Prague Summer School on Discrete Mathematics

Homeworks: Hardness of Approximation

Homework 0

The following would be very useful (and might be necessary!) as preparation towards this mini-course.

- Watching this video here:

  `https://www.youtube.com/watch?v=_RhTRqp1d74`

  This video provides the overall context/plan for the course.

- Recalling the basic notions of **P**, **NP**, **NP**-completeness (**NP**-hardness), approximation algorithms.

- Recalling simple approximation algorithms, e.g. Vertex Cover (factor 2), 3SAT (factor $\frac{7}{8}$), Set Cover (factor ln n).

- Recalling the Goemans-Williamson algorithm for Max-Cut. This is covered (and will be covered quickly) in Lecture L-5.

- Recalling background on basic Fourier analysis on the hypercube. This is covered (and will be covered quickly) in Lectures L-1 (page 8-10), L-6, L-7 (page 2-3, Beckner operator). You could also look at the analysis of the 3-bit linearity test and 3-bit dictatorship test with noise, Lecture L-2.

- Reading notes for the first talk.

  The first talk will cover the material in Lecture L-0. Those unfamiliar with the "proof checking" viewpoint might find it useful to take a look beforehand. The material also appears in typeset notes here (Lecture 1, 2, scroll down the page):

  `https://cs.nyu.edu/~khot/PCP-Spring20.html`

- Watching this video here:

  `https://www.youtube.com/watch?v=sVZnOgYTEtE`

  This material will be covered in the last two talks in the course, but since this is a bit involved, some preparation would help.

Homework 1

1. **2SAT**

   Show that 2SAT is in **P**, i.e. design a polynomial time algorithm that decides whether a 2SAT formula φ has a satisfying assignment.

2. **Maximum Acyclic Subgraph**

   Give a $\frac{1}{2}$-approximation algorithm for the Maximum Acyclic Subgraph problem. In this problem, given a directed graph $G(V, E)$, the goal is to find an *acyclic* subgraph $G(V, \tilde{E})$, $\tilde{E} \subseteq E$ so as to maximize $|\tilde{E}|$.

   Note: Assuming the Unique Games Conjecture, $\frac{1}{2}$-approximation is known to be optimal!

3. **Independent Set**

   Assume the PCP Theorem, i.e. that $\text{Gap3SAT}_{1,\theta}$ is **NP**-hard for some absolute constant $\theta < 1$.

   Using the standard reduction from 3SAT to Independent Set (IS) problem, show that $\text{GapIS}_{\alpha,\beta}$ is **NP**-hard for some absolute constants $\alpha > \beta$ (the gap problem asks, given an n-vertex graph, to distinguish whether there is an independent set of size αn or whether there is no independent set of size βn). To recall, the standard reduction constructs a triangle for every 3SAT clause whose three vertices are labeled by the literals appearing in that clause and then connects every pair of mutually opposite literals (this reduction, in particular, gives $\alpha = \frac{1}{3}$).

   Using the multiplicativity of the size of the maximum independent set under a suitable graph product, show that $\text{GapIS}_{\alpha^k,\beta^k}$ is **NP**-hard for every constant integer $k \geqslant 1$. Hence Independent Set is **NP**-hard to approximate within *any* constant factor.

   Now let n be the size of the initial graph G (before taking the graph product) in the question above. Let $k = \log n$ and let H be a random induced subgraph of the product graph G^k such that $|H|$ is polynomial in n. Show that with high probability, the fractional size of the maximum independent set in H is roughly the same as in G^k (here one uses a Chernoff bound and a union bound over all *maximal* independent sets in G^k and the fact that maximal independent sets in G^k themselves have a product structure and hence there aren't too many of them!). Conclude that Independent Set on N-vertex graphs is **NP**-hard to approximate within a factor N^δ for some constant $\delta > 0$ (under a randomized reduction).

4. **Maximum Coverage**

   Assume that it is **NP**-hard to approximate the Set Cover problem within a factor $(1 - \varepsilon) \ln n$ where n is the size of the universe for the set system. Show that the Maximum Coverage problem is **NP**-hard to approximate within a factor $1 - \frac{1}{e} + \delta$ where $\delta \to 0$ as $\varepsilon \to 0$. In this problem, given a set system and a number k, the goal is to find k sets so as to maximize the size of their union.

   Hint: Reduce approximating Set Cover to approximating Maximum Coverage. If there is a set cover of size k, one could, using an algorithm for the Maximum Coverage problem, find k sets that cover $1 - \frac{1}{e} + \delta$ fraction of the elements, delete these elements, and repeat.

   Note: The respective factors are known to be optimal for both the problems!

