

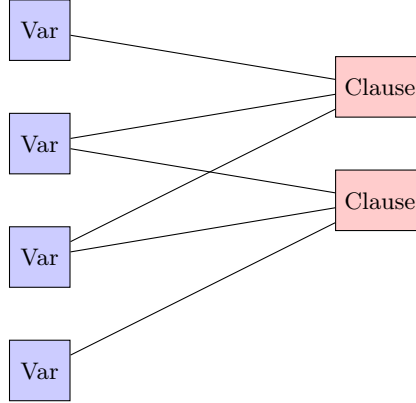
Homework 1

Lin Zejin

May 18, 2025

-
- **Collaborators:** I finish this homework by myself.
-

Problem 1. (a) Reduce from the instance of MAX-E3SAT-6.



Variables x_i have $\sigma(x_i) \in \{0, 1\}$ and Clauses $c_i = x_{j_i}^1 \wedge x_{j_i}^2 \wedge x_{j_i}^3$ have $\sigma(c_i) \in [7]$ to represent the state of c_i .

Therefore, constraint is naturally induced.

In the instance of MAX-E3SAT, the ratio of $|U|$ and $|V|$ is 2. So this is a regular Label-Cover Game for $K = 2, L = 7$ and $|V| = 2|U|$.

In the lecture we have proved that this is an instance of $\text{MAX-LC}_{1,1-\epsilon}$ for some ϵ .

So $\text{MAX-LC}_{1,1-\epsilon}$ is NP-Hard.

(b) We actually can construct another graph induced by (a).

We add \bar{x}_i to the graph in (a) and add the induced constraints from c_i contains variable x_i to \bar{x}_i .

Here the Label-Cover Game is regular and symmetric.

Then for the $\text{MAX-E3SAT-6}_{1,1-\epsilon}$ instance, the completeness is trivial.

Now we prove the soundness. That's because, if $\text{OPT}_{\text{MAX-E3SAT-6}} \leq 1 - \epsilon$, consider any $\sigma : U \rightarrow \{0, 1\}, V \rightarrow [7]$. At least $(1 - \epsilon)|V|$ clauses are not satisfied by $\sigma|_U$. For each clause, there exists at least one variable x_i/\bar{x}_i such that do not satisfy the constraint.

So Verifier rejects with probability at least $(1 - \epsilon)|V|/2|V| = (1 - \epsilon)/2$. So the soundness property holds if we set $\epsilon' = \frac{1+\epsilon}{2}$.

So we prove that $\text{GAP-LC}(K, L)_{1,1-\epsilon}$ is NP-Hard for some ϵ and K, L even if the graph is regular and symmetric.

By Raz' Paralled Repetition Theorem, we can reduce an instance of $\text{GAP} - \text{LC}(K, L)_{1, \delta}$ to the instance of $\text{GAP} - \text{LC}_{1, \exp(-\Omega(\frac{\delta^3 t}{\log t}))}$. Therefore, we finally prove that for any $\eta > 0$, there exists K, L such that $\text{GAP} - \text{LP}(K, L)_{1, \eta}$ is NP-Hard.

Problem 2. (a) For a regular Label-Cover problem $G = (U, V, E)$ that every veritce in U matches k vertices in V , $|U| = |V| = n$, consider the k -uniform hypergraph $H = (V', E')$ where $V' = E$ and k -tuples are all $[(u, v_1), (u, v_2), \dots, (u, v_k)]$ for $(u, v_i) \in E$. $[L']$ now represents the value of (u, v_i) , i.e. $[L'] = [L] \times [K]$. $[K] = [k + 1]$.

The maps are defined as: For the labeling $\sigma : [V] \rightarrow [L] \times [K]$, $\sigma(u, v_i) = (l, k)$. If $\pi_{(u, v_i)}(k) = l$ is matching in Labek-Cover problem, then we let $\pi_e^i(\sigma(u, v_i)) = k + 1$. Otherwise, if (l, k) does not satisfy the constraint, then we let $\pi_e^i(\sigma(l, k)) = i$.

So the constraint is weakly satisfied iff at least two edges in the k -tuples are satisfied in the constraint before. Also, the constraint is strongly satisfied iff all edges in the k -tuples are satisfied.

Completeness is trivial since if there is some label in the Label-Cover Game satisfy all constraint, then it can be naturally induced in the hypergraph.

Soundness is because: Assume $\text{OPT} \geq \epsilon$ in k -ary-Consistent-Labeling problem. Then we choose all edges (u_i, v_j) that are satisfied in the Label-Cover Game, denoted as S . There are at least $2\epsilon n$ edges. Now we label each u_i, v_j one by one.

Since the graph G is regular, at most $2k - 1$ edges in S have common vertice with an edge in S .

So each time we choose an arbitrary $e = (u, v) \in S$, label it with the label in k -ary-Consistent-Labeling and then we remove those edges in S who intersects with e .

In the end, for those vertices that have not been labeled yet, label it randomly.

Then at least $\frac{2\epsilon n}{2k} = \frac{\epsilon}{k}n$ edges are satisfied in Label-Cover-Game.

