Computational Complexity of Some Restricted Instances of 3SAT

Piotr Berman [*] Marek Karpinski [†] Alexander D. Scott [‡]

Abstract

Tovey [10] showed that it is NP-hard to decide the satisfiability of 3-SAT instances in which every variable occurs four times, while every instance of 3-SAT in which each variable occurs three times is satisfiable. We explore the border between these two problems. Answering a question of Iwama and Takaki, we show that, for every fixed $k \geq 0$, there is a polynomial time algorithm to determine the satisfiability of 3-SAT instances in which k variables occur four times and the remaining variables occur three times. On the other hand, it is NP-hard to decide the satisfiability of 3-SAT instances in which all but one variable occurs three times, and the remaining variable is allowed to occur an arbitrary number of times.

1 Introduction

An instance of k-SAT is a set of clauses that are disjunctions of exactly k literals. The problem is to determine whether there is an assignment of truth

[*]Dept. of Computer Science and Engineering, Pennsylvania State University. Research done in part while visiting Dept. of Computer Science, University of Bonn. Work partially supported by NSF grant CCR-9700053 and NIH grant 9R01HG02238-12. E-mail berman@cse.psu.edu.

[†]Dept. of Computer Science, University of Bonn. Research done in part while visiting the Institute for Advanced Study, Princeton. Work partially supported by DFG grants, Max-Planck Research Prize, DIMACS, and IST grant 14036 (RAND-APX). E-mail marek@cs.uni-bonn.de.

[‡]Dept. of Mathematics, University College London, Gower Street, London WC1E 6BT. Current address: Mathematical Institute, University of Oxford, 24-29 St Giles', Oxford OX1 3LB, UK; E-mail scott@maths.ox.ac.uk. Research supported in part by EPSRC grant GR/S26323/01.

values to the variables such that all the clauses are satisfied. It is well known that 2-SAT can be solved in polynomial time, while Cook [4] showed that k-SAT is NP-hard for $k \geq 3$. This leads to the general question of exploring the boundary region between polynomial time and NP-hard satisfiability problems, by considering more or less restricted problem instances (of course, this is most interesting if P$\neq$NP).

One way to restrict instances of k-SAT is to limit the number of times a variable can occur. Tovey [10] showed that instances of 3-SAT in which every variable occurs three times are always satisfiable (this is an immediate corollary of Hall's Theorem), while it is NP-hard to decide the satisfiability of 3-SAT instances in which every variable occurs four times. Results of this type for $k > 3$ were obtained by Dubois [5] and Kratochvíl, Savický and Tuza [7], and also in [2], [3]. The approximation hardness of the corresponding maximization problem was studied in [2], [3].

The boundary between three occurrences and four occurrences of variables in 3-SAT was further examined by Iwama and Takaki [6]. Let us write $(3, 4^{(k)})$-SAT for the set of 3-SAT instances in which k variables occur four times and the remaining variables occur three times, and $(3, 4^{(k)}, n)$-SAT for the set of instances of $(3, 4^{(k)})$-SAT with n variables. Thus $(3, 4^{(0)})$-SAT is the collection of 3-SAT instances in which every variable occurs exactly three times. Note that k must be divisible by three, as the total number of literals is three times the number of clauses. Iwama and Takaki showed that every instance of $(3, 4^{(3)})$-SAT is satisfiable, while there are unsatisfiable instances of $(3, 4^{(9)})$-SAT. They further asked whether there is a constant k such that $(3, 4^{(k)})$-SAT is NP-hard. In this paper, we give a negative answer to this question (under the $P \neq NP$ assumption). We also show that it is NP-hard to decide the satisfiability of 3-SAT instances in which all but one variable occurs three times, and the remaining variable is allowed to occur an arbitrary number of times.

We remark that there are other interesting ways to explore the border between polynomial-time and NP-hard satisfiability problems. For instance, instead of looking at restrictions of 3-SAT, one can consider extensions of 2-SAT. This line of investigation has been pursued by several authors, including Monasson and Zecchina [8], Monasson, Zecchina, Kirkpatrick, Selman and Troyansky [9], Anderson [1] and Zhao, Deng, Lee and Zhu [11].

