Recursive Tarskian Set Constraints without Functions*
(extended abstract)

Tomasz Charatonik and Leszek Pacholski

Institute of Computer Science, University of Wrocław
{tch,pacholsk}@ii.uni.wroc.pl

1 Introduction

There exist different formulations of a set constraint as well as of a μ-calculus. Following [1, 2] we define Tarskian μ-calculus, which subsumes Kozen's propositional μ-calculus [3], and we present Tarskian set constraints, which provide recursion and are interpreted in any first order structure.

The satisfiability of set constraints is an important problem in program verification. R. Givan, D. Kozen, D. McAllester and C. Witty presented in [1, 2] a complexity analysis of that problem for set expressions defined in Tarskian μ-calculus with different limits put on the syntax. For the constraints without recursion (without the μ operator) they proved that the satisfiability is:

- EXPTIME-complete, if expressions contain no deterministic operators,
- EXPTIME-complete, if there are no deterministic functions (constants are allowed),
- NEXPTIME-complete, if there are functions but not constants,
- decidable in 2-NEXPTIME, if both functions and constants are allowed.

The last case gap was closed by P. Mielniczuk and L. Pacholski who gave in [4] a NEXPTIME algorithm.

The authors of [1, 2] proved also in the papers that the satisfiability of recursive constraints without functions and constants is EXPTIME-complete, and that for the full Tarskian μ-calculus (with recursion, deterministic functions and constants) the constraint set satisfiability problem is undecidable.

The satisfiability problems of recursive Tarskian set constraints with constants but without functions, as well as with functions and without constants remained open. We give a polynomial time reduction of the former to a single set expression satisfiability, proving that it is EXPTIME-complete.

2 Tarskian μ-calculus

Let $\mathcal{V}$ be a set of set variables, $\mathcal{R}_k$ and $\mathcal{F}_k$ be sets of operation symbols of arity k, nondeterministic (relational) and deterministic (functional) respectively (we

* Sponsored by KBN grant 8T 11C 043 19

assume all the sets are pairwise disjoint), $\mathcal{R} = \bigcup_{k\geq 0} \mathcal{R}_k$, $\mathcal{F} = \bigcup_{k\geq 0} \mathcal{F}_k$. Assume the sets $\mathcal{R}_k$ and $\mathcal{F}_k$ for all k and the set $\mathcal{V}$ to be countably infinite. *Constants* are deterministic operations of arity 0, and deterministic operations of arity > 0 are called *functions*. We use the letters X, Y, ... to denote variables, P, Q, R, ... for nondeterministic operation symbols, f, g, ... for functional ones, and a, c for constant symbols.

Set expressions (or *expressions* in short) of the Tarskian μ-calculus are defined by the grammar:

$$E ::= X \mid \neg E \mid E_1 \cup E_2 \mid R(E_1, \ldots, E_k) \mid f(E_1, \ldots, E_k) \mid \mu X.E[X]$$

where $X \in \mathcal{V}$, $R \in \mathcal{R}_k$ and $f \in \mathcal{F}_k$, and all occurrences of X in $E[X]$ in the expression $\mu X.E[X]$ are in the scope of even number of negations. Expressions of the form $\mu X.E[X]$ are called *recursive*. Let $\mathrm{rel}(E)$, $\mathrm{fun}(E)$ and $\mathrm{const}(E)$ be the sets of all relational, functional and constant symbols in E, and $\mathrm{sym}(E) = \mathrm{rel}(E) \cup \mathrm{fun}(E) \cup \mathrm{const}(E)$. Let $\mathrm{var}(E)$ be the set of all variables of E.

An occurrence of a variable X is *bound* in E, if it is in the scope of some operator μX, otherwise it is *free*. An expression is *closed*, if it has no free variables. An expression E is in a *normal form*, if for each variable X in E there is no operator μX, or there is exactly one operator μX and all occurrences of X are bound.

