DIMACS Technical Report 96-49
November 1996

Ranking Arithmetic Proofs by Implicit Ramification[1]

by

Stephen J. Bellantoni

[1]Much of this work was carried out while the author was a visiting researcher at DIMACS, the Center for Discrete Mathematics and Theoretical Computer Science.

DIMACS is a partnership of Rutgers University, Princeton University, AT&T Research, Bellcore, and Bell Laboratories.

DIMACS is an NSF Science and Technology Center, funded under contract STC–91–19999; and also receives support from the New Jersey Commission on Science and Technology.

ABSTRACT

Proofs in an arithmetic system are ranked according to a ramification hierarchy based on occurrences of induction. It is shown that this ranking of proofs corresponds exactly to a natural ranking of the primitive recursive functions based on occurrences of recursion. A function is provably convergent using a rank r proof, if and only if it is a rank r function. The result is of interest to complexity theorists, since rank one corresponds to polynomial time. Remarkably, this characterization of polynomial-time provability admits induction over formulas having arbitrary quantifier complexity.

1 Introduction

The primitive recursive functions can be assigned ranks, based on an examination of the structure of their derivations as built up from the initial functions by the rules of composition and recursion. One of the hierarchies defined using such a ranking consists of the polynomial-time computable functions at level 1, and at higher levels consists of certain of the Grzegorczyk classes [4]. The current work shows how to rank arithmetic proofs containing inductions, such that the ranking of proofs corresponds exactly to this ranking of recursions. It is shown that a function can be proven convergent by a rank r proof if and only if it is a rank r function.

The ranking of proofs is entirely syntactic, without any obvious reference to computation time, the size of quantified values, etc. The ranking of proofs is not a disguise for the ranking of function symbols appearing in the proofs. Remarkably, induction formulas are allowed to have arbitrary alternations of unbounded quantifiers. Yet the first level of the logical hierarchy proves convergence of exactly the polynomial time computable functions. Thus, the fragment of arithmetic consisting of rank 1 proofs, passes a key test for being considered a "feasible" system.

Compared to the bounded induction rules of earlier weak subsystems of arithmetic, the present definitions permit a radical reduction in the syntactic restrictions placed on the induction formula. The only remaining restriction is a relatively mild one concerning occurrences of 0. In exchange, the overall amount of induction in the proof is restricted, by measuring the rank of the proof.

The fragment of arithmetic consisting of rank 1 proofs is not closed under modus ponens. On the one hand, this certainly seems awkward from the viewpoint of traditional approaches to logic and earlier approaches to feasible arithmetic. On the other hand, it is also precisely the kind of feature described by Fagin et. al. as desirable for reasoning about knowledge using limited computational resources [8] p. 309. The present work refers to modal concepts by reading "$x = y$" as "x is (feasibly, constructively) known to be equal to y". The logical connectives and quantifiers are treated classically. Proofs are trees of formulas with axioms at the leaves and instances of either modus ponens or induction at the interior nodes.

Modal concepts are incident on the idea of *ramification* that is central to this work. Generally, a ramified mathematical system is one in which objects are defined using levels such that the definition of an object in level i refers only to objects in levels strictly below i. Such definitions are "predicative" because they do not contain any circularity of reference. A prototypical example occurs in ramified set theory, where the universe of sets at level i is defined by comprehension over sets in the previous universes. In the present work, an 'object' is an integer together with an *depth of understanding* (or: amount of knowledge, strength of comprehension) of that integer. Ramification level r ($0 \leq r \leq \omega$) consists of the integers as known with a depth of understanding at least r. Each input position i of each primitive recursive function f is labeled with a 'rank' number $\rho(f, i)$ that, intuitively, tells how much understanding of the ith input is required in order to have any understanding at all of the output [4]. One defines $\rho(f) = \max_i \rho(f, i)$. Each variable x in each proof Π is labeled with a 'rank' number $\rho(\Pi, \ulcorner x \urcorner)$, that tells how much understanding is required of the integers that x ranges over, in order to understand the proof (i.e. to understand the truth of each step of the proof).[1] Then one defines $\rho(\Pi) = \max_{\ulcorner x \urcorner} \rho(\Pi, \ulcorner x \urcorner)$. Induction is a principle that may increase the rank of the induction variable. Although each step of the induction can be understood using only the depth of understanding required for the antecedents $\alpha[0]$ and $\alpha[x] \supset \alpha[x+1]$, the *completion* of the induction typically requires a greater amount of knowledge. Thus, induction is a principle that generates the next ramification level.

Ramification levels are analogous to the 'tiers' that Leivant defined in order to control impredicativity [13]. In Leivant's work, the strength of the logic is constrained by the explicit syntactic presence of special unary predicates. Indeed, from the presence of modal concepts in the present work one might expect to have explicit unary predicates $K_i x$ in the logic indicating that "x is understood to a depth of i". However, this is not the approach taken here. The "depth of understanding about the integers that x ranges over, required in order to understand the proof" is not something that has to be manifested within the proof itself. It can be viewed as an aspect of the proof that we are only interested in when performing an analysis of the proof. The number of ramification levels required to understand the proof, is implicit in the structure of the proof — specifically in the layering of induction instances. The purpose of this paper is to measure this implicit ramification in the syntax, using $\rho(\Pi)$, and to relate this measure to computational complexity.

The results of this paper are obtained by defining a *standard ramified model* which contains ω many copies of the integers $\mathbf{N}^0, \mathbf{N}^1, \ldots, \mathbf{N}^\omega$, each copy indicating a different depth of understanding of the integers. Let $\mathbf{K} = \cup_i \mathbf{N}^i$; the universe of the model is $\mathbf{K} \cup \{\bot\}$. Defining $\mathbf{U}^r = \cup_{i \geq r} \mathbf{N}^i \cup \{\bot\}$, the subuniverse $\mathbf{U}^r \cap \mathbf{K}$ corresponds to ramification level r: the integers as known with a depth of understanding at least r. The model is designed to satisfy: $\mathbf{f}(\vec{\mathbf{x}}) = \bot$ iff $\exists i, \mathbf{x}_i \notin \mathbf{U}^{\rho(f,i)} \cap \mathbf{K}$. This formalizes the intuition that one must have at least a depth $\rho(f, i)$ of understanding of the ith input, in order to know the output. The main result shows that the conclusion of a given proof Π holds in the standard ramified model when each variable x is relativized to $\mathbf{U}^{\rho(\Pi, \ulcorner x \urcorner)}$. This formalizes the intuition that the proof can be understood by having a depth of understanding only $\rho(\Pi, \ulcorner x \urcorner)$ of the integers that x ranges over. That is, proof Π involves a depth of understanding at most $\rho(\Pi)$. On the other hand, f requires an depth of understanding at least $\rho(f)$. By a simple semantic argument, it will follow that if Π proves the convergence of f, then $\rho(\Pi) \geq \rho(f)$. Therefore, low rank proofs have limited computational complexity: they cannot prove the convergence of high-rank functions.

The subuniverses $\mathbf{U}^r$ seem to be analogous to successive worlds of an intuitionistic model. However, their existence is not presupposed by the definition of the semantics itself. Indeed, the semantics are first-order semantics, and ordinary first-order models of arithmetic are models of the system used here. Ramification levels are implicit in an arithmetic proof, not explicitly required by the semantics.

[1] "The integers that x ranges over" means the integers *qua* what can be denoted by x, not simply the integers in an extensional sense. For two different variables x and y, "the integers that x ranges over" are understood differently — and to a different amount — than "the integers that y ranges over".

The definition of $\rho(\Pi, \ulcorner x \urcorner)$ proceeds using an analysis of proofs, including an analysis of modus ponens and $\supset$, that is entirely type level one. This contrasts with all the proof analysis methods that are based on the Curry-Howard isomorphism of formulas as types, using higher-type functionals. The present analysis works by defining a directed weighted graph, the "contribution graph" $G(\Pi)$ of the proof, which contains information about occurrences of $\supset$ in the proof. The vertices of the graph are (code numbers for) the variables of the language, plus certain extra vertices, and the graph consists of a set of directed edges weighted 0 or 1. The rank of a variable is the maximum weight of paths ending at that variable in the graph. Consider a modus ponens instance "α, $\alpha \supset \beta \mid \beta$". The derivation of $\alpha \supset \beta$ provides a mapping from the ranks of variables in the derivation of α, to the ranks of variables in the proof concluded by β. The mapping from the rank of one variable to the rank of another is specified by the maximum weight path from one to the other in the graph. The composition corresponding to a modus ponens rule is achieved simply by taking the union of the two graphs corresponding to the subproofs. Repeated applications of modus ponens leads to the repeated composition of mappings, not to higher-type abstraction and application. Although the specific definition of $G(\Pi)$ is based on the specific proof system used here, similar constructs should be definable for other proof systems. As Prawitz noted, "modus ponens is present also in normal derivations in the systems of natural deduction... it is really also present in cut-free derivations in the calculus of sequents... " ([18], p. 261).

The edges in $G(\Pi)$ are defined in such a way that a rank $r < \omega$ proof must contain at least r "nested" inductions. On the other hand, not every induction has significantly more computational content than was already in the subproofs. In a computationally significant use of induction, the subproofs leading up to the induction premisses add to earlier subproofs in a significant way. For example, an induction might have significant new computational content if its premisses are obtained by applying the conclusion of an earlier induction to a defining axiom for a function. This idea will be incorporated into the definition of $G(\Pi)$, via the choice of edges and weights. Proof rank, defined using the weights of paths in $G(\Pi)$, is a measure of those occurrences of induction that are interleaved with uses of primitive arithmetic facts. Function rank, in comparison, is a measure of the occurrences of recursion in the function derivation. The two correspond: a function has rank r, if and only if it can be proved convergent by a rank r proof. It is important to realize that proof rank is not a disguise for the rank of function symbols appearing in the proof. For example, a low-rank proof can use the defining axioms for a high-rank function.

Intuitively speaking, a low-rank proof cannot fully analyze a high-rank function, simply because there are not enough steps in the unrolled version of the proof to cover all of the steps in the unrolled computation of the function (by "unrolled" I mean, very loosely, the replacement of induction or recursion by a sequence of steps). For example, to prove the convergence of a function that is defined by a nontrivial recursion, it seems that one has to use at least one instance of induction. A proof by modus ponens alone won't be able to analyze the recursive function. This seems to be true even if one allows quantifier reasoning in the proof by modus ponens. We would like to generalize this idea to functions defined by various ranks of recursion and proofs having corresponding ranks of induction. As it has been shown that these ranks of recursion characterize various function classes with limited computational complexity [4], one expects the corresponding ranks of induction to characterize various proof systems having limited computational complexity. Remarkably, the result holds even when quantifiers are unrestricted in the induction formulas.

Why is it that the presence of quantifiers in an induction formula does not help a low-rank proof to analyze a high-rank function? I can think of two possible intuitions to hazard in response to this question. First, one can speculate that quantifiers in a *thoroughly* constructive and feasible system *should* be eliminable in favor of free variables and feasibly constructive type 1 Skolem functions. Stated differently, the natural state of affairs in a feasible system is to use feasible Skolem functions, not existential quantifiers. Alternating quantifiers are introduced when one throws away the definitions of the Skolem functions. From this perspective, the use of alternating quantifiers in an induction formula should not represent a computationally powerful new principle. On the contrary, the use of existential quantifiers actually represents a loss of information, since the definitions of the witnessing Skolem functions are no longer known.

A second conceivable response is to note that quantifiers in a classical arithmetic proof arise either from first-order rules or from induction rules. The classical first-order quantifier manipulations permit one to keep track of the scope of free variables, e.g. when they are in negative subformulas. The quantifiers introduced by these rules do not seem to me to have a lot of computational content: they mark the scope of variables, not the presence of a computation. On the other hand, a quantifier that is introduced in the conclusion of induction, is a notation indicating that the induction is *complete*. Such a quantifier marks a ramification step, and hence is a mark of computational content. But the computational content marked by such a quantifier is just whatever was obtained during the induction. Thus, the quantifier symbol itself is peripheral to the main issue, which is the occurrence of induction. Perhaps limiting the presence of such quantifiers in induction formulas just serves indirectly to restrict the layering of induction, by limiting induction formulas that carry the mark of a preceeding induction. I propose instead to directly restrict the layering of inductions.

A review of earlier work in weak subsystems of arithmetic is given by Krajicek [10]. For relevant background in recursion theory and computational complexity, see Clote's survey [7].

2 Ranking Recursions

The development begins by defining a recursive function class which is equivalent to the class of primitive recursive functions. A function derivation is a tree structure with leaves labeled by initial functions and internal nodes appropriately labeled by functions defined using one of the derivation rules. I frequently speak of a derivation interchangeably with the corresponding function, although of course each derivable function has many derivations. Write m_f for the arity of function f. The class $\mathcal{PR}_2$ is the class of derivations defined by:

- Initial functions: constant 0; successors $s_0x = 2x$ and $s_1x = 2x+1$; predecessor $p(x) = \lfloor x/2 \rfloor$; and conditional $c(x,y,z) =$ "if $x \bmod 2 = 0$ then y else z".
- Full composition: given h and $g_1, \ldots, g_m$ with $m = m_h \geq 1$, define f by $f(\vec{x}) = h(g_1(\vec{x}^1), \ldots, g_m(\vec{x}^m))$, where each $\vec{x}^j$ is a vector consisting of elements from the vector $\vec{x}$ (possibly with repetitions, omissions, or changes in order).
- Full recursion on notation: given g and h define f by $f(0,\vec{y}) = g(\vec{y})$ and $f(x,\vec{y}) = h(x,\vec{y}, f(px,\vec{y}))$ for $x \neq 0$.

