An intensional fixed point theory over first order arithmetic

Gerhard Jäger*

Abstract

The purpose of this article is to present a new theory $\mathsf{IPA}(\sigma)$ for fixed points over arithmetic which allows the building up of fixed points in a very nested and entangled way. But in spite of its great expressive power we can show that the proof-theoretic strength of our theory – which is intensional in a meaning to be described below – is characterized by the Feferman-Schütte ordinal Γ_0. Our approach is similar to the building up of fixed points over state spaces in the propositional modal μ-calculus.

1 Introduction

Fixed point theories play an important role in many branches of mathematical logic and computer science. Typically, an operator Φ mapping the power set $Pow(M)$ of some set M to $Pow(M)$ is given, and we are interested in fixed points of Φ. The logical structure of the generation or definition of such fixed points heavily depends on the complexity of (the description of) Φ and on whether Φ is supposed to be, for example, positive/monotone. A further principal distinction refers to the fact whether we look for minimal or arbitrary fixed points of Φ.

Formal systems for finite iterations of fixed points given by positive arithmetic formulas have first been studied by Feferman in connection with his proof of Hancock's conjecture; see [6]. In Jäger, Kahle, Setzer and Strahm [7] transfinite iterations of this form of fixed point theories are introduced and analyzed from a proof-theoretic point of view. A further good source of reading about theories for arbitrary, not necessarily minimal fixed points is

*Thanks to Andrea Cantini for arranging my stay at the Dipartimento di Filosofia of the Universitá degli Studi di Firenze during which a substantial part of this paper was written.

Cantini [5]. In this monograph, fixed points of inductive definitions provide a uniform way of dealing with truth definitions of various sorts.

The purpose of this article is to present a new theory $\mathsf{IPA}(\sigma)$ for fixed points over arithmetic which allows the building up of fixed points in a very nested and entangled way and goes beyond what has been possible so far. Our approach is similar to the building up of fixed points over state spaces in the propositional modal μ-calculus, which is elegantly described, for example, in Bradfield and Stirling [3] and Arnold and Niwinski [1].

To illustrate one of the possibilities of fixed point formation in $\mathsf{IPA}(\sigma)$, take two arithmetic formulas $A[X, Y, a]$ and $B[X, Y, b]$ of second arithmetic which are positive in X and negative in Y. We introduce a form of σ abstraction and let the expression $\sigma(X, a)A[X, Y, a]$ represent *some* fixed point of the operator $\Phi_{A,Y}$,

$$\Phi_{A,Y} : Pow(\mathbb{N}) \longrightarrow Pow(\mathbb{N}),$$

$$\Phi_{A,Y}(M) \;:=\; \{n \in \mathbb{N} : \mathbb{N} \models A[M, Y, n]\},$$

containing Y as a set parameter. The variable Y is said to be negative in the term $\sigma(X, a)A[X, Y, a]$, hence the formula $C[Y, b]$,

$$C[Y, b] \;:=\; B[Y, \sigma(X, a)A[X, Y, a], b],$$

is positive in Y. In our theory $\mathsf{IPA}(\sigma)$ also a fixed point $\sigma(Y, b)C[Y, b]$ of the new operator Ψ_C,

$$\Psi_C : Pow(\mathbb{N}) \longrightarrow Pow(\mathbb{N}),$$

$$\Psi_C(M) \;:=\; \{n \in \mathbb{N} : \mathbb{N} \models B[M, \sigma(X, a)A[X, M, a], n]\},$$

is available. Observe the impredicativity of this proceeding: the "meaning" of $\sigma(X, a)A[X, Y, a]$ depends on the value of the parameter Y; on the other hand, in order to determine a fixed point N of Ψ_C we have to know the set $\sigma(X, a)A[X, N, a]$. Hence the generation of fixed points in $\mathsf{IPA}(\sigma)$ cannot be nicely stratified as in Feferman's theories $\widehat{\mathsf{ID}}_n$.

Even more complex fixed point constructions can be carried through in $\mathsf{IPA}(\sigma)$. In spite of this significant expressive power we can show that the proof-theoretic strength of our theory – which is intensional in a meaning to be described below – is characterized by the Feferman-Schütte ordinal Γ_0.

2 The theories $\widehat{\mathsf{ID}}_n$: a repetition

In this section we repeat the definitions of the iterated fixed point theories $\widehat{\mathsf{ID}}_n$ considered, for example, in Feferman [6], and the second order fixed point theory FP_0, introduced in Avigad [2]. They form the starting point of our considerations and provide the lower proof-theoretic bound of the system $\mathsf{IPA}(\sigma)$, to be introduced later.

Let $\mathcal{L}_1$ denote any language of first order arithmetic with number variables $a, b, c, u, v, w, x, y, z, \ldots$ (possibly with subscripts), a constant 0 as well as function and relation symbols for all primitive recursive functions and relations. The number terms $(r, s, t, r_1, s_1, t_1, \ldots)$ and formulas of $\mathcal{L}_1$ are defined as usual.

In the following we make use of the usual coding machinery in $\mathcal{L}_1$: $\langle \ldots \rangle$ is a standard primitive recursive function for forming n-tuples $\langle t_1, \ldots, t_n \rangle$, so-called sequence numbers; $(t)_i$ is the ith component of (the sequence coded by) t if i is less than the length of t; i.e. $(t)_i = t_{i+1}$ for all $0 \leq i \leq n-1$, provided that $t = \langle t_1, \ldots, t_n \rangle$.

Given a language $\mathcal{L}'$ comprising $\mathcal{L}_1$ and a set constant P not occurring in $\mathcal{L}'$, we write $\mathcal{L}'[\mathsf{P}]$ for the extension of $\mathcal{L}'$ which is obtained by permitting formulas of the form $(t \in \mathsf{P})$, where $\in$ is the membership relation symbol, as further atomic formulas. A formula A of $\mathcal{L}'[\mathsf{P}]$ is positive in P if all subformulas of A of the form $(t \in \mathsf{P})$ are positive in A in the usual sense.

To present the syntax of the theories $\widehat{\mathsf{ID}}_n$ in some detail, we introduce, for all natural numbers n, *operator forms of level* n and, based on these operator forms, the languages $\mathcal{L}(n)$. We proceed inductively as follows:

1. $\mathcal{L}(0)$ is defined to be the language $\mathcal{L}_1$.

2. Let P be a fresh set constant which does not occur in the language $\mathcal{L}(n)$; we write $\mathcal{L}(n, \mathsf{P})$ for the extension of $\mathcal{L}(n)$ by P. An $\mathcal{L}(n, \mathsf{P})$ formula $A[\mathsf{P}, a]$ which is positive in P and contains at most the variable a free is called an operator form of level n.

3. To each operator form $A[\mathsf{P}, a]$ of level n we associate an new set constant P_A; the language $\mathcal{L}(n+1)$ is the extension of $\mathcal{L}(n)$ by all new set constants which are generated by operator forms of level n.

For every natural number $n \geq 1$, the theory $\widehat{\mathsf{ID}}_n$ is formulated in the language $\mathcal{L}(n)$. Its logical axioms are the usual axioms of classical first order logic with equality in the first sort. The non-logical axioms of $\widehat{\mathsf{ID}}_n$ comprise the axioms of Peano arithmetic PA with the schema of complete induction for all formulas of $\mathcal{L}(n)$ plus the fixed point axioms

$$\text{(Fix)} \qquad\qquad (r \in \mathsf{P}_A) \;\leftrightarrow\; A[\mathsf{P}_A, r]$$

for all number terms r and each set constant P_A which is associated to an operator form $A[\mathsf{P}, a]$ whose level is less than n.

The union of the theories $\widehat{\mathsf{ID}}_n$ for all natural numbers $n \geq 1$ is denoted by $\widehat{\mathsf{ID}}_{<\omega}$. Obviously, each theorem provable in $\widehat{\mathsf{ID}}_{<\omega}$ is already provable in some $\widehat{\mathsf{ID}}_n$ for n being sufficiently large.

(Fix) simply states that the set constant P_A describes *some* fixed point of the operator Φ_A,

$$\Phi_A : Pow(\mathbb{N}) \longrightarrow Pow(\mathbb{N}), \quad \Phi_A(M) \;:=\; \{n \in \mathbb{N} : \mathbb{N} \models A[M, n]\}.$$

Observe that an operator form $A[\mathsf{P}, a]$ of level n may contain set constants representing fixed points of operator forms of levels less than n. In order to fix the meaning of $A[M, n]$ in $\mathbb{N}$, these set constants have to be interpreted before.

