PSMT: Satisfiability Modulo Theories Meets Probability Distribution

Fuqi Jia, Rui Han, Xutong Ma, Baoquan Cui, Minghao Liu, Pei Huang, Feifei Ma, and Jian Zhang

September 13, 2023

A Story to PSMT

One day, when utilizing the SMT solver, my colleague who concentrates on static analysis asked: Why is the solution obtained by the SMT solver difficult to understand (too large, too small, or not realistic), especially for practical instances?
I think it is an interesting question, and our preliminary answer is to introduce probability distribution into the SMT-solving process.

A Simple Example

```
void HealthCare(double weight, double height){
    BMI = weight / (height * height);
    if(BMI >= 18.5 && BMI <= 24.9) {...} // healthy
    else {...} // treatment suggestion
}
```

Path Condition: $18.5 \leq \frac{w}{h^2} \leq 24.9 \Longrightarrow$ healthy.

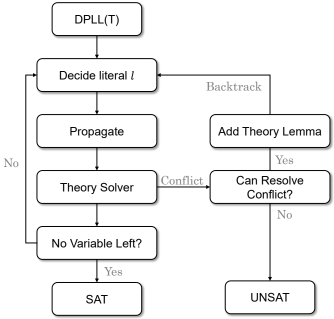
An SMT solver finds an assignment for $healthy = \top$, $w = 0.375, h = 0.125$, which is correct but not real. In practice, such variables usually obey some distributions, for example, $w \sim N(60, 10), h \sim N(1.7, 0.1)$. Our algorithm can obtain an assignment $w = 66.82, h = 1.75$ on the distributions.

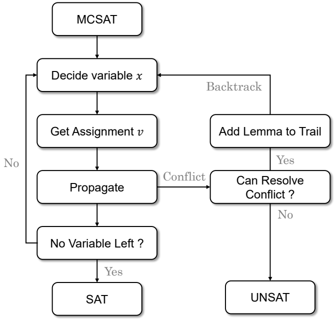
Contributions and Values

1. We introduce the probability distribution into SMT and define it as the PSMT problem. It allows the variables in solutions to satisfy certain probability distributions.
2. We construct the Prob-MCSAT algorithm, combining probability distribution and conflict-driven process.
3. Given distributions for variables, Prob-MCSAT can generate a plausible solution that is closer to reality.

SMT and MCSAT

- Satisfiability Modulo theories (SMT) is an area of automated deduction that studies methods for checking the satisfiability of first-order formulas with respect to some logical theory T of interest [1, 2].





SMT with Probability

- It has many applications in probabilistic program analysis [3], stochastic hybrid systems [4] and etc.
- Most of the works concentrate on the probability that the constraints are satisfied [5].
- Recent work on SMT sampling [6] approximates sample points by adding extra constraints after the solver obtains a solution.
- It seems an unexplored area that solves an SMT problem with probability for variables.

Domain under Constraint

Definition 1 (Domain under Constraint)

Given a constraint ψ and variable x, the domain under constraint $D(\psi, x)$ is the domain of x where when assigning x any value in $D(\psi, x)$, there exists a full assignment satisfying ψ.

Distribution under Constraint

Definition 2 (Distribution under Constraint)

Given a constraint ψ and a distribution for variable x, whose probability density function is $P(x), x \in \mathbb{R}$, the distribution of x under constraint ψ is a refined distribution whose probability density function is $\widetilde{P}(\psi, x)$,

$$\widetilde{P}(\psi, x) = \begin{cases} \frac{1}{\int_{D(\psi,x)} P(x)dx} P(x), x \in D(\psi, x), \\ 0, x \in \mathbb{R} - D(\psi, x). \end{cases}$$

Probability Satisfiability Modulo Theories

Definition 3 (Probability Satisfiability Modulo Theories)

Given an n variable SMT constraint ψ, a value distribution P, and a variable order σ, find an assignment $\alpha \models \psi$, i.e., α satisfies ψ. Meanwhile, the i-th variable x follows distribution under constraint, i.e., $\alpha[x_i] \sim \widetilde{P}(\psi(\{\alpha[x_{\sigma_1}], \cdots, \alpha[x_{\sigma_{i-1}}]\}), x_i)$ where $1 \leq i \leq n$.

