PCPs and Hardness of Approximation: Assignment
NYU Computer Science, Spring 2008

Due on Monday, May 5th, 3:30pm; Absolutely no extension

Solve all problems (or as much as you can). Collaboration is allowed, but you must write your own solutions and mention names of your collaborators.

Problem 1

Let \( \omega(G) \) denote the size of the largest clique in a graph \( G \). For any positive integer \( k \), let \( G^k \) denote the product graph whose vertices are \( k \)-tuples of vertices in \( G \) and tuples \((u_1, u_2, \ldots, u_k)\) and \((v_1, v_2, \ldots, v_k)\) are connected iff for every \( 1 \leq i \leq k \), either \( u_i = v_i \) or \((u_i, v_i)\) is an edge in \( G \).

Prove that \( \omega(G^k) = \omega(G)^k \).

We know from the PCP Theorem that for some constants \( \alpha < \beta \) there exists a polynomial time reduction that maps a SAT instance \( \phi \) of size \( n \) to an \( N \)-vertex graph \( G \) such that

- \( \phi \) is satisfiable \( \implies \omega(G) \geq \beta N \)
- \( \phi \) is unsatisfiable \( \implies \omega(G) \leq \alpha N \).

This shows that MAX-CLIQUE (i.e. the problem of finding the largest clique in a graph) cannot be approximated in polynomial time within factor \( \beta/\alpha \) unless \( P = NP \). Using the product graph construction above, show that MAX-CLIQUE cannot be approximated in polynomial time within any constant factor unless \( P = NP \). Hint: take the graph \( G^k \) for a large enough constant \( k \).

Now take \( k = (\log N)^C \) where \( C \) is a large enough constant and express the hardness factor (i.e. \((\beta/\alpha)^k\)) as a function of the size of the final (product) graph. Using this, prove that for arbitrarily small constant \( \varepsilon > 0 \), MAX-CLIQUE on \( m \) vertex graphs cannot be approximated in polynomial time within factor \( 2^{(\log m)^{1-\varepsilon}} \) unless SAT instances of size \( n \) can be solved in time \( 2^{(\log n)^{C'}} \) for some constant \( C' \) (\( C' \) could depend on \( \varepsilon \)).

Finally, take \( k = O(\log N) \) so that the size of the product graph is \( |G|^k = N^{O(\log N)} \) and the hardness gap is \((\beta/\alpha)^k = N^{\Omega(1)} \). Let \( G' \) be a random induced subgraph of \( G^k \) of size \( N' = N^2/\alpha^{2k} \).

Using this, show that MAX-CLIQUE on \( m \) vertex graphs cannot be approximated in polynomial time within factor \( m^\gamma \) for some constant \( \gamma > 0 \) unless SAT can be solved in randomized polynomial time. Hint: Using Chernoff bound, show that with high probability, taking random induced subgraph of \( G^k \) essentially preserves the fractional size of the maximum clique. In the “NO case”, you would need to take a union bound over all cliques in \( G^k \). A straightforward upper bound on the number of cliques in \( G^k \) is \( 2^{|G|} \) and this union bound would not work. Observe however that it suffices to take a union bound only over the maximal cliques in \( G^k \); these are product of maximal cliques in \( G \), and there are at most \( 2^{Nk} \) many of them.
Problem 2

Fix $\varepsilon > 0$ to be a small constant. Given a function $f : \{-1, 1\}^n \mapsto \{-1, 1\}$, consider the following test: Pick three inputs $x, y, z$ uniformly and independently at random from $\{-1, 1\}^n$ and an input $\mu \in \{-1, 1\}^n$ as follows: independently for every $1 \leq i \leq n$, the bit $\mu_i$ is chosen to be 1 with probability $1 - \varepsilon$ and $-1$ with probability $\varepsilon$. Let $w = -xyz\mu$, i.e. for every $1 \leq i \leq n$, $w_i = -x_i y_i z_i \mu_i$. Finally

Test: Accept iff $f(x)f(y)f(z)f(w) = -1$.

Show that if this test accepts with probability $\frac{1}{2} + \delta$, then there must exist $S \subseteq [n]$ such that

$$|S| \neq 0, \quad |S| \leq O\left(\frac{1}{\varepsilon} \log \left(\frac{1}{\delta}\right)\right), \quad |\widehat{f}(S)| \geq \delta.$$ 

Note that $\widehat{f}(S)$ denotes the Fourier coefficient of $f$ at $S$. If $f$ is a dictatorship function, what is the probability that the test accepts?

Remark: This test can be used to extend Håstad’s 3-bit PCP to a 4-bit PCP where the verifier reads 4 bits and accepts iff the XOR of the bits equals 1 and no “folding” is needed. In Håstad’s 3-bit PCP on the other hand, since the proof is “folded”, the acceptance predicate of the verifier is $a \oplus b \oplus c = 0$ half the times and $a \oplus b \oplus c = 1$ half the times.

Problem 3

The “tribes” function $f : \{-1, 1\}^{2^k} \mapsto \{-1, 1\}$ is defined as follows. The variables are indexed as $\{x_{ij} \mid 1 \leq i \leq 2^k, \ 1 \leq j \leq k\}$, and

$$\forall \ x \in \{-1, 1\}^{2^k}, \quad f(x) := \lor_{i=1}^{2^k} \left( \land_{j=1}^{k} x_{ij} \right).$$

Note that $-1$ is logical TRUE, 1 is logical FALSE, $\lor$ is logical OR, $\land$ is logical AND.

What is (approximately) $\Pr_x [f(x) = 1]$? What is the influence of each variable (correct upto a constant factor)? Recall that

$$\text{Influence}_i(f) := \Pr_x [f(x) \neq f(xe_i)]$$

where $e_i$ denotes the input that is $-1$ in the $i^{th}$ coordinate and 1 otherwise.

Remark: A celebrated result of Kahn, Kalai and Linial states that every balanced boolean function on $n$ variables must contain a variable with influence at least $\Omega(\log n/n)$. The tribes function shows that this lower bound is optimal upto a constant factor.