CS 153 Current topics in theoretical computer science		**Spring 2010**
	Problem Set 1	
Out: April 22		*Due:* **May 6**

Reminder: you are encouraged to work in small groups; however you must turn in your own write-up and note with whom you worked. The solutions to these problems can be found in various online course notes and research papers. Please do not search for or refer to these solutions.

1. In the MAX-K-COVER problem, we are given a collection of subsets $S_1, S_2, \ldots, S_m$, and the objective is to find the subset of k of them that covers the largest number of elements; i.e., the subset $I \subseteq [m]$ with $|I| = k$ and $|\cup_{i \in I} S_i|$ maximum. Show that the greedy algorithm achieves an approximation ratio of at least $1 - 1/e$.

2. We proved in class that for all constants $\epsilon, \delta > 0$, $[1 - \epsilon, 1/2 + \delta]$-GAP MAX-3-LIN is **NP**-hard. Consider the problem MAX-3-MAJ in which each "clause" is the MAJORITY predicate on 3 literals (variables or their negations), and the objective is to find an assignment that maximizes the number of simultaneously satisfied clauses. Prove by reduction from MAX-3-LIN that MAX-3-MAJ is **NP**-hard to approximate to better than a 2/3 factor. Hint: replace each linear equation with 4 MAJORITY clauses.

3. Inapproximability of CLIQUE. In this problem you will show that for every constant $\epsilon > 0$, CLIQUE cannot be approximated to within a $n^{1-\epsilon}$ factor, unless $\mathbf{P} = \mathbf{NP}$. This strong result was proved by Zuckerman in 2006, although the threshold of $n^{1-\epsilon}$ had been known for some time, predicated on slightly weaker complexity assumptions. There are two ingredients in this proof, described next.

   - A refinement of the *query complexity* parameter q of a PCP system is the *free-bit complexity*, denoted f. For each sequence of the verifier's coin tosses r, it computes q queries into the proof, and a predicate $\phi_r : \{0,1\}^q \to \{\text{accept}, \text{reject}\}$ that it applies to the answers to the q queries. The *free bit complexity* f is the log of the maximum number of distinct answers that make the verifier accept; i.e., it is the maximum over r of $\log_2 |\phi_r^{-1}(\text{accept})|$. Let us denote by $\mathbf{FPCP}_{c,s}(r, f)$ the set of languages with PCP systems having completeness c, soundness error s, randomness complexity r, and free-bit complexity f. Håstad proved the following theorem[1]:

     Theorem 1.1 *For all constants $\overline{f} > 0$, there is an ℓ such that*

     $$\mathbf{NP} \subseteq \mathbf{FPCP}_{1,2^{-\ell}}(O(\log n), \overline{f}\ell).$$

   - A function $E : [N] \times [D] \to [M]$ is called a (K, s) *disperser* if for all subsets $X \subseteq [N]$ of cardinality at least K, the set $E(X, [D]) = \{E(x, y) | x \in X, y \in [D]\}$ has cardinality at least sM. A disperser E is *efficient* if the function E can be computed in polynomial time in the length of its input. Zuckerman gave the following construction of dispersers:

[1] The quantity $\overline{f}$ is called the *amortized free-bit complexity*.

Theorem 1.2 *For any constant $\epsilon > 0$, and any $s = s(N) > 0$, there is an efficient family of (N^ϵ, s) dispersers $E : [N] \times [D] \to [M]$ with $D \leq O(\log N / \log(1/s))$ and $M \geq N^{\Omega(1)}$.*

(a) Prove that if $\mathbf{NP} \subseteq \mathbf{FPCP}_{1,s}(r, f)$ then it is **NP**-hard to distinguish whether a graph on 2^{r+f} vertices (i) has a clique of size at least 2^r or (ii) has no clique larger than $s2^r$.

