Notre Dame Journal of Formal Logic
Volume 35, Number 3, Summer 1994

Σ_1^1-Completeness of a Fragment of the Theory of Trees With Subtree Relation

P. CINTIOLI and S. TULIPANI

Abstract We consider the structure IT_S of all labeled trees, called also infinite terms, in the first order language $\mathcal{L}$ with function symbols in a recursive signature S of cardinality at least two and at least a symbol of arity two, with equality and a binary relation symbol $\sqsubseteq$ which is interpreted to be the subtree relation. The existential theory over $\mathcal{L}$ of this structure is decidable (see Tulipani [9]), but more complex fragments of the theory are undecidable. We prove that the $\exists\Delta$ theory of the structure is in Σ_1^1, where $\exists\Delta$ formulas are those in prenex form consisting of a string of unbounded existential quantifiers followed by a string of arbitrary quantifiers all bounded with respect to $\sqsubseteq$. Since the fragment of the theory was already known to be Σ_1^1-hard (see Marongiu and Tulipani [5]), it is now established to be Σ_1^1-complete.

1 Preliminaries and Introduction A *signature* S is a set of operation symbols on which is defined a function $ar : S \to \mathbb{N}$ into the set of natural numbers, called *arity*. Symbols of arity zero are called constant symbols. Throughout this paper we assume at least that S is a nonempty recursive set.

For every nonempty set A let A^* denote the free monoid of finite sequences of elements of A, including the empty sequence Λ. Let $\cdot$ be the operation of concatenation on A^*. A set $D \subseteq A^*$ is called *prefix-closed* if $p \cdot q \in D$ implies $p \in D$. A set D is called a *domain-tree* if:

(1) $D \subseteq \mathbb{N}_+^*$ and $\Lambda \in D$, where $\mathbb{N}_+$ is the set of positive integers;
(2) D is prefix closed.

When D is a domain-tree, the elements of D are called *positions*. Moreover, a mapping $t : D \to S$ is called a *tree* (or *infinite term*) over the signature S if

(3) $\forall p \in D$, if $t(p) = g$ and $ar(g) = k$ then $\forall j \in \mathbb{N}_+$, $\quad p \cdot j \in D \longleftrightarrow 1 \le j \le k$.

Received February 1, 1994; revised November 16, 1994

This also makes sense when g is a constant symbol. In such a case $k = 0$ and p is maximal in D, i.e., there is no $q \in D$ such that p is a proper prefix of q.

The *subtree* t' of a tree t at position p is a mapping $t' : D' \to S$ where $D' = \{q : p \cdot q \in D\}$ and $t'(q) = t(p \cdot q)$. A *rational tree* is a tree with a finite number of subtrees. A *finite tree* is a tree with a finite domain. We denote by IT_S the set of trees in the signature S. This can be made into an algebra of signature S, which we continue to denote IT_S, by defining, for every $f \in S$ of arity k and every $t_1, \ldots, t_k \in IT_S$, the tree t, denoted $f(t_1, \ldots, t_k)$, as the unique tree where $t(\Lambda) = f$ and, if $k > 0$, $t_1, \ldots, t_k$ are subtrees at positions $1, \ldots, k$, respectively. Moreover, a relation $\sqsubseteq$ is defined on IT_S by $t' \sqsubseteq t$ if and only if t' is a subtree of t at some position. The relation $\sqsubseteq$ is reflexive and transitive, i.e., a preorder, on IT_S. The set RT_S of rational trees is a substructure of IT_S and the set FT_S of finite trees is a substructure of RT_S; moreover, the preorder $\sqsubseteq$ is antisymmetric on FT_S, i.e., a partial order.

The first order theory $Th(IT_S)$, in the first order language for the signature S, is decidable, moreover $Th(IT_S) = Th(RT_S)$, (see Maher [2], and Marongiu and Tulipani [3]). This is no longer true when the preorder relation $\sqsubseteq$ is added and the signature has at least two symbols and a symbol is of arity at least two. In fact, under this hypothesis, every substructure of $(IT_S, \sqsubseteq)$ has an undecidable theory (see McCarthy [1], Marongiu and Tulipani [4], and Treinen [8]). However, it was proved in Tulipani [9] that the existential fragment of $Th(IT_S, \sqsubseteq)$ is decidable. This is the best result, since fragments more complex are undecidable, (see [4], [8], and [10]).

