Axiomatizations and Conservation Results for Fragments of Bounded Arithmetic[*]

Samuel R. Buss[†]
Department of Mathematics
University of California, San Diego

Abstract

This paper presents new results on axiomatizations for fragments of Bounded Arithmetic which improve upon the author's dissertation. It is shown that $(\Sigma^b_{i+1} \cap \Pi^b_{i+1})$-PIND and strong Σ^b_i-replacement are consequences of S^i_2. Also Δ^b_{i+1}-IND is a consequence of T^i_2. The latter result is proved by showing that S^{i+1}_2 is $\forall\exists\Sigma^b_{i+1}$-conservative over T^i_2. Furthermore, S^{i+1}_2 is conservative over $T^i_2 + \Sigma^b_{i+1}$-replacement with respect to Boolean combinations of Σ^b_{i+1}-formulas.

1 Introduction

In [1] we introduced weak first-order theories of arithmetic, called collectively *Bounded Arithmetic*. These theories have the non-logical symbols 0, S, $+$, $\cdot$, $\le$, $\lfloor \frac{1}{2}x \rfloor$, $|x|$ and $\#$ where 0, S, $+$, $\cdot$ and $\le$ have the usual interpretations of zero, successor, plus, times and less than or equal to, and where $|x| = \lceil \log_2(x) \rceil$ is the length of the binary representation of x, $\lfloor \frac{1}{2}x \rfloor$ is x divided by two rounded down, and $x\#y$ is $2^{|x|\cdot|y|}$. (The binary operator $\#$ is called the "smash" operation, see Nelson [6].)

[*] This paper is in final form and no version of it will be submitted for publication elsewhere.

[†] Work performed while at the Mathematical Sciences Research Institute and the University of California, Berkeley. Supported in part by NSF fellowships DMS-8511465 and DMS-8701828.

The syntax of first-order logic is enlarged to include *bounded quantifiers* of the forms $(\forall x \leq t)$ and $(\exists x \leq t)$ where t is an arbitrary term not containing x. Bounded quantifiers of the form $(\forall x \leq |t|)$ and $(\exists x \leq |t|)$ are called *sharply bounded quantifiers.* The usual first order quantifiers are called *unbounded quantifiers.*

A formula is *bounded* if all of its quantifiers are bounded. In [1], the bounded formulae are classified in a hierarchy of sets Σ_i^b and Π_i^b by counting alternations of bounded quantifiers, ignoring sharply bounded quantifiers. This is analogous to the definition of the arithmetical hierarchy where one counts the alternations of unbounded quantifiers, ignoring bounded quantifiers. It is well known that a predicate is definable by a Σ_i^b predicate if and only if it is a Σ_i^p predicate, where Σ_i^p is the set of predicates at the i-th level of the Meyer-Stockmeyer polynomial hierarchy; for example, Σ_1^p is NP, the set of non-deterministic polynomial time computable predicates.

Let Ψ be a set of formulae. The following axiom schemata are defined as follows where A may be any formula in Ψ:

$$\Psi\text{-IND}: \quad A(0) \wedge (\forall x)(A(x) \rightarrow A(Sx)) \rightarrow (\forall x)A(x)$$

$$\Psi\text{-PIND}: \quad A(0) \wedge (\forall x)(A(\lfloor \tfrac{1}{2}x \rfloor) \rightarrow A(x)) \rightarrow (\forall x)A(x)$$

$$\Psi\text{-LIND}: \quad A(0) \wedge (\forall x)(A(x) \rightarrow A(Sx)) \rightarrow (\forall x)A(|x|)$$

$$\Psi\text{-MIN}: \quad (\exists x)A(x) \rightarrow (\exists x)[A(x) \wedge (\forall y < x)(\neg A(y))]$$

$$\Psi\text{-LMIN}: \quad (\exists x)A(x) \rightarrow A(0) \vee (\exists x)[A(x) \wedge (\forall y \leq \lfloor \tfrac{1}{2}x \rfloor)(\neg A(y))]$$

Ψ-replacement :

$$\begin{aligned}&(\forall x \leq |t|)(\exists y \leq s)A(x,y) \leftrightarrow \\ &\qquad \leftrightarrow (\exists w \leq SqBd(t,s))(\forall x \leq |t|)(A(x,\beta(Sx,w)) \wedge \beta(Sx,w) \leq s)\end{aligned}$$

strong Ψ-replacement :

$$\begin{aligned}&(\exists w \leq SqBd(t,s))(\forall x \leq |t|)[(\exists y \leq s)A(x,y) \leftrightarrow \\ &\qquad\qquad \leftrightarrow A(x,\beta(Sx,w) \wedge \beta(Sx,w) \leq s]\end{aligned}$$

Here β is a variant of the Gödel sequence coding function, with $\beta(i,w)$ equal to the i-th element of the sequence coded by w, and $SqBd$ is a term which depends on the precise definition of the β function. It should be noted that the term $SqBd$ must use the # function symbol; indeed, # has precisely the growth rate necessary to make the replacement axioms valid.

Note the IND axioms are the usual induction axioms; both PIND and LIND are versions of induction on the length of a number. The MIN axioms express the least number principle; whereas LMIN is a *length* minimization axiom.

The theory T_2^i is a theory of Bounded Arithmetic axiomatized by the Σ_i^b-IND axioms and an additional finite set of open axioms. The theory S_2^i is axiomatized by the Σ_i^b-PIND axioms and the same finite set of open axioms. (The subscript 2 denotes the presence of # in the language.) S_2^1 is the weakest "nice" theory of Bounded Arithmetic; in particular, S_2^1 is strong enough to define any polynomial time computable function and to use induction on formulae containing symbols for the polynomial time computable functions [1]. Thus S_2^1 can define the Gödel β function and the replacement axioms are meaningful in S_2^1.

We say that a theory R can Σ_i^b-*define* a function $f\colon \mathbb{N}^k \to \mathbb{N}$ if there exists a Σ_i^b-formula $A(\vec{x},y)$ such that $R \vdash (\forall\vec{x})(\exists y)A(\vec{x},y)$ and such that $A(\vec{n},f(\vec{n}))$ is valid for all $\vec{n}$ in $\mathbb{N}^k$. In [1], it is shown that S_2^i can Σ_i^b-define precisely the $\square_i^p$ functions. The $\square_i^p$ functions are the functions at the i-th level of the polynomial time hierarchy; namely $\square_{i+1}^p$ is the set of functions which can be computed in polynomial time with an oracle for a Σ_i^p set, and $\square_1^p$ is the set of polynomial time computable functions. Hence S_2^1 can Σ_1^b-define precisely the polynomial time computable functions. Part of the motivation for studying Bounded Arithmetic comes from these connections to computational complexity. (See Nelson [6] for another motivation).

Many relationships among the various axiomatizations have been known. Firstly, for $i \geq 1$, the Σ_i^b-IND axioms imply the Σ_i^b-PIND axioms and the Σ_{i+1}^b-PIND axioms imply the Σ_i^b-IND axioms [1]. Hence the theory S_2^{i+1} contains T_2^i which in turn contains S_2^i. As a consequence, $S_2 = \bigcup_i S_2^i$ and $T_2 = \bigcup_i T_2^i$ are the same theory; they are also equivalent to the theory $I\Delta_0 + \Omega_1$ studied by Wilkie and Paris [9]. Secondly, relative to the base theory S_2^1, the Σ_i^b-IND, Π_i^b-IND and Σ_i^b-MIN axioms are equivalent. In addition Σ_i^b-PIND, Σ_i^b-LIND, Π_i^b-PIND, Π_i^b-LIND and Σ_i^b-LMIN axioms are equivalent over S_2^1. Finally, combining results of [1] and Ressayre [8] it was known that, relative to the base theory S_2^1, the Σ_{i+1}^b-replacement axioms imply the strong Σ_i^b-replacement axioms which imply the Σ_i^b-PIND axioms which in turn imply the Σ_i^b-replacement axioms for all i.

In this paper some further results of this type are proved. First we show that Σ_i^b-PIND implies strong Σ_i^b-replacement; i.e., that S_2^i proves strong Σ_i^b-replacement, for $i \geq 1$. In addition, we show that S_2^i proves $(\Sigma_{i+1}^b \cap \Pi_{i+1}^b)$-PIND and T_2^i proves Δ_{i+1}^b-IND. The results for S_2^i are not too difficult; however, for the other result we must first show that S_2^{i+1} is $\forall\exists\Sigma_{i+1}^b$ conservative over T_2^i. We also show that S_2^{i+1} is conservative over $T_2^i + \Sigma_{i+1}^b$-replacement for all Boolean combinations of Σ_{i+1}^b-formulae (possibly containing free variables).

The class $\Sigma_{i+1}^b \cap \Pi_{i+1}^b$ should not be confused with Δ_{i+1}^b. A formula A is Δ_i^b with respect to a theory R if and only if R proves A is equivalent both to a Σ_i^b-formula and a Π_i^b-formula; when it is clear from the context what the theory R is, we shall just say A is "Δ_i^b" instead of "Δ_i^b with respect to R". On the other hand, $\Sigma_{i+1}^b \cap \Pi_{i+1}^b$ is the class of formulae which are explicitly written in Σ_{i+1}^b and Π_{i+1}^b form simultaneously.[‡]

There are still a number of open problems concerning axiomatizations of Bounded Arithmetic, most notably, whether S_2 is finitely axiomatizable and whether the theories S_2^i and T_2^i are all distinct. Several other, less ambitious, open problems are posed at the end of this paper.

2 The Main Results

We begin by proving two theorems about the theories S_2^i.

Theorem 1 $(i \geq 1)$. *Let $A(v,x)$ be a Σ_i^b-formula and $t(v)$ be a term. Then*

$$S_2^i \vdash (\exists w)(\forall x \leq |t|)(A(v,x) \leftrightarrow Bit(x,w) = 1).$$

Thus $S_2^i \vdash$ strong Σ_i^b-replacement.

The final sentence of Theorem 1 is an easy consequence of the first part and of the fact that Σ_i^b-replacement is provable by S_2^i. The function symbol $Bit(i,y)$ is Σ_1^b-defined by S_2^1 to be equal to 0 or 1 depending on the value of the bit in the 2^i position of the binary representation of y. (Much of our notation is explained in detail in [1].)