PSMT Example

- Given a constraint $\psi = \{-2 \leq x \leq 2 \wedge x^2 + y \geq 1\}$, and x follows a uniform distribution in $[-2, 2]$, i.e., $U([-2, 2])$.
- If a partial assignment is $\alpha = \{y \leftarrow 0\}$, then we have $\psi(\alpha) = -2 \leq x \leq 2 \wedge x^2 \geq 1$.
- *Domain under Constraint* for x:

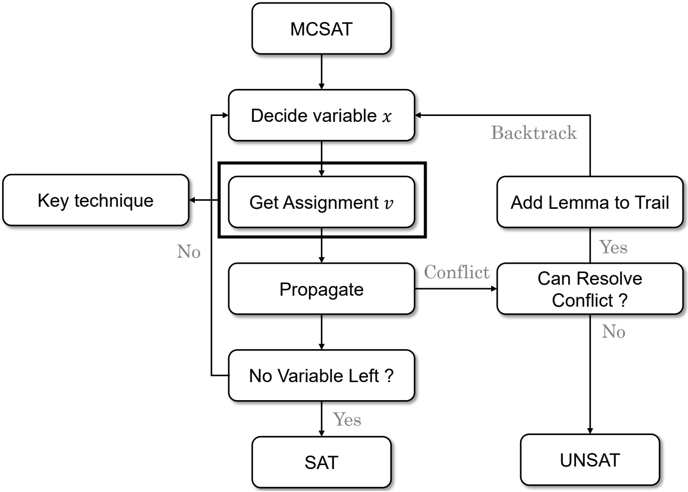
$$D(\psi(\alpha), x) = [-2, -1] \cup [1, 2].$$

- *Distribution under Constraint* for x:

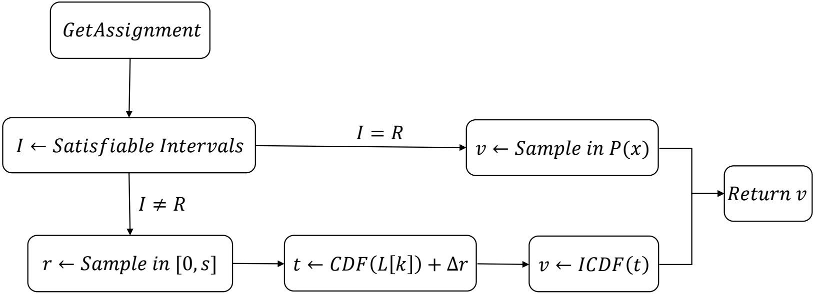
$$\widetilde{P}(\psi(\alpha), x) = \frac{1}{\int_{-2}^{-1} \frac{1}{4} dx + \int_{1}^{2} \frac{1}{4} dx} P(x) = 2P(x) = \frac{1}{2},$$

  where $x \in D(\psi(\alpha), x)$.
- Sample under the distribution and find an assignment $\{x \leftarrow v, y \leftarrow 0\}$ can be a solution to the PSMT problem.

Overall Framework



GetAssignment

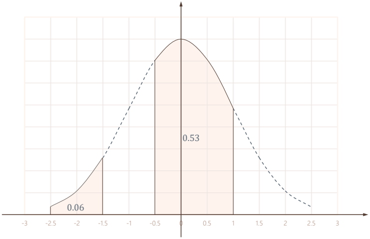


$$Q[i] \leftarrow CDF(U[i]) - CDF(L[i]),$$

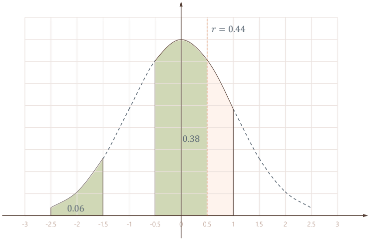
$$k \leftarrow \underset{r-\sum_{i=0}^{j} Q[i] \geq 0}{\arg\max}\ j,$$

$$\Delta r \leftarrow r - \sum_{i=0}^{k} Q[i].$$

Example



(a) Segmented Distribution



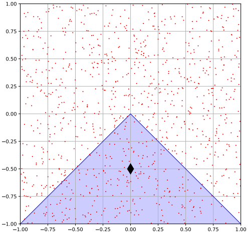
(b) Sample in Segmented Distribution

A normal distribution on fragmented satisfiable intervals ($[-2.5, -1.5] \cup [-0.5, 1]$) of a variable.

