CSE 533: The PCP Theorem and Hardness of Approximation **(Autumn 2005)**

Lecture 19: UGC-hardness of Max Cut

Dec. 5, 2005

Lecturer: Ryan ODonnell *Scribe: Ryan O'Donnell and Sridhar Srinivasan*

1 UGC-hardness of MAX-CUT

Our goal in this lecture is to show that, assuming the Unique Games Conjecture (UGC), it is NP-hard to improve upon the approximation algorithm of Goemans and Williamson for MAX-CUT.

We will need the following two "ingredients, discussed last lecture:

1. The Unique Games Conjecture: for all $\delta > 0$, there is a sufficiently large m such that, if $|\Sigma| \geq m$, then Gap-Unique-Label-Cover$(\Sigma)_{1-\delta,\delta}$ is NP-hard.

2. The Majority Is Stablest Theorem: Let $-1 < \rho < 0$, $\epsilon > 0$. Then, $\exists \tau > 0$, $C < \infty$, such that if $f : \{-1,1\}^m \to [-1,1]$ satisfies
$$\mathrm{Inf}_i^{\leq C}(f) := \sum_{|S| \leq C, S \ni i} \hat{f}(S)^2 \leq \tau$$
for all $1 \leq i \leq m$, then we have
$$\frac{1}{2} - \frac{1}{2}\mathrm{Stab}_\rho(f) = \frac{1}{2} - \frac{1}{2}\sum_S \rho^{|S|}\hat{f}(S)^2 < \frac{\cos^{-1}\rho}{\pi} + \epsilon.$$

In the remainder of this lecture, we show the following:

Theorem 1.1. *The Unique Games Conjecture implies that* $\forall\ -1 < \rho < 0$ *and* $\epsilon > 0$, *Gap-MAX-CUT*$_{\frac{1-\rho}{2}-\epsilon, \frac{\cos^{-1}\rho}{\pi}+\epsilon}$ *is NP-hard.*

Remark 1.2. *Recall that this result is optimal, because the Goemans-Williamson algorithm matches the above parameters (without the* ϵ*). In particular, by taking* $\rho = \rho^* \approx -.69$*, we can conclude that* $.878$*-approximating MAX-CUT is NP-hard.*

To begin the proof of Theorem 1.1, let τ and C be the parameters we get from Majority Is Stablest, using ρ and $\epsilon/2$. With these in hand, define
$$\delta = \frac{\epsilon\tau^2}{8C}.$$

Now by taking $m := |\Sigma|$ large enough, we can use the Unique Games Conjecture to assert that Gap-Unique-Label-Cover$(\Sigma)_{1-\delta,\delta}$ is NP-hard. We will now give a reduction from this Gap-ULC problem to Gap-MAX-CUT.

First try at a reduction. The instance of ULC is a bipartite graph G with vertex set V and W. The "first try" one would naturally think of for a PCP/reduction for MAX-CUT is the following: First, expect a long-code of the label of each vertex in V in W in the "proof". Equivalently, associate to each vertex v a block of vertices identified with $\{-1,1\}^m$, so that a cut in the resulting graph yields a boolean function $f_u : \{-1,1\}^m \to \{-1,1\}$ for each $u \in V \cup W$. The natural try for a test here (i.e., a natural definition for the edges in the graph we construct) is:

- Pick an edge (v, w) in G at random. Call the supposed long-codes of the vertices' labels f_v and f_w.
- Let π denote the permutation given by the constraint on the edge.
- Pick $x \in \{-1,1\}^m$ uniformly at random, and pick $\mu \in \{-1,1\}^m$ according to the $\frac{1-\rho}{2}$-biased distribution.
- Test $f_v(x) \neq f_w(x\mu \circ \pi)$.

Here $x \circ \mu$ denotes the string $(x_{\sigma(1)}, \ldots, x_{\sigma(m)})$.

However, there is an immediate flaw with this reduction — the graph constructed is bipartite! In particular, a "cheating prover" could make all of the f_v functions for $v \in V$ identically 1 and all of the f_w functions for $w \in W$ identically -1. This obviously does not correspond to any useful labeling of the original Label-Cover graph, and yet it passes the test with probability 1!

To get around this problem, we use a slightly more subtle reduction.

The actual reduction. In our actual reduction we will only require long-codes for the W vertices; we will use these to, in a way, "infer" supposed long-codes for the V vertices.

The $\neq$ test we will use is the following:

- Pick a vertex $v \in V$ uniformly at random and *two* random edges based at v, say (v, w) and (v, w'). Let $f_w, f_{w'} : \{-1,1\}^m \to \{-1,1\}$ denote the supposed long-codes for w and w'.
- Let π and π' be the permutations associated with the constraints on the two edges (v, w) and (v, w').
- Pick $x \in \{-1,1\}^m$ uniformly at random, and pick $\mu \in \{-1,1\}^m$ according to the $\frac{1-\rho}{2}$-biased distribution.
- Test $f_w(x \circ \pi) \neq f_{w'}((x\mu) \circ \pi')$.