In the recursion scheme: position 1 in f is the *recursion position*; position m_h in h is the *critical position*; the *step* function is h.

To rank all of these functions, one assigns a *rank*, $\rho(f,i)$, for each derivation $f \in \mathcal{PR}_2$ and each position $1 \leq i \leq m_f$. The rank is supposed to be a measure of how hard it is to understand the function's use of a given input. One can think of rank as being in some sense a measure of how "opaque" an input position is. In the logical setting, substitutivity of equals will fail, with the ranks of input positions being related to the amount of induction that is required to know the output of the function.

In defining $\rho(f)$, one uses the idea that when a function f is derived from subfunctions $\vec{h}$, then there is some sense in which the rank of input positions of $\vec{h}$ contribute to the rank of input positions of f. For example, if $f(x) = h(0, g(x))$, then one requires $\rho(f,1) \geq \rho(h,2)$ and $\rho(f,1) \geq \rho(g,1)$. The key point is that recursion makes it strictly more difficult to understand how a value is used. If $f(x) = h(x, f(px))$, then one requires $\rho(f,1) \geq 1 + \rho(h,2)$ as well as $\rho(f,1) \geq \rho(h,1)$. The definition of $\rho(f,i)$ is:

- If f is an initial function, then $\rho(f,i) = 0$ for $1 \leq i \leq m_f$.
- If f is defined by the composition $h(g_1(\vec{x}^1), \ldots, g_m(\vec{x}^m))$, then
$\rho(f,i) = \max(\{\rho(h,j) : \ x_i \in \vec{x}^j\} \cup \{\rho(g_j,k) : \ x_i \equiv (\vec{x}^j)_k\} \cup \{0\})$ for $1 \leq i \leq m_f$.
- If f is defined by recursion from step function h and base function g, then
$\rho(f,1) = \max\{\rho(h,1),\ 1+\rho(h,m_h)\}$ and
$\rho(f,i) = \max\{\rho(h,i),\ \rho(h,m_h),\ \rho(g,i-1)\}$ for $2 \leq i \leq m_f$.

Define the *rank* of derivations by: $\rho(f) = \max(\{\rho(f,i) : \ 1 \leq i \leq m_f\} \cup \{0\})$.

Define $\mathcal{PR}_2^r = \{f \in \mathcal{PR}_2 : \ \rho(f) \leq r\}$.

Rank differs from "tiers" (Leivant [14]) because rank is not a type system. This is not merely a syntactic distinction, as the present system admits more instances of composition and recursion than the stricter system in [14]. In particular, $\mathcal{PR}_2$ includes all the primitive recursive functions and separates at all levels, while the system in [14] collapses to level 2.

Theorem 2.1 (Bellantoni & Cook [5]) *$\mathcal{PR}_2^1$ consists of exactly the polynomial time computable functions.*

Theorem 2.2 (Bellantoni [4]) *For $r \geq 2$, $\mathcal{PR}_2^r = \mathcal{E}^{r+1}$, the Grzegorczyk class of level $r+1$. In particular, $\mathcal{PR}_2^2 =$ the Kalmar elementary functions.*

3 Synthetic Arithmetic

Now a proof system is developed that is suitable for counting the rank of inductions. I would like to be able to admit a function symbol for every primitive recursive function, without having convergence of all functions proved trivially by existential generalization from reflexivity instances $f(\vec{x}) = f(\vec{x})$. To accomplish this, I view '=' and '$\neq$' as relations that exclude each other but which do not satisfy $x = y \vee x \neq y$. These relations can be read constructively: "$x = y$" is "x is known to be the same as y" or "there is a constructive demonstration that x is the same as y", and "$x \neq y$" is similarly read as "x is known to be different from y" or "there is a constructive demonstration that x is different from y". Other than '=' and '$\neq$', the logic is entirely classical; and classical semantics are used. This approach is in some sense opposite to intuitionistic logic, in which '$\supset$' and other connectives are given a constructive semantics, and in which reflexivity and substitutivity of '=' are treated as analytic truths. Compared to free logic, the current approach has the advantage of classical quantifier rules together with correspondingly simpler semantics.

Let Kt be the formula $\exists w, (w = t \vee w \neq t)$, where w is the first variable not appearing in term t. Under the reading above, Kt asserts that there is something which is known to be equal to or known to be different from t. Shortening this phrase a little bit, "Kt" is read as "t is known". Convergence statements are statements of the form, $\forall \vec{x} \in K, \ Kf(\vec{x})$ (that is, $\forall \vec{x}, ((\wedge_i Kx_i) \supset Kf(\vec{x})))$.

For lack of a better word I call the system "synthetic", in respect of the synthetic (constructive, non-analytic) reading of "$t = t$". There already is such a thing as "epistemic" arithmetic; see Goodman [9] and Shapiro [20]. A constructibility predicate Kx was defined by Lifschitz in his contribution to Shapiro's volume [19]; the semantic meaning is given by a form of realizability. A special predicate letter is used to delimit the numbers in Peano's original formulation. Leivant extends this to a series of special predicate letters corresponding to ramification levels [13]. A typical constructive approach to non-convergence of functions is given in Beeson's logic of partial terms [1] p. 97. For free logic one can start with Lambert's collection [11], particularly the contributions by Garson (p. 111) and Posy (p. 49). The present development is, as far as I know, original.

Let $\mathcal{SFO}(F)$ (*synthetic first-order logic over F*) be the logic having the following axioms and rules, defined over the language with function symbols F, relation symbols '=' and '$\neq$', and logical connectives $\supset$, $\neg$, $\forall$. The other connectives are defined classically. Proofs are trees of formulas. Uniform substitution of term t for variable x in α is indicated by $\alpha[t/x]$. To improve readability I insert a comma after quantifiers; and I use abbreviations such as "$\forall x \in K, \ \alpha$" for "$\forall x, (Kx \supset \alpha)$". In the (FO) axioms below, $\vec{z}$ is a list consisting of some or all of the free variables in the subsequent formula. In (KS) below, $\vec{a}$ and $\vec{b}$ are terms with free variables $\vec{z}$.

The (FO) axioms are the axioms of classical first-order predicate calculus without equality (adapted from [16]). The other axioms are modifications of the usual axioms for equality. These are meant to formalize, first of all, the idea that $x = y \vee x \neq y$ is not valid; and second of all, that deductions involving $=$ and $\neq$ can be understood as ones involving constructive knowledge. For example, if one knows that $a = b$, and one knows what $f(a)$ and $f(b)$ are, then in fact one knows that $f(a) = f(b)$. But simply from knowing $a = b$, one does not know $f(a) = f(b)$: function inputs are opaque contexts.

$$
\begin{array}{ll}
\text{(FO.1)} & \forall\vec{z},\ (\alpha \supset (\beta \supset \alpha)) \\
\text{(FO.2)} & \forall\vec{z},\ ((\alpha \supset (\beta \supset \gamma)) \supset ((\alpha \supset \beta) \supset (\alpha \supset \gamma))) \\
\text{(FO.3)} & \forall\vec{z},\ ((\neg\alpha \supset \neg\beta) \supset (\beta \supset \alpha)) \\
\text{(FO.4)} & \forall\vec{z},\ ((\forall x,\ (\alpha \supset \beta)) \supset ((\forall x,\ \alpha) \supset (\forall x,\ \beta))) \\
\text{(FO.5)} & \forall\vec{z},\ ((\forall x,\ \alpha) \supset \alpha[t/x]) \ \ (t \text{ is free for } x \text{ in } \alpha) \\
\text{(FO.6)} & \forall\vec{z},\ (\alpha \supset \forall y,\ \alpha) \ \ (y \text{ is not free in } \alpha)
\end{array}
$$

$$
\begin{array}{ll}
\text{(E)} & \forall x, y,\ (x = y \supset \neg(x \neq y)) \\
(\neg\text{E}) & \forall x, y,\ (x \neq y \supset \neg(x = y)) \\
\text{(KR=)} & \forall x,\ (Kx \supset x = x) \\
\text{(S=)} & \forall x, y,\ (x = y \supset y = x) \\
(\neg\text{S=}) & \forall x, y,\ (x \neq y \supset y \neq x) \\
\text{(T=)} & \forall x, y, z,\ ((x = y \wedge y = z) \supset x = z) \\
(\neg\text{T=}) & \forall x, y, z,\ ((x = y \wedge y \neq z) \supset x \neq z) \\
\text{(KS)} & \forall\vec{z},\ ((\vec{a} = \vec{b} \wedge Kf(\vec{a}) \wedge Kf(\vec{b})) \supset f(\vec{a}) = f(\vec{b}))
\end{array}
$$

The only deduction rule of $\mathcal{SFO}(F)$ is (MP): $\alpha,\ \alpha \supset \beta \mid \beta$.

A very convenient feature of $\mathcal{SFO}(F)$ is that models of $\mathcal{SFO}(F)$ are just first-order models in which the interpretations of '=' and '≠', and the interpretation of the function symbols, satisfy the second group of axioms above. The system $\mathcal{SFO}(F)$ is complete for these models because $\mathcal{SFO}(F)$ is an extension of first-order logic without equality (FO.1-6 above). Of course, models of $\mathcal{SFO}(F)$ may interpret '=' in a non-normal way i.e. $\mathbf{x} = \mathbf{y}$ with $\mathbf{x}$ and $\mathbf{y}$ being different points in the model.

For the purposes of this work, it is appropriate to include extra copies of the function symbol for the initial function 0. Zero is special because it appears explicitly in the induction rule. Let F_2 consist of one function symbol for each $\mathcal{PR}_2$ derivation, except that for the initial function 0 put countably many function symbols, $\{0_i : 0 \le i < \omega\}$ into F_2. (Throughout, ω is the first limit ordinal.) One could use only a single symbol for zero if one were willing to accept that (ind) be restricted to induction over 0-free formulas; see the definition of (ind) below.

A base system $\mathcal{SQ}_2$ (*synthetic Q in base 2*) is defined by adding the following axioms to $\mathcal{SFO}(F_2)$, for $0 \le i, j \le \omega$. First are axioms for 0, s, p, and c:

$$
\begin{array}{llll}
\text{(0.1)} & 0_i = 0_j & \text{(s.6)} & \forall x, y \in K,\ (s_0 x = s_0 y \supset x = y) \\
\text{(s.1)} & s_0 0_i = 0_j & \text{(s.7)} & \forall x, y \in K,\ (s_1 x = s_1 y \supset x = y) \\
\text{(s.2)} & \forall x \in K,\ (s_1 x \neq 0_i) & \text{(p.1)} & p0_i = 0_j \\
\text{(s.3)} & \forall x \in K,\ (s_0 x \neq s_1 0_i) & \text{(p.2)} & \forall x \in K,\ (ps_1 x = x) \\
\text{(s.4)} & \forall x \in K,\ (s_0 x \neq s_1 x) & \text{(p.3)} & \forall x \in K,\ (ps_0 x = x) \\
\text{(s.5)} & \forall x \in K,\ (s_0 x \neq s_1 x) & \text{(p.4)} & \forall x \in K,\ (x = s_0 px \vee x = s_1 px) \\
 & & \text{(c.1)} & \forall x, y, z \in K,\ (x = s_0 px \supset c(x, y, z) = y) \\
 & & \text{(c.2)} & \forall x, y, z \in K,\ (x = s_1 px \supset c(x, y, z) = z)
\end{array}
$$

The remaining axioms are for f defined by composition or recursion. They are paraphrased below.

(comp) $(\forall\vec{x}\in K,\ Kh(g_1(\vec{x}^1),\ldots,g_m(\vec{x}^m)))$
$\supset(\forall\vec{x}\in K,\ (f(\vec{x})=h(g_1(\vec{x}^1),\ldots,g_m(\vec{x}^m))))$
where f is defined by composition from h and $g_1,\ldots,g_m$.

(rec.1) $((\forall\vec{y}\in K,\ Kg(\vec{y}))\wedge(\forall x,\vec{y},z\in K,\ Kh(x,\vec{y},z)))$
$\supset(\forall\vec{y}\in K,\ f(0_i,\vec{y})=g(\vec{y}))$
where f is defined by recursion from h and g.

(rec.2) $((\forall\vec{y}\in K,\ Kg(\vec{y}))\wedge(\forall x,\vec{y},z\in K,\ Kh(x,\vec{y},z)))$
$\supset(\forall x,\vec{y}\in K,\ (x\neq 0_j\supset(Kf(px,\vec{y})\supset f(x,\vec{y})=h(x,\vec{y},f(px,\vec{y})))))$
where f is defined by recursion from h and g.

