REPORTS ON MATHEMATICAL LOGIC
49 (2014), 23–34
doi:10.4467/20842589RM.14.002.2272

Rafael GRIMSON and Bart KUIJPERS

# SOME FRAGMENTS OF SECOND-ORDER LOGIC OVER THE REALS FOR WHICH SATISFIABILITY AND EQUIVALENCE ARE (UN)DECIDABLE

A b s t r a c t. We consider the $\Sigma^1_0$-fragment of second-order logic over the vocabulary $\langle +, \times, 0, 1, <, S_1, ..., S_k\rangle$, interpreted over the reals, where the predicate symbols $S_i$ are interpreted as semi-algebraic sets. We show that, in this context, satisfiability of formulas is decidable for the first-order $\exists^*$-quantifier fragment and undecidable for the $\exists^*\forall$- and $\forall^*$-fragments. We also show that for these three fragments the same (un)decidability results hold for containment and equivalence of formulas.

## 1. Introduction and summary

First-order logic over the vocabulary $\langle +, \times, 0, 1, <\rangle$, interpreted in the structure $\mathcal{R} = (\mathbf{R}, +, \times, 0, 1, <)$, the ordered field of the real numbers $\mathbf{R}$, has

*Received 15 November 2013*

received considerable interest in several areas of theoretical computer science. One particular such area is that of *constraint databases*, where this logic is used as a basis for first-order query languages ([7], in particular Chapter 2). Hereto, the vocabulary $\langle +, \times, 0, 1, < \rangle$ is extended to some vocabulary $\langle +, \times, 0, 1, <, S_1, ..., S_k \rangle$, where, for $i \in \{1, ..., k\}$, $S_i$ is a predicate of arity $ar(S_i)$, which is a natural number strictly larger than 0. The predicates $S_1, ..., S_k$ represent the *input relations* to a query. In the constraint database formalism, the predicates $S_i$ are interpreted by first-order definable relations over $\langle +, \times, 0, 1, < \rangle$, that is, by semi-algebraic subsets of $\mathbf{R}^{ar(S_i)}$, $i \in \{1, ..., k\}$ (see [1]). Since first-order logic over the reals admits quantifier elimination [8], in constraint databases, it is assumed that the input relations $S_1, ..., S_k$ are given by quantifier-free formulas.

First-order logic over $\langle +, \times, 0, 1, <, S_1, ..., S_k \rangle$ then allows the definition of new relations by means of formulas with free variables over $\langle +, \times, 0, 1, <, S_1, ..., S_k \rangle$, as well as the expression of properties of the query input relations $S_i$ by means of sentences over $\langle +, \times, 0, 1, <, S_1, ..., S_k \rangle$. These newly created relations or Boolean values represent the *output* to a query.

For example, the first-order query formula

$$\varphi(x, y) = \exists \varepsilon (0 < \varepsilon \wedge \forall x' \forall y' ((x - x')^2 + (y - y')^2 < \varepsilon^2 \rightarrow S(x', y'))),$$

defines the topological interior of the binary input relation $S$, viewed as a subset of $\mathbf{R}^2$. Likewise, the sentence

$$\begin{aligned} \psi = {} & \forall x \forall y \exists \varepsilon (0 < \varepsilon \wedge S(x, y) \\ & \rightarrow (\forall x' \forall y' ((x - x')^2 + (y - y')^2 < \varepsilon^2 \wedge S(x', y') \rightarrow x = x' \wedge y = y'))) \end{aligned}$$

expresses the Boolean (topological) property that all elements of the binary input relation $S$ are isolated points of $S$. When $S$ is restricted to be interpreted by semi-algebraic subsets of $\mathbf{R}^2$, this statement is equivalent to expressing that $S$ has *finite* cardinality.

From Tarski's theorem [8], which says that first-order logic over the reals is decidable (via quantifier elimination), we obtain, by plugging-in quantifier-free first-order descriptions of the input relations in the query formula and then eliminating the quantifiers from the obtained formula, a quantifier-free first-order description of the output. This amounts to an effective query evaluation strategy for constraint database queries of this type.