Feferman's article [6] provides a detailed proof-theoretic analysis of the theories $\widehat{\mathsf{ID}}_n$. Among other things it is shown there that the proof-theoretic strength of their union $\widehat{\mathsf{ID}}_{<\omega}$ can be characterized by the famous Feferman-Schütte ordinal Γ_0, which describes the limit of predicative mathematics.

Theorem 1 (Feferman) *The theory* $\widehat{\mathsf{ID}}_{<\omega}$ *has proof-theoretic strength* Γ_0.

The theories $\widehat{\mathsf{ID}}_n$ must not be confused with the theories ID_n which ask for least fixed points rather than arbitrary fixed points of the operator forms involved. They are significantly stronger than the $\widehat{\mathsf{ID}}_n$ – cf. e.g. Buchholz, Feferman, Pohlers and Sieg [4] for further details – and do not play a role in our present context.

Let us finish these remarks about the theories $\widehat{\mathsf{ID}}_n$ by pointing out their intensional character. Let $A[\mathsf{P}, a]$ and $B[\mathsf{P}, a]$ be two operator forms of level 0 and let $C[\mathsf{P}, \mathsf{Q}, a]$ be a P-positive formula of the language $\mathcal{L}(0, \mathsf{P}, \mathsf{Q})$, i.e. of $\mathcal{L}(0)$ extended by the set constants P and Q. Furthermore, set

$$D[\mathsf{P}, a] \;:=\; C[\mathsf{P}, \mathsf{P}_A, a] \quad \text{and} \quad E[\mathsf{P}, a] \;:=\; C[\mathsf{P}, \mathsf{P}_B, a].$$

Obviously, $D[\mathsf{P}, a]$ and $E[\mathsf{P}, a]$ are two operator forms of level 1 which differ in the parameters P_A and P_B only. However, there is no way to derive in $\widehat{\mathsf{ID}}_2$ (or even in $\widehat{\mathsf{ID}}_{<\omega}$) that the fixed points associated to $D[\mathsf{P}, a]$ and $E[\mathsf{P}, a]$ contain the same elements provided that P_A and P_B contain the same elements. In general we have that

$$\widehat{\mathsf{ID}}_{<\omega} \nvdash \forall x(x \in \mathsf{P}_A \leftrightarrow x \in \mathsf{P}_B) \rightarrow \forall x(x \in \mathsf{P}_D \leftrightarrow x \in \mathsf{P}_E).$$

This means that all theories $\widehat{\mathsf{ID}}_1, \widehat{\mathsf{ID}}_2, \ldots$ and $\widehat{\mathsf{ID}}_{<\omega}$ behave intensionally with respect to their parameters.

It is convenient for the proof-theoretic treatment of our system $\mathsf{IPA}(\sigma)$ in Section 5 to introduce extensions $\widehat{\mathcal{ID}}_n$ of the theories $\widehat{\mathsf{ID}}_n$ which allow the direct representation of simultaneous fixed points and to work with *systems* of operator forms of finite level rather than individual operator forms. They are formulated in the extensions $\mathcal{L}^e(n)$ of the languages $\mathcal{L}(n)$ defined as follows:

1. $\mathcal{L}^e(0)$ is defined to be the language $\mathcal{L}_1$.

2. Let $\vec{\mathsf{P}} = \mathsf{P}_1, \ldots, \mathsf{P}_p$ and $\vec{\mathsf{Q}} = \mathsf{Q}_1, \ldots, \mathsf{Q}_q$ be sequences of fresh set constants which do not occur in the language $\mathcal{L}^e(n)$ and write $\mathcal{L}^e(n, \vec{\mathsf{P}}, \vec{\mathsf{Q}})$ for the extension of $\mathcal{L}^e(n)$ by $\vec{\mathsf{P}}$ and $\vec{\mathsf{Q}}$. A list

$$\mathbb{S} \;=\; (\, A_1[\vec{\mathsf{P}}, \vec{\mathsf{Q}}, a],\; \ldots,\; A_{p+q}[\vec{\mathsf{P}}, \vec{\mathsf{Q}}, a]\,)$$

   of $\mathcal{L}^e(n, \vec{\mathsf{P}}, \vec{\mathsf{Q}})$ formulas $A_i[\vec{\mathsf{P}}, \vec{\mathsf{Q}}, a]$, for $1 \leq i \leq p+q$, which are positive in $\vec{\mathsf{P}}$ and negative in $\vec{\mathsf{Q}}$ and contain at most the variable a free is called an operator system of level n and signature (p, q).

3. To each operator system $\mathbb{S}$ of level n and signature (p, q) we associate new set constants $\mathsf{P}^1_{\mathbb{S}}, \ldots, \mathsf{P}^p_{\mathbb{S}}$ and $\mathsf{Q}^1_{\mathbb{S}}, \ldots, \mathsf{Q}^q_{\mathbb{S}}$; the language $\mathcal{L}^e(n{+}1)$ is the extension of $\mathcal{L}^e(n)$ by all new set constants which are generated by operator systems of level n.

Given a natural number $n \geq 1$, the theory $\widehat{\mathcal{ID}}_n$, formulated in $\mathcal{L}^e(n)$, is obtained from Peano arithmetic PA by permitting the schema of complete induction for all formulas of $\mathcal{L}^e(n)$ and by adding the following simultaneous fixed point axioms for all operator systems $\mathbb{S} = (A_1[\vec{\mathsf{P}}, \vec{\mathsf{Q}}, a], \ldots, A_{p+q}[\vec{\mathsf{P}}, \vec{\mathsf{Q}}, a])$ of signature (p, q) and level less than n:

$$(\mathrm{Fix}^e) \qquad \begin{array}{lcl}
(r \in \mathsf{P}^1_{\mathbb{S}}) & \leftrightarrow & A_1[\mathsf{P}^1_{\mathbb{S}}, \ldots, \mathsf{P}^p_{\mathbb{S}}, \mathsf{Q}^1_{\mathbb{S}}, \ldots, \mathsf{Q}^q_{\mathbb{S}}, r], \\
\vdots & & \vdots \\
(r \in \mathsf{P}^p_{\mathbb{S}}) & \leftrightarrow & A_p[\mathsf{P}^1_{\mathbb{S}}, \ldots, \mathsf{P}^p_{\mathbb{S}}, \mathsf{Q}^1_{\mathbb{S}}, \ldots, \mathsf{Q}^q_{\mathbb{S}}, r], \\
(r \notin \mathsf{Q}^1_{\mathbb{S}}) & \leftrightarrow & A_{p+1}[\mathsf{P}^1_{\mathbb{S}}, \ldots, \mathsf{P}^p_{\mathbb{S}}, \mathsf{Q}^1_{\mathbb{S}}, \ldots, \mathsf{Q}^q_{\mathbb{S}}, r], \\
\vdots & & \vdots \\
(r \notin \mathsf{Q}^q_{\mathbb{S}}) & \leftrightarrow & A_{p+q}[\mathsf{P}^1_{\mathbb{S}}, \ldots, \mathsf{P}^p_{\mathbb{S}}, \mathsf{Q}^1_{\mathbb{S}}, \ldots, \mathsf{Q}^q_{\mathbb{S}}, r].
\end{array}$$

As above, $\widehat{\mathcal{ID}}_{<\omega}$ is defined to be the union of the theories $\widehat{\mathcal{ID}}_n$ for all natural numbers $n \geq 1$.

It follows immediately from the syntactic setup that each theory $\widehat{\mathsf{ID}}_n$ is contained in $\widehat{\mathcal{ID}}_n$. On the other hand, by employing the usual techniques of coding systems of fixed point equations into fixed point equations it is easily proved that proof-theoretic strength is not improved by moving from $\widehat{\mathsf{ID}}_n$ to $\widehat{\mathcal{ID}}_n$ and that both theories prove the same arithmetic sentences. The detailed proof of the following lemma is left to the reader.

