Lecture 3: 1 vs $3/4 + \epsilon$ Hardness for Max-Coverage

Jan 22, 2008

Lecturer: Ryan O'Donnell   *Scribe: Ravishankar Krishnaswamy*

1 Outline

In this lecture, we prove that the 1 vs $3/4 + \epsilon$ decision problem of MAX-COVERAGE is NP-Hard. The proof is a reduction from the MAX-LABEL-COVER problem. Just to recap, let us begin by defining the problem instance.

The MAX-LABEL-COVER(K,L) Problem:

An instance to the MAX-LABEL-COVER(K,L) problem comprises of a bipartite graph $\mathcal{G}(U, V, E)$ with U and V being the two partitions, and a projection function $\pi_{v \to u} : L \to K$ on each edge (u, v). The objective is to obtain a labeling of the vertices (a function $f : U \to K, V \to L$) which satisfies maximum fraction of the edges. Edge (u, v) is satisfied by labeling f if and only if $f(u) = \pi_{v \to u}(f(v))$. Figure 1 describes such an instance.

The following theorem has been proved by Raz.

Theorem 1.1. *[3] Given any constant $\eta > 0$, there exists a constant $q = q(\eta)$ such that the 1 vs η decision problem for MAX-LABEL-COVER(K,L) is NP-Hard with $|K|, |L| \leq q$. Moreover, the constant q is actually bounded by $poly(1/\eta)$.*

Improved results on parallel repetition are proved by Holenstein [2] and Feige et al. [1]. Further, the hardness result holds (as we have seen in the first homework) even under the assumptions that $\mathcal{G}$ is *left and right regular* and $|U| = |V|$. We use this stronger result in our reduction to prove the soundness. Also, it is important to note that the size of the label sets K and L are bounded by a function which depends only on η. This means that for a fixed value of η, the sizes of K and L are at most a constant.

We now reduce an instance $\mathcal{G}$ of MAX-LABEL-COVER(K,L) to an instance $\mathcal{H}$ of MAX-COVERAGE such that

$$\begin{aligned} \mathsf{Opt}(\mathcal{G}) = 1 &\Rightarrow \mathsf{Opt}(\mathcal{H}) = 1 \text{ (Completeness)} \\ \mathsf{Opt}(\mathcal{G}) < \eta &\Rightarrow \mathsf{Opt}(\mathcal{H}) < 3/4 + \epsilon \text{ (Soundness)} \end{aligned}$$

where ϵ is a function of only η. We can see that once we prove completeness and soundness, we are through (by application of Theorem 1.1). Note that we have overloaded $\mathcal{G}$ to denote both the instance of the MAX-LABEL-COVER(K,L) problem, as well as the graph in the instance.
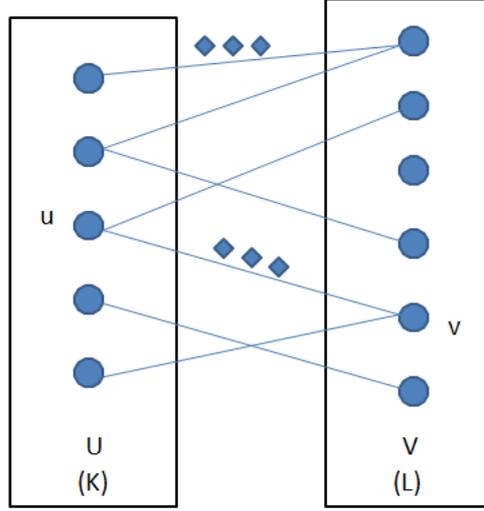
Figure 1: Max-Label-Cover Instance

2 The Reduction

Given an instance of the MAX-LABEL-COVER(K,L) problem (we are given a bipartite graph $\mathcal{G}(U, V, E)$), we create an instance $\mathcal{H}(\mathcal{E}, \mathcal{S}, m)$ of the MAX-COVERAGE problem, where $\mathcal{E}$ is the set of ground elements, $\mathcal{S}$ is the set system, and m is the number of sets we are allowed to pick. Henceforth, we let k and l denote the cardinality of sets K and L respectively.