We can also view the above example formulas as *second-order formulas (without second-order quantifiers)* over the vocabulary $\langle +, \times, 0, 1, < \rangle$, if we view $S$ as a binary relation variable. In a second-order context, we would write $\varphi(x, y, S)$ and $\psi(S)$ to indicate both the free first- and second-order variables of these formulas. To be precise, we consider formulas in $\Sigma^1_0$. A second-order formula belongs to this syntactic fragment if its quantifiers range only over first-order variables, although it may have free second-order variables.

If we stick to this second-order view of query formulas, the following definition specifies what we mean by a $\Sigma^1_0$ second-order formula over the reals being *satisfiable*. This definition uses the Henkin-semantics (that interprets relation symbols by semi-algebraic subsets of $\mathbf{R}^\ell$ rather than by arbitrary subsets of $\mathbf{R}^\ell$, where $\ell$ is the appropriate arity), that is also used in constraint databases [7].

**Definition 1.1.** We say that a formula $\varphi(x_1, ..., x_n, S_1, ..., S_k)$ in the the $\Sigma^1_0$-fragment of second-order logic over $\langle +, \times, 0, 1, < \rangle$ with free first-order variables $x_1, ..., x_n$ and free relation variables $S_1, ..., S_k$ of arities $ar(S_1), ..., ar(S_k)$, respectively, is *satisfiable* if there exist real numbers $a_1, ..., a_n$ and semi-algebraic subsets $A_1, ..., A_k$ of $\mathbf{R}^{ar(S_1)}, ..., \mathbf{R}^{ar(S_k)}$, respectively, such that

$$\mathcal{R} \models \varphi[a_1, ..., a_n, A_1, ..., A_n]$$

holds. □

We have a similar definition of the *containement* and the *equivalence* of two second-order formulas.

**Definition 1.2.** Let $\varphi(x_1, ..., x_n, S_1, ..., S_k)$ and $\psi(x_1, ..., x_n, S_1, ..., S_k)$ be two relational second-order formulas over $\langle +, \times, 0, 1, < \rangle$, which have the same free first-order variables $x_1, ..., x_n$ and the same free relation variables $S_1, ..., S_k$. We say that (the interpretation of) $\varphi(x_1, ..., x_n, S_1, ..., S_k)$ is *contained* in (the interpretation of) $\psi(x_1, ..., x_n, S_1, ..., S_k)$ denoted $\varphi(x_1, ..., x_n, S_1, ..., S_k) \subseteq \psi(x_1, ..., x_n, S_1, ..., S_k)$, if for all real numbers $a_1, ..., a_n$ and all semi-algebraic subsets $A_1, ..., A_k$ of $\mathbf{R}^{ar(S_1)}, ..., \mathbf{R}^{ar(S_k)}$ respectively, we have that

$$\mathcal{R} \models (\varphi \rightarrow \psi)[a_1, ..., a_n, A_1, ..., A_n]$$

holds.

We say that $\varphi(x_1, ..., x_n, S_1, ..., S_k)$ and $\psi(x_1, ..., x_n, S_1, ..., S_k)$ are *equivalent*, denoted

$$\varphi(x_1, ..., x_n, S_1, ..., S_k) \equiv \psi(x_1, ..., x_n, S_1, ..., S_k),$$

if both $\varphi(x_1, ..., x_n, S_1, ..., S_k) \subseteq \psi(x_1, ..., x_n, S_1, ..., S_k)$ and $\psi(x_1, ..., x_n, S_1, ..., S_k) \subseteq \varphi(x_1, ..., x_n, S_1, ..., S_k)$ hold. □

Clearly, the decidability of containment implies the decidability of equivalence. Since *finiteness* of the relations $S_1, ..., S_k$ is expressible in second-order logic over $\langle +, \times, 0, 1, < \rangle$, as illustrated by the above example, Proposition 2.6.4 in [7] says that, in general, satisfiability, containment and equivalence are undecidable properties of second-order formulas over $\langle +, \times, 0, 1, < \rangle$.