We will now analyze the completeness and soundness of this test.

2 Completeness of the reduction

Suppose the Unique-Label-Cover instance has an assignment $\sigma : (V \cup W) \to \Sigma$ satisfying a $(1-\delta)$-fraction of the edge-constraints in G. Now suppose each f_w is a proper long-encoding of $\sigma(w)$; i.e., it is the $\sigma(w)$th dictator function.

The test generates two edges (v, w) and (v, w'). By the left regularity of the graph G, each edge is, individually, a uniformly random edge. Hence by the union bound, with probability $\geq 1 - 2\delta$, σ satisfies the constraint on both edges.

So suppose both edges are satisfied by σ. What is the probability the test succeeds? By definition of f_w, we have

$$f_w(x \circ \pi) = (x \circ \pi)_{\sigma(w)} = x_{\pi(\sigma(w))}.$$

Since σ satisfies the edge (v, w), we have $\pi(\sigma(w)) = \sigma(v)$. Similarly, we have

$$f_{w'}((x\mu) \circ \pi') = ((x\mu) \circ \pi')_{\sigma(w')} = (x\mu)_{\pi'(\sigma(w'))},$$

and $\pi'(\sigma(w')) = \sigma(v)$ as well. So the test ends up testing

$$x_{\sigma(v)} \neq (x\mu)_{\sigma(v)}.$$

With the random choice of x and μ, this happens iff $\mu_{\sigma(v)} = -1$. This happens with probability precisely $\frac{1-\rho}{2}$.

Overall, we can conclude that the probability the test succeeds is at least $(1-2\delta) \cdot \frac{1-\rho}{2} \geq \frac{1-\rho}{2} - \epsilon$ (using the fact that $\epsilon > 2\delta$).

This completes the proof of the completeness claimed in Theorem 1.1.

3 Soundness of the reduction

As usual, we will prove the contrapositive of the soundness condition. To that end, suppose that the supposed long-codes $\{f_w\}$ are such that the test passes with probability greater than $\frac{\cos^{-1}\rho}{\pi} + \epsilon$. We will show how to "decode" these supposed long-codes into an assignment $\sigma : (V \cup W) \to \Sigma$ for the original ULC graph G which satisfies at least a δ fraction of the constraints.

The first step is the standard "averaging" argument, with respect to the initial random choice of $v \in V$: Specifically, since the overall test passes with probability at least $\frac{\cos^{-1}\rho}{\pi} + \epsilon$, there must be at least an $\epsilon/2$ fraction of the v's such that the test, conditioned on v, passes with probability at least $\frac{\cos^{-1}\rho}{\pi} + \epsilon/2$. (For otherwise, the test would succeed with probability less than $(\epsilon/2) \cdot 1 + (1-\epsilon/2) \cdot (\frac{\cos^{-1}\rho}{\pi} + \epsilon/2) < \frac{\cos^{-1}\rho}{\pi} + \epsilon$.) Let us call such v's *good*.

The next step is to write down the test's success probability. For this purpose, let us imagine that in the first step of the test we have picked a good v. (This happens with probability $\geq \epsilon/2$.) The subsequent success probability is then:

$$\begin{aligned} & \mathop{\mathbf{E}}_{w,w'}[\mathop{\mathbf{E}}_{x,\mu}[\tfrac{1}{2} - \tfrac{1}{2} f_w(x \circ \pi) f_{w'}((x\mu) \circ \pi')]] \\ = \quad & \tfrac{1}{2} - \tfrac{1}{2} \mathop{\mathbf{E}}_{x,\mu}[\mathop{\mathbf{E}}_{w,w'}[f_w(x \circ \pi) f_{w'}((x\mu) \circ \pi')]]. \end{aligned}$$

We now make the following definition:

Definition 3.1. *For every $v \in V$, define the function $g_v : \{-1,1\}^m \to [-1,1]$ by setting*

$$g_v(y) = \mathop{\mathbf{E}}_{w \sim v}[f_w(x \circ \pi_{v,w})].$$

Here w is chosen to be a random neighbor of v, and $\pi_{v,w}$ is the bijection on the edge-constraint (v,w).

The idea behind this definition is that if all the f_w's were proper long-codes of labels, each of which was consistent with a label for v, then g_v would be the long-code of that consistent label. Continuing our above calculation, we have that the success probability of the test, conditioned on having chosen v first, is

$$\tfrac{1}{2} - \tfrac{1}{2} \mathop{\mathbf{E}}_{x,\mu}[g_v(x) g_v(x\mu) = \tfrac{1}{2} - \tfrac{1}{2}\mathrm{Stab}_\rho(g_v).$$

We are now in good shape. Assuming v is a good vertex, we have that the above quantity is at least $\frac{\cos^{-1}\rho}{\pi} + \epsilon/2$. But then we can apply the MIS Theorem to show that g_v must have at least one coordinate with large, C-degree influence.

