Sun Yat-Sen University

SAT 2008

University of Paderborn

Theory and Applications of Satisfiability Testing

Eleventh International Conference May 12 - 15, 2008, Guangzhou, P. R. China

A new Bound for an NP-Hard Subclass of 3-SAT using Backdoors

Stephan Kottler, Michael Kaufmann, Carsten Sinz

University of Tuebingen, Germany

Eberhard Karls Universität Tübingen

Introduction

The boolean satisfiability problem (SAT) is one of the well known hard problems in theoretical computer science. Even when restricting the number of literals in each clause to a maximum of three (3-SAT), deciding satisfiability of a given instance is known to still be NP-complete. From the theoretical point of view the upper bound to solve 3-SAT could be improved steadily (see [13]). From the practical point of view we know by experience that many SAT instances evolving from real-world applications can be solved within nearly linear time. This is often due to some hidden structure that facilitates the solving process enormously. One possibility to measure this structure, namely *Backdoor Sets*, was introduced in 2003 by Williams, Gomes and Selman [16]. On the one hand it was shown that small backdoor sets are often related to real-world instances [16, 12], on the other hand minimal backdoors of randomized, hence unstructured 3-SAT instances contain from 30% to 65% of all variables [6].
We use backdoors not as a measure of structure but rather to guide an algorithm for an NP-hard subclass of 3-SAT and Mixed Horn Formulas (MHF). MHF denotes the set of all SAT instances in conjunctive normal form where each clause is either Horn or binary [10, 11].

Backdoor Sets

We use the definition of strong backdoor sets that is given in [8]. Note that there are also *weak backdoor sets* [16, 8], however, they are not relevant for this work. A backdoor is defined with respect to a class $\mathcal{C}$ of formulas that can be recognized and solved in polynomial time. A set $\mathcal{B}$ of variables $\mathcal{V}$ of a boolean formula F is a *strong backdoor set* of F with respect to $\mathcal{C}$ (strong $\mathcal{C}$-backdoor) if $F[\tau] \in \mathcal{C}$ for every truth assignment $\tau : \mathcal{B} \mapsto \{false, true\}$. $F[\tau]$ denotes the result of removing all clauses that contain a literal l with $\tau(l) = true$ and removing all literals m with $\tau(m) = false$ from F.
We particularly use a variant of strong backdoors, so-called *deletion backdoors* [9, 14]: $\mathcal{B}$ is a deletion backdoor if the formula $F - \mathcal{B}$ belongs to $\mathcal{C}$, where $F - \mathcal{B}$ denotes the result of removing all occurrences (both positive and negative) of the variables in B from the clauses of formula F. Every deletion backdoor is a strong backdoor, if class $\mathcal{C}$ is *clause-induced* ($F \in \mathcal{C} \Rightarrow F' \in \mathcal{C}$ for all $F' \subseteq F$) [9].
In this work we solely deal with the two clause-induced classes Horn and 2-SAT as base classes of backdoors.

The Class $2^\star$-CNF

Definition Let $2^\star$-CNF be the subclass of 3-SAT with the restriction that any clause C with $|C| = 3$ must only contain negative literals.

A small Example:
$(x_1 \vee x_2) \wedge (x_2 \vee \overline{x_3}) \wedge (\overline{x_4} \vee \overline{x_5}) \wedge (\overline{x_1} \vee \overline{x_2} \vee \overline{x_6}) \ldots$

$2^\star$-CNF is NP-complete.

The definition of $2^\star$-CNF as a subclass of 3-SAT $\in$ NP directly implies $2^\star$-CNF to be in NP. The NP-completeness of $2^\star$-CNF can be shown by the polynomial time reduction 3-SAT $\leq_p$ $2^\star$-CNF. For a detailed description of the proof please have a look at the proceedings version.

An alternative proof could adapt the idea to prove NP-hardness for MHF [10]. It turns out that $2^\star$-CNF $\subset$ MHF encodes the problem to decide whether the vertices of a graph can be colored with at most three different colors such that no vertices with the same color are connected by an edge.

