CSS.330.1 10 Feb 2023

Problem Set 1

- Due Date: **24 Feb 2023**
- Turn in your problem sets electronically (LaTeX, pdf or text file) by email. If you submit handwritten solutions, start each problem on a fresh page.
- Collaboration is encouraged, but all writeups must be done individually and must include names of all collaborators.
- Refering sources other than class notes is strongly discouraged. But if you do use an external source (eg., other text books, lecture notes, or any material available online), ACKNOWLEDGE all your sources (including collaborators) in your writeup. This will not affect your grades. However, not acknowledging will be treated as a serious case of academic dishonesty.
- The points for each problem are indicated on the side.
- Be clear in your writing.

---

1. **[weighted version of vertex cover]** (15)

   Consider the following weighted version of Vertex Cover (W-VC).

   Input: Undirected graph $G = (V, E)$ with weights $w : V \to \mathbb{Z}$ on the vertices.
   Output: A cover $C \subseteq V$ of the vertices such that for every edge $(u, v) \in E$ either $u \in C$ or $v \in C$.
   Objective: Minimize the weight of the cover (i.e., $\sum_{v \in C} w(v)$).

   Observe that the 2-approximation algorithm for the unweighted version discussed in lecture does not extend to this weighted version. Design an alternate deterministic 2-approximation algorithm for W-VC.

   [Hint: First design a LP relaxation of the problem with variables for each vertex in the graph and then deterministically round the LP to obtain a 2-approximate solution.]

2. **[gap preserving reductions]** (15)

   A reduction from one gap problem gap-A$_\alpha$ to gap-B$_\beta$ (for some $0 < \alpha, \beta < 1$) is said to be a gap preserving reduction if it maps YES instances of gap-A$_\alpha$ to YES instances of gap-B$_\beta$ and NO instances of gap-A$_\alpha$ to NO instances of gap-B$_\beta$. The existence of a gap preserving reduction from gap-A$_\alpha$ to gap-B$_\beta$ implies that if it is NP-hard to approximate problem A to within α, then it is NP-hard to approximate problem B to within β.

   For every $\alpha > 0$, show that there exists a and ε, β and a gap preserving reduction from gap-3SAT$_\alpha$ to gap-2SAT$_{1-\varepsilon,\beta}$. Hence, conclude that there exists a $\beta \in (0, 1)$ such that approximating MAX2SAT to within β is NP-hard.

Note:The gap problems gap-3SAT$_\alpha$ and gap-2SAT$_{1-\varepsilon,\beta}$ are defined as follows.
gap-3SAT$_\alpha$:

$$\begin{aligned}\textsf{YES} &= \{\phi|\phi \text{ is a satisfiable 3CNF formula}\}\\ \textsf{NO} &= \{\phi|\phi \text{ is a 3CNF formula such that no assignment satisfies more}\\ &\qquad \text{than } \alpha \text{ fraction of the clauses}\}\end{aligned}$$

gap-2SAT$_{1-\varepsilon,\beta}$:

$$\begin{aligned}\textsf{YES} &= \{\phi|\phi \text{ is a 2CNF formula with an assignment that satsifies}\\ &\qquad \text{at least } (1-\varepsilon)\text{-fraction of the clauses}\}\\ \textsf{NO} &= \{\phi|\phi \text{ is a 2CNF formula such that no assignment satisfies more}\\ &\qquad \text{than } \beta \text{ fraction of the clauses}\}\end{aligned}$$

3. **[three vs. two queries]** (15)

   In class, we stated that Håstad proved the following strengthening of the PCP Theorem which shows that every language in NP has a PCP with 3 queries and soundness error almost $1/2$.

   $$[\text{Håstad}]\ \forall \varepsilon > 0, \textsc{CircuitSAT} \in PCP_{1-\varepsilon,1/2+\varepsilon}[O(\log n), 3].$$

   Suppose we were able to further strengthen the above result to prove that CIRCUITSAT has a 2 query PCP (i.e., CircuitSAT $\in PCP_{1,s}[O(\log n), 2]$ for some $0 < s < 1$), then show that then $NP = P$!

   Thus, Håstad's PCP is optimal with respect to the number of queries till the status of the P vs. NP question is resolved.

