Solving p-Max-r-Satisfiability Above a Tight Lower Bound

Eun Jung Kim
Joint work with with N.Alon, G.Gutin, S.Szeider and A.Yeo

Department of Computer Science
Royal Holloway, University of London
eunjungkim78@gmail.com

23 December 2009

Outline

1. **Introduction**
   - Maximum Satisfiability
   - The class FPT
2. **p-Max-r-Sat for general r**
3. **p-Max-r-Sat for $r = 2$**

Maximum Satisfiability

Maximum Satisfiability

- Given a CNF formula, find a truth assignment that satisfies as many clauses as possible.

Max-r-Sat

- Given a CNF formula, each clause having exactly r literals, find a truth assignment that satisfies as many clauses as possible.

Even $\textsc{Max-2-Sat}$ is NP-hard and hard to approximate (i.e. cannot have $((C - \epsilon)$-approximation for some constant C unless P=NP), in strong contrast with $2\text{-}\textsc{Sat}$ which is solvable in linear time.

Parameterized Problems and FPT

FPT

A parameterized problem is fixed-parameter tractable if there is $f(k)n^c$ time algorithm for some constant c.

- Vertex Cover: Is there a vertex cover of size at most k? Best known $O^*(1.2832^k)$.
- Steiner Tree: Find a minimum weight subgraph connecting k prescribed terminal vertices of G. Best known $O^*(2^k)$.
- k-path: Does G contain a path of length k? Best $O^*(4^k)$.

Parameterized Intractability

There are also problems which are unlikely to be FPT. Such as: k-weighted CNF, Independent Set, Hitting Set, Dominating Set, Set Cover...

FPT and Kernel

Reduction To Problem Kernel

A kernelization of a parameterized problem Π is a many-to-one transformation from $(I, k) \in \Sigma^*$ to $(I', k') \in \Sigma^*$ s.t.

- $(I, k) \in \Pi$ if and only if $(I', k') \in \Pi$.
- $k' \leq k$ and $|I'| \leq g(k)$ for some computable function g.
- Transformation computable in time polynomial in $|I|$ and k.

Theorem

A parameterized problems belongs to FPT if and only if it allows a kernelization.

Example: Vertex Cover

- LP Relaxation of ILP formulation has a half-integral optimal solution.

Tight Bound and Parameterization

It is always possible to satisfy a $1 - 2^{-r}$ fraction of a given multiset of clauses.

- Random truth assignment satisfies a clause with $p = 1 - \frac{1}{2^r}$.
- Derandomization yields $(1 - 2^{-r})$-approximation. Best possible.

p-Max-r-Sat

Instance: A pair (F, k) where F is a set of m clauses of size r and k is a nonnegative integer.
Parameter: The integer k.
Question: Is $\#\text{sat}(F) \geq ((2^r - 1)m + k)/2^r$?

Algebraic Representation

Let F be an r-CNF formula with clauses $C_1, \ldots, C_m$ in the variables $x_1, x_2, \ldots, x_n$.

Random Variable X

- $X = \sum_{C \in F}[1 - \prod_{x_i \in \mathrm{var}(C)}(1 - \epsilon_i x_i)]$
- $\epsilon_i = 1$ if x_i is in C, $\epsilon_i = -1$ if $\overline{x_i}$ is in C.

Example 1: $F = \{x_1 x_2, x_1\overline{x_2}, \overline{x_1}x_3, \overline{x_2}x_3, x_2 x_3\}$: $X = x_1 + 3x_3 + x_1 x_3$

Example 2: $F = \{x_1 x_2, x_1\overline{x_2}, \overline{x_1 x_3}, \overline{x_2}x_3, x_2 x_3\}$: $X = x_1 + x_3 - x_1 x_3$

Lemma

For a truth assignment τ, $X = 2^r(\#\mathrm{sat}(\tau, F) - (1 - 2^{-r})m)$. That is, $X \geq k$ iff F is a YES-instance.

Probabilistic Inequality Come Into Play

Intuitively, if the variance of X is large enough we may have $X \geq k$.

