Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Survey on k-satisfiable Partial Satisfaction

Bachelor Thesis

Hu Qinheping

January 22, 2014

Advisors: Prof. Welzl, Dr. Hertli

Department of Computer Science, ETH Zürich

Abstract

A CNF formula F is k-satisfiable if each k clauses of F can be satisfied simultaneously. Define r_k to be the supremum of all real number r so that every weighted k-satisfiable CNF formula has an assignment that satisfies at least r-fraction of the weight. Define s_k analogously for unweighted CNF formulas.

We survey the results of the bound of r_k and s_k, i.e., under which conditions imposed on a CNF fomula which fraction of the clauses or weight is guaranteed to be simultaneously satisfied by some assignment, and consider whether such an assignment can be found algorithmically.

Furthermore, we discuss the problem of finding a assignment for k-satisfiable CNF formula satisfying at least given r-fraction of the total weight where $r > r_k$.

Contents

Chapter 1

Introduction

Satisfiability (SAT) is the problem of deciding whether a boolean formula in propositional logic has an assignment that evaluates to true. SAT occurs as a problem and is a tool in applications, e.g., Artificial Intelligence and circuits, and it is considered a fundamental problem in theory. A partial satisfaction problem is to consider how much fraction of a not-satisfiable formula can be satisfied. In this thesis, we discuss the satisfiability problem base a local satisfiability condition and we can see that how local satisfiability implies fractional global satisfiability.

In Boolean logic, a formula is in conjunctive normal form (CNF) if it is a conjunction of clauses , where a clause is a disjunction of literals. We use the simplified notation that

- a literal l is either a variable $x \in V$ or its negation $\overline{x}$;
- a clause C is a set of literals l;
- a CNF formula F over a finite variables set $V = \{x_1, ..., x_n\}$ is a set $\{C_1, ..., C_m\}$ of disjunctive clauses.

We say that a clause is "m-clause" if it contains exactly m literals. We require that no clause contains a variable and its negation simultaneously because such clause is always satisfied and meaningless.

We define a vocabulary function as follow: if l is a literal, then vbl(l) is its variable, i.e., vbl(x)=vbl($\overline{x}$) $= x$ for $x \in V$. In a formula F, switching a variable x means replacing each occurrence of x in F by $\overline{x}$ and vice versa.

For V a set of boolean variables, an assignment α on V is a mapping $\alpha : V \rightarrow \{0,1\}$, i.e. $\alpha \in \{0,1\}^V$. We say that an assignment α satisfies a literal l if either $l = x$ for $x \in V$ with $\alpha(x) = 1$ or $l = \overline{x}$ with $\alpha(x) = 0$, an assignment α satisfies a clause C if there exists a literal $l \in C$ such that α satisfies l, an assignment α satisfies a CNF formula F if α satisfies all clauses of F.

Sometime the considered formula can not be globally satisfied. So we want the considered formula weighted such that we can evaluate how much of it can be satisfied at the same time. The idea of weighted formulas is to assign a real number (weight) to each clause such that for an assignment α, we can see how much weight of the formula can be satisfied by this assignment. Here is a formal definition: for a CNF formula F on a variable set V, we define a clause weight function $\mu : F \to \mathbb{R}^+$, and the function μ extends to subsets $G \subset F$ by $\mu(G) = \sum_{C \in G} \mu(C)$. Given a truth assignment α, let $\mu^{[\alpha]}(F)$ denote $\mu(\{C \in F \mid \alpha \text{ satisfies } C\})$. Finally, let $\mu^*(F) := \max_\alpha \mu^{[\alpha]}(F)$. Note that an unweighted formula can be seen as a weighted formula in which each clause has same weight.

Instead of global satisfiable formulas, in this thesis we mainly discuss the formulas which are "locally satisfiable" in the sense that small subsets of the clauses are simultaneously, i.e., we deal with the notion of k-satisfiable CNF formula introduced and studied by Lieberherr and Specker [1].

Definition 1.1 *A CNF formula is k-satisfiable if any k clauses of it can be simultaneously satisfied.*

If a CNF formula is not satisfiable, it is sometimes desired to at least satisfy as much weight as possible. Here we use two ratios to investigate how much is the fraction of the weight and the clauses that can be always satisfied, given certain local satisfiable preconditions on the formulas.

Definition 1.2 *Define r_k as follow:*

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

and for formulas with unit clause weights, we define

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

Thus the ratio s_k is the supremum of the set of reals such that in any k-satisfiable formula with m clauses, at least $s_k m$ clauses can be simultaneously satisfied. Similarly, the ratio r_k is the supremum of the set of reals such that in any weighted k-satisfiable formula, at least fraction r_k of the total weight can be simultaneously satisfied.

