ON PROVING CONSISTENCY OF EQUATIONAL THEORIES IN BOUNDED ARITHMETIC

ARNOLD BECKMANN AND YORIYUKI YAMAGATA

Abstract. We consider equational theories based on axioms for recursively defining functions, with rules for equality and substitution, but no form of induction—we denote such equational theories as **PETS** for *pure equational theories with substitution*. An example is Cook's system PV without its rule for induction.

We show that the Bounded Arithmetic theory S^1_2 proves the consistency of **PETS**. Our approach employs model-theoretic constructions for **PETS** based on approximate values resembling notions from domain theory in Bounded Arithmetic, which may be of independent interest.

1. Introduction

The question whether the hierarchy of Bounded Arithmetic theories is strict or not, is an important open problem due to its connections to corresponding questions about levels of the Polynomial Time Hierarchy [2]. One obvious way to address this problem is to make use of Gödel's 2nd Incompleteness Theorem, using statements expressing consistency for Bounded Arithmetic theories. Early lines of research aimed to restrict the formulation of consistency suitably to achieve this aim [2, 8], with limited success.

One particular programme is to consider consistency statements based on equational theories. Buss and Ignjatović [4] have shown that the consistency of an induction free version of Cook's equational theory PV [5] is not provable in S^1_2, where S^1_2 is the Bounded Arithmetic theory related to polynomial time reasoning. Their version of induction free PV is formulated in a system that allows, in addition to equations, also inequalities between terms, and Boolean formulas. Furthermore, a number of properties have been stated as axioms.

On the other hand, in a pure equational setting, where lines in derivations are equations between terms, where axioms are restricted to recursive definitions of function symbols, and where induction is not allowed, consistency becomes provable in Bounded Arithmetic: The first author has shown in [1] that the consistency of pure equational theories, in which substitution is not allowed, is provable in S^1_2 – in particular this result applies to Cook's PV [5] without substitution and induction. The second author of this paper has improved on this result in [10] by showing that the consistency of PV without induction but with substitution is provable in S^2_2, the second level of the hierarchy of Bounded Arithmetic theories.

In this paper, we extend both our previous results [1, 10]. With **PETS**(Ax) we denote the pure equational theory with substitution but without induction, based on a 'set of nice axioms Ax' – we will make the notion of 'nice axioms' precise in Definition 6. Cook's original PV [5] without induction but with substitution is one example of a pure equational theory in this sense. The main aim of this paper

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

is to show that the consistency of **PETS**(Ax) is provable in S^1_2, thus improving on both [1, 10]. To this end we employ a novel method of defining pre-models in Bounded Arithmetic based on approximate values, which may be of independent interest. Our approach resembles elements from domain theory, however we leave a full treatment of domain theory in Bounded Arithmetic to future research.

In the next section, we briefly introduce Bounded Arithmetic and fix some notions used throughout the paper. In Section 3 we define pure equational theories **PETS**(Ax), which will be more general than PV without induction in that arbitrary recursive functions may be considered. This is followed in Section 4 by an introduction of approximate values and semantics based on approximation, leading to feasible evaluations of terms based on such approximation semantics. Section 5 then defines pre-models for equational theories based on approximation semantics, including a suitably restricted version which can be expressed as a bounded formula and used in induction arguments inside Bounded Arithmetic. A key notion will be a way of updating such pre-models with further information about approximate values of functions, in a way that preserves the notion of being a pre-model, provably in S^1_2. In Section 6 we prove our first main theorem showing a form of soundness for **PETS**(Ax) based on approximation semantics in S^2_2– an immediate consequence will be that S^2_2 proves the consistency of **PETS**(Ax). The final sections improve on this approach to obtain the main result in S^1_2: In Section 7 we introduce instructions which are extracted from derivations in **PETS**(Ax) and store key steps in the construction of updates of models. In this way we obtain an explicit way of describing pre-model constructions related to **PETS**(Ax) derivations, which allows us to reduce the bounded quantifier complexity of induction assertions in the proof of our second main theorem in Section 8 to show an improved soundness property for **PETS**(Ax) provable in S^1_2.

2. Bounded Arithmetic

2.1. **Language of Bounded Arithmetic.** We give a brief introduction to Bounded Arithmetic to support the developments in this paper. For more in depth discussions and results we refer the interested reader to the literature [2, 6]. Theories of Bounded Arithmetic are first order theories of arithmetic similar to Peano Arithmetic, their domain of discourse is the set of non-negative integers. For the purpose of this paper we can assume that the language of Bounded Arithmetic contains a symbol for each polynomial time computable function, including 0, 1, $+$, $\cdot$, $|.|$, #, representing zero, one, addition, multiplication, the binary length function $|x|$ that computes the number of bits in a binary representation of x and can be defined by $|x| = \lceil \log_2(x+1) \rceil$, and smash # computing $x \# y = 2^{|x|\cdot|y|}$.

2.2. **Theories of Bounded Arithmetic.** Theories of Bounded Arithmetic contain suitable defining axioms for their function symbols. The main differences are various forms of induction for various classes of bounded formulas, which we will define next.

Bounded quantifiers are defined as follows:

$$\begin{aligned} (\forall x \le t)A &\quad \text{abbreviates} \quad (\forall x)(x \le t \to A) \\ (\exists x \le t)A &\quad \text{abbreviates} \quad (\exists x)(x \le t \land A) \end{aligned}$$

If the bounding term t of a bounded quantifier is of the form $|t'|$, then the quantifier is called *sharply bounded*.

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

Classes of bounded formulas Σ^{b}_i and Π^{b}_i are defined in [2] by essentially counting alternations between existential and universal bounded quantifiers. Predicates defined by Σ^{b}_i and Π^{b}_i formulas define computational problems in corresponding classes of the Polynomial Time Hierarchy of decision problems Σ^{p}_i and Π^{p}_i, respectively. For example, those defined by Σ^{b}_1 correspond exactly to Σ^{p}_1, i.e. NP.

In particular, the Σ^{b}_i and Π^{b}_i classes include the following formulas:

- $\Sigma^{\mathrm{b}}_0 = \Pi^{\mathrm{b}}_0$ is the class of formulas built from atomic formulas and closed under Boolean connectives and sharply bounded quantification.
- $\Sigma^{\mathrm{b}}_{i+1}$ includes all formulas of the form $(\exists x \le t)A$ with $A \in \Pi^{\mathrm{b}}_i$.
- Π^{b}_{i+1} includes all formulas of the form $(\forall x \le t)A$ with $A \in \Sigma^{\mathrm{b}}_i$.

The theories S^i_2, $i \ge 0$, of Bounded Arithmetic have been defined in [2], establishing a close relationship between fragments of Bounded Arithmetic and levels of the Polynomial Time Hierarchy of functions. More precisely, the $\Sigma^{\mathrm{b}}_{i+1}$-definable functions of S^{i+1}_2, that is the functions whose graph can be described by a $\Sigma^{\mathrm{b}}_{i+1}$ formula, and whose totality is provable in S^{i+1}_2, form exactly the $i{+}1$-st level of the Polynomial Time Hierarchy of functions, $\mathrm{FP}^{\Sigma^{\mathrm{p}}_i}$.

Instead of defining the theory S^i_2, we state some characteristic properties about induction provable in them. Let Σ^{b}_i-IND be the schema of induction for Σ^{b}_i-properties, consisting of formulas of the form

$$A(0) \wedge (\forall x)(A(x) \to A(x+1)) \to (\forall x)A(x)$$

for $A \in \Sigma^{\mathrm{b}}_i$. The schema of *logarithmic induction* Σ^{b}_i-LIND is then obtained by restricting the conclusion of induction to logarithmic values, that is

$$A(0) \wedge (\forall x)(A(x) \to A(x+1)) \to (\forall x)A(|x|)$$

for $A \in \Sigma^{\mathrm{b}}_i$. We have the following:

Theorem 1 ([2]). *The instances of Σ^{b}_i-LIND and Π^{b}_i-LIND are provable in S^i_2.*

We already mentioned the intricate relationship between Bounded Arithmetic theories and the Polynomial Time Hierarchy in terms of definable functions. Furthermore, it is known that a collapse of the hierarchy of Bounded Arithmetic theories is equivalent to the provability in Bounded Arithmetic of a collapse of the Polynomial Time Hierarchy [3, 7, 11]. With T^i_2 denoting the theory defined by Σ^{b}_i-IND, we have that $\mathrm{T}^i_2 = \mathrm{S}^{i+1}_2$ is equivalent to $\Sigma^{\mathrm{p}}_{i+1} \subseteq \Pi^{\mathrm{p}}_{i+1}/poly$ provable in T^i_2.

The Bounded Arithmetic theory S^1_2 is able to formalize meta-mathematics and essential constructions to prove Gödel's Incompleteness Theorems [2]. A basis for such formalization is a feasible encoding of sequences of numbers. For this paper we assume that a suitable encoding of sequences and operations on them can be formalized in S^1_2 as done in [2]. We assume the following notation:

- With $\langle x_1, \ldots, x_k \rangle$ we denote the encoding of sequence $x_1, \ldots, x_k$. We will use σ, τ etc to range over sequence encodings.
- With '$::$' we denote concatenation of sequences:

$$\langle x_1, \ldots, x_k \rangle :: \langle x_{k+1}, \ldots, x_{k+\ell} \rangle \quad = \quad \langle x_1, \ldots, x_k, x_{k+1}, \ldots, x_{k+\ell} \rangle$$

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

- With ‘ : ’ we denote the function that adds an element to the left or right of a sequence:

$$\begin{aligned} x : \sigma &= \langle x \rangle :: \sigma \\ \sigma : x &= \sigma :: \langle x \rangle \end{aligned}$$

The predicate ‘being a sequence encoding’, as well as the operations ‘ : ’ and ‘ :: ’, can be defined in S^1_2 with their usual properties proven.

In the following, in order to argue that transformations and constructions involving syntax (like terms, proofs, etc) can be conducted in Bounded Arithmetic, we will focus on the length of objects, instead of their size as given by e.g. number of bits in a suitable Gödelization. For an object o we will define its length $\mathbf{l}(o)$ to be the number of symbols occurring in o. As all our constructions will happen in the context of a given derivation $\mathcal{D}$, we obtain that the size of the Gödelization of object o can then be bounded by $\mathbf{l}(o)$ times the size of the Gödelization of $\mathcal{D}$.

Furthermore, the constructions for defining o in the context of $\mathcal{D}$ will always be explicit and simple enough to be formalizable in S^1_2, similar to constructions in [2] dealing with meta-mathematical notions. In those cases where more involved induction is needed (as in Theorems 42 and 53) these will be analyzed carefully.

2.3. Notations.

In the remaining part of this section we will fix some notation used throughout this paper. We use $\equiv$ for equality of syntax.

- With $\#S$ we denote the cardinality of a set S.
- For a function f, $\mathrm{dom}(f)$ denotes its domain, $\mathrm{rng}(f)$ its range.
- For a set S, tuples in S^n are denoted with $(s_1, \ldots, s_n)$.
- We use the notation $\overline{s}$ for a tuple $(s_1, \ldots, s_n)$ as well as a sequence $s_1, \ldots, s_n$ of objects.
- $\max(X)$ computes the maximum (according to a given order) of the elements in X. max is applied to a tuple by computing the maximal component in it.

Remark. Technically, there is a difference between a tuple of the form $(s_1, \ldots, s_n)$, which is an element of S^n, and the sequence $s_1, \ldots, s_n$, which is a formal list used e.g. as the arguments to an n-ary function. We will identify both and write $\overline{s} \in S^n$ and $f(\overline{s})$ in the same context, as long as it does not lead to confusion, in which case we will choose a more precise differentiation.

3. Equational Theories

3.1. Domain of discourse.

The intended domain of discourse $\mathbb{B}$ are finite binary strings representing numbers. $\mathbb{B}$ can be defined inductively as follows:

$$v ::= \epsilon \mid v0 \mid v1$$

We will also use terms denoting binary strings, which are formed from constant ϵ using unary function symbols s_0 and s_1 to add a single digit to the right of a string. We also use $t0$ to denote $\mathrm{s}_0(t)$, and $t1$ for $\mathrm{s}_1(t)$ for terms t, following Cook [5]. We will drop ϵ when writing explicit binary strings, e.g. writing 1101 instead of $\epsilon 1101$, or $\mathrm{s}_1(\mathrm{s}_0(\mathrm{s}_1(\mathrm{s}_1(\epsilon))))$.

Remark. There is a choice in notation in that bits can go to the right or left of v. Going to the left would follow domain theory convention where the focus are infinite binary strings. We decided to put them to the right, following Bounded Arithmetic

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

convention as initiated by Cook [5], because the focus in this paper are finite binary strings representing numbers.

Remark. Our results are not restricted to finite binary strings, but can be applied to general free algebras as domains of discourse as done in [1]. However, for sake of simplicity we will only consider binary strings in this paper.

3.2. **Terms.** We fix the language we use for equational theories.

Definition 2 (Language for Equational Theories)**.** We have the following basic ingredients:

- A countable set $\mathcal{X}$ of *variables;* we use $x, y, x_1, x_2, \dots$ to denote variables.
- A countable set $\mathcal{F}$ of *function symbols;* we use $f, g, h, f_1, f_2, \dots$ to denote function symbols. Each function $f \in \mathcal{F}$ comes with a non-negative integer $\mathrm{ar}(f)$ called its *arity.* We assume that ϵ, s_0 and s_1 are included in $\mathcal{F}$; ϵ, s_0 and s_1 form the set $\mathcal{B}$ of *basic function symbols.*

Remark. A function with arity 0 is called a *constant.* For example, ϵ is a constant.

Definition 3 (Terms)**.** Let $X \subseteq \mathcal{X}$ and $F \subseteq \mathcal{F}$. The set $\mathcal{T}(X, F)$ of *terms over F and X*, or simply *terms*, is defined inductively as follows:

- All variables in X are terms.
- If $f \in F$ has arity n and $t_1, \dots, t_n$ are terms, then $f(t_1, \dots, t_n)$ is a term.

We will use s, t, u to denote terms.

The *length* $\mathbf{l}(t)$ *of term t* is defined in the following way: The length of a variable is 1, and, recursively,

$$\mathbf{l}(f(t_1, \dots, t_n)) = \mathbf{l}(t_1) + \cdots + \mathbf{l}(t_n) + 1 \ \ .$$

With $\mathrm{Var}(t)$ we denote the *set of variables* that are occurring in a term t. For S a set of terms, $\mathrm{Var}(S)$ denotes the union of $\mathrm{Var}(t)$ for $t \in S$.

Definition 4 (Substitution)**.** Let t, u be terms and x be a variable. The *substitution of u for x in t*, denoted $t[u/x]$, is obtained by replacing any occurrence of x in t by u.

We extend substitution to sequences of variables and terms of the same length by successively substituting terms: $t[\overline{u}/\overline{x}]$ stands for $t[u_1/x_1][u_2/x_2] \dots [u_n/x_n]$.

Definition 5 (Instance and injective renaming)**.** A *substitution instance*, or *instance*, of an equation $s = t$ is any $s[\overline{u}/\overline{x}] = t[\overline{u}/\overline{x}]$ for terms $\overline{u}$.