Lemma 2 *For any natural number n greater than 0 and for all sentences A of $\mathcal{L}_1$ we have:*

1. *The theory $\widehat{\mathcal{ID}}_n$ can be embedded into the theory $\widehat{\mathsf{ID}}_n$ so that all arithmetic assertions are preserved;*

$$\widehat{\mathcal{ID}}_n \vdash A \quad \Longleftrightarrow \quad \widehat{\mathsf{ID}}_n \vdash A.$$

2. *The theory $\widehat{\mathcal{ID}}_{<\omega}$ can be embedded into the theory $\widehat{\mathsf{ID}}_{<\omega}$ so that all arithmetic assertions are preserved;*

$$\widehat{\mathcal{ID}}_{<\omega} \vdash A \quad \Longleftrightarrow \quad \widehat{\mathsf{ID}}_{<\omega} \vdash A.$$

A further interesting fixed point theory is Avigad's system FP_0. It is formulated in the language $\mathcal{L}_2$ of second order arithmetic which is obtained from $\mathcal{L}_1$ by adding set variables $U, V, W, X, Y, Z, \ldots$ (possibly with subscripts) and the binary relation symbol $\in$ for membership. $\mathcal{L}_2$ formulas which do not contain bound set variables are called *arithmetic*.

Besides the usual axioms of classical logic with equality in the first sort, the theory FP_0 comprises the *induction axiom*

$$\text{(IA)} \qquad \forall X(0 \in X \ \land \ \forall y(y \in X \ \rightarrow \ y' \in X) \ \rightarrow \ \forall y(y \in X))$$

and, for all arithmetic formulas $A[X, a]$ which are positive in X, the *fixed point axioms*

$$\text{(FP)} \qquad \exists X \forall y(y \in X \ \leftrightarrow \ A[X, y]).$$

Note, however, that $A[X, a]$ may have additional number and set parameters. Hence, if X does not occur in A, the schema (FP) immediately implies comprehension for the arithmetic formula A.

Avigad [2] tells us a lot about the theory FP_0; among other things, it is shown there, that FP_0 is equivalent to Friedman's theory ATR_0 of arithmetic transfinite recursion. A further result states that FP_0 is a conservative extension of $\widehat{\mathsf{ID}}_{<\omega}$ with respect to all formulas of $\mathcal{L}_1$.

Theorem 3 (Avigad) *The theories* FP_0 *and* $\widehat{\mathsf{ID}}_{<\omega}$ *are proof-theoretically equivalent.*

Although being of the same proof-theoretic strength, the theory FP_0 is syntactically much more "flexible" than the rather static system $\widehat{\mathsf{ID}}_{<\omega}$; arbitrary X-positive arithmetic formulas of $\mathcal{L}_2$ with arbitrary number and set parameters are permitted in the build up of new fixed points. In the next section we introduce another theory which is even more liberal with respect to the generation of fixed points and allows rather intricate nestings of those.

3 The theory $\mathsf{IPA}(\sigma)$

The intensional fixed point calculus $\mathsf{IPA}(\sigma)$ will be formulated in the language $\mathcal{L}(\sigma)$. It is obtained from $\mathcal{L}_1$ by adding countably many set variables $U, V, W, X, Y, Z, \ldots$ (possibly with subscripts), the membership relation symbol $\in$ and the set term constructor σ.

The set terms $(R, S, T, R_1, S_1, T_1, \ldots)$ and formulas $(A, B, C, A_1, B_1, C_1, \ldots)$ of $\mathcal{L}(\sigma)$ as well as the collections $POS(U)$ and $NEG(U)$ of U-positive and U-negative $\mathcal{L}(\sigma)$ set terms and formulas are generated simultaneously by the following inductive definition:

(σ1) Every formula of $\mathcal{L}_1$ is a formula of $\mathcal{L}(\sigma)$ and belongs to $POS(U)$ and $NEG(U)$ for any U.

(σ2) Every set variable V is a set term of $\mathcal{L}(\sigma)$ and belongs to $POS(U)$ for any U; moreover, it belongs to $NEG(U)$ for all U different from V.

(σ3) If S is a set term of $\mathcal{L}(\sigma)$ and r a number term, then $(r \in S)$ is a formula of $\mathcal{L}(\sigma)$. If S belongs to $POS(U)$ [$NEG(U)$], then $(r \in S)$ belongs to $POS(U)$ [$NEG(U)$].

(σ4) If A is a formula of $\mathcal{L}(\sigma)$, then so also is $\neg A$. If A belongs to $POS(U)$ [$NEG(U)$], then $\neg A$ belongs to $NEG(U)$ [$POS(U)$].

(σ5) If A and B are formulas of $\mathcal{L}(\sigma)$, then so also is $(A \vee B)$. If A and B belong to $POS(U)$ [$NEG(U)$], then $(A \vee B)$ belongs to $POS(U)$ [$NEG(U)$].

(σ6) If A is a formula of $\mathcal{L}(\sigma)$, then so also is $\exists x A$. If A belongs to $POS(U)$ [$NEG(U)$], then $\exists x A$ belongs to $POS(U)$ [$NEG(U)$].

(σ7) If A is a formula from $POS(X)$, then $\sigma(X,a)A$ is a set term of $\mathcal{L}(\sigma)$, in which all free occurrences of the variables X and a are bound by the set term constructor σ. If A belongs to $POS(U)$ [$NEG(U)$], then $\sigma(X,a)A$ belongs to $POS(U)$ [$NEG(U)$].

Further expressions like $(A \wedge B)$, $(A \rightarrow B)$, $(A \leftrightarrow B)$ and $\forall x A$ are treated as abbreviations as usual. In the following we frequently omit parentheses and implicitly assume that all bound variables have been renamed to avoid conflict of variables. Moreover, we often make use of the vector notation $\vec{\mathcal{Z}}$ as shorthand for finite strings $\mathcal{Z}_1, \ldots, \mathcal{Z}_n$ of expressions whose length is not important or evident from the context.

Suppose now that $\vec{R} = R_1, \ldots, R_n$, $\vec{U} = U_1, \ldots, U_n$ and that $\mathcal{Z}$ is a term or formula of $\mathcal{L}(\sigma)$. Then $\mathcal{Z}[\vec{R}/\vec{S}]$ is the term or formula of $\mathcal{L}(\sigma)$ which is obtained from $\mathcal{Z}$ by simultaneously replacing all free occurrences of the variables $\vec{U}$ by the set terms $\vec{R}$; in order to avoid collision of variables, a renaming of bound variables may be necessary. If the $\mathcal{L}(\sigma)$ formula A is written as $B[\vec{U}]$, then we often simply write $B[\vec{R}]$ instead of $A[\vec{R}/\vec{U}]$. The notations $A[\vec{r}/\vec{u}]$, $B[\vec{r}]$, $A[\vec{R},\vec{r}/\vec{U},\vec{u}]$ and $B[\vec{R},\vec{r}]$ are always used in the corresponding sense.

In addition, if $A[U]$ and $B[a]$ are formulas of $\mathcal{L}(\sigma)$, we often write $A[\lambda x.B[x]]$ to indicate the result, for each number term t with an occurrence in an atomic formula $(t \in U)$ in A, of substituting $B[t]$ for that subformula. As above, it may be necessary to rename bound variables.

A formula or set term of $\mathcal{L}(\sigma)$ is called *positive in X* or *X-positive* in case that it belongs to the collection $POS(X)$. Thus for each X-positive formula $A[X,a]$ a set term $\sigma(X,a)A[X,a]$ is provided by the language $\mathcal{L}(\sigma)$. Within the theory $\mathsf{IPA}(\sigma)$ this set term will define a fixed point of the monotone operator Φ_A provided by $A[X,a]$ and, of course, suitable interpretations of all the σ-expressions occurring within $A[X,a]$,

$$\Phi_A : Pow(\mathbb{N}) \longrightarrow Pow(\mathbb{N}), \quad \Phi_A(M) := \{n \in \mathbb{N} : \mathbb{N} \models A[M,n]\}.$$

To make this idea precise, let us now fix the theory $\mathsf{IPA}(\sigma)$. It is formulated in the language $\mathcal{L}(\sigma)$, its logic is classical logic with equality in the first sort, and its non-logical axioms comprise the axioms of Peano arithmetic PA with the schema of complete induction for all formulas of $\mathcal{L}(\sigma)$ plus the following fixed point axioms.

Fixed point axioms of $\mathsf{IPA}(\sigma)$. For all X-positive formulas $A[X,a]$ of $\mathcal{L}(\sigma)$ and all number terms r we have

(FIX) $$(r \in \sigma(X,a)A[X,a]) \leftrightarrow A[\sigma(X,a)A[X,a], r].$$

It is an immediate and trivial observation that the theories $\widehat{\mathsf{ID}}_1, \widehat{\mathsf{ID}}_2, \ldots$ and $\widehat{\mathsf{ID}}_{<\omega}$ can be canonically embedded into $\mathsf{IPA}(\sigma)$, thus establishing a lower bound of the proof-theoretic strength of $\mathsf{IPA}(\sigma)$. Later we will see that this bound is sharp.

