,22

CMPUT 675: Approximation Algorithms **Fall 2011**

Lecture 21,22 (Nov 22&24, 2011): Label Cover, Hardness of Set Cover

Lecturer: Mohammad R. Salavatipour *Scribe: based on older notes*

Our goal from now (until the end of the course) is to present an $\Omega(\log n)$-hardness result for set cover. To achieve this goal we need to define a few other problems and prove hardness results for them. The first one is a variation of Max-3SAT.

21.0.1 Gap-Max-3SAT(5) problem

Given a 3SAT formula with the extra restriction that every variable belongs to 5 clauses.

Goal: To find the maximum number of clauses satisfied by an assignment.

Theorem 1 *There is a gap preserving reduction form Max-3SAT to Max-3SAT(5).*

Given a Max-3SAT(5) instance ϕ let $Opt(\phi)$ denote the number of clauses that can be satisfied by a truth assignment. Then it is NP-hard to decide if:

- $opt(\phi) = m$
- $opt(\phi) \leq (1-\epsilon)m$ for some constant $\epsilon > 0$.

21.0.2 Label Cover Problem

The label cover problem is a graph theoric modeling of a 2-pover 1-round proof system for NP. An instance of label cover consists of the followings:

- $G(V \cup W, E)$ is a bipartite graph.
- $[N] = \{1...N\}$, $[M] = \{1...M\}$ are 2 sets of labels, $[N]$ for the vertices in V and $[M]$ for the vertices in W.
- $\{\Pi_{v,w}\}_{(v,w)\in E}$ denotes a (partial) function on every edge (v,w) such that $\Pi_{v,w} : [M] \rightarrow [N]$

A labeling $l : V \rightarrow [N], W \rightarrow [M]$ is said to cover edge (v,w) if $\Pi_{v,w}(l(w)) = l(v)$.

Goal: Given an instance of label cover, find a labeling that covers maximum fraction of the edge.

Theorem 2 *Given an instance $\mathcal{L}(G, M = 7, N = 2, \{\Pi_{v,w}\})$ it is NP-hard to decide if*

- $opt(\mathcal{L}) = 1$, *or*
- $opt(\mathcal{L}) \leq 1 - \epsilon$

Proof. We use Theorem 1. Given a Max-3SAT(5) formula ϕ, we construct an instance $\mathcal{L}$ as follows.

Let the variables of ϕ be $\{x_1, \ldots, x_n\}$ and clauses be $\{C_1, \ldots, C_m\}$.

Define $V = \{x_1 ... x_n\}$, and $W = \{C_1, \ldots, C_m\}$, i.e. create a vertex in V for every variable of ϕ and a vertex in W for every clause of ϕ. Two vertices x_i and C_j are adjacent iff $x_i \in C_j$. Note that the degree of every vertex in V is 5, because each variable is in 5 clauses and the degree of every vertex in W is 3.

For $\Pi_{v,w}$: V gets labels from $\{0, 1\}$ and for every clause $C_j \in V$ let [7] be the set of seven satisfying assignments that a clause C_j can have. Then, $\Pi_{x_i,C_j} : [7] \rightarrow [2]$ is basically the bijection of the assignment of variable x_i in the given satisfying assignment of C_j.

Example: if $C_j = x_1 \vee x_2 \vee x_3$, then $\Pi_{x_1,C_j}(101) = 1$ and $\Pi_{x_2,C_j}(101) = 0$

If ϕ is satisfiable (i.e. a Yes instance), then there is a truth assignment satisfying all clauses. Consider the labeling of x_i's defined by this truth assignment and also every clause (of course satisfied) has one of those seven labels and these labels are consistent on every edge. This means, all the edges of $\mathcal{L}$ are covered. So $opt(\mathcal{L}) = 1$.