An instance $s[\overline{u}/\overline{x}] = t[\overline{u}/\overline{x}]$ is called an *injective renaming* of $s = t$, iff $\{\overline{x}\} = \mathrm{Var}(s) \cup \mathrm{Var}(t)$, and $\overline{u}$ is a list of pairwise distinct variables.

3.3. **Nice axiom systems.** We will consider axioms consisting of equations that satisfy particular conditions, which have been called *nice* in [1].

Definition 6 (Nice Axioms, [1])**.** Let Ax be a set of equations over $\mathcal{F}$. Ax is called a *set of nice axioms for $\mathcal{F}$*, if the following is satisfied: Each equation in Ax must be of one of the following forms, for some $f \in \mathcal{F} \setminus \mathcal{B}$, some $t, t_\epsilon \in \mathcal{T}(\mathcal{F}, \{\overline{y}\})$, and $t_0, t_1 \in \mathcal{T}(\mathcal{F}, \{x, \overline{y}\})$:

$$\begin{aligned} f(\overline{y}) &= t \\ f(\epsilon, \overline{y}) &= t_\epsilon \\ f(x0, \overline{y}) &= t_0 \\ f(x1, \overline{y}) &= t_1 . \end{aligned}$$

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

Furthermore, the left-hand side of an equation is occurring at most once among equations in Ax, also modulo injective renamings.

Henceforth, Ax is a set of nice axioms.

Remark. The definition implies that for any $t = u$ in Ax we have $\mathrm{Var}(u) \subseteq \mathrm{Var}(t)$.

Remark. Consider a term $f(\overline{s})$ with $f \in \mathcal{F} \setminus \mathcal{B}$. The property of Ax being nice implies that there is at most one axiom $t = u$ in Ax such that $f(\overline{s})$ can be written as a substitution instance of t.

Remark. For the reader familiar with term-rewriting we remark that the term rewriting system induced by nice axioms is orthogonal, which follows from the previous two remarks.

Remark. The definition of a nice axiom system in [1] also contains a completeness condition, requiring that each function symbol in $\mathcal{F} \setminus \mathcal{B}$ is defined by an equation, and that the case distinction in the latter part is complete. We omit this form of completeness, as it is not needed for our developments.

The left-hand side of an equation in Ax is of a very special form: an argument to the outermost function symbol can either be a variable, ϵ, $x0$, or $x1$ for some variable x. We capture these forms in the following definition.

Definition 7 (Generalized variable)**.** A *generalized variable* is a term which either is a variable, or ϵ, or of the form $x0$ or $x1$ for some variable x.

Remark. Consider an axiom $t = u$ in Ax. It follows that t must be of the form $f(\overline{s})$, that each s_i is a generalized variable, hence each s_i contains at most one variable. Furthermore, the same variable cannot occur simultaneously in s_i and s_j for $i \neq j$.

Definition 8 (Rules for equational reasoning)**.** Let s, t, u represent terms, and x a variable. The following are the rules that can be used to derive equations:

Axiom: $\vdash t = u$, where $t = u$ is an injective renaming of an equation in Ax.
Reflexivity: $\vdash t = t$
Symmetry: $t = u \vdash u = t$
Transitivity: $t = s, s = u \vdash t = u$
Compatibility: $t = u \vdash s[t/x] = s[u/x]$
Substitution: $t = u \vdash t[s/x] = u[s/x]$.

In the case of Substitution as stated above, we say that the application of Substitution binds the variable x.

Remark. We also make use of a display style for presenting rules, like

$$\mathrm{Re}\,\frac{}{t = t}$$

for Reflexivity Rule, or

$$\mathrm{Tr}\,\frac{t = s \qquad s = u}{t = u}$$

for Transitivity Rule. As indicated, we may abbreviate the rule that is used by its first two letters.

Definition 9 (Derivations)**.** A *derivation* is a finite tree whose nodes are labelled with rules for equational reasoning, such that for each node, the premises of the rule at that node coincide with the conclusions of rules at corresponding child nodes.

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

Derivations can also be defined inductively from rules for equational reasoning: Any Axiom or Reflexivity Rule is a derivation ending in the equation given by that rule. If R is a unary rule (like Symmetry, Compatibility or Substitution) with premise e' and conclusion e, and $\mathcal{D}'$ a derivation ending in e', then

$$R\,\frac{\begin{matrix}\mathcal{D}'\\ e'\end{matrix}}{e}$$

is a derivation ending in e. The only binary rule we are considering is the Transitivity Rule. If $\mathcal{D}_1$ a derivation ending in $t = s$, and $\mathcal{D}_2$ a derivation ending in $s = u$, then

$$\mathrm{Tr}\,\frac{\begin{matrix}\mathcal{D}_1\\ t = s\end{matrix}\qquad\begin{matrix}\mathcal{D}_2\\ s = u\end{matrix}}{t = u}$$

is a derivation ending in $t = u$.

We tacitly assume that labels of rules carry information to uniquely identify them. For example, in the case of Compatibility and Substitution in Definition 8, the label would contain x and s.

Building on Definition 3, the length $\mathbf{l}(t = u)$ of an equation $t = u$ is set to $\mathbf{l}(t) + \mathbf{l}(u) + 1$. The length $\mathbf{l}(l)$ of a label l of a rule R is set to either $\mathbf{l}(s) + \mathbf{l}(x) + 1$ if R is Compatibility or Substitution involving x and s, or 1 otherwise. The *length* $\mathbf{l}(\mathcal{D})$ *of a derivation* $\mathcal{D}$ is defined as the sum of the lengths of the equations and labels of rules occurring in it.

With $\mathrm{Var}(\mathcal{D})$ we denote the set of variables occurring in $\mathcal{D}$. $\mathrm{BVar}(\mathcal{D})$ is the set of variables occurring in $\mathcal{D}$ that are bound by an application of the Substitution Rule.

Definition 10 (Pure Equational Theories with Substitution)**.** The *pure equational theory with substitution* **PETS**(Ax) consists of all equations that can be derived using the Axiom rules for Ax, as well as the Reflexivity, Symmetry, Transitivity, Compatibility and Substitution Rules.

Definition 11 (Variable Normal Form)**.** A **PETS**(Ax) derivation $\mathcal{D}$ is in *variable normal form* if each variable occurring in $\mathcal{D}$ is either occurring in the equation in which the derivation ends, or is removed in exactly one application of Substitution (as the variable which is bound by that application of Substitution).

Proposition 12. *Assume $\mathcal{D} \vdash t = u$, then there exists $\mathcal{D}'$ in variable normal form of same length as $\mathcal{D}$ such that $\mathcal{D}' \vdash t = u$.*

Proof. We observe that we can replace all occurrences of a fixed variable by a fresh variable throughout a derivation ending in an equation e, obtaining a similar derivation of the equation e, potentially with the former variable being renamed to the latter fresh variable. This transformation works as our derivations are tree-like. The above transformation does not change the length of the derivation. □

Remark. While the length does not change when transforming a derivation to variable normal form, its size (in the sense of number of bits in a suitable Gödelisation) in general will change due to the need of choosing new indices for fresh variables. As remarked before, the size of the resulting derivation in variable normal form will be bounded by a constant times the product of the length of the new and the size of the original derivation.

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

Definition 13 (Formal Consistency)**.** With $\mathrm{Cons}(\mathbf{PETS}(\mathrm{Ax}))$ we denote the sentence in the language of Bounded Arithmetic which expresses that there is no derivation in $\mathbf{PETS}(\mathrm{Ax})$ ending in $0 = 1$, where 0, resp. 1, denotes $\epsilon 0$, resp. $\epsilon 1$.

We close the section with an example theory, $\mathbf{PETS}_{\mathrm{exp}}$. The theory defines a form of exponentiation, it will reoccur at later sections to explain terminology defined there.

Example. Let $\mathcal{F}_{\mathrm{exp}}$ be $\{\epsilon, \mathrm{s}_0, \mathrm{s}_1, \oplus, \otimes, \mathrm{e}\}$ where $\oplus, \otimes, \mathrm{e}$ are binary function symbols.

Let $\mathrm{Ax}_{\mathrm{exp}}$ be the nice set of axioms given by

$$\begin{array}{ccc} x \oplus \epsilon = x & x \otimes \epsilon = \epsilon & \mathrm{e}(x, \epsilon) = 1 \\ x \oplus yi = (x \oplus y)i & x \otimes yi = (x \otimes y) \oplus x & \mathrm{e}(x, yi) = e(x, y) \otimes x \end{array}$$

for $i \in \{0, 1\}$.

For a, b terms, consider the following $\mathbf{PETS}_{\mathrm{exp}}$ derivation of $\mathrm{e}(\epsilon, a \oplus b0) = \epsilon$. Let $\mathcal{D}_1$ denote the derivation

$$\begin{prooftree}
\AxiomC{}
\LeftLabel{Ax}
\UnaryInfC{$\mathrm{e}(x_2, y_2 0) \ = \ \mathrm{e}(x_2, y_2) \otimes x_2$}
\doubleLine
\LeftLabel{Su}
\UnaryInfC{$\mathrm{e}(\epsilon, (a \oplus b)0) \ = \ \mathrm{e}(\epsilon, a \oplus b) \otimes \epsilon$}
\AxiomC{}
\LeftLabel{Ax}
\UnaryInfC{$x_3 \otimes \epsilon \ = \ \epsilon$}
\LeftLabel{Su}
\UnaryInfC{$\mathrm{e}(\epsilon, a \oplus b) \otimes \epsilon \ = \ \epsilon$}
\LeftLabel{Tr}
\BinaryInfC{$\mathrm{e}(\epsilon, (a \oplus b)0) \ = \ \epsilon$}
\end{prooftree}$$

Then let $\mathcal{D}_{\mathrm{exp}}$ be

$$\begin{prooftree}
\AxiomC{}
\LeftLabel{Ax}
\UnaryInfC{$x_1 \oplus y_1 0 \ = \ (x_1 \oplus y_1)0$}
\doubleLine
\LeftLabel{Su}
\UnaryInfC{$a \oplus b0 \ = \ (a \oplus b)0$}
\LeftLabel{Co}
\UnaryInfC{$\mathrm{e}(\epsilon, a \oplus b0) \ = \ \mathrm{e}(\epsilon, (a \oplus b)0)$}
\AxiomC{$\mathcal{D}_1$}
\noLine
\UnaryInfC{$\mathrm{e}(\epsilon, (a \oplus b)0) \ = \ \epsilon$}
\LeftLabel{Tr}
\BinaryInfC{$\mathrm{e}(\epsilon, a \oplus b0) \ = \ \epsilon$}
\end{prooftree}$$

We use a double line to indicate multiple applications of the indicated rule. Here, two applications of substitution are used, one for each variable substituted for.

4. Approximation Semantics

A core idea in domain theory is to define finite approximations to function in a way that is consistent with axioms and equational reasoning. Although we do not develop domain theory fully in this paper, we will make use of some of its notions, and show in particular that those notions can be defined and reasoned with in S^1_2. As constant functions are approximated by themselves, this allows us to conclude that **PETS** will never derive $0 = 1$, showing that the consistency of **PETS** is provable in bounded arithmetic.

4.1. Approximate values.

The notion of approximate values v is defined in [10], which adds 'unknown value' of a term, as defined in [1] and denoted with '$*$', to bitstrings. Infeasible values, although finite, can be considered as infinite bit strings within theories of feasibility like Bounded Arithmetic. From the domain theoretic view, the space of approximate values is a domain formed by infinite strings of bits.

Definition 14 (Approximate values)**.** The set $\mathbb{D}$ of *approximate values* is defined inductively as follows:

$$v ::= \epsilon \mid v0 \mid v1 \mid *$$

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

For the proofs of main theorems we will need a refined approach to measuring length, so that inductive statements in those proofs lead to polynomial bounds. The general principle will be that we think of a typing structure for our objects, so that the i-length of an object of type i is defined in some natural way (like length of a word, or cardinality of a set), while for $j < i$ the j-length of a type i object is given as the maximum of the j-lengths of its components. For example, our basic objects will be words in $\mathbb{D}$. They will be of type 0, and their 0-length is just their standard length as words. A sequence of words is of type 1, its 1-length is its sequence length, i.e. the number of occurrences of words in it, and its 0-length the maximum of the lengths of words occurring in it. A set of sequences of words is of type 2, its 2-length is its cardinality, its 1-length the maximum of the sequence lengths of its elements, and its 0-length the maximum of the word lengths of its elements. Etc.

It is then obvious that the length $\mathbf{l}(o)$ of an object o of type i is bounded by a constant times the product of its j-lengths for $j \leq i$.

Definition 15. For an approximate value $v \in \mathbb{D}$, its 0-*length* $\mathbf{l}_0(v)$ satisfies

$$\mathbf{l}_0(\epsilon) = \mathbf{l}_0(*) = 1$$
$$\mathbf{l}_0(v0) = \mathbf{l}_0(v1) = \mathbf{l}_0(v) + 1$$

For a tuple $\overline{w} = (w_1, \ldots, w_n) \in \mathbb{D}^n$, its 0-*length* is given as

$$\mathbf{l}_0(\overline{w}) = \max\{\mathbf{l}_0(w_1), \ldots, \mathbf{l}_0(w_n)\} \ ,$$

and its 1-*length* as

$$\mathbf{l}_1(\overline{w}) = n \ .$$

The idea of an approximate value is that bits of higher significance may not be known and are thus replaced by $*$.

Example. 1101, $*01$, $*$ are examples of approximate values, $1{*}1$, $**$ are not. Following the above intuition that $*$ replaces 'unknown' bits of higher significance, $*01$ is an approximation to 1101. We will make this intuition precise in the next subsection.

4.2. **Approximation relation.**

A relation $\trianglelefteq$ has been defined in [1]. Here we will consider the converse $\sqsubseteq$ of $\trianglelefteq$.

Definition 16 (Approximation relation)**.** The *approximation relation* $\sqsubseteq$ is a binary relation over $\mathbb{D}$, defined inductively as follows:

- $* \sqsubseteq v$ for any $v \in \mathbb{D}$.
- $\epsilon \sqsubseteq \epsilon$.
- if $v_1 \sqsubseteq v_2$, then $v_1 0 \sqsubseteq v_2 0$ and $v_1 1 \sqsubseteq v_2 1$.

We pronounce '$v \sqsubseteq w$' as 'v approximates w'.

We extend $\sqsubseteq$ to tuples: $(v_1, \ldots, v_n) \sqsubseteq (w_1, \ldots, w_n)$ iff $v_i \sqsubseteq w_i$ for each i.

Example. $*$ approximates any value in $\mathbb{D}$. $*01$ approximates 1101, but does not approximate 1010.

Proposition 17. *$\sqsubseteq$ is a partial order on $\mathbb{D}^n$, that is, it is reflexive, anti-symmetric and transitive.* □

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

Definition 18 (Compatibility)**.** $u, v \in \mathbb{D}$ are *compatible* if $u \sqsubseteq v$ or $v \sqsubseteq u$. $(u_1, \dots, u_n)$ and $(v_1, \dots, v_n)$ in $\mathbb{D}^n$ are compatible if each u_i and v_i are. We denote $\overline{u}$ and $\overline{v}$ being compatible with $\overline{u} \mathbin{\triangle} \overline{v}$.

The following lemma follows immediately from the definition of compatibility, by induction on the size of the set.