Theorem 4 *There exists a canonical interpretation of $\widehat{\mathsf{ID}}_{<\omega}$ into the theory $\mathsf{IPA}(\sigma)$ so that, for example, all sentences of $\mathcal{L}_1$ are not affected.*

As in $\widehat{\mathsf{ID}}_1, \widehat{\mathsf{ID}}_2, \ldots$ and $\widehat{\mathsf{ID}}_{<\omega}$ we do not claim any extensional behavior of our fixed points in $\mathsf{IPA}(\sigma)$; in particular, we do not include axioms that imply that fixed points are extensional in their parameters. Remember that two sets U and V are considered as equal (in the sense of extensional set theory) provided that they contain the same elements,

$$(U = V) \ := \ \forall x (x \in U \leftrightarrow x \in V),$$

and let $A[U, X, a]$ be an X-positive formula of $\mathcal{L}(\sigma)$. Then in general the implication

$$(S = T) \wedge (r \in \sigma(X, a) A[S, X, x]) \ \rightarrow \ (r \in \sigma(X, a) A[T, X, x])$$

is not provable in $\mathsf{IPA}(\sigma)$. In $\mathsf{IPA}(\sigma)$ fixed points depend on the *syntactic form* of the corresponding fixed point clauses and the set parameters involved; logical relationships between these clauses and the extensional equality of their parameters are not respected. Therefore we call $\mathsf{IPA}(\sigma)$ an *intensional* fixed point theory.

The expressive strength of $\mathsf{IPA}(\sigma)$ stems from the fact that fixed point constructions can be iterated in a very nested way. Choose, for example, two arithmetic formulas $A[X, Y, a]$ and $B[X, Y, a]$ which are both positive in X and negative in Y and set

$$C[Z, b] \ := \ B[Z, \sigma(X, a) A[X, Z, a], b].$$

Then $C[Z, b]$ is positive in Z, and therefore the set term $\sigma(Z, b) C[Z, b]$ may be formed. According to our axioms, $\sigma(X, a) A[X, Y, a]$ can be regarded as an operation $\mathbf{f}_A$ which assigns to each set Y a fixed point $\mathbf{f}_A(Y)$ of the operator defined by $A[X, Y, a]$. This operation is then plugged in into the arithmetic $B[X, Y, a]$ to give the new – no longer arithmetic – formula $C[Z, b]$, which is positive in Z. Therefore it seems that the computation of a fixed point of (the operator corresponding to) this formula does not only depend on individual and previously computed sets, as it is the case for the traditional fixed point theories $\widehat{\mathsf{ID}}_n$ (cf., e.g. Feferman [6]), but on the whole operation $\mathbf{f}_A$.

However, as mentioned above, we will show in this paper that the proof-theoretic strength of the theory $\mathsf{IPA}(\sigma)$, in spite of its great expressive power, is exactly that of the theory $\widehat{\mathsf{ID}}_{<\omega}$, i.e. of the system of all finitely iterated fixed points of positive operator forms.

Lubarsky [8] introduced a μ-calculus $\mathsf{PA}(\mu)$ over Peano arithmetic in which least fixed points rather than arbitrary fixed points are required to exist. This additional requirement yields enormous proof-theoretic strength, and $\mathsf{PA}(\mu)$ is significantly stronger than theories like $\mathsf{ID}_{<\omega}$ or $(\Delta^1_2\text{-}\mathsf{CA}) + (\mathsf{BI})$; for a definition of these systems see, for example, Buchholz, Feferman, Pohlers and Sieg [4]. The theory $\mathsf{PA}(\mu)$ will not be considered further in this article.

4 The auxiliary system $\mathsf{IPA}^-(\sigma)$

$\mathsf{IPA}^-(\sigma)$ is an auxiliary subsystem of $\mathsf{IPA}(\sigma)$ which will be convenient later for determining the upper proof-theoretic bound of $\mathsf{IPA}(\sigma)$: first we show that $\mathsf{IPA}(\sigma)$ can be interpreted into $\mathsf{IPA}^-(\sigma)$; afterwards $\mathsf{IPA}^-(\sigma)$ will be reduced to $\widehat{\mathcal{ID}}_{<\omega}$.

$\mathsf{IPA}^-(\sigma)$ is obtained from $\mathsf{IPA}(\sigma)$ by restricting the formation of σ-terms $\sigma(X,a)A$ to those X-positive formulas A of $\mathcal{L}(\sigma)$ which do not contain number parameters besides a. In more detail, the set terms and formulas of the language $\mathcal{L}^-(\sigma)$, in which $\mathsf{IPA}^-(\sigma)$ has to be formulated, are defined as the set terms and formulas of $\mathcal{L}(\sigma)$ with clause (σ7) replaced as follows:

(σ^-7) If A is an $\mathcal{L}^-(\sigma)$ formula from $POS(X)$ with *at most the free number variable* a and possibly several free set variables besides X, then $\sigma(X,a)A$ is a set term of $\mathcal{L}^-(\sigma)$, in which all free occurrences of the variables X and a are bound by the set term constructor σ. If A belongs to $POS(U)$ [$NEG(U)$], then $\sigma(X,a)A$ belongs to $POS(U)$ [$NEG(U)$].

The set terms and formulas of $\mathcal{L}^-(\sigma)$ which do not contain free set variables are called *semiclosed*. Thus free number variables are permitted in semiclosed set terms and formulas. In closed set terms and formulas neither free number nor free set variables are allowed, but this notion will not be used in the following.

By a simple argument it can be shown that number parameters in σ-terms do not contribute to the expressive and proof-theoretic strength of our fixed point theories. In particular, we have the following reduction result.

Theorem 5 *There exists an interpretation $\mathcal{I}$ which maps each formula A of $\mathcal{L}(\sigma)$ to the formula $\mathcal{I}(A)$ of $\mathcal{L}^-(\sigma)$ so that the following two properties are satisfied:*

1. *For any formula A of $\mathcal{L}(\sigma)$ which does not contain σ-terms its translation $\mathcal{I}(A)$ is identical to A.*

2. *$\mathcal{I}$ reduces the theory $\mathsf{IPA}(\sigma)$ to the theory $\mathsf{IPA}^-(\sigma)$; i.e. we have for all formulas A of $\mathcal{L}(\sigma)$ that*

$$\mathsf{IPA}(\sigma) \vdash A \quad \Longrightarrow \quad \mathsf{IPA}^-(\sigma) \vdash \mathcal{I}(A).$$

Proof For any formula A of $\mathcal{L}(\sigma)$ the formula $\mathcal{I}(A)$ is defined by the following induction on A:

($\mathcal{I}1$) If A is a formula of $\mathcal{L}_1$, then $\mathcal{I}(A)$ is this formula A.

($\mathcal{I}2$) If A is the formula $\neg B$ and not a formula of $\mathcal{L}_1$, then $\mathcal{I}(A)$ is the formula $\neg\mathcal{I}(B)$.

($\mathcal{I}3$) If A is the formula $(B \vee C)$ and not a formula of $\mathcal{L}_1$, then $\mathcal{I}(A)$ is the formula $(\mathcal{I}(B) \vee \mathcal{I}(C))$.

($\mathcal{I}4$) If A is the formula $\exists x B$ and not a formula of $\mathcal{L}_1$, then $\mathcal{I}(A)$ is the formula $\exists x \mathcal{I}(B)$.