Now assume that ϕ is a "no" instance (i.e. $opt(\phi) \leq (1 - \epsilon)m$. Consider any solution to $\mathcal{L}$. Labels on V give a truth assignment to the variables. This means, the fraction of satisfied clauses of ϕ by this truth assignment is at most $(1 - \epsilon)m$. Consider any unsatisfied clauses $C_j = x_1 \vee x_2 \vee x_3$ by this truth assignment. Any label (satisfying truth assignment) to C_j is inconsistent with the truth assignment to at least one of its variables (or else clause C_j is satisfied). Therefore, at least one of the the 3 edges incident with C_j must not be covered. This implies that at last $\frac{\epsilon}{3}$ fraction of edges are not covered. Thus $Opt(\mathcal{L}) \leq (1 - \frac{\epsilon}{3})$. ■

21.1 2P1R Proof Systems

Raz's verifier is a 2 Prover 1-round proof system for a language L with parameters c and s (where c is usually 1 and s is usually 1-ϵ) is a probabilistic verifier V with access to two proofs Π_1 and Π_2 such that an input y for L, V sends one query to each of Π_1 and Π_2 and:

- if $y \in L \rightarrow \exists \Pi_1$ and Π_2 such that $Pr[V \text{ accepts}] = c$
- if $y \notin L \rightarrow \forall \Pi_1$ and Π_2 such that $Pr[V \text{ accepts}] \leq s$

The PCP Theorem shows that for every $L \in$ NP, there is a 2P1R with $c = 1$,$s = 1 - \delta$ for some $\delta > 0$.

Every problem in NP can be reduced to MAX-3SAT. We construct a 2P1R proof system with the above parameters for MAX-3SAT. Given a formula ϕ, the proofs Π_1 and Π_2 are supposed to encode a truth assignment to ϕ. For every variable $x_i \in \phi$, the value of $\Pi_1[i] \in \{0, 1\}$ is the value of x_i. For every $C_j \in \phi$, $\Pi_2[j] \in \{1, \ldots, 7\}$ is one of seven satisfying assignments for C_j. V picks a random clause C_j and a random variable, say x_i, from that clause accepts if and only if $\Pi_1[i]$ is consistent with $\Pi_2[j]$.

- if ϕ is a "yes" instance $\rightarrow$ proofs Π_1 and Π_2 form a satisfying truth assignment $\rightarrow$ V accepts with probability 1
- if ϕ is a "no" instance $\rightarrow$ at most $(1 - \epsilon)m$ clauses can be satisfied $\rightarrow$ there is a probability of at least $\frac{\epsilon}{3}$ that the answers from Π_1 and Π_2 are inconsistent $\rightarrow$ V accepts with probability $< 1 - \frac{\epsilon}{3}$ (where $\frac{\epsilon}{3} = \delta$)

Can we amplify this probability by repetition?

A k-repetition for this 2P1R proof system is as follows: verifier V^k chooses k clauses (randomly) and a variable (randomly) from each. We have proof entries $\Pi_1[i_1 \ldots i_k] \in \{0, 1\}^k$ (representing assignments to k-tuples of variables $i_1 \dots i_k$) and $\Pi_2[j_i \dots j_k] \in \{1, \dots, 7\}^k$ (representing satisfying assignments to k clauses $C_{j_1} \dots C_{j_k}$). V^k accepts if and only if all answers are consistent.

This corresponds to the following repetition of label cover: from an instance $\mathcal{L}(\mathcal{G}(\mathcal{V}, \mathcal{W}, \mathcal{E}), [\mathcal{M}], [\mathcal{N}], \{\Pi_{vw}\})$, we build $\mathcal{L}^k(G'(V', W', E'), [M'], [N'], \{\Pi\text{'}_{vw}\})$ where:

- $V' = V^k$ (k-tuples of V)
- $W' = W^k$
- $[M]' = [M]^k$
- $[N]' = [N]^k$
- $(V', W') \in E' \Leftrightarrow (v_{i_j}, w_{i_j}) \in E,\ \forall i, j\ 1 \le j \le k\ (V' = (v_{i_1}, \dots, v_{i_k}), W' = (w_{i_1}, \dots, w_{i_k}))$
- $\Pi'_{vw}(b_1, \dots, b_k) = \Pi_{v_{i_1}, w_{i_1}}(b_1), \Pi_{v_{i_2}, w_{i_2}}(b_2), \dots, \Pi_{v_{i_k}, w_{i_k}}(b_k)$
- if $\mathrm{OPT}(\mathcal{L}) = 1 \rightarrow \mathrm{OPT}(\mathcal{L}^k) = 1$

We expect that if $\mathrm{OPT}(\mathcal{L}) \le 1 - \delta$ then $\mathrm{OPT}(\mathcal{L}^k) \le (1 - \delta)^k$, but this is not true.