The understanding of (comp) is: if one knows what $h(g_1(\vec{x}^1),\ldots,g_m(\vec{x}^m))$ is for all known $\vec{x}$, then one knows that $f(\vec{x})=h(g_1(\vec{x}^1),\ldots,g_m(\vec{x}^m))$ for all known $\vec{x}$. For (rec.1): if the base function, g, is known, and the step function, h, is known, then for all known $\vec{y}$, one knows that $f(0,\vec{y})$ is equal to $g(\vec{y})$. For (rec.2): if the base function g is known, and the step function h is known, then for all values x that are known to be different from 0, and for all known $\vec{y}$, if one knows $f(px,\vec{y})$ then one knows that $f(x,\vec{y})$ equals $h(x,\vec{y},f(px,\vec{y}))$.

Note. In the recursion axiom (rec.1), the literal value 0_i is specified in the base term $f(0_i,\vec{y})$. Since models may be non-normal, this formulation does not constrain f on all x such that $x=0_i$. Definition: a *shadow of zero* is a point $\mathbf{x}$ in a model such that $\mathbf{x}=\mathbf{0}_i$ in the model, for some i. By (0.1) it doesn't matter which i one uses. On the syntactic side, the various function symbols 0_i, $i\geq 0$ are *variant names for* 0. By (0.1), every variant name for 0 denotes a shadow of 0. A shadow of 0 may be denoted by more than one variant name, and there may be shadows of 0 that are not denoted by any variant name. The use of variant names for zero does not constitute an explicit reference to ramification levels. The variant names of zero are interchangeable in the axioms. They are treated symmetrically rather than ordered into successive layers.

Finally one obtains $\mathcal{SA}_2$ (*synthetic arithmetic in base 2*) by adding an induction rule to $\mathcal{SQ}_2$. In this rule, one requires as usual that x and z do not already appear in α; furthermore, *it is required that* 0_i *does not already appear in* α. The (ind) rule is:

$$\frac{\alpha[0_i/y]\qquad\forall x\in K,\ (x\neq 0_j\supset(\alpha[px/y]\supset\alpha[x/y]))}{\forall z\in K,\ \alpha[z/y]}$$

Other than the restriction concerning 0_i, x and z, the induction formula can be arbitrary; for example, it can contain any number of unbounded alternating quantifiers.

Note. Of considerable interest in the induction rule is the fact that the conclusion of induction says that $\alpha[z/y]$ holds at all $z\in K$; in particular this includes all the shadows of zero. Thus, a use of induction constrains the truth value of α at all shadows of 0, even though in the premisses one is only required to have specified the truth value of α at one of the shadows of zero, namely 0_i. Compare with (rec.1). It is only through occurrences of induction that the value of a recursively defined function is constrained on all shadows of zero. This seems analogous to the situation in nonstandard model theory, where functions are only required to take on "correct" values at nonstandard numbers by the fact that induction must be satisfied.

The restriction that 0_i does not already appear in α, is mitigated by the fact that one can use other variant names 0_j $(j \neq i)$ in α. In fact this was the purpose of introducing variant names for zero; the alternative would be to have only one symbol for 0, and not allow it into the induction formula $\alpha[y]$. The variant names for zero are equivalent in the sense that they are all treated in the same way by the axioms. The point, then, is that the variant name 0_i used to form the base case of the induction has in some sense a kind of independence from all the other names in α: there is no reason why 0_i has to denote identically the same element as 0_j, $j \neq i$. In other words it is the axiom system (0.1, rec, etc), rather than just a coincidence of names, that determines the relationship between, on the one hand, 'zero used as the base value for the induction variable' and, on the other hand, 'zero used to refer to a fixed element of the universe independent of the value of the induction variable'. This use of 0_i and 0_j contrasts with the ordinary unramified formulation in which such a distinction cannot be made.

4 Ranking Proofs

Now a rank, $\rho(\Pi)$, is assigned for each proof Π. But before defining $\rho(\Pi)$, something should be said about fairness. One trivial way to have rank r provability correspond to rank r functions, would be to let the rank of a proof be the maximum of the ranks of the function symbols appearing in it. But this would be cheating, because then rank r proofs essentially could only use the definitions of rank r functions. Obviously such a system could not prove the convergence statement for a function symbol outside of $\mathcal{PR}_2^r$. One should exclude from consideration any definition of $\rho(\Pi)$ which refers to $\rho(f)$, or in which $\rho(\Pi)$ is known to be large simply due the use of $\mathcal{SQ}_2$ axioms for a lot of functions of high rank. To be fair, the use of an $\mathcal{SQ}_2$ defining axiom for a function should cost nothing in terms of the rank of the proof. This does not mean that the statement of the definition of $\rho(\Pi)$ has to be independent of how and where composition and recursion axioms are used; only that uses of composition and recursion axioms should not increase $\rho(\Pi)$, either by themselves or in combination with (MP) and the logical axioms. In other words, every $\mathcal{SQ}_2$ proof should have rank 0. Thus, induction must be critical to the definition of rank.

A graph $G(\Pi)$, the *constraint graph of* Π, is used to define $\rho(\Pi)$. The constraint graph represents auxiliary information about the proof, namely certain constraints that ρ must satisfy. Let V be the set of code numbers $\{\ulcorner x \urcorner, \ulcorner y \urcorner, \ldots\}$ of variables in the language of $\mathcal{FO}$, and let $\{\ulcorner 0_i \urcorner : i \geq 0\}$ be code numbers for the variant names of 0. The nodes of $G(\Pi)$ are $V \cup \{\ulcorner 0_i \urcorner : i \geq 0\}$. A 'rank' will be assigned to each $\ulcorner 0_i \urcorner$ as well as to each $\ulcorner x \urcorner \in V$. I write $\langle \ulcorner u \urcorner, \ulcorner v \urcorner, d \rangle$ for an edge from $\ulcorner u \urcorner$ to $\ulcorner v \urcorner$ of weight d.

The edge $\langle \ulcorner u \urcorner, \ulcorner v \urcorner, d\rangle$ represents a constraint, that the rank of v must be at least as great as d plus the rank of u.

Let $\Pi_1, \Pi_2, \ldots$ be the subproofs, in left to right order, leading to the premisses of the last step of Π. The notation $\ulcorner y \urcorner \in t$ or $\ulcorner 0_i \urcorner \in t$ means that the variable variable y or constant 0_i has an occurrence in the term t. Note that the quantified variables in (comp) and (rec.1-2) are not arbitrary but are in some cases the same for two quantifiers. This will save us a few edges in $G(\Pi)$, because edges of the form $\langle \ulcorner x \urcorner, \ulcorner x \urcorner, 0\rangle$ can be omitted; they would be redundant under the subsequent definition of ρ. The graph $G(\Pi)$ is defined by cases on the last rule in Π, as follows.

Ax Any axiom except (FO.5) or (rec.1).
(G.1) $G(\Pi) = \emptyset$

(FO.5) $\forall \vec{z}, ((\forall x, \alpha) \supset \alpha[t/x])$, where t is free for x in α
(G.2) $G(\Pi) = \{\langle \ulcorner x \urcorner, \ulcorner y \urcorner, 0\rangle : \ulcorner y \urcorner \in t, \ulcorner y \urcorner \neq \ulcorner x \urcorner\} \cup \{\langle \ulcorner x \urcorner, \ulcorner 0_i \urcorner, 0\rangle : 0_i \in t\}$

(rec.1) $((\forall \vec{y} \in K,\ Kg(\vec{y})) \wedge (\forall x, \vec{y}, z \in K,\ Kh(x, \vec{y}, z)))$
$\supset (\forall \vec{y} \in K,\ f(0_i, \vec{y}) = g(\vec{y}))$
(G.3) $G(\Pi) = \{\langle \ulcorner x \urcorner, \ulcorner 0_i \urcorner, 0\rangle\} \cup \{\langle \ulcorner z \urcorner, \ulcorner 0_i \urcorner, 1\rangle\} \cup \{\langle \ulcorner z \urcorner, \ulcorner y_j \urcorner, 0\rangle : 1 \leq j \leq m_f - 1\}$

(MP) $\alpha,\ \alpha \supset \beta \mid \beta$
(G.4) $G(\Pi) = G(\Pi_1) \cup G(\Pi_2)$.

(ind) $(\alpha[0_i/y]), (\forall x \in K, (x \neq 0_j \supset (\alpha[px/y] \supset \alpha[x/y])))\mid (\forall z \in K, \alpha[z/y])$
(G.5) $G(\Pi) = G(\Pi_1) \cup G(\Pi_2) \cup \{\langle \ulcorner 0_i \urcorner, \ulcorner z \urcorner, 0\rangle\} \cup \{\langle \ulcorner x \urcorner, \ulcorner z \urcorner, 0\rangle\}$

This completes the definition of $G(\Pi)$. Next I will define $\rho(\Pi, \ulcorner x \urcorner)$ for $\ulcorner x \urcorner \in V$. Approximately speaking, $\rho(\Pi, \ulcorner x \urcorner)$ is the supremum of the weights of all finite paths that end at $\ulcorner x \urcorner$. In this way, one has $\rho(\Pi, \ulcorner x \urcorner) \geq d + \rho(\Pi, \ulcorner y \urcorner)$ whenever there is a finite path from y to x of total weight d. Thus, ρ satisfies the constraints expressed by the edges in $G(\Pi)$. In fact, it will be simplest to define ρ to be the minimum function satisfying these constraints.

One hopes that a suitable renaming of variables in Π would guarantee that $G(\Pi)$ contains no positively weighted cycles. But instead of attempting such a renaming, I simply ignore any proofs whose constraint graph contains a positively weighted cycle. This is accomplished by assigning rank ω when a cycle occurs. Evidently, not too many proofs are ignored this way: later it will be shown that there are finite-rank proofs of convergence of all the primitive recursive functions.

A *ranking function*, q, is a total function from $V \cup \{\ulcorner 0_i \urcorner : i \geq 0\}$ to $\mathbf{N} \cup \{\omega\}$, where $\mathbf{N}$ is the non-negative integers. Say that q is *good for* G, written $q \geq G$, iff: $\langle \ulcorner u \urcorner, \ulcorner v \urcorner, d\rangle \in G$ implies $q(\ulcorner v \urcorner) \geq d + q(\ulcorner u \urcorner)$. If q is good for G, then in fact q is good for every subgraph of G. For a nonempty set of ranking functions Q, the pointwise minimum '$\min Q$' is the function q' such that $q'(\ulcorner u \urcorner) = \min\{q(\ulcorner u \urcorner) : q \in Q\}$, for all $\ulcorner u \urcorner \in V \cup \{\ulcorner 0_i \urcorner : i \geq 0\}$. It happens that if every $q \in Q$ is good for G, then so is $\min Q$. By assigning $q(\ulcorner u \urcorner) = \omega$ in case $\ulcorner u \urcorner$ is in a positively weighted cycle in $G(\Pi)$, one can guarantee that $\{q : q \geq G(\Pi)\}$ is nonempty. For details, see parts (1a) and (2) of the next lemma below.

Writing $\rho(\Pi, \cdot)$ for $\lambda n.\rho(\Pi, n)$, define:

$$\rho(\Pi, \cdot) = \min\{q : \ q \geq G(\Pi)\}.$$

One has that $\rho(\Pi, \ulcorner x \urcorner) = 0$ for all $\ulcorner x \urcorner \notin \Pi$. See (1b) of the next lemma below. Define the *rank* of proof Π by:

$$\rho(\Pi) = \max\{\rho(\Pi, \ulcorner x \urcorner) : \ulcorner x \urcorner \in V\}.$$

Notice that $\rho(\Pi)$ does not depend on $\rho(\Pi, \ulcorner 0_i \urcorner)$ unless there is an edge from $\ulcorner 0_i \urcorner$ to some $\ulcorner x \urcorner \in V$.