Lemma 19. *Any non-empty finite subset S of $\mathbb{D}$ of pairwise compatible elements has a maximal element w.r.t. $\sqsubseteq$. We denote the maximal element with* $\max_{\sqsubseteq}(S)$*. We also set* $\max_{\sqsubseteq}(\emptyset) = *$. □

The following lemma has already been proven in [1]:

Lemma 20. *If $\overline{u}, \overline{v}, \overline{w} \in \mathbb{D}^n$ and $\overline{u}, \overline{v} \sqsubseteq \overline{w}$, then $\overline{u} \mathbin{\triangle} \overline{v}$.* □

We introduce a couple of notions from the treatment of domain theory as introduced by Vickers [9].

Definition 21 (Generator, [9])**.** A *generator* for $\mathbb{D}^n \to \mathbb{D}$ is a tuple $(\overline{u}, v)$, denoted $\overline{u} \mapsto v$, with $\overline{u} \in \mathbb{D}^n$ and $v \in \mathbb{D} \setminus \{*\}$.

Definition 22 (Consistent set, [9])**.** A *consistent set* $\tilde{f}$ of $\mathbb{D}^n \to \mathbb{D}$ is a finite set of generators for $\mathbb{D}^n \to \mathbb{D}$ satisfying the following condition:

if $\overline{x} \mapsto y, \overline{u} \mapsto v \in \tilde{f}$ and $\overline{x} \mathbin{\triangle} \overline{u}$, then $y \mathbin{\triangle} v$.

We say that $\tilde{f}$ has arity n, denoted as $\mathrm{ar}(\tilde{f}) = n$.

Definition 23 (Finitely generated maps, [9])**.** A consistent set $\tilde{f}$ defines a mapping on $\mathbb{D}^{\mathrm{ar}(\tilde{f})} \to \mathbb{D}$, which we call a *finitely generated map*, via

$$\tilde{f}(\overline{x}) = \max_{\sqsubseteq}\{v \mid \exists \overline{w}, \overline{w} \sqsubseteq \overline{x} \text{ and } \overline{w} \mapsto v \in \tilde{f}\} \ .$$

We sometimes write $\tilde{f}[\overline{x}]$ to denote the set

$$\{v \mid \exists \overline{w}, \overline{w} \sqsubseteq \overline{x} \text{ and } \overline{w} \mapsto v \in \tilde{f}\}$$

so that $\tilde{f}(\overline{x}) = \max_{\sqsubseteq} \tilde{f}[\overline{x}]$.

To see that finitely generated maps are well-defined, consider two generators $\overline{w} \mapsto v$ and $\overline{w}' \mapsto v'$ in $\tilde{f}$ with $\overline{w}, \overline{w}' \sqsubseteq \overline{x}$. With Lemma 20 we obtain $\overline{w} \mathbin{\triangle} \overline{w}'$. Hence $v \mathbin{\triangle} v'$ as $\tilde{f}$ is consistent. Thus, using Lemma 19, the set

$$\tilde{f}[\overline{x}] = \{v \mid \exists \overline{w}, \overline{w} \sqsubseteq \overline{x} \text{ and } \overline{w} \mapsto v \in \tilde{f}\}$$

has a maximal element w.r.t. $\sqsubseteq$.

Example. Consider $\tilde{f} = \{\epsilon \mapsto \epsilon, *0 \mapsto *1, *00 \mapsto *11\}$. $\tilde{f}$ is a consistent set. We compute $\tilde{f}[000] = \{*0 \mapsto *1, *00 \mapsto *11\}$, as $*0, *00 \sqsubseteq 000$. Hence $\tilde{f}(000) = *11$, as $*1 \sqsubseteq *11$. The set $\{\epsilon \mapsto \epsilon, *0 \mapsto *1, *00 \mapsto *00\}$ is not consistent, because $*0$ and $*00$ are compatible, but $*1$ and $*00$ are not.

Lemma 24 (Expansion property of finitely generated maps)**.** *Let $\tilde{f}_1$ and $\tilde{f}_2$ be consistent sets of $\mathbb{D}^n \to \mathbb{D}$. If $\tilde{f}_1 \subseteq \tilde{f}_2$, then $\tilde{f}_1(\overline{v}) \sqsubseteq \tilde{f}_2(\overline{v})$ for $\overline{v} \in \mathbb{D}^n$.*

Proof. From $\tilde{f}_1 \subseteq \tilde{f}_2$ we immediately obtain $\tilde{f}_1[\overline{v}] \subseteq \tilde{f}_2[\overline{v}]$. Hence the assertion follows. □

For $\overline{x} \sqsubseteq \overline{y}$ we have $\tilde{f}[\overline{x}] \subseteq \tilde{f}[\overline{y}]$, thus we obtain that finitely generated maps are monotone.

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

Lemma 25 (Monotonicity of finitely generated maps)**.** *Finitely generated maps are monotone w.r.t.* $\sqsubseteq$*.* □

Remark. There are monotone, finitely generated maps which cannot be represented by finite consistent sets. For example, the identity function from $\mathbb{D}$ to $\mathbb{D}$ is monotone but cannot be represented by a finite consistent set.

Definition 26. For a consistent set $\tilde{f}$, we define its length measures as a type 2 object as follows:

- Its 0-length is given as
$$\mathbf{l}_0(\tilde{f}) \quad = \quad \max\{\mathbf{l}_0(\overline{v}), \mathbf{l}_0(w) \mid \overline{v} \mapsto w \in \tilde{f}\} \ .$$
- The 1-length is given as 1 plus its arity:
$$\mathbf{l}_1(\tilde{f}) \quad = \quad \mathrm{ar}(\tilde{f}) + 1 \ .$$
- The 2-length is given by its cardinality:
$$\mathbf{l}_2(\tilde{f}) \quad = \quad \#\tilde{f} \ .$$

Remark. As discussed before, the length of a consistent set $\tilde{f}$, is bounded by $\mathbf{l}(\tilde{f}) = \mathrm{O}(\mathbf{l}_2(\tilde{f}) \cdot \mathbf{l}_1(\tilde{f}) \cdot \mathbf{l}_0(\tilde{f}))$.

Definition 27 (Frame)**.** A *frame* F is a partial, finite mapping of function symbols $f \in \mathcal{F} \setminus \mathcal{B}$ to consistent sets. We extend F to all $f \in \mathcal{F} \setminus \mathcal{B}$ by setting $F(f) = \perp$ for $f \notin \mathcal{B} \cup \mathrm{dom}(F)$, where $\perp$ denotes the empty set $\emptyset$.

The set of frames is partially ordered by
$$F_1 \sqsubseteq F_2 \iff \forall f, F_1(f) \subseteq F_2(f) \ .$$

A frame F defines an *evaluation* $F(f)(\overline{v})$ for $f \in \mathcal{F}$ and $\overline{v} \in \mathbb{D}^{\mathrm{ar}(f)}$ as follows:

- If $f \in \mathcal{B}$, let $F(f)(\overline{v}) = f(\overline{v})$
- If $f \notin \mathcal{B}$ and $F(f) = \tilde{f}$, let $F(f)(\overline{v}) = \tilde{f}(\overline{v})$.

Observe that for $f \notin \mathrm{dom}(F) \cup \mathcal{B}$, we have $F(f) = \perp$, hence $F(f)(\overline{v}) = \perp(\overline{v}) = *$.

Definition 28. For a frame F, we use the following length measures, treating frames as type 3 objects:

- The 0-length of F is given by $\mathbf{l}_0(F) = \max\{\mathbf{l}_0(F(f)) \mid f \in \mathrm{dom}(F)\}$.
- The 1-length of F is given by $\mathbf{l}_1(F) = \max\{\mathrm{ar}(f) + 1 \mid f \in \mathrm{dom}(F)\}$.
- The 2-length of F is given by $\mathbf{l}_2(F) = \max\{\#F(f) \mid f \in \mathrm{dom}(F)\}$.
- The 3-length of F is given by $\mathbf{l}_3(F) = \mathbf{l}(\mathrm{dom}(F))$, where $\mathbf{l}(\mathrm{dom}(F))$ is the sum of the lengths of the objects in $\mathrm{dom}(F)$.

Remark. The length of F can be bounded by
$$\mathbf{l}(F) \quad = \quad \mathrm{O}(\mathbf{l}_3(F) \cdot \mathbf{l}_2(F) \cdot \mathbf{l}_1(F) \cdot \mathbf{l}_0(F)) \ .$$

Definition 29 (Assignments)**.** An *assignment* ρ is a partial, finite mapping from variables $\mathcal{X}$ to approximate values in $\mathbb{D}$. We extend assignments outside their domain to '$*$', i.e. $\rho(x) = *$ for $x \notin \mathrm{dom}(\rho)$.

We extend the approximation order $\sqsubseteq$ to assignments pointwise:
$$\rho_1 \sqsubseteq \rho_2 \quad \text{iff} \quad \forall x, \rho_1(x) \sqsubseteq \rho_2(x) \ .$$

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

With $\rho[x \mapsto v]$ we denote the assignment that behaves like ρ except that it maps x to v:

$$\rho[x \mapsto v](y) \quad = \quad \begin{cases} v & : \text{ if } y = x \\ \rho(y) & : \text{ otherwise.} \end{cases}$$

We apply assignments to generalized variables in the obvious way, e.g., $\rho(\mathrm{s}_i(x)) = \mathrm{s}_i(\rho(x))$.

Definition 30. For assignment ρ we define the following measures, treating assignments as type 1 objects:

- The 0-length of ρ is given by $\mathbf{l}_0(\rho) = \max\{\mathbf{l}_0(v) \mid v \in \mathrm{rng}(\rho)\}$.
- The 1-length of ρ is given by $\mathbf{l}_1(\rho) = \mathbf{l}(\mathrm{dom}(\rho))$, where $\mathbf{l}(\mathrm{dom}(\rho))$ is the sum of the lengths of the objects in $\mathrm{dom}(\rho)$.

Remark. Using the above measures, the length of ρ, $\mathbf{l}(\rho)$, can be bounded by

$$\mathbf{l}(\rho) \quad = \quad \mathrm{O}(\mathbf{l}_1(\rho) \cdot \mathbf{l}_0(\rho)) \ .$$

Definition 31 (Evaluation)**.** Let ρ be an assignment, F a frame, and t a term. The *evaluation* $[\![t]\!]_{F,\rho}$ *of t under F, ρ* is defined recursively as follows:

$$\begin{aligned} [\![x]\!]_{F,\rho} &= \rho(x) \quad \text{for a variable } x \in \mathcal{X}; \\ [\![f(t_1, \dots, t_n)]\!]_{F,\rho} &= F(f)([\![t_1]\!]_{F,\rho}, \dots, [\![t_n]\!]_{F,\rho}) \ . \end{aligned}$$

Example. We have $[\![0]\!]_{F,\rho} = 0$, $[\![1]\!]_{F,\rho} = 1$, as well as $[\![f(\bar{t})]\!]_{F,\rho} = *$ if $f \notin \mathrm{dom}(F)$.

We have the following immediate properties of evaluations.

Lemma 32. (1) $[\![t]\!]_{F,\rho} \in \mathbb{D}$
(2) $[\![t]\!]_{F,\rho}$ *is monotone in F and ρ w.r.t.* $\sqsubseteq$.

Proof. (1) follows immediately from the definition.

We prove (2) by induction on t, showing that for $F \sqsubseteq F'$ and $\rho \sqsubseteq \rho'$,

$$[\![t]\!]_{F,\rho} \sqsubseteq [\![t]\!]_{F',\rho'} \ .$$

If $t \equiv x$, the assertion holds because $\rho(x) \sqsubseteq \rho'(x)$. If $t \equiv f(t_1, \dots, t_n)$, we compute

$$\begin{aligned} [\![f(t_1, \dots, t_n)]\!]_{F,\rho} &= F(f)([\![t_1]\!]_{F,\rho}, \dots, [\![t_n]\!]_{F,\rho}) \\ &\sqsubseteq F'(f)([\![t_1]\!]_{F,\rho}, \dots, [\![t_n]\!]_{F,\rho}) \\ &\sqsubseteq F'(f)([\![t_1]\!]_{F',\rho'}, \dots, [\![t_n]\!]_{F',\rho'}) \ = \ [\![t]\!]_{F',\rho'} \ , \end{aligned}$$

where the first approximation uses the expansion property of finitely generated maps, Lemma 24, and the second the induction hypothesis and monotonicity of finitely generated maps, Lemma 25. □

Lemma 33. *Let ρ be an assignment, F a frame, and t a term. Then*

$$\mathbf{l}_0([\![t]\!]_{F,\rho}) \ \leq \ \max(\mathbf{l}_0(\rho), \mathbf{l}_0(F)) + \mathbf{l}(t) \ .$$

Proof. By induction on t. If t is a variable x, we have

$$\mathbf{l}_0([\![t]\!]_{F,\rho}) \ = \ \mathbf{l}_0(\rho(x)) \ \leq \ \mathbf{l}_0(\rho) \ .$$

If t is ϵ we compute $\mathbf{l}_0([\![t]\!]_{F,\rho}) = 1 = \mathbf{l}(t)$. For t of the form $\mathrm{s}_i(t_1)$ we have

$$\begin{aligned} \mathbf{l}_0([\![t]\!]_{F,\rho}) \ &= \ \mathbf{l}_0(\mathrm{s}_i([\![t_1]\!]_{F,\rho})) \ = \ \mathbf{l}_0([\![t_1]\!]_{F,\rho}) + 1 \\ &\leq \ \max(\mathbf{l}_0(\rho), \mathbf{l}_0(F)) + \mathbf{l}(t_1) + 1 \ = \ \max(\mathbf{l}_0(\rho), \mathbf{l}_0(F)) + \mathbf{l}(t) \ . \end{aligned}$$

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

Finally, assume t is of the form $f(t_1, \dots, t_n)$ with $f \notin \mathcal{B}$. Then we have $\mathbf{l}_0([\![t]\!]_{F,\rho}) \leq \mathbf{l}_0(F)$. □

Lemma 34 (Substitution Lemma)**.** $[\![t(u)]\!]_{F,\rho} = [\![t(x)]\!]_{F,\rho[x \mapsto [\![u]\!]_{F,\rho}]}$

Proof. The proof is by induction on t. □

We close the section with examples for computing evaluations.

Example. Consider the symbol **flip**, denoting the unary bit-flipping function.

(1) Let $F_0 = \emptyset$ be the empty frame, and $\rho_0 \colon x \mapsto *0$. Then $[\![\mathbf{flip}(x)1]\!]_{F_0,\rho_0} = F_0(\mathbf{flip})(*0)1 = \emptyset(*0)1 = *1$.
(2) Let $F_1 = \{\mathbf{flip} \colon *00 \mapsto *1\}$, and $\rho_1 \colon x \mapsto *$. Then $[\![\mathbf{flip}(x)1]\!]_{F_1,\rho_1} = F_1(\mathbf{flip})(*)1 = *1$.
(3) Let $F_2 = \{\mathbf{flip} \colon *00 \mapsto *1, \mathbf{flip} \colon *0 \mapsto *1\}$, and ρ_0 as before. Then $[\![\mathbf{flip}(x)1]\!]_{F_2,\rho_0} = F_2(\mathbf{flip})(*0)1 = *11$ as now $F_2(\mathbf{flip})[*0] = \{*1\}$.
(4) Let $F_3 = \{\mathbf{flip} \colon *00 \mapsto *1, \mathbf{flip} \colon *0 \mapsto *1, \mathbf{flip} \colon *00 \mapsto *11\}$. We can compute $F_3(\mathbf{flip})(*00) = *11$ and $F_3(\mathbf{flip})(*0) = *1$, as well as $F_3(\mathbf{flip})(1) = *$ and $F_3(\mathbf{flip})(1100) = *11$.

