Unsatisfiable CNF-formulas

Heidi Gebauer *

October 10, 2008

Abstract

A Boolean formula in a conjunctive normal form is called a \((k, s)\)-formula if every clause contains exactly \(k\) variables and every variable occurs in at most \(s\) clauses. We show that there are unsatisfiable \((k, 4 \cdot 2^k)\)-CNF formulas.

1 A better bound for unsatisfiable formulas

Theorem 1.1 For every sufficiently large \(k\) there is an unsatisfiable \((k, 4 \cdot 2^k)\)-CNF.

Note that due to Kratochvíl, Savický and Tuza [2] every \((k, 2^k)\)-CNF is satisfiable. So our result shows that this bound is tight up to a factor \(4e\).

Proof: We consider the class \(C\) of hypergraphs \(G\) whose vertices can be arranged in a binary tree \(T_G\) such that every hyperedge of \(G\) is a path of \(T_G\). For positive integers \(k, s \geq 1\) we denote by a \((k, s)\)-tree a \(k\)-uniform hypergraph \(G \in C\) such that

- every full branch of \(T_G\) contains a hyperedge of \(G\) and
- every vertex of \(T_G\) belongs to at most \(s\) hyperedges of \(G\)

When there is no danger of confusion we write \(G\) for \(T_G\). The following lemma is the core of our proof.

Lemma 1.2 For every sufficiently large \(k\) there is a \((k, 2 \cdot 2^k)\)-tree \(G\).

We first show that Lemma 1.2 implies Theorem 1.1. Suppose that there is a \((k, 2 \cdot 2^k)\)-tree \(G\) and let \(G'\) be a copy of \(G\). Let \(H\) be the hypergraph obtained by generating a new root \(v\) and attaching \(G\) as a left subtree and \(G'\) as a right subtree. Note that \(H\) is a \((k, 2 \cdot 2^k)\)-tree as well.

Let \((x_1, x'_1), (x_2, x'_2), \ldots, (x_r, x'_r)\) denote the pairs of siblings of \(H\). We set \(x'_i := \bar{x}_i\) for every \(i, i = 1, \ldots, r\) (i.e. each non-root vertex represents a literal \(x \in \{x_1, \bar{x}_1, x_2, \bar{x}_2, \ldots, x_r, \bar{x}_r\}\)). Let \(E(H)\)

*Institute of Theoretical Computer Science, ETH Zurich, CH-8092 Switzerland. Email: gebauerh@inf.ethz.ch.
denote the set of hyperedges of $\mathcal{H}$. Then for every hyperedge $\{y_1, y_2, \ldots, y_n\} \in E(\mathcal{H})$ we form the clause $C_{\{y_1, y_2, \ldots, y_n\}} = (y_1 \lor y_2 \lor \ldots \lor y_n)$ and set $\mathcal{F} := \bigwedge_{e \in E(\mathcal{H})} C_e$.

Note that every variable $x_i$ of $\mathcal{F}$ occurs in at most $2 \cdot \Delta(\mathcal{F})$ clauses with $\Delta(\mathcal{F})$ denoting the maximum degree a variable in $\mathcal{F}$. Indeed, the number of occurrences of the variable $x_i$ is bounded by the number of occurrences of the literal $x_i$ plus the number of occurrences of the literal $\overline{x_i}$, which is at most $2\Delta$. So $\mathcal{F}$ is a $(k, 2 \cdot \frac{2^k}{k})$-CNF.

It remains to show that $\mathcal{F}$ is not satisfiable. Let $\alpha$ be an assignment to $\{x_1, \ldots, x_r\}$.

Observation 1.3 Note that there is (at least) one full branch $b_{\text{full}}$ of $\mathcal{H}$ such that all literals along $b_{\text{full}}$ are set to FALSE by $\alpha$.

By assumption $b_{\text{full}}$ contains a hyperedge $h$. But $\alpha$ does not satisfy the clause $C_h$, implying that $\alpha$ does not satisfy $\mathcal{F}$. Since $\alpha$ was chosen arbitrarily, $\mathcal{F}$ is not satisfiable. □

It remains to prove our key lemma.

Proof of Lemma 1.2: We need some notation first. The vertex set and the hyperedge set of a hypergraph $\mathcal{H}$ are denoted by $V(\mathcal{H})$ and $E(\mathcal{H})$, respectively. By a slight abuse of notation we consider $E(\mathcal{H})$ as a multiset, i.e. every hyperedge $e$ can have a multiplicity greater than 1. By a bottom hyperedge of a tree $T_\mathcal{H}$ we denote a hyperedge covering a leaf of $T_\mathcal{H}$. Let $d = 2^k$. For simplicity we assume that $k$ is a power of 2, implying that $d$ is power of 2 as well.

To construct the required hypergraph $\mathcal{G}$ we establish first a (not necessarily $k$-uniform) hypergraph $\mathcal{H}$ and then successively modify its hyperedges and $T_\mathcal{H}$. The following lemma is about the first step.

Lemma 1.4 There is a hypergraph $\mathcal{H} \in C$ with maximum degree $2d$ such that every full branch of $T_\mathcal{H}$ has $2^i$ bottom hyperedges of size $\log d + 1 - i$ for every $i$ with $0 \leq i \leq \log d$.