Expressions are interpreted in any first order structure, denoted $\mathcal{M}$, with a valuation given. A structure $\mathcal{M}$ consists of a domain $D_{\mathcal{M}}$ (domain elements are called *states*) and an interpretation in $D_{\mathcal{M}}$ of relational and functional symbols, denoted $\mathcal{M}(\cdot)$, such that:

- if $R \in \mathcal{R}_k$, then $\mathcal{M}(R) \subseteq (D_{\mathcal{M}})^{k+1}$
- if $f \in \mathcal{F}_k$, then $\mathcal{M}(f)$ is a function $(D_{\mathcal{M}})^k \to D_{\mathcal{M}}$

A *valuation* is a function $\rho : \mathcal{V} \to 2^{D_{\mathcal{M}}}$, giving a meaning of subsets of $D_{\mathcal{M}}$ to free variables. By $\rho\,[X := S]$, where $S \subseteq D_{\mathcal{M}}$, we denote the valuation

$$\rho\,[X := S]\,(Y) = \begin{cases} S & \text{when} \quad Y = X \\ \rho\,(Y) & \text{when} \quad Y \neq X \end{cases}$$

An *interpretation* of a Tarskian μ-calculus expression E in a structure $\mathcal{M}$ with a valuation ρ is denoted $\mathcal{M}\,[\![E]\!]\,\rho$ (we will sometimes omit $\mathcal{M}$ or ρ, if the structure or the valuation is obvious from the context, or insignificant) and defined as follows:

$$\begin{aligned}
\mathcal{M}\,[\![X]\!]\,\rho &= \rho\,(X) \\
\mathcal{M}\,[\![\neg E]\!]\,\rho &= D_{\mathcal{M}} - \mathcal{M}\,[\![E]\!]\,\rho \\
\mathcal{M}\,[\![E_1 \cup E_2]\!]\,\rho &= \mathcal{M}\,[\![E_1]\!]\,\rho \cup \mathcal{M}\,[\![E_2]\!]\,\rho \\
\mathcal{M}\,[\![R(E_1, \ldots, E_k)]\!]\,\rho &= \{y \in D_{\mathcal{M}} \mid \exists x_1 \in \mathcal{M}\,[\![E_1]\!]\,\rho, \ldots, x_k \in \mathcal{M}\,[\![E_k]\!]\,\rho : \\
&\qquad \langle x_1, \ldots, x_k, y\rangle \in \mathcal{M}(R)\} \\
\mathcal{M}\,[\![f(E_1, \ldots, E_k)]\!]\,\rho &= \{\mathcal{M}(f)(x_1, \ldots, x_k) \mid \\
&\qquad x_1 \in \mathcal{M}\,[\![E_1]\!]\,\rho, \ldots, x_k \in \mathcal{M}\,[\![E_k]\!]\,\rho\} \\
\mathcal{M}\,[\![\mu X.E[X]]\!]\,\rho &= \bigcap \{S \subseteq D_{\mathcal{M}} \mid \mathcal{M}\,[\![E[X]]\!]\,\rho\,[X := S] \subseteq S\}
\end{aligned}$$

Constant symbols are interpreted as singleton subsets of $D_{\mathcal{M}}$, and 0-ary nondeterministic operation symbols as set constants. An expression $E[X]$ determines an operator $f_{E[X]} : 2^{D_{\mathcal{M}}} \to 2^{D_{\mathcal{M}}}$, such that $f_{E[X]}(S) = \mathcal{M}\,[\![E[X]]\!]\,\rho\,[X := S]$. When $E[X]$ is positive in X, then $f_{E[X]}$ is a monotone function (i.e.: if $S_1 \subseteq S_2$, then $f_{E[X]}(S_1) \subseteq f_{E[X]}(S_2)$), so it has the least and the greatest fix point (Knaster–Tarski theorem). The meaning of an expression $\mu X.E[X]$ is the least fix point of $f_{E[X]}$. We use abbreviations $E_1 \cap E_2$ for $\neg(\neg E_1 \cup \neg E_2)$, $\top$ for $P \cup \neg P$ and $\bot$ for $\neg\top$ to denote intersection, universal and empty set.

An expression E is *satisfied* by a model $\mathcal{M}$ with a valuation ρ, written $\langle \mathcal{M}, \rho \rangle \models E$, when $\mathcal{M}\,[\![E]\!]\,\rho \neq \emptyset$. We say that E is *satisfiable*, if there exist a model and a valuation that satisfy E. Determining whether a given set expression is satisfiable may be reduced ([1, 2]) in linear time to expression satisfiability in Herbrand μ-calculus, which is decidable in deterministic exponential time. This gives

Theorem 1. *The satisfiability of a set expression of full Tarskian μ-calculus is in EXPTIME.*

3 Set Constraints

Definition 2. *A formula of a form* $E_1 \subseteq E_2$ *or* $E_1 \not\subseteq E_2$, *where* E_1 *and* E_2 *are set expressions is called a* set constraint. *The former one is* positive, *and the latter is* negative. *A constraint is called* closed *if both its expressions are closed.*

