There is a prim. rec. θ_n transforming Θ_n^* into Θ_n, preserving HA-equivalence (we will suppress this).*
4. *If $\varphi(P;\bar{x}) \in \Theta_n^*$ is positive in P and $\vartheta \in \Theta_n^*$, then $\varphi(\vartheta;\bar{x}) \in \Theta_n^*$.*

Conjecture (2. Diagonal Lemma)

Θ_n^* is "stable" under the diagonal lemma: if $\varphi \in \Theta_n^*$ then the diag. lma. gives $\psi \in \Theta_n^*$ with

$$\mathrm{HA} \vdash \psi \leftrightarrow \varphi(\ulcorner\psi\urcorner).$$

An almost negative hierarchy Θ_n

Goal: $\widehat{\mathrm{ID}}^{\mathrm{i}}_1{}^* \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Lemma (1. Θ_n)

1. *Θ^*_n exhaust the almost negative formulae: $\bigcup_{n\in\mathbb{N}} \Theta^*_n = \mathsf{AN}$.*
2. *$n > 0$: Θ_n are provably equivalent in PA to Π_{n+1} and vice versa. Similarly, Θ_0 are PA-equivalent to Σ_1.*
3. *There is a prim. rec. θ_n transforming Θ^*_n into Θ_n, preserving HA-equivalence (we will suppress this).*
4. *If $\varphi(P;\bar{x}) \in \Theta^*_n$ is positive in P and $\vartheta \in \Theta^*_n$, then $\varphi(\vartheta;\bar{x}) \in \Theta^*_n$.*

Conjecture (2. Diagonal Lemma)

Θ^*_n is "stable" under the diagonal lemma: if $\varphi \in \Theta^*_n$ then the diag. lma. gives $\psi \in \Theta^*_n$ with

$$\mathrm{HA} \vdash \psi \leftrightarrow \varphi(\ulcorner \psi \urcorner).$$

An almost negative hierarchy Θ_n

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}}{}^* \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Lemma (1. Θ_n)

1. *Θ_n^* exhaust the almost negative formulae: $\bigcup_{n\in\mathbb{N}} \Theta_n^* = \mathsf{AN}$.*
2. *$n > 0$: Θ_n are provably equivalent in PA to Π_{n+1} and vice versa. Similarly, Θ_0 are PA-equivalent to Σ_1.*
3. *There is a prim. rec. θ_n transforming Θ_n^* into Θ_n, preserving HA-equivalence (we will suppress this).*
4. *If $\varphi(P;\bar{x}) \in \Theta_n^*$ is positive in P and $\vartheta \in \Theta_n^*$, then $\varphi(\vartheta;\bar{x}) \in \Theta_n^*$.*

Conjecture (2. Diagonal Lemma)

Θ_n^* is "stable" under the diagonal lemma: if $\varphi \in \Theta_n^*$ then the diag. lma. gives $\psi \in \Theta_n^*$ with

$$\mathrm{HA} \vdash \psi \leftrightarrow \varphi(\ulcorner\psi\urcorner).$$

An almost negative hierarchy Θ_n

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Lemma (1. Θ_n)

1. *Θ_n^* exhaust the almost negative formulae: $\bigcup_{n\in\mathbb{N}} \Theta_n^* = \mathsf{AN}$.*
2. *$n > 0$: Θ_n are provably equivalent in PA to Π_{n+1} and vice versa. Similarly, Θ_0 are PA-equivalent to Σ_1.*
3. *There is a prim. rec. θ_n transforming Θ_n^* into Θ_n, preserving HA-equivalence (we will suppress this).*
4. *If $\varphi(P;\bar{x}) \in \Theta_n^*$ is positive in P and $\vartheta \in \Theta_n^*$, then $\varphi(\vartheta;\bar{x}) \in \Theta_n^*$.*

Conjecture (2. Diagonal Lemma)

Θ_n^* is "stable" under the diagonal lemma: if $\varphi \in \Theta_n^*$ then the diag. lma. gives $\psi \in \Theta_n^*$ with

$$\mathrm{HA} \vdash \psi \leftrightarrow \varphi(\ulcorner \psi \urcorner).$$

An almost negative hierarchy Θ_n

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Lemma (1. Θ_n)