5. Frames as Pre-Models

As frames will only provide restricted approximations to function evaluation, we cannot expect to achieve $[\![t]\!]_{F,\rho} = [\![u]\!]_{F,\rho}$ for equations $t = u$ occurring in a derivation $\mathcal{D}$. Instead, we will show for a given derivation $\mathcal{D}$, suitable frame F, and equation $t = u$ occurring in $\mathcal{D}$, that there are frames F', F'' such that $F \sqsubseteq F', F''$, $[\![t]\!]_{F,\rho} \sqsubseteq [\![u]\!]_{F',\rho}$ and $[\![u]\!]_{F,\rho} \sqsubseteq [\![t]\!]_{F'',\rho}$. The idea is that the derivation $\mathcal{D}$ gives an "instruction" of how to build frames such F' and F'' from F. It then follows that $\mathcal{D}$ cannot end in $0 = 1$, because $[\![0]\!]_{F,\rho} = 0 \not\sqsubseteq 1 = [\![1]\!]_{F,\rho}$.

For axioms $t = u$ it suffices to demand $[\![t]\!]_{F,\rho} \sqsubseteq [\![u]\!]_{F,\rho}$, as their property of being nice implies that t will have the restricted form of a function symbol applied to generalized variables, hence the converse can always be enforced by constructing a larger frame F' such that $[\![u]\!]_{F,\rho} \sqsubseteq [\![t]\!]_{F',\rho}$. Frames of this form will be called *pre-models*.

Definition 35 (Pre-Model)**.** A frame F is a *pre-model of* Ax iff for any $t = u$ in Ax and any assignment ρ, $[\![t]\!]_{F,\rho} \sqsubseteq [\![u]\!]_{F,\rho}$.

Remark. As the notion of being a pre-model contains an unbounded quantification over assignments, it is not readily expressed as a bounded formula. However, we do not expect *being a pre-model* to be Π_1-complete, as evaluations based on finite frames can only produce finitely many different values. Instead of analyzing the complexity of *being a pre-model* further, we will define below a more restricted version which contains explicit bounds and can be directly used for reasoning in bounded arithmetic.

Lemma 36. *The empty frame is a pre-model of* Ax*.*

Proof. Let F be the empty frame. Consider $t = u$ in Ax, and assignment ρ. Then t is of the form $f(\overline{s})$ for some f in $\mathcal{F} \setminus \mathcal{B}$. We have $F(f) = \bot$, hence $[\![f(\overline{s})]\!]_{F,\rho} = * \sqsubseteq [\![u]\!]_{F,\rho}$. □

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

We restrict the notion of being a pre-model to obtain a bounded property. Let κ be a positive integer, which is intended to bound the 0-length of approximations occurring in frames and assignments that need to be considered in the definition of pre-models. Furthermore, we restrict the definition to a finite set of axioms, intended to represent those used in a given derivation $\mathcal{D}$.

Definition 37 (κ-Pre-Model)**.** Let Ax_0 be a finite subset of Ax. A frame F is a *κ-pre-model of* Ax_0 iff $\mathbf{l}_0(F) \leq \kappa$, and for any $t = u$ in Ax_0 and any assignment ρ with $\mathrm{dom}(\rho) \subseteq \mathrm{Var}(\mathrm{Ax}_0)$ and $\mathbf{l}_0(\rho) \leq \kappa$, we have $[\![t]\!]_{F,\rho} \sqsubseteq [\![u]\!]_{F,\rho}$.

Remark. The notion of F being a κ-pre-model of Ax_0 can be written as a Π^{b}_1 formula.

We will now define the notion of *updates* that can be used to expand pre-models based on Ax_0 as defined above. In line with our previous discussion, let $\mathbf{l}_0(v_1, \ldots, v_n) = \max\{\mathbf{l}_0(v_i) \colon i = 1, \ldots, n\}$.

Definition 38 (Updates)**.** Let Ax_0 be a finite subset of Ax, and F a κ-pre-model of Ax_0. An *update based on F, κ and* Ax_0 is any $f \in \mathcal{F} \setminus \mathcal{B}$ and generator $\overline{v} \mapsto w$, which we denote as $f \colon \overline{v} \mapsto w$, such that $\mathbf{l}_0(\overline{v}, w) \leq \kappa$, and there exists $t = u$ in Ax_0 and an assignment ρ satisfying that

- t is of the form $f(\overline{s})$,
- $v_i = \rho(s_i)$ for $i \leq \mathrm{ar}(f)$,
- and $w = [\![u]\!]_{F,\rho}$.

With $F \ast f \colon \overline{v} \mapsto w$ we denote the frame F' given by

$$F'(g) = F(g) \quad \text{if } g \neq f$$
$$F'(f) = F(f) \cup \{\overline{v} \mapsto w\}$$

The 0-length of $f \colon \overline{v} \mapsto w$, denoted $\mathbf{l}_0(f \colon \overline{v} \mapsto w)$, is given by $\mathbf{l}_0(\overline{v}, w)$, its 1-length, denoted $\mathbf{l}_1(f \colon \overline{v} \mapsto w)$, by $\mathrm{ar}(f) + 1$.

Remark. The arguments $\overline{s}$ to f above are generalized variables, as Ax is nice. Thus $\rho(s_i)$ is well-defined. Furthermore, for each s_i at most one variable can occur, and such variables are distinct for different s_j's as Ax is nice, as remarked before. Hence, an update uniquely determines an axiom in Ax on which it is based.

Remark. The length $\mathbf{l}(f \colon \overline{v} \mapsto w)$ of update $f \colon \overline{v} \mapsto w$ can be bounded by

$$\mathbf{l}(f \colon \overline{v} \mapsto w) \;=\; \mathrm{O}(\mathbf{l}(f) \cdot \mathbf{l}_1(f \colon \overline{v} \mapsto w) \cdot \mathbf{l}_0(f \colon \overline{v} \mapsto w)) \ .$$

Remark. For $F' = F \ast f \colon \overline{v} \mapsto w$ we compute

- $\mathbf{l}_0(F') = \max\{\mathbf{l}_0(F), \mathbf{l}_0(\overline{v}, w)\}$,
- $\mathbf{l}_1(F') = \max\{\mathbf{l}_1(F), \mathrm{ar}(f)\}$.
- $\mathbf{l}_2(F') \leq \mathbf{l}_2(F) + 1$,
- $\#(F') \leq \#(F) + 1$,

We now formulate and prove a crucial property of updates: They can be used to extend κ-pre-models for Ax_0.

Proposition 39 (S^1_2)**.** *Let* Ax_0 *be a finite subset of* Ax*, F a κ-pre-model of* Ax_0*, $f \colon \overline{v} \mapsto w$ an update based on F, κ and* Ax_0*, and $F' = F \ast f \colon \overline{v} \mapsto w$. Then F' is a κ-pre-model of* Ax_0*.*

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

Proof. We argue in S^1_2. Let the assumption of the proposition be given, and assume that $f\colon \overline{v} \mapsto w$ is given via $t = u$ in Ax_0 and assignment ρ, where t is of the form $f(\overline{s})$ and $\overline{v} = \rho(\overline{s})$. W.l.o.g., $\mathrm{dom}(\rho) = \mathrm{Var}(t)$. We have $\mathbf{l}_0(\rho) \leq \kappa$.

In order to show that $F' = F \ * \ f\colon \overline{v} \mapsto w$ is a κ-model for Ax_0, it suffices to show that

(1) $F'(f)$ is a consistent set, and
(2) for any $t' = u'$ in Ax_0, and any assignment ρ' with $\mathbf{l}_0(\rho') \leq \kappa$, we have $[\![t']\!]_{F',\rho'} \sqsubseteq [\![u']\!]_{F',\rho'}$.

For (1), consider $\overline{v}' \mapsto w' \in F(f)$ such that $\overline{v}' \mathbin{\triangle} \overline{v}$. Then there exists $\overline{y}$ such that $\overline{v}, \overline{v}' \sqsubseteq \overline{y}$ and $\mathbf{l}_0(\overline{y}) \leq \kappa$ – we can choose y_i to be $\max_{\sqsubseteq}\{v_i, v_i'\}$, hence $\mathbf{l}_0(y_i) \leq \kappa$ follows from assumption $\mathbf{l}_0(\overline{v}_i), \mathbf{l}_0(\overline{v}_i') \leq \kappa$. Choose $\hat{\rho}$ with $\mathrm{dom}(\hat{\rho}) = \mathrm{Var}(t)$ such that $y_i = \hat{\rho}(s_i)$, which is possible since $\overline{v} \sqsubseteq \overline{y}$. We observe that $\rho \sqsubseteq \hat{\rho}$ and that $\mathbf{l}_0(\hat{\rho}) \leq \kappa$.

Let S be $F(f)[\overline{y}]$, that is

$$S \quad = \quad \{\tilde{w} \mid \exists \tilde{v}, \tilde{v} \sqsubseteq \overline{y} \text{ and } \tilde{v} \mapsto \tilde{w} \in F(f)\} \ .$$

We have $w' \in S$ as $\overline{v}' \sqsubseteq \overline{y}$, hence

$$w' \ \sqsubseteq \ \max\nolimits_{\sqsubseteq} S \ = \ F(f)(\overline{y}) = [\![t]\!]_{F,\hat{\rho}} \ \sqsubseteq \ [\![u]\!]_{F,\hat{\rho}}$$

as F is a κ-pre-model of Ax_0. Furthermore,

$$w \ = \ [\![u]\!]_{F,\rho} \ \sqsubseteq \ [\![u]\!]_{F,\hat{\rho}}$$

as $\rho \sqsubseteq \hat{\rho}$. Hence $w \mathbin{\triangle} w'$ using Lemma 20.

For (2), let $t' = u'$ be in Ax_0 and ρ' be an assignment with $\mathbf{l}_0(\rho') \leq \kappa$. If $t' = u'$ is not identical to $t = u$, then the assertion follows from F being a κ-pre-model of Ax_0: Ax being nice implies $[\![t']\!]_{F',\rho'} = [\![t']\!]_{F,\rho'}$ in this case, hence

$$[\![t']\!]_{F',\rho'} \ = \ [\![t']\!]_{F,\rho'} \ \sqsubseteq \ [\![u']\!]_{F,\rho'} \ \sqsubseteq \ [\![u']\!]_{F',\rho'}$$

as F is a κ-pre-model of Ax_0, and $F \sqsubseteq F'$.

Now consider $t' = u'$ being identical to $t = u$. Let y_i be $\rho'(s_i)$. If $\overline{v} \not\sqsubseteq \overline{y}$, then again $[\![t']\!]_{F',\rho'} = [\![t']\!]_{F,\rho'}$ and the assertion follows from F being a κ-pre-model of Ax_0 as before.

So assume $\overline{v} \sqsubseteq \overline{y}$. Let $\overline{x}$ be the list of variables occurring in t, then we have $\rho{\restriction}_{\overline{x}} \sqsubseteq \rho'{\restriction}_{\overline{x}}$. We compute

$$F'(f)(\overline{y}) \ = \ \max\nolimits_{\sqsubseteq} F'(f)[\overline{y}] \ = \ \max\nolimits_{\sqsubseteq}(\{w\} \cup F(f)[\overline{y}]) \ = \ \max\nolimits_{\sqsubseteq}\{w, F(f)(\overline{y})\} \ .$$

We consider w and $F(f)(\overline{y})$ in turns: For $F(f)(\overline{y})$ we have

$$F(f)(\overline{y}) \ = \ [\![t]\!]_{F,\rho'} \ \sqsubseteq \ [\![u]\!]_{F,\rho'}$$

as F is a κ-pre-model of Ax_0. In case of w we have,

$$w \ = \ [\![u]\!]_{F,\rho} \ = \ [\![u]\!]_{F,\rho{\restriction}_{\overline{x}}} \ \sqsubseteq \ [\![u]\!]_{F,\rho'{\restriction}_{\overline{x}}} \ = \ [\![u]\!]_{F,\rho'}$$

using $\rho{\restriction}_{\overline{x}} \sqsubseteq \rho'{\restriction}_{\overline{x}}$. Hence $F'(f)(\overline{y}) \sqsubseteq [\![u]\!]_{F,\rho'}$. Thus

$$[\![t]\!]_{F',\rho'} \ = \ F'(f)(\overline{y}) \ \sqsubseteq \ [\![u]\!]_{F,\rho'} \ \sqsubseteq \ [\![u]\!]_{F',\rho'}$$

as $F \sqsubseteq F'$. □

Definition 40. Let Ax_0 be a finite subset of Ax, and F a κ-pre-model of Ax_0. A *sequence of updates based on* F, κ *and* Ax_0 is a sequence σ of the form

$$\langle\, f_1\colon \overline{v}_1 \mapsto w_1\,, \ldots,\, f_\ell\colon \overline{v}_\ell \mapsto w_\ell \,\rangle$$

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

such that for

$$\begin{aligned} F_0 &:= F \\ F_{i+1} &:= F_i * f_{i+1} \colon \overline{v}_{i+1} \mapsto w_{i+1} \end{aligned}$$

we have that

$$f_{i+1} \colon \overline{v}_{i+1} \mapsto w_{i+1} \quad \text{is an update based on } F_i,\ \kappa \text{ and } \mathrm{Ax}_0.$$

Let $F * \sigma$ denote F_ℓ.

The 0-length of σ is given by

$$\mathbf{l}_0(\sigma) \;=\; \max\{\mathbf{l}_0(\overline{v}_1, w_1), \ldots, \mathbf{l}_0(\overline{v}_\ell, w_\ell)\} \;\;,$$

its 1-length by

$$\mathbf{l}_1(\sigma) \;=\; \max\{\mathrm{ar}(f_1), \ldots, \mathrm{ar}(f_\ell)\} + 1 \;\;.$$

The *domain sequence length of* σ, denoted $\mathbf{dom\text{-}l}(\sigma)$, is given as $\mathbf{l}(f_1) + \cdots + \mathbf{l}(f_\ell)$.

Remark. The length of σ, $\mathbf{l}(\sigma)$, can be bounded by

$$\mathbf{l}(\sigma) = \mathrm{O}(\mathbf{l}_1(\sigma) \cdot \mathbf{l}_0(\sigma) \cdot \mathbf{dom\text{-}l}(\sigma)) \;\;.$$

Remark. For $F' = F * \sigma$ we compute

- $\mathbf{l}_0(F') = \max\{\mathbf{l}_0(F), \mathbf{l}_0(\sigma)\}$,
- $\mathbf{l}_1(F') = \max\{\mathbf{l}_1(F), \mathbf{l}_1(\sigma)\}$.
- $\mathbf{l}_2(F') \leq \mathbf{l}_2(F) + \mathbf{dom\text{-}l}(\sigma)$,
- $\#(F') \leq \#(F) + \mathbf{dom\text{-}l}(\sigma)$,

Corollary 41 (S^1_2)**.** *Assuming the notions given by the previous definition, all F_i's are κ-pre-models of Ax_0, for $i \leq \ell$.*

Proof. The proof is by induction on $i \leq \ell$ using Proposition 39. □

We extend our example of the bit-flipping function from the previous section to compute updates.