In [5] the fragment $\exists\Delta$ of existential quantification of Δ-formulas was investigated. Let $\mathcal{T}$ be the set of first order terms in signature S, then Δ-formulas are defined recursively as the smallest set of first order formulas satisfying:

- $t_1 = t_2$ and $t_1 \sqsubseteq t_2$ are Δ-formulas, for $t_1, t_2 \in \mathcal{T}$;
- if φ, ψ are Δ-formulas then $\varphi \wedge \psi$, $\varphi \vee \psi$, $\neg\varphi$, $\varphi \to \psi$ are Δ-formulas;
- if φ is a Δ-formula then $(\exists x \sqsubseteq t)\varphi$ and $(\forall x \sqsubseteq t)\varphi$ are Δ-formulas, for every $t \in \mathcal{T}$ and every variable x not in t.

It was observed in [4], and it is not difficult to prove, that the fragment $Th_{\exists\Delta}(RT_S, \sqsubseteq)$ of $\exists\Delta$-formulas, which are true in the structure of rational trees with subterm relation, is recursively enumerable. Moreover, this fragment is no longer equal to the fragment $Th_{\exists\Delta}(IT_S, \sqsubseteq)$ as in the case when $\sqsubseteq$ is not present (see [3]). In fact, in [5] it was proved that, when S has at least a constant symbol and a symbol of arity at least two, the fragment $Th_{\exists\Delta}(IT_S, \sqsubseteq)$ is Σ_1^1-hard. One may easily note that the result continues to hold also in the more general case when the signature S contains two function symbols and one of them is of arity at least two. Here, we prove membership in Σ_1^1 of the fragment $Th_{\exists\Delta}(IT_S, \sqsubseteq)$ under the hypothesis that S is recursive. So, we may conclude that the fragment is Σ_1^1-complete when S is recursive, is of cardinality at least two, and has a symbol of arity at least two (for terminology see Odifreddi [6]).

2 Main Result We assume that the signature S is a recursive set. Our goal is to prove that the fragment $Th_{\exists\Delta}(IT_S, \sqsubseteq)$ is in Σ_1^1. We observe that we may also assume that S has cardinality greater than one and has a symbol of arity at least two. Otherwise the statement follows since the first order theory of IT_S is decidable. In fact, when S has cardinality one, the first order theory of IT_S is clearly decidable since IT_S is trivial with only one element, whereas, when no element in S has arity greater than one, the first order theory of IT_S is decidable by the celebrated Rabin's Theorem on two successors (see Rabin [7]).

Now, we are going to transform effectively every first order $\exists\Delta$-sentence φ, in the signature S possibly with the predicate symbol $\sqsubseteq$, into a second order Σ^1_1-sentence Ψ in the language of arithmetic such that

(4) $(IT_S, \sqsubseteq) \models \varphi$ if and only if $AR2 \models \Psi$

where $AR2$ is the second order arithmetic.

Theorem 2.1 *Without loss of generality we can restrict ourselves to the case of sentences of the following kind:*

(5) $\exists x_0\, Q_1 x_1\, Q_2 x_2 \ldots Q_n x_n\, \alpha$ *where;*
(6) $Q_i \in \{\forall, \exists\}$ *and the quantifiers* $Q_i x_i$*, for* $i = 1, \ldots, n$*, are all bounded with respect to* $\sqsubseteq$ *to variables* x_j *with* $0 \leq j < i$*;*
(7) α *is quantifier-free with only atomic subformulas of the following two types* $x = t$, $x \sqsubseteq y$

where x, y *are variables and* t *is a first order term.*

Proof: Start with any φ in $\exists\Delta$. It is straightforward to transform φ, by adding existential quantifiers, if necessary, into a logically equivalent prenex formula $Q'_1 x_1 \ldots Q'_n x_n \beta$ where the atomic subformulas are as in (7) and, for $i = 1, \ldots, n$, the quantifier $Q'_i x_i$ can be $\exists x_i$, $\exists x_i \sqsubseteq x_j$ or $\forall x_i \sqsubseteq x_j$, for some $j < i$.