[‡] Louise Hay and the author [2] have shown that the predicates definable by $\Sigma_2^b \cap \Pi_2^b$-formulae are precisely the predicates which are polynomial time truth table reducible to SAT. More generally, a predicate is definable by a $\Sigma_{i+1}^b \cap \Pi_{i+1}^b$-formula if and only if it is polynomial time truth table reducible to a set in Σ_i^p.

$$\begin{tikzcd}
\Sigma^b_i\text{-IND} \arrow[r, Leftrightarrow] \arrow[d, Rightarrow] & \Pi^b_i\text{-IND} \arrow[r, Leftrightarrow] & \Sigma^b_i\text{-MIN} \arrow[r, Leftrightarrow] & \Delta^b_{i+1}\text{-IND} \\
\Sigma^b_i\text{-PIND} \arrow[r, Leftrightarrow] \arrow[d, Leftrightarrow] & \Pi^b_i\text{-PIND} \arrow[r, Leftrightarrow] & \Sigma^b_i\text{-LIND} \arrow[r, Leftrightarrow] & \Pi^b_i\text{-LIND} \\
\Sigma^b_i\text{-LMIN} \arrow[r, Leftrightarrow] \arrow[d, Rightarrow] & \text{strong } \Sigma^b_i\text{-replacement} \arrow[r, Leftrightarrow] & (\Sigma^b_{i+1} \cap \Pi^b_{i+1})\text{-PIND} \\
\Sigma^b_{i-1}\text{-IND}
\end{tikzcd}$$

$$\Sigma^b_{i+1}\text{-MIN} \iff \Pi^b_i\text{-MIN}$$

$$\Sigma^b_{i+1}\text{-replacement} \implies \Sigma^b_i\text{-PIND} \implies \Sigma^b_i\text{-replacement}$$

$$S^{i+1}_2 \underset{\Sigma^b_{i+1}}{\succ} T^i_2$$

$$S^{i+1}_2 \underset{\mathcal{B}(\Sigma^b_{i+1})}{\succ} T^i_2 + \Sigma^b_{i+1}\text{-replacement}$$

Relationships among axiomatizations for Bounded Arithmetic
relative to the base theory S^1_2 with $i \geq 1$
(including the results of this paper)

Proof Let $\mathit{Numones}(w)$ be the Σ^b_1-defined function symbol of S^1_2 which is equal to the number of ones in the binary representation of w. That is,

$$\mathit{Numones}(w) = (\#i < |w|)(\mathit{Bit}(i, w) = 1),$$

so *Numones* is a kind of Hamming metric. Let $B(k, v)$ be the formula

$$(\exists w < 2^{|t|+1})[\mathit{Numones}(w) = k \wedge (\forall x \le |t|)(\mathit{Bit}(x, w) = 1 \rightarrow A(v, x))].$$

Clearly $S^i_2 \vdash B(0, v)$ and $S^i_2 \vdash k > j \wedge B(k, v) \rightarrow B(j, v)$. Since $B \in \Sigma^b_i$ and $S^i_2 \vdash \neg B(|t| + 2, v)$, it follows from Σ^b_i-LIND that

$$S^i_2 \vdash (\exists k \le |t| + 1)(B(k, v) \wedge \neg B(k + 1, v)).$$

Thus S^i_2 proves that there exists a maximum value for k such that $B(k, v)$ holds. The w associated with this k is the desired w which makes Theorem 1 true. □

Definition *Let $A(b)$ be a formula with free variable b and possibly other free variables. Then $\mathit{PIND}_A(b)$, $\mathit{IND}_A(b)$ and $\mathit{MIN}_A(b)$ are the formulae:*

$$\mathit{PIND}_A(b) : \quad A(0) \wedge (\forall x \le b)(A(\lfloor \tfrac{1}{2} x \rfloor) \rightarrow A(x)) \rightarrow A(b)$$

$$\mathit{IND}_A(b) : \quad A(0) \wedge (\forall x < b)(A(x) \rightarrow A(x + 1)) \rightarrow A(b)$$

$$\mathit{MIN}_A(b) : \quad A(b) \rightarrow (\exists x \le b)[A(x) \wedge (\forall y < x)(\neg A(y))].$$

Theorem 2 $(i \ge 1)$. *Suppose $A \in \Sigma^b_{i+1} \cap \Pi^b_{i+1}$. Then $S^i_2 \vdash \mathit{PIND}_A$.*

In other words, $S^i_2 \vdash (\Sigma^b_{i+1} \cap \Pi^b_{i+1})$-PIND for $i \ge 1$. It is important to recall the distinction between $\Sigma^b_{i+1} \cap \Pi^b_{i+1}$ and Δ^b_{i+1}; it is open whether S^i_2 proves Δ^b_{i+1}-PIND.

Proof First note that every $A(b, \vec{v}) \in \Sigma^b_{i+1} \cap \Pi^b_{i+1}$ can be put in the form

$$(Q_1 x_1 \le |t_1|) \cdots (Q_k x_k \le |t_k|)\mathcal{B}(A_1, \ldots, A_s)$$

where each A_j is a Σ^b_i-formula and $\mathcal{B}(A_1, \ldots, A_s)$ denotes a Boolean combination of $A_1, \ldots, A_s$; this is readily shown by induction on the complexity of A.[§] Note especially that each $(Q_i x_i \le |t_i|)$ is a sharply bounded quantifier. Without loss of generality, we can assume that each term t_j contains as variables only b and the parameters $\vec{v}$; also, the formulae $A_j(b,\vec{v},\vec{x})$ have free variables as indicated.

Let $C(y,b,\vec{v})$ be the formula $A(\mathit{MSP}(b,|b| \dot{-} y),\vec{v})$ where $\dot{-}$ is subtraction and where $\mathit{MSP}(b,z)$ is the Σ_1^b-defined function of S_2^1 which is equal to the integer part of $b/2^z$. Thus it will suffice to show that S_2^1 proves $\mathrm{LIND}_C(y)$ (where now b becomes a parameter). Towards this end, let $C_j(y,b,\vec{v},\vec{x})$ be $A_j(\mathit{MSP}(b,|b| \dot{-} y),\vec{v},\vec{x})$. By a trivial extension of Theorem 1, S_2^i can prove the existence of numbers $w_1,\ldots,w_k$ such that

$$(\forall y \le |b|)(\forall x_1 \le |t_1|)\cdots(\forall x_k \le |t_k|)[\mathit{Bit}(\langle y,\vec{x}\rangle, w_j) = 1 \leftrightarrow C_j(y,b,\vec{v},\vec{x})].$$

Here $\langle y,x_1,\ldots,x_k\rangle$ denotes the Gödel number of the sequence of integers. Sequences are coded in an efficient manner [1]; in particular, there is a term $r(b,\vec{v})$ so that if $y \le |b|$ and if for all j, $x_j \le |t_j|$, then $\langle y,x_1,\ldots,x_k\rangle \le |r|$. Given these $w_1,\ldots,w_k$, the formula $C(y,b,\vec{v})$ is actually equivalent to a Δ_1^b-formula using the w_j's as parameters. Clearly, S_2^1 proves Δ_1^b-LIND since a Δ_1^b-formula is by definition provably equivalent to a Σ_1^b-formula. Thus it follows that S_2^i proves $\mathrm{LIND}_C(y)$ and hence $\mathrm{PIND}_A(b)$. □

The theory $\mathrm{I}\Sigma_n$ is the fragment of Peano arithmetic axiomatized by a simple base theory plus induction on Σ_n formulae (see Paris-Kirby [7]). It is well-known that $\mathrm{I}\Sigma_n$ proves induction for formulae in $\Sigma_{n+1} \cap \Pi_{n+1}$; indeed, the proof is very similar to the above proofs (although our proof of Theorem 1 seems to necessarily be slightly more complicated than the analogous proof for $\mathrm{I}\Sigma_n$). However there seems to be no way to apply the proofs of Theorems 1 and 2 to the theories T_2^i. In fact, T_2^i does prove $(\Sigma_{i+1}^b \cap \Pi_{i+1}^b)$-IND; the proof is presented below. But first we shall show that T_2^i proves induction for Boolean combinations of Σ_i^b-formulae (mostly because this has a short elegant proof).

Theorem 3 ($i \ge 1$). *Suppose* $T_2^i \vdash \mathit{MIN}_{\neg A}$ *and* $T_2^i \vdash \mathit{IND}_B$. *Then* $T_2^i \vdash \mathit{IND}_{A\wedge B}$.

Proof Let $A(b,\vec{v})$ and $B(b,\vec{v})$ have the indicated free variables and let $\mathrm{HYP}_{A\wedge B}$ be the hypothesis of $\mathrm{IND}_{A\wedge B}$, namely, the formula

$$A(0,\vec{v}) \wedge B(0,\vec{v}) \wedge (\forall x)[A(x,\vec{v}) \wedge B(x,\vec{v}) \to A(x+1,\vec{v}) \wedge B(x+1,\vec{v})].$$

[§] Louise Hay and the author [2] have strengthened this to show that every A in $\Sigma_{i+1}^b \cap \Pi_{i+1}^b$ is equivalent to a formula of the form $B = (\exists x \le |t|)(A_1 \wedge \neg A_2)$ where A_1 and A_2 are Σ_i^b-formulae. The equivalence of A and B is provable in S_2^i.

