Advanced Approximation Algorithms **(CMU 15-854B, Spring 2008)**

Lecture 9: Hardness of Max-Ek-Indep.-Set and A.-k-Center

February 12, 2008

Lecturer: Ryan O'Donnell, Anupam Gupta *Scribe: Eric Blais*

In this lecture, we complete the proof of hardness of approximation for the Max-Ek-Hypergraph-Indepedent-Set problem. We also give a proof of hardness of approximation for the Asymmetric k-Center problem.

1 Max-Ek-Hypergraph-Independent-Set

In this section, we complete the proof of hardness of approximation for the Max-Ek-Hypergraph-Independent-Set problem.

Theorem 1.1. *For all constant even-valued $k \geq 4$ and all constants $\epsilon, \delta > 0$, the Weighted Max-Ek-Hypergraph-Independent-Set $(1 - \frac{2}{k} - \delta)$ vs. ϵ decision problem is NP-hard.*

The proof of Theorem 1.1 uses a reduction from the Label-Cover(K,L) problem. Given an instance $\mathcal{G} = (U, V, E)$ of the Label-Cover(K,L) problem, we construct a k-regular hypergraph $\mathcal{H}$. For each vertex $v \in V$, we add a corresponding "block" $\{0,1\}^L_v$ of 2^L vertices in $\mathcal{H}$. We assign a p-biased weight to each vertex, with $p = 1 - \frac{2}{k} - \delta$. For each pair of edges $(u, v), (u, v') \in E$, we add a hyperedge on the set of vertices $\{A^v_1, \ldots, A^v_{k/2}, B^{v'}_1, \ldots, B^{v'}_{k/2}\}$ if and only if $\pi_{v \to u}(\bigcap_{i=1}^{k/2} A^v_i)$ and $\pi_{v' \to u}(\bigcap_{i=1}^{k/2} B^{v'}_i)$ are disjoint sets of keys. For more details on the reduction and the intuition behind it, see Lecture 8.

The reduction takes polynomial time. In the previous lecture, we used the reduction to prove the completeness of Theorem 1.1.

Lemma 1.2 (Completeness)**.** *Under the above reduction, if* $\mathsf{Opt}(\mathcal{G}) = 1$ *then* $\mathsf{Opt}(\mathcal{H}) \geq p - \frac{2}{k} - \delta$.

Proof. See Section 3 of Lecture 8. □

In order to complete the proof of Theorem 1.1, all that remains is to show the soundness of the theorem.

1.1 Proof of soundness

We will use the following tool to prove the soundness of Theorem 1.1.

Theorem 1.3. *For $t = t(s, \epsilon, \delta) = O(\frac{1}{\delta^2} \log(\frac{1}{s\epsilon\delta}))$, if $\mathcal{F} \subseteq \{0,1\}^L$ has p-biased weight $\geq \epsilon$, where $p = 1 - \frac{1}{s} - \delta$ then $\mathcal{F}$ is too big to be s-wise t-intersecting.*

Proof. See Homework 3. □

Lemma 1.4 (Soundness)**.** *Under the above reduction, if* $\mathsf{Opt}(\mathcal{G}) < \frac{\epsilon}{2t(\frac{k}{2},\frac{\epsilon}{2},\delta)^2}$*, then* $\mathsf{Opt}(\mathcal{H}) < \epsilon$*.*

Define $\eta := \epsilon/2t(\frac{k}{2},\frac{\epsilon}{2},\delta)^2$. Note that $\eta > 0$ and that η does not depend on $|K|$ or $|L|$. Take $|K|, |L| \le \mathrm{poly}(\frac{1}{\eta})$.

Proof. We prove the contrapositive of Lemma 1.4. Let $\mathcal{I}$ be an independent set in $\mathcal{H}$ with weight at least ϵ. Let $\mathcal{F}_v = \mathcal{I} \cap \{0,1\}_v^L$. By an averaging argument, at least an $\epsilon/2$ fraction of the vertices in V have weight $wt(\mathcal{F}_v) \ge \epsilon/2$. Call these vertices the "good" vertices.