Example. The bit-flipping function **flip** can be defined by a set of nice axioms $\mathrm{Ax}_{\mathbf{flip}}$:

$$\{\ \mathbf{flip}(\epsilon) = \epsilon, \quad \mathbf{flip}(x0) = \mathbf{flip}(x)1, \quad \mathbf{flip}(x1) = \mathbf{flip}(x)0\ \}\;\;.$$

Let Ax_0 be a finite subset of $\mathrm{Ax}_{\mathbf{flip}}$ given by $\mathrm{Ax}_0 = \{\mathbf{flip}(\epsilon) = \epsilon, \mathbf{flip}(x0) = \mathbf{flip}(x)1\}$. Let κ a large enough integer.

(1) Let $F_0 = \emptyset$ be the empty frame. $\mathbf{flip} \colon {*}00 \mapsto {*}1$ is an update based on F_0, κ and Ax_0, by virtue of $\rho_0 \colon x \mapsto {*}0$ and $\mathbf{flip}(x0) = \mathbf{flip}(x)1$ in Ax_0, because $\rho_0(x0) = {*}00$ and $[\![\mathbf{flip}(x)1]\!]_{F_0,\rho_0} = {*}1$.

(2) Let $F_1 = F_0 * \mathbf{flip} \colon {*}00 \mapsto {*}1$. $\mathbf{flip} \colon {*}0 \mapsto {*}1$ is an update based on F_1, κ and Ax_0, by virtue of $\rho_1 \colon x \mapsto {*}$ and $\mathbf{flip}(x0) = \mathbf{flip}(x)1$ in Ax_0, because $\rho_1(x0) = {*}0$ and $[\![\mathbf{flip}(x)1]\!]_{F_1,\rho_1} = {*}1$.

(3) Let $F_2 = F_1 * \mathbf{flip} \colon {*}0 \mapsto {*}1$. $\mathbf{flip} \colon {*}00 \mapsto {*}11$ is an update based on F_2, κ and Ax_0, by virtue of ρ_0 and $\mathbf{flip}(x0) = \mathbf{flip}(x)1$ in Ax_0, because $\rho_0(x0) = {*}00$ and $[\![\mathbf{flip}(x)1]\!]_{F_2,\rho_0} = {*}11$.

We have that

$$\sigma = \langle \mathbf{flip} \colon {*}00 \mapsto {*}1, \mathbf{flip} \colon {*}0 \mapsto {*}1, \mathbf{flip} \colon {*}00 \mapsto {*}11 \rangle$$

is a sequence of updates based on F_0, κ and Ax_0.

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

6. Soundness in S^2_2

We prove a soundness property for equational reasoning using approximation semantics. The proof will be formalizable in S^2_2. This will be improved in the remaining sections to a proof formalizable in S^1_2 by introducing an additional property. To keep the exposition clearer, we first prove soundness based on the notions introduced so far.

Theorem 42 (S^2_2)**.** *Assume $\mathcal{D} \vdash t = u$ is in Variable Normal Form (see Definition 11). Let Ax_0 consist of those axioms in Ax that occur as injective renamings in $\mathcal{D}$. Let ρ be an assignment, and F a pre-model for Ax. Let κ be $\max\{\mathbf{l}_0(F), \mathbf{l}_0(\rho)\} + \mathbf{l}(\mathcal{D})$. Then there are sequences σ_1 and σ_2 of updates based on F, κ and Ax_0 such that*

$$\begin{aligned} [\![t]\!]_{F,\rho} &\sqsubseteq [\![u]\!]_{F \,*\, \sigma_1,\rho} \\ [\![u]\!]_{F,\rho} &\sqsubseteq [\![t]\!]_{F \,*\, \sigma_2,\rho} \end{aligned}$$

The idea behind the proof of this theorem is that each subtree in $\mathcal{D}$ extends a given sequence of updates with further updates linked to the subtree so that the desired approximation properties hold. The proof will proceed by induction on the structure of the derivation tree. The reason that the considered size measures stay appropriately bounded is because the induction hypothesis is used for immediate subtrees, and used only once for each subtree.

To prove the previous theorem, we consider the following more general claim.

Claim 43 (S^2_2)**.** *Let $\mathcal{D}$ be a derivation in Variable Normal Form. Let Ax_0 consist of those axioms in Ax that occur as injective renamings in $\mathcal{D}$. Let F be a pre-model for Ax, and let U be an integer larger than $\mathbf{l}_0(F) + \mathbf{l}(\mathcal{D})$.*

Let $\mathcal{D}_0 \vdash t = u$ be a sub-derivation of $\mathcal{D}$. Let ρ be an assignment, and σ a sequence of updates based on F, U, Ax_0, satisfying

$$\begin{aligned} \mathrm{dom}(\rho) &\subseteq \mathrm{Var}(\mathcal{D}) \\ \mathbf{l}_0(\rho), \mathbf{l}_0(\sigma), \mathbf{l}_1(\sigma), \mathbf{dom}\text{-}\mathbf{l}(\sigma) &\leq U - \mathbf{l}(\mathcal{D}_0) \end{aligned}$$

Then there are sequences σ_1 and σ_2 of updates based on F, U, Ax_0 with

$$\begin{aligned} \mathbf{l}_1(\sigma_i), \mathbf{dom}\text{-}\mathbf{l}(\sigma_i) &\leq \mathbf{l}(\mathcal{D}_0) \\ \mathbf{l}_0(\sigma_i) &\leq \max\{\mathbf{l}_0(F), \mathbf{l}_0(\sigma), \mathbf{l}_0(\rho)\} + \mathbf{l}(\mathcal{D}_0) \end{aligned}$$

such that

$$\begin{aligned} [\![t]\!]_{F',\rho} &\sqsubseteq [\![u]\!]_{F' \,*\, \sigma_1,\rho} \\ [\![u]\!]_{F',\rho} &\sqsubseteq [\![t]\!]_{F' \,*\, \sigma_2,\rho} \end{aligned}$$

for $F' = F * \sigma$.

The Theorem follows from the Claim by letting $\mathcal{D}_0 = \mathcal{D}$, ρ as given, $\sigma = \langle\rangle$, and $U = \max\{\mathbf{l}_0(F), \mathbf{l}_0(\rho)\} + \mathbf{l}(\mathcal{D})$.

Proof of Claim 43. We argue in S^2_2. Fix $\mathcal{D}$, F, Ax_0 and U as in the Claim. We prove that for any $\mathcal{D}_0$, ρ, σ satisfying the conditions of the Claim, there are σ_1 and σ_2 satisfying the assertion of the Claim, by induction on $\mathbf{l}(\mathcal{D}_0)$. Thus this is proven by logarithmic induction (LIND) on a Π^b_2-property, which is available in S^2_2 by Theorem 1.

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

Let $\mathcal{D}_0$, ρ, σ be given, that are satisfying the conditions in the Claim. Let F' be $F * \sigma$. then F' is a U-pre-model of Ax_0 by Corollary 41.

We now consider cases according to the last rule applied in $\mathcal{D}_0$. If that is the **Reflexivity Rule** $\vdash t = t$, we can choose $\sigma_1 = \sigma_2 = \langle\rangle$ to satisfy the assertion of the Claim.

Axiom Rule: More interesting is the case of Axiom Rule $\vdash t = u$ with $t = u$ an injective renaming of an equation in Ax_0. W.l.o.g. we can assume that $t = u$ is in Ax, as renamings of variables would make no difference to the following argument. As Ax is nice, we have that t is of the form $f(\overline{s})$ for some $f \in \mathcal{F} \setminus \mathcal{B}$ and generalized variables $\overline{s}$. Let v_i be $\rho(s_i)$ and $w = \llbracket u \rrbracket_{F',\rho}$. We compute $\mathbf{l}_0(v_i) \leq \mathbf{l}_0(\rho) + 1$ and, using Lemma 33,

$$\mathbf{l}_0(w) \ \leq\ \max\{\mathbf{l}_0(\rho), \mathbf{l}_0(F')\} + \mathbf{l}(u) \ <\ \max\{\mathbf{l}_0(\rho), \mathbf{l}_0(F), \mathbf{l}_0(\sigma)\} + \mathbf{l}(\mathcal{D}_0) \ .$$

Let $\sigma_1 = \langle\rangle$ and $\sigma_2 = \langle f \colon \overline{v} \mapsto w\rangle$, then

$$\begin{aligned} \mathbf{l}_1(\sigma_i) \ &\leq\ \mathrm{ar}(f) \ <\ \mathbf{l}(\mathcal{D}_0) \ , \\ \mathbf{dom\text{-}l}(\sigma_i) \ &\leq\ \mathbf{l}(f) \ <\ \mathbf{l}(\mathcal{D}_0) \ , \text{ and} \\ \mathbf{l}_0(\sigma_i) \ &\leq\ \max\{\mathbf{l}_0(\sigma), \mathbf{l}_0(\rho), \mathbf{l}_0(F)\} + \mathbf{l}(\mathcal{D}_0) \ . \end{aligned}$$

Furthermore, $\llbracket t \rrbracket_{F',\rho} \sqsubseteq \llbracket u \rrbracket_{F',\rho}$ as F' is a U-pre-model of Ax_0, which proves the assertion for σ_1. For σ_2, let F'' be $F' * \sigma_2$, then we have

$$\begin{aligned} \llbracket u \rrbracket_{F',\rho} \ = \ w \ \sqsubseteq \ \max\nolimits_{\sqsubseteq} F''(f)[\overline{v}] \ &= \ F''(f)(\overline{v}) \\ &= \ F''(f)(\ldots, \rho(s_i), \ldots) \ = \ \llbracket t \rrbracket_{F'',\rho} \ . \end{aligned}$$

Symmetry Rule: For the case of Symmetry Rule, let $\mathcal{D}_1$ be the sub-derivation of $\mathcal{D}_0$ ending in $u = t$. By induction hypothesis we obtain σ_1' and σ_2' satisfying the assertion for $\mathcal{D}_1$. By choosing $\sigma_1 = \sigma_2'$ and $\sigma_2 = \sigma_1'$ we immediately fulfill the assertion for $\mathcal{D}_0$.

Transitivity Rule: If $\mathcal{D}_0$ ends with an application of the Transitivity Rule, it must be of the form

$$\dfrac{\begin{array}{cc} \mathcal{D}_1 & \mathcal{D}_2 \\ t = s & s = u \end{array}}{t = u}$$

By induction hypothesis applied to $\mathcal{D}_1$, ρ and σ, we obtain some σ_1^1 satisfying

$$\begin{aligned} \mathbf{l}_1(\sigma_1^1), \mathbf{dom\text{-}l}(\sigma_1^1) \ &\leq\ \mathbf{l}(\mathcal{D}_1) \ , \\ \mathbf{l}_0(\sigma_1^1) \ &\leq\ \max\{\mathbf{l}_0(\sigma), \mathbf{l}_0(\rho), \mathbf{l}_0(F)\} + \mathbf{l}(\mathcal{D}_1) \ , \text{ and} \\ \llbracket t \rrbracket_{F',\rho} \ &\sqsubseteq\ \llbracket s \rrbracket_{F_1^1,\rho} \ \text{ for } F_1^1 = F' * \sigma_1^1 \ . \end{aligned}$$

We compute

$$\begin{aligned} \mathbf{l}_1(\sigma * \sigma_1^1) \ &\leq\ \max\{\mathbf{l}_1(\sigma), \mathbf{l}_1(\sigma_1^1)\} \ \leq\ U - \mathbf{l}(\mathcal{D}_0) + \mathbf{l}(\mathcal{D}_1) \ <\ U - \mathbf{l}(\mathcal{D}_2) \\ \mathbf{dom\text{-}l}(\sigma * \sigma_1^1) \ &\leq\ \mathbf{dom\text{-}l}(\sigma) + \mathbf{dom\text{-}l}(\sigma_1^1) \ \leq\ U - \mathbf{l}(\mathcal{D}_0) + \mathbf{l}(\mathcal{D}_1) \ <\ U - \mathbf{l}(\mathcal{D}_2) \end{aligned}$$

and

$$\begin{aligned} \mathbf{l}_0(\sigma * \sigma_1^1) \ \leq\ &\max\{\mathbf{l}_0(F), \mathbf{l}_0(\rho), \mathbf{l}_0(\sigma)\} + \mathbf{l}(\mathcal{D}_1) \\ &\leq\ U - \mathbf{l}(\mathcal{D}_0) + \mathbf{l}(\mathcal{D}_1) \ <\ U - \mathbf{l}(\mathcal{D}_2) \end{aligned}$$

because $\mathbf{l}(\mathcal{D}_0) > \mathbf{l}(\mathcal{D}_1) + \mathbf{l}(\mathcal{D}_2)$. Thus, we can apply i.h. to $\mathcal{D}_2$, ρ and $\sigma * \sigma_1^1$, obtaining σ_1^2 satisfying

$$\begin{aligned}
\mathbf{l}_1(\sigma_1^2), \textbf{dom-l}(\sigma_1^2) &\leq \mathbf{l}(\mathcal{D}_2) \ , \\
\mathbf{l}_0(\sigma_1^2) &\leq \max\{\mathbf{l}_0(\rho), \mathbf{l}_0(F), \mathbf{l}_0(\sigma * \sigma_1^1)\} + \mathbf{l}(\mathcal{D}_2) \ , \text{ and} \\
\llbracket s \rrbracket_{F_1^1,\rho} &\sqsubseteq \llbracket u \rrbracket_{F_1^1 * \sigma_1^2,\rho} \ .
\end{aligned}$$

Let σ_1 be $\sigma_1^1 :: \sigma_1^2$, then we compute

$$\begin{aligned}
\mathbf{l}_1(\sigma_1) &= \max\{\mathbf{l}_1(\sigma_1^1), \mathbf{l}_1(\sigma_1^2)\} \leq \mathbf{l}(\mathcal{D}_1) + \mathbf{l}(\mathcal{D}_2) < \mathbf{l}(\mathcal{D}_0) \\
\textbf{dom-l}(\sigma_1) &= \textbf{dom-l}(\sigma_1^1) + \textbf{dom-l}(\sigma_1^2) \leq \mathbf{l}(\mathcal{D}_1) + \mathbf{l}(\mathcal{D}_2) < \mathbf{l}(\mathcal{D}_0)
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{l}_0(\sigma_1) &= \max\{\mathbf{l}_0(\sigma_1^1), \mathbf{l}_0(\sigma_1^2)\} \\
&\leq \max\{\mathbf{l}_0(\sigma_1^1), \max\{\mathbf{l}_0(\rho), \mathbf{l}_0(F), \mathbf{l}_0(\sigma :: \sigma_1^1)\} + \mathbf{l}(\mathcal{D}_2)\} \\
&= \max\{\mathbf{l}_0(\rho), \mathbf{l}_0(F), \mathbf{l}_0(\sigma), \mathbf{l}_0(\sigma_1^1)\} + \mathbf{l}(\mathcal{D}_2) \\
&\leq \max\{\mathbf{l}_0(F), \mathbf{l}_0(\rho), \mathbf{l}_0(\sigma)\} + \mathbf{l}(\mathcal{D}_1) + \mathbf{l}(\mathcal{D}_2) \\
&< \max\{\mathbf{l}_0(F), \mathbf{l}_0(\rho), \mathbf{l}_0(\sigma)\} + \mathbf{l}(\mathcal{D}_0)
\end{aligned}$$

Furthermore, $\llbracket t \rrbracket_{F',\rho} \sqsubseteq \llbracket s \rrbracket_{F_1^1,\rho} \sqsubseteq \llbracket u \rrbracket_{F_1^1 * \sigma_1^2,\rho} = \llbracket u \rrbracket_{F' * \sigma_1,\rho}$, because

$$F_1^1 * \sigma_1^2 = (F' * \sigma_1^1) * \sigma_1^2 = F' * (\sigma_1^1 :: \sigma_1^2) = F' * \sigma_1$$

which proves the assertion for σ_1. The construction for σ_2 is similar, starting with $\mathcal{D}_2$.