N.Alon, G.Gutin and M.Krivelevich

Let X be a random variable. Suppose that X satisfies

1. $\mathbb{E}(X) = 0$
2. $\mathbb{E}(X^2) = \sigma^2 > 0$ and
3. $\mathbb{E}(X^4) \leq b\sigma^4$

Then Prob($X > \frac{\sigma}{4\sqrt{b}}$) $\geq \frac{1}{4^{4/3}b} > 0$

If we can express $\mathbb{E}(X^2)$ in terms of $|I|$, say $\mathbb{E}(X^2) \geq |I|^2$, then

- either $|I|$ is large enough ($\geq k$) and I is a YES-instance
- or $|I|$ is upper-bounded in terms of k, meaning quadratic kernel.

Probabilistic Inequality Come Into Play

$X = X(x_1, x_2, \ldots, x_n)$ can be written as $X = \sum_{I \in \mathcal{S}} X_I$, where

- $X_I = c_I \prod_{i \in I} x_i$, each c_I is a nonzero integer and
- $\mathcal{S}$ is a family of nonempty subsets of $\{1, \ldots, n\}$ each with at most r elements.

1st Moment

$\mathbb{E}(X) = 0$

2nd Moment: Parseval's Theorem says

For any $f : \{-1, 1\} \to \mathbb{R}$, $\mathbb{E}(f(x)^2) = \sum_{I \subseteq [n]} \hat{f}(I)^2$.
(in our case, $\mathbb{E}(X^2) = \sum_{I \in \mathcal{S}} c_I^2$)

Probabilistic Inequality Come Into Play

4th Moment: Hypercontractive Inequality says

Let $f(x) = \sum_{|I| \le r} \hat{f}(I) \prod_{i \in I} x_i$ denote an arbitrary multilinear polynomial over $x_1, \ldots, x_n$ of degree at most r. Define a random variable $X = f(x_1, \ldots, x_n)$ by choosing a vector $(x_1, \ldots, x_n) \in \{-1, 1\}^n$ uniformly at random and setting $X = f(x_1, \ldots, x_n)$. Then $\mathbb{E}(X^4) \le 9^r (\mathbb{E}(X^2))^2$.

The Main Result for General r

Theorem

The problem p-Max-r-Sat *is fixed-parameter tractable and can be solved in time* $O(m) + 2^{O(k^2)}$*. Moreover, there exists a kernel of size* $O(k^2)$.

Outline

- $\mathrm{Prob}(X \geq \frac{\sqrt{\mathbb{E}(X^2)}}{2\sqrt{b}}) > 0$, where $b = 9^r$
- $\mathbb{E}(X^2) = \sum_{I \in \mathcal{S}} c_I^2 \geq |\mathcal{S}|$
- Therefore $\mathrm{Prob}(X \geq \frac{\sqrt{|\mathcal{S}|}}{2 \cdot 3^r}) > 0$.
- Now, if $k \leq \frac{\sqrt{|\mathcal{S}|}}{2 \cdot 8^r}$ then YES-instance.
- Otherwise, $|\mathcal{S}| = O(k^2)$...

Semicomplete Reduction

- Two clauses Y, Z has a conflict if there is a literal $p \in Y$ such that $\overline{p} \in Z$.
- r-CNF formula F is semicomplete if the number of clauses is $m = 2^r$ and every pair of distinct clauses of F has a conflict. ex) $\{xy, x\overline{y}, \overline{x}z, \overline{xz}\}$.

Lemma

Every truth assignment to a semicomplete r-CNF formula satisfies exactly $2^r - 1$ clauses.

Semicomplete Reduction Rule & Lemma

Given an r-CNF formula F that contains a semicomplete subset $F' \subseteq F$, delete F' from F and consider $F \setminus F'$ instead. The semicomplete reduction is *safe*.

The Main Result for $r = 2$

Terminology

- A variable $x \in \mathrm{var}(F)$ is insignificant if for each literal y the numbers of occurrences of the two clauses xy and $\overline{x}y$ in F are the same.
  ex) $F = \{x_1x_2, x_1\overline{x_2}, \overline{x_1}x_3, \overline{x_2}x_3, x_2x_3\}$
- A variable $x \in \mathrm{var}(F)$ is significant if it is not insignificant.