Let $t = t(\frac{k}{2},\frac{\epsilon}{2},\delta)$. When v is a good vertex, Theorem 1.3 tells us that $\mathcal{F}_v$ is too big to be $\frac{k}{2}$-wise t-intersecting. Therefore, there exist $A_1^v, \dots, A_{k/2}^v \in \mathcal{F}_v$ such that $|\bigcap_{i=1}^{k/2} A_i^v| < t$. For every good vertex v, define $\mathrm{Sugg}(v) = \bigcap_{i=1}^{k/2} A_i^v$. We claim that any two good vertices v, v' that share a common neighbor have consistent sets of suggestions:

Claim 1.5. *If* $(u,v),(u,v') \in E$ *and* v, v' *are good vertices, then*

$$\pi_{v\to u}(\mathrm{Sugg}(v)) \cap \pi_{v'\to u}(\mathrm{Sugg}(v')) \neq \emptyset \ .$$

Proof. By contradiction. Assume that $\pi_{v\to u}(\mathrm{Sugg}(v))$ and $\pi_{v'\to u}(\mathrm{Sugg}(v'))$ are disjoint. Then $\{A_1^v, \dots, A_{k/2}^v, B_1^{v'}, \dots, B_{k/2}^{v'}\}$ is a hyperedge. But all A's and B's are in $\mathcal{I}$, so they form an independent set. Contradiction! □

Having defined "goodness" for vertices in V, let us now extend the definition for vertices in U. Specifically, we say that a vertex $u \in U$ is "good" if it has a good neighbor: $u \in U$ is a good vertex if there exists a good vertex $v \in V$ such that $(u,v) \in E$. For each good vertex $u \in U$, define the "buddy" $b(u) \in V$ of u to be one of the good vertices in V adjacent to u (choosing arbitrarily when there are multiple possible choices).

We now randomly construct the labeling function $f : V \to L, U \to K$ according to the following rules:

- For good v's, choose $f(v)$ randomly from $\mathrm{Sugg}(v)$.
- For good u's, choose $f(u)$ randomly from $\pi_{b(u)\to u}(\mathrm{Sugg}(b(u)))$.
- For the rest of the vertices, choose the key or label arbitrarily.

To complete the proof, we want to show that $\mathbf{Pr}_{(u,v)\in E,f}[f \text{ satisfies } \pi_{v\to u}] \ge \epsilon/2t^2$. Note that instead of choosing an edge $(u,v) \in E$ randomly, we can first choose $v \in V$ randomly, then choose a random $u \in U$ uniformly from the neighbors of v. The resulting edge (u,v) is still chosen uniformly from all edges by the right-regularity of $\mathcal{G}$. When we pick $v \in V$ randomly, v will be good with probability at least $\epsilon/2$. Then any neighbor $u \in U$ of v that we select is good too. By Claim 1.5, there exists a label key $a \in \pi_{v\to u}(\mathrm{Sugg}(v)) \cap \pi_{b(u)\to u}(\mathrm{Sugg}(b(u)))$. With probability at least $1/t$, v is assigned the label $f(v)$ such that $\pi_{v\to u}(f(v)) = a$, and with probability at least $1/t$, u is assigned the key $f(u) = a$. So the probability that f satisfies $\pi_{v\to u}$ is at least $\epsilon/2 \cdot 1/t \cdot 1/t = \epsilon/2t^2$, as we wanted to show. □

1.2 Max-Ek-Hypergraph-Vertex-Cover

For the proof of hardness of the Asymmetric k-Center problem that we will cover in the next section, we are particularly interested in a special case of Theorem 1.1.

Corollary 1.6. *For all constant even-valued $k \geq 4$, the Weighted Max-Ek-Hypergraph-Independent-Set $(1 - \frac{3}{k})$ vs. $\frac{1}{k}$ decision problem is NP-hard.*

Proof. Apply Theorem 1.1 with $\delta = \epsilon = 1/k$. □

Theorem 1.1 also gives a hardness result for the vertex cover problem.

Corollary 1.7. *For all constant even-valued $k \geq 4$, the Weighted Min-Ek-Hypergraph-Vertex-Cover $(1 - \frac{3}{k})$ vs. $\frac{1}{k}$ decision problem is NP-hard.*

Proof. This follows directly from Corollary 1.6 and the complementarity of the independent set and vertex cover problems. □

1.3 When k is not constant

Theorem 1.1 and its corollaries apply only when k is constant. When k is not constant, we can prove similar results but we need to keep track of the size of the problem instances after reductions.

Fact 1.8. *There is a reduction of the SAT problem to the Unweighted Min-Ek-Hypergraph-Vertex-Cover $\frac{3}{k}$ vs. $(1 - \frac{1}{k})$ decision problem in which an instance of the SAT problem of size n reduces to an instance of the Min-Ek-Hypergraph-Vertex-Cover problem with $|V| = n^{O(\log k)}$ vertices and $|E| \leq 2^{\mathrm{poly}(k)} \cdot |V|$ edges.*

We omit the technical details of the reduction.