In this paper, we are interested in first-order quantifier-prefix fragments of second-order logic over $\langle +, \times, 0, 1, < \rangle$, for which satisfiability, containment and equivalence are (un)decidable. By a first-order quantifier-prefix fragment of second-order logic over $\langle +, \times, 0, 1, < \rangle$, we mean a subclass of formulas that can be written in prenex-form

$$\mathcal{Q}(y_1, ..., y_m)\varphi(y_1, ..., y_m, x_1, ..., x_n, S_1, ..., S_k),$$

where $\mathcal{Q}(y_1, ..., y_m)$ is a sequence of first-order quantifiers—belonging to some syntactic family—over $y_1, ..., y_m$ and $\varphi(y_1, ..., y_m, x_1, ..., x_n, S_1, ..., S_k)$ is a quantifier-free second-order formula over $\langle +, \times, 0, 1, < \rangle$, with free first-order variables $y_1, ..., y_m, x_1, ..., x_n$ and the free relation variables $S_1, ..., S_k$. Again, in such classes of formulas, relation variables are not quantified.

The three last lines of the following table summarize the (un)decidability results of this paper. For completeness, we have added, in the first line of the table, the known results concerning conjunctive formulas in the $\exists^*$-fragment (that is, conjunctions of possibly negated atomic formulas—see Chapter 2 of [7]).

We remark that a substantial differences with the classical decision problem [3], is that we consider a logic in which certain functions, relations and constants $(+, \times, <, 0, 1)$ have a fixed interpretation in the reals and for which the remaining predicate symbols are also restricted to range

| quantifier-prefix | satisfiability | containment | equivalence |
|---|---|---|---|
| $\exists^*(conjunctive)$ | decidable[7] | decidable [7] | decidable [7] |
| $\exists^*$ | decidable (FMP[1]) | undecidable | undecidable |
| $\exists^*\forall$ | undecidable | undecidable | undecidable |
| $\forall^*$ | undecidable | undecidable | undecidable |

over semi-algebraic sets. It is not clear if other results concerning the classical decision problem can be carried over to our setting of the reals.

This paper is organized as follows. We give, in Section 2, an elementary proof that satisfiability is decidable for the $\exists^*$-fragment of the $\Sigma^1_0$-fragment of second-order logic over $\langle +, \times, 0, 1, < \rangle$. We show, in particular, that in case of satisfiability, the predicates $S_1, ..., S_k$ can be interpreted by finite sets. In Section 3, we show that satisfiability is undecidable for the $\exists^*\forall$-fragment of second-order logic over $\langle +, \times, 0, 1, < \rangle$ and in Section 4, we show the same for the $\forall^*$-sentences. In Section 5, we show the results in the above table for containment and equivalence.

## 2. For the $\exists^*$-fragment satisfiability is decidable

It is known that satisfiability, containment and equivalence of conjunctive formulas, i.e., conjunctions of possibly negated atomic formulas preceded by a first-order $\exists^*$-prefix, in the $\Sigma^1_0$-fragment of second-order logic over the vocabulary $\langle +, \times, 0, 1, <, S_1, ..., S_k \rangle$ are decidable [7, Chapter 2]. We first show that this is no longer the case if also disjunctions are allowed.

In this section, we prove the following result. Although this result is already known [7, Chapter 2], we provide an elementary proof and also give a complexity result.

**Theorem 2.1.** *For the (first-order) $\exists^*$-quantifier fragment of the $\Sigma^1_0$-fragment of second-order logic over the vocabulary $\langle +, \times, 0, 1, <, S_1, ..., S_k \rangle$, satisfiability is decidable. Furthermore, in case of satisfiability, the relations $S_1, ..., S_k$ may be interpreted by finite sets. Our decision procedure requires exponential time in the length of the formula.*

[1]Finite model property: in case of satisfiability, the relations may be interpreted by finite sets.