It is easy to see that

$$T_2^i \vdash (\forall x < a)A(x,\vec{v}) \land \mathrm{HYP}_{A\land B}(\vec{v}) \to (\forall x \le a)B(x,\vec{v})$$

since $T_2^i \vdash \mathrm{IND}_B$. So it will suffice to show that T_2^i proves $\mathrm{HYP}_{A\land B}(\vec{v}) \to (\forall x)A(x,\vec{v})$. Let us argue informally in T_2^i: suppose $\mathrm{HYP}_{A\land B}(\vec{v})$ but $(\exists x)(\neg A(x,\vec{v}))$. By $\mathrm{MIN}_{\neg A}$ there is a minimum a such that $\neg A(a,\vec{v})$. Thus $(\forall x \le a)B(x,\vec{v})$. In particular, $A(a \dot- 1,\vec{v})$ and $B(a \dot- 1,\vec{v})$ both hold. But now by $\mathrm{HYP}_{A\land B}$ we have that $A(a,v)$ holds, which is a contradiction. □

It is a corollary of Theorem 3 and of a result of Hausdorff that T_2^i proves induction for Boolean combinations of Σ_i^b-formulae:

Corollary 4 $(i \ge 1)$. *Suppose A is a Boolean combination of Σ_i^b-formulae. Then $T_2^i \vdash IND_A$.*

Proof By Hausdorff's characterization of Boolean combinations into a difference hierarchy [3], A is tautologically equivalent to a formula of the form

$$A_1 \land \neg(A_2 \land \neg(A_3 \land \cdots \neg(A_{k-1} \land \neg A_k)\cdots))$$

where each $A_j \in \Pi_i^b$. Let $\mathbb{A}_k$ be the set of formulae which are tautologically equivalent to a formula in this form. We prove by induction on k that T_2^i proves IND_A for every $A \in \mathbb{A}_k$. This is already known for $k=1$ since $\mathbb{A}_1$ is Π_i^b. Now suppose $T_2^i \vdash \mathbb{A}_k\text{-IND}$. First we show that if A is an arbitrary formula in $\mathbb{A}_k$ then $T_2^i \vdash \mathrm{IND}_{\neg A}(b)$. Well this can be done by letting $B(a,b,\vec{v})$ be the formula $A(b \dot- a,\vec{v})$ and using IND_B. Since subtraction is a Σ_1^b-defined function symbol of T_2^i the formula B can be picked to be a formula in $\mathbb{A}_k$; hence $T_2^i \vdash \mathrm{IND}_B(a)$. Now let D be a formula in $\mathbb{A}_{k+1}$ of the form $C \land \neg A$ where C is a Π_i^b-formula. It is known that $T_2^i \vdash \Sigma_i^b\text{-MIN}$ (see [1]), so by Theorem 3, $T_2^i \vdash \mathrm{IND}_D$. □

Interestingly, the methods of proof of Theorem 3 and Corollary 4 do apply to the theories S_2^i and $\mathrm{I}\Sigma_n$. This gives an alternative proof that S_2^i (respectively, $\mathrm{I}\Sigma_n$) proves PIND_A (respectively, IND_A) for A a Boolean combination of Σ_i^b-formulae (respectively, Σ_n-formulae). Of course this is not as strong as Theorems 1 and 2 above.

We are now ready to state our main theorems.

Theorem 5 $(i \ge 1)$. *Suppose $A(\vec{v})$ is a Σ_{i+1}^b-formula and that $S_2^{i+1} \vdash A(\vec{v})$. Then $T_2^i \vdash A(\vec{v})$.*

In other words, S_2^{i+1} is $\forall\Sigma_{i+1}^b$-conservative over T_2^i.

By a well-known theorem of Parikh's, Theorem 5 implies that S_2^{i+1} is $\forall\exists\Sigma_{i+1}^b$-conservative over T_2^i.

Corollary 6 *If $i \geq 1$ then $T_2^i \vdash \Delta_{i+1}^b$-IND. Hence $T_2^i \vdash (\Sigma_{i+1}^b \cap \Pi_{i+1}^b)$-IND.*

Proof of Corollary 6 from Theorem 5:

Let A be Δ_{i+1}^b with respect to T_2^i. This means there is a Σ_{i+1}^b-formula A_Σ and a Π_{i+1}^b-formula A_Π which are T_2^i-provably equivalent to A. The IND axiom for A can be reexpressed as

$$A_\Pi(0) \wedge (\forall x < b)(A_\Sigma(x) \rightarrow A_\Pi(Sx)) \rightarrow A_\Sigma(b).$$

This is a Σ_{i+1}^b-formula and since $S_2^{i+1} \vdash \Delta_{i+1}^b$-IND by Theorem 2.22 of [1], it is a consequence of S_2^{i+1}. Hence by Theorem 5 it is a consequence of T_2^i. □

Theorem 7 ($i \geq 1$). *Suppose $A(\vec{v})$ is a Boolean combination of Σ_{i+1}^b-formulae and $S_2^{i+1} \vdash (\forall\vec{v})A(\vec{v})$. Then $T_2^i + \Sigma_{i+1}^b$-replacement $\vdash (\forall\vec{v})A(\vec{v})$.*

In other words, S_2^{i+1} is conservative over $T_2^i + \Sigma_{i+1}^b$-replacement with respect to Boolean combinations of Σ_{i+1}^b-formulae.

Again, Parikh's theorem and Theorem 7 imply that S_2^{i+1} is conservative over T_2^i with respect to $\forall\exists\mathcal{B}(\Sigma_{i+1}^b)$-formulas, where $\mathcal{B}(\Sigma_{i+1}^b)$ is the set of Boolean combinations of Σ_{i+1}^b-formulas. We give the proofs of Theorems 5 and 7 in the next three sections.

3 $\square_{i+1}^p$-functions are definable by T_2^i.

The main theorem of [1] showed that the Σ_{i+1}^b-definable functions of S_2^{i+1} are precisely the $\square_{i+1}^p$ functions. This together with the conservation result, Theorem 5, which is proved below implies that the Σ_{i+1}^b-definable functions of T_2^i are also precisely the $\square_{i+1}^p$ functions. However, in order to prove the conservation result we will first prove directly that every $\square_{i+1}^p$ function is Σ_{i+1}^b-definable in T_2^i. (The converse, that every Σ_{i+1}^b-definable function of T_2^i is in $\square_{i+1}^p$, follows by the same result for the stronger theory S_2^{i+1}.)

Theorem 8 ($i \geq 1$): *Suppose $U(a, b, \vec{v})$ is a Σ_i^b-formula and $s(\vec{v})$ is a term. The following is a theorem of T_2^i:*

$$(\exists w)(\forall j \leq |s|)[Bit(j, w) = 1 \leftrightarrow U(LSP(w, j), j, \vec{v})].$$

Recall that $LSP(w,j)$ is the Σ^b_i-defined function of S^1_2 which is equal to $w \bmod 2^j$. Hence the above formula specifies the value of $Bit(j,w)$ as a Σ^b_i-predicate of the j lower order bits of w.

Proof The idea of the proof is to use the Σ^b_i-MIN axioms to get such a w. However, instead of minimizing w, we must minimize the complement of the bit-reversal of w. So let $Flip_s(w,\vec{v})$ be the function such that for all w,

$$|Flip_s(w,\vec{v})| \le |s(\vec{v})| + 1$$

and

$$(\forall j \le |s|)(Bit(j, Flip_s(w,\vec{v})) = 1 \dot{-} Bit(|s| \dot{-} j, w))\,.$$

Clearly $Flip_s$ is a polynomial time function and using techniques from Chapter 2 of [1], it can easily be Σ^b_1-defined in S^1_2.

Now let $B(u,a,\vec{v})$ be the formula

$$(\forall j \le |s|)[Bit(j, Flip_s(u,\vec{v})) = 1 \rightarrow U(LSP(Flip_s(u,\vec{v}),j),j,\vec{v})].$$

So B is a Σ^b_i-formula and $T^i_2 \vdash MIN_B(u)$. Let us argue informally in T^i_2: there exists a u such that $B(u,a,\vec{v})$, namely, $u = 2^{|s|+1} \dot{-} 1$; hence there exists a minimal such u. Given a minimal such u, we claim that $w = Flip_s(u,\vec{v})$ satisfies the desired condition of Theorem 8. Clearly for all j, if $Bit(j,w) = 1$ then $U(LSP(w,j),j,\vec{v})$ holds. So it suffices to show that if $Bit(j,w) = 0$ then $\neg U(LSP(w,j),j,\vec{v})$. Suppose not, we claim that changing the bit at the 2^j position in w gives a smaller u satisfying B; more precisely, let $w^* = 2^j + LSP(w,j)$ and let $u^* = Flip_s(w^*,\vec{v})$. So $u^* < u$ and $B(u^*,a,\vec{v})$ holds by our supposition. But this contradicts the minimality of u. $\Box$

Theorem 9 $(i \ge 1)$. *Let $k \in \Box^p_{i+1}$. Then T^i_2 can Σ^b_{i+1}-define k.*

Proof Let M be a deterministic Turing machine with an oracle for a Σ^p_i predicate Ω which computes $k(x)$ in time bounded by a polynomial p. We first claim that T^i_2 can prove that for all x there exists a w such that $Bit(j,w) = 1$ if and only if the j-th oracle query of M on input yields a "yes" answer. Of course, the w will be defined as in Theorem 8. Towards this end, let $f(w,j,x)$ be the polynomial time function computed as follows:

> $f(w,j,x)$ simulates M on input x for up to $p(|x|)$ steps. When the $(n+1)$-st oracle query is made (for $n < j$), the simulation uses $Bit(n,w)$ as the oracle's answer. When either $p(|x|)$ steps

have elapsed or just as the $(j+1)$-st query is to be attempted, the simulation terminates and $f(w,j,x)$ outputs the Gödel number of the final instantaneous description (ID) of the simulation.

We also define $g(v)$ to be the polynomial time function which accepts as input a Gödel number of an ID of M and outputs the value on the query tape of M (i.e., outputs the number which is ready to be used as a query to the oracle). Finally, $h(v)$ also accepts as input an ID of M, but outputs the value on the output tape of M. The defining equation, $DEF_\Omega(w,u)$, of w can now be given as

$$(\forall j < p(|x|))[Bit(j,w) = 1 \leftrightarrow \Omega(g(f(LSP(w,j),j,x)))]$$

with $\Omega(\cdots)$ a Σ_i^b-formula. By Theorem 8, $T_2^i \vdash (\exists w) DEF_\Omega(w,x)$ so T_2^i can Σ_{i+1}^b-define k by proving

$$(\exists y \le 2^{p(|x|)})(\exists w \le 2^{p(|x|)})(DEF_\Omega(w,x) \wedge y = h(f(w,p(|x|),x))).$$

Of course, the $y = k(x)$ can readily be proved to be unique. □

The proof of Theorem 9 gave a very special kind of Σ_{i+1}^b-definition for k; we call this a *Q_i-definition*:

Definition *A theory R can Q_i-*define *the function $f(\vec{x})$ if and only if there is a Σ_i^b-formula $U(w,j,\vec{x})$, a term $t(\vec{x})$ and a Σ_1^b-defined function f^* of S_2^1 such that $R \vdash (\forall x)(\exists w) DEF_{U,t}(w,\vec{x})$, where $DEF_{U,t}$ is the formula*

$$(\forall j < |t|)[Bit(j,w) \leftrightarrow U(LSP(w,j),j,\vec{x})],$$

and such that, for all $\vec{n}, w \in \mathbb{N}$, if $DEF_{U,t}(w,\vec{n})$ then $f(\vec{n}) = f^*(w,\vec{n})$.

