The \textit{k-Satisfiability} problem remains NP-complete for dense families

Ingo Schiermeyer

\textit{Lehrstuhl C für Mathematik, Technische Hochschule Aachen, Templergraben 55, W-52056 Aachen, Germany}

Received 12 July 1991
Revised 23 December 1991

Abstract

We consider the \textit{k-Satisfiability} problem (\textit{k-SAT}): Given a family \( F \) of \( n \) clauses \( c_1, \ldots, c_n \) in conjunctive normal form, each consisting of \( k \) literals corresponding to \( k \) different variables of a set of \( r \geq k \geq 1 \) boolean variables, is \( F \) satisfiable? By \( \textit{k-SAT}(< n_0) \) we denote the \( k \)-\textit{SAT} problem restricted to families with \( n > n_0(r) \) clauses. We prove that for each \( k > 3 \) and each integer \( l \geq 4 \) such that \( r > l k^3 \), the \( \textit{k-SAT}(< (2^l - 1/4/l)) \) problem is NP-complete.

1. Introduction

Let \( V = \{v_1, v_2, \ldots, v_r\} \) be a set of boolean variables. A \textit{truth assignment} for \( V \) is a function \( t: V \rightarrow \{\text{TRUE, FALSE}\} \). If \( t(v) = \text{TRUE} \) we say that \( v \) is 'true' under \( t \); if \( t(v) = \text{FALSE} \) we say that \( v \) is 'false'. If \( v \) is a variable in \( V \), then \( v \) and \( \bar{v} \) are \textit{literals} over \( V \). The \textit{positive} literal \( v \) is true under \( t \) if and only if the variable \( v \) is true under \( t \); the \textit{negative} literal \( \bar{v} \) is true if and only if the variable \( v \) is false.

A \textit{k-clause} over \( V \) is a set of \( k \) literals corresponding to \( k \) different variables over \( V \). It represents the disjunction of those literals and is \textit{satisfied} by a truth assignment if and only if at least one of its members is true under that assignment. A family \( F \) of \( k \)-\textit{clauses} over \( V \) is \textit{satisfiable} if and only if there exists some truth assignment for \( V \) that simultaneously satisfies all the clauses in \( F \). Such a truth assignment is called a \textit{satisfying truth assignment} for \( F \). The \textit{k-Satisfiability} problem (\textit{k-SAT}) is specified as follows.

Correspondence to: Ingo Schiermeyer, Lehrstuhl C für Mathematik, Technische Hochschule Aachen, Templergraben 55, W-52056 Aachen, Germany.
The k-SAT problem

Instance: A set $V$ of $r \geq k \geq 1$ variables and a family $F$ of $k$-clauses over $V$.

Question: Is there a satisfying truth assignment for $F$?

The $k$-SAT problem has been shown to be NP-complete for all $k \geq 3$ and solvable in polynomial time for $k = 1$ and $k = 2$ (cf. [1]).

Kratochvíl et al. [2] called a family $F$ a $(k, s)$-formula if every variable occurs in at most $s$ $k$-clauses. By $(k, s)$-SAT they denote the $k$-SATISFIABILITY problem restricted to $(k, s)$-formulas and obtained the following results.

Theorem 1.1 (Kratochvíl et al. [2]). For every $k \geq 3$, there is an integer $f(k)$ such that

(i) every $(k, s)$-formula with $s \leq f(k)$ has a satisfying truth assignment;
(ii) $(k, s)$-SAT is NP-complete for every $s > f(k) + 1$.

Note that (ii) remains valid if each variable is supposed to occur in exactly $s$ $k$-clauses. This theorem extends a result of Tovey [5], who proved that $f(3) = 3$. The bounds in [2] for $f(k)$ are $\left\lceil \frac{2^k}{e^k} \right\rceil \leq f(k) \leq 2^{k-1} - 2^{k-4} - 1$ for every $k \geq 4$. Hence, Theorem 1.1 may be considered as a result for sparse families.

In the following we restrict ourselves to those instances of families having $n > n_0(k)$ $k$-clauses over $V$, where $n_0 = n_0(r)$ is an integer depending on $r$ with $0 \leq n_0 < \binom{r}{k} 2^k$. We then define the $k$-SATISFIABILITY problem for families with $n > n_0$ $k$-clauses as follows.

The $k$-SAT($> n_0$) problem

Instance: A set $V$ of $r \geq k \geq 1$ variables and a family $F$ of $n > n_0$ $k$-clauses over $V$.

Question: Is there a satisfying truth assignment for $F$?

Clearly, $k$-SAT($> n_0$) is solvable in polynomial time for $k = 1$ and $k = 2$ since this is the case for $k$-SAT.

Our work was motivated by the following results.

Theorem 1.2 (Schiermeyer [3]). No instance of the $k$-SAT($> \binom{r}{k} (2^k - 1)$) problem has any satisfying truth assignment.

Theorem 1.3 (Schiermeyer [4]). Any instance of the $k$-SAT($> \binom{r}{k} (2^k - 1 - k/r)$) problem has at most one satisfying truth assignment.

Furthermore, each instance of Theorems 1.2 and 1.3 can be decided in polynomial time and a satisfying truth assignment can be determined in polynomial time provided there exists one (in the case of Theorem 1.3).

Our main result, which will be proved in the next section, is the following.
Theorem 1.4. For each \( k \geq 3 \) and each integer \( l \geq 4 \) the \( k\text{-SAT}(> (\frac{k}{2}) (2^{k-1} - 4/l)) \) problem is NP-complete for \( r \geq lk^2 \).