**Proof.** We remark that it is sufficient to prove the theorem for any quantifier-free formula of the $\Sigma^1_0$-fragment of second-order logic over the vocabulary $\langle +, \times, 0, 1, <, S_1, ..., S_k\rangle$. Let $\varphi(x_1, ..., x_n)$ be such a quantifier-free formula over $\langle +, \times, 0, 1, <, S_1, ..., S_k\rangle$. We can write $\varphi(x_1, ..., x_n)$ in disjunctive normal form as

$$\bigvee_{i=1}^{d} \varphi_i(x_1, ..., x_n), \tag{1}$$

for some $d \in \mathbf{N} \setminus \{0\}$ ($\mathbf{N}$ denotes the set of natural numbers), where for $i \in \{1, ..., d\}$, $\varphi_i(x_1, ..., x_n)$ has the form

$$\bigwedge_{j=1}^{k} (\bigwedge_{r=1}^{p_{ij}} S_j(\mathbf{u}_{ijr}) \wedge \bigwedge_{s=1}^{q_{ij}} \neg S_j(\mathbf{v}_{ijs})) \wedge \bigwedge_{j=1}^{p_i} P_{ij}(x_1, ..., x_n)\ \theta_{ij}\ 0, \tag{2}$$

where $p_{ij}, q_{ij}, p_i \in \mathbf{N}$ and $P_{ij}(x_1, ..., x_n)$ are polynomials in the variables $x_1, ..., x_n$ with integer coefficients, with $\theta_{ij} \in \{<, >, \leq, \geq\}$ and where $\mathbf{u}_{ijr}$ and $\mathbf{v}_{ijs}$ are $ar(S_j)$-tuples of terms over the vocabulary $\langle +, \times, 0, 1, <\rangle$ with variables from $x_1, ..., x_n$.

Clearly, $\varphi(x_1, ..., x_n)$ is satisfiable if and only if $\varphi_i(x_1, ..., x_n)$ is satisfiable for some $i \in \{1, ..., d\}$.

*Notation*: In the following we denote by $\mathbf{u}_{ijr}[a_1, ..., a_n]$ the $ar(S_j)$-tuple that has as $\ell$-th component, for $\ell \in \{1, ..., ar(S_j)\}$, the $\ell$-th term of $\mathbf{u}_{ijr}$ with the variables $x_1, ..., x_n$ instantiated to $a_1, ..., a_n$.

To the formulas $\varphi_i(x_1, ..., x_n)$, for $i \in \{1, ..., d\}$, written in this normal form, we associate a formula

$$\psi_i(x_1, ..., x_n) := \bigwedge_{j=1}^{k} (\bigwedge_{r=1}^{p_{ij}} \bigwedge_{s=1}^{q_{ij}} \mathbf{u}_{ijr} \neq \mathbf{v}_{ijs}) \wedge \bigwedge_{j=1}^{p_i} P_{ij}(x_1, ..., x_n)\ \theta_{ij}\ 0,$$

where for vectors of terms $\mathbf{t} = (t_1, .., t_N), \mathbf{s} = (s_1, ..., s_N)$, $\mathbf{t} \neq \mathbf{s}$ abbreviates the formula $\vee_{i=1}^{N} \neg(t_i = s_i)$.

*Claim*: The formula $\varphi_i(x_1, ..., x_n)$ is satisfiable if and only if the first-order quantifier-free formula $\psi_i(x_1, ..., x_n)$ is satisfiable.

Proof of the claim: For the only-if direction, assume that $\varphi_i(x_1, ..., x_n)$ is satisfiable. This means that there exists real numbers $a_1, ..., a_n$ and $\langle +,$

$\times, 0, 1, <\rangle$-definable relations $A_1, ..., A_k$ such that $\mathcal{R} \models \varphi_i[a_1, ..., a_n, A_1, ..., A_k]$. Therefore, for all $j \in \{1, ..., k\}$, $r \in \{1, ..., p_{ij}\}$ and $s \in \{1, ..., q_{ij}\}$ we have $\mathbf{u}_{ijr}[a_1, ..., a_n] \in A_j$ and $\mathbf{v}_{ijs}[a_1, ..., a_n] \notin A_j$. Then it follows that for any $i$, $r$ and $s$, $\mathbf{u}_{ijr}[a_1, ..., a_n] = \mathbf{v}_{ijs}[a_1, ..., a_n]$ is impossible. Since $P_{ij}(a_1, ..., a_n)\ \theta_{ij}\ 0$, for all $j \in \{1, ..., p_i\}$ this implication is proven.

