

Decision Procedures for Multisets with Cardinality Constraints

Ruzica Piskac Viktor Kuncak

École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

Schloss Dagstuhl, October 2007

Multisets

Definition

- ▶ **Multiset** (bag) is a collection of elements where an element can occur several times
- ▶ Formally, multisets is a function $f : E \rightarrow \mathbb{N}$ (E - finite universe)

Selected operations and relations on multisets:

- ▶ Plus $(m_1 \uplus m_2)(e) = m_1(e) + m_2(e)$
- ▶ Union $(m_1 \cup m_2)(e) = \max\{m_1(e), m_2(e)\}$
- ▶ Intersection $(m_1 \cap m_2)(e) = \min\{m_1(e), m_2(e)\}$
- ▶ Subset $m_1 \subseteq m_2 \iff \forall e. m_1(e) \leq m_2(e)$

Sets in Software Analysis and Verification

We can use sets to abstractly describe data structure operations.

content - abstract set representing list content

```
public void add(Object o1)
ensures content = old content  $\cup$  {o1}
{
    Node n = new Node();
    n.data = o1;
    n.next = first;
    first = n;
}
```

⇒ PROBLEM: what if the copy of an element already exists in the list?

Multisets in Software Analysis and Verification

```
public void add(Object o1)
ensures content = old content  $\uplus$  {o1}
{
    Node n = new Node();
    n.data = o1;
    n.next = first;
    first = n;
}
```

Multisets correctly capture repeated elements.

Cardinality Constraints on Multisets

```
public void add(Object o1)
ensures content = old content  $\uplus$  {o1}
{ ... }
```

Consider invariant:

$$\text{size} = |\text{content}|$$

saying that 'size' variable stores number of list elements.

Verification conditions for preserving invariants:

- ▶ Insertion: $|x| = 1 \rightarrow |L \uplus x| = |L| + 1$
- ▶ Deletion: $x \subseteq L \wedge |x| = 1 \rightarrow |L \setminus x| = |L| - 1$

(x denotes the singleton set representing the element)

Multisets in interactive theorem proving

- ▶ Multiset library of the interactive theorem prover Isabelle
- ▶ Several Isabelle developments build on the Multiset library
 - ▶ permutation library
 - ▶ specifications of sorting algorithms
 - ▶ specification of UNITY parallel programming approach

Our Results

- ▶ describe new decision procedure for a quantifier-free multiset formulas with cardinality operator
- ▶ show that it is solvable in polynomial space
- ▶ show that adding quantifiers yields undecidability

Example Formulas in Our Language

Inner linear arithmetic formulas:

▶ $a \subseteq b$, meaning $\forall e. a(e) \leq b(e)$

▶ $a = \text{setof}(b)$, meaning

$$\forall e. ((b(e) = 0 \wedge a(e) = 0) \vee (b(e) > 0 \wedge a(e) = 1))$$

In fact, we allow any expression $\forall e. F(m_1(e), \dots, m_k(e))$
(F is quantifier-free Presburger arithmetic formula)

Outer linear arithmetic formulas:

▶ $|a| = k$, meaning $\sum_e a(e) = k$

We allow any expression $\sum_F t = k$ where

▶ F - quantifier-free Presburger arithmetic formula

▶ t - Presburger arithmetic term (or tuple of terms)

We allow conditional expression $\text{ite}(F, t_1, t_2)$ within terms.