Theorem

Let F be a 2-CNF formula which is irreducible w.r.t. semicomplete reduction *and let $k \geq 0$ be an integer. If F has more than $3k - 2$ significant variables, then $\#\mathrm{sat}(F) \geq (3|F| + k)/4$, i.e. F is a YES-instance.*

Proof Outline of the Main Theorem

1. Transform p-MAX-2-SAT to the problem of finding max-weighted subgraph in G^0
2. Switch some vertices of G^0, if necessary, to ensure the existence of a subgraph of G^0 with weight k.
3. When we cannot ensure a weight-k subgraph of G^0: Use Tutte-Berge formula to show that G^0 is small enough (kernel).

Switching a variable

We switch a variable x in F by replacing x with $\overline{x}$ and $\overline{x}$ with x for each occurrence of x.
Switching a set X of variables in F is naturally defined. We say that F_X is obtained from F by *switching* X.

Lemma

$\#\text{sat}(F) = \#\text{sat}(F_X)$.

We are allowed to *switch* a variable x instead of assigning $x = FALSE$.

Construction of Auxiliary Graph G^0

Let $c(x)$ ($c(xy)$ respectively) denote the number of clauses in F containing x (xy respectively).

Transformation Lemma

For each subset $R = \{x_1, \ldots, x_q\} \subseteq \text{var}(F)$ we have $\#\text{sat}(F) \geq (3m + k_R)/4$, for k_R equals

$$\sum_{1\leq i\leq q} (c(x_i) - c(\overline{x_i})) + \sum_{1\leq i<j\leq q} (c(x_i\overline{x_j}) + c(\overline{x_i}x_j) - c(x_ix_j) - c(\overline{x_i}\overline{x_j})).$$

We construct a canonical *auxiliary graph* $G = (V, E)$ from F with weights:

- $w(x) := c(x) - c(\overline{x})$
- $w(xy) := c(x\overline{y}) + c(\overline{x}y) - c(xy) - c(\overline{x}\overline{y})$

Operations on G^0

Let G^0 be the graph obtained from G by removing all edges of weight zero.
Switching a variable x in F corresponds to the followings.

- Reversing the sign of $w(x)$
- Reversing the sign of all edges adjacent with x.

Switching a set of variables X corresponds to the followings.

- Reversing the signs of $w(x)$ for all $x \in X$.
- Reversing the signs of all edges between X and $V \setminus X$.

Lemma

If there exists a set $X \subset V(G)$ and an induced subgraph $Q = (U, H)$ of G with $w_X(Q) \geq k$, then $\text{sat}(F) \geq (3m + k)/4$.

Induced Tree allows nice strategy

Let Q be an induced tree. Then

- We can find a set $X \subset V(Q)$ such that $w_X(Q) \geq |E(Q)|$.
- Search and switch, if necessary, and take a random switching of the whole Q.

Let $Q_1, \ldots, Q_m$ be a collection of vertex-disjoint induced trees. Then

- We can find a set $X \subset V(Q)$ such that $w_X(\bigcup_{i=1}^m Q_i) \geq \sum_{i=1}^m |E(Q_i)|$
- Perform a sequence of independent random switchings of $Q_1, \ldots, Q_m$.

YES-instance, or a small matching

If NO-instance, we do NOT have a collection of vertex-disjoint induced trees with at least k edges in total.
Consequently, there is no matching of size k.

Tutte-Berge Formula

The size of a maximum matching in G^0 equals

$$\min_{S \subseteq V(G^0)} \frac{1}{2}\{|V(G^0)| + |S| - oc(G^0 - S)\}$$

where $oc(G^0 - S)$ is the number of odd components (connected components with an odd number of vertices) in $G^0 - S$.

It suffices to prove that $oc(G^0 - S) \leq k - 1 + \#$ insig. vars.

We consider the parameterized MAX-r-SAT problem and proved that

- the problem is fixed-parameter tractable for each fixed r.
- the problem can be reduced to an equivalent problem of size $O(k^2)$ for each fixed r.
- when $r = 2$, there is a problem kernel with at most $3k$ variables.

Some interesting problems to consider includes

- Establishing FPT / W[t]-hardness for parameterization-above/below type problems such as: Planar independent set problem, Max-Lin-p, Max-Cut, Max-SAT... Only a few results available.
- In particular, Satisfiability Type Problems from parameterized perspective: been studied a lot, but not coping with inapproximability results yet.