1. *Θ_n^* exhaust the almost negative formulae: $\bigcup_{n\in\mathbb{N}} \Theta_n^* = \mathsf{AN}$.*
2. *$n > 0$: Θ_n are provably equivalent in PA to Π_{n+1} and vice versa. Similarly, Θ_0 are PA-equivalent to Σ_1.*
3. *There is a prim. rec. θ_n transforming Θ_n^* into Θ_n, preserving HA-equivalence (we will suppress this).*
4. *If $\varphi(P; \bar{x}) \in \Theta_n^*$ is positive in P and $\vartheta \in \Theta_n^*$, then $\varphi(\vartheta; \bar{x}) \in \Theta_n^*$.*

Conjecture (2. Diagonal Lemma)

Θ_n^* is "stable" under the diagonal lemma: if $\varphi \in \Theta_n^*$ then the diag. lma. gives $\psi \in \Theta_n^*$ with

$$\mathrm{HA} \vdash \psi \leftrightarrow \varphi(\ulcorner \psi \urcorner).$$

An almost negative hierarchy Θ_n

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Lemma (1. Θ_n)

1. Θ_n^* *exhaust the almost negative formulae:* $\bigcup_{n\in\mathbb{N}} \Theta_n^* = \mathsf{AN}$.
2. $n > 0$*:* Θ_n *are provably equivalent in* PA *to* Π_{n+1} *and vice versa. Similarly,* Θ_0 *are* PA*-equivalent to* Σ_1*.*
3. *There is a prim. rec.* θ_n *transforming* Θ_n^* *into* Θ_n*, preserving* HA*-equivalence (we will suppress this).*
4. *If* $\varphi(P;\bar{x}) \in \Theta_n^*$ *is positive in* P *and* $\vartheta \in \Theta_n^*$*, then* $\varphi(\vartheta;\bar{x}) \in \Theta_n^*$*.*

Conjecture (2. Diagonal Lemma)

Θ_n^* *is "stable" under the diagonal lemma: if* $\varphi \in \Theta_n^*$ *then the diag. lma. gives* $\psi \in \Theta_n^*$ *with*

$$\mathrm{HA} \vdash \psi \leftrightarrow \varphi(\ulcorner\psi\urcorner).$$

A satisfaction predicate

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Lemma (3. Satisfaction)

There are $\mathsf{Sat}_n(e, F) \in \Theta_n^* \cap L_0$ *with*

$$\mathrm{HA} \vdash \mathsf{Eval}(e, \varphi) \to (\mathsf{Sat}_n(e, \ulcorner \varphi(x) \urcorner) \leftrightarrow \varphi(\mathsf{apl}(e, x)))$$

for each $\varphi \in \Theta_n \cap \mathcal{L}_0$.

Proof.

A satisfaction predicate

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Lemma (3. Satisfaction)

There are $\mathsf{Sat}_n(e, F) \in \Theta_n^* \cap L_0$ *with*

$$\mathrm{HA} \vdash \mathsf{Eval}(e, \varphi) \to (\mathsf{Sat}_n(e, \ulcorner \varphi(x) \urcorner) \leftrightarrow \varphi(\mathsf{apl}(e, x)))$$

for each $\varphi \in \Theta_n \cap \mathcal{L}_0$.

Proof.

Note that the predicates "$F \in \Theta_n$*" are prim. rec., i.e. atomic in* HA. *Define* $\mathsf{Deconstruct}_{n+1}(F, v, G, s, i, f, g)$ *as essentially*

$$F = [\forall v \bigwedge_j s_j] \wedge s_i = [g \to f] \wedge \Theta_n(g) \wedge \Theta_0(f).$$

A satisfaction predicate

Goal: $\widehat{\mathrm{ID}}^{\mathrm{i}}_1{}^* \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Lemma (3. Satisfaction)

There are $\mathsf{Sat}_n(e, F) \in \Theta^*_n \cap L_0$ *with*

$$\mathrm{HA} \vdash \mathsf{Eval}(e, \varphi) \to (\mathsf{Sat}_n(e, \ulcorner \varphi(x) \urcorner) \leftrightarrow \varphi(\mathsf{apl}(e, x)))$$

for each $\varphi \in \Theta_n \cap \mathcal{L}_0$.