The letter "Q" stands for "query" and the idea (as in the above proof) is that a function is Q_i-definable if and only if it is computable by a polynomial time Turing machine with a Σ_i^p-oracle. Note that every Q_i-definable function is Σ_{i+1}^b-definable by the definition. Also, by Corollary 10 and the main theorem of [1], every Σ_{i+1}^b-definable function of S_2^{i+1} is Q_i-definable by T_2^i.

Corollary 10 *Every $\square_{i+1}^p$-function is Q_i-definable by T_2^i (and conversely).*

Proof This was what the proof of Theorem 9 showed. The converse follows from the fact that the Σ_{i+1}^b-definable functions of S_2^{i+1} are precisely the $\square_{i+1}^p$-functions. □

In spite of the fact that in T^i_2, the notions of Q_i-definable and Σ^b_{i+1}-definable coincide, we must work extensively with the Q_i-definable functions. The reason is that there is no a priori reason why the notions provably coincide — for instance, given a Σ^b_{i+1}-definable function T^i_2 is there a Q_i-defined function which is T^i_2-provably the same function? The answer is yes, but it will take a lot of work to show it. Also, the author knows no simple proof of the (true) variant of Theorem 11 concerning Σ^b_{i+1}-definable functions.

Theorem 11 $(i \geq 1)$
(a) *Suppose g and h are Q_i-defined by T^i_2. Then there is a Q_i-defined function f such that T^i_2 can prove $(\forall\vec{x})(f(\vec{x}) = g(h(\vec{x}),\vec{x}))$.*

(b) *Suppose $g(\vec{x})$ and $h(y,z,\vec{x})$ are Q_i-defined by T^i_2 and let t be a term. Then there is a Q_i-defined function f such that T^i_2 can prove that for all $\vec{x}$ and all $y \neq 0$,*

$$f(0,\vec{x})=\min\{g(\vec{x}),t(0,\vec{x})\}$$
$$f(y,\vec{x})=\min\{h(y,f(\lfloor\tfrac{1}{2}y\rfloor,\vec{x}),\vec{x}),t(y,\vec{x})\}.$$

In other words, the Q_i-definable functions are provably closed under composition and limited iteration. (This gives a second proof of Corollary 10, using the alternate definition of polynomial time functions via composition and limited iteration instead of via Turing machines.)

Proof Let g and h have Q_i-definitions via Σ^b_i-formulae V and W, terms r and s, and Σ^b_1-defined functions g^* and h^* of S^1_2, respectively. So $g(\vec{x}) = y$ is true if and only if $(\exists w)(DEF_{V,r}(w,\vec{x}) \wedge g^*(w,\vec{x}) = y)$ and similarly for h. Also, let r^* and s^* be terms which dominate g and h so that $r^* > g$ and $s^* > h$ are provable by T^i_2. Without loss of generality, assume r, s, r^* and s^* are increasing in each of their variables.

To define f by composition, let the term $t(\vec{x})$ be equal to

$$2^{|r(s^*(\vec{x}),\vec{x})| + |s(\vec{x})|}$$

and let $U(w,j,\vec{x})$ be the Σ^b_i-formula

$$[j < |s(\vec{x})| \rightarrow V(w,j,\vec{x})] \wedge$$
$$\wedge [j \geq |s(\vec{x})| \rightarrow W(MSP(w,|s(\vec{x})|), j \dot{-} |s(\vec{x})|, h^*(w,\vec{x}),\vec{x})].$$

(Recall that $MSP(w,j)$ is defined to the equal to the integer part of $w/2^j$.) Now let $f^*(w,\vec{x})$ be $g^*(MSP(w,|s(\vec{x})|), h^*(w,\vec{x}),\vec{x})$. It is clear that f is properly Q_i-defined by

$$f(\vec{x}) = y \leftrightarrow (\exists w \le 2^{|t|})(DEF_{U,t}(w,\vec{x}) \wedge f^*(w,\vec{x}) = y)$$

and that T^i_2 proves that f is formed by composition from g and h.

(b) is proved with a similar but more complicated construction. The idea is that for the Q_i-definition of f we make w be the concatenation of the w's required for the computation of g and the repeated computations of h. Towards this end we define simultaneously functions $k(n,y,\vec{x})$, $f^{**}(n,w,y,\vec{x})$ and $E(n,w,y,\vec{x})$ by iteration on $n = 0,\ldots,|y|$. This will be done so that for an appropriate w, $f^{**}(n,w,y,\vec{x})$ is equal to $f(MSP(y,|y| \dot{-} n),\vec{x})$, i.e., the n-th intermediate result in the calculation of $f(y,\vec{x})$. The $k(n,y,\vec{x})$ will denote the first bit position of w for coding "oracle answers" for the computation of $f^{**}(n,w,y,\vec{x})$ from $f^{**}(n-1,w,y,\vec{x})$ and $E(n,w,y,\vec{x})$ will be the substring of w consisting of the oracle answers for the computation of $f^{**}(n,w,y,\vec{x})$. We define, for $n \le |y|$,

$$\begin{aligned}
k(0,y,\vec{x}) &= 0\\
E(0,w,y,\vec{x}) &= w\\
f^{**}(0,w,y,\vec{x}) &= \min\{g^*(w,\vec{x}), t(0,\vec{x})\}\\
k(1,y,\vec{x}) &= |r(\vec{x})|\\
k(n+2,y,\vec{x}) &= k(n+1,y,\vec{x}) + |s^*(y,t(y,\vec{x}),\vec{x})|\\
E(n,w,y,\vec{x}) &= MSP(w,k(n,y,\vec{x}))\\
f^{**}(n+1,w,y,\vec{x}) &= \min\{t(MSP(y,|y| \dot{-} (n+1)),\vec{x}),\\
&\qquad h^*(E(n+1,w,y,\vec{x}), MSP(y,|y| \dot{-} (n+1)),\\
&\qquad\qquad f^{**}(n,w,y,\vec{x}),\vec{x})\}
\end{aligned}$$

For larger values of n, we may define k, E and f^{**} arbitrarily. It is clear that k, E, and f^{**} can be Σ^b_1-defined by S^1_2. We define $U(w,j,y,\vec{x})$ to be

$$\begin{aligned}
&[j < k(1,y,\vec{x}) \rightarrow V(w,j,\vec{x})] \wedge\\
&\qquad \wedge (\forall n < |y|)[k(n+1,y,\vec{x}) \le j < k(n+2,y,\vec{x}) \rightarrow\\
&\qquad\qquad \rightarrow W(E(n+1,w,y,\vec{x}), j \dot{-} k(n+1,y,\vec{x}), f^{**}(n,w,y,\vec{x}),\vec{x})]
\end{aligned}$$

and choose the term $v(y,\vec{c})$ to bound $k(|y|+1,y,\vec{x})$ and set $f^*(w,y,\vec{x}) = f^{**}(|y|,w,y,\vec{x})$. It is now straightforward to check that U, v and f^* provide a Q_i-definition of f which is provably formed by limited iteration from g and h. □

The next theorem shows that minimization for Σ_i^b-formulae can be Q_i-defined in T_2^i. This is needed for the proof of Theorem 17 below.

Theorem 12 $(i \geq 1)$. *Let $A(a, \vec{v})$ be a Σ_i^b-formula. Then there is a Q_i-defined function f such that*

$$T_2^i \vdash (\exists x \leq z) A(x, \vec{v}) \rightarrow A(f(z, \vec{v}), \vec{v}) \wedge (\forall y < f(z, \vec{v}))(\neg A(y, \vec{v})).$$

Proof Let $U(w, j, z, \vec{v})$ be the Σ_i^b-formula

$$(\exists x < 2^{|z| \dot{-} j})[A(x + Flip_z(w, z), \vec{v})].$$

Let the term $t(z, \vec{v}) = z$ and let $f^*(w, z, \vec{v}) = Flip_z(w, z)$. The reader may check that the desired function f is Q_i-defined by U, t and f^*; note that the idea of computing f is to do a binary search for the least x such that $A(x, \vec{v})$ holds. □

The function $f(z, \vec{v})$ of Theorem 12 is denoted by $(\mu x \leq z) A(x, \vec{v})$.

4 The *Witness* Formula

We next review briefly a definition from [1] which is necessary for the proof of the main theorems. Let $i \geq 1$ be fixed, and let $A(\vec{a})$ be a Σ_i^b-formula. A formula $Witness_A^{i,\vec{a}}(w, \vec{a})$ is defined which has quantifier complexity less than that of A and which states that w is a number "witnessing" the truth of $A(\vec{a})$.

Definition *Suppose $i \geq 1$ and $A(\vec{a}) \in \Sigma_i^b$ and $\vec{a}$ is a vector of variables including all those free in A. The formula $Witness_A^{i,\vec{a}}$ is defined below, inductively on the complexity of A:*

(1) *If $A \in \Sigma_{i-1}^b \cup \Pi_{i-1}^b$ then $Witness_A^{i,\vec{a}}$ is just A itself.*

(2) *If A is $B \wedge C$ then define*

$$Witness_A^{i,\vec{a}}(w, \vec{a}) \iff Witness_B^{i,\vec{a}}(\beta(1, w), \vec{a}) \wedge Witness_C^{i,\vec{a}}(\beta(2, w), \vec{a}).$$

(3) *If A is $B \vee C$ then define*

$$Witness_A^{i,\vec{a}}(w, \vec{a}) \iff Witness_B^{i,\vec{a}}(\beta(1, w), \vec{a}) \vee Witness_C^{i,\vec{a}}(\beta(2, w), \vec{a}).$$

(4) *If A is $B \to C$ then we define*

$$\mathit{Witness}_A^{i,\vec{a}}(w,\vec{a}) \iff \mathit{Witness}_{\neg B}^{i,\vec{a}}(\beta(1,w),\vec{a}) \vee \mathit{Witness}_C^{i,\vec{a}}(\beta(2,w),\vec{a}).$$

(5) *If $A \notin \Sigma_{i-1}^b \cup \Pi_{i-1}^b$ and $A(\vec{a})$ is $(\forall x \le |s(\vec{a})|)B(\vec{a},x)$ then define*

$$\begin{aligned}\mathit{Witness}_A^{i,\vec{a}}(w,\vec{a}) \iff & \mathit{Seq}(w) \wedge \mathit{Len}(w) = |s(\vec{a})| + 1 \wedge \\ & \wedge (\forall x \le |s(\vec{a})|)\, \mathit{Witness}_{B(\vec{a},b)}^{i,\vec{a},b}(\beta(x+1,w),\vec{a},x).\end{aligned}$$

In words, w witnesses $A(\vec{a})$ if $w = \langle w_0,\ldots,w_{|s|}\rangle$ and each w_i witnesses $B(\vec{a},i)$. The formula $\mathit{Seq}(w)$ says w is a valid Gödel number of a sequence and $\mathit{Len}(w)$ is a function giving the number of entries in the sequence w.