The Ground Elements: For each edge $(u, v) \in E$, introduce a block of 2^k ground elements (the elements from the set $\{0, 1\}^k$). Each element is therefore a boolean string of length k. For being able to distinguish between ground elements corresponding to different blocks, we refer to the ground elements from the block corresponding to edge (u, v) by the set $\{0, 1\}^k_{(u,v)}$. Therefore, the total number of elements in $\mathcal{E}$ is $2^k|E|$. Once again, note that for every element $e \in \mathcal{E}$, we can associate a unique edge (u, v) to it. These elements are the diamonds in Figure 1.

Notation 2.1. *Given an element $e \in \{0, 1\}^k_{(u,v)}$ and a label $a \in K$, define e_a to be the bit corresponding the ath position of e.*

The Set System: For each $u \in U, a \in K$, we create a set $S_{u,a}$ that covers a subset of the elements which are associated with edges of the form (u, v'). Specifically, $e \in \{0, 1\}^k_{(u,v')}$ belongs to $S_{u,a}$ if and only if $e_a = 0$.

Similarly, for $v \in V, \alpha \in L$, we create a set $S_{v,\alpha}$ that covers a subset of the elements which are associated with edges of the form (u', v). Specifically, $e \in \{0, 1\}^k_{(u',v)}$ belongs to $S_{v,\alpha}$ if and only if $e_{\pi_{v \to u'}(\alpha)} = 1$. The number of sets we therefore create is $k|U| + l|V|$.

2

Covering Constraint: We also set the maximum number of sets that can be picked to be $m = |U| + |V|$.

Because k and l are constants, it is easy to see that this reduction is poly-time. It is also worth noticing that the reduction is designed with the idea that whenever a consistent labeling exists for an edge (u, v), the whole block of ground elements $\{0, 1\}^k_{(u,v)}$ can be covered by using just 2 sets (those corresponding to the satisfying labels).

We now prove Completeness and Soundness of this reduction.

Theorem 2.2. *Completeness:* $\mathsf{Opt}(\mathcal{G}) = 1 \Rightarrow \mathsf{Opt}(\mathcal{H}) = 1$

Proof. Since $\mathsf{Opt}(\mathcal{G}) = 1$, there exists a function $f : U \to K, V \to L$ s.t for each edge (u, v), $f(u) = \pi_{v \to u}(f(v))$. Let us choose (as our $m = |U| + |V|$ sets) a set corresponding to each vertex. In particular, let us choose the sets $S_{u,f(u)}$ for $u \in U$ and $S_{v,f(v)}$ for $v \in V$.

We show that these m sets cover all the ground elements. Let us focus on an edge (u, v). All the ground elements associated with this belong to the set $\{0, 1\}^k_{(u,v)}$.

By definition of the set system, the set $S_{u,f(u)}$ covers all the elements $e \in \{0, 1\}^k_{(u,v)}$ with $e_{f(u)} = 0$. Similarly, the set $S_{v,f(v)}$ covers all the elements $e \in \{0, 1\}^k_{(u,v)}$ with $e_{\pi_{v \to u}(f(v))} = 1$.