5. **Independent Set via FGLSS Reduction**

   Assume the PCP Theorem, i.e. that **NP** has a PCP verifier that uses logarithmic randomness, makes a constant number of queries to the proof, has perfect completeness, and soundness $s = \frac{1}{2}$ (i.e. a proof of an incorrect statement is accepted with probability at most s).

   Let $r = O(\log n)$ be the number of random bits and $q = O(1)$ be the number of queries used by the verifier. Using the "FGLSS reduction", show that it is **NP**-hard to distinguish whether a graph has an independent set of size 2^r or whether it has no independent set of size $s2^r$. The FGLSS reduction constructs a blob of at most 2^q vertices for every choice τ of the randomness. The vertices in this blob correspond to the query patterns that the verifier may accept on randomness τ and they are connected to each other forming a clique. Further, two vertices in different blobs are connected if the corresponding query patterns have a proof location in common and assign different bit-values to this location (i.e. if they are mutually *contradictory*).

Do you know the error amplification technique using random walks on an expander? If so, show that **NP** has a PCP verifier that uses logarithmic randomness and *logarithmic* number of queries, has perfect completeness, and soundness $\tilde{s}$ that is inverse polynomial in n (this verifier is obtained by running the earlier verifier logarithmic number of times, but in a dependent manner using a walk on an expander rather than as independent runs). Conclude via the FGLSS reduction that Independent Set on N-vertex graphs is **NP**-hard to approximate within a factor N^δ for some constant $\delta > 0$.

Homework 2

1. **4-bit Linearity Test**

   Given a function $f : \{-1,1\}^n \to \{-1,1\}$, consider a test that picks $x, y, z \in \{-1,1\}^n$ uniformly at random and accepts if and only if

   $$f(xyz) = f(x)f(y)f(z).$$

   Show that if the test accepts with probability $\frac{1}{2} + \varepsilon$, then there is a Fourier coefficient $\hat{f}_\alpha$ such that $|\hat{f}_\alpha| \geqslant \Omega(\sqrt{\varepsilon})$.

2. **3LIN to 3SAT**

   Show how to write a set of four clauses whose conjunction is equivalent to the linear constraint $a \cdot b \cdot c = 1$ (in the $\{-1,1\}$ or multiplicative notation). Using this, show that Håstad's result on 3LIN implies optimal hardness result for 3SAT as well, albeit with imperfect completeness, i.e. that $\text{Gap3SAT}_{1-\varepsilon, \frac{7}{8}+\varepsilon}$ is **NP**-hard for every constant $\varepsilon > 0$.

3. **Graph Linearity Test**

   Let $k \geqslant 2$ be an integer thought of as a constant. The "graph linearity test" works as follows. Given a function $f : \{-1,1\}^n \to \{-1,1\}$, the test picks $x_1, \ldots, x_k \in \{-1,1\}^n$ uniformly at random and accepts if and only if

   $$f(x_i x_j) = f(x_i)f(x_j) \quad \forall 1 \leqslant i < j \leqslant k.$$

   Thus the test consists of $\binom{k}{2}$ runs of the basic 3-bit linearity test, but these runs are very much dependent (as opposed to independent). It is known nevertheless that these runs *behave* as if they were independent in the sense that if the test accepts with probability $2^{-\binom{k}{2}} + \varepsilon$, then f must have a Fourier coefficient that is at least $\Omega(\varepsilon)$ in magnitude! Moreover, this result implies an alternate proof of Håstad's well-known result that it is **NP**-hard to approximate Independent Set in N-vertex graph within a factor $N^{1-\delta}$ for every constant $\delta > 0$!

While the proof for $k \geqslant 4$ is tricky, show that the claim holds for $k = 3$, i.e. that if the test accepts with probability $\frac{1}{8} + \varepsilon$, then f must have a Fourier coefficient that is at least $\Omega(\varepsilon)$ in magnitude.

4. **Optimality of Graph Linearity Test**

   For this problem, it would be more natural to use the $\mathbb{F}_2$ or additive notation. We show here that the "soundness" probability of $2^{-\binom{k}{2}}$ in the question above is optimal.

   Consider the function

$$g(y_1, \ldots, y_{2n}) = y_1 y_2 \oplus y_3 y_4 \oplus \cdots \oplus y_{2n-1} y_{2n}.$$

   Show that all Fourier coefficients of g are small, i.e. at most $2^{-\Omega(n)}$. Let $k \geqslant 2$ be an integer thought of as a constant. Show that for a random k-dimensional subspace $L \subseteq \mathbb{F}_2^{2n}$, the restriction $g|_L$ of the function g to the subspace L is linear with probability (at least) $2^{-\binom{k}{2}} - o_n(1)$.