Therefore $\text{OPT} \geq \frac{\epsilon}{k}$ for Label-Cover Game.

As a result, if $\text{OPT} \leq \eta$ in Label-Cover Game, then $\text{OPT} \leq k\eta$ in k -ary-Consistent-Labeling problem.

Since $\text{MAX} - \text{LC}_{1, \eta}$ is NP-Hard, to distinguish instance with strong value 1 and weak value less than $k\eta$ k -ary-Consistent-Labeling problem is NP-Hard $\forall \eta > 0$.

Here we end the proof.

(b)

Problem 3. Consider all values $d(r, v) \pmod{\frac{1}{2}}$. They divide $[0, \frac{1}{2})$ into $|V| + 1$ pieces of interval.(including the interval $[v, v]$ if exists) If we choose θ in each interval, edges that will be removed are the same, so the cost is the same.

As a result, we can try θ in each interval and find the minimum cost. This will be less than 2OPT .

Problem 4. (a) If a connected component has diameter at most k in the $(10, 0.1, 1, 1)$ -expandar G , we prove that it has at most 10^k vertices.

By induction, $k = 1$ is trivial. Assume $k - 1$ holds for it. Assume subgraph G' has the maximum number of vertices. There isn't any vertex in G' that has distance less than $k - 1$ with each vertex in G' and also connects with

other vertex u outside. Otherwise, u can be added to g' , which causes contradiction with the maximum property. Then for k , any vertex in the graph with diameter $k - 1$ has degree 10 so at most 10^k vertices are connected to the graph. Since any vertex beyond G' has distance larger than k with some vertices in G' as we proved before, the expanded graph has at most 10^k vertices.

So each connected component has at most $10^{1/2 \log_{10} n} = n^{1/2}$ vertices in this problem. As n large enough, $n^{1/2} < 0.1n$. For those connected components S_1, \dots, S_k , removed edges are

$$|\partial S_1 \cup \partial S_2 \cup \dots \partial S_k| = \frac{1}{2} \sum_{t=1}^k |\partial S_t| \geq \frac{1.01}{2} \sum_{t=1}^k |S_t| > 0.5n$$

So we must have deleted $\Omega(n)$ edges.

Now we set the pair (s_i, t_i) to be all (u, v) where $u, v \in G$ and distance between u and v is k .

Then for any possible connected component in multicut, vertices u, v in it have distance is less than k .

For a $(10, 0.1, 1.1)$ -expander graph, by (a) we removed at least $\frac{1}{2}n$ if $k = \frac{1}{2} \log_{10} n$.

However, in LP case, we can set $x_e = \frac{1}{k}$ for any edge e . Then the cost will be

$$\frac{1}{k} \cdot |E| = \frac{5}{k} |V| = \frac{5n}{k}$$

So the integral gap is $\Omega(\log n)$.

Problem 5. (a)

$$\begin{aligned} \mathbb{E}(\text{cut value}) &= \sum_{(i,j) \in E} \omega_{ij} \cdot \frac{\arccos \langle v_i, v_j \rangle}{\pi} \\ &= \sum_{(i,j) \in E} \omega_{ij} - \sum_{(i,j) \in E} \frac{\frac{\pi}{2} + \arcsin \langle v_i, v_j \rangle}{\pi} \\ &\geq \sum_{(i,j) \in E} \omega_{ij} - \beta \cdot \sum_{(i,j) \in E} \omega_{ij} \sqrt{\frac{1 + \langle v_i, v_j \rangle}{2}} \\ &\geq \sum_{(i,j) \in E} \omega_{ij} - \beta \left(\sum_{(i,j) \in E} \omega_{ij} \frac{1 + \langle v_i, v_j \rangle}{2} \right)^{1/2} \left(\sum_{(i,j) \in E} \omega_{ij} \right)^{1/2} \\ &= 1 - \beta(1 - \text{SDP})^{1/2} \\ &\geq 1 - \beta(1 - \text{OPT})^{1/2} \end{aligned}$$

where $\beta = \sup_{\alpha \in (-1, 1)} \frac{\frac{\pi}{2} + \arcsin \alpha}{\sqrt{1 + \alpha}} < +\infty$ by L'hospital rule.

Therefore, it is a $1 - \epsilon$ vs. $1 - \beta\sqrt{\epsilon}$ search algorithm.

(b)

Similar to max-cut. If we set $\mathbb{F}_2 = \{\pm 1\}$, then

$$\frac{1 - bx_i x_j}{2} = \begin{cases} 1 & x_i \oplus x_j = b \\ 0 & x_i \oplus x_j \neq b \end{cases}$$

where $1 \oplus 1 = -1 \oplus -1 = -1, 1 \oplus -1 = 1 \oplus 1 = 1$.

So the problem is to maximize the objective

$$\sum_{(i,j) \in E} \omega_{ij} \frac{1 - b_{ij} x_i x_j}{2}$$

Similarly, we set the SDP relaxation:

$$\min \sum_{(i,j) \in E} \omega_{ij} \frac{1 - b_{ij} \langle v_i, v_j \rangle}{2}$$

conditioned on $\|v_i\| = 1$.