Now, if the sentence does not have the desired form, then take a term t which contains all the variables x_i in $\{x_1, \ldots, x_n\}$ which are quantified by nonbounded quantifier $\exists x_i$. Such a t exists since the signature has a symbol of arity at least two. Hence, the following sentence

(8) $\exists x_0 Q_1 x_1 \ldots Q_n x_n (x_0 = t \wedge \beta)$

satisfies properties (6)–(7), where $Q_i x_i$ is $Q'_i x_i$ if $Q_i x_i$ was bounded and $Q_i x_i$ is $\exists x_i \sqsubseteq x_0$ otherwise.

Our aim is to code trees of IT_S as functions $F : \mathbb{N} \to \mathbb{N}$. This can be achieved easily, in a standard way, by fixing an encoding $\langle\, \rangle_k : \mathbb{N}^k \to \mathbb{N}$ of k-tuples of natural numbers. Then, a finite sequence $s_1 \cdot s_2 \cdots s_l \in \mathbb{N}^*_+$ will be coded by the integer $a = \langle l, \langle s_1, \ldots, s_l\rangle\rangle$. As usual, we denote the number l by $\mathrm{length}(a)$. It is assumed that $\langle n\rangle = n$, for every n, that $\langle 0\rangle = 0$ codes the empty sequence and $\mathrm{length}(0) = 0$. So, every domain-tree D can be thought of, by coding, as a subset of $\mathbb{N}$. Moreover, we can assume, for convenience, that $S \subseteq \mathbb{N}_+$. Hence every tree $t : D \to S$ can be determined by a function $F : \mathbb{N} \to \mathbb{N}$ where the following hold

(9) $\mathrm{Im}\, F \subseteq S \cup \{0\}$;
(10) $\{a : F(a) \in S\}$ encodes a domain-tree;
(11) the analogous property of (3) obtained by encoding.

Now, we write a second order formula $\mathrm{Tree}(F)$ for defining in $AR2$ functions which code elements of IT_S. We need the following primitive recursive relations on $\mathbb{N}$:

- $\mathrm{conc}(a, b, c)$, the concatenation, which holds iff $a = \langle l, \langle a_1, \ldots, a_l \rangle\rangle$, $b = \langle k, \langle b_1, \ldots, b_k \rangle\rangle$, $c = \langle l + k, \langle a_1, \ldots, a_l, b_1, \ldots, b_k \rangle\rangle$;
- $\mathrm{prec}(a, b)$, which holds iff $\exists z\, \mathrm{conc}(a, z, b)$;
- $\mathrm{notseq}(a)$, which holds iff a does not code any sequence in $\mathbb{N}^*_+$.

Then, we consider the language $\mathcal{L}^*_2$ of second order arithmetic with symbols for all primitive functions. Then, the formula $\mathrm{Tree}(F)$ will be the universal quantification, over all first order variables x, y, p, j, g, k, of the conjunction of the following formulas, which clearly can be written in $\mathcal{L}^*_2$. Remember that S and the arity function $ar : S \to \mathbb{N}$ are both recursive.

(12) $F(x) \in S \cup \{0\}$

(13) $F(0) \neq 0 \wedge \forall x(\mathrm{notseq}(x) \to F(x) = 0)$

(14) $F(x) \neq 0 \wedge \mathrm{prec}(y, x) \longrightarrow F(y) \neq 0$

(15) $ar(g) = k \wedge F(p) = g \longrightarrow (\exists q(\mathrm{conc}(p, \langle 1, j \rangle, q) \wedge F(q) \neq 0) \leftrightarrow (1 \leq j \wedge j \leq k))$.

Note that (12) takes care of (9); (13) and (14) take care of (10) (see (1) and (2) and remember that 0 codes the empty sequence). Moreover, (15) takes care of (11), (see (3)). Note also that $ar(c) = 0$, for every constant symbol c, hence (15) means that $F(p) = c$ implies $F(q) = 0$ for every immediate successor q of p and by (14), $F(q) = 0$ for every q such that $\mathrm{prec}(p, q)$ and $q \neq p$.