(b) Prove that if $\mathbf{NP} \subseteq \mathbf{FPCP}_{1,s}(r, f)$, then $\mathbf{NP} \subseteq \mathbf{FPCP}_{1,K/2^R}(R, Df)$, by using a (K, s) disperser $E : [2^R] \times [D] \to [2^r]$. Hint: repeat the verifier D times, using (non-independent) random strings dictated by E.

(c) Prove that it is **NP**-hard to approximate CLIQUE to within $n^{1-\epsilon}$, for any constant $\epsilon > 0$.

4. Inapproximability of SET COVER. Recall that the greedy algorithm for SET COVER over a universe of size n produces a $O(\log n)$ approximation. In this problem you will show that SET COVER cannot be approximated to within a $\Omega(\log n)$ factor, unless $\mathbf{NP} \subseteq \mathbf{DTIME}(n^{O(\log \log n)})$. There are two ingredients in this proof, described next.

- First, we need a version of LABEL COVER with *subconstant* soundness error, as in the following theorem (which uses Raz's Parallel Repetition Theorem applied to an instance of LABEL COVER with constant soundness error):

  Theorem 1.3 *For every $\epsilon = \epsilon(n) > 0$ there is a reduction from 3-SAT to $[1, \epsilon]$-GAP LABEL COVER, running in time $n^{O(\log(1/\epsilon))}$, that produces an instance of LABEL COVER with graph size $n^{O(\log(1/\epsilon))}$ and with alphabet size at most $(1/\epsilon)^{O(1)}$. Moreover the bipartite graph produced by this reduction is (left-) and (right-) regular.*

- Second, we need explicit constructions of the following combinatorial object: a (m, ℓ) *set system* over a universe U is a collection $C_1, C_2, \ldots, C_m$, each C_i a subset of U, with the property that if the union of at most ℓ subsets from among

  $$\{C_1, C_2, \ldots, C_m, \overline{C_1}, \overline{C_2}, \ldots, \overline{C_m}\}$$

  equals U, then the union must contain both C_i and $\overline{C_i}$ for some i. It is not hard to construct such set systems:

  Theorem 1.4 *For all $m \geq \ell \geq 2$, there exist (m, ℓ) set systems over a universe of size $O(2^{2\ell} m^2)$, and these can be constructed in time $2^{O(\ell)} m^{O(1)}$.*

(a) Let $\{C_1, C_2, \ldots, C_m\}$ be a (m, ℓ) set system over universe U. Fix a function $\pi : [m] \to [m]$, and define

$$A = \{\overline{C_{\pi(i)}} : i \in [m]\} \quad B = \{C_j : j \in [m]\}.$$

Given subsets $A' \subseteq A$ and $B' \subseteq B$ with $|A'| \leq \ell/2$ and $|B'| \leq \ell/2$ for which $A' \cup B'$ constitutes a cover of U, describe a randomized procedure that produces from A' an $i \in [m]$ and from B' a $j \in [m]$ that satisfy $j = \pi(i)$ with probability at least $4/\ell^2$.

(b) Let $G = (V_1, V_2, E)$ be an instance of $[1, \epsilon]$-GAP LABEL COVER with alphabet $[m]$. Fix a parameter ℓ. Using an (m, ℓ) set system over universe U, give a reduction to SET COVER and prove (completeness) if there exists a labeling satisfying all edges simultaneously, then there is a set cover of size at most $|V_1| + |V_2|$, and

(soundness) if there exists a set cover of size at most $(\ell/8)(|V_1|+|V_2|)$, then there exists a labeling satisfying more than a $2/\ell^2$ fraction of the edges.

Hint: the SET COVER instance will be over the universe $E \times U$, and it will have a set for each element of $V_1 \times \Sigma$ and a set for each element of $V_2 \times \Sigma$.

(c) Prove that SET COVER cannot be approximated to within a $c \log N$ factor, for some constant c (where N is the size of the SET COVER instance), unless $\mathbf{NP} \subseteq \mathbf{DTIME}(n^{O(\log \log n)})$. Hint: choose ϵ, ℓ appropriately and apply the previous part.