The definition of $\rho(\Pi)$ is "fair" in the sense discussed earlier, even though weighted edges are associated with the defining axioms of the functions. Although it is true that a positively weighted edge is added for an instance of a function definition (rec.1, G.3), such an edge ends at a name for zero rather than at a variable. The weight of this edge does not contribute to $\rho(\Pi)$ until a path containing the edge is extended by a (G.5) edge to reach the induction variable in an (ind) conclusion. This is because the definition of $\rho(\Pi)$ only refers to $q(\ulcorner x \urcorner)$ for variables $\ulcorner x \urcorner \in V$, and an (ind) instance is the only place where an edge passes from some $\ulcorner 0_i \urcorner$ to some variable $\ulcorner x \urcorner \in V$. Uses of the axioms defining the functions do not increase the rank either by themselves or in combination with (MP) and the logical axioms: all $\mathcal{SQ}_2$ proofs have rank 0. Thus, induction is critical. Say that two occurrences of (ind) are *nested* if there is a path from any literal in the conclusion of one, to any literal in any antecedent of the other. If a proof has rank $r < \omega$, then it must contain instances of induction nested r deep: the graph $G(\Pi)$ contains a path passing through r different names for zero, corresponding to r interleavings of (G.3/rec.1) and (G.5/ind).[2]

The converse is not true in general. In a rank 1 proof there may be arbitrarily many (ind) occurrences, and these may be arbitrarily deeply nested. This is analogous to the fact that a rank 1 function derivation may contain arbitrarily deeply "nested" recursions, for example if each recursion is on a different input. Not all occurrences of induction have new computational content. Viewed differently, an occurrence of induction on a variable x might happen to generate a ramification level that is less than the ramification levels already generated by some other inductions in other parts of the proof. We wouldn't want to increase the overall proof rank for such an occurrence of induction. Essentially, the presence of weighted edges in some of the non-induction parts of the proof, is a way of determining

[2]Consider a rank r proof, $r < \omega$. Its graph contains a path of weight r ending at a variable, in which each node of the path is labeled with a different element of $V \cup \{\ulcorner 0_i \urcorner : \ i \geq 0\}$. The graph doesn't contain any weighted cycles, because $r < \omega$. But the weighted edges in a graph are exactly the edges that pass from some variable $\ulcorner x \urcorner$ to some name for zero, $\ulcorner 0_i \urcorner$ (they are (G.3) edges). Consider any two weighted edges on the path, such that no weighted edge appears between them. In order to form a path, there must be a "return" subpath from the literal $\ulcorner 0_i \urcorner$ to the variable $\ulcorner y \urcorner$ that starts the next weighted edge. There also must be a similar "return" subpath from the last $\ulcorner 0_j \urcorner$ to the variable at the end of the path. These "return" subpaths must each start with a (G.5) edge corresponding to an instance of induction, because (G.5) is the only place where an edge is added from a zero to a variable. These r instances of (ind) are nested, and they are distinct because they use different names 0_i in their base cases.

whether or not subsequent inductions have new computational content i.e. generate a new ramification level. An induction that does not use a new statement of fact (i.e. a defining axiom for a function) is not one that needs to be accounted for.

In summary, although the definition of proof rank uses the recursion axioms for the primitive recursive functions, proof rank is not a disguise for function rank. Proof rank relates to the layering of inductions. Function rank relates to the layering of recursions. These are two independent concepts. A low-rank proof can refer to high-rank functions, and a high-rank proof might refer only to low-rank functions.

The following lemma was promised above.

Lemma 4.1 *For every proof* Π*, (1a) there is a ranking function* q *that is good for* $G(\Pi)$*, such that (1b) if* x *is a variable not in* Π*, then* $q(\ulcorner x \urcorner) = 0$*. Furthermore, (2) if* Q *is any nonempty set of ranking functions such that every* $q \in Q$ *is good for* $G(\Pi)$*, then* $\min Q$ *is also good for* $G(\Pi)$*.*

Proof. The required ranking function assigns the first limit ordinal, ω, when there is a cycle in $G(\Pi)$. *Fact:* if $0 \leq d < \omega$ then $d + \omega = \omega$, although $\omega + d > \omega$. See Mendelson [17] p. 204.

(1). For $\ulcorner u \urcorner \in V \cup \{\ulcorner 0_i \urcorner : i \geq 0\}$, let $P_{\ulcorner u \urcorner}$ be the set of finite-length paths terminating at $\ulcorner u \urcorner$ in $G(\Pi)$. Note $P_{\ulcorner u \urcorner}$ is nonempty because it contains the trivial path terminating at $\ulcorner u \urcorner$. Let $W(p)$ be the sum of the edge weights in finite path p. Define $q(\ulcorner u \urcorner) = \sup\{W(p) : p \in P_{\ulcorner u \urcorner}\}$. Note $0 \leq q(\ulcorner u \urcorner) \leq \omega$ because $W(p) < \omega$ for $p \in P_{\ulcorner u \urcorner}$ and because $W(p) = 0$ for the trivial path $p \in P_{\ulcorner u \urcorner}$.

Now consider any edge $\langle \ulcorner u \urcorner, \ulcorner v \urcorner, d \rangle$. Every $p \in P_{\ulcorner u \urcorner}$ can be extended to a finite path $p' \in P_{\ulcorner v \urcorner}$ by adding the edge $\langle \ulcorner u \urcorner, \ulcorner v \urcorner, d \rangle$. Therefore, $\{W(p) : p \in P_{\ulcorner v \urcorner}\} \supseteq \{d + W(p) : p \in P_{\ulcorner u \urcorner}\}$. Two cases are considered: either $q(\ulcorner v \urcorner) < \omega$ or $q(\ulcorner v \urcorner) = \omega$. If $q(\ulcorner v \urcorner) < \omega$, then $\{W(p) : p \in P_{\ulcorner v \urcorner}\}$ is finite and nonempty, and therefore so is $\{d + W(p) : p \in P_{\ulcorner u \urcorner}\}$. In this case, $q(\ulcorner v \urcorner) = \max\{W(p) : p \in P_{\ulcorner v \urcorner}\} \geq \max\{d + W(p) : p \in P_{\ulcorner u \urcorner}\} \geq d + \max\{W(p) : p \in P_{\ulcorner u \urcorner}\} = d + q(\ulcorner u \urcorner)$. On the other hand, if $q(\ulcorner v \urcorner) = \omega$, then using the Fact, $q(\ulcorner v \urcorner) = \omega \geq d + \omega \geq d + q(\ulcorner u \urcorner)$. The two cases together prove that if $\langle \ulcorner u \urcorner, \ulcorner v \urcorner, d \rangle \in G$ then $q(\ulcorner v \urcorner) \geq d + q(\ulcorner u \urcorner)$. Therefore (1a): q is good for $G(\Pi)$. Also (1b): if u does not appear in Π then $P_{\ulcorner u \urcorner}$ contains only the trivial path of length 0 terminating at $\ulcorner u \urcorner$, and therefore $q(\ulcorner u \urcorner) = 0$.

(2) Assume that Q is non-empty and that every $q \in Q$ is good for $G(\Pi)$. Let $q' = \min Q$ and consider any edge $\langle \ulcorner u \urcorner, \ulcorner v \urcorner, d \rangle \in G(\Pi)$. Let q be a function in Q minimizing $q(\ulcorner v \urcorner)$; then $q'(\ulcorner v \urcorner) = q(\ulcorner v \urcorner)$ by the definition of $\min Q$. Since q is good for $G(\Pi)$, one has $q(\ulcorner v \urcorner) \geq d + q(\ulcorner u \urcorner)$. By the definition of $\min Q$ again, $q(\ulcorner u \urcorner) \geq q'(\ulcorner u \urcorner)$. Therefore, $q'(\ulcorner v \urcorner) = q(\ulcorner v \urcorner) \geq d + q(\ulcorner u \urcorner) \geq d + q'(\ulcorner u \urcorner)$. It follows that q' is good for $G(\Pi)$. $\square$

5 Standard Ramified Model

Now the development turns to semantic issues.

Let $\perp$ be a unique element, and let $\mathbf{N}^0, \mathbf{N}^1, \mathbf{N}^2, \ldots$ together with $\mathbf{N}^\omega$ be countably many copies of the non-negative integers.

Define a *depth* for each point in $\cup_j \mathbf{N}^j \cup \{\perp\}$ as follows: $\delta(\mathbf{x}) = j$ for all $\mathbf{x} \in \mathbf{N}^j$, and $\delta(\perp) = -1$. The intuitive meaning is that $\delta(\mathbf{x})$ is the depth of knowledge that one has of $\mathbf{x}$: the greater the depth, the greater one's understanding of $\mathbf{x}$, and the more able one is to perform induction with $\mathbf{x}$ or to compute functions with $\mathbf{x}$ as an input. The multiple copies of an integer do not represent different *values*, but different *amounts of understanding* of a single integer value.

A vector $\vec{\mathbf{x}} \in (\cup_j \mathbf{N}^j \cup \{\perp\})^{(m_f)}$ of length m_f is *adequate* for $f \in \mathcal{PR}_2$ if $\delta(\mathbf{x}_i) \geq \rho(f, i)$ for all i, $1 \leq i \leq m_f$. For example, for every function f, every vector consisting of m_f many elements from $\mathbf{N}^\omega$ is adequate for f. The empty vector is adequate for every zero-ary function. A vector containing $\perp$ is inadequate for every function.

Some more notation will be helpful. For a point $\mathbf{x} \in \cup_j \mathbf{N}^j$, let $\underline{\mathbf{x}}$ be the corresponding integer (i.e. the integer without indication of which set $\mathbf{N}^j$ it is in). For an integer $\mathbf{n}$, let $[\underline{\mathbf{n}}]^j$ be the copy of $\mathbf{n}$ in $\mathbf{N}^j$. For a function symbol f, one writes $\mathbf{f}$ for the interpretation of f in a given model, and $\underline{\mathbf{f}}$ for the standard function on the integers defined by the $\mathcal{PR}_2$ derivation. Given an assignment A and a model M, one writes $\mathbf{t}$ for the interpretation of term t under (M, A).

The *standard ramified model* is the model M_0 defined as follows. The universe of M_0 is $\cup_j \mathbf{N}^j \cup \{\perp\}$. Define $\mathbf{x} = \mathbf{y}$ iff: $\delta(\mathbf{x}) \geq 0$, and $\delta(\mathbf{y}) \geq 0$, and $\underline{\mathbf{x}}$ is identically $\underline{\mathbf{y}}$. Similarly, define $\mathbf{x} \neq \mathbf{y}$ iff: $\delta(\mathbf{x}) \geq 0$, and $\delta(\mathbf{y}) \geq 0$, and $\underline{\mathbf{x}}$ is different from $\underline{\mathbf{y}}$. The interpretation $\mathbf{f}$ of function symbol f is defined by:

$$\mathbf{f}(\vec{\mathbf{x}}) \equiv \begin{cases} [\underline{\mathbf{f}}(\underline{\vec{\mathbf{x}}})]^{\min(\{\delta(\mathbf{x}_i):\, 1 \leq i \leq m_f\} \cup \{\omega\})} & \text{if } \vec{\mathbf{x}} \text{ is adequate for } f \\ \perp & \text{otherwise.} \end{cases}$$

Under this definition, if $\vec{\mathbf{x}}$ is adequate for f and $m_f \geq 1$, then $\delta(\mathbf{f}(\vec{\mathbf{x}})) = \min \delta(\vec{\mathbf{x}})$. If f is zero-ary, then $\delta(\mathbf{f}()) = \omega$. For example, in M_0 every variant name 0_i is interpreted as $[\underline{\mathbf{0}}]^\omega$. For every j, the predecessor of $[\underline{\mathbf{0}}]^j$ is $[\underline{\mathbf{0}}]^j$; in fact $\delta(\mathbf{px}) = \delta(\mathbf{x})$ for every $\mathbf{x}$.

Recall $Kt \equiv \exists w, (w = t \vee w \neq t)$; the interpretation $\mathbf{K}$ of K is $\cup_j \mathbf{N}^j$. The model does not satisfy $\forall x, Kx$, because $\perp \notin \mathbf{K}$. Say that a function *diverges* on a given vector of points if the output of the function is not in $\mathbf{K}$; otherwise it *converges*. In these models, functions converge iff the input is adequate. For example, if any input is $\perp$ then the output is also $\perp$.

M_0 has been presented to make the basic structure apparent. In fact the following will be used instead. A *variant standard ramified model* is a model M that is the same as M_0 except the variant names 0_i are interpreted by any shadows of 0, not just by $[\underline{\mathbf{0}}]^\omega$. Other constant symbols are still given interpretations in $\mathbf{N}^\omega$. One must be careful to remember that the subscript i in 0_i refers to the variant name, not to the depth $\delta(\mathbf{0}_i)$ of the shadow $\mathbf{0}_i$ denoted by 0_i. Note that M_0 itself is a variant model.

Given a ranking function q, say that M is *good for* q, written $M \geq q$, iff M is a variant standard ramified model such that $\delta(\mathbf{0}_i) \geq q(\ulcorner 0_i \urcorner)$ for all i. Note $M_0 \geq q$ for every ranking function q. When one considers all possible models $M \geq q$ for a given q, the interpretation of 0_i is essentially relativized to $\{[\underline{\mathbf{0}}]^j : q(\ulcorner 0_i \urcorner) \leq j \leq \omega\}$.

None of the variant models is a model of the theorems of $\mathcal{SA}_2$. However, it will be shown that a suitable relativization of each $\mathcal{SA}_2$ theorem is satisfied.

In order to relativize variables, let U^j $(1 \leq j \leq \omega)$ be new unary predicate letters. One should realize that the symbols U^j are only being introduced in order to carry out the analysis, and they do not appear in the definition of $\mathcal{SA}_2^r$ or in the final statement of the results. Give U^j the interpretation $\mathbf{U}^j \equiv \cup_{i \geq j} \mathbf{N}^i \cup \{\perp\}$ in every variant model M. Thus, $j \geq k$ implies $\mathbf{U}^j \subseteq \mathbf{U}^k$. As j increases, the subuniverses $\mathbf{U}^j$ get smaller and smaller, and more and more functions are total on $\mathbf{U}^j \cap \mathbf{K}$; in some sense, more and more is known over $\mathbf{U}^j \cap \mathbf{K}$. This seems to be similar in spirit to the successive worlds of an intuitionistic Kripke model.

