CMPUT 675: Approximation Algorithms **Winter 2005**

Lecture 22: April 4

Lecturer: Mohammad R. Salavatipour *Scribe: Zhipeng Cai*

Today we will present some improved versions of the PCP Theorem and show how to derive hardness results for Vertex Cover, Max-3SAT, and Independent Set.

22.1 Improved PCP Theorems

As we saw in the last lecture, PCP theorem says: $NP = PCP_{1,\frac{1}{2}}(O(\log n), O(1))$. We are interested to reduce the number of query bits and also decrease the the probability of failure (i.e. better soundness). Our first theorem says that the number of query bits can be as small as 3 bits (and as you show in assignment 3 this is best possible).

Theorem 22.1 *For some $s < 1 : NP = PCP_{1,s}(O(\log n), 3)$*

Proof: Let $L \in NP$ be an arbitrary language and y be an instance of L. By PCP theorem we can construct a 3CNF formula F such that:

- if $y \in L$ there is a truth assignment for F such that all clauses of F are satisfied.
- if $y \notin L$ then for any truth assignment for F at most $(1 - \epsilon)$ fractions of clauses are satisfied.

We assume that the proof π given for y is the truth assignment to formula F above. The verifier V given y and proof π, first computes F in polytime. Then uses $O(\log n)$ bits to pick a random clause of F and query the truth assignment to its 3 variables (which of course requires only 3 bits of the proof). The verifier accepts if and only if the clause is satisfied. It is easy to see that:

- If $y \in L$ then there is a proof π s.t. V accepts with probability 1.
- If $y \notin L$ then for any proof π, V accepts with probability at most $1 - \epsilon$.

■

The downside of this theorem is that, although the number of query bits is best possible, the soundness probability is much worse than $\frac{1}{2}$. The following theorem shows that we can actually keep the soundness probability arbitrary close to $\frac{1}{2}$ while reading only 3 bits of the proof.

Theorem 22.2 (Guruswami/Sudan/Lewin/Trevisan/93) *For all $\epsilon > 0$*

$$NP = PCP_{1,\frac{1}{2}+\epsilon}(O(\log n), 3).$$

Note that the 3 bits of the proof are selected by the verifier adaptively. That is, the position of the second bit checked may depend on the truth value of the first bit read. This theorem is tight by the following result:

Theorem 22.3 *(Karloff/Zwick)*

$$P = PCP_{1,\frac{1}{2}}(O(\log n), 3).$$

Earlier, Håstad proved the following result from which several tight hardness results follow:

Theorem 22.4 (Håstad'97) *$NP = PCP_{1-\epsilon,\frac{1}{2}+\epsilon}(O(\log n), 3)$ where the verifier selects the 3 bits of the proof a priori. That is, the verifier uses $O(\log n)$ random bits to choose 3 positions, i_1, i_2, i_3 of the proof and a bit b and accepts if and only if $\pi(i_1) \oplus \pi(i_2) \oplus \pi(i_3) = b$.*

The following results follow from Håstad's theorem.

Corollary 22.5 *For any $\epsilon > 0$, it is NP-hard to approximate:*

- *Max-3SAT within a factor of $(\frac{7}{8} + \epsilon)$,*
- *Vertex cover within a factor of $(\frac{7}{6} + \epsilon)$,*
- *Maximum independent set is within a factor of $(\frac{1}{2} + \epsilon)$.*

Definition 22.6 *Given a system of linear equations module 2 with 3 distinct variable's per equation (e.g $X_1 \oplus X_2 \oplus X_3 = b$). The value of an assignment to the variable's is the number of equations satisfied by that assignment.*

Max-E3LIN2: Given a system of linear equations module 2 with 3 distinct variable's per equation, find an assignment with maximum number of satisfied equations.

Clearly, given an assignment we can find the value of the assignment in polynomial time. There is also a trivial $\frac{1}{2}$-approximation algorithm for this: consider the two assignments $x_i = 0$ for all i and $x_i = 1$ for all i. At least one of these two assignments satisfies at least half the equations. Surprisingly, this is the best possible approximation factor one can hope for.

Theorem 22.7 *Unless P=NP, there is no $(\frac{1}{2} + \epsilon)$ -approximation for Max-E3LIN2 for any $\epsilon > 0$.*

Proof: Given a language $L \in NP$ and an instance y for L consider the verifier V for L from Håstad PCP theorem. Assume that for every random string V is going to read 3 positions i_1, i_2, i_3 and accepts if and only if $\pi(i_1) \oplus \pi(i_2) \oplus \pi(i_3) = b$. So for every random bit string we have one linear equation.

We generate all the n^d random bits and from V and we get n^d equations, one for each string. For random string r_i, V accepts if and only if the truth assignment $\pi(i_1) \oplus \pi(i_2) \oplus \pi(i_3) = b$ satisfy that equation.

If y is yes instance then there is a proof for y such that V accepts with probability $\geq (1 - \epsilon)$, i.e. a fraction of $\geq (1 - \epsilon)$ of equations are satisfied.

If y is a no instance then for any any proof V accepts with probability $\leq \frac{1}{2} + \epsilon$, i.e. at most a fraction of $\leq \frac{1}{2} + \epsilon$ of equation are satisfied.

This implies a hardness gap of $\frac{\frac{1}{2}+\epsilon}{1-\epsilon} \leq \frac{1}{2} + 2\epsilon$. ■

Using this theorem we prove the hardness results mentioned in Corollary 22.5.