2 Asymmetric k-Center

In Lecture 7, we introduced the Asymmetric k-Center problem. In the same lecture, we saw an $O(\log^* n)$-approximation algorithm for the problem. In this section, we give a matching lower bound of $\Omega(\log^* n)$.

2.1 Overview

We obtain a hardness result for the Asymmetric k-Center problem through by a reduction from the Min-Ek-Hypergraph-Vertex-Cover problem. In order to apply Fact 1.8, we will consider instances of the Min-Ek-Hypergraph-Vertex-Cover problem with $|M| = n^{O(\log k)}$ vertices and $|N| \leq |M| \cdot 2^{k^\beta}$ edges.

Recall that we can look at the vertex cover problem as a problem on a bipartite graph $G = (U, V, E)$: vertices in U represent vertices (or elements) in the original problem, vertices in V represent a k-hyperedge (or set) in the original problem, and the edge (u, v) is in E iff v contains

u in the original problem. The minimum vertex cover problem is now equivalent to finding the smallest subset $U' \subseteq U$ such that every vertex in V is covered by a vertex in U'.

The YES and NO instances of the Ek-Vertex-Cover problem that we use in the reduction are chosen to satisfy Corollary 1.7: the YES instances have a vertex cover of size at most $3/k \cdot |M|$, while the NO instances have no vertex cover of size less than $(1 - 1/k) \cdot |M|$. One observation that we use in the proof of hardness of the Asymmetric k-Center problem is that in NO instances, any vertex cover with fewer than $(1 - 2/k) \cdot |M|$ vertices leaves a large number of hyperedges uncovered.

Claim 2.1. *In a NO instance of the Ek-Vertex-Cover problem, any set of vertices of size at most* $(1 - 2/k) \cdot |M|$ *leaves at least* $1/k2^{k^\beta} \cdot |N|$ *hyperedges uncovered.*

Proof. By contradiction. Suppose on the contrary that more than $(1 - 1/k2^{k^\beta}) \cdot |N|$ edges are covered by a set of $(1 - 2/k) \cdot |M|$ vertices. Cover the remaining edges with one vertex each. This process requires fewer than $1/k2^{k^\beta} \cdot |N| \leq |M|/k$ vertices. Thus, with less than $(1 - 2/k + 1/k) \cdot |M| = (1 - 1/k) \cdot |M|$ vertices, all the edges are covered. This contradicts the fact that the instance is a NO instance. □

Let us define $f(k) = k2^{k^\beta}/2$. Then we can restate our claim as saying that in NO instances, any set of $1 - 2/k$ vertices covers at most $1 - 2/f(k)$ of the edges.

2.2 Reduction

We build instances of the Asymmetric k-Center problem by putting many copies of the Ek-Vertex-Cover problem in parallel. Specifically, we will build $t + 1$ layers of vertices $L_0, \ldots, L_t$, with layer L_i containing v_i vertices, each having d_i in-edges from vertices in layer L_{i-1}.[1] The layer L_0 will contain a single vertex (with no in-edges), and the vertices in layer L_1 all have a single in-edge from the vertex in L_0.

When we take YES instances of the Ek-Vertex-Cover problem, we can build an instance of the Asymmetric k-Center problem with the above construction for which $3v_i/d_i$ vertices from the layer L_i cover all vertices in layer L_{i+1}. So we can cover every vertex in the construction with $1 + \frac{3v_1}{d_1} + \cdots + \frac{3v_t}{d_t}$. Let us define $k := 1 + \frac{3v_1}{d_1} + \cdots + \frac{3v_t}{d_t}$.

For instances of the Asymmetric k-Center problem generated with NO instances of the Min-Ek-Vertex-Cover problem, we argue that when we pick any k centers, there is always at least one vertex at distance t of any center. In our construction, we can always t-cover all the vertices by picking the single vertex in L_0. So we are arguing that we can't do any better even if we include another k centers. For this argument, it suffices to show that it is impossible to $t - 1$ cover all the vertices in L_t with $k - 1$ vertices from L_1.[2] For the argument, assume that $k \leq (1 - 2/d_1) \cdot v_1$ and define $d_i = f(d_{i-1})$ for all $1 < i \leq t$. By Claim 2.1, at most a $1 - 2/f(d_1) = 1 - 2/d_2$ fraction

[1] Note that the vertices in layer L_i play two different roles in the two instances of the vertex cover problem that include L_i: in the instance of the problem between layers L_{i-1} and L_i, the vertices in L_i represent hyperedges, while in the instance of the problem between layers L_i and L_{i+1}, they represent the vertices in the original problem.