For a formula α, let α^q be obtained from α by relativizing each variable x to $U^{q(\ulcorner x \urcorner)}$. The relativization replaces each quantifier $\forall x,\ \alpha$ with $\forall x,\ (U^{q(\ulcorner x \urcorner)}x \supset \alpha)$, and prefixes the formula with "$U^{q(\ulcorner x \urcorner)}x \supset$" for each free variable x.

The following closure property of $\mathbf{U}^j$ is used for analyzing substitution (FO.5).

Lemma 5.1 (Closure of $\mathbf{U}^j$) *Given a term t and any $j \geq 0$, let A be an assignment for a variant M such that $\mathbf{y} \in \mathbf{U}^j$ for all $y \in t$ and $\mathbf{0}_i \in \mathbf{U}^j$ for all $0_i \in t$. Then $\mathbf{t} \in \mathbf{U}^j$.*

Proof. It is proved by induction on the term structure of t. If t is just a variable y, then the statement of the lemma follows trivially by the assumption $\mathbf{y} \in \mathbf{U}^j$. If t is a zero-ary function symbol f other than some 0_i, then $\delta(\mathbf{t}) = \omega$ and the statement again follows. If t is some 0_i then it follows trivially because $\mathbf{0}_i \in \mathbf{U}^j$ by assumption.

Otherwise, t is $f(\vec{a})$ and the arity of f is $m_f \geq 1$. By the induction hypothesis on each term in $\vec{a}$, one has $\vec{\mathbf{a}} \in \mathbf{U}^j$, i.e. $\delta(\mathbf{a}_i) = -1$ or $\delta(\mathbf{a}_i) \geq j$ (for $1 \leq i \leq m_f$). Under the definition of M_0, if $\vec{\mathbf{a}}$ is inadequate for f then $\mathbf{f}(\vec{\mathbf{a}})$ is $\perp$, in which case $\mathbf{f}(\vec{\mathbf{a}}) \in \mathbf{U}^j$. On the other hand, if $\vec{\mathbf{a}}$ is adequate for f then $\delta(\mathbf{a}_i) \geq j$ for each i, because $\delta(\mathbf{a}_i) < j \Rightarrow \delta(\mathbf{a}_i) = -1 \Rightarrow \vec{\mathbf{a}}$ is inadequate for f. Therefore, if $\vec{\mathbf{a}}$ is adequate for f (with $m_f \geq 1$) one has $\delta(\mathbf{f}(\vec{\mathbf{a}})) = \min\{\delta(\mathbf{a}_i) :\ 1 \leq i \leq m_f\} \geq j$, again implying $\mathbf{f}(\vec{\mathbf{a}}) \in \mathbf{U}^j$. □

6 Implicit Ramification Theorem

Now we have arrived at the key "implicit ramification" theorem relating the ranking of induction to the ramification subuniverses U^j. The theorem says that if Π proves α and $q \geq G(\Pi)$, then $M \models \alpha^q$ for every $M \geq q$ (including M_0). A simple semantic argument (see §7) will then show that if Π proves the convergence of f, then $\rho(\Pi) \geq \rho(f)$.

The proof of the implicit ramification theorem is unfortunately long, since we must consider each axiom and rule of $\mathcal{SA}_2$. I have broken it into two lemmas. The first lemma considers the simpler cases: all axioms other than (FO.5) and (rec.1). This lemma does involve some work, for (comp) and (rec.2). The second lemma handles the most interesting cases, finishing the theorem by treating the relativizations of (FO.5), (rec.1), (MP), and (ind).

In variant models, $(Kt)^q$ holds if and only if Kt holds: there is something equal or not equal to t, iff there is something in $\mathbf{N}^\omega$ that is equal or not equal to t. Therefore, when making semantic arguments one can ignore relativizations of the quantifier in K. In the following I drop the relativization and refer only to $\mathbf{K}$.

Lemma 6.1 *Let Π be an instance of an $\mathcal{SA}_2$ axiom other than (FO.5) or (rec.1), let $q \geq G(\Pi)$ and let $M \geq q$. Then M satisfies the q-relativization of the axiom instance.*

Proof. The relativizations of the axioms (FO.1), (FO.2), (FO.3), (FO.4), and (FO.6) follow by standard first-order semantic reasoning.

Consider the equality axioms of $\mathcal{SFO}$. The relativization of each of these axioms consists of the same axiom with the universal quantifiers relativized to a subuniverse. The restriction to a subuniverse only weakens the assertion. So, if the unrelativized form is satisfied by M then the relativized form is also satisfied by M. By the construction, the unrelativized axioms are satisfied by M. I take substitution (sub) as a typical example. Assume $\vec{a} = \vec{b}$ and $Kf(\vec{a})$ and $Kf(\vec{b})$ are satisfied. Since $\mathbf{Kf}(\vec{\mathbf{a}})$ and $\mathbf{Kf}(\vec{\mathbf{b}})$ hold, it must be that $\mathbf{f}(\vec{\mathbf{a}})$ is $[\underline{\mathbf{f}}(\underline{\vec{\mathbf{a}}})]^j$ and $\mathbf{f}(\vec{\mathbf{b}})$ is $[\underline{\mathbf{f}}(\underline{\vec{\mathbf{b}}})]^k$ for some j and k. Since $\vec{\mathbf{a}} = \vec{\mathbf{b}}$, it must be that the underlying integers are the same: $\underline{\vec{\mathbf{a}}}$ is identically $\underline{\vec{\mathbf{b}}}$. Therefore $\underline{\mathbf{f}}(\underline{\vec{\mathbf{a}}})$ is identically $\underline{\mathbf{f}}(\underline{\vec{\mathbf{b}}})$, implying $\mathbf{f}(\vec{\mathbf{a}}) = \mathbf{f}(\vec{\mathbf{b}})$ in M, as required by the conclusion of substitutivity.

Now consider the $\mathcal{SQ}_2$ axioms for the initial functions (0.1-2), (s.1-7), (p.1-3) and (c.1-2). Without belaboring details, these axioms are all satisfied by the construction of M, because: the initial functions 0_i, s_0, s_1, p and c converge everywhere in $\mathbf{K}$ and output a value in $\mathbf{K}$ whose underlying integer is the output of the standard function; the depth of the value in $\mathbf{K}$ does not matter to the '$=$' and '$\neq$' relations; and the corresponding statements are true in the standard model of arithmetic. The relativization of the universal quantifiers in these axioms only weakens the statement, so the relativizations also hold.

What remains is (comp) and (rec.2). There are no free variables in the relativized form of the (comp) axiom:

$$\begin{aligned}&(\forall \vec{x} \in U^{q(\ulcorner \vec{x} \urcorner)} \cap K,\ Kh(g_1(\vec{x}^1), \ldots, g_m(\vec{x}^m)))\\ &\supset (\forall \vec{x} \in U^{q(\ulcorner \vec{x} \urcorner)} \cap K,\ f(\vec{x}) = h(g_1(\vec{x}^1), \ldots, g_m(\vec{x}^m)))\end{aligned}$$

The argument can be summarized as follows. The relativized antecedent asserts that vectors of depth $q(\ulcorner \vec{x} \urcorner)$ are adequate for $\lambda \vec{x}.h(g_1(\vec{x}^1), \ldots, g_m(\vec{x}^m))$. Under the definition of ρ, it will follow that vectors of depth $q(\ulcorner \vec{x} \urcorner)$ are adequate for f. In M_0, the convergence of both sides of "$f(\vec{x}) = h(g_1(\vec{x}^1), \ldots, g_m(\vec{x}^m))$" ensures that the equality holds, because f was defined in $\mathcal{PR}_2$ by composition from h and $g_1, \ldots, g_m$.

- To show that $(\text{comp})^q$ is satisfied in M, let A be any assignment. Assume that for all $\vec{\mathbf{x}} \in \mathbf{U}^{q(\ulcorner \vec{x} \urcorner)} \cap \mathbf{K}$ one has $M, A|^{\vec{\mathbf{x}}}_{\vec{x}} \models Kh(g_1(\vec{x}^1), \ldots, g_m(\vec{x}^m))$. One must show, for $\vec{\mathbf{x}} \in \mathbf{U}^{q(\ulcorner \vec{x} \urcorner)} \cap \mathbf{K}$, that $M, A|^{\vec{\mathbf{x}}}_{\vec{x}} \models f(\vec{x}) = h(g_1(\vec{x}^1), \ldots, g_m(\vec{x}^m))$. Choose any $\vec{\mathbf{x}} \in \mathbf{U}^{q(\ulcorner \vec{x} \urcorner)} \cap \mathbf{K}$. Let $\mathbf{G}_j$ be the interpretation of $g_j(\vec{x}^j)$ and let $\mathbf{F}$ be the interpretation of $f(\vec{x})$. Recall that
  $\rho(f, i) = \max(\{\rho(h, j) :\ x_i \in \vec{x}^j\} \cup \{\rho(g_j, k) :\ x_i \equiv \langle \vec{x}^j \rangle_k\} \cup \{0\})$ for $1 \leq i \leq m_f$.

Since $Kh(g_1(\vec{x}^1), \ldots, g_m(\vec{x}^m)))$ holds, it must be that $\vec{\mathbf{G}}$ is adequate for h, because h only converges on adequate inputs. It follows that $x_i \in \vec{x}^j \Rightarrow \delta(\mathbf{x}_i) \geq \rho(h, j)$ for all i and j, because: if $x_i \in \vec{x}^j$ and $\delta(\mathbf{x}_i) < \rho(h, j)$ then $\delta(\mathbf{G}_j) = \delta(\mathbf{g}_j(\vec{\mathbf{x}}^j)) = \min(\{\delta(\mathbf{x}_{i'}) : \mathbf{x}_{i'} \in \vec{\mathbf{x}}^j\} \cup \{\omega\}) \leq \delta(\mathbf{x}_i) < \rho(h, j)$, contradicting that $\vec{\mathbf{G}}$ is adequate for h. The adequacy of $\vec{\mathbf{G}}$ also implies $\vec{\mathbf{G}} \in \mathbf{K}$, which implies that for each j, $\vec{\mathbf{x}}^j$ is adequate for g_j. Therefore, $\delta(\mathbf{x}_i) \geq \rho(g_j, k)$ for all j and k such that $x_i \equiv (\vec{x}^j)_k$. One also has $\delta(\mathbf{x}_i) \geq 0$ for all i, because $\mathbf{x}_i \in \mathbf{K}$. Using the definition of $\rho(f, i)$, these three statements about $\delta(\mathbf{x}_i)$ imply that $\mathbf{x}_i$ is adequate for f. Therefore, $\mathbf{F} \in \mathbf{K}$ and by the definition of M, $\mathbf{F} = [\underline{\mathbf{f}}(\underline{\vec{\mathbf{x}}})]^{\delta(\mathbf{F})}$. Since $\vec{\mathbf{G}}$ is adequate for h and $\vec{\mathbf{x}}^j$ is adequate for g_j $(1 \leq j \leq m)$, the interpretation of $h(g_1(\vec{x}^1), \ldots, g_m(\vec{x}^m)))$ is $[\underline{\mathbf{h}}(\underline{\mathbf{g}}_1(\underline{\vec{\mathbf{x}}}^1), \ldots, \underline{\mathbf{g}}_m(\underline{\vec{\mathbf{x}}}^m))]^d$ for some d. By the definition of M this is equal to $[\underline{\mathbf{f}}(\underline{\vec{\mathbf{x}}})]^{\delta(\mathbf{F})}$, completing the proof that $M, A|^{\vec{\mathbf{x}}}_{\vec{x}} \models f(\vec{x}) = h(g_1(\vec{x}^1), \ldots, g_m(\vec{x}^m))$.

Finally, the relativized form of (rec.2) is:

$$
\begin{aligned}
&((\forall \vec{y} \in U^{q(\ulcorner \vec{y} \urcorner)} \cap K,\ Kg(\vec{y})) \wedge (\forall x, \vec{y}, z \in U^{q(\ulcorner x \urcorner), q(\ulcorner \vec{y} \urcorner), q(\ulcorner z \urcorner)} \cap K,\ Kh(x, \vec{y}, z))) \\
&\supset \forall x, \vec{y} \in U^{q(\ulcorner x \urcorner), q(\ulcorner \vec{y} \urcorner)} \cap K,\ (x \neq 0_j \supset (Kf(px, \vec{y}) \supset f(x, \vec{y}) = h(x, \vec{y}, f(px, \vec{y})))).
\end{aligned}
$$

The argument is summarized as follows. The convergence statement $Kf(px, \vec{y})$ implies that $\langle \mathbf{px}, \vec{\mathbf{y}} \rangle$ is adequate for f. Since $\delta(\mathbf{x}) = \delta(\mathbf{px})$, this also means that $\langle \mathbf{x}, \vec{\mathbf{y}} \rangle$ is adequate for f. It will follow using the definition of ρ that $\langle \mathbf{x}, \vec{\mathbf{y}}, \mathbf{f}(\mathbf{px}, \vec{\mathbf{y}}) \rangle$ is adequate for h. As in the case of (comp), convergence of the two sides implies that the equality holds. The antecedent clauses actually are not required to show that $(\text{rec.2})^q$ is satisfied.