Definition 3. *A positive constraint* $E_1 \subseteq E_2$ *(negative* $E_1 \not\subseteq E_2$ *) is* satisfied *in a model* $\mathcal{M}$ *with a valuation* ρ, *denoted* $\langle \mathcal{M}, \rho \rangle \models E_1 \subseteq E_2$ *(* $\langle \mathcal{M}, \rho \rangle \models E_1 \not\subseteq E_2$ *), if* $\mathcal{M}\,[\![E_1]\!]\,\rho \subseteq \mathcal{M}\,[\![E_2]\!]\,\rho$ *(* $\mathcal{M}\,[\![E_1]\!]\,\rho \not\subseteq \mathcal{M}\,[\![E_2]\!]\,\rho$ *). A constraint C is* satisfiable *if there exist a model $\mathcal{M}$ and a valuation ρ, such that* $\langle \mathcal{M}, \rho \rangle \models C$ *.*

Definition 4. *A set Γ of set constraints is* satisfied *by a pair* $\langle \mathcal{M}, \rho \rangle$ *, written* $\langle \mathcal{M}, \rho \rangle \models \Gamma$ *, if for every constraint* $C \in \Gamma$ *we have* $\langle \mathcal{M}, \rho \rangle \models C$ *. A set Γ is* satisfiable, *if there exist $\mathcal{M}$ and ρ such that* $\langle \mathcal{M}, \rho \rangle \models \Gamma$ *.*

We consider satisfiability of finite sets of constraints in nonempty models only. If a constraint C (a set of constraints Γ) is satisfied in a model $\mathcal{M}$ with any valuation, we will write it $\mathcal{M} \models C$ ($\mathcal{M} \models \Gamma$).

Let Γ_{pos} denote the set of positive constraints in Γ, and Γ_{neg} be the set of negative constraints in Γ. Let $\mathrm{rel}(\Gamma)$ denote the set of all relational symbols occurring in Γ, $\mathrm{const}(\Gamma)$ be the set of all constants in Γ, and $\mathrm{ar}(R)$ be the arity of a symbol R. Note that $\mathrm{ar}(R) = k$ means that R takes k arguments, and thus is interpreted as a $(k+1)$-ary relation.

Now we present our result:

Theorem 5. *The satisfiability of recursive Tarskian set constraints with no function symbols of arity > 0 is in EXPTIME.*

Proof (5, sketch). The problem is polynomially reducible to the satisfiability problem of a single set expression of Tarskian μ-calculus. The proof presented in [1, 2] for nondeterministic constraints may be extended to the case with constants.

Let $\mathit{\Gamma}$ be the given set of constraints. For $\mathit{\Gamma}_{\text{pos}} = \{E_1 \subseteq E'_1, \ldots, E_n \subseteq E'_n\}$ define the following set expressions

$$E_{\text{cex}}(\mathit{\Gamma}) = (E_1 \cap \neg E'_1) \cup \ldots \cup (E_n \cap \neg E'_n),$$

$$E_{\text{viol}}(\mathit{\Gamma}) = \mu X. E_{\text{cex}}(\mathit{\Gamma}) \cup \bigcup_{\substack{R \in \text{rel}(\mathit{\Gamma}) \\ 1 \leq i \leq \text{ar}(R)}} R(\top, \ldots, \top, X, \top, \ldots, \top),$$

where X is a fresh variable and occurs in the i-th position in the last union (the union is over symbolss of arity ≥ 1).

The expression $E_{\text{cex}}(\mathit{\Gamma})$ defines a set of counter examples for $\mathit{\Gamma}_{\text{pos}}$, i.e. those states that break the positive constraints in $\mathit{\Gamma}$, and $E_{\text{viol}}(\mathit{\Gamma})$ describes the set of states available from $\mathcal{M}\,[\![E_{\text{cex}}(\mathit{\Gamma})]\!]$ by applying operations denoted by symbols of $\text{rel}(\mathit{\Gamma})$.

It can be shown that for any model $\mathcal{M}$

$$\mathcal{M} \models \mathit{\Gamma}_{\text{pos}} \iff \mathcal{M}\,[\![E_{\text{cex}}(\mathit{\Gamma})]\!] = \emptyset \iff \mathcal{M}\,[\![E_{\text{viol}}(\mathit{\Gamma})]\!] = \emptyset. \tag{1}$$

