A New Lower Bound on the Maximum Number of Satisfied Clauses in Max-SAT and Its Algorithmic Application

Robert Crowston, Gregory Gutin, Mark Jones, and Anders Yeo
Royal Holloway, University of London, United Kingdom
{robert,gutin,markj,anders}@cs.rhul.ac.uk

Abstract. For a formula $F$ in conjunctive normal form (CNF), let $\text{sat}(F)$ be the maximum number of clauses of $F$ that can be satisfied by a truth assignment, and let $m$ be the number of clauses in $F$. It is well-known that for every CNF formula $F$, $\text{sat}(F) \geq m/2$ and the bound is tight when $F$ consists of conflicting unit clauses ($x$) and ($\bar{x}$). Since each truth assignment satisfies exactly one clause in each pair of conflicting unit clauses, it is natural to reduce $F$ to the unit-conflict free (UCF) form. If $F$ is UCF, then Lieberherr and Specker (J. ACM 28(2):411-421, 1981) proved that $\text{sat}(F) \geq \hat{\varphi}m$, where $\hat{\varphi} = (\sqrt{5} - 1)/2$.

We introduce another reduction that transforms a UCF CNF formula $F$ into a UCF CNF formula $F'$, which has a complete matching, i.e., there is an injective map from the variables to the clauses, such that each variable maps to a clause containing that variable or its negation. The formula $F'$ is obtained from $F$ by deleting some clauses and the variables contained only in the deleted clauses. We prove that $\text{sat}(F) \geq \hat{\varphi}m + (1 - \hat{\varphi})(m - m') + n'(2 - 3\hat{\varphi})/2$, where $n'$ and $m'$ are the number of variables and clauses in $F'$, respectively. This improves the Lieberherr-Specker lower bound on $\text{sat}(F)$.

We show that our new bound has an algorithmic application by considering the following parameterized problem: given a UCF CNF formula $F$, decide whether $\text{sat}(F) \geq \hat{\varphi}m + k$, where $k$ is the parameter. This problem was introduced by Mahajan and Raman (J. Algorithms 31(2):335–354, 1999) who conjectured that the problem is fixed-parameter tractable, i.e., it can be solved in time $f(k)(nm)^{O(1)}$ for some computable function $f$ of $k$ only. We use the new bound to show that the problem is indeed fixed-parameter tractable by describing a polynomial-time algorithm that transforms any problem instance into an equivalent one with at most $\lfloor (7 + 3\sqrt{5})k \rfloor$ variables.

1 Introduction

Let $F = (V, C)$ be a CNF formula, with a set $V$ of variables and a multiset $C$ of clauses, $m = |C|$, and $\text{sat}(F)$ the maximum number of clauses that can be satisfied by a truth assignment. With a random assignment of truth values to the variables, the probability of a clause being satisfied is at least $1/2$. Thus, $\text{sat}(F) \geq m/2$ for any $F$. This bound

* Research of Gutin, Jones and Yeo was supported in part by an EPSRC grant. Research of Gutin was also supported in part by the IST Programme of the European Community, under the PASCAL 2 Network of Excellence.
is tight when $F$ consists of pairs of conflicting unit clauses ($x$) and ($\bar{x}$). Since each truth assignment satisfies exactly one clause in each pair of conflicting unit clauses, it is natural to reduce $F$ to the unit-conflict free (UCF) form by deleting all pairs of conflicting clauses. If $F$ is UCF, then Lieberherr and Specker [8] proved that $\text{sat}(F) \geq \hat{\phi}m$, where $\hat{\phi} = (\sqrt{5} - 1)/2$ (golden ratio inverse), and that for any $\epsilon > 0$ there are UCF CNF formulae $F$ for which $\text{sat}(F) < m(\hat{\phi} + \epsilon)$. Yannakakis [13] gave a short probabilistic proof that $\text{sat}(F) \geq \hat{\phi}m$ by showing that if the probability of every variable appearing in a unit clause being assigned True is $\hat{\phi}$ (here we assume that for all such variables $x$ the unit clauses are of the form $(x)$) and the probability of every other variable being assigned True is $1/2$, then the expected number of satisfied clauses is $\hat{\phi}m$.

In this paper, we introduce another reduction. We say that a UCF CNF formula $F = (V, C)$ has a complete matching if there is an injective map from the variables to the clauses, such that each variable maps to a clause containing that variable or its negation. We show that if a UCF CNF formula $F = (V, C)$ has no complete matching, then there is a subset $C'$ of clauses that can be found in polynomial time, such that $F' = (V \setminus V', C \setminus C')$ has a complete matching, where $V'$ is the set of variables of $V$ appearing only in the clauses $C'$ (in positive or negative form). In addition, we show that all clauses of $C'$ can be satisfied by assigning appropriate truth values to variables in $V'$. We also prove that for any UCF CNF $F'' = (V'', C'')$ with a complete matching, $\text{sat}(F'') \geq \hat{\phi}|C''| + (2 - 3\hat{\phi})|V''|/2$. These results imply that for a UCF CNF formula $F = (V, C)$, we have $\text{sat}(F) \geq \hat{\phi}m + (1 - \hat{\phi})(m - m') + n'(2 - 3\hat{\phi})/2$, where $m = |C|$, $m' = |C \setminus C'|$ and $n' = |V \setminus V'|$. The last inequality improves the Lieberherr-Specker lower bound on $\text{sat}(F)$.

Mahajan and Raman [10] were the first to recognize both practical and theoretical importance of parameterizing maximization problems above tight lower bounds. (We give some basic terminology on parameterized algorithms and complexity in the next section.) They considered Max-SAT parameterized above the tight lower bound $m/2$. The problem is to decide whether we can satisfy at least $m/2 + k$ clauses, where $k$ is the parameter. Mahajan and Raman proved that this parameterization of Max-SAT is fixed-parameter tractable by obtaining a problem kernel with $O(k)$ variables.

Since $\hat{\phi}m$ rather than $m/2$ is an asymptotically tighter lower bound for UCF CNF formulae, Mahajan and Raman [10] introduced the following parameterization of Max-SAT: given a UCF CNF formula $F$, decide whether $\text{sat}(F) \geq \hat{\phi}m + k$, where $k$ is the parameter. Mahajan and Raman conjectured that this parameterized problem is fixed-parameter tractable. To solve the conjecture in the affirmative, we show the existence of a proper $O(k)$-variable kernel for Max-SAT parameterized above $\hat{\phi}m$ which follows from our improvement of the Lieberherr-Specker lower bound. Here we try to optimize the number of variables rather than the number of clauses in the kernel as the number of variables is more important than the number of clauses from the computational point of view.

The rest of this paper is organized as follows. In Section 2, we give further terminology and notation. Section 3 proves the improvement of the Lieberherr-Specker lower bound on $\text{sat}(F)$ assuming correctness of the following lemma: if $F = (V, C)$ is a compact CNF formula, then $\text{sat}(F) \geq \hat{\phi}|C| + (2 - 3\hat{\phi})|V|/2$ (we give definition of a compact CNF formula in the next section). This non-trivial lemma is proved in Section 4. In