- Letting A be any assignment, assume the antecedent of $(\text{rec.2})^q$ is satisfied, and choose any $\mathbf{x}, \vec{\mathbf{y}} \in \mathbf{U}^{q(\ulcorner x \urcorner), q(\ulcorner \vec{y} \urcorner)} \cap \mathbf{K}$ such that $\mathbf{x} \neq \mathbf{0}_j$. One must show that $M, A|^{\mathbf{x}, \vec{\mathbf{y}}}_{x, \vec{y}} \models (Kf(px, \vec{y}) \supset f(x, \vec{y}) = h(x, \vec{y}, f(px, \vec{y}))$. Recall that $\rho(f, 1) = \max\{\rho(h, 1), 1 + \rho(h, m_h)\}$, and $\rho(f, i) = \max\{\rho(h, i), \rho(g, i-1), \rho(h, m_h)\}$ for $2 \leq i \leq m_f$.

  Assume that $M, A|^{\mathbf{x}, \vec{\mathbf{y}}}_{x, \vec{y}} \models Kf(px, \vec{y})$. Then $\langle \mathbf{px}, \vec{\mathbf{y}} \rangle$ is adequate for f, because functions in M only converge on adequate inputs. Therefore $\delta(\mathbf{f}(\mathbf{px}, \vec{\mathbf{y}})) = \min(\{\delta(\mathbf{px})\} \cup \{\delta(\mathbf{y}_i) : 1 \leq i < m_f - 1\}) \geq \min\{\rho(f, i) : 1 \leq i \leq m_f\} \geq \rho(h, m_f)$ by the adequacy of $\langle \mathbf{px}, \vec{\mathbf{y}} \rangle$ and the definition of $\rho(f)$. Also one has $\delta(\mathbf{x}) = \delta(\mathbf{px}) \geq \rho(f, 1) \geq \rho(h, 1)$ and $\delta(\mathbf{y}_i) \geq \rho(f, i+1) \geq \rho(h, i+1)$ for $1 \leq i \leq m_f - 1$, again using the adequacy of $\langle \mathbf{px}, \vec{\mathbf{y}} \rangle$ and the definition of $\rho(f)$. It follows that $\langle \mathbf{x}, \vec{\mathbf{y}}, \mathbf{f}(\mathbf{px}, \vec{\mathbf{y}}) \rangle$ is adequate for h. Easily, $\langle \mathbf{x}, \vec{\mathbf{y}} \rangle$ is adequate for f because $\langle \mathbf{px}, \vec{\mathbf{y}} \rangle$ is adequate for f and because $\delta(\mathbf{x}) = \delta(\mathbf{px})$. In the variant model M, under the definition of f for $\mathbf{x} \neq \mathbf{0}_j$ with f derived in $\mathcal{PR}_2$ by recursion from g and h, the adequacy of the inputs on each side of the equation implies that $f(x, \vec{y}) = h(x, \vec{y}, f(px, \vec{y}))$ holds.

□

Now we have reached the more interesting cases, involving edges of $G(\Pi)$.

Theorem 6.2 (Implicit Ramification) *Let Π be an $\mathcal{SA}_2$ proof with conclusion β, let $q \geq G(\Pi)$, and let $M \geq q$. Then $M \models \beta^q$.*

Proof. The proof is by structural induction on Π, considering each rule or axiom by which the conclusion of Π could have been reached. Applying the preceding lemma, what remains is (FO.5), (rec.1), (MP) and (ind). Let A be any assignment.

Consider (FO.5): $\forall\vec{z}, (\forall x,\, \alpha \supset \alpha[t/x])$. The relativization has the form

$$\gamma \supset \forall\vec{z} \in U^{q(\ulcorner\vec{z}\urcorner)},\ ((\forall x \in U^{q(\ulcorner x\urcorner)},\ \alpha') \supset \alpha'[t/x]),$$

where γ consists of the antecedents relativizing any free variables. To show that (M, A) satisfies the relativized statement, assume $M, A \models \gamma$, let $\vec{\mathbf{z}} \in \mathbf{U}^{q(\ulcorner\vec{z}\urcorner)}$, and assume $M, A|^{\vec{\mathbf{z}}}_{\vec{z}} \models \forall x \in U^{q(\ulcorner x\urcorner)},\ \alpha'$. One must show that $M, A|^{\vec{\mathbf{z}}}_{\vec{z}} \models \alpha'[t/x]$. Because of the edge $\langle\ulcorner x\urcorner, \ulcorner y\urcorner, 0\rangle$ added in (G.2), one has $q(\ulcorner x\urcorner) \le q(\ulcorner y\urcorner)$ for all variables $\ulcorner y\urcorner \in t$ (it is trivial if $\ulcorner y\urcorner = \ulcorner x\urcorner$). Similarly, the edge $\langle\ulcorner x\urcorner, \ulcorner 0_i\urcorner, 0\rangle$ ensures $q(\ulcorner x\urcorner) \le q(\ulcorner 0_i\urcorner)$ for all $0_i \in t$. Let $j = q(\ulcorner x\urcorner)$; one has $\mathbf{U}^j \supseteq \mathbf{U}^{q(\ulcorner y\urcorner)}$ and $\mathbf{U}^j \supseteq \mathbf{U}^{q(\ulcorner 0_i\urcorner)}$. Each variable $y \in t$ is either free or is among $\vec{z}$; in either case one has $\mathbf{y} \in \mathbf{U}^j$ because $(M, A|^{\vec{\mathbf{z}}}_{\vec{z}})$ satisfies a relativizing clause $U^{q(\ulcorner y\urcorner)}y$ that implies $U^j y$. Similarly, each $0_i \in t$ is interpreted by a shadow $\mathbf{0}_i \in \mathbf{U}^j$ because $M \ge q$ and $\mathbf{U}^j \supseteq \mathbf{U}^{q(\ulcorner 0_i\urcorner)}$. Now the closure lemma for $\mathbf{U}^j$ gives $\mathbf{t} \in \mathbf{U}^j$. Since $M, A|^{\vec{\mathbf{z}}}_{\vec{z}} \models \forall x \in U^j,\ \alpha'$, one has $M, A|^{\vec{\mathbf{z}}}_{\vec{z}} \models \alpha'[t/x]$ as required.

The relativized form of (rec.1) is:

$$\begin{aligned}&((\forall\vec{y} \in U^{q(\ulcorner\vec{y}\urcorner)} \cap K,\ Kg(\vec{y})) \wedge (\forall x, \vec{y}, z \in U^{q(\ulcorner x\urcorner), q(\ulcorner\vec{y}\urcorner), q(\ulcorner z\urcorner)} \cap K,\ Kh(x, \vec{y}, z)))\\ &\supset (\forall\vec{y} \in U^{q(\ulcorner\vec{y}\urcorner)} \cap K,\ (f(0_i, \vec{y}) = g(\vec{y})))\end{aligned}$$

The argument can be summarized as follows. The relativized antecedents assert that points at the depths given by q, are adequate for g and h. This, together with the (G.3) edges and the definition of $\rho(f)$, ensures that appropriate vectors $\langle\mathbf{0}_i, \vec{\mathbf{y}}\rangle$ are adequate for f. Using the definition of M_0, the adequacy of inputs on both sides of the equality "$f(0_i, \vec{y}) = g(\vec{y})$" ensures that the equality holds, because f is defined in $\mathcal{PR}_2$ by a recursion having g as the base function.

- Assume the antecedent of $(\text{rec.1})^q$ is satisfied, and choose any $\vec{\mathbf{y}} \in \mathbf{U}^{q(\ulcorner\vec{y}\urcorner)} \cap \mathbf{K}$; one must show $M, A|^{\vec{\mathbf{y}}}_{\vec{y}} \models f(0_i, \vec{y}) = g(\vec{y})$. In M, functions only converge on adequate inputs; therefore, satisfaction of $(\forall\vec{y} \in U^{q(\ulcorner\vec{y}\urcorner)} \cap K,\ Kg(\vec{y}))$ (under any assignment) implies that $q(\ulcorner y_j\urcorner) \ge \rho(g, j)$ for $1 \le j \le m_g$. Similarly, the satisfaction of $(\forall x, \vec{y}, z \in U^{q(\ulcorner x\urcorner), q(\ulcorner\vec{y}\urcorner), q(\ulcorner z\urcorner)} \cap K,\ Kh(x, \vec{y}, z)))$ implies that $q(\ulcorner x\urcorner) \ge \rho(h, 1)$, and $q(\ulcorner y_j\urcorner) \ge \rho(h, j+1)$ for $1 \le j < m_h - 2$, and $q(\ulcorner z\urcorner) \ge \rho(h, m_h)$.

  Now, the edges added in (G.3) for the (rec.1) axiom are

  $$\{\langle\ulcorner x\urcorner, \ulcorner 0_i\urcorner, 0\rangle\} \cup \{\langle\ulcorner z\urcorner, \ulcorner 0_i\urcorner, 1\rangle\} \cup \{\langle\ulcorner z\urcorner, \ulcorner y_j\urcorner, 0\rangle :\ 1 \le j \in m_f - 1\}$$

  Since $q \ge G(\Pi)$, these give $q(\ulcorner 0_i\urcorner) \ge q(\ulcorner x\urcorner) \ge \rho(h, 1)$ and, critically, $q(\ulcorner 0_i\urcorner) \ge 1 + q(\ulcorner z\urcorner) \ge 1 + \rho(h, m_h)$. Under the definition of $\rho(f, 1) = \max\{\rho(h, 1), 1 + \rho(h, m_h)\}$ it follows that $q(\ulcorner 0_i\urcorner) \ge \rho(f, 1)$. We already have that $q(\ulcorner y_j\urcorner) \ge \max\{\rho(g, j), \rho(h, j+1)\}$ from the antecedents, and using the third set of edges one also has $q(\ulcorner y_j\urcorner) \ge q(\ulcorner z\urcorner) \ge \rho(h, m_h)$, for $1 \le j \le m_f - 1$. Since $\rho(f, j+1) = \max\{\rho(h, j+1), \rho(h, m_h), \rho(g, j)\}$ by definition, one has $q(\ulcorner y_j \urcorner) \geq \rho(f, j+1)$ for $1 \leq j \leq m_f - 1$. In summary, therefore, the conditions $M \geq q$ and $\vec{\mathbf{y}} \in \mathbf{U}^{q(\ulcorner \vec{y} \urcorner)} \cap \mathbf{K}$ (together with the antecedents) ensure that $\langle \mathbf{0}_i, \vec{\mathbf{y}} \rangle$ is adequate for f. Similarly, $\vec{\mathbf{y}}$ is adequate for g. Therefore, under the definition of M, one has $f(0_i, \vec{y}) = [\underline{\mathbf{f}}(\underline{\mathbf{0}_i}, \underline{\vec{\mathbf{y}}})]^{\delta(\mathbf{f}(0_i, \vec{\mathbf{y}}))} = [\underline{\mathbf{g}}(\underline{\vec{\mathbf{y}}})]^{\delta(\mathbf{g}(\vec{\mathbf{y}}))} = g(\vec{y})$ as required.

For (MP), let Π_1 be a proof of α, and let Π_2 be a proof of $\alpha \supset \beta$. The relevant clause is (G.4): $G(\Pi) = G(\Pi_1) \cup G(\Pi_2)$. Since q is good for $G(\Pi)$, q is also good for the subgraphs: $q \geq G(\Pi_1)$ and $q \geq G(\Pi_2)$. Applying the induction hypothesis on Π_1 and Π_2 with M, one has $M, A \models \alpha^q$ and $M, A \models (\alpha \supset \beta)^q$. By standard first-order semantic reasoning, $M, A \models \beta^q$.

Finally we have arrived at (ind). Let Π_1 and Π_2 be the subproofs for the premisses of an induction with conclusion $\forall z \in K,\ \alpha[z/y]$. Recall the (G.5) edges:

$$G(\Pi) = G(\Pi_1) \cup G(\Pi_2) \cup \{\langle \ulcorner 0_i \urcorner, \ulcorner z \urcorner, 0 \rangle\} \cup \{\langle \ulcorner x \urcorner, \ulcorner z \urcorner, 0 \rangle\}.$$

Since $G(\Pi) \supseteq G(\Pi_1) \cup G(\Pi_2)$, the assumption $q \geq G(\Pi)$ gives $q \geq G(\Pi_1)$ and $q \geq G(\Pi_2)$. Therefore the induction hypothesis can be applied to Π_1 and Π_2. Letting γ be the relativizing antecedents for the free variables of α (other than y), and letting α' be obtained from α by relativizing the quantifiers only, the induction hypothesis implies that for all $M' \geq q$,

$$\begin{array}{lll} (1) & M', A & \models \gamma \supset \alpha'[0_i/y] \\ (2) & M', A & \models \gamma \supset \forall x \in U^{q(\ulcorner x \urcorner)} \cap K,\ (x \neq 0_j \supset (\alpha'[px/y] \supset \alpha'[x/y])). \end{array}$$

Note that γ does not contain x or z since these are not free in α. Let $d = q(\ulcorner z \urcorner)$. It is sufficient to show that, for the original M given in the statement of the theorem,

$$(3) \qquad M, A \ \models \gamma \supset \forall z \in U^d \cap K,\ \alpha'[z/y].$$

Since $q \geq G(\Pi)$, the existence of the edges $\langle \ulcorner 0_i \urcorner, \ulcorner z \urcorner, 0 \rangle$ and $\langle \ulcorner x \urcorner, \ulcorner z \urcorner, 0 \rangle$ implies $d = q(\ulcorner z \urcorner) \geq q(\ulcorner 0_i \urcorner)$ and similarly $d \geq q(\ulcorner x \urcorner)$.