   Hint: Restriction to a random k-dimensional subspace is (essentially) substituting

$$(y_1, \ldots, y_{2n})^{\mathsf{T}} = M(x_1, \ldots, x_k)^{\mathsf{T}},$$

   for a random $2n \times k$ matrix M. After the substitution, one needs coefficient of all $\binom{k}{2}$ terms $x_i x_j, i < j$ to vanish. Show that these events are (essentially) independent.

Homework 3

1. **Dictatorship Test towards 3SAT**

   Let -1 correspond to logical True and $+1$ correspond to logical False. Show how to write a clause $a \vee b \vee c$, $a, b, c \in \{-1, 1\}$, as an arithmetic expression that evaluates to 1 or 0 depending on whether the clause evaluates to True or False.

   Hint: the corresponding arithmetic expression for $a \vee b$ would be $\frac{3-a-b-ab}{4}$.

   Let a probability distribution $\mathcal{D}$ on three bits (a, b, c) be defined as follows. Pick bits a, b uniformly at random. If $a = 1$, then let $c = -b$. If $a = -1$, then let $c = b$ with probability $1 - \delta$ and let $c = -b$ with probability δ. One can choose δ to be a sufficiently small constant as needed. Note that the bit c is unbiased, is independent of the bit a, but is correlated to the bit b with $\mathbb{E}[bc] = -\delta$. Also $\mathbb{E}[abc] = -1 + \delta$ and $(a, b, c) \neq (1, 1, 1)$.

   Given a function $f : \{-1, 1\}^n \to \{-1, 1\}$, $\mathbb{E}[f] = 0$, consider the following test. Pick $x, y, z \in \{-1, 1\}^n$ such that for every co-ordinate $1 \leqslant i \leqslant n$, independently, (x_i, y_i, z_i) is chosen according to the distribution $\mathcal{D}$ above. The test accepts if and only if $f(x) \vee f(y) \vee f(z)$ evaluates to True.

   Show that the test has perfect completeness, i.e. that if f is a dictatorship, it passes the test with probability 1. Show that the test has soundness $\frac{7}{8}$, i.e. that if f passes the test with probability $\frac{7}{8} + \varepsilon$, then f must have a Fourier coefficient $\hat{f}_\alpha$ that is at least $\gamma = \gamma(\varepsilon, \delta)$ in magnitude and $|\alpha|$ is at most $C = C(\varepsilon, \delta)$.

2. **Max-Cut on Almost Bipartite Graphs (Hardness)**

   Under the Unique Games Conjecture, show that it is **NP**-hard to distinguish whether a given graph has a cut of size $1 - \varepsilon$ (as a fraction of the total number of edges) or whether there is no cut of size $1 - \Omega(\sqrt{\varepsilon})$. Think of $\varepsilon > 0$ as a very small constant.

   Just use a different setting of the correlation parameter ρ in the reduction. In the next question, we show that this is optimal.

3. **Max-Cut on Almost Bipartite Graphs (Algorithm)**

   Let $G(V, E)$ be a graph that has a cut of size (as a fraction of the total number of edges) at least $1 - \varepsilon$. Think of $\varepsilon > 0$ as a very small constant. Show that the Goemans-Williamson's algorithm finds a cut of size $1 - O(\sqrt{\varepsilon})$.

   You could first assume, for the sake of simplicity, that the SDP solution yields vectors $\{v_i\}_{i \in V}$ such that for *every* edge $(i, j) \in E$, $\langle v_i, v_j \rangle \leqslant -1 + 2\varepsilon$. Towards the general case, note that the function $\cos^{-1} x$ is convex in the vicinity of -1.

4. **Max-Cut to 2SAT**

   Using a simple reduction that writes a Max-Cut constraint as a pair of clauses, show that under the Unique Games Conjecture, it is **NP**-hard to distinguish whether a given 2SAT instance has an assignment that satisfies $1-\varepsilon$ fraction of the clauses or whether no assignment satisfies more than a $1 - \Omega(\sqrt{\varepsilon})$ fraction of clauses. Think of $\varepsilon > 0$ as a very small constant.

   Note: This is optimal, i.e. there is a matching algorithm!

Homework 4

1. **Expansion in "Doubled" Graph**

   Given a regular bipartite graph $G(U, V, E)$, construct a "doubled" graph $H(V, \tilde{E})$ as follows: a random edge $(v, w) \in \tilde{E}$ is sampled by sampling a random vertex $u \in U$ and sampling two of its neighbors $v, w \in V$ independently (i.e. $(u, v), (u, w) \in E$). Note that H has self-loops (since v, w could coincide), but it is a minor issue.