Theorem 3 (Raz 1998) *Parallel Repetition Theorem*
if $OPT(\mathcal{L}) \le 1 - \delta \rightarrow OPT(\mathcal{L}^k) \le (1 - \delta)^{\Omega(k)}$
i.e. if ϕ is a no instance of SAT, V^k accepts with probability $2^{-\Omega(k)}$

Note: $[M'] = [7^k]$ and $[N'] = [2^k]$

Theorem 4 *There is a reduction from SAT to an instance $\mathcal{L}(G(V, W, E), [7^k], [2^k], \{\Pi_{vw}\})$ of label cover such that:*

- *if ϕ is a yes instance $\rightarrow$ $OPT(\mathcal{L}) = 1$*
- *if ϕ is a no instance $\rightarrow$ $OPT(\mathcal{L}) = 2^{-ck}$ for some constant $c < 1$*

and $\mathcal{L} = n^{O(k)}$

Corollary 1 *if NP is not a subset of $O(n^{poly\ log(n)})$ then there is no $O(2^{log^{1-\epsilon} n})$-approximation for labelcover for any $\epsilon > 0$.*

Today we complete the proof of $\Omega(\log n)$-hardness of set cover. In the last lecture, we introduced the following theorem:

Theorem 5 *There is a reduction mapping an instance ϕ of SAT to an instance $L(G(V, W, E), M = [7^k], N = [2^k], \{\Pi_{v,w}\})$ of label cover, such that:*

- *If ϕ is a yes instance, then opt$(L) = 1$.*
- *If ϕ is a no instance, then opt$(L) \le 2^{-\delta k}$, for some $\delta > 0$, $|L| = n^{O(k)}$.*

To prove the hardness of approximation of set cover, we need the following set system.

Definition 1 *A set system $(U, C_1, \cdots, C_m, \overline{C_1}, \cdots, \overline{C_m})$ with parameters m and l, where U is the a universe of elements of size $O(l \cdot logm \cdot 2^l)$ and $C_1, \cdots, C_m$ are subsets of U. This set system has the property that any collection of $\leq l$ subsets from $C_i's$ that cover U must contain a set and its complement.*

There are explicit construction of such (m, l)-set systems. There are also easy probabilisitic constructions. Consider a label cover instance $\text{Ł}(G(V, W, E), M = [7^k], N = [2^k], \{\Pi_{v,w}\})$. We can assume $|V| = |W|$ (e.g. if not, we can create copies of the vertices in V with the same neighbours). We build an instance of set cover § such that:

- If $opt(\text{Ł}) = 1$, then $opt(S) \leq |V| + |W|$.
- If $opt(\text{Ł}) < \frac{2}{l^2}$, then $opt(\S) > \frac{l}{16}(|V| + |W|)$.

Consider a set system with $m = N = 2^k$ and l to be specified later. For every edge $e = (v, w) \in G$, we have a (disjoint) (m, l)-set system with universe U_e. Let $C_1^{vw}, \cdots, C_{N=m}^{vw}$ be the subsets of U_e. The union of all $U_e's$ (for all the edges e) is the universe of the set cover instance, denoted as

$$U = \bigcup_{(v,w) \in G} U_{vw}.$$

Now we define the subsets in our set cover instance. For every $v \in V$ ($w \in W$) and every label $i \in [2^k]$ ($j \in [7^k]$), we have a set

$$S_{v,i} = \bigcup_{w:(v,w) \in E} C_i^{vw} \qquad S_{w,j} = \bigcup_{v:(v,w) \in E} \overline{C_{\Pi_{vw}(j)}^{vw}}$$

This completes the construction of § from Ł.