In particular, at a more intuitive level, g_v must look vaguely like some jth dictator. We will decode v to that label, j. Further, note that g_v is an average of supposed long-codes (after appropriate permutations). If g_v looks a little like the jth dictator, then we can show it must be that at least a small fraction of the w's neighboring v look a little like the $\pi^{-1}_{v,w}(j)$th dictator. Hence, if we list-decode all f_w's to the set of dictators they vaguely resemble, and then pick at random from this list (we will also show this list is small), then there's at least a slight chance of getting $\pi^{-1}(j)$.

Let us make this intuition rigorous. We know from the MIS Theorem that for every good v, there is at least one coordinate $j = j_v$ such that

$$\mathrm{Inf}_j^{\leq C}(g_v) \geq \tau. \tag{1}$$

We will set $\sigma(v) = j_v$ in the overall assignment σ we are constructing. As for the w's in W, since $\mathrm{Inf}_j^{\leq C}(g_v)$ refers to the Fourier expansion of g_v, let's work out what this is. We have $g_v = \mathrm{avg}_{w \sim v} f_w \circ \pi_{v,w}$. Now

$$\begin{aligned} f_w &= \sum_S \hat{f}_w(S)\chi_S \\ \Rightarrow f_w \circ \pi &= \sum_S \hat{f}_w(S)(\chi_S \circ \pi) \\ &= \sum_T \hat{f}_w(\pi^{-1}(T))\chi_T. \end{aligned}$$

Therefore, we conclude that

$$g_v = \sum_T \left(\mathop{\mathrm{avg}}_w \hat{f}_w(\pi^{-1}_{v,w}(T)) \right) \chi_T.$$

Having computed the Fourier expansion of g_v, let's examine (1):

$$\begin{aligned}
\tau &\leq \mathrm{Inf}_j^{\leq C}(g_v) \\
&= \sum_{|S|\leq C, S\ni j} \hat{g}_v(S)^2 \\
&= \sum_{|S|\leq C, S\ni j} \left(\mathop{\mathbf{E}}_w[\hat{f}_w(\pi^{-1}(S))]\right)^2 \\
&\leq \sum_{|S|\leq C, S\ni j} \mathop{\mathbf{E}}_w[\hat{f}_w(\pi^{-1}(S))^2] \qquad \text{(Cauchy-Schwarz)} \\
&= \mathop{\mathbf{E}}_w[\sum_{|S|\leq C, S\ni j} \hat{f}_w(\pi^{-1}(S))^2] \\
&= \mathop{\mathbf{E}}_w[\mathrm{Inf}_{\pi^{-1}(j)}^{\leq C}(f_w)].
\end{aligned}$$

Thus by another averaging argument, we can conclude that at least a $\tau/2$ fraction of v's neighbors w have $\mathrm{Inf}_{\pi_{v,w}^{-1}(j)}^{\leq C} \geq \tau/2$.

We now want to "list-decode" each f_w into

$$\mathcal{S}_w := \{k : \mathrm{Inf}_k^{\leq C}(f_w) \geq \tau/2\}$$

and choose $\sigma(w)$ randomly from this set. Assuming this set is always not too large — say, at most R in size — then we're happy. The reason is that in this case, the expected fraction of ULC constraints σ satisfies will be at least

$$\frac{\epsilon}{2} \cdot \frac{\tau}{2} \cdot \frac{1}{R}.$$

The justification of this is that we imagine that both σ is chosen at random and that (v, w) is chosen as a random edge. The first term above is the probability that v is good. The second term is the probability that w is one of the neighbors of v with C-degree influence in the $\pi^{-1}(j_v)$ coordinate of at least $\tau/2$. The final term is the probability that the $\pi^{-1}(j_v)$ is chosen out of $\mathcal{S}_w$ in the construction of σ.

Since we took $\delta = \epsilon\tau^2/(8C)$, we can show that there exists a labeling σ satisfying at least a δ fraction of the constraints (and thus complete the proof) as soon as we can show $R \leq (2C)/\tau$.

This follows immediately from the following easy lemma:

Lemma 3.2. *For any* $h : \{-1,1\}^m \to \{-1,1\}$,

$$\#\{k : \mathrm{Inf}_k^{\leq C}(h) \geq \eta\} \leq C/\eta.$$

Proof. The following stronger inequality completes the proof:

$$\sum_{i=1}^m \mathrm{Inf}_i^{\leq C}(h) = \sum_{i=1}^m \sum_{|S|\leq C, S\ni i} \hat{h}(S)^2 = \sum_{|S|\leq C} |S|\hat{h}(S)^2 \leq C \sum_S \hat{h}(S)^2 = C.$$

□