Proof.

Note that the predicates "$F \in \Theta_n$" are prim. rec., i.e. atomic in HA*. Define* $\mathsf{Deconstruct}_{n+1}(F, v, G, s, i, f, g)$ *as essentially*

$$F = [\forall v \bigwedge_j s_j] \wedge s_i = [g \to f] \wedge \Theta_n(g) \wedge \Theta_0(f).$$

A satisfaction predicate

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Lemma (3. Satisfaction)

There are $\mathsf{Sat}_n(e, F) \in \Theta_n^* \cap L_0$ *with*

$$\mathrm{HA} \vdash \mathsf{Eval}(e, \varphi) \to (\mathsf{Sat}_n(e, \ulcorner \varphi(x) \urcorner) \leftrightarrow \varphi(\mathsf{apl}(e, x)))$$

for each $\varphi \in \Theta_n \cap \mathcal{L}_0$.

Proof.

Induction on n.

A satisfaction predicate

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Lemma (3. Satisfaction)

There are $\mathsf{Sat}_n(e, F) \in \Theta_n^* \cap L_0$ *with*

$$\mathrm{HA} \vdash \mathsf{Eval}(e, \varphi) \to (\mathsf{Sat}_n(e, \ulcorner\varphi(x)\urcorner) \leftrightarrow \varphi(\mathsf{apl}(e, x)))$$

for each $\varphi \in \Theta_n \cap \mathcal{L}_0$.

Proof.

Induction on n.

Sat_0 *is essentially the classical* Sat_{Σ_1}.

A satisfaction predicate

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Lemma (3. Satisfaction)

There are $\mathsf{Sat}_n(e, F) \in \Theta_n^* \cap L_0$ *with*

$$\mathrm{HA} \vdash \mathsf{Eval}(e, \varphi) \to (\mathsf{Sat}_n(e, \ulcorner \varphi(x) \urcorner) \leftrightarrow \varphi(\mathsf{apl}(e, x)))$$

for each $\varphi \in \Theta_n \cap \mathcal{L}_0$.

Proof.

Induction on n.
We define $\mathsf{Sat}_{n+1}(e, F)$ *as*

$$\begin{aligned} &\mathsf{Eval}(e, \varphi) \wedge \Theta_{n+1}(F) \wedge \\ &\quad \forall x \forall v, G, s, i, f, g\, \mathsf{Deconstruct}_{n+1}(F, v, G, s, i, f, g) \\ &\qquad \to (\mathsf{Sat}_n(e_v^x, g) \to \mathsf{Sat}_0(e_v^x, f))) \end{aligned}$$

A satisfaction predicate

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Lemma (3. Satisfaction)

There are $\mathsf{Sat}_n(e, F) \in \Theta_n^* \cap L_0$ *with*

$$\mathrm{HA} \vdash \mathsf{Eval}(e, \varphi) \to (\mathsf{Sat}_n(e, \ulcorner \varphi(x) \urcorner) \leftrightarrow \varphi(\mathsf{apl}(e, x)))$$

for each $\varphi \in \Theta_n \cap \mathcal{L}_0$.

Proof.

Induction on n.
We define $\mathsf{Sat}_{n+1}(e, F)$ *as*

$$\begin{aligned}
&\mathsf{Eval}(e, \varphi) \wedge \Theta_{n+1}(F) \wedge \\
&\qquad \forall x \forall v, G, s, i, f, g \; \mathsf{Deconstruct}_{n+1}(F, v, G, s, i, f, g) \\
&\qquad\qquad\qquad \to (\mathsf{Sat}_n(e_v^x, g) \to \mathsf{Sat}_0(e_v^x, f)))
\end{aligned}$$