($\mathcal{I}5$) Finally let A be a formula of the form $(t \in \sigma(X, a)B[X, \vec{Y}, a, \vec{b}])$ so that $B[X, \vec{Y}, a, \vec{b}]$ contains at most $X, \vec{Y}, a, \vec{b}$ free. Then we set

$$\mathcal{I}(A) \ := \ (\langle t, \vec{b}\rangle \in \sigma(Z, c)B^\circ[Z, \vec{Y}, c])$$

where $B^\circ[Z, \vec{Y}, c]$ is defined to be the formula

$$\exists x \exists \vec{y}(c = \langle x, \vec{y}\rangle \ \wedge \ \mathcal{I}(B)[\lambda z.(\langle z, \vec{y}\rangle \in Z), \vec{Y}, x, \vec{y}]).$$

In view of this definition, the first assertion of our theorem is trivially satisfied. In order to verify the second assertion, we only have to check that all fixed point axioms are respected by our translation $\mathcal{I}$. So let $B[X, \vec{Y}, a, \vec{b}]$ be an X-positive formula of $\mathcal{L}(\sigma)$ which contains at most $X, \vec{Y}, a, \vec{b}$ free. We have to show that the axiom

$$r \in \sigma(X, a)B[X, \vec{Y}, a, \vec{b}] \ \leftrightarrow \ B[\sigma(X, a)B[X, \vec{Y}, a, \vec{b}], \vec{Y}, r, \vec{b}]$$

of $\mathsf{IPA}(\sigma)$ is interpreted appropriately. However, with the notations of ($\mathcal{I}5$) and with $S[\vec{Y}]$ denoting the set term $\sigma(Z, c)B^\circ[Z, \vec{Y}, c]$ we immediately obtain that the equivalence

$$\langle r, \vec{b}\rangle \in S[\vec{Y}] \ \leftrightarrow \ B^\circ[S[\vec{Y}], \vec{Y}, \langle r, \vec{b}\rangle]$$

is provable in $\mathsf{IPA}^-(\sigma)$ and thus also

$$\langle r, \vec{b}\rangle \in S[\vec{Y}] \;\leftrightarrow\; \mathcal{I}(B)[\lambda z.(\langle z, \vec{b}\rangle \in S[\vec{Y}]), \vec{Y}, r, \vec{b}].$$

Hence the projection of the set term $S[\vec{Y}]$ on $\langle \vec{b}\rangle$ takes over the role of the set term $\sigma(X,a)B[X,\vec{Y},a,\vec{b}]$ and provides the required interpretation of the fixed point axiom. □

5 The reduction of $\mathsf{IPA}^-(\sigma)$

In view of Theorem 4 and the previous result, the proof-theoretic analysis of $\mathsf{IPA}(\sigma)$ is completed if we manage to reduce the auxiliary system $\mathsf{IPA}^-(\sigma)$ to the theory $\widehat{\mathsf{ID}}_{<\omega}$. Although we do not encounter great conceptual difficulties in this reduction, it requires a careful bookkeeping with respect to the elimination of σ-terms.

A particular role will be played by specific semiclosed σ-terms – we will call them prime σ-terms later. Interesting examples of prime σ-terms are those of the form $\sigma(X,a)A$ which do not contain semiclosed σ-terms as proper subterms but permit subexpressions of the form $\sigma(Y,b)B$ in which X appears as a free parameter.

To describe our reduction procedure, we need a series of auxiliary notions. First, a *term list* $\mathbb{D}$ *of length* n is defined to be a finite sequence

$$\mathbb{D} \;=\; (S_1, \dots, S_n)$$

of semiclosed set terms of $\mathcal{L}^-(\sigma)$ which satisfies for all i, j from $\{1, \dots, n\}$ with $i \neq j$ the following two properties:

(T1) S_i and S_j are (syntactically) different set terms.

(T2) If S_i is a proper subterm of S_j, then $i < j$.

Secondly, depending on a given term list $\mathbb{D} = (S_1, \dots, S_n)$ we introduce the binary 1-*step reachability relation* $\rhd_{\mathbb{D},1}$ on the components $S_1, \dots, S_n$ of $\mathbb{D}$ by setting

$$S_i \rhd_{\mathbb{D},1} S_j \quad :\Longleftrightarrow \quad \begin{cases} S_i \text{ has the form } \sigma(X,a)A \text{ and there is} \\ \text{a proper subterm } T \text{ of } A \text{ containing } X \\ \text{so that } S_j \text{ is the term } T[S_i/X]. \end{cases}$$

Finally, we say that S_j is *reachable* from a S_i via $\mathbb{D}$ – correspondingly denoted by $S_i \rhd_{\mathbb{D}} S_j$ – in case that the pair (S_i, S_j) belongs to the transitive closure of the relation $\rhd_{\mathbb{D},1}$ of 1-step reachability. We write $S_i \not\rhd_{\mathbb{D}} S_j$ to express that S_j cannot be reached from S_i via $\mathbb{D}$.

Lemma 6 *Let $\mathbb{D}$ be the term list $(S_1, \ldots, S_n)$, and let i, j be two different elements of $\{1, \ldots, n\}$. Then we have:*

1. *If $S_i \rhd_{\mathbb{D}} S_j$, then S_i is a proper subterm of S_j, hence also $i < j$.*
2. *If $S_i \rhd_{\mathbb{D},1} S_j$ and S_i has the form $\sigma(X, a)A[X, a]$, then there exists exactly one proper subterm T of $A[X, a]$, which contains X, so that S_j is the term $T[S_i/X]$; this term T is either X-positive or X-negative. In the first case we say that S_j is* positively *1-step reachable from S_i via $\mathbb{D}$, in the second case that S_j is* negatively *1-step reachable from S_i via $\mathbb{D}$.*

Proof The first assertion of this lemma immediately follows from the definition of 1-step reachability and the fact that $\rhd_{\mathbb{D}}$ is its transitive closure.

The uniqueness of the term T which is claimed in the second assertion is obvious. For the remaining part of the second assertion we only have to exploit the fact that the formula $A[X, a]$ is positive in X and X has to occur in T. $\square$

Let $\mathbb{D}$ be the term list $(S_1, \ldots, S_n)$ and choose two set terms S_i and S_j from $\mathbb{D}$. Then S_j is said to be *positively* reachable from S_i via $\mathbb{D}$ if there exists a sublist $(S_{\ell_1}, \ldots, S_{\ell_m})$ of $\mathbb{D}$ satisfying

$$S_i \rhd_{\mathbb{D},1} S_{\ell_1} \rhd_{\mathbb{D},1} \ldots \rhd_{\mathbb{D},1} S_{\ell_m} \rhd_{\mathbb{D},1} S_j$$

so that the number of negative 1-step reductions in this sequence is even. If there exists such a reduction sequence, leading from S_i to S_j, with an odd number of 1-step reductions, then S_j is *negatively* reachable from S_i via $\mathbb{D}$. The corresponding notations are $S_i \rhd^+_{\mathbb{D}} S_j$ and $S_i \rhd^-_{\mathbb{D}} S_j$, respectively. The following property of these two refined reachability relations should be obvious.

Lemma 7 *Let $\mathbb{D}$ be the term list $(S_1, \ldots, S_n)$, and let i, j be two different elements of $\{1, \ldots, n\}$. If $S_i \rhd^+_{\mathbb{D}} S_j$, then S_i occurs only positively in S_j, and if $S_i \rhd^-_{\mathbb{D}} S_j$, then S_i occurs only negatively in S_j. As a consequence, it is not possible that $S_i \rhd^+_{\mathbb{D}} S_j$ and $S_i \rhd^-_{\mathbb{D}} S_j$.*

In a term list $\mathbb{D} = (S_1, \ldots, S_n)$ those set terms S_i, $1 \leq i \leq n$, which cannot be reached via $\mathbb{D}$ from other terms in $\mathbb{D}$ are now called *prime terms* with respect to $\mathbb{D}$.

Lemma 8 *Let $\mathbb{D}$ be the term list $(S_1, \dots, S_n)$, and let i, j, k be elements of $\{1, \dots, n\}$. Then we have:*

1. *If $S_i \rhd_{\mathbb{D}} S_j$ and if S_k occurs in S_j, then S_k occurs in the term S_i or S_k is the term S_i or $S_i \rhd_{\mathbb{D}} S_k$.*

2. *If S_i and S_j are different prime terms with respect to $\mathbb{D}$, then it cannot be the case that $S_i \rhd_{\mathbb{D}} S_k$ as well as $S_j \rhd_{\mathbb{D}} S_k$.*

Proof Turning to the first assertion, we obtain from $S_i \rhd_{\mathbb{D}} S_j$ that there exist finitely many set terms $S_{\ell_1}, \dots, S_{\ell_m}$, all occurring in the term list $\mathbb{D}$, so that S_{ℓ_1} is the term S_i, S_{ℓ_m} is the term S_j and

$$S_{\ell_1} \rhd_{\mathbb{D},1} S_{\ell_2} \rhd_{\mathbb{D},1} \dots \rhd_{\mathbb{D},1} S_{\ell_{m-1}} \rhd_{\mathbb{D},1} S_{\ell_m}.$$

Now it is sufficient for our first assertion to prove by induction on $m \geq 2$ that, if S_k occurs in S_{ℓ_m}, then either S_k occurs in S_{ℓ_1} or S_k is the term S_{ℓ_1} or $S_{\ell_1} \rhd_{\mathbb{D}} S_k$.