Compatibility Rule. In case of the last rule being the Compatibility Rule, $\mathcal{D}_0$ will have the form:

$$\frac{\begin{array}{c}\mathcal{D}_1 \\ t = u\end{array}}{s[t/x] = s[u/x]}$$

Applying the i.h. to $\mathcal{D}_1$, ρ and σ, we obtain σ_1 and σ_2 satisfying the assertion for $\mathcal{D}_1$. Let $\rho_1^1 = \rho[x \mapsto \llbracket t \rrbracket_{F,\rho}]$ and $\rho_1^2 = \rho[x \mapsto \llbracket u \rrbracket_{F' * \sigma_1,\rho}]$. Then we have $\rho_1^1 \sqsubseteq \rho_1^2$. Employing the Substitution Lemma 34, we obtain

$$\llbracket s[t/x] \rrbracket_{F',\rho} = \llbracket s \rrbracket_{F',\rho_1^1} \sqsubseteq \llbracket s \rrbracket_{F' * \sigma_1,\rho_1^2} = \llbracket s[u/x] \rrbracket_{F' * \sigma_1,\rho}$$

which shows that σ_1 also satisfies the assertion for $\mathcal{D}_0$. Similar for σ_2 and $\mathcal{D}_0$.

Substitution Rule. If $\mathcal{D}_0$ ends in an application of Substitution, it will have the following form:

$$\frac{\begin{array}{c}\mathcal{D}_1 \\ t = u\end{array}}{t[s/x] = u[s/x]}$$

We only consider the case that x is occurring in $t = u$, the other case is trivial.

Let ρ' be $\rho[x \mapsto [\![s]\!]_{F',\rho}]$, then clearly $\mathrm{dom}(\rho') \subseteq \mathrm{dom}(\rho) \subseteq \mathrm{Var}(\mathcal{D})$. Furthermore, using Lemma 33, we obtain

$$\begin{aligned} \mathbf{l}_0(\rho') &\leq \max\{\mathbf{l}_0(\rho), \mathbf{l}_0([\![s]\!]_{F',\rho})\} \\ &\leq \max\{\mathbf{l}_0(\rho), \max\{\mathbf{l}_0(\rho), \mathbf{l}_0(F')\} + \mathbf{l}(s)\} \\ &= \max\{\mathbf{l}_0(\rho), \mathbf{l}_0(F')\} + \mathbf{l}(s) \\ &= \max\{\mathbf{l}_0(\rho), \mathbf{l}_0(F), \mathbf{l}_0(\sigma)\} + \mathbf{l}(s) \end{aligned}$$

hence

$$\mathbf{l}_0(\rho') \leq U - \mathbf{l}(\mathcal{D}_0) + \mathbf{l}(s) < U - \mathbf{l}(\mathcal{D}_1)$$

because

$$\mathbf{l}(\mathcal{D}_0) = \mathbf{l}(\mathcal{D}_1) + \mathbf{l}(t[s/x] = u[s/x]) > \mathbf{l}(\mathcal{D}_1) + \mathbf{l}(s) .$$

Thus we can apply the i.h. to $\mathcal{D}_1$, ρ' and σ, obtaining σ_1 and σ_2 such that

$$\mathbf{l}_1(\sigma_i), \textbf{dom-l}(\sigma_i) \leq \mathbf{l}(\mathcal{D}_1) < \mathbf{l}(\mathcal{D}_0)$$

and

$$\begin{aligned} \mathbf{l}_0(\sigma_i) &\leq \max\{\mathbf{l}_0(F), \mathbf{l}_0(\sigma), \mathbf{l}_0(\rho')\} + \mathbf{l}(\mathcal{D}_1) \\ &\leq \max\{\mathbf{l}_0(F), \mathbf{l}_0(\sigma), \mathbf{l}_0(\rho)\} + \mathbf{l}(s) + \mathbf{l}(\mathcal{D}_1) \\ &< \max\{\mathbf{l}_0(F), \mathbf{l}_0(\sigma), \mathbf{l}_0(\rho)\} + \mathbf{l}(\mathcal{D}_0) . \end{aligned}$$

Furthermore,

$$[\![t]\!]_{F',\rho'} \sqsubseteq [\![u]\!]_{F' * \sigma_1,\rho'} .$$

Now we can compute, employing the Substitution Lemma 34,

$$\begin{aligned} [\![t[s/x]]\!]_{F',\rho} &= [\![t]\!]_{F',\rho'} \\ &\sqsubseteq [\![u]\!]_{F' * \sigma_1,\rho'} = [\![u]\!]_{F' * \sigma_1,\rho[x \mapsto [\![s]\!]_{F',\rho}]} \\ &\sqsubseteq [\![u]\!]_{F' * \sigma_1,\rho[x \mapsto [\![s]\!]_{F' * \sigma_1,\rho}]} = [\![u[s/x]]\!]_{F' * \sigma_1,\rho} , \end{aligned}$$

which proves the assertion for σ_1 and $\mathcal{D}_0$. Similar for σ_2 and $\mathcal{D}_0$. □

Corollary 44. *The consistency of* $\mathbf{PETS}(\mathrm{Ax})$ *is provable in* S^2_2.

Proof. We argue in S^2_2. Assume for sake of contradiction, that $\mathcal{D}$ is a $\mathbf{PETS}(\mathrm{Ax})$ derivation ending in $0 = 1$. Using Proposition 12 we can assume that $\mathcal{D}$ is in Variable Normal Form. Let Ax_0 consist of those axioms in Ax that occur as injective renamings in $\mathcal{D}$. Let ρ be the empty assignment, and F be the empty pre-model for Ax. Let $\kappa = \mathbf{l}(\mathcal{D})$. By the previous Theorem 42, there is a sequence σ_1 of updates based on F, κ and Ax_0 such that

$$0 = [\![0]\!]_{F,\rho} \sqsubseteq [\![1]\!]_{F * \sigma_1,\rho} = 1$$

which is impossible. □

To finish this section, we apply the construction of the proof of Claim 43 to the example derivation $\mathcal{D}_{\exp}$ from Section 3.

Example. Consider the derivation $\mathcal{D}_{\exp}$ of $\mathrm{e}(\epsilon, a \oplus b0) = \epsilon$ from Section 3. Let $F_0 = \emptyset$ and $\rho_0 = \emptyset$. We construct an update σ such that

$$\epsilon = [\![\epsilon]\!]_{F_0,\rho_0} \sqsubseteq [\![\mathrm{e}(\epsilon, a \oplus b0)]\!]_{F_0 * \sigma,\rho_0}$$

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

by following the proof of Claim 43. Hence, as we are considering the approximation assertion from right to left in relation to the conclusion of $\mathcal{D}_{\mathrm{exp}}$, we also need to traverse the derivation in the same order from right to left.

Starting from the right in $\mathcal{D}_{\mathrm{exp}}$, consider

$$\text{Su}\,\dfrac{\text{Ax}\,\dfrac{}{x_3 \otimes \epsilon \;=\; \epsilon}}{\mathrm{e}(\epsilon, a \oplus b) \otimes \epsilon \;=\; \epsilon}$$

For the application of substitution rule, we let

$$\rho_1(x_3) = \llbracket \mathrm{e}(\epsilon, a \oplus b) \rrbracket_{F_0, \rho_0} = * \;\; .$$

From the axiom we obtain the update $\sigma_1 = \langle \otimes\colon (*, \epsilon) \mapsto \epsilon \rangle$ and compute

$$\llbracket x_3 \otimes \epsilon) \rrbracket_{F_0 * \sigma_1, \rho_1} = \epsilon \;\; ,$$

hence

$$\llbracket \mathrm{e}(\epsilon, a \oplus b) \otimes \epsilon) \rrbracket_{F_0 * \sigma_1, \rho_0} = \epsilon \;\; .$$

To continue in the derivation according to transitivity, consider

$$\text{Su}\,\dfrac{\text{Ax}\,\dfrac{}{\mathrm{e}(x_2, y_2 0) \;=\; \mathrm{e}(x_2, y_2) \otimes x_2}}{\mathrm{e}(\epsilon, (a \oplus b)0) \;=\; \mathrm{e}(\epsilon, a \oplus b) \otimes \epsilon}$$

and define ρ_2 via $x_2 \mapsto \epsilon$ and $y_2 \mapsto \llbracket a \oplus b \rrbracket_{F_0 * \sigma_1, \rho_0} = *$ following substitution. The axiom gives rise to an update $\mathrm{e}\colon (\epsilon, *0) \mapsto \epsilon$, as $\llbracket \mathrm{e}(x_2, y_2) \otimes x_2 \rrbracket_{F_0 * \sigma_1, \rho_2} = \epsilon$. Let $\sigma_2 = \langle \otimes\colon (*, \epsilon) \mapsto \epsilon,\ \mathrm{e}\colon (\epsilon, *0) \mapsto \epsilon \rangle$, then

$$\llbracket \mathrm{e}(\epsilon, (a \oplus b)0)) \rrbracket_{F_0 * \sigma_2, \rho_0} = \epsilon \;\; .$$

Again continuing in the derivation to the left according to transitivity, consider

$$\text{Co}\,\dfrac{\text{Su}\,\dfrac{\text{Ax}\,\dfrac{}{x_1 \oplus y_1 0 \;=\; (x_1 \oplus y_1)0}}{a \oplus b0 \;=\; (a \oplus b)0}}{\mathrm{e}(\epsilon, a \oplus b0) \;=\; \mathrm{e}(\epsilon, (a \oplus b)0)}$$

From the application of composition and substitution rules, we let ρ_3 be given by $x_1 \mapsto \llbracket a \rrbracket_{F_0 * \sigma_2, \rho_0} = *$ and $y_1 \mapsto \llbracket b \rrbracket_{F_0 * \sigma_2, \rho_0} = *$. Then the axiom gives rise to an update $\oplus\colon (*, *0) \mapsto *0$ as $\llbracket (x_1 \oplus y_1)0 \rrbracket_{F_0 * \sigma_2, \rho_3} = *0$. Let

$$\sigma \quad = \quad \langle\ \otimes\colon (*, \epsilon) \mapsto \epsilon,\ \ \mathrm{e}\colon (\epsilon, *0) \mapsto \epsilon,\ \ \oplus\colon (*, *0) \mapsto *0\ \rangle \;\; ,$$

then $\llbracket \mathrm{e}(\epsilon, a \oplus b0)) \rrbracket_{F_0 * \sigma, \rho_0} = \epsilon$.

7. Instructions and Frame Pre-Models

In this and the next section, we prove our main theorem in S^1_2. The idea is that the proof of Claim 43 follows a simple, explicit procedure to construct the new update sequences σ_1 and σ_2, which can replace the existential quantifier of the induction statement of Claim 43. In fact, the derivation from which the update sequences are constructed, contains explicitly all information needed in the constructions. We make this extraction explicit in this section, in form of so called 'instructions'. In the following Section 8, we will use instructions to transform the statement and proof of Claim 43 into a similar statement and proof (in Claim 54), where the proof now is in S^1_2.

We start by naming the instructions that will be considered.

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

Definition 45 (Instructions)**.** We define a set of *instructions* and their *length* as follows:

Axiom: $\mathrm{A}[t \to u]$ and $\mathrm{A}[t \leftarrow u]$ are instructions, for any axiom $t = u \in \mathrm{Ax}$. Their length is $\mathbf{l}(t) + \mathbf{l}(u) + 1$.

Substitution: $\mathrm{S}{\uparrow}[s, t/x]$ and $\mathrm{S}{\downarrow}[s, t/x]$ are instructions, for terms s, t and variable x. Their length is $\mathbf{l}(s) + \mathbf{l}(t) + 1$.

Sequences of instructions will be denoted with τ. With **dom-l**(τ) we denote the sequence length of τ, that is the number of instructions occurring in τ. With $\mathbf{l}(\tau)$ we denote the length of τ given as the sum of lengths of instructions occurring in them.

We now define the process of turning derivations into sequences of related instructions.

Definition 46. For a derivation $\mathcal{D}$, we define *sequences of instructions* $\overrightarrow{\mathrm{Inst}}_{\mathcal{D}}$ and $\overleftarrow{\mathrm{Inst}}_{\mathcal{D}}$ by recursion on $\mathcal{D}$ as follows.

Axiom Rule: If D is of the form

$$\mathrm{Ax}\,\frac{}{t = u}$$

let

$$\begin{aligned} \overrightarrow{\mathrm{Inst}}_{\mathcal{D}} &:= \langle \mathrm{A}[t \to u] \rangle \\ \overleftarrow{\mathrm{Inst}}_{\mathcal{D}} &:= \langle \mathrm{A}[t \leftarrow u] \rangle \ . \end{aligned}$$

Reflexivity Rule: If D is of the form

$$\mathrm{Re}\,\frac{}{t = t}$$

let

$$\overrightarrow{\mathrm{Inst}}_{\mathcal{D}} := \overleftarrow{\mathrm{Inst}}_{\mathcal{D}} := \langle \rangle \ .$$

Symmetry Rule: Consider $\mathcal{D}$ of the form

$$\mathrm{Sy}\,\frac{\begin{matrix}\mathcal{D}_1 \\ u = t\end{matrix}}{t = u}$$

Let $\overrightarrow{\mathrm{Inst}}_{\mathcal{D}_1}$ and $\overleftarrow{\mathrm{Inst}}_{\mathcal{D}_1}$ be given by i.h., then define

$$\begin{aligned} \overrightarrow{\mathrm{Inst}}_{\mathcal{D}} &:= \overleftarrow{\mathrm{Inst}}_{\mathcal{D}_1} \\ \overleftarrow{\mathrm{Inst}}_{\mathcal{D}} &:= \overrightarrow{\mathrm{Inst}}_{\mathcal{D}_1} \ . \end{aligned}$$

Transitivity Rule: Consider $\mathcal{D}$ of the form

$$\mathrm{Tr}\,\frac{\begin{matrix}\mathcal{D}_1 \\ t = s\end{matrix} \qquad \begin{matrix}\mathcal{D}_2 \\ s = u\end{matrix}}{t = u}$$

Let $\overrightarrow{\mathrm{Inst}}_{\mathcal{D}_1}$, $\overleftarrow{\mathrm{Inst}}_{\mathcal{D}_1}$, $\overrightarrow{\mathrm{Inst}}_{\mathcal{D}_2}$ and $\overleftarrow{\mathrm{Inst}}_{\mathcal{D}_2}$ be given by i.h. Define

$$\begin{aligned} \overrightarrow{\mathrm{Inst}}_{\mathcal{D}} &:= \overrightarrow{\mathrm{Inst}}_{\mathcal{D}_1} :: \overrightarrow{\mathrm{Inst}}_{\mathcal{D}_2} \\ \overleftarrow{\mathrm{Inst}}_{\mathcal{D}} &:= \overleftarrow{\mathrm{Inst}}_{\mathcal{D}_2} :: \overleftarrow{\mathrm{Inst}}_{\mathcal{D}_1} \ . \end{aligned}$$

Compatibility Rule: Consider $\mathcal{D}$ of the form

$$\mathrm{Co}\,\frac{\begin{matrix}\mathcal{D}_1 \\ t = u\end{matrix}}{s[t/x] = s[u/x]}$$

Let $\overrightarrow{\mathrm{Inst}}_{\mathcal{D}_1}$ and $\overleftarrow{\mathrm{Inst}}_{\mathcal{D}_1}$ be given by i.h., then define

$$\begin{aligned}\overrightarrow{\mathrm{Inst}}_{\mathcal{D}} &:= \overrightarrow{\mathrm{Inst}}_{\mathcal{D}_1}\\ \overleftarrow{\mathrm{Inst}}_{\mathcal{D}} &:= \overleftarrow{\mathrm{Inst}}_{\mathcal{D}_1} \quad .\end{aligned}$$

Substitution Rule: If $\mathcal{D}$ is of the form

$$\mathrm{Su}\,\frac{\begin{matrix}\mathcal{D}_1\\ t = u\end{matrix}}{t[s/x] = u[s/x]}$$

then let

$$\begin{aligned}\overrightarrow{\mathrm{Inst}}_{\mathcal{D}} &:= \mathrm{S}{\uparrow}[t, s/x] : \overrightarrow{\mathrm{Inst}}_{\mathcal{D}_1} : \mathrm{S}{\downarrow}[u, s/x]\\ \overleftarrow{\mathrm{Inst}}_{\mathcal{D}} &:= \mathrm{S}{\uparrow}[u, s/x] : \overleftarrow{\mathrm{Inst}}_{\mathcal{D}_1} : \mathrm{S}{\downarrow}[t, s/x] \quad .\end{aligned}$$

Remark. We observe that $\mathbf{l}(\overrightarrow{\mathrm{Inst}}_{\mathcal{D}}) = \mathbf{l}(\overleftarrow{\mathrm{Inst}}_{\mathcal{D}}) \leq \mathbf{l}(\mathcal{D})$.