A Backdoor Driven Algorithm

Based on the concept of backdoor sets we can specify a simple deterministic algorithm to decide satisfiability for arbitrary formulas of the class $2^\star$-CNF within the time $\mathbf{O(1.427^n * poly(n))}$:

Function `bd_solve`(F)
- $c \leftarrow 0.513$
- $C^+ \leftarrow \{(x_i \vee x_j) \in F : \ x_i, x_j \text{ positive}\}$
- Choose minimum $\mathcal{B}^+ \subseteq \mathcal{V}$, such that $\forall\, C \in C^+ \ \exists\, x_i \in \mathcal{B}^+ : x_i \in C$
- **if** $|\mathcal{B}^+| \leq c * |\mathcal{V}|$ **then**
  - **return** Solve F by using the Horn-Backdoor $\mathcal{B}^+$
- $C^- \leftarrow \{(\overline{x_h} \vee \overline{x_i} \vee \overline{x_j}) \in F : \ \overline{x_h}, \overline{x_i}, \overline{x_j} \text{ negative}\}$
- Choose minimum $\mathcal{B}^- \subseteq \mathcal{V}$, such that $\forall\, C \in C^- \ \exists\, x_i \in \mathcal{B}^- : \overline{x_i} \in C$
- **if** $|\mathcal{B}^-| \leq (1-c) * |\mathcal{V}|$ **then**
  - **return** Solve F by using the Binary-Backdoor $\mathcal{B}^-$
- **return** F Unsatisfiable

We first aim to find the smallest possible set of variables $\mathcal{B}^+$ such that every clause in C^+ (positive binary clauses) contains at least one variable of the set $\mathcal{B}^+$. Note that $\mathcal{B}^+$ constitutes a Horn-backdoor for F. Solving the instance F by using a Horn-backdoor $\mathcal{B}^+$ may in the worst case imply to examine all possible truth assignments of the variables in $\mathcal{B}^+$. More precisely this might mean that for each of the $2^{c*n} = 1.427^n$ truth assignments a Horn instance has to be solved. Hence, the complexity of this part is limited by $O(1.427^n * |F|)$.

If there is no Horn-backdoor of the desired size, we then consider the set of all clauses (C^-) consisting of three (negative) literals. We aim to find a smallest possible set of variables $\mathcal{B}^-$ such that each clause within the set C^- contains at least one variable of the set $\mathcal{B}^-$. Since any clause in $F \setminus C^-$ consists of at most two literals the set $\mathcal{B}^-$ constitutes a Binary-backdoor. To determine satisfiability of F by using the Binary-backdoor $\mathcal{B}^-$ with at most $(1-c)*n = 0.487*n$ variables may in the worst case imply to solve a 2-SAT instance for each possible truth assignment of the variables in $\mathcal{B}^-$. Since 2-SAT can be solved in linear time [1, 3] the complexity of this part can be limited by $O(1.402^n * |F|)$.

Vertex Cover

Since all clauses within C^+ are binary clauses the problem to find the smallest possible set $\mathcal{B}^+$ can be seen as a VERTEX-COVER-problem. Considering the fact that we are only interested in a set $\mathcal{B}^+$ up to a particular size, we can make use of a parameterized algorithm:
Given a graph $G = (V_G, E_G)$ the parameterized VERTEX COVER problem asks if there is a subset of vertices $C \subseteq V_G$ with k or fewer vertices such that each edge in E_G has at least one of its endpoints in C. According to [7] there are algorithms solving the parameterized VERTEX COVER in time $O(k * |V_G| + 1.29^k)$. With parameter k being $c * |\mathcal{V}| = 0.513 * n$ the complexity of this task can be limited by $O(n^2 + 1.14^n)$.