For the if-direction, the satisfiability of $\psi_i(x_1, ..., x_n)$ implies that there exists real numbers $a_1, ..., a_n$ such that $\mathcal{R} \models \psi_i[a_1, ..., a_n]$. Since any non-empty semi-algebraic set contains points with real algebraic coordinates, we may assume that $a_1, ..., a_n$ are real algebraic numbers. Let

$$A_j := \{\mathbf{u}_{ijr}[a_1, ..., a_n] \mid r \in \{1, ..., p_{ij}\}\},$$

for $j \in \{1, ..., k\}$. We remark that, being finite sets of points with real algebraic coordinates, these $A_j$ are first-order definable over $\langle +, \times, 0, 1, <\rangle$. Then, $\mathcal{R} \models \varphi_i[a_1, ..., a_n, A_1, ..., A_k]$ because for all $j \in \{1, ..., k\}$ and $r \in \{1, ..., p_{ij}\}$, $\mathbf{u}_{ijr}[a_1, ..., a_n] \in A_j$ per definition of $A_j$ and $\mathbf{v}_{ijs}[a_1, ..., a_n] \notin A_j$, for all $s \in \{1, ..., q_{ij}\}$ , because $\mathbf{v}_{ijs}[a_1, ..., a_n]$ differs from all $\mathbf{u}_{ijr}[a_1, ..., a_n]$ for $r \in \{1, ..., p_{ij}\}$. Also $P_{ij}(a_1, ..., a_n)\ \theta_{ij}\ 0$, for all $j \in \{1, ..., p_i\}$ remains true. This concludes the proof of the claim.

From the claim it is immediately clear that satisfiability of formulas in the $\exists^*$-fragment of the $\Sigma^1_0$-fragment of second-order logic $\langle +, \times, 0, 1, <, S_1, ..., S_k\rangle$ is decidable. Indeed, given a quantifier-free formula

$$\varphi(x_1, ..., x_n) = \vee_{i=1}^{d} \varphi_i(x_1, ..., x_n)$$

in this fragment, the formula

$$\psi(x_1, ..., x_n) = \vee_{i=1}^{d} \psi_i(x_1, ..., x_n)$$

is created and satisfiability of this formula adds up to deciding the truth of the $\langle +, \times, 0, 1, <\rangle$-sentence $\exists x_1 \cdots \exists x_n \psi(x_1, ..., x_n)$, which is possible because of the decidability of first-order logic over the reals (for example, via quantifier elimination), first proven by Tarski [8].

From the proof of the claim above, it is also clear that, in case of satisfiability, the semi-algebraic sets $A_j := \{\mathbf{u}_{ijr}[a_1, ..., a_n] \mid r \in \{1, ..., p_{ij}\}\}$ are finite.

For what concerns the complexity of the decision procedure, we remark that it might take exponential time and space to put the original formula in the disjunctive normal form given by Equations (1) and (2) [2]. Afterwards, the procedure described by Grigoriev and Vorobjov [4] to decide emptiness of semi-algebraic sets described by first-order formulas can be applied to the formulas $\psi_i$, for $i \in \{1,...,d\}$. This last step is simply exponential in the number of variables of the formula. Since the number of variables of the original formula is not increased in the transformation to the normal form, we obtain that the whole decision procedure for an arbitrary input formula of length $L$ can be performed within $L^{L^{\mathcal{O}(1)}}$-time, that is, within exponential time. This finishes the proof. □

## 3. For the $\exists^*\forall$-fragment satisfiability is undecidable

In this section, we prove the following result.

**Theorem 3.1.** *For the $\exists^*\forall$-fragment of the $\Sigma^1_0$-fragment of second-order logic over $\langle +, \times, 0, 1, <, S_1, ..., S_k\rangle$ satisfiability is undecidable.*

First we give a lemma.