4. **[inapproximability of clique via graph products]** (8+7=15)

   In class, we proved the following theorem showing the inapproximability of clique. 3-COLOR $\in PCP_{c,s}[r,q]$ implies it is NP-hard to approximate MAXCLIQUE to within a factor s/c as long as $2^{r+q} = \text{poly}(\cdot)$. This resulted in the following inapproximability result for MAXCLIQUE assuming the PCP Theorem (i.e., 3-COLOR $\in PCP_{1,1/2}[O(\log n), O(1)]$).

   $$\exists \alpha \in (0,1),\ \text{it is NP-hard to approximate CLIQUE to within } \alpha \tag{1}$$

   We then applied sequential repetition on the PCP (i.e., $PCP_{c,s}[r,q] \subseteq PCP_{c^k,s^k}[kr,kq]$ for all $k \in \mathbb{Z}$) to obtain the following strengthening of the above result.

   $$\forall \alpha \in (0,1),\ \text{it is NP-hard to approximate CLIQUE to within } \alpha \tag{2}$$

In this problem, we will discuss an alternative approach to prove this result using graph products. For a graph $G = (V, E)$ we define the square of G, $G^2 = (V', E')$, as follows: The vertex set V' equals V^2, the set of pairs of vertices of G. Two distinct vertices (u_1, u_2) and (v_1, v_2) are adjacent in E' if and only if $(u_1, v_1) \in E$ and $(u_2, v_2) \in E$.

(a) Prove that the squaring operation defined above satisfies $\omega(G^2) = (\omega(G))^2$ where $\omega(G)$ denotes the size of the largest clique in G.

(b) Use (a) to given an alternate proof of (2) from (1).

5. **[linearity test of 3 functions]** (10)

Consider the following modification of the BLR-linearity test towards testing linearity of 3 functions $f, g, h : \{0,1\}^n \to \{1, -1\}$ simultaneously.

$$\begin{aligned}\mathsf{BLR\text{-}3\text{-}Test}^{f,g,h} : \text{“ } & 1. \text{ Choose } y, z \in_R \{0,1\}^n \text{ independently}\\ & 2. \text{ Query } f(y), g(z), \text{ and } h(y+z)\\ & 3. \text{ Accept if } f(y)g(z)h(y+z) = 1. \qquad \text{”}\end{aligned}$$

Clearly, if the three functions f, g, h are the same linear function, then the above test accepts with probability 1. Suppose one of the three functions f, g, h (say f) and its negation (i.e., $-f$) is δ-far from linear (this means $\max_\alpha |\widehat{f}_\alpha| \le 1 - 2\delta$), show that

$$\Pr_{y,z}[\mathsf{BLR\text{-}3\text{-}Test}^{f,g,h} \text{ rejects }] \ge \delta.$$

[Hint: The Cauchy-Schwarz inequality $(\sum a_i b_i)^2 \le (\sum a_i^2) \cdot (\sum b_i^2)$ may come useful.]

6. **[recycling queries in linearity test]** $(3 + 6 + 3 + 3 = 15)$

In lecture, we analyzed the soundness of the BLR-Test to show that if f is $(1/2 - \varepsilon)$-far from linear, then the test accepts with probability at most $1/2 + \varepsilon$. If we repeat this test k times, we obtain a linearity test which makes $3k$ queries and has the following property: if f is $(1/2 - \varepsilon)$-far from linear, then the test accepts with probability at most $(1/2 + \varepsilon)^k = 1/2^k + \delta$. Thus every additional 3 queries improves the soundness by a factor of $1/2$. In this problem, we show that this can be considerably improved.

Assume that both f and $-f$ are $(1 - \varepsilon)/2$-far from linear (i.e., $\max_\alpha |\widehat{f}_\alpha| \le \varepsilon$). Consider the following linearity test (parameterized by k).

$$\begin{aligned}\mathsf{Test}_k^f : \text{“ } & 1. \text{ Choose } z_1, z_2, \ldots, z_k \in_R \{0,1\}^n\\ & 2. \text{ For each distinct pair } (i,j) \in \{1, \ldots, k\}\\ & \qquad \text{Check if } f(z_i)f(z_j)f(z_i + z_j) = 1.\\ & 3. \text{ Accept if all the tests pass.} \qquad \text{”}\end{aligned}$$

Observe that this test makes at most $k + \binom{k}{2}$ queries. We will show below that the soundness of the test is roughly $2^{-\binom{k}{2}}$, thus showing that every additional query improves the soundness by a factor of $1/2$ (almost).

Assume that both f and $-f$ are $(1-\varepsilon)/2$-far from linear.

