Partial Satisfaction of k-Satisfiable Formulas

Claudia Käppeli and Dominik Scheder

Institut für Theoretische Informatik
ETH Zürich
dominik.scheder@inf.ethz.ch

Abstract. A CNF formula is called *k-satisfiable*, if every subformula containing at most k clauses is satisfiable. What is the largest ratio r such that for any k-satisfiable formula F, there is an assignment satisfying at least a fraction r of F? This question can be asked for formulas with weighted and formulas with unweighted clauses. For weighted k-satisfiable formulas, denote that ratio by r_k. For unweighted formulas, denote it by s_k. The numbers r_k have already been studied, but little has been known for s_k. We show that s_k and r_k differ for $k = 2, 3$. For $k = 2$, we show that $s_2 = \frac{2}{3}$, which is larger than $r_2 = (\sqrt{5}-1)/2 \approx 0.619$, the inverse of the golden ratio. Further, we show that $r_3 = \frac{2}{3} < 21/31 \leq s_3 \leq 7/10$.

1 Introduction

A CNF formula F over a finite variable set V is a set of clauses; a clause C is a set of literals; a literal is either a variable $x \in V$ or its negation $\bar{x}$. We require that no clause contains a variable and its negation simultaneously. If l is a literal, then $\mathrm{vbl}(l)$ is its variable, i.e., $\mathrm{vbl}(x) = \mathrm{vbl}(\bar{x}) = x$ for $x \in V$. Let F be a CNF formula over variables set V and $\mu : F \to \mathbb{R}^+$ a clause weight function. The function μ extends to subsets $G \subseteq F$ by $\mu(G) := \sum_{C \in G} \mu(C)$. For a truth assignment $\alpha : V \to \{0, 1\}$, let $\mu^\alpha(F)$ denote $\mu(\{C \in F | \alpha \text{ satisfies } C\})$. Finally, let $\mu^*(F) = \max_\alpha \mu^\alpha(F)$.

Definition 1.1. *A CNF formula F is called* k-satisfiable *if any subformula $G \subseteq F$ with $|G| \leq k$ is satisfiable.*

Definition 1.2. *Define r_k as follows:*

$$r_k := \inf \left\{ \frac{\mu^*(F)}{\mu(F)} \mid F \text{ is } k\text{-satisfiable}, \ \mu : F \to \mathbb{R}^+ \right\} , \tag{1}$$

and for formulas with unit clause weights, we define

$$s_k := \inf \left\{ \frac{\mu^*(F)}{\mu(F)} \mid F \text{ is } k\text{-satisfiable}, \ \mu : F \to \mathbb{R}^+, \ \mu(C) = 1 \ \forall \ C \in F \right\} . \tag{2}$$

2 Previous Results

k-satisfiable formulas and the numbers r_k were first introduced and studied by Lieberherr and Specker [1]. The asymptotic behavior of the r_k was determined by Trevisan [2] and later Král [3].

Theorem 2.1 (Lieberherr and Specker [1, 4]). $r_1 = \frac{1}{2}$, $r_2 = \frac{\sqrt{5}-1}{2}$ *and* $r_3 \geq \frac{2}{3}$.

Proof. We give a proof by Yannakakis [5], which is simpler than the original proof of Lieberherr and Specker. It is not difficult to see that $r_1 = 1/2$. For r_2 and r_3, fix some p, $0 \leq p \leq 1$, and choose a random truth assignment α with

$$\Pr(\alpha(x) = 1) = \begin{cases} p, & \text{if } \{x\} \in F, \\ 1-p, & \text{if } \{\bar{x}\} \in F, \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

independently for each variable. Note that since $\{\{x\}, \{\bar{x}\}\} \not\subseteq F$ for any variable x, the probability distribution is well-defined. If F is 2-satisfiable, choose $p = \frac{\sqrt{5}-1}{2}$ and observe that each clause is satisfied with probability at least p. If F is 3-satisfiable, choose $p = \frac{2}{3}$, and again, each clause is satisfied with probability at least $\frac{2}{3}$. The claim that $r_2 \geq \frac{\sqrt{5}-1}{2}$ and $r_3 \geq \frac{2}{3}$ follows from linearity of expectation. We omit details. $\square$

Lieberherr and Specker [1] also defined a family $(F_i)_{i \in \mathbb{N}}$ of 2-satisfiable formulas for which $\frac{\sqrt{5}-1}{2}$ is asymptotically tight. They conjectured that $\lim_{k\to\infty} r_k = 1$, which was disproven by Huang and Lieberherr [6], who showed that $\lim_{k\to\infty} r_k \leq 3/4$. Trevisan [2], showed that $\lim_{k\to\infty} r_k = 3/4$. Trevisan [2] later prvoed that $\lim_{k\to\infty} r_k = 3/4$ and gives lower bounds on the r_k. These were later improved by Král [3], who in addition determined the exact value of r_4.