Chapter 2

Previous Results

In [4], Käppeli and Scheder introduced a proof idea of $r_3 \leq \frac{2}{3}$, but they did not provide a complete proof. In this part, we first present a complete construction of a 3-satisfiable formula $SAT_3(n)$. Then we show that the bound $\frac{2}{3}$ of r_3 is asymptotically tight.

The formula $SAT_3(n)$ for n even is defined as follows: it has n variables x_i, $1 \leq i \leq n$, and 2^n variables y_{UV} where UV is one of 2^n partitions on $X := \{x_1, ..., x_n\}$, i.e., $X = U \uplus V$. The clauses of the formula $SAT_3(n)$ are the following:

- n clauses $\{x_i\}$ for $1 \leq i \leq n$ each of weight $\frac{1}{n}$.
- 2^n sets of clauses $G_{UV} := \{\{\overline{u}, y_{UV}\} | u \in U\} \cup \{\{\overline{v}, \overline{y}_{UV}\} | v \in V\}$. Each of these $n2^n$ clauses has the weight equal to $2/(n2^n)$, i.e., the sum of their weights is equal to 2.

Note that the total weight of $SAT_3(n)$ is equal to 3. Now We state two lemmas on formulas $SAT_3(n)$. We first show that it is always 3-satisfiable such that we can use it to bound r_3.

Lemma 2.1 *The formula $SAT_3(n)$ is 3-satisfiable for all n even.*

Proof Let W be a minimal subformula of a formula $SAT_3(n)$, which cannot be simultaneously satisfied. By the minimality of W, any variable either does not appear in the clauses of W at all or it has both a positive and a negative occurrence in the clauses of W, if not, i.e., a variable just have positive literals in W, then we let $W^{[x \mapsto 1]}$ be a smaller subset in which all clauses cannot be simultaneously satisfied, which is a contradiction to the precondition that the size of W is minimum. We can use this fact to prove that the size of W should be greater than 3.

Assume for the sake of contradiction that $|W| \leq 3$. Since W cannot be satisfied, it contain at least one 1-clause $\{x_k\}$ for some $1 \leq k \leq n$. Therefore W should also have a clause containing the negation of x_k, i.e., clause $\{\overline{x_k}, y_{UV}\}$ (or clause $\{\overline{x_k}, \overline{y}_{UV}\}$) for some UV. Then W should have a clause containing the negation of y_{UV}, i.e., clause $\{\overline{x_l}, \overline{y}_{UV}\}$. Because a variable appear in one partition just once, we know that $x_k \neq x_l$. At last, the variable x_l appear in W just with its negative occurrence, which is contradiction to the minimality of W. □

Then we show that every assignment satisfy almost same fraction of the weight in the formula $SAT_3(n)$.

Lemma 2.2 *The maximum weight of the satisfied clauses of a formula $SAT_3(n)$ is equal to* $2 + O(n^{-\frac{1}{4}})$.

Proof Consider a fixed assignment which set k of the variables x_i, $1 \leq i \leq n$ to true. We deal with each type of clause contained in $SAT_3(n)$ separately and see the satisfied probability:

- $\{x_i\}$ for $1 \leq i \leq n$

  There are exactly k clauses satisfied. Hence the weight of the satisfied clauses of this type is exactly $\frac{k}{n}$, note that the total weight of this type is 1.

- Clauses in all G_{UV},

  Choose a random partition UV such that for each $x_i \in X$, variable x_i is in the part U with probability $\frac{1}{2}$. Let T be the number of variables of U which are set to be true by the considered assignment. Note that T is a random variable and it follows a binomial distribution with the expected value equal to $\frac{k}{2}$.

  To satisfy a maximum number of clauses in G_{UV}, we set y_{UV} to 1 if $T \geq \frac{k}{2}$ and 0 otherwise. For the random partition UV, let $E_{UV}[unsatisfied]$ be the expected value of the number of unsatisfied clauses in G_{UV}. Then for $0 < \delta < 1$ we have

$$\begin{aligned} E_{UV}(unsatisfied) &= \sum_{t=0}^{k} \Pr[T=t]\min(t,k-t) \\ &\geq \sum_{|t-\frac{k}{2}|<\delta\frac{k}{2}} \Pr[T=t]\min(t,k-t) \\ &\geq (1-\delta)\frac{k}{2} \sum_{|t-\frac{k}{2}|<\delta\frac{k}{2}} \Pr[T=t]. \end{aligned}$$