$m = 2$: The term S_{ℓ_1} has the form $\sigma(X, a)A[X, a]$, and there exists a proper subterm R of $A[X, a]$ containing X so that S_{ℓ_2} is the term $R[S_{\ell_1}/X]$. Each subterm S_k of S_{ℓ_2} therefore is a subterm of S_{ℓ_1} or is identical to S_{ℓ_1} or has the form $R_0[S_{\ell_1}/X]$ where R_0 is a subterm of R containing X. Since R_0 is also a subterm of $A[X, a]$, the last case implies that $S_{\ell_1} \rhd_{\mathbb{D},1} S_k$.

$m > 2$: The term $S_{\ell_{m-1}}$ can be written as $\sigma(Y, b)B[Y, b]$, and there exists a proper subterm T of $B[Y, b]$ containing Y so that S_{ℓ_m} is the term $T[S_{\ell_{m-1}}/Y]$. If S_k is a subterm of $S_{\ell_{m-1}}$ or identical to $S_{\ell_{m-1}}$, then our assertion follows from the induction hypothesis. Otherwise, there exists a subterm T_0 of T containing Y so that S_k is the term $T_0[S_{\ell_{m-1}}/Y]$. This T_0 has to be a proper subterm of $B[Y, b]$, and, as a consequence, we have $S_{\ell_{m-1}} \rhd_{\mathbb{D},1} S_k$. This implies $S_{\ell_1} \rhd_{\mathbb{D}} S_k$ and finishes our proof by induction.

To show the second assertion, we can assume $S_i \rhd_{\mathbb{D}} S_k$ and $S_j \rhd_{\mathbb{D}} S_k$; without loss of generality we may also assume that $i < j$. From $S_j \rhd_{\mathbb{D}} S_k$ we conclude with Lemma 6 that S_j is a proper subterm of S_k. Thus $S_i \rhd_{\mathbb{D}} S_k$ together with our first assertion yields that S_j occurs in S_i or S_j is the term S_i or $S_i \rhd_{\mathbb{D}} S_j$. And therefore we arrive at a contradiction with the facts that S_j is prime with respect to $\mathbb{D}$ and $i < j$. Hence it is not possible that $S_i \rhd_{\mathbb{D}} S_k$ and $S_j \rhd_{\mathbb{D}} S_k$. $\square$

We call a term list $\mathbb{D} = (S_1, \dots, S_n)$ an *ordered term list* if it satisfies for all natural numbers i, j, k the following property:

$$1 \leq i < j < k \leq n \text{ and } S_i \rhd_{\mathbb{D}} S_k \quad \Longrightarrow \quad S_i \rhd_{\mathbb{D}} S_j.$$

The previous lemma is now immediately used to show that each term list $\mathbb{D}$ can be permuted into an ordered term list.

Lemma 9 (Ordering lemma) *For every term list $\mathbb{D} = (S_1, \ldots, S_n)$ there exists a permutation π of the set $\{1, \ldots, n\}$ so that $(S_{\pi(1)}, \ldots, S_{\pi(n)})$ is an ordered term list.*

Proof For natural numbers i, j from $\{1, \ldots, n\}$, call j an *irregular $\mathbb{D}$-successor* of i if

- $S_i \rhd_{\mathbb{D}} S_j$,
- there exists a natural number k so that $i < k < j$ and $S_i \not\rhd_{\mathbb{D}} S_k$.

If $\mathbb{D}$ is not an ordered term list, we can choose the least i which has an irregular $\mathbb{D}$-successor. Furthermore, let j be the least irregular $\mathbb{D}$-successor of i. Hence there exists a natural number k so that

$$S_i \rhd_{\mathbb{D}} S_p \quad \text{and} \quad S_i \not\rhd_{\mathbb{D}} S_q$$

for $p = i + 1, \ldots, i + k$ and $q = i + k + 1, \ldots, j - 1$. The first assertion of the previous lemma thus implies that the terms $S_{i+k+1}, \ldots, S_{j-1}$ cannot be subterms of S_j. Therefore the reordering

$$\mathbb{D}' := (S_1, \ldots, S_i, \ldots, S_{i+k}, S_j, S_{i+k+1}, \ldots, S_{j-1}, S_{j+1}, \ldots, S_n)$$

of $\mathbb{D}$ is again a term list. If $\mathbb{D}'$ is an ordered term list, our lemma is already proved; otherwise we iterate the previous procedure until we have obtained an ordered term list. □

To continue our reduction procedure we choose a sequence $\mathcal{P}_1, \mathcal{P}_2, \ldots$ of set constants not occurring in the language $\mathcal{L}^-(\sigma)$. Depending on this sequence we write $\mathcal{L}^-(\sigma, \mathcal{P}_1, \ldots, \mathcal{P}_n)$ for the extension of $\mathcal{L}^-(\sigma)$ by the constants $\mathcal{P}_1, \ldots, \mathcal{P}_n$. Later we will use these set constants to mark specific σ-terms stemming from a given term list.

Given a term list $\mathbb{D} = (S_1, \ldots, S_n)$ and an expression $\mathcal{Z}$ which is either a σ-term or a formula of $\mathcal{L}^-(\sigma)$, we define the stages $\mathcal{Z}(\mathbb{D}, i)$ of the unwinding of $\mathcal{Z}$ via $\mathbb{D}$ by induction on $i \leq n$ as follows:

1. $\mathcal{Z}(\mathbb{D}, 0)$ is the expression $\mathcal{Z}$.
2. If $i > 0$, then we obtain $\mathcal{Z}(\mathbb{D}, i)$ from $\mathcal{Z}(\mathbb{D}, i-1)$ by substituting the set constant $\mathcal{P}_i$ for each occurrence of $S_i(\mathbb{D}, i-1)$ in $\mathcal{Z}(\mathbb{D}, i-1)$.

Clearly the so defined expressions $\mathcal{Z}(\mathbb{D}, i)$ are terms or formulas of the language $\mathcal{L}^-(\sigma, \mathcal{P}_1, \ldots, \mathcal{P}_i)$. If Γ is a (finite) set of $\mathcal{L}^-(\sigma)$ formulas, then we write

$\Gamma(\mathbb{D}, i)$ for the collection of all i-stages of the unwindings of the elements of Γ via $\mathbb{D}$, i.e. for all natural numbers i from $\{0, \ldots, n\}$ we set

$$\Gamma(\mathbb{D}, i) \;:=\; \{A(\mathbb{D}, i) : A \in \Gamma\}.$$

If S is a set term and $\mathcal{Z}$ a set term or a formula of $\mathcal{L}^-(\sigma, \mathcal{P}_1, \ldots, \mathcal{P}_n)$, then $\mathcal{Z}[\mathcal{P}_i \| S]$ denotes the term or formula of $\mathcal{L}^-(\sigma, \mathcal{P}_1, \ldots, \mathcal{P}_n)$ which is obtained from $\mathcal{Z}$ by replacing the set term S by the set constant $\mathcal{P}_i$.

In this unwinding of $\mathcal{Z}$ via $\mathbb{D}$ we replace in the first step all terms S_1 by the constant $\mathcal{P}_1$, and so we may regard S_1 as the "definition" of $\mathcal{P}_1$. In the next step $\mathcal{P}_2$ is introduced and substituted, though not for the terms S_2, but for those expressions which we obtain from S_2 if we replace all occurrences of S_1 within S_2 by $\mathcal{P}_1$; in this sense $S_2[\mathcal{P}_1 \| S_1]$ "defines" $\mathcal{P}_2$. Then we continue with $\mathcal{P}_3, \mathcal{P}_4, \ldots$ accordingly.