(6) *If $A \notin \Sigma_{i-1}^b \cup \Pi_{i-1}^b$ and A is $(\exists x \le t(\vec{a}))B(\vec{a},x)$ then define*

$$\begin{aligned}\mathit{Witness}_A^{i,\vec{a}}(w,\vec{a}) \iff & \mathit{Seq}(w) \wedge \mathit{Len}(w) = 2 \wedge \beta(1,w) \le t(\vec{a}) \wedge \\ & \wedge \mathit{Witness}_{B(\vec{a},b)}^{i,\vec{a},b}(\beta(2,w),\vec{a},\beta(1,w)).\end{aligned}$$

So w witnesses $A(\vec{a})$ if $w = \langle n,v\rangle$ where $n \le t(\vec{a})$ and v witnesses $B(\vec{a},n)$.

(7) *If $A \notin \Sigma_{i-1}^b \cup \Pi_{i-1}^b$ and A is $\neg B$ then use prenex operations to push the negation sign "into" the formula so that it can be handled by cases (1)–(6).*

The purpose of defining *Witness* is to give a canonical way of verifying that $A(\vec{a})$ is true. It is easy to see that $(\exists w)\,\mathit{Witness}_A^{i,\vec{a}}(w,\vec{a})$ is equivalent to $A(\vec{a})$. The next propositions express some properties of *Witness*; these are proved mostly by induction on the complexity of A.

Proposition 13 *For $i \ge 1$, and $A \in \Sigma_i^b$, $\mathit{Witness}_A^{i,\vec{a}}$ is a Δ_i^b-formula with respect to S_2^1. If $i \ge 2$ then it is in fact either a Σ_{i-1}^b-formula or a Π_{i-1}^b formula.*

Proposition 14 $(i \ge 1)$. *Let $A(\vec{a})$ be a Σ_i^b-formula. then there is a term $t_A(\vec{a})$ such that*

$$S_2^i \vdash A(\vec{a}) \leftrightarrow (\exists w \le t_A)\,\mathit{Witness}_A^{i,\vec{a}}(w,\vec{a}).$$

Also there is a Σ_1^b-defined function $g_A(w)$ such that

$$S_2^1 \vdash \mathit{Witness}_A^{i,\vec{a}}(w,\vec{a}) \to \mathit{Witness}_A^{i,\vec{a}}(g_A(w),\vec{a}) \wedge g_A(w) \le t_A.$$

Proposition 15 $(i \geq 1)$. *Let A be a Σ_i^b-formula. The predicate represented by $Witness_A^{i,\vec{a}}$ is a Δ_i^p-predicate.*

The above propositions are proved in [1]. We shall also need the following strengthened version of Proposition 14:

Proposition 16 $(i \geq 1)$. *Let $A(\vec{a})$ be a Σ_{i+1}^b-formula. Then:*

(a) $T_2^i \vdash (\exists w)\, Witness_A^{i+1,\vec{a}}(w, \vec{a}) \rightarrow A(\vec{a})$.

(b) *There is a term t_A so that*

$$T_2^i + \Sigma_{i+1}^b\textit{-replacement} \vdash A(\vec{a}) \rightarrow (\exists w \leq t_A)\, Witness_A^{i+1,\vec{a}}(w, \vec{a}).$$

This is easily proved by induction on the complexity of A. Note that for the proof of part (b) the Σ_{i+1}^b-replacement axiom is exactly what we need to handle Case (5) of the definition of the *Witness* formula.

5 The Main Proof

In this section the proofs of Theorems 5 and 7 are given. The arguments are proof-theoretic and hence constructive; however, they use cut elimination and thus may not be feasibly constructive (since they involve superexponential growth rates). We use a Gentzen-style sequent calculus: each line in a proof is a *sequent* of the form

$$A_1, \ldots, A_k \longrightarrow B_1, \ldots, B_\ell$$

where each A_j and B_j is a formula. The intended meaning of this sequent is that the conjunction of the *antecedent* $A_1, \ldots, A_k$ implies the disjunction of the *succedent* $B_1, \ldots, B_\ell$. Note that the sequent connective symbol $\longrightarrow$ is distinct from the logical connective $\rightarrow$. Capital Greek letters $\Gamma, \Delta, \Pi, \Lambda, \ldots$ will be used to denote a series of formulae separated by commas, these are called *cedents.*

There are about 23 rules of inference for the sequent calculus; in addition, there are induction rules which replace the induction axioms. The *initial sequents* (i.e., axioms) of a sequent calculus proof must be equality axioms, logical axioms or non-logical axioms. The theories S_2^i and T_2^i all have the same set of non-logical axioms; namely, a finite set of open (i.e., quantifier free) sequents. There are no induction axioms as initial sequents since induction rules are used instead. An important theorem (due to Gentzen) concerning the sequent calculus is that many instances of the cut rule may be eliminated from proofs — more precisely, all *free cuts* may be eliminated from a proof. Rather than define precisely what a free cut is, let us merely say that for a proof of a Σ_i^b-formula in a theory S_2^i or T_2^i, we may assume that every formula appearing in the proof is a Σ_i^b- or a Π_i^b-formula. For more information on the sequent calculus for theories of Bounded Arithmetic, consult chapter 4 of [1] and the references cited there.

If T is a cedent we write $\bigwedge\Gamma$ and $\bigvee\Gamma$ to denote the conjunction and disjunction, respectively, of the formulae in Γ. Conjunction and disjunction associate from right to left; for example, if Γ is A, B, C then $\bigwedge\Gamma$ denotes $A \wedge (B \wedge C)$.

We have already mentioned the function and predicate symbols β, *Seq*, and *Len* which manipulate Gödel numbers of sequences. We use $\langle a_1, \ldots, a_n\rangle$ to denote the Gödel number of the sequence $a_1, \ldots, a_n$. Also, $*$ is a binary function defined so that

$$\langle a_1, \ldots, a_n\rangle * a_{n+1} = \langle a_1, \ldots, a_n, a_{n+1}\rangle.$$

Finally $\langle\!\langle a_1, \ldots, a_n\rangle\!\rangle$ is equal to $\langle a_1, \langle a_2, \ldots, \langle a_{n-1}, a_n\rangle \ldots\rangle\,\rangle$.

These conventions allow us to conveniently discuss witnessing a cedent. For example, suppose Γ is $A_1, \ldots, A_n$ and that $w = \langle\!\langle w_1, \ldots, w_n\rangle\!\rangle$. Then $\mathit{Witness}^{i,\vec{a}}_{\bigwedge\Gamma}(w, \vec{a})$ holds if and only if $\mathit{Witness}^{i,\vec{a}}_{A_j}(w_j, \vec{a})$ holds for all $1 \le j \le n$.

Instead of proving Theorems 5 and 7 directly, we prove a stronger theorem:

Theorem 17 $(i \ge 1)$. *Suppose the sequent $\Gamma, \Pi \longrightarrow \Delta, \Lambda$ is a theorem of S_2^{i+1} and each formula in $\Gamma \cup \Delta$ is Σ_{i+1}^b and each formula in $\Pi \cup \Lambda$ is Π_{i+1}^b. Let $c_1, \ldots, c_p$ be the free variables in the sequent and let G and H be the formulae*

$$G = \left(\bigwedge\Gamma\right) \wedge \bigwedge\{\neg C : C \in \Lambda\}$$

and

$$H = \left(\bigvee\Delta\right) \vee \bigvee\{\neg C : C \in \Pi\}.$$

Then there is a Q_i-defined function f of T_2^i such that

$$T_2^i \vdash \mathit{Witness}^{i+1,\vec{c}}_G(w, \vec{c}) \rightarrow \mathit{Witness}^{i+1,\vec{c}}_H(f(w, \vec{c}), \vec{c}).$$

Proof of Theorem 5 from Theorem 17. Let $A(\vec{c})$ be a Σ^b_{i+1}-formula which is provable in S^{i+1}_2. By Theorem 17, $T^i_2 \vdash \mathit{Witness}^{i,\vec{c}}_A(f(\vec{c}),\vec{c})$ for some Q_i-defined function f. By Proposition 16(a) $T^i_2 \vdash A(\vec{c})$. $\Box$

Proof of Theorem 7 from Theorem 17. Let $A(\vec{c})$ be a Boolean combination of Σ^b_{i+1}-formulae which is provable by S^{i+1}_2. Thus A is tautologically equivalent to a conjunction of disjunctions $\bigwedge_j \bigvee_k A_{jk}$ with each A_{jk} a Σ^b_{i+1}- or a Π^b_{i+1}-formula. Hence S^{i+1}_2 proves each disjunct $\bigvee_k A_{jk}$. Fix a value for j and let Δ_j be the cedent containing the Σ^b_i-formulae among A_{jk} and let Λ_j be the rest of the A_{jk}'s. Hence S^{i+1}_2 proves the sequent $\longrightarrow \Delta_j, \Lambda_j$. Let G be the formula $\neg(\bigvee \Lambda_j)$ and let H be $\bigvee \Delta_j$. By Theorem 17

$$T^i_2 \vdash (\exists w)\, \mathit{Witness}^{i+1,\vec{c}}_G(w,\vec{c}) \rightarrow (\exists w)\, \mathit{Witness}^{i+1,\vec{c}}_H(w,\vec{c}).$$

By Proposition 16(b),

$$T^i_2 + \Sigma^b_{i+1}\text{-replacement} \vdash G(\vec{c}) \rightarrow H(\vec{c}).$$

Hence $T^i_2 + \Sigma^b_{i+1}$-replacement proves the sequent $\longrightarrow \Delta_j, \Lambda_j$, or equivalently, the formula $\bigvee_k A_{jk}$. Hence $A(\vec{c})$ is a consequence of $T^i_2 + \Sigma^b_{i+1}$-replacement. $\Box$

We next prove Theorem 17: the outline of the proof is identical to the proof of Theorem 5.5 of [1]. Indeed, this proof is a strengthened version of that proof.