Proof of Lemma 1.4: Let $T$ be a binary tree with $\log d + 1$ levels. In order to construct the desired hypergraph $\mathcal{H}$ we proceed for each vertex $v$ of $T$ as follows. For each leaf descendant $w$ of $v$ we let the path from $v$ to $w$ be a hyperedge of multiplicity $2^{l(v)}$ where $l(v)$ denotes the level of $v$. Figure 1 shows an illustration. The construction yields that each full branch of $T_\mathcal{H}$ has $2^i$ bottom hyperedges.

![Figure 1: An illustration of $\mathcal{H}$ for $d = 4$. The hyperedge $\{a, b, c\}$ has multiplicity 1, $\{b, c\}$ has multiplicity 2 and $\{c\}$ has multiplicity 4.](image-url)
of size $\log d + 1 - i$ for every $i$ with $0 \leq i \leq \log d$. So it remains to show that $d(v) \leq 2d$ for every vertex of $v \in V(T)$. Note that every vertex $v$ has $2^{\log d - l(v)}$ leaf descendants in $T_H$, implying that $v$ is the start node of $2^{\log d - l(v)} \cdot 2^{l(v)} \leq d$ hyperedges. So the degree of the root is at most $d \leq 2d$. We then apply induction. Suppose that $d(u) \leq 2d$ for all nodes $u$ with $l(u) \leq i - 1$ for some $i$ with $1 \leq i \leq \log d$ and let $v$ be a vertex on level $i$. By construction exactly half of the hyperedges containing the ancestor of $v$ also contain $v$ itself. Hence $v$ occurs in at most $1/2 \cdot 2d = d$ hyperedges as non-start node. Together with the fact that $v$ is the start node of at most $d$ hyperedges this implies that $d(v) \leq d + d \leq 2d$. □

The next lemma deals with the second step of the construction of the required hypergraph $G$.

**Lemma 1.5** There is a hypergraph $H' \in C$ with maximum degree $2d$ such that each full branch of $T_{H'}$ has $2^i$ bottom hyperedges of size $\log d + 1 - i + \lfloor \log \log d \rfloor$ for some $i$ with $0 \leq i \leq \log d$.

**Proof:** Let $H \in C$ be a hypergraph with maximum degree $2d$ such that every leaf $u$ of $T_H$ is the end node of a set $S_i(u)$ of $2^i$ hyperedges of size $\log d + 1 - i$ for every $i$ with $0 \leq i \leq \log d$. (Lemma 1.4 guarantees the existence of $H$.) To each leaf $u$ of $T_H$ we then attach a binary tree $T_u'$ of height $\lfloor \log \log d \rfloor$ in such a way that $u$ is the root of $T_u'$. Let $v_0, \ldots, v_{2^{\lfloor \log \log d \rfloor} - 1}$ denote the leaves of $T_u'$. For every $i$ with $0 \leq i \leq 2^{\lfloor \log \log d \rfloor} - 1$ we then augment every hyperedge of $S_i(u)$ with the set of vertices different from $u$ along the full branch of $T_u'$ ending at $v_i$.

After repeating this procedure for every leaf $u$ of $T_H$ we get the desired hypergraph $H'$. It remains to show that every vertex in $H'$ has degree at most $2d$. To this end note first that during our construction the vertices of $H$ did not change their degree. Secondly, let $u$ be a leaf of $T_H$. By assumption $u$ has degree at most $2d$ and by construction $d(v) \leq d(u)$ for all vertices $v \in V(H') \setminus V(H)$, which completes our proof. □

**Lemma 1.6** There is a hypergraph $H'' \in C$ with maximum degree $2d$ such that every full branch of $T_{H''}$ has one bottom hyperedge of size $\log d + 1 + \lfloor \log \log d \rfloor$.

Note that due to our choice of $d$, Lemma 1.6 directly implies Lemma 1.2 □

**Proof of Lemma 1.6:** By Lemma 1.5 there is a hypergraph $H' \in C$ with maximum degree $2d$ such that each full branch of $T_{H'}$ has $2^i$ bottom hyperedges of size $\log d + 1 - i + \lfloor \log \log d \rfloor$ for some $i$ with $0 \leq i \leq \log d$. For every leaf $u$ of $T_{H'}$ we proceed as follows. Let $e_1, \ldots, e_{2^i}$ denote the bottom hyperedges of $H'$ ending at $u$. We then attach a binary tree $T''$ of height $i$ to $u$ in such a way that $u$ is the root of $T''$. Let $p_1, \ldots, p_{2^i}$ denote the full branches of $T''$. We finally augment $e_j$ with the vertices along $p_j$, for $j = 1 \ldots 2^i$.

After repeating this procedure for every leaf $u$ of $T_{H'}$ we get the resulting graph $H''$. By construction every full path of $T_{H''}$ has one bottom hyperedge of size $\log d + 1 + \lfloor \log \log d \rfloor$. A similar argument as in the proof of Lemma 1.5 shows that the maximum degree of $H''$ is at most $2d$. □
References

[1] N. Alon and J.H. Spencer, The Probabilistic Method *J. John Wiley & Sons* (2002).

[2] J. Kratochvíl, P. Savický and Z. Tuza, One more occurrence of variables makes satisfiability jump from trivial to NP-complete *SIAM Journal of Computing* 22(1) **22**(1) (1993) 203210