Theoretical Computer Science ••• (••••) •••–•••

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Computing and estimating the volume of the solution space of SMT(LA) constraints

Cunjing Ge [a,c], Feifei Ma [a,b,c,*], Peng Zhang [d], Jian Zhang [a,d]

[a] State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, China
[b] Laboratory of Parallel Software and Computational Science, Institute of Software, Chinese Academy of Sciences, China
[c] University of Chinese Academy of Sciences, China
[d] School of Computer Science and Technology, Shandong University, China

ARTICLE INFO

Article history:
Received 4 November 2015
Received in revised form 16 October 2016
Accepted 28 October 2016
Available online xxxx

Keywords:
SMT
Volume
Counting
Convex polytope

ABSTRACT

The satisfiability modulo theories (SMT) problem is a decision problem, i.e., deciding the satisfiability of logical formulas with respect to combinations of background theories (like reals, integers, arrays, bit-vectors). In this paper, we study the counting version of SMT with respect to linear arithmetic – SMT(LA), which generalizes both model counting and volume computation of convex polytopes. We describe a method for estimating the volume of convex polytopes based on the Multiphase Monte-Carlo method. It employs a new technique to reutilize random points, so that the number of random points can be significantly reduced. We prove that the reutilization technique has no side-effect on the error. We also investigate a simplified version of hit-and-run method: the coordinate directions method. Based on volume estimation method for polytopes, we present an approach to estimating the volume of the solution space of SMT(LA) formulas. It employs a heuristic strategy to accelerate the volume estimation procedure. In addition, we devise some specific techniques for instances that arise from program analysis.

1. Introduction

The satisfiability (SAT) problem in the propositional logic is a fundamental problem in computer science. But in practice, many problems cannot be expressed by propositional formulas directly or naturally. In recent years, there have been a lot of works on solving the Satisfiability Modulo Theories (SMT) problem, which try to decide the satisfiability of logical formulas with respect to combinations of background theories (like reals, integers, arrays, bit-vectors). SMT can be regarded as an extension to SAT, as well as a kind of constraint satisfaction problem (CSP). Quite efficient SMT solvers have been developed, such as CVC3/CVC4, Z3 and Yices [1,10,11].

The counting version of CSP, i.e., #CSP, has been studied by various researchers [5,6]. There has also been much work on the model counting problem in the propositional logic, i.e., counting the number of models of a propositional formula. It is closely related to approximate reasoning [33,8].

On the other hand, the counting version of SMT, i.e., #SMT, has not been studied much. In this paper, we focus on the #SMT problem with respect to the theory of linear arithmetic – #SMT(LA). Given a set of SMT(LA) constraints, we would like to know how many solutions there are. Or, in other words, how large the solution space is. The problem can be regarded

* Corresponding author.
E-mail addresses: gecj@ios.ac.cn (C. Ge), maff@ios.ac.cn (F. Ma), algzhang@sdu.edu.cn (P. Zhang), zj@ios.ac.cn (J. Zhang).

http://dx.doi.org/10.1016/j.tcs.2016.10.019
0304-3975/

as an extension to SMT solving, and also a generalization of both the model counting problem in the propositional logic and volume computation of convex polytopes. It has recently gained some attention in the software engineering community [23,16].

An SMT(LA) formula is satisfiable if and only if there exists a Boolean assignment to its linear inequalities such that the SMT formula is evaluated to true in Boolean level, and the conjunction of inequalities is also consistent. Such Boolean assignment is called feasible assignment. The linear system corresponding to a feasible assignment forms a convex polytope. Ma et al. [32] proposed an exact approach for #SMT(LA) problem which integrates SMT solving with volume computation for convex polytopes. However, exact volume computation in general is a difficult problem. It has been proved to be #P-hard, even for explicitly described polytopes [12,21,22]. Yet, in many applications, it suffices to have an approximate value of the volume of the solution space. Therefore, it is desirable to study highly efficient methods for *estimating* the volume of the solution space.