Lemma 10 *Let $\mathbb{D}$ be the term list $(S_1, \ldots, S_n)$ and i any natural number from $\{0, \ldots, n\}$.*

1. *If $\mathcal{Y}$ and $\mathcal{Z}$ are terms or formulas of $\mathcal{L}^-(\sigma)$ and if the expressions $\mathcal{Y}(\mathbb{D}, i)$ and $\mathcal{Z}(\mathbb{D}, i)$ are identical, then $\mathcal{Y}$ and $\mathcal{Z}$ are identical.*

2. *If A and R are a formula and a term of $\mathcal{L}^-(\sigma)$, respectively, and if $R(\mathbb{D}, i)$ occurs in $A(\mathbb{D}, i)$, then R is a subterm of A.*

Proof This assertion is proved by induction on $i \leq n$ and obvious for $i = 0$. Now assume that $i > 0$ and that $\mathcal{Y}(\mathbb{D}, i)$ and $\mathcal{Z}(\mathbb{D}, i)$ are identical. By the definition of our unwindings, this implies that the expressions

$$\mathcal{Y}(\mathbb{D}, i{-}1)[\mathcal{P}_i \| S_i(\mathbb{D}, i{-}1)] \quad \text{and} \quad \mathcal{Z}(\mathbb{D}, i{-}1)[\mathcal{P}_i \| S_i(\mathbb{D}, i{-}1)]$$

are identical. Therefore, $\mathcal{Y}(\mathbb{D}, i{-}1)$ and $\mathcal{Z}(\mathbb{D}, i{-}1)$ have to be identical as well, and so the induction hypothesis implies that $\mathcal{Y}$ and $\mathcal{Z}$ are identical. The proof of the second assertion of this lemma can be carried through analogously. $\square$

Lemma 11 (Occurrence lemma) *Let $\mathbb{D} = (S_1, \ldots, S_n)$ be an arbitrary ordered term list and S_k a prime term with respect to $\mathbb{D}$. In addition, assume that the following assumptions are satisfied:*

(A1) For any i, $1 \leq i \leq n$, the term S_i has the form $\sigma(X_i, a_i)A_i[X_i, a_i]$.

(A2) m is a natural number with $k + m \leq n$, and $(S_k, S_{k+1}, \ldots, S_{k+m})$ is the sublist of $\mathbb{D}$ which comprises exactly S_k and all set terms which are reachable from S_k via $\mathbb{D}$.

(A3) For any j, $k \leq j \leq k+m$, it is

$$B_j[X_j, a_j] \;:=\; \begin{cases} A_j[X_j, a_j] & \text{if } j = k \text{ or } S_k \rhd^+_{\mathbb{D}} S_j, \\ \neg A_j[X_j, a_j] & \text{if } S_k \rhd^-_{\mathbb{D}} S_j. \end{cases}$$

Then we have:

1. *It is not possible that a set constant $\mathcal{P}_k, \dots, \mathcal{P}_n$ occurs in one of the formulas*
$$A_1[S_1, a_1](\mathbb{D}, n),\ \dots,\ A_{k-1}[S_{k-1}, a_{k-1}](\mathbb{D}, n).$$

2. *For any natural numbers i, j satisfying $k \leq i, j \leq k+m$ the formula $A_i[S_i, a_i](\mathbb{D}, n)$ behaves with respect to the set constant $\mathcal{P}_j$ so that*
$$\begin{array}{lcl} S_k \rhd^+_{\mathbb{D}} S_i \text{ and } S_k \rhd^+_{\mathbb{D}} S_j & \Longrightarrow & A_i[S_i, a_i](\mathbb{D}, n) \text{ is positive in } \mathcal{P}_j, \\ S_k \rhd^+_{\mathbb{D}} S_i \text{ and } S_k \rhd^-_{\mathbb{D}} S_j & \Longrightarrow & A_i[S_i, a_i](\mathbb{D}, n) \text{ is negative in } \mathcal{P}_j, \\ S_k \rhd^-_{\mathbb{D}} S_i \text{ and } S_k \rhd^+_{\mathbb{D}} S_j & \Longrightarrow & A_i[S_i, a_i](\mathbb{D}, n) \text{ is negative in } \mathcal{P}_j, \\ S_k \rhd^-_{\mathbb{D}} S_i \text{ and } S_k \rhd^-_{\mathbb{D}} S_j & \Longrightarrow & A_i[S_i, a_i](\mathbb{D}, n) \text{ is positive in } \mathcal{P}_j. \end{array}$$

3. *Any set constant $\mathcal{P}_j$, $k \leq j \leq k+m$, occurs positively in all formulas*
$$B_k[S_k, a_k](\mathbb{D}, n),\ \dots,\ B_{k+m}[S_{k+m}, a_{k+m}](\mathbb{D}, n)$$
if $j = k$ or $S_k \rhd^+_{\mathbb{D}} S_j$; otherwise, if $S_k \rhd^-_{\mathbb{D}} S_j$, then the set constant $\mathcal{P}_j$ occurs negatively in all these formulas.

Proof To show the first assertion, we choose arbitrary natural numbers $i \in \{1, \dots, k-1\}$ and $j \in \{k, \dots, n\}$. By applying Lemma 6 we obtain that

(*) S_i is not reachable form S_j via $\mathbb{D}$.

Now assume that $\mathcal{P}_j$ occurs in $A_i[S_i, a_i](\mathbb{D}, n)$. Then the term $S_j(\mathbb{D}, j-1)$, which is replaced by the constant $\mathcal{P}_j$ in the unwinding process, has to be a subterm of the formula $A_i[S_i, a_i](\mathbb{D}, j-1)$. Making use of the previous lemma, this implies that S_j is a subterm of $A_i[S_i, a_i]$.
Recall that $i < j$ and that S_i is the term $\sigma(X_i, a_i)A_i[X_i, a_i]$. Since $\mathbb{D}$ is a term list, the term S_j must not occur in $\sigma(X_i, a_i)A_i[X_i, a_i]$. But this implies that there exists a term T containing X_i so that T is in $A_i[X_i, a_i]$ and $T[S_i/X_i]$ is the term S_j. However, then S_j is reachable from S_i. This is a contradiction to (*), and therefore $\mathcal{P}_j$ cannot occur in $A_i[S_i, a_i](\mathbb{D}, n)$.

Concerning our second assertion, we confine ourselves to proving the third implication; the other three can be treated accordingly. From the assumptions $S_k \rhd^-_{\mathbb{D}} S_i$ and $S_k \rhd^+_{\mathbb{D}} S_j$ we obtain by Lemma 6 and Lemma 7 that S_k is a proper subterm of S_i and S_j and that all occurrences of S_k in S_i are negative and all occurrences of S_i in S_j are positive. Since S_i is the term $\sigma(X_i, a_i)A_i[X_i, a_i]$ and the formula $A_i[X_i, a_i]$ is X_i-positive, $A_i[S_i, a_i]$ contains no positive occurrences of S_k. Consequently, $A_i[S_i, a_i]$ does not have any positive occurrences of S_j.

The constant $\mathcal{P}_j$ is introduced into the formula $A_i[S_i, a_i](\mathbb{D}, n)$ – if at all – at stage j of the unwinding process by substituting $\mathcal{P}_j$ in $A_i[S_i, a_i](\mathbb{D}, j-1)$ for all occurrences of $S_j(\mathbb{D}, j-1)$. But since there are no positive occurrences of S_j in $A_i[S_i, a_i]$, there are no positive occurrences of $S_j(\mathbb{D}, j-1)$ in $A_i[S_i, a_i](\mathbb{D}, j-1)$. This implies that $A_i[S_i, a_i](\mathbb{D}, j)$ does not contain positive occurrences of the constant $\mathcal{P}_j$. In the following transformations, leading from $A_i[S_i, a_i](\mathbb{D}, j)$ to $A_i[S_i, a_i](\mathbb{D}, n)$, the occurrences of $\mathcal{P}_j$ may be only replaced, but not affected otherwise. Hence $\mathcal{P}_j$ does not occur positively in $A_i[S_i, a_i](\mathbb{D}, n)$, i.e. $A_i[S_i, a_i](\mathbb{D}, n)$ is negative in $\mathcal{P}_j$.

The third assertion, finally, follows immediately from the second assertion and the definition of the formulas $B_i[X_i, a_i]$ for $k \le i \le k + m$. □

A further definition is needed: A term list $\mathbb{D}$ of length n is called *complete* for a (finite) set Γ of $\mathcal{L}^-(\sigma)$ formulas if there are no occurrences of σ-terms within the formulas which belong to $\Gamma(\mathbb{D}, n)$.

Given a finite set Γ of semiclosed $\mathcal{L}^-(\sigma)$ formulas, we can now systematically replace all σ-terms by fresh set constants, always choosing those which do not contain other closed σ-terms as proper subterms. Proceeding in this way immediately yields the following lemma.

Lemma 12 *For any finite set Γ of semiclosed $\mathcal{L}^-(\sigma)$ formulas there exists a term list $\mathbb{D}$ of finite length which is complete for Γ.*

Now the stage is set for the reduction of the theory $\mathsf{IPA}^-(\sigma)$ to the system $\widehat{\mathcal{ID}}_{<\omega}$. So take a formula A of $\mathcal{L}^-(\sigma)$ and suppose that there exists a proof Π of A in $\mathsf{IPA}^-(\sigma)$. We carry through the following steps:

R1. We first transform all formulas F in Π into semiclosed formulas $F^\star$ by simply replacing in every F each subformula $(t \in U)$ with a free occurrence of the set variable U by the formula $(t = t)$. Obviously, this transformation yields a proof $\Pi^\star$ in $\mathsf{IPA}^-(\sigma)$ of the formula $A^\star$.