We wish to transform every sentence as in (5) into a sentence of $AR2$ where there exists a unique second order quantifier $\exists F$ and F is constrained to satisfy $\mathrm{Tree}(F)$. Moreover, we manage in such a way that all the quantifiers $Q_1 x_1, \ldots, Q_n x_n$ range over natural numbers which are constrained to represent positions in the tree F and, on the other hand, positions determine subtrees of F; x_0 represents the root of F. So, we need to define formulas $Ug(F, x, t)$ of $\mathcal{L}^*_2$ for every first order variable x and every term t in signature S. Such formulas will have $\{F, x\} \cup \mathrm{var}(t)$ as free variables and F is the only second order variable. The definition is by structural induction on t as follows:

$$\text{(16)} \quad \begin{aligned} Ug(F, x, c) &= F(x) = c \\ Ug(F, x, y) &= \forall u \forall v \forall z(\mathrm{conc}(x, u, v) \wedge \mathrm{conc}(y, u, z) \\ &\quad \to F(v) = F(z)) \\ Ug(F, x, g(t_1, \ldots, t_k)) &= F(x) = g \wedge \exists z_1 \ldots \exists z_k \bigwedge_{1 \leq i \leq k} \\ &\quad (\mathrm{conc}(x, \langle 1, i \rangle, z_i) \wedge Ug(F, z_i, t_i)), \end{aligned}$$

where c is a constant symbol, y is a first order variable, g is an operation symbol of arity $k > 0$ and $z_1, \ldots, z_k$ are new first order variables not used elsewhere. Now, we are ready to prove our theorem.

Theorem 2.2 *Let S be a nonempty recursive signature. Then, the fragment $Th_{\exists\Delta}(IT_S, \sqsubseteq)$ is Σ^1_1-complete if and only if S is of cardinality greater than one and with a symbol of arity at least two.*

Proof: Going from left to right, suppose S is of cardinality one or all the symbols in S have arity less than two. Then the first order theory of IT_S is decidable, as we discussed before. Hence, under our hypothesis our fragment is decidable and it cannot be Σ^1_1-complete. Therefore, the left to right condition of the the theorem is necessary.

For the other direction, given the result of [5], we have to prove membership in Σ^1_1 of our fragment. Start with a formula as in (5). First, transform the matrix α into a formula Θ_α of $AR2$ by replacing all its atomic subformulas according to the following rules:

$$(17)\quad \begin{array}{lll} x = t & \text{replaced by} & Ug(F, x, t) \\ x \sqsubseteq y & \text{replaced by} & \exists z(\,\mathrm{prec}(y, z) \wedge Ug(F, z, x)\,), \end{array}$$

where z is a new first order variable not used elsewhere. Then, according to (7) and to (16), Θ_α is a formula for $AR2$ where the free first order variables are the same as in α. Now, let Φ be the formula

$$(18)\quad \mathrm{Tree}(F) \wedge \Theta_\alpha\,[x_0 \leftarrow 0]$$

where $[x_0 \leftarrow 0]$ is the substitution of x_0 for 0 which is the code for the empty position. By (12)–(15) there exists a bijection $\delta : IT_S \longrightarrow \{F : \mathbb{N} \to \mathbb{N},\ AR2 \models \mathrm{Tree}(F)\}$. Fix $T \in IT_S$ and positions $p_1, \ldots, p_n$ in the domain of T. Denote by T/p the subtree of T at position p and by F_T the corresponding of T under δ. Then, we can prove

Claim 2.3

$$IT_S \models \alpha[x_0 \leftarrow T, x_1 \leftarrow T/p_1, \ldots, x_n \leftarrow T/p_n]$$

if and only if

$$AR2 \models \Phi[F \leftarrow F_T, x_1 \leftarrow \overline{p_1}, \ldots, x_n \leftarrow \overline{p_n}].$$

where $\overline{p}$ denotes the code of the position p.

Proof: The proof of Claim 2.3 follows simply from (17) and from the meaning of the formula $Ug(F, x, t)$ determined by (16).

Note that, if p, q are positions in T, then $AR2 \models Ug(F, x, y)[F_T, \overline{p}, \overline{q}]$, (where x and y are distinct), means that $T/p = T/q$. Moreover, note that, if p is prefix of q, then $T/q \sqsubseteq T/p$.