□

Constructing fix-points

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Let $\Phi(P; x)$ be almost negative with P (strictly) positively.
By Lemma 1.1 $\Phi \in \Theta^*_n$ for some $n > 0$.
By Lemma 1.4 $\Phi(\mathrm{Sat}_n(\emptyset^{\cdot}_x, d); x) \in \Theta^*_n$.
Apply the Diagonal Lemma conjecture (2) on d: there is $\psi \in \Theta^*_n$ with

$$\mathrm{HA} \vdash \psi(x) \leftrightarrow \Phi(\mathrm{Sat}_n(\emptyset^{\cdot}_x, \ulcorner \psi \urcorner); x).$$

By Lemma 3 Sat_n is disquotational, so

$$\mathrm{HA} \vdash \psi(x) \leftrightarrow \Phi(\psi; x).$$

Hence any finte fragment of $\widehat{\mathrm{ID}}_1^{\mathrm{i}*}$ is interpretable in HA.
In particular $\widehat{\mathrm{ID}}_1^{\mathrm{i}*}$ is conservative over HA. $\square$

Constructing fix-points

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Let $\Phi(P; x)$ be almost negative with P (strictly) positively.
By Lemma 1.1 $\Phi \in \Theta_n^*$ for some $n > 0$.
By Lemma 1.4 $\Phi(\mathrm{Sat}_n(\emptyset_x^\cdot, d); x) \in \Theta_n^*$.
Apply the Diagonal Lemma conjecture (2) on d: there is $\psi \in \Theta_n^*$ with

$$\mathrm{HA} \vdash \psi(x) \leftrightarrow \Phi(\mathrm{Sat}_n(\emptyset_x^\cdot, \ulcorner \psi \urcorner); x).$$

By Lemma 3 Sat_n is disquotational, so

$$\mathrm{HA} \vdash \psi(x) \leftrightarrow \Phi(\psi; x).$$

Hence any finte fragment of $\widehat{\mathrm{ID}}_1^{\mathrm{i}*}$ is interpretable in HA.
In particular $\widehat{\mathrm{ID}}_1^{\mathrm{i}*}$ is conservative over HA. $\square$

Constructing fix-points

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Let $\Phi(P; x)$ be almost negative with P (strictly) positively.
By Lemma 1.1 $\Phi \in \Theta_n^*$ for some $n > 0$.
By Lemma 1.4 $\Phi(\mathrm{Sat}_n(\emptyset_x^{\cdot}, d); x) \in \Theta_n^*$.
Apply the Diagonal Lemma conjecture (2) on d: there is $\psi \in \Theta_n^*$ with

$$\mathrm{HA} \vdash \psi(x) \leftrightarrow \Phi(\mathrm{Sat}_n(\emptyset_x^{\cdot}, \ulcorner \psi \urcorner); x).$$

By Lemma 3 Sat_n is disquotational, so

$$\mathrm{HA} \vdash \psi(x) \leftrightarrow \Phi(\psi; x).$$

Hence any finte fragment of $\widehat{\mathrm{ID}}_1^{\mathrm{i}*}$ is interpretable in HA.
In particular $\widehat{\mathrm{ID}}_1^{\mathrm{i}*}$ is conservative over HA. $\square$

Constructing fix-points

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Let $\Phi(P; x)$ be almost negative with P (strictly) positively.
By Lemma 1.1 $\Phi \in \Theta_n^*$ for some $n > 0$.
By Lemma 1.4 $\Phi(\mathsf{Sat}_n(\emptyset_x^{\cdot}, d); x) \in \Theta_n^*$.
Apply the Diagonal Lemma conjecture (2) on d: there is $\psi \in \Theta_n^*$ with

$$\mathrm{HA} \vdash \psi(x) \leftrightarrow \Phi(\mathsf{Sat}_n(\emptyset_x^{\cdot}, \ulcorner\psi\urcorner); x).$$

By Lemma 3 Sat_n is disquotational, so

$$\mathrm{HA} \vdash \psi(x) \leftrightarrow \Phi(\psi; x).$$

Hence any finte fragment of $\widehat{\mathrm{ID}}_1^{\mathrm{i}*}$ is interpretable in HA.
In particular $\widehat{\mathrm{ID}}_1^{\mathrm{i}*}$ is conservative over HA. $\square$