**Lemma 3.2.** *Let $S$ be unary predicate symbol. Then the $\langle +, \times, 0, 1, <, S\rangle$-formula*

$$I(S) := \forall x(S(0) \wedge (x < 1 \wedge S(x) \to x = 0) \wedge (x \geq 1 \wedge S(x) \to S(x-1)))$$

*expresses that $S$ is an initial segment of $\mathbf{N}$.*

**Proof.** If $A = \{0, 1, ..., n\}$ for some $n \in \mathbf{N}$, then $0 \in A$ and $x < 1$ and $x \in A$ imply $x = 0$ and $x \geq 1$ and $x \in A$ imply $x - 1 \in A$.

On the other hand if $\mathcal{R} \models I[A]$, then $0 \in A$ and no other $x < 1$ is in $A$. Suppose $x \geq 1$ belongs to $A$. We can write $x = \lfloor x \rfloor + r$, with $\lfloor x \rfloor \in \mathbf{N} \setminus \{0\}$ and $0 \leq r < 1$. If we assume $0 < r$, then also $x - 1, x - 2, ..., x - \lfloor x \rfloor = r$ belong to $A$, which is impossible. If $r = 0$, then $x \in \mathbf{N} \setminus \{0\}$ and this implies that also $1, 2, ..., x \in A$. Therefore $A = \{0\}$, $A$ is an initial segment of $\mathbf{N}$ or $A = \mathbf{N}$. Since a discrete semi-algebraic subset of $\mathbf{R}$ is finite, the above argument implies that $A$ is an initial segment of $\mathbf{N}$. □

**Proof of Theorem 3.1.** Suppose that, for the sake of contradiction, satisfiability of $\exists^*\forall$-formulas is decidable. Let $P(x_1, ..., x_9)$ be a polynomial in $\mathbf{Z}[x_1, ..., x_9]$ (here $\mathbf{Z}$ denotes the set of integers).

*Claim*: The $\langle +, \times, 0, 1, <, S \rangle$-formula

$$H_P(x_1, ..., x_9, S) := I(S) \wedge \bigwedge_{i=1}^{9} S(x_i) \wedge P(x_1, ..., x_9) = 0$$

is satisfiable if and only if $P(x_1, ..., x_9) = 0$ has a solution in $\mathbf{N}^9$.

Proof of the claim: If $H_P(x_1, ..., x_9, S)$ is satisfiable, there exists an $A \subset \mathbf{R}$ that satisfies $I(S)$ and there exist $a_1, ..., a_9 \in A$ such that $P(a_1, ..., a_9) = 0$. By Lemma 3.2, $A$ is an initial segment of $\mathbf{N}$ and $a_1, ..., a_9$ are therefore natural numbers that satisfy $P(a_1, ..., a_9) = 0$.

On the other hand, if $P(x_1, ..., x_9) = 0$ has a solution $(a_1, ..., a_9) \in \mathbf{N}^9$, then we set $A = \{0, 1, ..., \max\{a_1, ..., a_9\}\}$ and observe that $\mathcal{R} \models (I(S) \wedge \wedge_{i=1}^{9} S(x_i) \wedge P(x_1, ..., x_9) = 0)[a_1, .., a_9, A]$ because $a_1, ..., a_9$ belong to $A$. This proves the claim.

Since Hilbert's 10th problem is undecidable for polynomials in 9 variables [5, 6], by the claim, satisfiability of the formula $H_P(x_1, ..., x_9, S)$, which can be rewritten into a formula

$$\forall x(S(0) \wedge (x < 1 \wedge S(x) \rightarrow x = 0) \wedge (x > 1 \wedge S(x) \rightarrow S(x-1)) \wedge \bigwedge_{i=1}^{9} S(x_i) \wedge P(x_1, ..., x_9) = 0)$$

of the $\exists^*\forall$-fragment of the $\Sigma_0^1$-fragment of second-order logic over $\langle +, \times, 0, 1, <, S \rangle$, must be undecidable. □

## 4. For $\forall^*$-sentences satisfiability is undecidable

**Theorem 4.1.** *For $\forall^*$-sentences of the $\Sigma_0^1$-fragment of second-order logic over $\langle +, \times, 0, 1, <, S_1, ..., S_k \rangle$, satisfiability is undecidable.*

**Proof.** Suppose that, for the sake of contradiction, satisfiability of $\forall^*$-formulas is decidable. Let $P(x_1, ..., x_9)$ be a polynomial in $\mathbf{Z}[x_1, ..., x_9]$.