We will now describe a process of evaluating terms using approximations along sequences of instruction. We start with the most basic and also most interesting step of the reverse direction of an axiom instruction.

For the remainder of this section, let κ and $\mathcal{D}$ be given. Let Ax_0 consist of those axioms in Ax that occur as injective renamings in $\mathcal{D}$.

Definition 47. Let $t = u$ be an axiom in Ax_0, ρ an assignment, and F a κ-pre-model of Ax_0. Define $\Psi(t \leftarrow u, \langle F, \rho\rangle)$ to be $f\colon \overline{v} \mapsto w$ satisfying

- t is of the form $f(\overline{s})$ for some terms $\overline{s}$;
- $v_i = \rho(s_i)$ for $i \leq \mathrm{ar}(f)$;
- and $w = [\![u]\!]_{F,\rho}$.

For a sequence σ of updates based on F, κ and Ax_0, let $\Psi(t \leftarrow u, \langle F, \sigma, \rho\rangle)$ be $\Psi(t \leftarrow u, \langle F * \sigma, \rho\rangle)$.

Lemma 48. *Let $t = u$ be an axiom in Ax_0, κ' a positive integer with $\kappa' \leq \kappa - \mathbf{l}(u)$, ρ an assignment with $\mathbf{l}_0(\rho) \leq \kappa'$, and F a κ-pre-model of Ax_0 with $\mathbf{l}_0(F) \leq \kappa'$. Let $f\colon \overline{v} \mapsto w$ be given by $\Psi(t \leftarrow u, \langle F, \rho\rangle)$. Then $f\colon \overline{v} \mapsto w$ is an update based on F, κ and Ax_0, satisfying that $\mathbf{l}_0(\overline{v}, w) \leq \kappa' + \mathbf{l}(u)$ and*

$$[\![u]\!]_{F,\rho} \;\sqsubseteq\; [\![t]\!]_{F * f:\overline{v}\mapsto w,\ \rho} \quad .$$

Proof. As Ax is nice, we have that t is of the form $f(\overline{s})$ for some $f \in \mathcal{F} \setminus \mathcal{B}$ and generalized variables $\overline{s}$ (see Definition 7). Then $v_i = \rho(s_i)$ and $w = [\![u]\!]_{F,\rho}$. We compute $\mathbf{l}_0(v_i) \leq \mathbf{l}_0(\rho) + 1 \leq \kappa' + \mathbf{l}(u) \leq \kappa$, and, using Lemma 33,

$$\mathbf{l}_0(w) \;\leq\; \max\{\mathbf{l}_0(\rho), \mathbf{l}_0(F)\} + \mathbf{l}(u) \;\leq\; \kappa' + \mathbf{l}(u) \;\leq\; \kappa$$

Hence, $f\colon \overline{v} \mapsto w$ is an update based on F, κ and Ax_0.

Furthermore, for $F' = F * f\colon \overline{v} \mapsto w$, we have

$$\begin{aligned}[\![u]\!]_{F,\rho} \;=\; w \;\sqsubseteq\; \max\nolimits_{\sqsubseteq} F'(f)[\overline{v}] \;&=\; F'(f)(\overline{v})\\ &=\; F'(f)(\dots, \rho(s_i) \dots) \;=\; [\![t]\!]_{F',\rho} \quad .\end{aligned}$$

□

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

Definition 49. Let τ be a sequence of instructions, ρ an assignment, F a κ-pre-model for Ax_0, and σ a sequence of updates based on F, κ and Ax_0. Let $\alpha = \langle F, \sigma, \rho\rangle$. We define $\Phi(\tau, \alpha) = \langle F, \sigma', \rho'\rangle$ by induction on τ:

If τ is the empty sequence, let $\Phi(\langle\rangle, \alpha) = \alpha$

Otherwise, τ is of the form $\tau' : I$ for some instruction I. Let $\langle F, \sigma', \rho'\rangle = \Phi(\tau', \alpha)$ by i.h. We consider cases according to the form of I:

Axiom: For $I = \mathrm{A}[t \to u]$ let $\Phi(\tau, \alpha) = \langle F, \sigma', \rho'\rangle$.
For $I = \mathrm{A}[t \leftarrow u]$ let $\nu = \Psi(t \leftarrow u, \langle F, \sigma', \rho'\rangle)$, and define

$$\Phi(\tau, \alpha) \;=\; \langle F, \sigma' * \nu, \rho'\rangle \enspace .$$

Substitution: If $I = \mathrm{S}{\uparrow}[t, s/x]$, let

$$\Phi(\tau, \alpha) \;=\; \langle F, \sigma', \rho'[x \mapsto [\![s]\!]_{F*\sigma',\,\rho'}]\rangle \enspace .$$

If $I = \mathrm{S}{\downarrow}[t, s/x]$, let ρ'' be ρ' with x removed from its domain: $\rho'' = \rho' \restriction_{\mathrm{dom}(\rho')\setminus\{x\}}$. Then let

$$\Phi(\tau, \alpha) \;=\; \langle F, \sigma', \rho''\rangle \enspace .$$

Lemma 50. *Let τ be a sequence of instructions for $\mathcal{D}$, ρ an assignment, F a κ-pre-model of* Ax_0*, and σ a sequence of updates based on F, κ and* Ax_0*, satisfying*

$$\max\{\mathbf{l}_0(\rho), \mathbf{l}_0(F), \mathbf{l}_0(\sigma)\} + \mathbf{l}(\tau) \;\le\; \kappa \enspace .$$

Let $\langle F, \sigma', \rho'\rangle$ be $\Phi(\tau, \langle F, \sigma, \rho\rangle)$. Then we have

1. *σ' is a sequence of updates based on F, κ and* Ax_0*;*
2. $\mathbf{dom}\text{-}\mathbf{l}(\sigma') \;\le\; \mathbf{dom}\text{-}\mathbf{l}(\sigma) + \mathbf{l}(\tau)$*;*
3. $\mathbf{l}_0(\rho'), \mathbf{l}_0(\sigma') \;\le\; \max\{\mathbf{l}_0(\rho), \mathbf{l}_0(F), \mathbf{l}_0(\sigma)\} + \mathbf{l}(\tau)$*.*
4. $\mathbf{l}_1(\sigma') \;\le\; \mathbf{l}_1(\sigma) + \mathbf{l}(\tau)$*;*

Proof. Assume $\tau = \langle I_1, \ldots, I_\ell\rangle$. Let τ_i be the sequence consisting of the first i elements in τ, for $i = 0, \ldots, \ell$. Let $\langle F, \sigma_i, \rho_i\rangle$ be $\Phi(\tau_i, \langle F, \sigma, \rho\rangle)$. We can show by induction on i that

1. σ_i is a sequence of updates based on F, κ and Ax_0;
2. $\mathbf{dom}\text{-}\mathbf{l}(\sigma_i) \;\le\; \mathbf{dom}\text{-}\mathbf{l}(\sigma) + \mathbf{l}(\tau_i)$;
3. $\mathbf{l}_0(\rho_i), \mathbf{l}_0(\sigma_i) \;\le\; \max\{\mathbf{l}_0(\rho), \mathbf{l}_0(F), \mathbf{l}_0(\sigma)\} + \mathbf{l}(\tau_i)$.
4. $\mathbf{l}_1(\sigma_i) \;\le\; \mathbf{l}_1(\sigma) + \mathbf{l}(\tau_i)$;

For $i = 0$ there is nothing to show as $\sigma_0 = \sigma$ and $\rho_0 = \rho$.

In the induction step from i to $i+1$ we have $\tau_{i+1} = \tau_i : I$ for some instruction I. We consider cases according to I.

If $I = \mathrm{A}[t \to u]$ or $I = \mathrm{S}{\downarrow}[t, s/x]$, there is nothing to show as $\sigma_{i+1} = \sigma_i$ and $\mathbf{l}_0(\rho_{i+1}) \le \mathbf{l}_0(\rho_i)$.

In case $I = \mathrm{S}{\uparrow}[t, s/x]$ we have $\sigma_{i+1} = \sigma_i$ and $\rho_{i+1} = \rho_i[x \mapsto [\![s]\!]_{F\,*\,\sigma, \rho_i}]$. Thus assertion (1) and (2) follow immediately from i.h. For assertion (3) we compute, using Lemma 33,

$$\mathbf{l}_0([\![s]\!]_{F\,*\,\sigma, \rho_i}) \;\le\; \max\{\mathbf{l}_0(F), \mathbf{l}_0(\sigma_i), \mathbf{l}_0(\rho_i)\} + \mathbf{l}(s) \enspace .$$

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

Hence, using i.h.

$$\begin{aligned}
\mathbf{l}_0(\rho_{i+1}) &\leq \max\{\mathbf{l}_0(\rho_i), \mathbf{l}_0([\![s]\!]_{F * \sigma, \rho_i})\} \\
&\leq \max\{\mathbf{l}_0(F), \mathbf{l}_0(\sigma_i), \mathbf{l}_0(\rho_i)\} + \mathbf{l}(s) \\
&\leq \max\{\mathbf{l}_0(F), \mathbf{l}_0(\sigma), \mathbf{l}_0(\rho)\} + \mathbf{l}(\tau_i) + \mathbf{l}(s) \\
&< \max\{\mathbf{l}_0(F), \mathbf{l}_0(\sigma), \mathbf{l}_0(\rho)\} + \mathbf{l}(\tau_{i+1}) \ .
\end{aligned}$$

In case $I = \mathrm{A}[t \leftarrow u]$ we have $\rho_{i+1} = \rho_i$. Let $\nu = \Psi(t \leftarrow u, \langle F, \sigma_i, \rho_i \rangle)$. By Lemma 48 we obtain that ν is an update based on F, κ and $\mathcal{D}$, and that

$$\mathbf{l}_0(\nu) \leq \max\{\mathbf{l}_0(F), \mathbf{l}_0(\sigma_i), \mathbf{l}_0(\rho_i)\} + \mathbf{l}(u) \ .$$

The former immediately implies assertion (1) for σ_{i+1}. The latter implies, using i.h.

$$\begin{aligned}
\mathbf{l}_0(\sigma_{i+1}) &\leq \max\{\mathbf{l}_0(\sigma_i), \mathbf{l}_0(\nu)\} \\
&\leq \max\{\mathbf{l}_0(F), \mathbf{l}_0(\sigma_i), \mathbf{l}_0(\rho_i)\} + \mathbf{l}(u) \\
&\leq \max\{\mathbf{l}_0(F), \mathbf{l}_0(\rho), \mathbf{l}_0(\sigma)\} + \mathbf{l}(\tau_i) + \mathbf{l}(u) \\
&< \max\{\mathbf{l}_0(F), \mathbf{l}_0(\rho), \mathbf{l}_0(\sigma)\} + \mathbf{l}(\tau_{i+1}) \ .
\end{aligned}$$

Thus assertion (3) follows.

For assertion (2) we compute using i.h.

$$\begin{aligned}
\textbf{dom-l}(\sigma_{i+1}) &= \textbf{dom-l}(\sigma_i) + \mathbf{l}(I_{i+1}) \\
&\leq \textbf{dom-l}(\sigma) + \mathbf{l}(\tau_i) + \mathbf{l}(I_{i+1}) = \textbf{dom-l}(\sigma) + \mathbf{l}(\tau_{i+1}) \ .
\end{aligned}$$

For assertion (4) we compute using i.h.

$$\mathbf{l}_1(\sigma_{i+1}) = \max\{\mathbf{l}_1(\sigma_i), \mathbf{l}_1(\nu)\} \leq \mathbf{l}_1(\sigma) + \mathbf{l}(\tau_i) + \mathbf{l}(t) < \mathbf{l}_1(\sigma) + \mathbf{l}(\tau_{i+1}) \ .$$

□

Lemma 51. *Consider $\tau = \tau_1 :: \tau_2$. Then*

$$\Phi(\tau, \langle F, \sigma, \rho \rangle) \quad = \quad \Phi(\tau_2, \Phi(\tau_1, \langle F, \sigma, \rho \rangle))$$

Proof. Immediate by induction on τ_1. □

We finish this section by applying the construction of sequences of instructions and their evaluation (Definitions 45, 46 47, 49) to the example derivation $\mathcal{D}_{\exp}$ from Section 3.

Example. To align to the application of the proof of Claim 43 the example derivation $\mathcal{D}_{\exp}$ from Section 3 as worked out at the end of Section 6, we compute $\overleftarrow{\mathrm{Inst}}_{\mathcal{D}_{\exp}}$ as defined in Definitions 45, 46.