3 Our Results

We show that $r_3 \leq \frac{2}{3}$. It is a curious fact that most authors cite Lieberherr and Specker [4] to have proven that $r_3 = \frac{2}{3}$, but actually they only proved $r_3 \geq \frac{2}{3}$, and state that "Unfortunately we have not been able to determine τ_3 exactly". We remedy this by giving a family $(F_i)_{i\in\mathbb{N}}$ of 3-satisfiable formulas for which $\frac{2}{3}$ is tight in the limit. Further, we show that this limit cannot be achieved by a single 3-satisfiable formula, i.e. we show that $\mu^*(F)/\mu(F) > \frac{2}{3}$ for all 3-satisfiable formulas F.

We study the numbers s_k, the "unweighted" counterpart of r_k. Surely, $s_k \geq r_k$, for any k, and Trevisan's proof [2] extends to s_k, showing that $\lim_{k\to\infty} s_k = 3/4$. For particular values of k however, r_k and s_k can differ. We show that $s_2 = \frac{2}{3}$ and $\frac{2}{3} = r_3 < s_3 \leq 7/10$.

Theorem 3.1. $r_3 \leq \frac{2}{3}$.

Proof Idea. Choose variables $X = \{x_1, \ldots, x_n\}$ for n even. Add the unit clauses $\{x_i\}, i = 1, \ldots, n$ to F. For any balanced partition $X = U \uplus V$, i.e., $|U| = |V| = n/2$, introduce a variable y_{UV} and set $G_{UV} := \{\{\bar{u}, y_{UV}\} \mid u \in U\} \cup \{\{\bar{v}, \bar{y}_{UV}\} \mid v \in V\}$. Add all these G_{UV} to F. Fix an assignment α to the variables x_i, which sets k of the x_i to 1. Choose a balanced partition at random, and with high probability, roughly $k/2$ variables in U and roughly $k/2$ in V are set to 1. Hence setting y_{UV} to 0 or 1 satisfies roughly the same number of clauses in G_{UV}, namely all but $k/2$. Choose the clause weights such that the total weight of all G_{UV} is twice the total weight of X. Then no matter how you choose k, you will satisfy roughly $\frac{2}{3}$ of the weight. To fill out the details, one uses standard tail inequalities of probability theory. $\square$

Theorem 3.2. *For any 3-satisfiable formula F and any weight function μ, $\frac{\mu^*(F)}{\mu(F)} > \frac{2}{3}$. This means that the lower bound $r_3 \geq \frac{2}{3}$ is only tight in the limit.*

Proof. Consider the probability distribution from the proof of Theorem 2.1 with $p = \frac{2}{3}$. Every clause is satisfied with probability $\geq \frac{2}{3}$, and some with strictly larger probability. If such a clause is in F, the expectation of the number of satisfied clauses is $> \frac{2}{3}|F|$, and we are done. Otherwise, one can show that the number of satisfied clauses is a random variable with positive variance. Hence with positive probability, it takes on values above its expectation, which is $\frac{2}{3}$. Therefore, there is an assginment satisfying strictly more than $\frac{2}{3}$. □

We do not know whether there are $k \geq 2$ which have a tight formula F with $\mu^*(F)/\mu(F) = r_k$. However, we know that this can only be the case if r_k is rational. Therefore, Theorem 3.2 is interesting because it shows that r_k being rational is not sufficient for a tight formula to exist.

Theorem 3.3. *For any $k \geq 1$, if $r_k \notin \mathbb{Q}$, then*

$$\frac{\mu^*(F)}{\mu(F)} > r_k$$

for any k-satisfiable formula F.