[2] To see this, consider any potential vertex v in L_k for $k > 1$ that we may include as one of our centers. This vertex has a parent $v' \in L_1$ that $(k-1)$-covers v. So any vertex that is $(t - k - 1)$-covered by v is $(t-1)$-covered by v'.

of the vertices in L_2 are covered by the centers in L_2. Applying the same claim once more, we then get that at most a $1 - 2/f(d_2) = 1 - 2/d_3$ fraction of the vertices in L_3 are 2-covered by the vertices in L_1. Repeating this argument for every layer, we eventually get that at most a $1 - 2/d_t$ fraction of the vertices in L_t are $(t-1)$-covered by any $k-1$ vertices in L_1. When $(1-2/d_t) < 1$, then at least one vertex in L_t is not covered by the vertices in L_1.

2.3 Analysis of the reduction

Our reduction gives YES instances of the Asymmetric k-Center problem for which each vertex is within distance 1 of a center and NO instances of the same problem for which some vertex is at distance t from any center. How large can t be in the NO instances? Let us examine the fraction of vertices covered by vertices from L_1 in each layer:

- The fraction of vertices in L_1 selected to be centers is $\leq 1 - \frac{2}{d_1}$.
- The fraction of vertices covered in L_2 is $\leq 1 - \frac{2}{f(d_1)} = 1 - 2/2^{d_1^\beta}$.
- The fraction of vertices covered in L_3 is $\leq 1 - \frac{2}{f(d_2)} = 1 - 2/2^{2^{d_1^\beta}}$.
- The fraction of vertices covered in L_4 is $\leq 1 - \frac{2}{f(d_3)} = 1 - 2/2^{2^{2^{d_1^\beta}}}$.
- ...
- The fraction of vertices covered in L_t is $\leq 1 - \frac{2}{f(d_{t-1})} = 1 - 2/2^{2^{\cdot^{\cdot^{\cdot^{2^{d_1^\beta}}}}}}$, where the tower of exponentials has height t.

So we can continue increasing the number t of layers until $2^{2^{\cdot^{\cdot^{\cdot^{d_1^\beta}}}}} \leq 1$, where the tower has height t. Thus, we can add a total of $t \simeq \log^* v_t + O(1)$ layers.

2.4 Technicalities

There is a small but important detail in the reduction that we have not yet considered. In the construction of the instances of the Min-Ek-Vertex-Cover problem, we provide the values of n and k, and the construction returns an instance with $M_{n,k}$ vertices (or left-vertices in the bipartite view of the problem) and $N_{n,k}$ hyperedges (right-vertices in the bipartite view of the problem). Let's say we use this instance for the ith layer of the construction of the Asymmetric k-Center problem instance. Then for the $(i+1)$th level, we give the value k' and obtain a new instance of the Min-Ek-Vertex-Cover problem of size $M'_{n,k'}$, $N'_{n,k'}$. For our construction to work out, we would like $M'_{n,k'} = N_{n,k}$. How can we enforce this condition?

One way around the problem is to take M' copies of the (M, N) construction and N copies of the (M', N') construction. The resulting structure satisfies $M'N = NM'$. This solves the

Furthermore, v' may also cover other vertices that are not covered by v, so we are better off including v' than v in our choice of centers. We also need to include the vertex in L_0 in our set of centers to cover the vertices in L_1.

matching-sizes problem, but we also need to show that Claim 2.1 still holds when we have duplicated instances of the Min-Ek-Vertex-Cover problem.

Claim 2.2. *When we have duplicated instance of the Min-Ek-Vertex-Cover problem, any set of at most a $(1-3/k)$ fraction of vertices leaves at least a $1/k^2 2^{k^\beta}$ fraction of the hyperedges uncovered.*

Proof. The proof uses an averaging argument, which we leave as an exercise to the reader. □

Lastly, to complete the proof of hardness for the Asymmetric k-Center problem, we need to use the duplication construction on all levels so that every adjacent layer has compatible sizes. This is done by using the appropriate parameters. For more details on this and other aspects of the proof, we refer the reader to the original article by Chuzhoy et al.[1].

References

[1] J. Chuzhoy, S. Guha, E. Halperin, S. Khanna, G. Kortsarz, R. Krauthgamer, and J. Naor. Asymmetric k-center is log* n-hard to approximate. *J. ACM*, 52(4):538–551, 2005.