Proof of Theorem 17:

By the free-cut elimination theorem there is a S^{i+1}_2-proof P of $\Gamma, \Pi \longrightarrow \Delta, \Lambda$ such that every cut in P has a Σ^b_{i+1} principal formula and such that P is in free variable normal form (see [1] for definitions). The proof of Theorem 17 is by induction on the number of sequents in the proof P.

To simplify notation we shall henceforth assume Π and Λ are the empty cedent. We can always fulfill this requirement by using ($\neg$:left) and ($\neg$:right) to move formulae from side to side and no essential cases are ignored under this assumption since each inference has a dual; for example, the dual of ($\exists \leq$:left) is ($\forall \leq$:right) and the dual of ($\wedge$:right) is ($\vee$:left).

To begin, consider the case where P has no inferences and consists of a single sequent. This sequent must be a nonlogical axiom of S^{i+1}_2 or a logical axiom or an equality axiom. In any event, it contains only atomic formulae and is also an axiom of T^i_2. For atomic formulae A, $\mathit{Witness}^{i+1,\vec{c}}_A$ is just A itself; hence this case is completely trivial.

The argument for the induction step splits into thirteen cases depending on the final inference of P.

Case (1): Suppose the last inference of P is ($\neg$:left) or ($\neg$:right). These are "cosmetic" inferences; see also the discussion above about assuming Π and Λ are empty.

Case (2): ($\wedge$:left). Suppose the last inference of P is:

$$\frac{B, \Gamma^* \longrightarrow \Delta}{B \wedge C, \Gamma^* \longrightarrow \Delta}$$

Let D be the formula $B \wedge (\bigwedge \Gamma^*)$ and let E be $(B \wedge C) \wedge (\bigwedge \Gamma^*)$. By the induction hypothesis, there is a Q_i-defined function symbol g of T^i_2 such that

$$T^i_2 \vdash \mathit{Witness}^{i+1,\vec{c}}_D(w, \vec{c}) \rightarrow \mathit{Witness}^{i+1,\vec{c}}_{\bigvee \Delta}(g(w, \vec{c}), \vec{c}).$$

Let h be the function defined by $h(w) = \langle \beta(1, \beta(1, w)), \beta(2, w) \rangle$ so that

$$T^i_2 \vdash \mathit{Witness}^{i+1,\vec{c}}_E(w, \vec{c}) \rightarrow \mathit{Witness}^{i+1,\vec{c}}_D(h(w), \vec{c})$$

follows immediately from the definition of *Witness*. Now let $f(w, \vec{c}) = g(h(w), \vec{c})$. By Theorem 11(a) and since h is Σ^b_1-defined by S^1_2, the function f is Q_i-defined by T^i_2. Also

$$T^i_2 \vdash \mathit{Witness}^{i+1,\vec{c}}_E(w, \vec{c}) \rightarrow \mathit{Witness}^{i+1,\vec{c}}_{\bigvee \Delta}(f(w, \vec{c}), \vec{c}),$$

so f fulfills the desired conditions.

Case (3): ($\vee$:left) Suppose the last inference of P is

$$\frac{B, \Gamma^* \longrightarrow \Delta \qquad C, \Gamma^* \longrightarrow \Delta}{B \vee C, \Gamma^* \longrightarrow \Delta}$$

Let D be the formula $B \wedge (\bigwedge \Gamma^*)$ and let E be $C \wedge (\bigwedge \Gamma^*)$ and let F be $(B \vee C) \wedge (\bigwedge \Gamma^*)$. By the induction hypothesis, there are Q_i-defined functions g and h such that

$$T^i_2 \vdash \mathit{Witness}^{i+1,\vec{c}}_D(w, \vec{c}) \rightarrow \mathit{Witness}^{i+1,\vec{c}}_{\bigvee \Delta}(g(w, \vec{c}), \vec{c})$$

and

$$T^i_2 \vdash \mathit{Witness}^{i+1,\vec{c}}_E(w, \vec{c}) \rightarrow \mathit{Witness}^{i+1,\vec{c}}_{\bigvee \Delta}(h(w, \vec{c}), \vec{c}).$$

Since $i \geq 1$, Proposition 13 states that $\mathit{Witness}^{i+1,\vec{c}}_{\bigvee\Delta}$ is either a Σ^b_i- or a Π^b_i-formula. Hence the function k defined by

$$k(w,a,b,\vec{c}) = \begin{cases} a & \text{if } \mathit{Witness}^{i+1,\vec{c}}_{\bigvee\Delta}(w,\vec{c}) \\ b & \text{otherwise} \end{cases}$$

is Q_i-defined. Let f be the function

$$f(w,\vec{c}) = k(\beta(1,\beta(1,w)), \overline{g}(w,\vec{c}), \overline{h}(w,\vec{c}), \vec{c})$$

where

$$\overline{g}(w,\vec{c}) = g(\langle\beta(1,\beta(1,w)),\beta(2,w)\rangle,\vec{c})$$

and

$$\overline{h}(w,\vec{c}) = h(\langle\beta(2,\beta(1,w)),\beta(2,w)\rangle,\vec{c}).$$

Since f is defined as the composition of Q_i-defined functions, f is itself Q_i-defined. Clearly f satisfies the desired conditions of Theorem 17.

Case (4): ($\exists\leq$:left). Suppose the last inference of P is

$$\frac{a \leq s, B(a), \Gamma^* \longrightarrow \Delta}{(\exists x \leq s)B(x), \Gamma^* \longrightarrow \Delta}$$

The free variable a is the *eigenvariable* and appears only as indicated. Let D be the formula $a \leq s \wedge (B(a) \wedge (\bigwedge \Gamma^*))$ and let E be $(\exists x \leq s)B(x) \wedge (\bigwedge \Gamma^*)$. By the induction hypothesis, there is a Q_i-defined function g such that

$$T^i_2 \vdash \mathit{Witness}^{i+1,\vec{c},a}_D(w,\vec{c},a) \rightarrow \mathit{Witness}^{i+1,\vec{c}}_{\bigvee\Delta}(g(w,\vec{c},a),\vec{c}).$$

(Note that the variable a can be omitted from the superscript in the right hand side of the implication since it does not appear free in Δ.)

This case splits into three subcases: first, if $(\exists x \leq s)B$ is not in $\Sigma^b_i \cup \Pi^b_i$, let h be the function Σ^b_1-defined by S^1_2 so that $h(w,\vec{c}) = \beta(1,\beta(1,w))$. Second, if $(\exists x \leq s)B \in \Sigma^b_i$, let $h(w,\vec{c}) = (\mu x \leq s)B(x,\vec{c})$; by Theorem 12, h is Q_i-defined by T^i_2. Third, if $(\exists x \leq s)B \in \Pi^b_i \setminus \Sigma^b_i$ then the quantifier $(\exists x \leq s)$ must be sharply bounded, so $h(w,\vec{c}) = (\mu x \leq s)B(x,\vec{c})$ is again Q_i-defined (to prove this, define h by limited iteration and use Theorem 11). In any case we have that

$$T^i_2 \vdash \mathit{Witness}^{i+1,\vec{c}}_E(w,\vec{c}) \rightarrow B(h(w,\vec{c}),\vec{c}) \wedge h(w,\vec{c}) \leq s(\vec{c})$$

and, indeed, that

$$T^i_2 \vdash \mathit{Witness}^{i+1,\vec{c}}_E(w,\vec{c}) \rightarrow \mathit{Witness}^{i+1,\vec{c},a}_B(\beta(2,\beta(1,w)),\vec{c},h(w,\vec{c})).$$

The desired Q_i-defined function $f(w,\vec{c})$ is given by

$$f(w,\vec{c}) = g(\langle\!\langle 0,\beta(2,\beta(1,w)),\beta(2,w)\rangle\!\rangle,\vec{c},h(w,\vec{c})).$$

Case (5): ($\forall \leq$:left). We omit the proof of this case, as it is fairly easy and exactly like case (5) of the proof of Theorem 5.5 of [1].

Case (6): ($\rightarrow$:left) and ($\rightarrow$:right). These cases are also omitted: they are very similar to ($\vee$:left) and ($\vee$:right).

Case (7): ($\vee$:right). This case is very simple; see Case (7) of Theorem 5.5 of [1].

Case (8): ($\wedge$:right). Suppose the last inference of P is

$$\frac{\Gamma \longrightarrow B, \Delta^* \qquad \Gamma \longrightarrow C, \Delta^*}{\Gamma \longrightarrow B \wedge C, \Delta^*}$$

Let D be the formula $B \vee (\bigvee \Delta^*)$, let E be $C \vee (\bigvee \Delta^*)$ and let F be $(B \wedge C) \vee (\bigvee \Delta^*)$. The induction hypothesis is that there are Q_i-defined functions g and h so that

$$T_2^i \vdash \mathit{Witness}_{\bigwedge \Gamma}^{i+1,\vec{c}}(w, \vec{c}) \rightarrow \mathit{Witness}_{D}^{i+1,\vec{c}}(g(w, \vec{c}), \vec{c})$$

and

$$T_2^i \vdash \mathit{Witness}_{\bigwedge \Gamma}^{i+1,\vec{c}}(w, \vec{c}) \rightarrow \mathit{Witness}_{E}^{i+1,\vec{c}}(h(w, \vec{c}), \vec{c}).$$

Let k be the function such that

$$k(v, w, \vec{c}) = \begin{cases} v & \text{if } \mathit{Witness}_{\bigvee \Delta^*}^{i+1,\vec{c}}(v, \vec{c}) \\ w & \text{otherwise.} \end{cases}$$

By Proposition 13, $\mathit{Witness}_{\bigvee \Delta^*}^{i+1,\vec{c}}$ is either a Σ_i^b- or a Π_i^b-formula; hence k is Q_i-defined. Let f be the function

$$f(w, \vec{c}) = \langle\, \langle \beta(1, g(w, \vec{c})), \beta(1, h(w, \vec{c})) \rangle, k(\beta(2, g(w, \vec{c})), \beta(2, h(w, \vec{c})), \vec{c}) \rangle.$$

By Theorem 11(a) f is Q_i-defined; furthermore, it is clear that

$$T_2^i \vdash \mathit{Witness}_{\bigwedge \Gamma}^{i+1,\vec{c}}(w, \vec{c}) \rightarrow \mathit{Witness}_{F}^{i+1,\vec{c}}(f(w, \vec{c}), \vec{c}).$$