Considering $\mathcal{D}_1$ given by

$$\mathrm{Su}\;\dfrac{\mathrm{Ax}\;\dfrac{}{x_3 \otimes \epsilon \ = \ \epsilon}}{\mathrm{e}(\epsilon, a \oplus b) \otimes \epsilon \ = \ \epsilon}$$

we compute $\overleftarrow{\mathrm{Inst}}_{\mathcal{D}_1}$ as

$$\langle \ \mathrm{S}{\uparrow}[\epsilon, \mathrm{e}(\epsilon, a \oplus b)/x_3], \ \mathrm{A}[x_3 \otimes \epsilon \to \epsilon], \ \mathrm{S}{\downarrow}[x_3 \otimes \epsilon, \mathrm{e}(\epsilon, a \oplus b)/x_3] \ \rangle$$

To continue we consider $\mathcal{D}_2$ given by

$$\mathrm{Su}\;\begin{array}{c} \mathrm{Ax}\;\dfrac{}{\mathrm{e}(x_2, y_2 0) \ = \ \mathrm{e}(x_2, y_2) \otimes x_2} \\ \hline\hline \mathrm{e}(\epsilon, (a \oplus b)0) \ = \ \mathrm{e}(\epsilon, a \oplus b) \otimes \epsilon \end{array}$$

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

and compute $\overleftarrow{\mathrm{Inst}}_{\mathcal{D}_2}$ as

$$\begin{aligned}
\langle\; &\mathrm{S}{\uparrow}[\mathrm{e}(\epsilon, y_2) \otimes \epsilon, a \oplus b/y_2],\; \mathrm{S}{\uparrow}[\mathrm{e}(x_2, y_2) \otimes x_2, \epsilon/x_2],\\
&\mathrm{A}[\mathrm{e}(x_2, y_2 0) \to \mathrm{e}(x_2, y_2) \otimes x_2],\\
&\mathrm{S}{\downarrow}[\mathrm{e}(x_2, y_2) \otimes x_2, \epsilon/x_2],\; \mathrm{S}{\downarrow}[\mathrm{e}(\epsilon, y_2) \otimes \epsilon, a \oplus b/y_2] \;\rangle
\end{aligned}$$

Similarly, for derivation $\mathcal{D}_3$ given by

$$\mathrm{Co}\;\dfrac{\mathrm{Su}\;\dfrac{\mathrm{Ax}\;\dfrac{}{x_1 \oplus y_1 0 \;=\; (x_1 \oplus y_1)0}}{a \oplus b0 \;=\; (a \oplus b)0}}{\mathrm{e}(\epsilon, a \oplus b0) \;=\; \mathrm{e}(\epsilon, (a \oplus b)0)}$$

we compute $\overleftarrow{\mathrm{Inst}}_{\mathcal{D}_3}$ as

$$\begin{aligned}
\langle\; &\mathrm{S}{\uparrow}[(a \oplus y_1)0, b/y_1],\; \mathrm{S}{\uparrow}[(x_1 \oplus y_1)0, a/x_1],\\
&\mathrm{A}[x_1 \oplus y_1 0 \to (x_1 \oplus y_1)0],\\
&\mathrm{S}{\downarrow}[(x_1 \oplus y_1)0, a/x_1],\; \mathrm{S}{\downarrow}[(a \oplus y_1)0, b/y_1] \;\rangle
\end{aligned}$$

As $\mathcal{D}_{\mathrm{exp}}$ is formed from derivations $\mathcal{D}_1$, $\mathcal{D}_2$, $\mathcal{D}_3$ by applications of transitivity, we obtain $\overleftarrow{\mathrm{Inst}}_{\mathcal{D}_{\mathrm{exp}}}$ as

$$\overleftarrow{\mathrm{Inst}}_{\mathcal{D}_1} :: \overleftarrow{\mathrm{Inst}}_{\mathcal{D}_2} :: \overleftarrow{\mathrm{Inst}}_{\mathcal{D}_3} \;\;.$$

Let α_0 be $\langle \emptyset, \emptyset, \emptyset \rangle$, then we can compute $\Phi(\overleftarrow{\mathrm{Inst}}_{\mathcal{D}_{\mathrm{exp}}}, \alpha_0)$ according to Definitions 47 and 49: We have

$$\Phi(\langle \mathrm{S}{\uparrow}[\epsilon, \mathrm{e}(\epsilon, a \oplus b)/x_3] \rangle, \alpha_0) \quad = \quad \langle \emptyset, \emptyset, \{x_3 \mapsto *\} \rangle$$

as $[\![\mathrm{e}(\epsilon, a \oplus b)]\!]_{\emptyset,\emptyset} = *$.

For $\mathrm{A}[x_3 \otimes \epsilon \to \epsilon]$ we consider

$$\Psi(x_3 \otimes \epsilon \to \epsilon, \{x_3 \mapsto *\}) \quad = \quad \otimes\colon (*, \epsilon) \mapsto \epsilon$$

hence

$$\begin{aligned}
&\Phi(\langle\; \mathrm{S}{\uparrow}[\epsilon, \mathrm{e}(\epsilon, a \oplus b)/x_3],\; \mathrm{A}[x_3 \otimes \epsilon \to \epsilon],\; \rangle, \alpha_0)\\
&\quad = \quad \langle \emptyset, \langle \otimes\colon (*, \epsilon) \mapsto \epsilon \rangle, \{x_3 \mapsto *\} \rangle \;\;.
\end{aligned}$$

Hence

$$\Phi(\overleftarrow{\mathrm{Inst}}_{\mathcal{D}_1}, \alpha_0) \quad = \quad \langle \emptyset, \langle \otimes\colon (*, \epsilon) \mapsto \epsilon \rangle, \emptyset \rangle \;\;.$$

Observe that this matches with the construction in the example at the end of Section 6. Indeed, continuing the computation yields

$$\Phi(\overleftarrow{\mathrm{Inst}}_{\mathcal{D}_{\mathrm{exp}}}, \alpha_0) \quad = \quad \langle \emptyset, \sigma, \emptyset \rangle$$

for the same

$$\sigma \quad = \quad \langle\; \otimes\colon (*, \epsilon) \mapsto \epsilon,\;\; \mathrm{e}\colon (\epsilon, *0) \mapsto \epsilon,\;\; \oplus\colon (*, *0) \mapsto *0 \;\rangle$$

as computed at the end of Section 6. We will make use of this relation in the next section.

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

8. Soundness in S^1_2

We are now in the position to prove a form of soundness of pure equational reasoning in S^1_2. As a reminder, $\mathrm{BVar}(\mathcal{D})$ denotes the set of variables occurring in $\mathcal{D}$ that are bound by an application of substitution, see Definition 9.

Lemma 52. *Let $\mathcal{D}$ be a derivation in Variable Normal Form (see Definition 11), ρ an assignment such that $\mathrm{dom}(\rho)$ and $\mathrm{BVar}(\mathcal{D})$ are disjoint. Let τ be $\overrightarrow{\mathrm{Inst}}_{\mathcal{D}}$ or $\overleftarrow{\mathrm{Inst}}_{\mathcal{D}}$, and let $\langle F, \sigma', \rho' \rangle$ be $\Phi(\tau, \langle F, \sigma, \rho \rangle)$. Then $\rho' = \rho$.*

Proof. By induction on $\mathcal{D}$. We only consider the case for $\overrightarrow{\mathrm{Inst}}_{\mathcal{D}}$, the case of $\overleftarrow{\mathrm{Inst}}_{\mathcal{D}}$ will be similar. The only rule which changes ρ is an application of Substitution. In this case, $\mathcal{D}$ will be of the form

$$\dfrac{\begin{array}{c}\mathcal{D}_1\\ t = u\end{array}}{t[s/x] = u[s/x]}$$

and $\overrightarrow{\mathrm{Inst}}_{\mathcal{D}}$ has the form

$$\mathrm{S}{\uparrow}[t, s/x] : \overrightarrow{\mathrm{Inst}}_{\mathcal{D}_1} : \mathrm{S}{\downarrow}[u, s/x]$$

By assumption we obtain $x \notin \mathrm{dom}(\rho)$ as $x \in \mathrm{BVar}(\mathcal{D})$. The i.h. shows that the evaluation of $\overrightarrow{\mathrm{Inst}}_{\mathcal{D}_1}$ does not change the assignment. Evaluating $\mathrm{S}{\uparrow}[t, s/x]$ changes ρ by mapping x to some value, while evaluating $\mathrm{S}{\downarrow}[u, s/x]$ removes x from the domain of the assignment. Hence, the resulting overall assignment will be ρ again. □

The following theorem is a refinement of Theorem 42 in that the claimed existence of update sequences is replaced by an explicit computation based on instruction sequences extracted from the derivation.

Theorem 53 (S^1_2)**.** *Let $\mathcal{D}$ be a derivation of $t = u$ in Variable Normal Form. Let Ax_0 consist of those axioms in Ax that occur as injective renamings in $\mathcal{D}$, ρ an assignment with $\mathrm{dom}(\rho) \subseteq \mathrm{Var}(t, u)$, and F a κ-pre-model for Ax_0. Let σ_1, σ_2 be given by*

$$\begin{array}{rcl}\langle F, \sigma_1, \rho \rangle & = & \Phi(\overrightarrow{\mathrm{Inst}}_{\mathcal{D}}, \langle F, \langle\rangle, \rho \rangle)\\ \langle F, \sigma_2, \rho \rangle & = & \Phi(\overleftarrow{\mathrm{Inst}}_{\mathcal{D}}, \langle F, \langle\rangle, \rho \rangle)\end{array}$$

Then

$$\begin{array}{rcl}[\![t]\!]_{F,\rho} & \sqsubseteq & [\![u]\!]_{F * \sigma_1, \rho}\\ [\![u]\!]_{F,\rho} & \sqsubseteq & [\![t]\!]_{F * \sigma_2, \rho}\end{array}$$

Instead of proving the theorem directly, we prove the following stronger claim, similar to Claim 43.

Claim 54 (S^1_2)**.** *Let $\mathcal{D}$ be a derivation in Variable Normal Form. Let Ax_0 consist of those axioms in Ax that occur as injective renamings in $\mathcal{D}$. Fix some κ-pre-model F for Ax_0, and some integer U such that $\mathbf{l}_0(F) + \mathbf{l}(\mathcal{D}) \leq U$. Let $X = \mathrm{Var}(\mathcal{D})$.*

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

Let $\mathcal{D}_0 \vdash t = u$ be a sub-derivation of $\mathcal{D}$. Let ρ be an assignment, and σ a sequence of updates based on F, U and Ax_0 such that

$$\begin{array}{rcl} \mathrm{dom}(\rho) & \subseteq & X \setminus \mathrm{BVar}(\mathcal{D}_0) \\ \mathbf{l}_0(\rho), \mathbf{l}_1(\sigma), \mathbf{l}_0(\sigma), \mathbf{dom}\text{-}\mathbf{l}(\sigma) & \leq & U - \mathbf{l}(\mathcal{D}_0) \end{array}$$

Let σ_1, σ_2 be given by

$$\begin{array}{rcl} \langle F, \sigma_1, \rho \rangle & = & \Phi(\overrightarrow{\mathrm{Inst}}_{\mathcal{D}_0}, \langle F, \sigma, \rho \rangle) \\ \langle F, \sigma_2, \rho \rangle & = & \Phi(\overleftarrow{\mathrm{Inst}}_{\mathcal{D}_0}, \langle F, \sigma, \rho \rangle) \end{array}$$

Then

$$\begin{array}{rcl} [\![t]\!]_{F * \sigma, \rho} & \sqsubseteq & [\![u]\!]_{F * \sigma_1, \rho} \\ [\![u]\!]_{F * \sigma, \rho} & \sqsubseteq & [\![t]\!]_{F * \sigma_2, \rho} \ . \end{array}$$

Theorem 53 follows from Claim 54 by letting $\mathcal{D}_0 = \mathcal{D}$, ρ as given, $\sigma = \langle\rangle$, and $U = \max\{\mathbf{l}_0(F), \mathbf{l}_0(\rho)\} + \mathbf{l}(\mathcal{D})$,

Proof of Claim 54. We argue in S^1_2. Let $\mathcal{D}$, F, κ, and X be given as in the Claim. We prove that for any $\mathcal{D}_0$, ρ, σ, σ_1 and σ_2 satisfying the conditions of the Claim, the assertion of the Claim holds, by induction on $\mathbf{l}(\mathcal{D}_0)$. Thus this is proven by logarithmic induction (LIND) on a Π^b_1-property, which is available in S^1_2 by Theorem 1.

We consider cases according to the last rule applied in $\mathcal{D}_0$. The details for each case follow the same lines as in the proof of Claim 43, except that now σ_1 and σ_2 are not chosen but given by the Φ-function applied to sequences of instances that are extracted from derivations. Details are left to the reader. □

Corollary 55. *The consistency of* $\mathbf{PETS}(\mathrm{Ax})$ *is provable in* S^1_2.

Proof. We argue in S^1_2. Assume $\mathcal{D}$ is a $\mathbf{PETS}(\mathrm{Ax})$ derivation ending in $0 = 1$. Using Proposition 12 we can assume that $\mathcal{D}$ is in Variable Normal Form. Let ρ be the empty assignment, and F the empty pre-model for Ax. Let σ_1 be given by

$$\langle F, \sigma_1, \rho \rangle \quad = \quad \Phi(\overrightarrow{\mathrm{Inst}}_{\mathcal{D}}, \langle F, \langle\rangle, \rho \rangle)$$

By the previous Theorem 53, we obtain

$$0 \quad = \quad [\![0]\!]_{F, \rho} \quad \sqsubseteq \quad [\![1]\!]_{F * \sigma_1, \rho} \quad = \quad 1$$

which is impossible. □

Acknowledgment. For the purpose of open access the authors have applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press

References

[1] Arnold Beckmann, *Proving consistency of equational theories in bounded arithmetic*, Journal of Symbolic Logic **67** (2002mar), no. 1, 279–296.

[2] Samuel R. Buss, *Bounded arithmetic*, Bibliopolis, Naples, Italy, 1986. Revision of 1985 Princeton University Ph.D. thesis.

[3] ______, *Relating the bounded arithmetic and polynomial-time hierarchies*, Annals of Pure and Applied Logic **75** (1995), 67–77.

[4] Samuel R. Buss and Aleksandar Ignjatović, *Unprovability of consistency statements in fragments of bounded arithmetic*, Ann. Pure Appl. Logic **74** (1995), no. 3, 221–244.

[5] Stephen A. Cook, *Feasibly constructive proofs and the propositional calculus*, Proceedings of the seventh annual acm symposium on theory of computing, 1975, pp. 83–97.

[6] Jan Krajíček, *Bounded arithmetic, propositional calculus and complexity theory*, Cambridge University Press, Heidelberg, 1995.

[7] Jan Krajíček, Pavel Pudlák, and Gaisi Takeuti, *Bounded arithmetic and the polynomial hierarchy*, Annals of Pure and Applied Logic **52** (1991), 143–153.

[8] Pavel Pudlák, *A note on bounded arithmetic*, Fundamenta Mathematicae **136** (1990), no. 2, 85–89.

[9] Steven Vickers, *Topology via logic*, Cambridge University Press, 1996.

[10] Yoyuki Yamagata, *Consistency proof of a fraement of pv with substitution in bounded arithmetic*, The Journal of Symbolic Logic **83** (2018), no. 3, 1063–1090.

[11] Domenico Zambella, *Notes on polynomially bounded arithmetic*, Journal of Symbolic Logic **61** (1996), 942–966.

Department of Computer Science, Swansea University, Swansea SA2 8PP, UK
Email address: `a.beckmann@swansea.ac.uk`

Department of Electrical, Electronic and Computer Engineering, University of Fukui, Bunkyo 3-9-1, Fukui
Email address: `yoriyuki@u-fukui.ac.jp`

https://doi.org/10.1017/jsl.2025.6 Published online by Cambridge University Press