   Let $S \subseteq V$ be a subset of fractional size δ. Show that at least a δ^2 fraction of edges in $\tilde{E}$ are inside S and hence the expansion $\Phi(S) \leqslant 1 - \delta$ in the graph H.

2. **Unique Games and Small Set Expansion**

   Let $\mathcal{U}(G(V, E), [R], \{\pi_{v,w}\}_{(v,w) \in E})$ be an instance of Unique Games where $G(V, E)$ is a regular graph, its vertices are to be assigned labels from the set $[R] = \{1, \ldots, R\}$, and $\pi_{v,w}$ are the (1-to-1) permutation constraints on the edges. Specifically, $\pi_{v,w} : [R] \to [R]$ is a permutation and a labeling $\ell : V \to [R]$ satisfies the edge (v, w) if and only if $\pi_{v,w}(\ell(v)) = \ell(w)$ (here it is convenient to think of an implicit direction for each edge).

   The "label extended graph" $H(V \times [R], \tilde{E})$ is defined on the set of vertices $V \times [R]$ and its edges are defined as

   $$\tilde{E} = \{((v, i), (w, \pi_{v,w}(i))) \mid (v, w) \in E,\ i \in [R]\}.$$

For any labeling $\ell : V \to [R]$, let $\mathsf{Sat}(\ell)$ be the fraction of edges (constraints) of the Unique Games instance satisfied by the labeling and let

$$S_\ell = \{(v, \ell(v)) \mid v \in V\},$$

be a subset of vertices in H. Let $\Phi(S)$ denote the expansion of the set S_ℓ in the graph H. Show that (this is very easy; the exercise is meant to simply point out this connection):

$$\mathsf{Sat}(\ell) = 1 - \Phi(S_\ell).$$

3. **Small Set Expansion and Large Eigenvalues**

   Let $G(V, E)$ be a regular n-vertex graph and $1 = \lambda_1 \geqslant \cdots \geqslant \lambda_n$ be the eigenvalues of its normalized adjacency matrix. Let $S \subseteq V$ be a subset with expansion $\Phi(S) \leqslant \delta$. Think of δ as a small constant. Show that the indicator vector $\mathbf{1}_S$ of the set S is nearly contained in the span of eigenvectors with eignevalues close to 1. Specifically, show that for some $\gamma, \eta > 0$, if $\lambda_1, \ldots, \lambda_m$ are all the eigenvalues $\geqslant 1-\gamma$, and if $v_1, \ldots, v_m$ are the corresponding eigenvectors, then there exists a vector $w \in \mathsf{Span}\{v_1, \ldots, v_m\}$ such that

   $$\|\mathbf{1}_S - w\|_2 \leqslant \eta \cdot \|\mathbf{1}_S\|_2.$$

   You can try to get as good bounds on γ, η as possible as function of δ.

4. **Expansion in Noisy Hypercube**

   Let $\varepsilon \in (0, \frac{1}{2})$ be a noise parameter thought of as a fixed constant . The noisy hypercube is a (weighted) graph with vertex set $\{-1, 1\}^n$ and its random edge (x, y) is sampled by first sampling $x \in \{-1, 1\}^n$ uniformly at random and obtaining y by flipping every coordinate of x independently with probability ε.

   Let $S \subseteq \{-1, 1\}^n$ be a subset of vertices of fractional size δ. Show that its expansion $\Phi(S) \to 1$ as $\delta \to 0$.

   Hint: Use the Bonami-Beckner's hyper-contractive inequality.

5. **Correlated Gaussians**

   This could be a challenging probability question (that arises in analysis of an approximation algorithm for the Unique Games problem).

   Let $(g_1, \ldots, g_k)$ and $(h_1, \ldots, h_k)$ be standard (mean zero, unit variance) Gaussians such that for all i, $\mathbb{E}[g_i h_i] = 1 - \varepsilon$ and for all $i \neq j$, g_i is independent of g_j, h_j. Let i_0, j_0 be such that

   $$g_{i_0} = \max\{g_1, \ldots, g_k\}, \qquad h_{j_0} = \max\{h_1, \ldots, h_k\}.$$

   Show that $\mathsf{Pr}[i_0 = j_0] = 1 - \Theta(\sqrt{\varepsilon \log k})$. You can assume $\varepsilon \log k \ll 1$.

Homework 5

Read the survey article titled "*On the 2-to-2 Games Conjecture*" (posted on the webpage) and specifically the proof in Section 5 (I do hope to cover some of it during the course).

Also, do work on the earlier homeworks if you did not finish before!