Lemma 1 *If $opt(L) = 1$, then $opt(\S) \leq |V| + |W|$.*

Proof. Consider an optimal labeling $l : V \rightarrow [2^k], W \rightarrow [7^k]$ for Ł. Because it is covering every edge $(v, w) \in E$, $\Pi_{vw}(l(w)) = l(v)$. This labeling defines a label for every vertex and every pair of vertex/label corresponds to a set in §. From $C_{l(v)}^{vw} \subseteq S_{v,l(v)}$ and $\overline{C_{l(v)}^{vw}} = \overline{C_{\Pi_{vw}(l(w))}^{vw}} \subseteq S_{w,l(w)}$, we have that $S_{v,l(v)} \cup S_{w,l(w)} \supseteq U_{vw}$. Because all U_e's for $e \in E$ are covered, U is covered. So we have a set cover of size $|V| + |W|$. ∎

Lemma 2 *if $opt(\S) \leq \frac{l}{16}(|V| + |W|)$, then $opt(L) \geq \frac{2}{l^2}$.*

Proof. From the set cover solution, we assign labels (maybe more than one label) to the vertices. If $S_{v,i}$ is in the solution, v gets label i. Since there are at most $\frac{l}{16}(|V| + |W|)$ sets and $|V| + |W|$ vertices, the average number of labels per vertex is $\leq \frac{l}{16}$. We discard vertices with more than $\frac{l}{2}$ labels. Afterwords, at most $\frac{|V|}{4}$ vertices from each side of V and W are discarded. Let V' and W' be the vertices remaining in V and W respectively. Then, $|V'| > \frac{3}{4}|V|$ and $|W'| > \frac{3}{4}|W|$. Pick an edge $e = (v, w)$ from G randomly. Then $\Pr[v \in V' and w \in W'] \geq 1 - (\frac{1}{4} + \frac{1}{4}) = \frac{1}{2}$. This means at least half of the edges of G are between V' and W'. Let $T_v = \{S_{v,i} : i\, is\, a\, label\, of\, v\}$ and $T_w = \{S_{w,j} : j\, is\, a\, label\, of\, w\}$. We have $|T_v| \leq \frac{l}{2}$ and $|T_w| \leq \frac{l}{2}$. Note that sets in $T_v \cup T_w$ cover U_{vw}. To be more precise, sets in $X_1 = \{C_i^{vw} : i\, is\, a\, label\, of\, v\} \cup X_2 = \{\overline{C_{\Pi_{vw}(j)}^{vw}} : j\, is\, a\, label\, of\, w\}$ cover universe U_{vw} ($|X_1 \leq \frac{l}{2}$ and $|X_2 \leq \frac{l}{2}$). Because U_{vw} is covered by at most l sets (i.e. $|X_1| + |X_2| \leq l$), there must be a set $C_i^{vw} \in X_1$ and $\overline{C_{\Pi_{vw}(j)}^{vw}} \in X_2$, such that they are complement, i.e. $i = \Pi_{vw}(j)$. Because we pick labels of v and w randomly, with probability $(\frac{2}{l})^2 = \frac{4}{l^2}$ we have set C_i^{vw} for v and $\overline{C_j^{vw}}$ for w, i.e. the labels i for v and j for w cover edge $e \in E$. Thus the expected fraction of edges between V' and W' that are covered is $\geq \frac{4}{l^2}$. Therefore, at least a fraction of $\frac{2}{l^2}$ of edges of G are covered. ∎

This lemma is equivalent to saying that if $opt(\text{\L}) < \frac{2}{l^2}$ then $opt(\S) > \frac{l}{16}(|V| + |W|)$.

Let $l \in \Theta(2^{\frac{\delta k}{2}})$. Then, $l^2 \in \Theta(2^{\delta k})$. We get a hardness of $\Omega(l)$ for $\S$. The size of $\S$ is $n^{O(k)} \cdot O(l \cdot \log m \cdot 2^l)$. If $k = c \log\log n$ for sufficiently large c, $l = O(2^{O(\log\log n)}) \geq \log n \log\log n$. $\log|\S| = O(\log\log n \cdot \log n + \log l + \log\log\log n + l) = \Theta(l)$.

We have the following hardness result for set cover:

Theorem 6 *Unless $NP \subseteq DTIME(n^{O(\log\log n)})$, set cover has no $\Omega(\log n)$-approximation algorithm.*