2 Results

We begin by answering the question raised by Iwama and Takaki.

Theorem 1. *Satisfiability of instances of* $(3, 4^{(k)}, n)$*-SAT can be determined in time* $2^{k/3} n^{k/3} \mathrm{poly}(n)$.

Thus for any fixed k, $(3, 4^{(k)}, n)$-SAT instances can be solved in polynomial time.

In order to prove Theorem 1, we shall rely on the fact that satisfiable instances of $(3, 4^{(k)})$-SAT have satisfying assignments with a particular structure. Let I be a satisfiable instance of $(3, 4^{(k)})$-SAT with clauses $\mathcal{C}$ and variables V, and let ϕ be a satisfying assignment. A *witness function* $w : \mathcal{C} \to V$ *for* ϕ is a function that, for each $C \in \mathcal{C}$, chooses a variable $w(C)$ which occurs (possibly negated) as a literal in C evaluating to true under ϕ. Thus $w(C)$ is a variable that "witnesses" the satisfaction of C in ϕ.

Note that if w is a witness function for some satisfying assignment and $w(C_1) = w(C_2) = x$ then x must occur as a literal with the same sign in C_1 and C_2. On the other hand, any function satisfying this condition can be used to find a satisfying assignment of I. We shall call such a function *consistent*

Lemma 1. *If* $I = (\mathcal{C}, V)$ *is a satisfiable instance of* $(3, 4^{(k)})$*-SAT then there is a satisfying assignment* ϕ *with a surjective witness function* $w : \mathcal{C} \to V$.

Proof. Let (ϕ, w) be a satisfying assignment and a witness function chosen such that the size of the image of w is maximal (over all such pairs). If w is surjective then we are done. Otherwise, let us consider the bipartite graph G with vertex classes $\mathcal{C}$ and V and an edge from $C \in \mathcal{C}$ to $v \in V$ if and only if v occurs (with either sign) in C. We shall say that an edge vC is *used* if $w(C) = v$; otherwise vC is *unused.*

Let $U = \{v \in V : |w^{-1}(v)| \geq 2\}$ be the set of variables that are used as witnesses by more than one clause (note that, since w is not surjective, $|U| > 0$). Let $s := \sum_{v \in U} |w^{-1}(v)| \geq 2|U|$ be the number of used edges incident with U. An *alternating path* in V is a path $v_1 C_1 v_2 C_2 \cdots v_k$ or $v_1 C_1 v_2 \cdots C_k$ for some $k \geq 1$ such that $v_1 \in U$, all edges $v_i C_i$ are used and all edges $C_i v_{i+1}$ are unused. We shall show that if w is not surjective then we can use a suitable alternating path to construct a new witness function w' with a larger image.

Let us say that a vertex of G is *reachable* if it belongs to any alternating path. Let U' be the set of reachable vertices in $V \setminus U$, and let $\mathcal{B}$ be the set of reachable vertices in $\mathcal{C}$. Note that if $u \in U \cup U'$ and the edge uC is used then it is easy to find an alternating path that contains C (take an alternating path as far as u, and if C has not already been visited then extend the path with the edge uC); it follows that $C \in \mathcal{B}$.

There are s used edges incident with U and $|U'|$ used edges incident with U'. Since every clause is incident with exactly one used edge, we have

$$|\mathcal{B}| \geq s + |U'|. \tag{1}$$

On the other hand, each clause is incident with at least 2 unused edges, and so there are at least $2|\mathcal{B}|$ unused edges incident with $\mathcal{B}$. It is easy to check that if $C \in \mathcal{B}$ and vC is unused then $v \in U \cup U'$ (we argue as before: take an alternating path as far as C and add the edge Cv if necessary). Thus there are at least $2|\mathcal{B}|$ unused edges incident with $U \cup U'$.