  Chernoff bounds can be used to bound the probability

$$\sum_{|t-\frac{k}{2}|<\delta\frac{k}{2}} \Pr[T=t] = \Pr_{UV}\left[|T-\frac{k}{2}| < \delta\frac{k}{2}\right] \geq 1 - e^{-\delta^2 k/6} - e^{-\delta^2 k/4}.$$

We set $\delta = k^{-1/3}$ and observe that

$$\begin{aligned} e^{k^{\frac{1}{3}}/6} \geq 2k^{\frac{1}{3}} \quad &\Rightarrow \quad 1 - 2e^{-k^{\frac{1}{3}}/6} \geq 1 - k^{-\frac{1}{3}} \\ &\Rightarrow \quad 1 - e^{-k^{\frac{1}{3}}/6} - e^{-k^{\frac{1}{3}}/4} \geq 1 - k^{-\frac{1}{3}} \\ &\Rightarrow \quad E_{UV}(unsatisfied) \geq (1 - k^{-\frac{1}{3}})^2 \frac{k}{2} \end{aligned}$$

Therefore the weight of the satisfied clauses of this type is

$$2^n \cdot \frac{2}{n2^n} \cdot (n - E_{UV}(unsatisfied)) \leq 2 - \frac{k}{n}(1 - k^{-\frac{1}{3}})^2.$$

Now the total weight of the satisfied clauses in $SAT_3(n)$ is

$$w := \frac{k}{n} + 2 - \frac{k}{n}(1 - k^{-\frac{1}{3}})^2.$$

Suppose that $k \leq n^{\frac{3}{4}}$, it is easy to see that the total weight $w = 2 + O(n^{-\frac{1}{4}})$.

Otherwise, i.e., $k > n^{\frac{3}{4}}$, the total weight of the satisfied clauses

$$w = 2 + \frac{k}{n} - \frac{k}{n} + O(k^{-\frac{1}{3}}) = 2 + O(n^{-\frac{1}{4}}).$$

□

These two lemmas lead us to show that the bound of r_3 is tight.

Theorem 2.3 *For each $\varepsilon > 0$, there exists a 3-satisfiable formula F with weight function μ such that $\mu^*(F) \leq \frac{2}{3}\mu(F) + \varepsilon$.*

Proof From the Lemma 2.5 and the Lemma 2.6, we can choose the formula $F := SAT_3(n)$ for a sufficiently large number n such that the error term $O(n^{-1/4})$ from the Lemma 2.6 is smaller than the considered ε. Therefore, the maximum total weight of satisfied clauses of $SAT_3(n)$ is greater than $\frac{2}{3}\mu(F) + \varepsilon$. □

Note that Theorem 2.7 indicate that the bound $r_3 \geq 2/3$ is tight.

Chapter 3

The Hardness of $> r_k$ Problem.

In this section, we discuss the hardness of a sequence of $> r_k$ problems. The $> r_k$ problem is to determinate for a k-satisfiable formula if there exists an assignment satisfying more than r fraction of weight where $r > r_k$. In [1], Lieberherr first raised a conjecture about the unweighted version of this problem and gave a proof idea for the 2-satisfiable case. In the following part, we will formally definite a weighted version of the conjecture and show that this conjecture is hold for $k = 2, 3$.

Conjecture 3.1 *If $r = r_k + \varepsilon$ for $\varepsilon > 0$, the set of pairs (F, μ) ,where μ is the weight function on the k-satisfiable CNF formula F which have an assignment satisfying at least $r\mu(F)$ weight, is NP-complete.*

First we state a lemma as the reduce basis.

Lemma 3.2 *For k constant, the set of k-satisfiable CNF formula which is globally satisfiable is NP-complete.*

Proof Here we describe a reduction from the well-known NP-complete problem 3SAT to k-satisfiable and satisfiable CNF problem. That is, given a 3-CNF formula F, we can construct a k-satisfiable CNF formula F' in polynomial time such that F is satisfiable iff F' is satisfiable. Assume that the size of the given 3-CNF formula should greater than k to make the problem well defined.

In fact, we just need to check whether F is k-satisfiable. If F is k-satisfiable, we just let $F' := F$, otherwise let F' be an arbitrary unsatisfiable k-satisfiable CNF formula.