Proof. For a formula F, let the *optimal* weight function $\tilde{\mu}$ be the weight function minimizing $\mu^*(F)/\mu(F)$. It is not difficult to see that this optimization problem can be modelled as a linear program with integer coefficients. Observe that $\mu^*(F)/\mu(F) \geq \tilde{\mu}^*(F)/\tilde{\mu}(F)$. Hence if there is a single tight formula F and a weight function μ with $\mu^*(F)/\mu(F) = r_k$, then r_k is the solution of the described linear program, and hence it is rational. □

Unweighted Formulas

As for formulas with unit clause weights, it is easy to see that $s_k \geq s_2$, and $s_1 = r_1 = \frac{1}{2}$. However, it turns out that $s_2 = \frac{2}{3}$, which is larger than $r_2 = \frac{\sqrt{5}-1}{2} \approx 0.619$.

Theorem 3.4. $s_2 = \frac{2}{3}$.

Proof. Let F be a 2-satisfiable formula. For a variable x, define

$$d^+(x) := |\{C \in F \mid x \in C \text{ and } C \setminus \{x\} \text{ contains only negative literals}\}|$$
$$d^-(x) := |\{C \in F \mid \bar{x} \in C \text{ and } C \text{ contains only negative literals}\}|$$

Switching a variable x means replacing each occurrence of x in F by $\bar{x}$ and vice versa. This, of course, does not change $\mu^*(F)$. As long as possible, apply the following rules:

1. If $d^-(x) > d^+(x)$, switch x
2. If $d^-(x) = d^+(x)$ and $\{\bar{x}\} \in F$, switch x.

It is easy to see that this process terminates after a finite number of steps. In the end, $d^+(x) \geq d^-(x)$ for any variable x, and $d^+(x) \geq d^-(x)+1$ if $\{\bar{x}\} \in F$. Let $F^- := \{C \in F \mid C$ contains only negative literals$\}$, and $F^+ := \{C \in F \mid C$ contains exactly one positive literal$\}$.

We calculate:

$$\begin{aligned}
|F^+| &= \sum_{x\in V,\{\bar{x}\}\in V} d^+(x) + \sum_{x\in V,\{\bar{x}\}\notin V} d^+(x) \\
&\geq \sum_{x\in V,\{\bar{x}\}\in V} (1+d^-(x)) + \sum_{x\in V,\{\bar{x}\}\notin V} d^-(x) \\
&= |F^-| + \sum_{x\in V} d^-(x) \\
&= \sum_{C\in F^-} (1+|C|) \quad \geq \quad 2|F^-| \ .
\end{aligned}$$

Hence $|F^-| \leq |F|/3$, and the assignment $\alpha = (1,\ldots,1)$, which sets each variable to 1, satisfies at least $\frac{2}{3}$ of all clauses of F. For an upper bound, the formula $\{\{x\},\{y\},\{\bar{x},\bar{y}\}\}$ demonstrates that $s_2 \leq \frac{2}{3}$. □

Theorem 3.5. *For 3-satisfiable formulas, we obtain the following bounds:*

$$\frac{21}{31} \leq s_3 \leq \frac{7}{10} \ ,$$

where $21/31 \approx 0.677 > \frac{2}{3} = r_3$.

Proof. To show that $s_3 \leq 7/10$, consider the following formula:

$$\big\{\{a\},\{b\},\{c\},\{d\},\{\bar{a},w\},\{\bar{b},w\},\{\bar{c},\bar{w}\},\{\bar{d},\bar{w}\},\{\bar{a},\bar{b},\bar{w}\},\{\bar{c},\bar{d},w\}\big\} \ .$$

This formula is 3-satisfiable and has 10 clauses, but no assignment satisfies more than 7 clauses.

To prove the lower bound, let F be any 3-satisfiable formula. We can assume F contains only positive unit clauses (if $\{\bar{x}\} \in F$, switch x). Let $X := \{x \in V \mid \{x\} \in F\}$. Write $F = F_1 \uplus F_2$ with

$$F_1 := \{\{x\} \in F\} \cup \{\{\bar{x}, l\} \in F \mid x \in X, \mathrm{vbl}(l) \in V \setminus X\}$$

and $F_2 := F \setminus F_1$. We will define two random assignments α and β. Define α such that $\Pr(\alpha(x) = 1) = \frac{2}{3}$ for $x \in X$, and $\Pr(\alpha(u) = 1) = 1/2$ for $u \in V \setminus X$. It is not difficult to see that $\Pr(\alpha \text{ satisfies } C) = \frac{2}{3}$ for $C \in F_1$ and $\Pr(\alpha \text{ satisfies } D) \geq 19/27 \approx 0.703$ for $D \in F_2$.