Now suppose that every vertex of $U \cup U'$ is in the image of w. Since the vertices of $U \cup U'$ are incident with $s + |U'|$ used edges, and at most k vertices have degree 4, it follows that the number of unused edges incident with $U \cup U'$ is at most

$$3(|U| + |U'|) + k - (s + |U'|) = 3|U| + 2|U'| + k - s.$$

Since this is at least $2|\mathcal{B}|$, it follows from (1) that

$$2(s + |U'|) \leq 3|U| + 2|U'| + k - s,$$

and so

$$s \leq |U| + k/3. \tag{2}$$

Now let $U'' = \mathrm{Im}(w) \setminus U$. Since $U \subset \mathrm{Im}(w)$ and $\mathrm{Im}(w) \neq V$, we have

$$|U| + |U''| \leq |V| - 1. \tag{3}$$

Since every vertex of U'' is incident with exactly one used edge, and there are $|\mathcal{C}| = |V| + k/3$ used edges in total, we have

$$s + |U''| = |V| + k/3. \tag{4}$$

Therefore, by (3) and (4),

$$s = |V| + k/3 - |U''| \geq |U| + k/3 + 1,$$

which contradicts (2).

It follows that there is some vertex $u' \in U'$ that is not contained in the image of w. Let $P = v_1 C_1 v_2 \cdots v_k C_k u'$ be a shortest alternating path from U to u': note that v_1 is in U, but $v_i \notin U$ for $i > 1$. Exchanging used and unused edges in P, we obtain a consistent witness function w' with a larger image than w. This contradicts the maximality of the image of w, and we therefore deduce that w is surjective. □

We can now prove Theorem 1.

Proof of Theorem 1. If I is satisfiable then there are a satisfying assignment ϕ and a surjective witness function w for ϕ. Since there are $n + k/3$ clauses, and n variables, it follows that at most $k/3$ variables are covered more than once by w. We can therefore search for such a w by explicitly examining every set of $\lfloor k/3 \rfloor$ variables and every assignment of those variables, and then checking for a matching from the remaining unsatisfied clauses to the remaining unassigned variables. □

We have shown that for any constant, $(3, 4^{(k)})$-SAT instances can be solved in polynomial time. What if we allow more than four occurrences of some variables? The following theorem shows that if we allow an unbounded number of occurrences of even one variable, then the problem becomes NP-hard.

Theorem 2. *The restriction of 3-SAT to the set of instances in which all but one variable occur exactly three times is NP-hard.*

We shall prove Theorem 2 by reduction from another problem. We define $(O3, L \leq 3)$-SAT to be the set of instances of satisfiability in which every variable occurs three times and every clause has length at most 3. This problem is NP-hard, as was shown by Tovey [10]. We give a proof for completeness.

Theorem 3. *Determining the satisfiability of instances of* $(O3, L \leq 3)$*-SAT is NP-hard.*

Proof. We give a reduction from 3-SAT due to Tovey [10]. Given an instance of 3-SAT, we run through its variables in turn, modifying the instance as follows. If a variable x occurs at most three times we do nothing. If x occurs $d > 3$ times, we introduce new variables $x_1, \ldots, x_d$ and 2-clauses $x_1 \vee \neg x_2, \ldots, x_{d-1} \vee \neg x_d, x_d \vee \neg x_1$. We then replace the d occurrences of x

by $x_1, \ldots, x_d$ in turn and remove x. It is easily checked that this preserves satisfiability/unsatisfiability, and when we have dealt with all the variables we have an equivalent instance of $(O3, L \leq 3)$-SAT. $\square$