There are $\binom{|F|}{k}$ number of subformulas of F with size k, and for each of these subformulas there are at most 2^{3k} possible assignments. So we can check whether F is k-satisfiable in $\binom{|F|}{k}2^{3k}$ time, which is polynomial since k is constant. □

Now we prove a lemma give us a process which can produced a 2-satisfiable CNF formula with a known maximum satisfied weight fraction $r' < r$. The idea is to construct a sequence of 2-satisfiable CNF formulas whose maximum satisfied weight fraction can be calculated from the size n and that fraction converge to r_2. Then we can choose a proper n such that the fraction r' is less than the given r.

Lemma 3.3 *For $r = r_2 + \varepsilon$ where $\varepsilon > 0$, we can construct a 2-satisfiable CNF formula F with the weight function μ and return γ' in time $O(\varepsilon^{-2})$, such that $\mu^*(F) = r'\mu(F)$ for some $r' < r$.*

Proof We choose some variables set V with $|V| = n$ and construct 2-satisfiable formula

$$F := \{\{x\} \mid x \in V\} \cup \{\{\overline{x}, \overline{y}\} \mid \{x, y\} \in \binom{V}{2}\};$$

and the weight function μ assigns $v = \frac{-1+\sqrt{5}}{2}n = \Phi n$ to every 1-clause and 1 to every 2-clause of F. Note that the overall weight is $nv + \binom{n}{2}$.

If we set k of variables in V to 1, then we can satisfy weight

$$s(k) := kv + \binom{n}{2} - \binom{k}{2},$$

which attains its maximum for $k = v + \frac{1}{2}$ because $s(k)$ is a quadratic function. Therefore no more than

$$\max(s(v), s(v+1)) = \frac{v^2}{2} + \frac{v}{2} + \binom{n}{2}$$

weight of F can be satisfied. That is,

$$\frac{\mu^*(F)}{\mu(F)} = \frac{v^2 + v + n(n-1)}{2nv + n(n-1)} = \frac{\Phi^2 + 1 + \frac{\Phi - 1}{n}}{2\Phi + 1 - \frac{1}{n}} > \frac{\Phi^2 + 1}{2\Phi + 1} = r_2,$$

to make this fraction less than r, we choose

$$n := \lfloor \frac{r + \Phi - 1}{2\Phi r + r - \Phi^2 - 1} \rfloor + 1,$$

note that this n is from the inequality $\mu^*(f)/\mu(F) < r$. Since $r_2 = \Phi$, we rewrite

$$n = \lfloor \frac{\sqrt{5} - 2 + \varepsilon}{\sqrt{5}\varepsilon} \rfloor + 1 = O(\frac{1}{\varepsilon})$$

At last, we get the formula F over the variables set $|V| = \lfloor \frac{r+\Phi-1}{2\Phi r + r - \Phi^2 - 1} \rfloor + 1$ and the weight function μ such that $\mu^*(F) = r'\mu(F)$ for the ratio $r' := \frac{\Phi^2+1+\frac{\Phi-1}{n}}{2\Phi+1-\frac{1}{n}}$. Because the size of the formula is $O(n + n^2) = O(\varepsilon^{-2})$, this construction can finish in time $O(\varepsilon^{-2})$.

Now we prove the 2-satisfiable case of the conjecture using the lemma. The idea of the deduction is mixing two formulas with different maximum satisfied weight fraction to get a new formula with the maximum satisfied weight fraction exactly r_2.

Theorem 3.4 *For $r = r_2 + \varepsilon$, the set of pairs (F, μ) ,where μ is the weight function on the 2-satisfiable CNF formula F which have an assignment satisfying at least $r\mu(F)$ weight, is NP-complete if $O(\varepsilon^{-2}) = poly(l)$ where l is the total length of string (F, μ).*

Proof Deduce from the Lemma 4.2, we just need to give a polynomial transformation T which transforms a 2-satisfiable CNF F' to a pair $T(F) = (F, \mu)$ where F is 2-satisfiable CNF formula and μ is the weight function on F, such that F' is satisfiable iff F has an assignment satisfying at least the fraction r of the total weight $\mu(F)$.

According to the Lemma 4.3, we can construct a 2-satisfiable CNF formula F_0 and the weight function μ_0 on it in polynomial time since $O(\varepsilon^{-2}) = \text{poly}(l)$, such that $\mu_0^*(F_0) = r'\mu_0(F_0)$ for $r' < r$.

Then let F contain F_0 concatenated with F' and μ extend μ_0 to formula F' such that $\mu(C) = \mu_0(C)$ for all $C \in F_0$ and $\mu(F') = \frac{r-r'}{1-r}\mu_0(F_0)$.