Formal Definition of the Language

$$F ::= A \mid F \wedge F \mid \neg F$$

$$A ::= M=M \mid M \subseteq M \mid \forall e.F^i \mid A^o$$

$$F^o ::= A^o \mid F^o \wedge F^o \mid \neg F^o$$

$$A^o ::= t^o \leq t^o \mid t^o = t^o \mid (t^o, \dots, t^o) = \sum_{F^i} (t^i, \dots, t^i)$$

$$t^o ::= k \mid |M| \mid C \mid t^o + t^o \mid C \cdot t^o \mid \text{ite}(F^o, t^o, t^o)$$

$$F^i ::= A^i \mid F^i \wedge F^i \mid \neg F^i$$

$$A^i ::= t^i \leq t^i \mid t^i = t^i$$

$$t^i ::= m(e) \mid C \mid t^i + t^i \mid C \cdot t^i \mid \text{ite}(F^i, t^i, t^i)$$

$$M ::= m \mid \emptyset \mid M \cap M \mid M \cup M \mid M \uplus M \mid M \setminus M \mid M \setminus\setminus M \mid \text{setof}(M)$$

Decision Procedure

1. reduce to normal form
2. replace multiset sums with PA sums
3. find semilinear sets characterizing the set of solutions of formulas under the sum
4. generate PA formula for the results of sums
5. check satisfiability of resulting formula

Sum Normal Form

Definition

A multiset formula is in **sum normal form** iff it is of the form

$$P \wedge (u_1, \dots, u_n) = \sum_{\text{true}} (t_1, \dots, t_n)$$

where

- ▶ P is a quantifier-free Presburger arithmetic formula without any multiset variables
- ▶ the variables in t_1, \dots, t_n occur only as expressions of the form $m(e)$ for m a multiset variable and e the fixed index variable

Reduction to Sum Normal Form

- ▶ express all multiset expressions using $\forall e. F$

Example:

$$m_3 = m_1 \setminus m_2 \rightsquigarrow$$

$$\forall e. \text{ite}(m_1(e) \leq m_2(e), 0, m_1(e) - m_2(e)) = m_3(e)$$

- ▶ replace all the expressions of the form $\forall e. F$ with sum expressions:

$$\forall e. F \rightsquigarrow \sum \text{ite}(F, 0, 1) = 0$$

- ▶ group all sums into one, using vectors:

$$\sum t_1 = k_1 \wedge \sum t_2 = k_2 \rightsquigarrow \sum (t_1, t_2) = (k_1, k_2)$$

Multisets Elimination

A formula in the sum normal form looks as follows:

$$P \wedge (u_1, \dots, u_n) = \sum_{x \in E} (t_1, \dots, t_n)$$

where free variables of t_1, \dots, t_n are multiset variables m_1, \dots, m_q and P is a quantifier-free Presburger arithmetic formula.

Goal: replace this sum with construct involving only PA formulas.

\rightsquigarrow semilinear sets

Semilinear Sets

Definition

Let $C_1, C_2 \subseteq \mathbb{N}^k$ be sets of vectors of non-negative integers. We define:

$$C_1 + C_2 = \{x_1 + x_2 \mid x_1 \in C_1 \wedge x_2 \in C_2\}$$

$$C_1^* = \{x_1 + \dots + x_n \mid x_i \in C_1 \wedge n \geq 0\}$$

Semilinear sets

Linear set = set of form $\{x\} + C^*$ for $x \in \mathbb{N}^n$ and $C \subseteq \mathbb{N}^n$ finite

Semilinear set = finite union of linear sets

Multisets Elimination

A formula in the sum normal form:

$$P \wedge (u_1, \dots, u_n) = \sum_{x \in E} (t_1, \dots, t_n)$$

is equisatisfiable with the formula

$$P \wedge (u_1, \dots, u_n) \in \{(t'_1, \dots, t'_n) \mid x_1, \dots, x_p \in \mathbb{N}\}^*$$

where t'_i is t_i in which each $m_k(e)$ is replaced by fresh var x_k

Can we describe $(u_1, \dots, u_n) \in \{(t'_1, \dots, t'_n) \mid x_1, \dots, x_p \in \mathbb{N}\}^*$ by PA formula?