However, since f is a consistent labeling on (u, v), we have $f(u) = \pi_{v \to u}(f(v))$. Hence, $S_{v,f(v)}$ covers all the elements $e \in \{0, 1\}^k_{(u,v)}$ with $e_{f(u)} = 1$. Therefore, the above 2 sets covers the entire block. Applying this argument to each edge (u, v), we get the desired result. $\square$

Theorem 2.3. *Contrapositive of Soundness:* $\mathsf{Opt}(\mathcal{H}) \geq 3/4 + \epsilon \Rightarrow \mathsf{Opt}(\mathcal{G}) \geq \epsilon^3/2000$

Before we prove this theorem, notice that the result stated above is actually the contrapositive of the original soundness claim that we had mentioned following Theorem 1.1. We will now show how proving this result would suffice. Given any fixed $\epsilon > 0$ (it depends on how close to $3/4$ we need the hardness result to be), we first set $\eta = \epsilon^3/2000$. Therefore, by Theorem 1.1, we can create an instance $\mathcal{G}$ of MAX-LABEL-COVER with a constant bound q on the cardinality of the sets K and L where the 1 vs η decision problem is hard. Also, recall that q depends only on η. We then reduce $\mathcal{G}$ to $\mathcal{H}$ using the reduction stated above.

Now, if $\mathsf{Opt}(\mathcal{G}) < \eta$, it must be the case that $\mathsf{Opt}(\mathcal{H}) < 3/4 + \epsilon$, because we have assumed that we know Theorem 2.3 to be true. This completes the proof of soundness.

We now actually prove Theorem 2.3.

Proof. Let $\mathsf{Opt}(\mathcal{H}) > 3/4 + \epsilon$. Let $\mathcal{S}^*$ be a collection of m sets which achieves a coverage of more than $3/4 + \epsilon$. Our task now is to "decode" $\mathcal{S}^*$ into a reasonably good labeling for $\mathcal{G}$.

Idea: Firstly, the collection of sets can be partitioned according to which vertex each one is associated with. The idea is to essentially show that, if there is large element coverage, it must be the case that there exists a decent fraction of the edges for which the sets chosen contain two which correspond to a consistent labeling. Therefore, a random assignment of labels (from the sets that are in each vertex's partition) should hopefully lead to one which satisfies a non-negligible fraction of the edges. We formally prove along similar lines.

3

Definition 2.4. For vertex $u \in U$, set $Suggest(u)$ to be the set $\{a \in K : S_{u,a} \in \mathcal{S}^*\}$. Likewise, for vertex $v \in V$, set $Suggest(v)$ to be the set $\{\alpha \in L : S_{v,\alpha} \in \mathcal{S}^*\}$. We also overload the definition of $Suggest$ for edges too by defining $Suggest(u,v)$ to be $Suggest(u) \cup Suggest(v)$.

Since we have picked $m = |U| + |V|$ sets, we expect on an average $Suggest(u,v)$ to be 2 for edge (u,v). We show this in the following lemma.

Lemma 2.5. $\mathbf{E}_{(u,v)}[|Suggest(u,v)|] = 2$

Proof.

$$
\begin{align}
\mathop{\mathbf{E}}_{(u,v)}[|Suggest(u,v)|] &= \mathop{\mathbf{E}}_{(u,v)}[|Suggest(u)|] + \mathop{\mathbf{E}}_{(u,v)}[|Suggest(v)|] \tag{1} \\
&= \mathop{\mathbf{E}}_{u}[|Suggest(u)|] + \mathop{\mathbf{E}}_{v}[|Suggest(v)|] \tag{2} \\
&= (\sum_{u \in U} Suggest(u))/|U| + (\sum_{v \in V} Suggest(v))/|V| \tag{3} \\
&= (\sum_{u \in U} Suggest(u) + \sum_{v \in V} Suggest(v))/|U| \tag{4} \\
&= m/(m/2) \tag{5} \\
&= 2 \tag{6}
\end{align}
$$

Here, the jump from step 1 to step 2 follows because of the initial assumption we had made that $\mathcal{G}$ is left regular and right regular (we can pick a random vertex by picking a random edge as well). Similarly, the jump from step 3 to step 4 follows because $|U| = |V|$ in $\mathcal{G}$. $\qquad \square$

We now use a probabilistic argument to show the existence of a good labeling, given m sets which achieve coverage more than $3/4 + \epsilon$.