Constructing fix-points

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Let $\Phi(P; x)$ be almost negative with P (strictly) positively.
By Lemma 1.1 $\Phi \in \Theta_n^*$ for some $n > 0$.
By Lemma 1.4 $\Phi(\mathrm{Sat}_n(\emptyset_x^{\cdot}, d); x) \in \Theta_n^*$.
Apply the Diagonal Lemma conjecture (2) on d: there is $\psi \in \Theta_n^*$ with

$$\mathrm{HA} \vdash \psi(x) \leftrightarrow \Phi(\mathrm{Sat}_n(\emptyset_x^{\cdot}, \ulcorner \psi \urcorner); x).$$

By Lemma 3 Sat_n is disquotational, so

$$\mathrm{HA} \vdash \psi(x) \leftrightarrow \Phi(\psi; x).$$

Hence any finte fragment of $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,*}$ is interpretable in HA.
In particular $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,*}$ is conservative over HA. □

Constructing fix-points

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Let $\Phi(P; x)$ be almost negative with P (strictly) positively.
By Lemma 1.1 $\Phi \in \Theta_n^*$ for some $n > 0$.
By Lemma 1.4 $\Phi(\mathrm{Sat}_n(\emptyset_x^{\cdot}, d); x) \in \Theta_n^*$.
Apply the Diagonal Lemma conjecture (2) on d: there is $\psi \in \Theta_n^*$ with

$$\mathrm{HA} \vdash \psi(x) \leftrightarrow \Phi(\mathrm{Sat}_n(\emptyset_x^{\cdot}, \ulcorner\psi\urcorner); x).$$

By Lemma 3 Sat_n is disquotational, so

$$\mathrm{HA} \vdash \psi(x) \leftrightarrow \Phi(\psi; x).$$

Hence any finte fragment of $\widehat{\mathrm{ID}}_1^{\mathrm{i}*}$ is interpretable in HA.
In particular $\widehat{\mathrm{ID}}_1^{\mathrm{i}*}$ is conservative over HA. $\square$

Constructing fix-points

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Let $\Phi(P; x)$ be almost negative with P (strictly) positively.
By Lemma 1.1 $\Phi \in \Theta_n^*$ for some $n > 0$.
By Lemma 1.4 $\Phi(\mathsf{Sat}_n(\emptyset^{\cdot}_x, d); x) \in \Theta_n^*$.
Apply the Diagonal Lemma conjecture (2) on d: there is $\psi \in \Theta_n^*$ with

$$\mathrm{HA} \vdash \psi(x) \leftrightarrow \Phi(\mathsf{Sat}_n(\emptyset^{\cdot}_x, \ulcorner\psi\urcorner); x).$$

By Lemma 3 Sat_n is disquotational, so

$$\mathrm{HA} \vdash \psi(x) \leftrightarrow \Phi(\psi; x).$$

Hence any finte fragment of $\widehat{\mathrm{ID}}_1^{\mathrm{i}*}$ is interpretable in HA.
In particular $\widehat{\mathrm{ID}}_1^{\mathrm{i}*}$ is conservative over HA. $\square$

Constructing fix-points

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}}{}^* \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Let $\Phi(P; x)$ be almost negative with P (strictly) positively.
By Lemma 1.1 $\Phi \in \Theta_n^*$ for some $n > 0$.
By Lemma 1.4 $\Phi(\mathsf{Sat}_n(\emptyset_x^{\cdot}, d); x) \in \Theta_n^*$.
Apply the Diagonal Lemma conjecture (2) on d: there is $\psi \in \Theta_n^*$ with

$$\mathrm{HA} \vdash \psi(x) \leftrightarrow \Phi(\mathsf{Sat}_n(\emptyset_x^{\cdot}, \ulcorner\psi\urcorner); x).$$

By Lemma 3 Sat_n is disquotational, so

$$\mathrm{HA} \vdash \psi(x) \leftrightarrow \Phi(\psi; x).$$

Hence any finte fragment of $\widehat{\mathrm{ID}}_1^{\mathrm{i}}{}^*$ is interpretable in HA.
In particular $\widehat{\mathrm{ID}}_1^{\mathrm{i}}{}^*$ is conservative over HA. $\square$