After finding a minimum, we design a randomize algorithm as follows:

Uniformly sample $\vec{r} \sim S^{n-1}$.

Set $x_i = \text{sgn} \langle \vec{r}, \vec{v}_i \rangle$.

Then

$$\begin{aligned} \mathbb{E}(\text{cut value}) &= \sum_{(i,j) \in E} \omega_{ij} \cdot \frac{\arccos b_{ij} \langle v_i, v_j \rangle}{\pi} \\ &= \sum_{(i,j) \in E} \omega_{ij} - \sum_{(i,j) \in E} \frac{\frac{\pi}{2} + \arcsin b_{ij} \langle v_i, v_j \rangle}{\pi} \\ &\geq \sum_{(i,j) \in E} \omega_{ij} - \beta \cdot \sum_{(i,j) \in E} \omega_{ij} \sqrt{\frac{1 + b_{ij} \langle v_i, v_j \rangle}{2}} \\ &\geq \sum_{(i,j) \in E} \omega_{ij} - \beta \left(\sum_{(i,j) \in E} \omega_{ij} \frac{1 + b_{ij} \langle v_i, v_j \rangle}{2} \right)^{1/2} \left(\sum_{(i,j) \in E} \omega_{ij} \right)^{1/2} \\ &= 1 - \beta(1 - \text{SDP})^{1/2} \\ &\geq 1 - \beta(1 - \text{OPT})^{1/2} \end{aligned}$$

where $\beta = \sup_{\alpha \in (-1,1)} \frac{\frac{\pi}{2} + \arcsin \alpha}{\sqrt{1+\alpha}} < +\infty$ by L'hospital rule.

Therefore, it is a $1 - \epsilon$ vs. $1 - \beta\sqrt{\epsilon}$ search algorithm.

Problem 6.

Problem 7. hyperplane cuts $\frac{\alpha}{\pi}$ edges in G_d with angle α .

Then totally, hyperplane cuts

$$\frac{\int_{\arccos \rho^*}^{\pi} \sin^{d-2} \alpha \frac{\alpha}{\pi} d\alpha}{\int_{\arccos \rho^*}^{\pi} \sin^{d-2} \alpha d\alpha} < \frac{\int_{\arccos \rho^*}^{\pi} (\pi - \alpha)^{(d-2)/2} \frac{\alpha}{\pi} d\alpha}{\int_{\arccos \rho^*}^{\pi} (\pi - \alpha)^{(d-2)/2} d\alpha} = \frac{\arccos \rho^*}{\pi} + O\left(\frac{1}{d}\right)$$

The first inequality is because $\frac{\sin \alpha}{\sin \beta} > \frac{\sqrt{\pi - \alpha}}{\sqrt{\pi - \beta}}$ if $\alpha < \beta$. Thus the probability of α in the left is less than the probability of β in the right if $\alpha < \beta$.

Problem 8. $f : \{\pm 1\}^n \rightarrow \mathbb{R}$ is a linear combination of function $f : \{\pm 1\}^n \rightarrow \{\pm 1\}$, which can be written in the form

of linear combination of Fourier base functions:

$$f(x) = \sum_{S \subset [n]} \hat{f}(S) \chi_S(x)$$

where $\chi_S(x) = \prod_{i \in S} x_i$ is a multilinear polynomial.

So it is expressible as a multilinear polynomial.

The uniqueness is because, if there is some multilinear polynomial g such that $g(x) = f(x), \forall x \in \{\pm 1\}^n$. Then using Parseval's Theorem we obtain that

$$\sum_{S \subset [n]} (\hat{f}(S) - \hat{g}(S))^2 = \mathbb{E}_{\vec{x} \sim \{\pm 1\}^n} (f(\vec{x}) - g(\vec{x}))^2 = 0$$

$$\text{So } f - g = \sum_{S \subset [n]} (\hat{f}(S) - \hat{g}(S)) \chi_S = 0.$$

Problem 9.

$$\langle f, g \rangle = \left\langle \sum_{S \subset [n]} \hat{f}(S) \chi_S, \sum_{S \subset [n]} \hat{g}(S) \chi_S \right\rangle = \sum_{S_1, S_2 \subset [n]} \hat{f}(S_1) \hat{g}(S_2) \langle \chi_{S_1}, \chi_{S_2} \rangle = \sum_{S \subset [n]} \hat{f}(S) \hat{g}(S)$$

However, if we let $f = \chi_{\{x\}}, g = \chi_{\{y\}}, h = \chi_{\{x,y\}}$, then

$$\mathbb{E}_t \chi_{\{x\}}(t) \chi_{\{y\}}(t) \chi_{\{x,y\}}(t) = \mathbb{E}_t t_x^2 t_y^2$$

But

$$\hat{f}(S) \hat{g}(S) \hat{h}(S) \equiv 0, \forall S \subset [n]$$

due to they are Fourier basis.