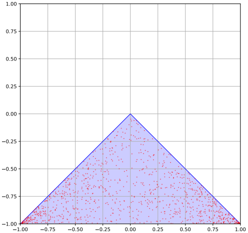
An Example

Consider a constraint,

$$x - y \geq 0 \wedge x + y \leq 0 \wedge y \geq -1.$$



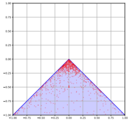
(c) Uniform Sampler



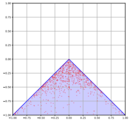
(d) Uniform Prob-MCSAT

When running 1000 times, the difference between Uniform Sampler, Z3, and Uniform Prob-MCSAT.

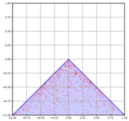
Ablation



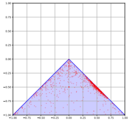
(a) $N(0, 0.1)$



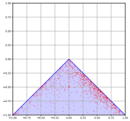
(b) $N(0, 0.25)$



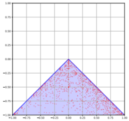
(c) $N(0, 0.5)$



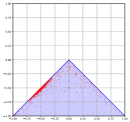
(d) $N(0.5, 0.1)$



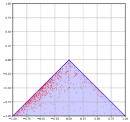
(e) $N(0.5, 0.25)$



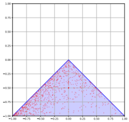
(f) $N(0.5, 0.5)$



(g) $N(-0.5, 0.1)$



(h) $N(-0.5, 0.25)$



(i) $N(-0.5, 0.5)$

Conclusion and Future Work

In this work in progress, we initially proposed PSMT and designed an algorithm, Prob-MCSAT. The visualized examples show that Prob-MCSAT can produce a clear trend in the satisfiable space. There are several future works that should be explored:

- Extending and refining the definition of PSMT;
- Identifying realistic applications for PSMT;
- Exploring distributions to enhance SMT solving speed;
- Utilizing Prob-MCSAT solutions for testing and teaching;
- Considering the bias in SMT solver solutions.

Thanks!

References I

[1] Clark Barrett, Pascal Fontaine, and Cesare Tinelli. *The SMT-LIB Standard: Version 2.6*. Tech. rep. Available at www.SMT-LIB.org. Department of Computer Science, The University of Iowa, 2017.

[2] Daniel Kroening and Ofer Strichman. "Bit Vectors". In: *Decision Procedures: An Algorithmic Point of View*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016, pp. 135–156. ISBN: 978-3-662-50497-0. DOI: 10.1007/978-3-662-50497-0_6. URL: https://doi.org/10.1007/978-3-662-50497-0_6.

References II

[3] Mateus Borges et al. "Iterative distribution-aware sampling for probabilistic symbolic execution". In: *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering*. 2015, pp. 866–877.

[4] Alessandro Abate et al. "Approximate model checking of stochastic hybrid systems". In: *European Journal of Control* 16.6 (2010), pp. 624–641.

[5] Martin Fränzle, Holger Hermanns, and Tino Teige. "Stochastic satisfiability modulo theory: A novel technique for the analysis of probabilistic hybrid systems". In: *Hybrid Systems: Computation and Control: 11th International Workshop, HSCC 2008, St. Louis, MO, USA, April 22-24, 2008. Proceedings 11*. Springer. 2008, pp. 172–186.

References III

[6] Matan Peled, Bat-Chen Rothenberg, and Shachar Itzhaky. "SMT Sampling via Model-Guided Approximation". In: *Formal Methods - 25th International Symposium, FM 2023, Lübeck, Germany, March 6-10, 2023, Proceedings*. Ed. by Marsha Chechik, Joost-Pieter Katoen, and Martin Leucker. Vol. 14000. Lecture Notes in Computer Science. Springer, 2023, pp. 74–91. DOI: 10.1007/978-3-031-27481-7_6.