(a) Show that the acceptance probability of the above test is given by

$$\begin{aligned}
\Pr[\text{acc}] &= \mathbb{E}_{z_1,\ldots,z_k}\left[\prod_{i,j}\left(\frac{1+f(z_i)f(z_j)f(z_i+z_j)}{2}\right)\right] \\
&= \frac{1}{2^{\binom{k}{2}}}\cdot\sum_{S\subseteq\binom{[k]}{2}}\mathbb{E}_{z_1,\ldots,z_k}\left[\prod_{(i,j)\in S} f(z_i)f(z_j)f(z_i+z_j)\right]
\end{aligned}$$

(b) Consider any term in the above summation corresponding to a non-empty S (i.e., $\mathbb{E}_{z_1,\ldots,z_k}\left[\prod_{(i,j)\in S} f(z_i)f(z_j)f(z_i+z_j)\right]$). Suppose $(1,2)\in S$. Show that

$$\mathbb{E}_{z_1,\ldots,z_k}\left[\prod_{(i,j)\in S} f(z_i)f(z_j)f(z_i+z_j)\right] \le \mathbb{E}_{z_1,z_2}[f(z_1+z_2)g(z_1)h(z_2)]$$

for some functions $g,h:\{1,-1\}^n\to\{\pm1\}$.

[Hint: Fix all the variables other than z_1 and z_2 such that the expectation is maximized.]

(c) Use the result of item 5 to conclude that the expression in the above (for non-empty sums) is at most ε (i.e., $\mathbb{E}_{z_1,\ldots,z_k}\left[\prod_{(i,j)\in S} f(z_i)f(z_j)f(z_i+z_j)\right]\le\varepsilon$ for non-empty S).

(d) Conclude that $\Pr[\text{acc}]$ is at most $2^{-\binom{k}{2}}+\varepsilon$.

7. **[derandomized linearity testing]** $(3+2+4+4+2=15)$

A subset $S\subseteq\{0,1\}^n$ is said to be an ε-biased set if for all $\alpha\in\{0,1\}^n\setminus\{0^n\}$, we have $|\Pr_{x\in S}[\langle x,\alpha\rangle=1]-\Pr_{x\in S}[\langle x,\alpha\rangle=0]|\le\varepsilon$.

Consider the following modification of the BLR test to check if $f:\{0,1\}^n\to\{\pm1\}$ is linear:

S-derandomized BLR-Test^f : " 1. Choose $y\in_R\{1,-1\}^n$ and $z\in_R S$ independently
2. Query $f(y), f(z)$, and $f(y+z)$
3. Accept if $f(y)f(z)f(y+z)=1$. "

Observe that the number of random coins required for this test is only $n+\log_2|S|$. There exist explicit constructions of ε-biased sets S of size at most $O(n^2/\varepsilon^2)$. Thus, the randomness is at most $n+O(\log n+\log(1/\varepsilon))$ as opposed to $2n$ for the (non derandomized) BLR test. In this problem, we will show that this S-derandomized test performs as well as the BLR test in terms of soundness. More precisely, we will show that $\Pr[\text{acc}]\ge(1+\delta)/2$, then there exists a Fourier coefficient of absolute value at least $\sqrt{\delta^2-\varepsilon}$, thus matching the soundness of the BLR test but for the ε loss factor.

(a) Show that if S is an ε-biased set then $|\mathbb{E}_{x\in S}[\chi_\alpha(x)]|\le\varepsilon$ forall $\alpha\neq 0^n$.

(b) Show that if f is a linear function (i.e, $f = \chi_\beta$), f passes the S-derandomized BLR-Test with probability 1.

For two functions $f, g : \{0,1\}^n \to \mathbb{R}$, define the inner product $\langle f, g\rangle_S$ and S-norm $\|f\|_S$ as follows:

$$\langle f, g\rangle_S = \mathbb{E}_{z\in S}\left[f(z)g(z)\right]; \qquad \|f\|_S = \sqrt{\langle f, f\rangle_S}.$$

(c) For an arbitrary $f : \{0,1\}^n \to \{\pm 1\}$, show that the acceptance probability of the above test is given by

$$\begin{aligned}\Pr[\text{acc}] &= \frac{1}{2}\left(1 + \sum_\alpha \widehat{f_\alpha}^2 \cdot \langle f, \chi_\alpha\rangle_S\right) \\ &= \frac{1}{2}\left(1 + \left\langle f, \sum_\alpha \widehat{f_\alpha}^2 \chi_\alpha \right\rangle_S\right).\end{aligned}$$

(d) Use the fact that S is an ε-biased set and f is a $\{\pm 1\}$-valued function to prove that

$$\left|\left\langle f, \sum_\alpha \widehat{f_\alpha}^2 \chi_\alpha \right\rangle_S\right| \le \sqrt{(1-\varepsilon)\sum_\alpha \widehat{f_\alpha}^4 + \varepsilon}.$$

(e) Conclude that if the S-derandomized BLR-Test accepts with probability at least $(1+\delta)/2$, then there exists an α such that $|\widehat{f_\alpha}| \ge \sqrt{\delta^2 - \varepsilon}$.