Consider the sentence

$$\forall x_1 \cdots \forall x_9 \left( \bigwedge_{i=1}^{9} (S_i(0) \wedge (x_i < 1 \wedge S_i(x_i) \rightarrow x_i = 0) \right.$$

$$\wedge(x_i \geq 1 \wedge S_i(x_i) \to S_i(x_i - 1))) \wedge$$
$$\left(\bigwedge_{i=1}^{9}(S_i(x_i) \wedge \neg S_i(x_i + 1)) \to P(x_1, ..., x_9) = 0\right)\Bigg).$$

By Lemma 3.2, the sentences on the first two lines of this formula express that the $S_i$ are initial segments of $\mathbf{N}$. The last line states that their maxima are a solution of the equation $P(x_1, ..., x_9) = 0$. Therefore, this formula is satisfiable if and only if there are natural numbers (the maxima of the $S_i$) that satisfy the equation $P(x_1, ..., x_9) = 0$, which contradicts again the fact that Hilbert's 10th problem is undecidable for polynomials in 9 variables [5, 6]. □

We remark that the previous proof changes the proof of Theorem 3.1 in that the existential quantifiers that express the existence of a solution of the equation $P(x_1, ..., x_9) = 0$ have been moved to the existence of the sets $S_1, ..., S_9$.

We also remark, that instead of using 9 unary relation names we could use one binary relation $S(x, y)$ in which $x$ runs from 1 to 9, and for which for each of these $x$-values, the $y$ values give an initial segment of $\mathbf{N}$.

## 5. Undecidability results for equivalence

The results of the two previous sections have some corollaries concerning the decidability of equivalence of formulas. It is well-known that containment and equivalence of formulas in the $\Sigma^1_0$-fragment of second-order logic over the vocabulary $\langle +, \times, 0, 1, <, S_1, ..., S_k\rangle$ are undecidable [7, Chapter 2]. It is also known that containment and equivalence of conjunctive formulas (that is, conjunctions of possibly negated atomic formulas preceded by a first-order $\exists^*$-prefix) in this logic $\langle +, \times, 0, 1, <, S_1, ..., S_k\rangle$ are decidable [7, Chapter 2]. Here, we show that this is no longer the case if also disjunctions are allowed.

**Corollary 5.1.** *For the $\exists^*$-fragment of the $\Sigma^1_0$-fragment of second-order logic over the vocabulary $\langle +, \times, 0, 1, <, S_1, ..., S_k\rangle$ equivalence, and hence containment, are undecidable.*

**Proof.** Obviously, the undecidability of equivalence implies the undecidability of containment. Assume, for the sake of contradiction, that equivalence of formulas in the $\exists^*$-fragment of the $\Sigma^1_0$-fragment of $\langle +, \times, 0, 1, <, S_1, ..., S_k\rangle$ is decidable. We show that it follows that satisfiabiliy of formulas in the $\exists^*\forall^*$-fragment of the $\Sigma^1_0$-fragment of second-order logic over $\langle +, \times, 0, 1, <, S_1, ..., S_k\rangle$ is decidable, which contradicts Theorems 3.1 and 4.1. Indeed, let $\forall y_1 \cdots \forall y_m \varphi(y_1, ..., y_m, x_1, ..., x_n, S_1, ..., S_k)$ be a formula in the $\exists^*\forall^*$-fragment, with $\varphi(y_1, ..., y_m, x_1, ..., x_n, S_1, ..., S_k)$ quantifier-free. It is clear from the definitions that $\forall y_1 \cdots \forall y_m \varphi(y_1, ..., y_m, x_1, ..., x_n, S_1, ..., S_k)$ is not satisfiable if and only if $\neg\varphi(y_1, ..., y_m, x_1, ..., x_n, S_1, ..., S_k)$ is equivalent to the formula $0 = 0$ (true). Since, $\neg\varphi(y_1, ..., y_m, x_1, ..., x_n, S_1, ..., S_k)$ belongs to the $\exists^*$-fragment, this finishes the proof. □