It is proved by an induction on $\underline{\mathbf{z}}$ that $M, A \models (\gamma \wedge U^d z \wedge Kz) \supset \alpha'[z/y]$. This implies (3) as required. (Recall that $\underline{\mathbf{z}}$ is the underlying integer for the element $\mathbf{z}$ that A assigns for z. We can assume $\mathbf{z} \in \mathbf{K}$ because of the antecedent Kz.)

The induction is summarized as follows. When $\mathbf{z}$ is a shadow of zero of depth at least d, let M' be M but with 0_i interpreted as $[\underline{\mathbf{0}}]^{\delta(\mathbf{z})}$. By (1), $\gamma \supset \alpha'[0_i/y]$ holds in M'. Since $\mathbf{z}$ is $\mathbf{0}_i$, this means that $\gamma \supset \alpha'[z/y]$ holds in M'. Since M and M' only differ on 0_i, which doesn't appear in $\alpha'[z/y]$, this means that M satisfies $\gamma \supset \alpha'[z/y]$. The base case of the induction on $\mathbf{z}$ follows. The step case follows by a slightly simpler argument using (2) with $M' \equiv M$.

- In the base case of this induction, $\underline{\mathbf{z}}$ is zero. If $\delta(\mathbf{z}) < d$ then $(U^d z \wedge Kz)$ is falsified and the entire formula is satisfied. Otherwise, $\mathbf{z}$ is a shadow of zero of depth at least $d \geq q(\ulcorner 0_i \urcorner)$. Consider the variant model M' obtained from M by setting the interpretation of 0_i equal to $[\underline{\mathbf{0}}]^{\delta(\mathbf{z})}$. Since $M \geq q$ and $\delta(\mathbf{z}) \geq q(\ulcorner 0_i \urcorner)$, one has $M' \geq q$, enabling us to use the induction hypothesis (1). Since $\mathbf{z}$ is identically $\mathbf{0}_i$ under (M', A), one has $M', A \models (\gamma \wedge U^d z \wedge Kz) \supset \alpha'[z/y]$ if and only if $M', A \models (\gamma \wedge U^d 0_i \wedge K0_i) \supset \alpha'[0_i/y]$.

The latter statement is implied by (1). Now we must replace M' with M again. By the definition of (ind), 0_i does not appear in α (nor in α', γ, $U^d z$ or Kz). But M and M' differ only in the interpretation of 0_i. Therefore $M, A \models (\gamma \wedge U^d z \wedge Kz) \supset \alpha'[z/y]$ if and only if $M', A \models (\gamma \wedge U^d z \wedge Kz) \supset \alpha'[z/y]$. The satisfaction by (M', A) has already been proved, finishing the base case of the induction on $\underline{\mathbf{z}}$.

- In the induction step case, the underlying integer $\underline{\mathbf{z}}$ is greater than zero. Again the case of interest is when $\delta(\mathbf{z}) \geq q(\ulcorner z \urcorner)$. Observe that $\underline{\mathbf{pz}} < \underline{\mathbf{z}}$. The subsidiary induction hypothesis is $M, A|_z^{\mathbf{pz}} \models (\gamma \wedge U^d z \wedge Kz) \supset \alpha'[z/y]$. By the definition of the (ind) rule, z is not free in α (or γ). So, the free occurrences of z are all displayed when we write "$(\gamma \wedge U^d z \wedge Kz) \supset \alpha'[z/y]$". Therefore, the subsidiary induction hypothesis is equivalent to $M, A \models (\gamma \wedge U^d pz \wedge Kpz) \supset \alpha'[pz/y]$. Since $\delta(\mathbf{x}) = \delta(\mathbf{px})$, this is equivalent to
$$(4) \quad M, A \models (\gamma \wedge U^d z \wedge Kz) \supset \alpha'[pz/y].$$
At the same time, we have the main induction hypothesis (2) with $M' \equiv M$. Assume $M, A \models (\gamma \wedge U^d z \wedge Kz)$. By (4), $M, A \models \alpha'[pz/y]$. Then using (2) with $M' \equiv M$, one has $M, A \models \alpha'[z/y]$ (the condition $\mathbf{z} \neq \mathbf{0}_j$ is satisfied because $\underline{\mathbf{z}} > 0$). Therefore, $M, A \models (\gamma \wedge U^d z \wedge Kz) \supset \alpha'[z/y]$ as required to complete the subsidiary induction on $\underline{\mathbf{z}}$.

□

7 Correspondence of $\mathcal{SA}_2^r$ and $\mathcal{PR}_2^r$

One direction of the main result can now be completed. Recall that the convergence statement for f is $\forall \vec{x} \in K,\ Kf(\vec{x})$.

Theorem 7.1 *Let $f \in \mathcal{PR}_2$ and let $r < \omega$. If $\mathcal{SA}_2^r$ proves the convergence of f, then $f \in \mathcal{PR}_2^r$.*

Proof. Let $\Pi \in \mathcal{SA}_2^r$ be a proof of $\forall \vec{x} \in K,\ Kf(\vec{x})$. Recall that $\rho(\Pi, \cdot) \geq G(\Pi)$ by definition. Also, $M_0 \geq \rho(\Pi, \cdot)$. Using the preceding theorem (and the fact that K is equivalent to its own relativization), $M_0 \models \forall \vec{x} \in U^{\rho(\Pi, \ulcorner \vec{x} \urcorner)} \cap K,\ Kf(\vec{x})$. Choose any $\vec{\mathbf{x}}$ such that $\delta(\mathbf{x}_i) = \rho(\Pi, \ulcorner x_i \urcorner)$ for $1 \leq i \leq m_f$. Then $\vec{\mathbf{x}} \in \mathbf{U}^{\rho(\Pi, \ulcorner \vec{x} \urcorner)} \cap \mathbf{K}$. By the relativized convergence statement, $\mathbf{f}(\vec{\mathbf{x}}) \in \mathbf{K}$. But by the construction of M_0, if $\vec{\mathbf{x}}$ were not adequate for f then $\mathbf{f}(\vec{\mathbf{x}})$ would diverge (i.e. would not be in $\mathbf{K}$). Therefore, $\vec{\mathbf{x}}$ must be adequate for f, implying $\rho(\Pi, \ulcorner x_i \urcorner) = \delta(\mathbf{x}_i) \geq \rho(f, i)$ for $1 \leq i \leq m_f$. This implies $\rho(\Pi) \geq \rho(f)$, so $f \in \mathcal{PR}_2^{\rho(\Pi)}$. Here, the expression "$\mathcal{PR}_2^{\rho(\Pi)}$" is defined because $\rho(\Pi) \leq r < \omega$ by assumption. □

The other direction of the main result is not too difficult. A fact will pave the way.

Fact. Any first-order proof that uses substitution (FO.5) only to replace a quantified variable $\forall x$, with a free instance of the same variable x, has $G(\Pi) = \emptyset$. Specifically, the following derived rules leave the constraint graph unchanged: (UE) if $\vdash \forall x, \alpha$ then $\vdash \alpha$; and (UI) if $\vdash \alpha$ then $\vdash \forall x, \alpha$.

Theorem 7.2 *Let $f \in \mathcal{PR}_2$, and let $r < \omega$. If $f \in \mathcal{PR}_2^r$, then there is an $\mathcal{SA}_2^r$ proof of the convergence of f.*

Proof. In fact, there is a proof Π of $\forall \vec{x} \in K,\ Kf(\vec{x})$ such that $\rho(\Pi, \ulcorner x_i \urcorner) = \rho(f, i)$ and such that $G(\Pi)$ has no edge starting at $\ulcorner x_i \urcorner$, for $1 \leq i \leq m_f$. It is shown by induction on the structure of the derivation of f in $\mathcal{PR}_2$.

If f is one of the initial functions, then the convergence statement for f is proved using first-order logic from the defining axioms for f. For example, to prove convergence of c, one uses (p.4) together with (c.1-2). Other than (UI) and (UE), the only quantifier manipulation is of the form $t_1 = t_2 \supset Kt_1$ or $t_1 \neq t_2 \supset Kt_1$. It follows that the only edges in the constraint graph are weight 0 edges originating at the quantified variable in an occurrence of Kt (for some term t).

If f is defined by the composition $h(g_1(\vec{x}^1), \ldots, g_m(\vec{x}^m))$, then let Π_0 and $\Pi_1, \ldots, \Pi_m$ be the proofs of convergence obtained by the induction hypothesis on h and $g_1, \ldots, g_m$. Renaming the variables in Π_0, say Π_0 is a proof of $\forall \vec{y} \in K,\ Kh(\vec{y})$. Eliminating each $\forall y_j$ with $g_j(\vec{x}^j)$, using (UE) on each $\forall x_i$ in each Π_j, $j \geq 1$, and then making propositional moves followed by (UI), one obtains $\forall \vec{x} \in K,\ Kh(g_1(\vec{x}^1), \ldots, g_m(\vec{x}^m))$. In performing the replacement of y_j with $g_j(\vec{x}^j)$, one creates an edge of weight zero from $\ulcorner y_j \urcorner$ to $\ulcorner x_i \urcorner$ for each $x_i \in \vec{x}^j$. Essentially this forces $\rho(\Pi, x_i) \geq \rho(\Pi_0, y_j)$. The edges of $G(\Pi_j)$ $(1 \leq j \leq m)$ remain in the proof, as do the edges of $G(\Pi_0)$. For this construction, $\rho(\Pi, x_i) = \max\{\rho(\Pi_0, y_j) : x_i \in \vec{x}^j\} \cup \{\rho(\Pi_j, (\vec{x}^j)_k) : x_i \equiv (\vec{x}^j)_k\} \cup \{0\}$. Since $\rho(\Pi_0, y_j) = \rho(h, j)$ and $\rho(\Pi_j, (\vec{x}^j)_k) = \rho(g_j, k)$ by the induction hypothesis, under the definition of $\rho(f, i)$ this gives $\rho(\Pi, x_i) = \rho(f, i)$.

If f is defined by recursion from step function h and base function g, then let Π_h and Π_g be obtained by the induction hypothesis on h and g. Rename the variables as shown in the following outline sketch of Π. First-order deductions have been abbreviated here.

First,

$$\dfrac{\dfrac{\dfrac{\Pi_g}{\forall \vec{y} \in K,\ Kg(\vec{y})} \qquad \dfrac{\Pi_h}{\forall x, \vec{y}, z \in K,\ Kh(x, \vec{y}, z)} \qquad \text{(rec.1) for } f}{\forall \vec{y} \in K,\ (f(0_i, \vec{y}) = g(\vec{y}))}}{\forall \vec{y} \in K,\ Kf(0_i, \vec{y})}$$

Second,

$$\dfrac{\dfrac{\dfrac{\dfrac{\Pi_g}{\forall \vec{y} \in K,\ Kg(\vec{y})} \qquad \dfrac{\Pi_h}{\forall x, \vec{y}, z \in K,\ Kh(x, \vec{y}, z)} \qquad \text{(rec.2) for } f}{\forall x, \vec{y} \in K,\ (x \neq 0_j \supset (Kf(px, \vec{y}) \supset f(x, \vec{y}) = h(x, \vec{y}, f(px, \vec{y}))))}}{\forall x, \vec{y} \in K,\ x \neq 0_j \supset (Kf(px, \vec{y}) \supset Kf(x, \vec{y}))}}{\forall x \in K,\ (x \neq 0_j \supset (\forall \vec{y} \in K,\ Kf(px, \vec{y}) \supset \forall \vec{y} \in K,\ Kf(x, \vec{y})))}$$

These two pieces are joined together by an (ind) instance,

$$\dfrac{\forall \vec{y} \in K,\ Kf(0_i, \vec{y}) \qquad \forall x \in K,\ (x \neq 0_j \supset (\forall \vec{y} \in K,\ Kf(px, \vec{y}) \supset \forall \vec{y} \in K,\ Kf(x, \vec{y})))}{\forall u \in K, \forall \vec{y} \in K,\ Kf(u, \vec{y})}$$

Examining this outline, one sees that a path of weight one is created from the critical variable $\ulcorner z \urcorner$ to the induction variable $\ulcorner u \urcorner$ through the intermediary $\ulcorner 0_i \urcorner$, due to the (ind) following the (rec.1). A path of weight zero is similarly created from the recursion variable $\ulcorner x \urcorner$ to $\ulcorner 0_i \urcorner$ and from there to the induction variable $\ulcorner u \urcorner$. For the other variables $\vec{y}$, the instances of (rec.1) for f result in edges of weight 0 from $\ulcorner z \urcorner$ to $\ulcorner y_i \urcorner$, for $1 \leq i \leq m_f - 1$.