At last, the formula F' is satisfiable iff we can satisfy the fraction

$$\frac{\mu^*(F)}{\mu(F)} = \frac{\mu(F') + \mu^*(F_0)}{\mu(F)} = \frac{\frac{r-r'}{1-r}\mu_0(F_0) + r'\mu_0(F_0)}{\frac{r-r'}{1-r}\mu_0(F_0) + \mu_0(F_0)} = r$$

of the total weight of F. □

In general, we can prove the k-satisfiable case of the conjecture 4.1 if we have two conditions:

- the set of k-satisfiable CNF formulas which is satisfiable is NP-complete;
- we can construct a k-satisfiable formula F with weight function μ and return the fraction r' "efficiently", such that $\mu^*(F)/\mu(F) = r$ for $r' < r$.

Then, similar to the Lemma 4.3, we construct a 3-satisfiable CNF formula using the same idea.

Lemma 3.5 *For $r = r_3 + \varepsilon$, we can construct a 3-satisfiable CNF formula F with the weight function μ and return the fraction r' in time $O(\varepsilon^{-4}2^{\varepsilon^{-4}})$, such that $\mu^*(F) = r'\mu(F)$ for some $r' < r$.*

Proof Let $F := SAT_3(n)$ which is introduced in the proof of theorem, and we just need to chose proper n such that $\mu^*(F) = r'\mu(F)$ for some $r' < r$. We have shown that at least $r_3 + cn^{-1/4}$ fraction of weight can be satisfied in $SAT_3(n)$ where c is constant. So we can choose a large $n = O(1/\varepsilon^4)$ such that $\mu^*(F) < r\mu(F)$. So the only problem is to calculate $r' = \frac{\mu^*(F)}{\mu(F)}$ for that n. Since r is constant, the size n is also constant. Assume that k of variables x_i are set to 1 in $SAT_3(n)$, the ratio

$$r' = \max_k \frac{\max_{\alpha(k)} \mu^{[\alpha(k)](F)}}{\mu(F)} = \max_k \sum_{t=0}^{k} \binom{k}{t} \frac{1}{2^k} \min(t, k-t),$$

where $\alpha(k)$ is an assignment which set exactly k of x_i to true. To get this r' we need to calculate n terms of $\sum_{t=0}^{k} \binom{k}{t} \frac{1}{2^k} \min(t, k-t)$ and find the maximum one, which need no more than $O(n^2)$ time. Note that the size of $SAT_3(n)$ is equal to $n(1+2^n) = O(n2^n)$. So at last, we construct a formula F and weight function μ such that $\mu^*(F) = r'\mu(F)$ in time $O(n2^n) = O(\varepsilon^{-4}2^{\varepsilon^{-4}})$ and calculate r' in time

$$O(n^2) = O(\varepsilon^{-8})$$

. □

Theorem 3.6 *For $r = r_3 + \varepsilon$, the set of pairs (F, μ) ,where μ is the weight function on the 3-satisfiable CNF formula F having an assignment which satisfies at least $r\mu(F)$ weight, is NP-complete if $O(\varepsilon^{-4}2^{\varepsilon^{-4}}) = poly(l)$ where l is the total length of the string (F, μ).*

Proof Use the same method in the proof of Theorem 4.4 and deduce from Lemma 4.2 and Lemma 4.5, this theorem follows. □

Bibliography

[1] Lieberherr, K.J., Specker, E.: *Complexity of partial satisfaction*. J. ACM 28(2) 1981 411-421.

[2] Lieberherr, K.J., Specker, E.: *Complexity of partial satisfaction ii*. Technical Report 293, Department of EEC, Princeton University (1982)

[3] Yannakakis, M.: *On the approxiamation of maximum satisfiability*. J. Algorithms 17(3) 1994 475-502

[4] Käppeli C, Scheder D.: *Partial Satisfaction of k-Satisfiable Formulas* J. Electronic Notes in Discrete Mathematics, 2007, 29: 497-501.

[5] 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

[6] Trevisan, L.: *On local versus global satisfiability.* SIAM J. Discret. Math. 17(4) 2004 541-547

[7] Carmichael, R.D: *Introduction on the Theory of Groups of Finite Order.* Dover, New York 1937.

[8] Landau, E. *Handbuch der lehre von der vertetlung der primzahlen.* Chelsea Publishing, New York, 1974

[9] Huang, M.D.A, Lieverherr, K.J.: *Implications of forbidden structures for extremal algorithmic problems.* Theor.Comput.Sci.40(2-3)(1985)195-210