CMSC 39600 Oct 23, 2007

Problem Set 1

- Due Date: **Nov 8 (Thurs), 2007**
- It is recommended that you try to solve all 10 problems, but it is sufficient if you submit the writeup for any 8 problems.
- Collaboration is encouraged, but all writeups must be done individually.
- Indicate names of all collaborators.
- Indicate all other sources (text books, lecture notes, other material available online) other than the course lecture notes.

1. **[gap preserving reductions]**

   A reduction from one gap problem gap-A_α to gap-B_β (for some $0 < \alpha, \beta < 1$) is said to be a gap preserving reduction if it maps YES instances of gap-A_α to YES instances of gap-B_β and NO instances of gap-A_α to NO instances of gap-B_β. The existence of a gap preserving reduction from gap-A_α to gap-B_β 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$_{1-\epsilon,\beta}$. Hence, conclude that there exists a $\beta \in (0,1)$ such that approximating MAX2SAT to within β is NP-hard.

2. **[three vs. two queries]**

   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.

   $$\text{[Håstad] } \forall \epsilon > 0, \textsf{Circuit-SAT} \in PCP_{1-\epsilon,1/2+\epsilon}[O(\log n), 3].$$

   Suppose we were able to further strengthen the above result to prove that Circuit-SAT has a 2 query PCP (i.e., Circuit-SAT $\in PCP_{1,s}[O(\log n), 2]$ 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.

3. **[optimal inapproximability of MAX3SAT]**

   The optimal query PCP theorem of Håstad stated in the earlier problem has the following additional property: the verifier's predicate for all random coins is of the form "$\pi_{i_1} \oplus \pi_{i_2} + \pi_{i_3} = b \mod 2$", for some $b \in \{0,1\}$, where π_{i_1}, π_{i_2} and π_{i_3} are the three (3) proof bits queried by the verifier. Using this strong form of the PCP

Theorem, show that for every $\delta > 0$, approximating MAX3SAT to within $7/8 + \delta$ is NP-hard.

[Recall that MAX3SAT is the problem of computing the maximum number of clauses of the given 3CNF Boolean formula satisfied by any assignment.]

4. **[inapproximability of clique via graph products]**

   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 (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 $k \in \mathbb{Z}^{\geq 0}$) 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. **[recycling randomness via random walks on an expander]**

   In lecture, we showed that by sequential repetition of PCPs (i.e., $PCP_{c,s}[r,q] \subseteq PCP_{c^k,s^k}[kr,kq]$ for all $k \in \mathbb{Z}^{\geq 0}$) can be used to improve the hardness factor of approximating clique (also see earlier problem). In this problem, we will discuss a more efficient way to perform repetition by recycling randomness using expander walks.

   Let $G = (V,E)$ be an (n,d,λ)-expander with $\lambda < d$. Let $B \subseteq V$ be a set of vertices with $|B| = \mu n$, where $0 < \mu < 1$. Suppose we pick a uniformly random vertex in G and then perform a t-step random walk in G starting from this vertex. We wish to upper-bound the probability p that all vertices encountered along this random walk are in the set B.

   (a) Let A denote the normalized adjacency matrix of G, and let P denote the matrix corresponding to *projection* onto B; in other words, P is the $n \times n$ diagonal matrix with 1's in the positions corresponding to B. Show that $p = \|P(AP)^t \pi\|_1$, where π is the vector $(1/n, \ldots |V| \text{times} \ldots, 1/n)$ (ie., the probability distribution of a random vertex in V), and $\|z\|_1$ denotes the l_1-norm of z (i.e., $\|z\|_1 = \sum_{i=1}^n |z_i|$).

(b) The matrix 2-norm of a matrix C is defined to be $\|C\|_2 = \max_{y \neq 0} \|Cy\|_2/\|y\|_2$. Show that $p \leq \mu \|PAPAP\ldots AP\|_2 \leq \|AP\|_2^t$.

(c) Show that $\|AP\|_2 \leq \sqrt{\mu + (\lambda/d)^2}$, and conclude $p \leq (\mu + (\lambda/d)^2)^{t/2}$.

[Hint: given arbitrary $y \neq 0$, write $z = Py$ and express $z = z_\parallel + z_\perp$ as in Lecture 8]

Extra Credit: show that in fact $\|PAP\|_2 \leq (\lambda/d) + \mu(1 - \lambda/d)$ and show how this can be used to conclude the sharper upper bound $p \leq \mu(\lambda/d) + \mu(1 - \lambda/d)^t$.

(d) Use the earlier part (c) to conclude that $\mathrm{PCP}_{1,1-s}[r, q] \subseteq \mathrm{PCP}_{1,2^{-k}}[r+O(k), O(kq)]$ for all $k \in \mathbb{Z}^{\geq 0}$.