Note that the number of \( k \)-clauses of any family does not exceed \( (\frac{k}{2})2^k \). Hence Theorem 1.4 may be considered as a result for dense families.

2. Proof of Theorem 1.4

We shall show that \( k\text{-SAT} \) can be transformed to \( k\text{-SAT}(> (\frac{k}{2}) (2^{k-1} - 4/l)) \) for each \( k \geq 3 \) and each \( l \geq 4 \) such that \( r \geq lk^2 \). Let \( F \) be a family of \( n \) \( k \)-clauses over a set \( V = \{v_1, v_2, \ldots, v_r\} \) of \( r \) variables. We now extend \( F \) to a family \( F' \) of \( k \)-clauses over a set \( V' = V \cup \{v_{r+1}, v_{r+2}, \ldots, v_{rlk}\} \) of \( r' = rlk \) variables as follows.

For each \( i \) with \( 1 \leq i \leq k \) we add all possible \( k \)-clauses over \( V' \) having exactly \( i \) literals over \( W = \{v_{r+1}, v_{r+2}, \ldots, v_{rlk}\} \), not all being negative. Thus, putting \( m = kl - 1 \), we add \( \sum_{i=1}^{k} \binom{rm}{i} (2^{i-1}) (\frac{r}{k-i})^{2k-i} \) \( k \)-clauses to \( F \). Equivalently, \( F' \) has \( \sum_{i=0}^{k} \binom{rm}{i} (2^{i-1}) (\frac{r}{k-i})^{2k-i} \) \( k \)-clauses. For \( 0 \leq i \leq k \) let \( t(i) := \binom{rm}{i} (\frac{r}{k-i})^{2k-i} \). Then for \( 0 \leq i \leq k-1 \),

\[
\frac{t(i+1)}{t(i)} = \frac{(rm-i)(k-i)}{(i+1)(r-k+i+1)2^{k-i}}
\]

is decreasing with increasing \( i \). Furthermore,

\[
\frac{t(k)}{t(k-1)} = \frac{(rm-k+1)}{2kr} > \frac{r(m-1)}{2kr} = \frac{m-1}{2k} = \frac{kl-2}{2k} > 1.
\]

Thus, \( t(1) < t(2) < \cdots < t(k) \) and we obtain

\[
\sum_{i=0}^{k} t(i) < \binom{rm}{k} \sum_{i=0}^{k} \left( \frac{2^{i-1}}{m-1} \right)^i = \binom{rm}{k} \left( \frac{1-(2k/(m-1))^{k+1}}{1-2k/(m-1)} \right) < \binom{rm}{k} \frac{m-1}{m-1-2k},
\]

since

\[
\binom{rm}{k} \left( \frac{m}{m+1} \right)^k > \binom{rm}{k} \quad \text{and} \quad \left( \frac{m+1}{m} \right)^k \geq \frac{m+k}{m}.
\]

Now let

\[
\frac{m}{k+m} \frac{m-1}{m-1-2k} - \frac{kl-2}{kl+k-1} \frac{kl-2}{kl-2-2k} < g(l).
\]
Then
\[ g(l) = 1 + \frac{k^2 l + 2k^2}{k^2 l^2 - k^2 l - 2k^2 - 3kl + 2} < 1 + \frac{kl + 2k}{kl^2 - kl - 2k - 3l} \]
\[ \leq 1 + \frac{kl + kl/2}{l(kl - k - 3) - kl/2} = 1 + \frac{3k}{2kl - 3k - 6} < 1 + 4/l \quad \text{for } l > 4. \]

For \( l = 4 \) the computation of \( g(4) \) also shows that \( g(4) < 2 = 1 + 4/l \) for all \( k \geq 3 \). Therefore, \( n' > (r^{(m+1)})(2^k - 1 - 4/l) \) for each \( k \geq 3 \) and each \( l \geq 4 \) such that \( r(m+1) = rkl \geq lk^2 \).

We now show that there exists a satisfying truth assignment for a family \( F \) of \( k\text{-SAT} \) if and only if there exists a satisfying truth assignment for the corresponding family \( F' \) of \( k\text{-SAT(>n)} \).

Clearly, any satisfying truth assignment for a family \( F' \) of \( k\text{-SAT(>n)} \) is also a satisfying truth assignment for the corresponding family \( F \) of \( k\text{-SAT} \), since \( F \) is a subfamily of \( F' \) by the construction. Conversely, if there is a satisfying truth assignment for a family \( F \) of \( k\text{-SAT} \), we then assign value TRUE to all \( r \) variables of \( W = \{v_{r+1}, v_{r+2}, \ldots, v_{r(m+1)}\} \). Since \( F \) is a subfamily of \( F' \), all clauses of \( F \) are satisfied. Any clause of \( F' - F \) contains literals of \( W \) for some \( 1 \leq i \leq k \). By the construction, at least one of these \( i \) literals is positive and has therefore value TRUE. Thus, all clauses of \( F' - F \) are also satisfied and we have a truth assignment for the family \( F' \) of \( k\text{-SAT(>n)} \).

Finally, we show that our transformation can be computed in polynomial time. For each \( k \geq 3 \) and each \( l \geq 4 \) we have \( n' \leq (r^k l)2^k = O(r^k) \) \( k \)-clauses. The number of symbols required to describe an individual literal need only add an additional \( \log r \) factor and thus our transformation is bounded by a polynomial function of \( r \). This completes the proof. \( \square \)

Acknowledgment

The author would like to thank the two referees for their valuable suggestions concerning the presentation.

References