3-Hitting Set

The task to find the smallest possible set $\mathcal{B}^-$ corresponds to a 3-HITTING-SET problem (see [6]). Analogously, to determine the set $\mathcal{B}^+$, we can use a parameterized algorithm in order to solve the 3-HITTING-SET problem to detect whether there is a set $\mathcal{B}^-$ with at most $(1 - 0.513) * n$ variables:
Given a collection Q of subsets of size at most three of a finite set S and a non-negative integer k, the parameterized 3-HITTING-SET problem asks if there is a subset $S' \subseteq S$ with $|S'| \leq k$ which allows S' to contain at least one element from each subset in Q [7]. Algorithms to solve this problem have been steadily improved in the last years. Wahlström recently gave an algorithm with an upper bound $O(poly(n) * 2.0755^k)$ with a polynomial $poly(n)$ [15]. With $k := 0.487 * n$ in our case the complexity can be bounded by $O(1.427^n * poly(n))$.

How to Conclude Unsatisfiability?

When reaching the last line of the algorithm we know that there exists neither a set of variables $\mathcal{B}^+$ nor a set $\mathcal{B}^-$ with the desired size. In this case we can conclude unsatisfiability of F. Since the considered clauses within C^+ solely consist of positive literals we need to set the values of at least $|\mathcal{B}^+|$ variables to true in order to satisfy all the clauses in C^+. Analogously the size $\mathcal{B}^-$ indicates the number of variables whose values have to be set to $false$ in order to satisfy all clauses in C^-. This is impossible with $|\mathcal{B}^+|$ being greater than $c * |\mathcal{V}|$ and $|\mathcal{B}^-|$ being greater than $(1-c) * |\mathcal{V}|$ at the same time.
Note that a similar argument to prove unsatisfiability of big random 3-SAT instances has been used by Franco and Swaminathan in [5]. The authors show that an approximation algorithm for 3-HITTING-SET can determine bounds on how many variables must be set to $true$ and how many must be set to $false$.

References

[1] B. Aspvall, M. F. Plass, and R. E. Tarjan. A linear-time algorithm for testing the truth of certain quantified boolean formulas. *Inf. Proc. Lett.*, 8:121–123, 1979.

[2] T. Brüggemann and W. Kern. An improved deterministic local search algorithm for 3-sat. *Theor. Comput. Sci.*, 329(1-3):303–313, 2004.

[3] A. del Val. On 2-SAT and renamable horn. In *AAAI: 17th National Conference on Artificial Intelligence*. AAAI / MIT Press, 2000.

[4] H. Fernau. A top-down approach to search-trees: Improved algorithmics for 3-hitting set. *Electronic Colloquium on Computational Complexity*, TR04-073, 2004.

[5] J. Franco and R. Swaminathan. On good algorithms for determining unsatisfiability of propositional formulas. *Discrete Appl. Math.*, 130(2):129–138, 2003.

[6] Y. Interian. Backdoor sets for random 3-sat. In *SAT*, 2003.

[7] R. Niedermeier. *Invitation to Fixed-Parameter Algorithms*. Universität Tübingen, October 2002.

[8] N. Nishimura, P. Ragde, and S. Szeider. Detecting backdoor sets with respect to horn and binary clauses. In *SAT*, 2004.

[9] N. Nishimura, P. Ragde, and S. Szeider. Solving #SAT using vertex covers. In *SAT*, 2006.

[10] S. Porschen and E. Speckenmeyer. Worst case bounds for some NP-complete modified Horn-SAT problems. In *SAT*, 2004.

[11] S. Porschen and E. Speckenmeyer. Satisfiability of mixed Horn formulas. *Discrete Applied Mathematics*, 155(11):1408–1419, 2007.

[12] Y. Ruan, H. A. Kautz, and E. Horvitz. The backdoor key: A path to understanding problem hardness. In *AAAI*, pages 124–130, 2004.

[13] U. Schöning. A probabilistic algorithm for k-sat and constraint satisfaction problems. In *Symposium on Foundations of Computer Science*, 1999.

[14] S. Szeider. Matched formulas and backdoor sets. In *SAT*, 2007.

[15] M. Wahlström. *Algorithms, measures, and upper bounds for satisfiability and related problems*. PhD thesis, Linköping University, 2007. Dissertation no 1079.

[16] R. Williams, C. Gomes, and B. Selman. Backdoors to typical case complexity. In *IJCAI*, 2003.