Case (9): ($\exists \leq$:right). Suppose the last inference of P is

$$\frac{\Gamma^* \longrightarrow B(r), \Delta^*}{r \leq s, \Gamma^* \longrightarrow (\exists x \leq s) B(x), \Delta^*}$$

We assume $r \leq s$ is in Γ; a similar argument works for $r \leq s$ in Π. Let D be the formula $B(r) \vee (\bigvee \Delta^*)$, let E be $r \leq s \wedge (\bigwedge \Gamma^*)$ and let F

be $(\exists x \leq s)B(x) \vee (\bigvee \Delta^*)$. The induction hypothesis is that there is a Q_i-defined function g such that

$$T_2^i \vdash \mathit{Witness}^{i+1,\vec{c}}_{\bigwedge \Gamma^*}(w, \vec{c}) \rightarrow \mathit{Witness}^{i+1,\vec{c}}_{D}(g(w, \vec{c}), \vec{c})$$

By the definition of *Witness*,

$$T_2^i \vdash \mathit{Witness}^{i+1,\vec{c}}_{E}(w, \vec{c}) \rightarrow r \leq s \wedge \mathit{Witness}^{i+1,\vec{c}}_{\bigwedge \Gamma^*}(\beta(2, w), \vec{c}).$$

So define f by

$$f(w, \vec{c}) = \langle\, \langle r(\vec{c}), \beta(1, g(\beta(2, w), \vec{c}))\rangle, \beta(2, g(\beta(2, w), \vec{c}))\rangle.$$

By Theorem 11(a), f is Q_i-defined and clearly f satisfies the conditions of Theorem 17.

Case (10): ($\forall \leq$:right). Suppose the last inference of P is

$$\frac{a \leq s, \Gamma \longrightarrow B(a), \Delta^*}{\Gamma \longrightarrow (\forall x \leq s)B(x), \Delta^*}$$

The free variable a is the *eigenvariable* and must appear only as indicated. Let D be the formula $a \leq s \wedge (\bigwedge \Gamma)$, let E be $B(a) \vee (\bigvee \Delta^*)$ and let $F(\vec{c}, d)$ be $(\forall x \leq d)B(x) \vee (\bigvee \Delta^*)$. The induction hypothesis is that there is a Q_i-defined function g such that

$$T_2^i \vdash \mathit{Witness}^{i+1,\vec{c},a}_{D}(w, \vec{c}, a) \rightarrow \mathit{Witness}^{i+1,\vec{c},a}_{E}(g(w, \vec{c}, a), \vec{c}, a).$$

First, consider the case where $(\forall x \leq s)B(x)$ is not in $\Sigma_i^b \cup \Pi_i^b$; recall that all formulas are assumed to be in Σ_{i+1}^b. So $(\forall x \leq s)$ is sharply bounded and $s = |r|$ for some term r. Let k be the function defined by

$$k(v, w, \vec{c}) = \begin{cases} v & \text{if } \mathit{Witness}^{i+1,\vec{c}}_{\bigvee \Delta^*}(v, \vec{c}) \\ w & \text{otherwise.} \end{cases}$$

Since $\mathit{Witness}^{i+1,\vec{c}}_{\bigvee \Delta^*}$ is either a Σ_i^b- or a Π_i^b-predicate (by Proposition 13) k is Q_i-defined. Let $p(w, \vec{c}, d)$ be defined by limited iteration as:

$$\begin{aligned} p(w, \vec{c}, 0) &= \langle\, \langle \beta(1, g(w, \vec{c}, 0))\rangle, \beta(2, g(w, \vec{c}, 0))\rangle \\ p(w, \vec{c}, m) &= \langle \beta(1, p(w, \vec{c}, \lfloor \tfrac{1}{2} m \rfloor)) * \beta(1, g(w, \vec{c}, |m|)), \\ &\qquad k(\beta(2, p(w, \vec{c}, \lfloor \tfrac{1}{2} m \rfloor)), \beta(2, g(w, \vec{c}, |m|)), \vec{c})\rangle \end{aligned}$$

for all $m \neq 0$. By Theorem 11(b), p can be Q_i-defined and

$$T_2^i \vdash \mathit{Witness}_D^{i+1,\vec{c}}(w,\vec{c}) \to \mathit{Witness}_F^{i+1,\vec{c},d}(p(w,\vec{c},0),\vec{c},0)$$

and

$$T_2^i \vdash \mathit{Witness}_D^{i+1,\vec{c}}(w,\vec{c}) \land \mathit{Witness}_F^{i+1,\vec{c},d}(p(w,\vec{c},\lfloor \tfrac{1}{2}m \rfloor),\vec{c},|m| \dot{-} 1) \to \\ \to \mathit{Witness}_F^{i+1,\vec{c},d}(p(w,\vec{c},m),\vec{c},|m|).$$

We now wish to use induction on the length of m to deduce that

$$T_2^i \vdash \mathit{Witness}_D^{i+1,\vec{c}}(w,\vec{c}) \to \mathit{Witness}_F^{i+1,\vec{c},d}(p(w,\vec{c},r),\vec{c},s).$$

However, $\mathit{Witness}_F^{i+1,\vec{c},d}(p(w,\vec{c},m),\vec{c},|m|)$ is a Δ_{i+1}^b-formula (since p is a Σ_{i+1}^b-defined function) and we have not yet shown that T_2^i has induction for Δ_{i+1}^b-formulae. To circumvent this problem, define the function h so that

$$h(w,\vec{c},0)=\langle p(w,\vec{c},0)\rangle$$
$$h(w,\vec{c},m)=h(w,\vec{c},\lfloor \tfrac{1}{2}m \rfloor) * p(w,\vec{c},m)$$

for $m \neq 0$. Again by Theorem 11(b), h is Q_i-defined, say by V, q and h^*. Note that T_2^i proves

$$b = MSP(a,j) \to p(w,\vec{c},b) = \beta(|b|+1, h(w,\vec{c},a)).$$

Thus, the formula above can be re-expressed as

$$T_2^i \vdash DEF_{V,q}(w^*,w,\vec{c},t) \land 0 \leq j \land j < |t| \land \mathit{Witness}_D^{i+1,\vec{c}}(w,\vec{c}) \land \\ \land \mathit{Witness}_F^{i+1,\vec{c},d}(\beta(j+1,h^*(w^*,w,\vec{c},t)),\vec{c},j) \to \\ \to \mathit{Witness}_F^{i+1,\vec{c},d}(\beta(j+2,h^*(w^*,w,\vec{c},t)),\vec{c},j+1).$$

Since h^* is Σ_1^b-defined by S_2^1 and since $i \geq 1$,

$$\mathit{Witness}_F^{i+1,\vec{c},d}(\beta(j+1,h^*(w^*,w,\vec{c},t)),\vec{c},j)$$

is either a Σ_i^b- or a Π_i^b-formula by Proposition 13. Hence by Π_i^b-LIND, T_2^i can prove

$$\mathit{Witness}_D^{i+1,\vec{c}}(w,\vec{c}) \to \mathit{Witness}_F^{i+1,\vec{c},d}(f(w,\vec{c}),\vec{c},s)$$

where $f(w,\vec{c}) = p(w,\vec{c},r(\vec{c}))$. So f is Q_i-defined and it follows readily from the definition of $\mathit{Witness}$ that

$$T_2^i \vdash \mathit{Witness}_D^{i+1,\vec{c}}(w,\vec{c}) \to \mathit{Witness}_{F(\vec{c},s)}^{i+1,\vec{c}}(f(w,\vec{c}),\vec{c}).$$

Second, consider the case where $(\forall x \leq s)B(x)$ is a formula in $\Sigma_i^b \cup \Pi_i^b$. Similar to the argument in Case (4), T_2^i can Q_i-define the function $h(w, \vec{c}) = (\mu x \leq s)(\neg B(x))$. Now we let

$$f(w, \vec{c}) = \langle 0, \beta(2, g(\langle 0, w \rangle, \vec{c}, h(w, \vec{c}))\rangle.$$

It is easy to verify that f satisfies the desired conditions and thus this case is also done.

Case (11): *Cut.* Suppose the last inference of P is

$$\frac{\Gamma \longrightarrow B, \Delta \qquad B, \Gamma \longrightarrow \Delta}{\Gamma \longrightarrow \Delta}$$

By the assumption that P is free-cut free, B must be a Σ_{i+1}^b-formula. Let D be the formula $B \vee (\bigvee \Delta)$ and let E be $B \wedge (\bigwedge \Gamma)$. The induction hypothesis is that there are Q_i-defined functions g and h such that

$$T_2^i \vdash \mathit{Witness}_{\bigwedge\Gamma}^{i+1,\vec{c}}(w, \vec{c}) \rightarrow \mathit{Witness}_{D}^{i+1,\vec{c}}(g(w, \vec{c}), \vec{c})$$

and

$$T_2^i \vdash \mathit{Witness}_{E}^{i+1,\vec{c}}(w, \vec{c}) \rightarrow \mathit{Witness}_{\bigvee\Delta}^{i+1,\vec{c}}(h(w, \vec{c}), \vec{c}).$$

We define the function f so that

$$f(w, \vec{c}) = \begin{cases} \beta(2, g(w, \vec{c})) & \text{if } \mathit{Witness}_{\bigvee\Delta}^{i+1,\vec{c}}(\beta(2, g(w, \vec{c})), \vec{c}) \\ h(\langle \beta(1, g(w, \vec{c})), w \rangle, \vec{c}) & \text{otherwise.} \end{cases}$$

By Proposition 13, $\mathit{Witness}_{\bigvee\Delta}^{i+1,\vec{c}}$ is a Σ_i^b- or a Π_i^b-formula, hence f is Q_i-defined. Also, it is clear that

$$T_2^i \vdash \mathit{Witness}_{\bigwedge\Gamma}^{i+1,\vec{c}}(w, \vec{c}) \rightarrow \mathit{Witness}_{\bigvee\Delta}^{i+1,\vec{c}}(f(w, \vec{c}), \vec{c}).$$

Case (12): (Σ_{i+1}^b-PIND). Suppose the last inference of P is

$$\frac{B(\lfloor \frac{1}{2} a \rfloor), \Gamma^* \longrightarrow B(a), \Delta^*}{B(0), \Gamma^* \longrightarrow B(t), \Delta^*}$$

where a is the *eigenvariable* and must not appear in the lower sequent.