Volume computation for convex polytopes is a classical problem in mathematics. The high complexity of exact volume computation procedure for convex polytopes is the bottleneck of the approach in [32]. On the other hand, volume estimation methods for convex bodies have been extensively studied in theory. The Monte-Carlo method is a straightforward way to estimate the volume of a convex body. However, it suffers from the curse of dimensionality, which means the possibility of sampling inside a certain space in the target object decreases very quickly while the dimension increases. As a result, the sample size has to grow exponentially to achieve a reasonable estimation. To avoid the curse of dimensionality, Dyer et al. proposed a polynomial time randomized approximation algorithm (called Multiphase Monte-Carlo Algorithm) [13]. At first, the theoretical complexity of this algorithm is $O^*(n^{23})$,[1] it was reduced to $O^*(n^4)$ at last by Lovász, Kannan et al. [28,19,26,31]. Despite the polynomial time results and reduced complexity, there is still a lack of practical implementation.

In this paper, we first describe an algorithm for estimating the volume of convex polytopes which is based on the Multiphase Monte-Carlo method. The algorithm is augmented with a new technique to reutilize random points, so that the number of random points can be significantly reduced. We prove that the reutilization technique has no side-effect on the error. We also investigate a simplified version of hit-and-run method: the coordinate directions method, which has never been employed in volume estimation before. Then we integrate our volume estimation method for convex polytopes into the framework of solving #SMT(LA) problems. We propose a heuristic improvement called two-round strategy, which automatically adjusts the number of random points for each invocation of polytope volume estimation. Besides, for instances arise from program analysis, we also introduce some effective techniques.

The rest of this paper is organized as follows. We first describe some basic concepts and notations, as well as some essential techniques and tools in Section 2. Then Section 3 reviews some related works. In Section 4, we present our volume estimation method for convex polytopes, with theoretical analysis. Section 5 presents our approach to volume computation and estimation for SMT(LA) formulas. In Section 6, we further discuss how to improve our approach for the instances generated from program analysis. Section 7 presents some experimental results. Finally, we conclude in Section 8.

This article is an extension of a conference paper [15] presented at the 9th International Workshop of Frontiers in Algorithmics.

2. Preliminaries

This section describes some basic concepts and notations. We also mention some existing techniques and tools that will be used later.

2.1. SMT(LA) formulas

Definition 1. A **linear arithmetic (LA)** constraint is an expression that may be written in the form $a_1x_1 + a_2x_2 + \cdots + a_nx_n \ op \ a_0$. Here $x_1, x_2, \ldots, x_n$ are numeric variables, $a_0, a_1, a_2, \ldots, a_n$ are constant coefficients, and $op \in \{<, \leq, >, \geq, =, \neq\}$.

Definition 2. An SMT formula ϕ over LA constraints, i.e., an **SMT(LA) formula**, can be represented as a Boolean formula $PS_\phi(b_1, \ldots, b_n)$ together with definitions in the form: $b_i \equiv c_i$. Here c_is are LA constraints. PS_ϕ is the *propositional skeleton* of the formula ϕ.

The propositional skeleton contains logical operators, like AND, OR, NOT. A simple example of SMT(LA) formulas is

$$(x + y < 1 \text{ OR } x \geq y) \text{ AND } (x + y < 1 \text{ OR } x < y \text{ OR } b).$$

Let the Boolean variables b_1 and b_2 represent the linear inequalities $x + y < 1$ and $x < y$ respectively. Then we obtain the propositional skeleton

$$(b_1 \text{ OR } (\text{NOT } b_2)) \text{ AND } (b_1 \text{ OR } b_2 \text{ OR } b).$$

[1] The "soft-O" notation O^* indicates that we suppress factors of $\log n$ as well as factors depending on other parameters like the error bound.

Please cite this article in press as: C. Ge et al., Computing and estimating the volume of the solution space of SMT(LA) constraints, Theoret. Comput. Sci. (2016), http://dx.doi.org/10.1016/j.tcs.2016.10.019