(e) Conclude from (d) (setting $k = \log n$) that it is NP-hard to approximate to within $n^{-\delta}$ for some $\delta \in (0, 1)$.

6. **[linearity test of 3 functions]**

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.

```
BLR-3-Test^{f,g,h} : " 1. Choose $y, z \in_R \{0,1\}^n$ independently
                      2. Query $f(y), g(z),$ and $h(y+z)$
                      3. Accept if $f(y)g(z)h(y+z) = 1$.          "
```

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 $\max_\alpha |\hat{f}_\alpha| \leq 1 - 2\delta$), show that

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

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

7. **[recycling queries in linearity test]**

In lecture, we analyzed the soundness of the BLR-Test to show that if f is $(1/2-\epsilon)$-far from linear, then the test accepts with probability at most $1/2+\epsilon$. If we repeat this test k times, we obtain a linearity test which makes $3k$ queries and has the following property: if f is $(1/2-\epsilon)$-far from linear, then the test accepts with probability at most $(1/2+\epsilon)^k = 1/2^k + \delta$. 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 $(1-\epsilon)/2$-far from linear (i.e., $\max_\alpha |\hat{f}_\alpha| \leq \epsilon$). Consider the following linearity test (parameterized by k).

```
Test_k^f : " 1. Choose $z_1, z_2, \ldots, z_k \in_R \{0,1\}^n$
             2. For each distinct pair $(i, j) \in \{1, \ldots, k\}$
                    Check if $f(z_i)f(z_j)f(z_i + z_j) = 1$.
             3. Accept if all the tests pass.                  "
```

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 $(1-\epsilon)/2$-far from linear.

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

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

(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 $\mathbb{E}_{z_1,\ldots,z_k}\left[\prod_{(i,j)\in S} f(z_i)f(z_j)f(z_i+z_j)\right]$ is upper bounded by $\mathbb{E}_{z_1,z_2}[f(z_1+z_2)g(z_1)h(z_2)]$ for some functions $g,h:\{0,1\}^n\to\{0,1\}$.

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

(c) Use the result of Problem 6 to conclude that the expression in the above (for non-empty sums) is at most ϵ (i.e., $\mathbb{E}_{z_1,\ldots,z_k}\left[\prod_{(i,j)\in S} f(z_i)f(z_j)f(z_i+z_j)\right]\le\epsilon$ for non-empty S).

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

8. **[parallelization: Boolean alphabet to non-Boolean alphabet]**

Prove that for all functions $r,q:\mathbb{Z}^{\geq 0}\to\mathbb{Z}^{\geq 0}$ and $\epsilon:\mathbb{Z}^{\geq 0}\to[0,1]$, we have

$$\mathrm{PCP}_{1,1-\epsilon}[r,q]\subseteq\mathrm{PCP}^{\Sigma}_{1,1-\epsilon/q}[r+\log_2 q,2],$$

where Σ is any alphabet of size at least 2^q (i.e, $\log_2|\Sigma|\geq q$).

9. **[gap amplification towards parallel repetition theorem?]**

In class, we proved the following version of the gap amplification lemma: There exists $\alpha\in(0,1)$ and an alphabet Σ such that

$$PCP^{\Sigma}_{1,1-\epsilon}[r,2]\subseteq PCP^{\Sigma}_{1,1-\epsilon'}[r+f(t),2],$$

where $\epsilon'=\min\{t\epsilon,\alpha\}$ and $f(\cdot)$ is some function.

The fact that $\epsilon'=\min\{t\epsilon,\alpha\}$ allowed us to increase the gap all the way to α but not any more. If we had instead had $\epsilon'=t\epsilon$, then we could potentially increase the gap all the way to 1, obtaining some form of a parallel repetition theorem. In this problem, we will explore that this approach is bound to fail showing that the "min" is necessary and not merely an artifact of the proof of the gap amplification lemma.

(a) It is known that for infinitely many constants d there exist (n, d, λ)-expanders G for infinitely many n, with the following two properties: (a)$\lambda(G) \leq 2\sqrt{d}$ and (b) G has *girth* at least $\frac{2}{3}\log_d n$, where the girth of a graph is the length of the smallest cycle in it. Suppose we make such a graph G into a constraint graph over the alphabet $\{0, 1\}$ by putting an inequality constraint on every edge. Show that $UNSAT(G) \geq \frac{1}{2} - O(\frac{1}{\sqrt{d}})$.

(b) On the other hand, show that for any t, if n is large enough, then the graph G' obtained via powering in Dinur's gap amplification procedure satisfies $UNSAT(G') \leq \frac{1}{2}$, thus showing that the gap cannot be increased beyond $1/2$ for this constraint graph

[Hint: Consider a random assignment, and use linearity of expectation]