First consider the case where B is not in $\Sigma_i^b \cup \Pi_i^b$ and hence $B(0)$ is in Γ and $B(t)$ is in Δ. The general idea is to treat the Σ_{i+1}^b-PIND inference as if it were $|t| \dot{-} 1$ cuts. So, in effect, this case is handled by formally iterating the method of Case (11).

Let D be the formula $B(\lfloor\frac{1}{2}a\rfloor)\wedge(\bigwedge\Gamma^*)$, let $E(\vec{c},a)$ be $B(a)\vee(\bigvee\Delta^*)$, let F be $B(0)\wedge(\bigwedge\Gamma^*)$ and let A be $B(t)\vee(\bigvee\Delta^*)$. The induction hypothesis is that there is a Q_i-defined function g such that

$$T_2^i \vdash \mathit{Witness}_D^{i+1,\vec{c},a}(w,\vec{c},a) \rightarrow \mathit{Witness}_E^{i+1,\vec{c},a}(g(w,\vec{c},a),\vec{c},a).$$

Let k and h be the functions Q_i-defined so that

$$k(v,w,\vec{c}) = \begin{cases} v & \text{if } \mathit{Witness}_{\bigvee\Delta^*}^{i+1,\vec{c}}(v,\vec{c}) \\ w & \text{otherwise} \end{cases}$$
$$h(v,w,\vec{c},a) = g(\langle\beta(1,v),\beta(2,w)\rangle,\vec{c},a).$$

By Proposition 14 there is a term t_E and a Σ_1^b-defined function q of S_2^1 such that

$$T_2^i \vdash \mathit{Witness}_E^{i,\vec{c},a}(w,\vec{c},a) \rightarrow \mathit{Witness}_E^{i+1,\vec{c},a}(q(w),\vec{c},a) \wedge q(w) \leq t_E(\vec{c},a)$$

Define p by limited iteration so that

$$\begin{aligned} p(w,\vec{c},0) &= q(g(w,\vec{c},0)) \\ p(w,\vec{c},m) &= q(\langle\beta(1,h(p(w,\vec{c},\lfloor\tfrac{1}{2}m\rfloor),w,\vec{c},m)), \\ &\qquad k(\beta(2,h(p(w,\vec{c},\lfloor\tfrac{1}{2}m\rfloor),w,\vec{c},m)),\beta(2,p(w,\vec{c},\lfloor\tfrac{1}{2}m\rfloor)),\vec{c})\rangle) \end{aligned}$$

for all $m > 0$. This is a valid definition by limited iteration since the use of the q function gives a provable bound on the size of p; namely, $p(w,\vec{c},m) \leq t_E(\vec{c},m)$. Thus by Theorem 11(b), p is Q_i-defined by T_2^i. Now it is easy to see that

$$T_2^i \vdash \mathit{Witness}_F^{i+1,\vec{c}}(w,\vec{c}) \rightarrow \mathit{Witness}_E^{i+1,\vec{c},a}(p(w,\vec{c},0),\vec{c},0)$$

and

$$\begin{aligned} T_2^i \vdash \mathit{Witness}_F^{i+1,\vec{c}}(w,\vec{c}) \wedge \mathit{Witness}_E^{i+1,\vec{c},a}(p(w,\vec{c},\lfloor\tfrac{1}{2}a\rfloor),\vec{c},\lfloor\tfrac{1}{2}a\rfloor) \rightarrow \\ \rightarrow \mathit{Witness}_E^{i+1,\vec{c},a}(p(w,\vec{c},a),\vec{c},a). \end{aligned}$$

By the same trick as we used in Case (10), we can replace $\mathit{Witness}_E^{i+1,\vec{c},a}(p(w,\vec{c},a),\vec{c},a)$ by a Σ_i^b- or a Π_i^b-formula and use the PIND axioms to get that

$$T_2^i \vdash \mathit{Witness}_F^{i+1,\vec{c}}(w,\vec{c}) \rightarrow \mathit{Witness}_E^{i+1,\vec{c},a}(p(w,\vec{c},a),\vec{c},a).$$

So f is Q_i-defined by $f(w,\vec{c}) = p(w,\vec{c},t)$ and then it is obvious from the definition of *Witness* that

$$T_2^i \vdash \mathit{Witness}_F^{i+1,\vec{c}}(w,\vec{c}) \rightarrow \mathit{Witness}_A^{i+1,\vec{c}}(f(w,\vec{c}),\vec{c}).$$

Second, consider the case where $B \in \Sigma_i^b \cup \Pi_i^b$. Here we cannot use the simplifying assumption that Π and Λ are empty and must consider the four subcases for $B(0)$ in Γ or Π and $B(t)$ is Δ or Λ. All four subcases are handled similarly: let h be the Q_i-defined function

$$h(w,\vec{c}) = (\mu x \leq |t|)(\neg B(\mathit{MSP}(t, |t| \dot{-} x))$$

as in the last paragraph of Case (10). The function f first checks if $\neg B(0)$ or $B(t)$ is true; if so, it is trivial to give a witness for it; if not, the function g given by the induction hypothesis is applied to $a = MSP(t, |t| \dot{-} h(w,\vec{c}))$ to get a witness for $(\bigvee \Delta^*)$. The details are left to the reader.

Case (13): (Structural Inferences). The cases where the last inference is an exchange inference, a contraction inference or a weak inference are all trivial and their proofs are omitted.

Q.E.D. Theorem 17. □

6 Conclusion

J. P. Ressayre [8] introduced the strong Σ_i^b-replacement axioms (he called them strong Σ_i^b-collection axioms) and showed that the theory $S_2^1 + \Sigma_{i+1}^b$-replacement is $\forall\exists\Sigma_{i+1}^b$-conservative over $S_2^1 +$ strong Σ_i^b-replacement. In view of Theorem 1, this means that $S_2^1 + \Sigma_{i+1}^b$-replacement is $\forall\exists\Sigma_{i+1}^b$-conservative over S_2^i; furthermore, by Theorem 5 this implies that for $i \geq 1$, $S_2^1 + \Sigma_{i+2}^b$-replacement is $\forall\Sigma_{i+1}^b$-conservative over T_2^i and is $\forall\exists\mathcal{B}(\Sigma_{i+1}^b)$-conservative over $T_2^i + \Sigma_{i+1}^b$-replacement, where $\mathcal{B}(\Sigma_{i+1}^b)$ denotes the set of Boolean combinations of Σ_{i+1}^b-formulae (possibly containing free variables).

The obvious question arises of what the exact strength of the Σ_{i+1}^b-replacement axioms is relative to S_2^i and T_2^i. From [1] we know that, relative to the base theory S_2^1,

$$\Sigma_{i+1}^b\text{-PIND} \implies \Sigma_{i+1}^b\text{-replacement} \implies \Sigma_i^b\text{-PIND}.$$

But do either of the arrows reverse? Note that if Σ^b_{i+1}-replacement implies Σ^b_{i+1}-PIND then by Ressayre's result, S_2^{i+1} is Σ^b_{i+1}-conservative over S_2^i — which seems unlikely. On the other hand, there seems to be no reason why it should not be the case that the Σ^b_{i+1}-replacement axioms are theorems of S_2^i.

Another open question is whether Δ^b_{i+1}-PIND is a consequence of S_2^i. Of course Δ^b_{i+1} means with respect to S_2^i. Possibly $S_2^i + \Delta^b_{i+1}$-PIND is conservative in some way over S_2^i?

Finally, it is still completely open whether the theories

$$S_2^1 \subseteq T_2^1 \subseteq S_2^2 \subseteq T_2^2 \subseteq \cdots$$

are all distinct. Can our result that S_2^{i+1} is Σ^b_{i+1}-conservative over T_2^i be strengthened to $S_2^{i+1} \equiv T_2^i$? Is T_2^i conservative over S_2^i? We remark that it is unlikely that T_2^1 is Σ^b_1-conservative over S_2^1 since this has surprising consequences for the computational complexity of linear programming. It is straightforward to see that T_2^1 can Σ^b_1-define a function which solves linear programming problems; hence, if T_2^1 is Σ^b_1-conservative over S_2^1 then so can S_2^1 and thus by the main theorem of [1], linear programming has a polynomial time algorithm. Of course this latter fact is well-known, but it would be very surprising to have a purely logical proof which did not depend on the geometry of linear programming — note that Khachiyan's and Karmarkar's algorithms do depend strongly on geometric considerations [5, 4].

In closing, let us remark that it is expected to be difficult to actually prove that the theories S_2^i and T_2^j are distinct; in part because it involves the same problems that arise in trying to prove $P \neq NP$. For example, if $S_2^1 \not\equiv S_2$ then S_2^1 does not prove $NP = \text{co-}NP$. However, there are known separation results for relativized theories: if we add a new function symbol f to the language of Bounded Arithmetic, then we have by Theorem 5.15 of [1] that $S_2^1(f) \not\equiv T_2^2(f)$. But no separation results are known for the unrelativized theories, and it seems that until new techniques are developed we shall have to content ourselves with proving equivalence and conservation results for fragments of Bounded Arithmetic.

References

[1] Samuel R. Buss. *Bounded Arithmetic*. Bibliopolis, 1986. Revision of 1985 Princeton University Ph.D. thesis.

[2] Samuel R. Buss and Louise Hay. On truth-table reducibility to SAT and the difference hierarchy over NP. In *Proceedings of the Structure in Complexity Conference*, pages 224–233, June 1988.

[3] Felix Hausdorff. *Set Theory.* Chelsea, third edition, 1978.

[4] N. Karmarkar. A new polynomial time algorithm for linear programming. *Combinatorica*, 4:373–395, 1984.

[5] L. G. Khachiyan. A polynomial algorithm in linear programming. *Soviet Math Doklady*, 20:191–194, 1979.

[6] Edward Nelson. *Predicative Arithmetic.* Princeton University Press, 1986.

[7] J. B. Paris and L. A. S. Kirby. Σ_n-collection schemes in arithmetic. In *Logic Colloquium '77*, pages 199–210. North-Holland, 1978.

[8] J. Pierre Ressayre. A conservation result for systems of bounded arithmetic. Handwritten manuscript, June 1986.

[9] A. J. Wilkie and J. B. Paris. On the scheme of induction for bounded arithmetic formulas. *Annals of Pure and Applied Logic*, 35:261–302, 1987.