22.2 Hardness results for Max-3SAT, Vertex Cover, and Max Independent Set

Theorem 22.8 *Unless P=NP, there is no $(\frac{7}{8}+\epsilon)$-approximation algorithm for Max-3SAT, for any $\epsilon > 0$.*

Proof: We give a gap-preserving reduction form Max-E3LIN2 problem to Max-3SAT. The reduction maps every equation $x \oplus y \oplus z = b$ in the instance of Max-E3LIN2 into a 3CNF formula with 4 clauses depending on whether $b = 0$ or $b = 1$ as follows:

$x \oplus y \oplus z = 0, \Leftrightarrow (\overline{x} \vee \overline{y} \vee \overline{z}) \wedge (\overline{x} \vee y \vee z) \wedge (x \vee \overline{y} \vee z) \wedge (x \vee y \vee \ \overline{z})$

$x \oplus y \oplus z = 1, \Leftrightarrow (x \vee y \vee z) \wedge (x \vee \overline{y} \vee \overline{z}) \wedge (\overline{x} \vee y \vee \overline{z}) \wedge (\overline{x} \vee \overline{y} \vee \ z)$

Let I be the instance of Max-E3LIN2 and $\phi(I)$ be the 3CNF formula obtained. Clearly if I has m equations, then $\phi(I)$ has $4m$ 3-clauses.

If I is a yes instance, i.e. there is an assignment with value at least $(1-\epsilon)m$ then the same assignment satisfies at least $4(1-\epsilon)m + 3\epsilon m$ of clauses.

If I is a no instance, then every assignment has value at most $(\frac{1}{2}+\epsilon)m$. Therefore at most $4(\frac{1}{2}+\epsilon)m+3(\frac{1}{2}-\epsilon)m$ of clauses of $\phi(I)$ are satisfied by any truth assignment.

Thus the hardness gap is

$$\frac{4(\frac{1}{2}+\epsilon)+3(\frac{1}{2}-\epsilon)}{4(1-\epsilon)} = \frac{7}{8}+\epsilon',$$

where ϵ' is a constant depending on ϵ only.

■

Finally, we use Theorem 22.7 to prove hardness results for vertex cover and maximum independent set. Note that since a clique is the complement of an independent set, our hardness factor for maximum independent set is the hardness of Max-Clique too.

Theorem 22.9 *Unless P=NP, there is no $\frac{7}{6}+\epsilon$ -approximation for vertex cover problem and no $\frac{1}{2}+\epsilon$ -approximation for Clique/Independent set, for any $\epsilon > 0$.*

Proof: Note that the complement of a vertex cover is an independent set. We give a gap-preserving reduction from an instance of Max-E3LIN2. Starting from an instance I for Max-E3LIN2 we construct a graph G with $4m$ vertices where m is the number of equations of I.

For every equation there are exactly four assignments that satisfy the equation. We have one vertex for each of these 4 assignments. All four vertices of each equation are connected to each other.

We also add edges between pairs from different groups of 4 vertices that are "inconsistent". A pair of vertices from different equation are inconsistent if a variable is assigned different values in the assignment corresponding to those vertices. (see Figure 22.1).

If I is a yes instance then there is an assignment for I which satisfies at least $(1-\epsilon)m$ of equations. Consider the equations satisfied by this truth assignment and consider the assignment of these equations. Each of these assignments corresponds to a vertex of G. It is easy to see that these $(1-\epsilon)m$ vertices are all independent (i.e. no edge between them) because they are *not* inconsistent (come from the same truth assignment). Thus G has an independent set of size $\geq (1-\epsilon)m$, and so a vertex cover of size size $\leq (3+\epsilon)m$.

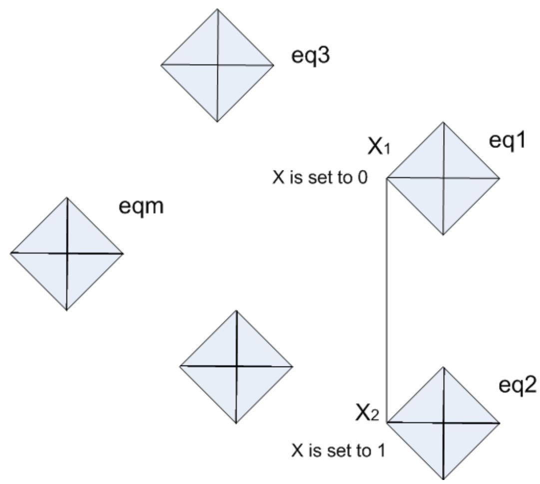

Figure 22.1: Reduction from Max-E3LIN2

Now suppose that I is a no instance, i.e. every assignment has value at most $(\frac{1}{2}+\epsilon)m$. This implies that every set of size $(\frac{1}{2}+\epsilon)m+1$ in G must have an edge. The reason is that: clearly such a set has an edge if it has two vertices from the same set of 4 vertices. Also, if there is an independent set of size $(\frac{1}{2}+\epsilon)m+1$ then they correspond to a set of consistent assignments to equations that satisfy at least $(\frac{1}{2}+\epsilon)m+1$ of the equations, which cannot be true. Thus the every independent set of G has size at most $(\frac{1}{2}+\epsilon)m$, and every vertex cover has size at least $(3.5-\epsilon)m$.

The hardness gap for vertex cover is $\frac{\frac{(3.5-\epsilon)}{4}}{\frac{(3+\epsilon)}{4}} = \frac{7}{6}+\epsilon'$, and the hardness gap for independent set (and also clique) is $\frac{\frac{(\frac{1}{2}+\epsilon)}{4}}{\frac{1-\epsilon}{4}} = \frac{1}{2}+\epsilon'$.

∎