These are the only significant new paths created in the constraint graph. Although some paths are created due to the implicit use of $f(x, \vec{y}) = h(x, \vec{y}, f(px, \vec{y})) \supset Kf(x, \vec{y})$, these have weight 0 and terminate at the quantified variable in $Kf(x, \vec{y})$. That variable does not appear in Π_h or Π_g, hence these paths contribute zero to the rank of the overall proof.

By the induction hypothesis, $\rho(\Pi_h, \ulcorner x \urcorner) = \rho(h, 1)$ and $\rho(\Pi_h, \ulcorner z \urcorner) = \rho(h, m_h)$. Due to the new paths ending at $\ulcorner u \urcorner$, and the fact that $\ulcorner z \urcorner$ and $\ulcorner x \urcorner$ do not appear in Π_g, it follows that $\rho(\Pi, \ulcorner u \urcorner) = \max\{1 + \rho(\Pi_h, \ulcorner z \urcorner), \rho(\Pi_h, \ulcorner x \urcorner)\} = max\{1 + \rho(h, m_h), \rho(h, 1)\} = \rho(f, 1)$.

Due to the paths from $\ulcorner z \urcorner$ to $\vec{y}$, one has $\rho(\Pi, \ulcorner y_i \urcorner) = \max\{\rho(\Pi_h, \ulcorner z \urcorner), \rho(\Pi_h, \ulcorner y_{i-1} \urcorner), \rho(\Pi_g, \ulcorner y_{i-1} \urcorner)\}$ for $2 \leq i \leq m_f$. By the induction hypothesis on Π_g and Π_h one has $\rho(\Pi_g, \ulcorner y_{i-1} \urcorner) = \rho(g, i-1)$ and $\rho(\Pi_h, \ulcorner y_{i-1} \urcorner) = \rho(h, i)$ $(2 \leq i \leq m_f)$ and $\rho(\Pi_h, \ulcorner z \urcorner) = \rho(h, m_h)$. Therefore $\rho(\Pi, \ulcorner y_{i-1} \urcorner) = \max\{\rho(g, i-1), \rho(h, i), \rho(h, m_h)\} = \rho(f, i)$ for $2 \leq i \leq m_f$. $\square$

Altogether, the following has been proved.

Corollary 7.3 *Let $f \in \mathcal{PR}_2$ and let $r < \omega$. Then $\mathcal{SA}_2^r$ proves the convergence of f, if and only if $f \in \mathcal{PR}_2^r$.*

Corollary 7.4 *A function is polynomial-time computable if and only if it is $\mathcal{SA}_2^1$-provably convergent. A function is Kalmar elementary if and only if it is $\mathcal{SA}_2^2$-provably convergent. For $r \geq 2$, a function is in $\mathcal{E}^{r+1}$ if and only if it is $\mathcal{SA}_2^r$-provably convergent.*

8 Ordinary Induction and Recursion

There is nothing special about the base 2 (induction-on-notation) formulation of the above results.

One can just as well define $\mathcal{PR}_1$ instead of $\mathcal{PR}_2$, by replacing s_0, s_1, p, and c with the initial functions $S(x) = x+1$ and $P(x) = x-1$ and $C(x, y, z) =$ "if $x = 0$ then y else z". The rule of "full recursion on notation" now becomes "full primitive recursion", replacing p with P. The definition of $\rho(f)$ is carried over unchanged, except that the set of initial functions is different. It is shown in the companion paper [4] that for $r \geq 1$, one has $\mathcal{PR}_1^r = \mathcal{E}^{r+1}$, the Grzegorczyk functions of level $r+1$. By Ritchie's theorem, $\mathcal{PR}_1^1$ consists of exactly the linear-space computable functions.

For a corresponding logic, one can define $\mathcal{SQ}_1$ and $\mathcal{SA}_1$ exactly as $\mathcal{SQ}_2$ and $\mathcal{SA}_2$ were defined, but using a signature based on $\mathcal{PR}_1$, changing p to P in (ind) and (rec.1-2), and replacing the initial axioms for s_0, s_1, p, and c with new axioms corresponding to the initial functions of $\mathcal{PR}_1$. The results are carried over literally by reading p as P — except for the proof that M satisfies the relativized axioms of the initial functions, which can be directly adapted to suit the new initial functions.

Theorem 8.1 *Let $f \in \mathcal{PR}_1$, and let $r < \omega$. Then $\mathcal{SA}_1^r$ proves the convergence of f, if and only if $f \in \mathcal{PR}_1^r$.*

Corollary 8.2 *A function is linear-space computable if and only if it is $\mathcal{SA}_1^1$-provably convergent. A function is Kalmar elementary if and only if it is $\mathcal{SA}_1^2$-provably convergent. For $r \geq 2$, a function is $\mathcal{SA}_1^r$-provably convergent (using a $\mathcal{PR}_1$ derivation) if and only if it is $\mathcal{SA}_2^r$-provably convergent (using a $\mathcal{PR}_2$ derivation).*

9 Conclusion

The current work provides a contrast to earlier weak subsystems of arithmetic such as those of Buss [6] (for an overview see Krajicek [10]). The system $\mathcal{SA}_2^1$ is equivalent to Buss's S_2^1 in the sense that they both prove convergence of exactly the polytime computable functions. It seems quite remarkable that this definition of polynomial time provability admits unrestricted quantifier complexity in the induction formula. The correspondence of $\mathcal{PR}_j^i$ and $\mathcal{SA}_j^i$ ($j \in \{1,2\}$, $0 \leq i < \omega$) suggests that a very strong relationship holds between recursion and induction despite the presence of complicating logical elements such as implication and quantification.

It is interesting that ramification levels, as identified in this work by the relativizations $\mathbf{U}^j$, can be inferred from the syntactic structure of a proof by induction. They do not need to be treated explicitly, either in the logical syntax via special unary predicate symbols or in the semantics via successive worlds with nested domains. Instead, they emerge naturally from an examination of the ramification that seems to be implicit in ordinary induction. Although they are implicit rather than explicit, the ramification levels $\mathbf{U}^j$ seem quite analogous to the successive worlds that were presupposed by constructivists when inventing intuitionistic semantics.

Only symbols for the primitive recursive functions were included in $\mathcal{SA}_2$. If one were to admit general recursion or higher-type primitive recursion, then one could define e.g. a function with exponential output length, using only a single recursion on notation. Correspondingly, such a function could be proved convergent using a single induction (see Leivant [13] lemma 3.7). It would be interesting to see whether a polytime logic is obtained when one restricts Leivant's "intrinsic theory" over W_0 and W_1 by requiring that the recursive equations be primitive recursive, rather than by requiring only existential formulas in induction rules [13].

A distinction has been made between 'zero' used as the base case of an induction and unrelated references to 'zero' in the induction formula. It is not clear whether this distinction is fundamental or can be eliminated.

A drawback to the definition of rank is that $\mathcal{SA}_2^r$ is not closed under modus ponens. This contrasts with the fact that $\mathcal{PR}_2^r$ is closed under composition. Although $\mathcal{SA}_2^r$ forms a proof "system" in a loose sense of the word, the approach has really been to view $\mathcal{SA}_2$ as a powerful given system and then to measure the complexity of all $\mathcal{SA}_2$ proofs. In practice, a feeling for the definition of rank should allow one to write down proofs with confidence that they will fall into a particular system such as $\mathcal{SA}_2^1$. An alternative viewpoint is that a system formalizing the idea of reasoning with limited resources *shouldn't* be closed under modus ponens [8].

Connections remain to be drawn between $\mathcal{SA}_2^r$, $\mathcal{SA}_1^r$, and earlier arithmetics. It should happen that the full systems $\mathcal{SA}_2$ and $\mathcal{SA}_1$ are both equivalent to Peano arithmetic, under a suitable translation of signatures and relativization to K. The issue is complicated by the presence of 0_i in the induction rule. Also, it would be nice to know more about $\mathcal{SA}_2^\omega$ and $\mathcal{SA}_1^\omega$. A method of relabeling of variables in Π such that $G(\Pi)$ is guaranteed to be acyclic, would show that each $\mathcal{SA}_2^\omega$ proof can be carried out in $\mathcal{SA}_2^r$ for some finite r. Finally, it would be interesting to carry out an analysis similar to the present one, but on a system having, like S_2^1, only a finite number of predefined function symbols.

There seems to be a conceptual relationship between large non-standard integers and shadows of $\mathbf{N}^\omega$ at low depth. For a recursively defined function $\mathbf{f}$, the descending chain $\ldots, \mathbf{f}(\mathbf{n}{+}2), \mathbf{f}(\mathbf{n}{+}1), \mathbf{f}(\mathbf{n}), \ldots$ is unfounded in a nonstandard model, where $\mathbf{n}$ is a nonstandard integer that corresponds to 0 in the natural way. In a ramified model where $\mathbf{n}$ is a shadow of 0, the chain is also unfounded in the sense that $\mathbf{f}(\mathbf{n})$ is unconstrained by the axiom defining $\mathbf{f}$. In both cases, it is only induction that forces $\mathbf{f}$ to take on a 'correct' value at points in these chains. Under this analogy, the subuniverses $\mathbf{U}^0, \mathbf{U}^1, \mathbf{U}^2, \ldots$ are like smaller and smaller initial segments of a nonstandard model, with $\mathbf{U}^\omega$ being the standard integers. It would be quite interesting to find an exact relationship between nonstandard integers in nonstandard unramified models, and shadow integers in standard ramified models.

References

[1] M. Beeson, *Foundations of Constructive Mathematics*, Metamathematical Studies, Springer-Verlag (1985).

[2] S. Bellantoni, "Predicative Recursion and Computational Complexity", T.R. 264/92, Department of Computer Science, University of Toronto, September 1992.

[3] S. Bellantoni, "Predicative Recursion and the Polytime Hierarchy", in *Feasible Mathematics II*, P. Clote and J. B. Remmel eds.; Birkhauser, Boston, (1995) p. 15-29.

[4] S. Bellantoni, "Ranking Primitive Recursions: The Low Grzegorczyck Classes Revisited", manuscript, 1996.

[5] S. Bellantoni and S. Cook, "A New Recursion-Theoretic Characterization of the Polytime Functions", in *Computational Complexity* v. 2, p. 97-110, 1992.

[6] S. Buss, *Bounded Arithmetic*, Ph.D. Dissertation, Princeton University, 1985; reprinted Bibliopolis, Naples, 1986.

[7] P. Clote, "Computation Models and Function Algebras", in *Handbook of Recursion Theory*, Ed Griffor, ed., Elsevier, 1996.

[8] R. Fagin, J. Halpern, Y. Moses, M. Vardi, *Reasoning About Knowledge*, The MIT Press, Cambridge, MA. 1995.

[9] N. Goodman, "Epistemic Arithmetic is a Conservative Extension of Intuitionistic Arithmetic", in *The Journal of Symbolic Logic*, v. 49, n. 1, March 1984.

[10] J. Krajicek, *Bounded Arithmetic, Propositional Logic, and Complexity Theory*, Cambridge University Press, 1995.

[11] K. Lambert, ed., *Philosophical Applications of Free Logic*, Oxford University Press, 1991.

[12] D. Leivant, ed., *Logic and Computational Complexity: International Workshop LCC '94, Indianapolis, IN, USA, October 1994, Selected Papers*, D. Leivant, ed.; Lecture Notes in Computer Science 960, Springer (1995).

[13] D. Leivant, "Intrinsic Theories and Computational Complexity", in [12], p 177-194.

[14] D. Leivant, "Ramified Recurrence and Computational Complexity I: Word recurrence and poly-time", in *Feasible Mathematics II*, P. Clote and J. B. Remmel eds.; Birkhauser, Boston, (1995) p. 321-343.

[15] D. Leivant, "Stratified Functional Programs and Computational Complexity" in *Conference Record of the Twentieth Annual ACM Symposium on Principles of Programming Languages, New York, 1993*, Association for Computing Machinery.

[16] A. Margaris, *First Order Mathematical Logic*, Blaisdell Publishing Co., 1967.

[17] E. Mendelson, *Introduction To Mathematical Logic*, Van Nostrand, New York, 1979.

[18] D. Prawitz, "Ideas and Results In Proof Theory", in *Proceedings of the Second Scandinavian Logic Symposium*, J. E. Fenstad, ed., Studies in Logic and the Foundations of Mathematics, v. 63, North Holland, 1971.

[19] S. Shapiro, ed., *Intensional Mathematics*, Studies in Logic V113, North-Holland, 1985.

[20] S. Shapiro, "Epistemic and Intuitionistic Arithmetic", in [19].