Random Decoding: If $Suggest(u) \neq \emptyset$ for $u \in U \cup V$, choose $f(u)$ to be an element uniformly randomly picked from $Suggest(u)$. If $Suggest(u) = \emptyset$, pick $f(u)$ arbitrarily.

We now need to show that $\mathbf{E}_f[Coverage_{\mathcal{G}}(f)] \geq \epsilon^3/2000$. Once we show this, we know by using a probabilistic argument that there exists at least one such labeling f which also achieves a coverage of at least $\epsilon^3/2000$, thereby establishing that $\mathsf{Opt}(\mathcal{G}) \geq \epsilon^3/2000$.

Before we get to proving it, here are some preliminaries.

Definition 2.6. An edge (u,v) is said to have consistent suggestions if there exist $a \in Suggest(u)$ and $\alpha \in Suggest(v)$ such that $a = \pi_{v \to u}(\alpha)$.

Claim 2.7. *Suppose (u,v) doesn't have consistent suggestions. Then, coverage of $\mathcal{S}^*$ on $\{0,1\}^k_{(u,v)}$ is at most $1 - 2^{-t}$ where $t = |Suggest(u,v)|$.*

Proof. Let $Suggest(u) = \{\kappa_1, \kappa_2, \ldots, \kappa_c\}$, and $Suggest(v) = \{\lambda_1, \lambda_2, \ldots, \lambda_d\}$. Also let $t = c + d$. For convenience, let us denote $\pi_{v \to u} : L \to K$ as just π.

An element $e \in \{0,1\}^k{}_{(u,v)}$ is covered if and only if

$$
\begin{aligned}
e_{\kappa_1} &= 0 \text{ (or)} \\
e_{\kappa_2} &= 0 \text{ (or)} \\
&\vdots \\
e_{\kappa_c} &= 0 \text{ (or)} \\
e_{\pi(\lambda_1)} &= 1 \text{ (or)} \\
e_{\pi(\lambda_2)} &= 1 \text{ (or)} \\
&\vdots \\
e_{\pi(\lambda_d)} &= 1
\end{aligned}
$$

Note that some $\pi(\lambda_i)$ may be the same as $\pi(\lambda_{i'})$, but what we are guaranteed is that none of the elements $\pi(\lambda_i)$ is equal to any term κ_j (else (u,v) would have consistent suggestions).

Let $t' = c + d'$ be the number of distinct indices that appear in the above equations. We are now interested in the fraction of the elements (these would be the elements not covered by the two sets) which have constraints on all of those t' distinct indices. This is exactly equal to $2^{-t'}$ which is at least 2^{-t}. Hence the fraction of elements covered is at most $1 - 2^{-t}$, hence proving the claim. $\quad\square$

Remark 2.8. On an average t is 2 (Lemma 2.5). This gives us the $3/4$ which we were looking for !

To gather some intuition, we first prove the expectation result assuming $Suggest(u,v) = 2$ for each $(u,v) \in E$.

By claim 2.7, $\mathcal{S}^*$ covers at most $3/4$ fraction of the blocks on all inconsistent edges. However, the total coverage of $\mathcal{S}^*$ is at least $3/4 + \epsilon$, which means that there exists $\Omega(\epsilon)$ fraction of the edges that have consistent suggestions. Also, if an edge (u,v) has consistent suggestions, it has to be the case that it has one labeling each from $Suggest(u)$ and $Suggest(v)$ (as $|Suggest(u,v)| = 2$, by assumption). Hence, the only possible assignment for f would satisfy the edge. This proves that the expected fraction of edges satisfied is non-negligible ($\Omega(\epsilon)$).

Now, we move on to the actual proof, by first proving the following concavity lemma.