**Corollary 5.2.** *For the $\exists^*\forall$-fragment , the $\exists^*\forall^*$-fragment and for the $\forall^*$-sentences of the $\Sigma^1_0$-fragment of second-order logic over $\langle +, \times, 0, 1, <, S_1, ..., S_k\rangle$ equivalence, and hence containment, are undecidable.*

**Proof.** First, assume, for the sake of contradiction, that equivalence of formulas in the $\exists^*\forall$-fragment of of the $\Sigma^1_0$-fragment of second-order logic over $\langle +, \times, 0, 1, <, S_1, ..., S_k\rangle$ is decidable. We show that it follows that satisfiabiliy of $\exists^*\forall$-fragment of the $\Sigma^1_0$-fragment of second-order logic over $\langle +, \times, 0, 1, <, S_1, ..., S_k\rangle$ is decidable, which contradicts Theorem 3.1. Indeed, let $\forall y \varphi(y, x_1, ..., x_n)$ be a formula, with $\varphi(y, x_1, ..., x_n)$ quantifier-free. It is clear from the definitions that that $\forall y \varphi(y, x_1, ..., x_n)$ is not satisfiable if and only if $\forall y \varphi(y, x_1, ..., x_n) \equiv 0 < 0$. This finishes the proof for the $\exists^*\forall$-fragment.

For the $\exists^*\forall^*$-fragment and for $\forall^*$-sentences, the proof is similar (now contradicting Theorem 4.1), since for a sentence $\forall x_1 \cdots \forall x_n \varphi(x_1, ..., x_n)$, we have that $\forall x_1 \cdots \forall x_n \varphi(x_1, ..., x_n)$ is not satisfiable if and only if $\forall x_1 \cdots \forall x_n \varphi(x_1, ..., x_n) \equiv 0 < 0$. This finishes the proof. □

**Acknowledgement.** Research partially supported by the following Argentinian and Belgian grants: PICT 2012-2403 and FWO G.0344.05.

## References

[1] J. Bochnak, M. Coste, M.-F. Roy, Real Algebraic Geometry, Volume 36 of Ergebenisse der Mathematik und ihrer Grenzgebiete, Folge 3, Springer-Verlag, Berlin, 1998.

[2] G. Jeronimo and J. Sabia, On the number of sets definable by polynomials, *Journal Algebra* **227**:2 (2000), 633–644.

[3] E. Börger, E. Grädel and Y. Gurevich, The Classical Decision Problem, Monographs in Mathematical Logic, Springer-Verlag, 2000.

[4] D. Grigoriev and N. N. (jr.) Vorobjov, Solving systems of polynomial inequalities in subexponential time, *Journal of Symbolic Computation* **5** (1988), 37–64.

[5] J. P. Jones and Y. V. Matijasevich, Exponential Diophantine representation of recursively enumerable sets, In J. Stern, editor, Proceedings of the Herbrand Symposium: Logic Colloquium '81, volume 107 of Studies in Logic and the Foundations of Mathematics, Amsterdam. North Holland, 1982, pp.159–177.

[6] Y. V. Matijasevich, Hilbert's Tenth Problem, MIT Press, Cambridge, MA, 1993.

[7] J. Paredaens, G. Kuper and L. Libkin, editors, Constraint databases, Springer-Verlag, 2000.

[8] A. Tarski, A Decision Method for Elementary Algebra and Geometry, University of California Press, 1951.

Rafael Grimson (corresponding author)
CONICET and Escuela de Ciencia y Tecnología,
Universidad de San Martín,
Campus Miguelete (CP1650).
San Martín, Provincia de Buenos Aires, Argentina
rgrimson@unsam.edu.ar

Bart Kuijpers
Databases and Theoretical Computer Science Group,
Hasselt University,
Agoralaan, Gebouw D, 3590 Diepenbeek, Belgium.