Note that a typical variable in the instance constructed above will belong to two clauses of length 2. We sketch a slightly more complicated construction that allows us to insist that every variable belongs to at most *one* clause of length two. Given an instance of 3-SAT, we first perform the construction in the proof above to obtain an instance I of $(O3, L \leq 3)$-SAT. We introduce new variables $a_i, b_i, \ldots, i = 1, 2, 3$. We build a new instance as follows: take the clause $\neg a_1 \vee \neg a_2 \vee \neg a_3$, and for each i add the following chain of clauses:

$$
\begin{array}{cccccc}
a_i & b_i & c_i & & & \\
a_i & \neg b_i & & & & \\
 & b_i & \neg c_i & d_i & & \\
 & & & \neg d_i & e_i & \\
 & & & \neg d_i & \neg e_i & f_i \\
 & & & & & \ldots
\end{array}
$$

For each i, take a copy I_i of I (on a new set of variables), and extend all of its 2-clauses to 3-clauses by adding the negation of some variable that occurs twice in the corresponding chain (ie one of $\{\neg c_i, \neg e_i, \ldots\}$). This gives the required instance. Note that in any satisfying assignment of the resulting instance, one of the variables a_i must be false, and so the variables $c_i, e_i, \ldots$ must be true, which means that I_i (and hence I) is satisfiable. On the other hand, if I is satisfiable, then the new instance is easily seen to be satisfiable.

We now return to the proof of Theorem 2.

Proof. We give a reduction from $(O3, L \leq 3)$-SAT. Let I be an instance of $(O3, L \leq 3)$-SAT. We take two copies I_1, I_2 of I (on disjoint sets of variables) and a new variable x. We construct an instance of 3-SAT by adding x to every 2-clause in I_1 and $\neg x$ to every 2-clause in I_2. The resulting instance is clearly equivalent to I. $\square$

3 Conclusion

We have given a polynomial time algorithm for instances of 3-SAT in which a constant number of variables occur four times and the remainder occur three times. What happens if we allow the number of variables occurring four times to grow slowly?

Problem. *Is there a function $f(n)$ that tends to ∞ as $n \to \infty$ such that the satisfiability of instances of $(3, 4^{(f(n))}, n)$-SAT is solvable in polynomial time? (We assume here that $f(n)$ is a multiple of three, to exclude trivial cases.)*

We conjecture that for sufficiently slow-growing f the problem can be solved in polynomial time.

References

[1] W. Anderson, *Solving problems in finite time*, Nature **400** (1999), 115–116

[2] P. Berman, M. Karpinski and A. D. Scott, *Approximation Hardness and Satisfiability of Bounded Occurrence Instances of SAT*, ECCC TR03-022 (2003), *submitted for publication*

[3] P. Berman, M. Karpinski and A. D. Scott, *Approximation Hardness of Short Symmetric Instances of MAX-3SAT*. ECCC TR03-049 (2003).

[4] S. C. Cook, *The complexity of theorem-proving procedures*, Proc. 3rd ACM STOC, 151-158, 1971

[5] O. Dubois, *On the r, s-SAT satisfiability problem and a conjecture of Tovey*, Discrete Applied Math. **26** (1990), 51–60

[6] K. Iwama and K. Takaki, *Satisfiability of 3CNF formulas with small clause/variable ratio*, *in* Satisfiability Problem: Theory and Applications, Dingzhu Du, Jun Gu, and Panos M. Pardalos, eds, American Mathematical Society, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, **35** (1997), 315–333

[7] J. Kratochvíl, P. Savický and Z. Tuza, *One more occurence of variable makes satisfiability jump from trivial to NP-complete*, SIAM J. Comput. **8** (1993), 203–210

[8] R. Monasson and R. Zecchina, *Tricritical points in random combinatorics: the $2+p$ SAT case*, J. Phys. A **31** (1998), 9209–9217

[9] R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman and L. Troyansky, *Determining computational complexity from characteristic 'phase transitions'*, Nature **400** (1999), 133–137

[10] C. A. Tovey, *A simplified satisfiability problem*, Discr. Appl. Math. **8** (1984), 85–89

[11] Y. Zhao, X. Deng, C. H. Lee and H. Zhu, $(2 + f(n))$-*SAT and its properties*, Discrete Applied Mathematics **136** (2004), 3–11