For $\mathit{\Gamma}_{\text{neg}} = \{F_1 \not\subseteq F'_1, \ldots, F_m \not\subseteq F'_m\}$ and $\text{const}(\mathit{\Gamma}) = \{c_1, \ldots, c_k\}$ define

$$\begin{aligned} E(\mathit{\Gamma}) = P\Big(& F_1 \cap \neg F'_1 \cap \neg E_{\text{viol}}(\mathit{\Gamma}), \ldots, F_m \cap \neg F'_m \cap \neg E_{\text{viol}}(\mathit{\Gamma}), \\ & c_1 \cap \neg E_{\text{viol}}(\mathit{\Gamma}), \ldots, c_k \cap \neg E_{\text{viol}}(\mathit{\Gamma})\Big) \cap \neg E_{\text{viol}}(\mathit{\Gamma}), \end{aligned}$$

for a fresh symbol $P \in \mathcal{R}_{m+k}$ (i.e. $P \notin \text{rel}(\mathit{\Gamma})$).

We can show

Lemma 6. *$\mathit{\Gamma}$ is satisfiable $\iff$ $E(\mathit{\Gamma})$ is satisfiable.*

Proof (6). For any constant symbol c and model $\mathcal{M}$ let $c^{\mathcal{M}}$ denote the state of $\mathcal{M}$ such that $\mathcal{M}\,[\![c]\!] = \{c^{\mathcal{M}}\}$.

$\boxed{\Longrightarrow}$

Assume $\mathcal{M} \models \mathit{\Gamma}$. $\mathcal{M} \models \mathit{\Gamma}_{\text{pos}}$ implies that $\mathcal{M}\,[\![\neg E_{\text{viol}}(\mathit{\Gamma})]\!] = D_{\mathcal{M}}$. Choose any $s_i \in \mathcal{M}\,[\![F_i \cap \neg F'_i]\!]$ for all $1 \leq i \leq m$, and any $s \in D_{\mathcal{M}}$. Define

$$\mathcal{M}(P) = \{\langle s_1, \ldots, s_m, c_1^{\mathcal{M}}, \ldots, c_k^{\mathcal{M}}, s\rangle\}.$$

Hence $s \in \mathcal{M}\,[\![E(\mathit{\Gamma})]\!]$.

$\boxed{\Longleftarrow}$

Suppose now that $s \in \mathcal{M}\,[\![E(\mathit{\Gamma})]\!]$. So for all $1 \leq i \leq m$ there exist $s_i \in \mathcal{M}\,[\![F_i \cap \neg F'_i \cap \neg E_{\text{viol}}(\mathit{\Gamma})]\!]$. Let $\mathcal{M}'$ be a model identical to $\mathcal{M}$ excepting the interpretation of the symbol P limited to one tuple:

$$\mathcal{M}'(P) := \{\langle s_1, \ldots, s_m, c_1^{\mathcal{M}'}, \ldots, c_k^{\mathcal{M}'}, s\rangle\}.$$

The model $\mathcal{M}'$ preserves essential properties:

$$s_i \in \mathcal{M}' \llbracket F_i \cap \neg F_i' \cap \neg E_{\mathrm{viol}}(\Gamma) \rrbracket \text{ for all } 1 \leq i \leq m$$

and

$$c_j^{\mathcal{M}'} \notin \mathcal{M}' \llbracket E_{\mathrm{viol}}(\Gamma) \rrbracket \text{ for all } 1 \leq j \leq k,$$

because P does not occur in Γ. Thus $\{s\} = \mathcal{M}' \llbracket E(\Gamma) \rrbracket$.

If we restrict the structure $\mathcal{M}'$ to those states for which s is in reflexive transitive closure of relations from $\mathrm{rel}(\Gamma) \cup \{P\}$, then the obtained *inverse image structure*† models Γ. □

□

It is shown in [1, 2] that the problem requires exponential time. Thus, we are given

Corollary 7. *The satisfiability of recursive Tarskian set constraints containing no function symbols of arity* > 0 *is complete for EXPTIME.*

References

1. Robert Givan, Dexter Kozen, David A. McAllester and Carl Witty, *Tarskian Set Constraints*, Proc. LICS (1996), pp. 138–147
2. Robert Givan, Dexter Kozen, David A. McAllester and Carl Witty, *Tarskian Set Constraints*, `http://dynamo.ecn.purdue.edu/~givan/brown/papers/lics96-journal-99.ps`
3. Dexter Kozen, *Results on the propositional μ-calculus*, Theoretical Computer Science 27 (1983), pp. 333–354
4. Paweł Mielniczuk and Leszek Pacholski, *Tarskian Set Constraints Are in NEXPTIME*, Proc. MFCS'98, LNCS 1450, pp. 589–596

† The notion and properties of inverse image structures can be found in [1]; in [2] they are called *inverse closure substructures*.