Constructing fix-points

Goal: $\widehat{\mathrm{ID}}^{\mathrm{i}}_1{}^* \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Let $\Phi(P; x)$ be almost negative with P (strictly) positively.
By Lemma 1.1 $\Phi \in \Theta^*_n$ for some $n > 0$.
By Lemma 1.4 $\Phi(\mathsf{Sat}_n(\dot{\emptyset_x}, d); x) \in \Theta^*_n$.
Apply the Diagonal Lemma conjecture (2) on d: there is $\psi \in \Theta^*_n$ with

$$\mathrm{HA} \vdash \psi(x) \leftrightarrow \Phi(\mathsf{Sat}_n(\dot{\emptyset_x}, \ulcorner \psi \urcorner); x).$$

By Lemma 3 Sat_n is disquotational, so

$$\mathrm{HA} \vdash \psi(x) \leftrightarrow \Phi(\psi; x).$$

Hence any finte fragment of $\widehat{\mathrm{ID}}^{\mathrm{i}}_1{}^*$ is interpretable in HA.
In particular $\widehat{\mathrm{ID}}^{\mathrm{i}}_1{}^*$ is conservative over HA. $\square$

Constructing fix-points

Goal: $\widehat{\mathrm{ID}}^{\mathrm{i}}_1{}^* \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Let $\Phi(P; x)$ be almost negative with P (strictly) positively.
By Lemma 1.1 $\Phi \in \Theta^*_n$ for some $n > 0$.
By Lemma 1.4 $\Phi(\mathsf{Sat}_n(\dot{\emptyset}_x, d); x) \in \Theta^*_n$.
Apply the Diagonal Lemma conjecture (2) on d: there is $\psi \in \Theta^*_n$ with

$$\mathrm{HA} \vdash \psi(x) \leftrightarrow \Phi(\mathsf{Sat}_n(\dot{\emptyset}_x, \ulcorner\psi\urcorner); x).$$

By Lemma 3 Sat_n is disquotational, so

$$\mathrm{HA} \vdash \psi(x) \leftrightarrow \Phi(\psi; x).$$

Hence any finte fragment of $\widehat{\mathrm{ID}}^{\mathrm{i}}_1{}^*$ is interpretable in HA.
In particular $\widehat{\mathrm{ID}}^{\mathrm{i}}_1{}^*$ is conservative over HA. $\square$

Constructing fix-points

Goal: $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,*} \vdash \varphi \Rightarrow \mathrm{HA} \vdash \varphi$.

Let $\Phi(P; x)$ be almost negative with P (strictly) positively.
By Lemma 1.1 $\Phi \in \Theta_n^*$ for some $n > 0$.
By Lemma 1.4 $\Phi(\mathsf{Sat}_n(\emptyset_x^{\cdot}, d); x) \in \Theta_n^*$.
Apply the Diagonal Lemma conjecture (2) on d: there is $\psi \in \Theta_n^*$ with

$$\mathrm{HA} \vdash \psi(x) \leftrightarrow \Phi(\mathsf{Sat}_n(\emptyset_x^{\cdot}, \ulcorner\psi\urcorner); x).$$

By Lemma 3 Sat_n is disquotational, so

$$\mathrm{HA} \vdash \psi(x) \leftrightarrow \Phi(\psi; x).$$

Hence any finte fragment of $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,*}$ is interpretable in HA.
In particular $\widehat{\mathrm{ID}}_1^{\mathrm{i}\,*}$ is conservative over HA. $\square$

Thank you for your attention!

References

Toshiyasu Arai.
Some results on cut-elimination, provable well-orderings, induction and reflection.
Annals of Pure and Applied Logic, 95(1–3):93–184, 1998.

Toshiyasu Arai.
Quick cut-elimination for strictly positive cuts.
Annals of Pure and Applied Logic, 162(10):807–815, 2011.

Wilfried Buchholz.
An intuitionistic fixed point theory.
Archive for Mathematical Logic, 37(1):21–27, 1997.

Christian Rüede and Thomas Strahm.
Intuitionistic fixed point theories for strictly positive operators.
Mathematical Logic Quaterly, 48(2):195–202, 2002.