Values of Tuples of Terms

Can we describe $(u_1, \dots, u_n) \in \{(t'_1, \dots, t'_n) \mid x_1, \dots, x_p \in \mathbb{N}\}^*$ by PA formula?

$$\{(t'_1, \dots, t'_n) \mid x_1, \dots, x_p \in \mathbb{N}\} = \{(z_1, \dots, z_n) \mid z_1 = t'_1 \wedge \dots \wedge z_n = t'_n\}$$

Need to describe sums of solutions of PA formulas:

Can a given vector (u_1, \dots, u_n) be expressed as a sum of some number of solutions of a given PA formula?

Semilinear Sets

In [GinsburgSpanier1968] it was shown:

- ▶ semilinear sets are closed under union, intersection and negation
- ▶ a solution of PA formula is a semilinear set

We can also show

- ▶ if S is semilinear, then S^* is given by PA formula

This proves that

$$(u_1, \dots, u_n) \in \{(t'_1, \dots, t'_n) \mid x_1, \dots, x_p \in \mathbb{N}\}^*$$

is effectively expressible as PA formula.

Consequently:

Satisfiability of our multiset constraints reduces to satisfiability of quantifier-free PA formulas.

PSPACE Algorithm for Deciding Linear Arithmetics with Sum Constraints

Given the formula $P \wedge (u_1, \dots, u_n) \in \{(z_1, \dots, z_n) \mid F\}^*$,

1. guess the values c_1, \dots, c_n of variables u_1, \dots, u_n such that P holds (in NP using Papadimitriou and Pottier)
2. checks whether the constraint $(c_1, \dots, c_n) \in \{(z_1, \dots, z_n) \mid F\}^*$ has a solution.

We found an algorithm for step 2 that runs in PSPACE.

Undecidability of quantified multiset constraints

Hilbert's Tenth Problem

Given a polynomial Diophantine equation with integer coefficient, is there a general algorithm for deciding whether the equation has a solution in integers.

Reduction

Since Hilbert's tenth problem is undecidable, we show its reduction to quantified multiset constraints.

Reduction

► Addition

$$x + y = z \iff \exists a, b. a = |x| \wedge b = |y| \wedge |a \uplus b| = z$$

► Multiplication

$$x \cdot y = z \iff \exists p. z = |p| \wedge x = |\text{setof}(p)| \wedge (\forall m. |m| = z \wedge |\text{setof}(m)| = 1 \wedge \text{setof}(m) \subseteq p \implies |m \cap p| = y)$$

Related Work

Zarba 2002 - decision procedures for quantifier-free multisets
but without the cardinality operator

Related Work

- [Zarba 2002](#) - decision procedures for quantifier-free multisets but without the cardinality operator
- [Lugiez 2005](#) - showed the decidability of quantified multiset formulas with a weaker form of cardinality operator that counts only distinct elements in a multiset

Related Work

- [Zarba 2002](#) - decision procedures for quantifier-free multisets but without the cardinality operator
- [Lugiez 2005](#) - showed the decidability of quantified multiset formulas with a weaker form of cardinality operator that counts only distinct elements in a multiset
 - showed decidability of certain quantifier-free expressible multiset formulas with cardinality operator

Related Work

- [Zarba 2002](#) - decision procedures for quantifier-free multisets but without the cardinality operator
- [Lugiez 2005](#) - showed the decidability of quantified multiset formulas with a weaker form of cardinality operator that counts only distinct elements in a multiset
- showed decidability of certain quantifier-free expressible multiset formulas with cardinality operator
 - stated that the decidability of quantified multiset formulas with the general cardinality operator is an open problem

Related Work

- ▶ The algorithm we presented shows decidability, but produces exponentially large QFPA formulas \Rightarrow non-deterministic exponential bound

Pottier 1991 - showed that the solution set of $Ax = b$ is a semilinear set and also showed how to construct it efficiently (singly exponential)

Papadimitriou 1981 - showed the bounds on solutions in Presburger arithmetic formula

Conclusions

- ▶ we describe novel decision procedure for all quantifier-free multiset formulas with cardinality operator
- ▶ we showed that it is solvable in polynomial space
- ▶ we showed that adding quantifiers yields undecidability