Lemma 2.9. *If $t = t_1 + t_2 + \ldots t_r$, the values of t_i maximizing $(\sum_i (1 - 2^{-t_i}))/r$ are $t_i = t/r$.*

Proof. Maximizing $(\sum_i (1 - 2^{-t_i}))/r$ is equivalent to minimizing $(1/r) \sum_i 2^{-t_i}$. This quantity is at least $2^{-t/r}$ (AM $\geq$ GM). Further, setting $t_i = t/r$ achieves this lower bound, thus proving the lemma. $\quad\square$

It is worth noting that if t suggestions are to be distributed to r edges, the above lemma proves that the optimal way to distribute them in order to maximize coverage is by giving each edge the average suggestions per edge (t/r).

Definition 2.10. Define an edge (u,v) to be "frugally suggested" if $Suggest(u,v) \leq 10/\epsilon$. Also, define an edge to be "good" if it is both frugally suggested and consistently suggested.

5

Suppose the fraction of good edges (u, v) is at least $\epsilon/20$.
We argue that this implies $\mathbf{E}_f[Coverage_{\mathcal{G}}(f)] \geq \epsilon^3/2000$. Suppose (u, v) is good, it has consistent suggestions and hence there exist $a \in Suggest(u)$ and $\alpha \in Suggest(v)$ such that the labeling (a, α) satisfies the edge (u, v). Further, since (u, v) is good, it is frugally suggested, so both $|Suggest(u)|$ and $|Suggest(v)|$ are at most $10/\epsilon$. Thus when f is randomly chosen, there is at least an $(\epsilon/10)^2$ chance that it will get both $f(u) = a$ and $f(v) = \alpha$, thus satisfying (u, v). Hence in expectation, f satisfies at least an $(\epsilon/20)(\epsilon/10)^2 = \eta$ fraction of edges (u, v), as desired.

We now show that the fraction of good edges is at least $\epsilon/20$. We assume the contrary. Then, for a randomly chosen edge (u, v),

$$
\begin{aligned}
\mathrm{P}[(u, v) \text{ has consistent suggestions}] \quad &\leq \quad \mathrm{P}[(u, v) \text{ is good}] + \mathrm{P}[(u, v) \text{ is not frugally suggested}] \\
&< \quad \epsilon/20 + \mathrm{P}[(u, v) \text{ has at least } 10/\epsilon \text{ suggestions}] \\
&\leq \quad \epsilon/20 + (2\epsilon)/10 \\
&= \quad \epsilon/4
\end{aligned}
$$

Here, $\mathrm{P}[(u, v)$ has at least $10/\epsilon$ suggestions$]$ is at most $(2\epsilon)/10$ by Markov's inequality.

This implies that the contribution of consistent edges to the $Coverage$ is at most $\epsilon/4$. Therefore, the average coverage on inconsistently suggested edges needs to be at least $3/4 + (3/4)\epsilon$. Further, the average number of labels per inconsistently suggested edge is at most $2/(1 - \epsilon/4) \leq 2 + \epsilon$ (as there are at least $(1 - \epsilon/4)$ fraction of inconsistently suggested edges). Therefore, by applying the concavity lemma (Lemma 2.9), the coverage on inconsistently suggested edges is at most $1 - 2^{-(2+\epsilon)} = (1 - (1/4)2^{-\epsilon}) \leq 1 - 1/4(1 - \epsilon \ln 2)$. This quantity is at most $3/4 + 0.17\epsilon$ which is strictly lesser than $3/4 + 3/4\epsilon$. This is a contradiction ! $\qquad\square$

This results in the following theorem, which was the objective of this lecture.

Theorem 2.11. *Given any constant $\epsilon > 0$, the 1 vs $3/4 + \epsilon$ decision problem of* MAX-COVERAGE *is* NP-*Hard.*

References

[1] U. Feige and J. Kilian. Two-prover protocols — low error at affordable rates. *SIAM Journal on Computing*, 30(1):324–346, Feb. 2000.

[2] T. Holenstein. Parallel repetition: simplifications and the no-signaling case. In *STOC '07: Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 411–419, New York, NY, USA, 2007. ACM.

[3] R. Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998.