Now, observe that a is the code of some position p in T if and only if $F_T(a) \neq 0$. The truth of the sentence (5) is determined by the existence of a tree $T \in IT_S$ assigned to x_0 and by trees $T_1, \ldots, T_n$ which interpret $x_1, \ldots, x_n$. Since every quantifier $Q_i x_i$, for $i = 1, \ldots, n$, is bounded to x_j for some $0 \leq j < i$, the trees $T_1, \ldots, T_n$ must be all subtrees of T. Then, to get our result, we have to put before the formula Φ the same list of quantifiers $Q_1 x_1 \ \ldots\ Q_n x_n$ which are before α in (5) and the variables have to be constrained to range on codes of positions in T. So, the Σ^1_1-sentence Ψ which works in (4) is

$$(19)\quad \exists F\ \tilde{Q}_1 x_1 \ \ldots\ \tilde{Q}_n x_n\ \Phi$$

where $\tilde{Q}_i x_i$, for $i = 1, \ldots, n$ denote the quantifiers $Q_i x_i$ of sentence in (5) relativized to the predicate $F(x_i) \neq 0$. In fact, the following claim holds for all $k = 0, \ldots, n$, for every T in IT_S and all positions $p_1, \ldots, p_k$ in T:

Claim 2.4

$$IT_S \models Q_{k+1} x_{k+1} \ldots Q_n x_n\, \alpha\,[x_0 \leftarrow T, x_1 \leftarrow T/p_1, \ldots, x_k \leftarrow T/p_k]$$

if and only if

$$AR2 \models \tilde{Q}_{k+1}x_{k+1} \ldots \tilde{Q}_n x_n \, \Phi\,[F \leftarrow F_T, x_1 \leftarrow \overline{p_1}, \ldots, x_k \leftarrow \overline{p_k}].$$

Proof: The proof is an easy induction on the number of quantifiers; the case of no quantifiers corresponds to $k = n$ and it is Claim 2.3.

Acknowledgments We thank anonymous referees for comments which improved the paper.

REFERENCES

[1] McCarthy, J., *Representation of Recursive Programs in First Order Logic*, Computer Science Department, Stanford University, Stanford, 1977. 1

[2] Maher, J., "Complete axiomatizations of the algebras of finite, infinite and rational trees," pp. 348–357 in *Proceedings of the Third Annual Symposium on Logic in Computer Science*, IEE Computer Society, New York, 1988. 1

[3] Marongiu, G., and S. Tulipani, "Quantifier elimination for infinite terms," *Archives of Mathematical Logic*, vol. 31 (1991), pp. 1–17. Zbl 0723.03016 MR 92i:03029 1, 1

[4] Marongiu, G., and S. Tulipani, *Decidability results for term algebras*, AILA Preprint, 1991. 1, 1, 1

[5] Marongiu, G., and S. Tulipani, "Undecidable fragments of term algebras with subterm relation," *Fundamenta Informaticae*, vol. 19 (1993), pp. 371–382. Zbl 0788.03011 MR 95f:68145 1, 1, 1, 2

[6] Odifreddi, P., *Classical Recursion Theory*, North-Holland, Amsterdam, 1989. Zbl 0661.03029 MR 90d:03072 1

[7] Rabin, M., "Decidable Theories," pp. 596–629 in *Handbook of Mathematical Logic*, edited by J. Barwise, North-Holland, Amsterdam, 1977. 2

[8] Treinen R., "A New Method for Undecidability Proofs of First Order Theories," *Journal of Symbolic Computation*, vol. 14 (1992), pp. 437–457. Zbl 0769.03026 MR 93m:03077 1, 1

[9] Tulipani, S., "Decidability of the Existential Theory of Infinite Terms with Subterm Relation," *Information and Computation*, vol. 108 (1994), pp. 1–33. Zbl 0801.68115 MR 94k:03012 1, 1

[10] Venkataraman, K., "Decidability of the purely existential fragment of the theory of term algebras," *Journal of the Association for Computing Machinery*, vol. 34 (1987), pp. 410–471. MR 88g:68042 1

Department of Mathematics
University of Siena, via del Capitano
53100 Siena
Italy

Department of Mathematics and Physics
University of Camerino, via Madonna delle Carceri

62032 Camerino
Italy
email: tulipani@camvax.cineca.it