R2. Then we identify the finite set Γ of all fixed point axioms which occur in $\Pi^\star$. So we have finitely many formulas $B_1[X_1, a_1], \ldots, B_m[X_m, a_m]$ of $\mathcal{L}^-(\sigma)$

which are positive in $X_1, \dots, X_m$, respectively, and Γ is the collection of all semiclosed formulas, for $1 \le i \le m$,

$$(\mathrm{FIX}_i) \qquad (r_i \in \sigma(X_i, a_i)B_i[X_i, a_i]) \;\leftrightarrow\; B_i[\sigma(X_i, a_i)B_i[X_i, a_i], r_i].$$

R3. By applying Lemma 12, we know that there exists a term list $\mathbb{D}_\Gamma$ of finite length which is complete for Γ. Because of Lemma 9 we may even assume that $\mathbb{D}_\Gamma$ is an ordered term list. To fix the notation, let $\mathbb{D}_\Gamma$ be the list $(S_1, \dots, S_n)$ with each S_j, for $1 \le j \le n$, being of the form

$$\sigma(X_j, a_j)C_j[X_j, a_j].$$

This implies, clearly, that there is an injection α which tells us for each i, where $1 \le i \le m$, the position of the term $\sigma(X_i, a_i)B_i[X_i, a_i])$ within the list $\mathbb{D}_\Gamma$, i.e. $S_{\alpha(i)}$ is the term $\sigma(X_i, a_i)B_i[X_i, a_i])$.

R4. The next step is to determine all terms which are prime with respect to $\mathbb{D}_\Gamma$. Thus we find natural numbers $\ell_1, \dots \ell_k$, $1 = \ell_1 < \dots < \ell_k \le n$, so that

$$(S_{\ell_1},\, \dots,\, S_{\ell_k})$$

is the sublist of the term list $\mathbb{D}_\Gamma$ comprising exactly the set terms from $\mathbb{D}_\Gamma$ which are prime with respect to $\mathbb{D}_\Gamma$. For purely notational reasons, also define $\ell_{k+1} := n$.

R5. It follows the introduction of the new set constants $\mathcal{P}_1, \dots, \mathcal{P}_n$ and the unwinding of Γ and of all formulas F in $\Pi^\star$ to $\Gamma(\mathbb{D}_\Gamma, n)$ and $F(\mathbb{D}_\Gamma, n)$, respectively, as described above.

R6. We want to set things up to apply Lemma 11 and so let for all natural numbers p with $1 \le p \le k$ and q_p with $\ell_p \le q_p \le \ell_{p+1}$

$$D_{q_p}[X_{q_p}, a_{q_p}] \;:=\; \begin{cases} C_{q_p}[X_{q_p}, a_{q_p}] & \text{if } q_p = \ell_p \text{ or } S_{\ell_p} \rhd^+_{\mathbb{D}} S_{q_p}, \\ \neg C_{q_p}[X_{q_p}, a_{q_p}] & \text{if } S_{\ell_p} \rhd^-_{\mathbb{D}} S_{q_p}. \end{cases}$$

From Lemma 11 we therefore obtain that for all natural numbers p, where $1 \le p \le k$, that

$$\mathbb{S}_p \;:=\; (\, D_{\ell_p}[S_{\ell_p}, a_{\ell_p}](\mathbb{D}_\Gamma, n) \,\dots,\, D_{\ell_{p+1}-1}[S_{\ell_{p+1}-1}, a_{\ell_{p+1}-1}](\mathbb{D}_\Gamma, n)\,)$$

can be considered as an operator system of level $p-1$, as introduced towards the end of Section 2.

R7. We also observe that the nth stage of the unwinding of the fixed point axiom (FIX_i) from Γ is the formula (see reduction step **R3**)

$$(\mathrm{FIX}'_i) \qquad (r_{\alpha(i)} \in \mathcal{P}_{\alpha(i)}) \;\leftrightarrow\; C_{\alpha(i)}[S_{\alpha(i)}, r_{\alpha(i)}](\mathbb{D}_\Gamma, n).$$

Since $\alpha(i)$ belongs to an interval $[\ell_p, \ell_{p+1})$ for a suitable natural number p, the formula in (FIX'_i) is logically equivalent to one of the fixed point axioms generated by the operator system $\mathbb{S}_p$.

R8. As a consequence, the full unwindings of all fixed point axioms which occur in the proof $\Pi^\star$ are derivable in the theory $\widehat{\mathcal{ID}}_k$. Induction along $\Pi^\star$ therefore yields the provability of the unwinded formula $A(\mathbb{D}_\Gamma, n)$ in $\widehat{\mathcal{ID}}_k$.

These reduction steps **R1** – **R8** provide a proof of the desired reduction of $\mathsf{IPA}^-(\sigma)$ to $\widehat{\mathcal{ID}}_{<\omega}$, a consequence of which is stated in the following lemma.

Lemma 13 (Reduction lemma) *For every sentence A of the language $\mathcal{L}_1$ we have that*

$$\mathsf{IPA}^-(\sigma) \vdash A \quad \Longrightarrow \quad \widehat{\mathcal{ID}}_{<\omega} \vdash A.$$

For determining the upper bound of $\mathsf{IPA}(\sigma)$ it only remains to combine this lemma with Theorem 5, concerning the embedability of $\mathsf{IPA}(\sigma)$ into $\mathsf{IPA}^-(\sigma)$, and with Lemma 2, which states that $\widehat{\mathcal{ID}}_{<\omega}$ and $\widehat{\mathsf{ID}}_{<\omega}$ have the same proof-theoretic strength. Since this bound agrees with the lower bound of Theorem 4, the proof-theoretic characterization of $\mathsf{IPA}(\sigma)$ is complete.

Theorem 14 *The theory $\mathsf{IPA}(\sigma)$ is a conservative extension of the system $\widehat{\mathsf{ID}}_{<\omega}$ with respect to all sentences of $\mathcal{L}_1$; in particular, the proof-theoretic ordinal of $\mathsf{IPA}(\sigma)$ is the Feferman-Schütte ordinal Γ_0.*

This finishes our present analysis of intensional fixed point theories. Corresponding theories whose fixed points posses more structure will be studied in a following publication.

References

[1] André Arnold and Damian Niwinski, *Rudiments of μ-Calculus*, Studies in Logic and Foundations of Mathematics, vol. 146, North-Holland, Amsterdam, 2001.

[2] Jeremy Avigad, *On the relationship between* ATR_0 *and* $\widehat{\mathsf{ID}}_{<\omega}$, Journal of Symbolic Logic **61** (1996), no. 3, 768–779.

[3] Julian Bradfield and Colin Stirling, *Modal logics and mu-calculi: an introduction*, Handbook of Process Algebra (Jan A. Bergstra, Alban Ponse, and Scott A. Smolka, eds.), Elsevier, Amsterdam, 2001, pp. 293–332.

[4] Wilfried Buchholz, Solomon Feferman, Wolfram Pohlers, and Wilfried Sieg, *Iterated Inductive Definitions and Subsystems of Analysis: Recent Proof-Theoretical Studies*, Lecture Notes in Mathematics, vol. 897, Springer, Berlin, 1981.

[5] Andrea Cantini, *Logical Frameworks for Truth and Abstraction*, North-Holland, Amsterdam, 1996.

[6] Solomon Feferman, *Iterated inductive fixed-point theories: application to Hancock's conjecture*, The Patras Symposion (G. Metakides, ed.), North-Holland, Amsterdam, 1982, pp. 171–196.

[7] Gerhard Jäger, Reinhard Kahle, Anton Setzer, and Thomas Strahm, *The proof-theoretic analysis of transfinitely iterated fixed point theories*, Journal of Symbolic Logic **64** (1999), no. 1, 53–67.

[8] Robert S. Lubarsky, *μ-definable sets of integers*, Journal of Symbolic Logic **58** (1993), no. 1, 291–313.

Address
Gerhard Jäger
Institut für Informatik und angewandte Mathematik
Universität Bern
Neubrückstrasse 10, CH-3012 Bern, Switzerland
`jaeger@iam.unibe.ch`