Next, we define the second random assignment β. If $x \in X$ and $\bar{x}$ does not occur in F_1, set $\Pr(\beta(x) = 1) = 3/4$. Otherwise, there is a clause $C_x = \{\bar{x}, l\} \in F_1$. Arbitrarily choose such a clause and call $\mathrm{vbl}(l)$ the *master variable* of x and C_x the *master clause.* Choose $\beta(u)$ to be 1 with probability 1/2, independently for each variable $u \in V \setminus X$. Depending on $\beta(u)$, set each variable x with master u such that its master clause is satisfied. Note that unit clauses having a master clause are satisfied with probability 1/2, and master clauses are satisfied with probability 1. Each remaining 2-clauses is satisfied with probability 3/4 (here we use that if $\{\bar{x}, l\}$ is the master clause of x, $\{\bar{x}, \bar{l}\} \notin F$, because of 3-satisfiability). A clause in F_2 is satisfied with probability $\geq 1/2$. Hence, the expected number of satisfied clauses is $3/4|F_1| + 1/2|F_2|$. We summarize:

$$\mathbf{E}[\mu^\alpha(F)] \geq \frac{2}{3}|F_1| + \frac{19}{27}|F_2|$$

$$\mathbf{E}[\mu^\beta(F)] \geq \frac{3}{4}|F_1| + \frac{1}{2}|F_2|$$

With probability $p = 27/31$, choose α, and with probability $4/31$, choose β. The expected number of satisfied clauses is $21/31|F|$. We obtain

$$s_3 \geq \frac{21}{31} > \frac{2}{3}$$

Note that $21/31 \approx 0.677$, which is larger than $\frac{2}{3}$. $\square$

Using a more sophisticated approach, we were actually able to derive a better lower bound of $s_3 \geq \frac{57}{82} \approx 0.695$. However, the proof is quite technical and requires some case analysis. We therefore do not include it in this work.

4 Conclusion

We have demonstrated that in partial satisfaction of k-satisfiable formulas, the results are different for weighted and unweighted formulas. Further, not only are the results different, also the methods employed to obtain these results differ greatly. In the weighted case, results are usually obtained by probabilistic arguments, whereas for s_2 we followed a completely different approach. Moreover, for s_3, we did use probabilistic methods, but had to introduce dependencies between variables. We would like to find a uniform approach for deriving lower bounds on s_k, for every k. We think the methods one has to develop to study the s_k will prove useful for different problems involving boolean satisfiability. Therefore, we think this topic is worth further effort. Let us conclude this paper by stating three open problems.

Problem 4.1. Devise a uniform method to prove lower bounds on s_k. Give families of k-satisfiable formulas providing upper bounds.

Conjecture 4.2. For every $k \geq 2$, $s_k > r_k$.

Problem 4.3. Is $s_k \in \mathbb{Q}$, for every k? For which k can the upper bound s_k be achieved by a single k-satisfiable formula, rather than an infinite family?

Acknowledgements

We thank Emo Welzl, who initially pointed out to the problem of partial satisfaction in unweighted formulas, for the fruitful suggestions and discussions on the topics of this paper.

References

1. Lieberherr, K.J., Specker, E.: Complexity of partial satisfaction. J. ACM **28**(2) (1981) 411–421
2. Trevisan, L.: On local versus global satisfiability. SIAM J. Discret. Math. **17**(4) (2004) 541–547
3. Král, D.: Locally satisfiable formulas. In: SODA '04: Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms, Philadelphia, PA, USA, Society for Industrial and Applied Mathematics (2004) 330–339
4. Lieberherr, K.J., Specker, E.: Complexity of partial satisfaction ii. Technical Report 293, Department of EECS, Princeton University (1982)
5. Yannakakis, M.: On the approximation of maximum satisfiability. J. Algorithms **17**(3) (1994) 475–502
6. Huang, M.D.A., Lieberherr, K.J.: Implications of forbidden structures for extremal algorithmic problems. Theor. Comput. Sci. **40**(2-3) (1985) 195–210