CS 153 Current topics in theoretical computer science	**Spring 2010**

Problem Set 2

Out: May 21 — *Due:* **June 1**

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. **UG-hardness from dictatorship tests.** Suppose we have a Long Code test of the following form: a k-tuple $z_1, z_2, \ldots, z_k$, with each $z_i \in \{-1,+1\}^R$, is chosen according to a joint distribution μ. The purported codeword $f : \{-1,+1\}^R \to \{-1,+1\}$ is queried at locations $z_1, z_2, \ldots, z_k$, and a "payoff"
$$P(f(z_1), f(z_2), \ldots, f(z_k))$$
is returned, where $P : \mathbb{R}^k \to \mathbb{R}$ is a multilinear function, and it is guaranteed that on the domain $[-1,+1]^k$, P returns values between -1 and $+1$. Denote by $V(f)$ the payoff returned when the test is applied to function f (so $V(f)$ is a random variable).

   Suppose further that the test satisfies the following completeness and soundness conditions, which are described with reference to three constants s, c, and τ, all of which lie in $[0,1]$.

   Completeness If f is a codeword (i.e., $f = \chi_{\{i\}}$ for some $i \in \{1,\ldots,R\}$), then $\mathbf{E}_\mu[V(f)] \geq c$.

   Soundness If $\mathbf{E}_\mu[V(f)] > s$, then there exists an τ-influential coordinate; i.e., there is some $i \in \{1,2,\ldots,R\}$ for which $I_i(f) \geq \tau$. Here the I_i are maps that take functions
$$g : \{-1,+1\} \to [-1,+1]$$
to $[0,1]$, and they satisfies three axioms:

   (a) there is an absolute positive constant t such that for any g, $\sum_i I_i(g) \leq t$ (bounded sum),

   (b) for each i, for any distribution on inputs g given by the random variable G, $I_i(\mathbf{E}[G]) \leq \mathbf{E}[I_i(G)]$ (convexity), and

   (c) for any permutation $\pi : [R] \to [R]$, if g' is the function g with its input coordinates permuted by π, then $I_i(g') = I_{\pi(i)}(g)$.

   You will show (assuming the UGC) that the existence of such a test is enough to prove that it **NP**-hard to approximate the associated generalized constraint satisfaction problem (GCSP) to better than a s/c factor. An instance of the associated GCSP has N $\{+1,-1\}$ variables and nonnegative weights w_S for each k-tuples S of the variables, satisfying $\sum_S w_S = 1$; the value of the instance is the maximum over assignments $\phi : [N] \to \{-1,+1\}$ to the N variables, of the quantity
$$\sum_{S=\{i_1,i_2,\ldots,i_k\}} w_S P(\phi(i_1), \phi(i_2), \ldots, \phi(i_k)).$$

Notice that MAX-CUT and the associated dictatorship test we discussed in class are captured by the above setup, with $k = 2$, and payoff function $P(x, y) = -xy$.

The starting point for the reduction is an instance $G = (U, V, E)$ of Unique Label Cover, with label set $\{1, 2, \ldots, R\}$ and edge constraints $\pi_{u,v}$ for each $(u, v) \in E$, for which it is **NP**-hard to distinguish between two cases:

YES there is an assignment such that for at least a $(1 - \delta)$ fraction of vertices $u \in U$, *all* edges[1] incident to u are satisfied, and

NO no assignment satisfies more that δ fraction of the edges.

As usual, we can assume δ is a sufficiently small constant.

For an assignment $A : U \cup V \to [R]$, the expected proof will be given by Long Code encodings f_v of $A(v)$ for each $v \in V$. The verifier's actions are as follows: choose a random vertex $u \in u$, select $z_1, z_2, \ldots, z_k$ according to μ, choose random u-neighbors of $v_1, v_2, \ldots, v_k \in V$, and return the payoff

$$P(f_{v_1}(\pi_{u,v_1}(z_1)), f_{v_2}(\pi_{u,v_2}(z_2)), \ldots, f_{v_k}(\pi_{u,v_k}(z_k))).$$

(a) For a vertex $u \in U$, define $f_u : \{-1, +1\}^R \to [-1, +1]$ by $f_u(z) = \mathbf{E}_{v \in N(u)}[f_v(\pi_{u,v}(z))]$. Here $N(u)$ denotes the neighbors of u, and v is assumed to be chosen randomly from among $N(u)$; as usual $\pi(z)_i = z_{\pi^{-1}(i)}$. Show that the expected payoff returned by the verifier equals $\mathbf{E}_{u \in U}[V(f_u)]$. (So, in effect, the verifier is querying the "average" codeword of the neighbors of a random left-vertex). Hint: use the fact that P is multilinear.

(b) Show that when given the proof associated with an assignment A such that for at least a $(1 - \delta)$ fraction of vertices $u \in U$, *all* edges incident to u are satisfied, the expected payoff returned by the verifier is at least $(1 - \delta)c - \delta$.

(c) Show that if the expected payoff returned by the verifier is at least $s + \eta$, then there exists an assignment $A : U \cup V \to [R]$ satisfying at least a

$$\eta \cdot \frac{\tau}{t+1} \cdot \frac{\tau}{t} \cdot \frac{\tau}{t(t+1)}$$

fraction of the edge constraints. Hint: define candidate label sets

$$L(u) = \{i : I_i(f_u) \geq \tau\}$$

for $u \in U$, and

$$L(v) = \{i : I_i(f_v) \geq \tau/(t+1)\}$$

for $v \in V$, and argue that there are a $\eta \cdot \tau/(t+1)$ fraction of "good" edges (u, v) for which a random labeling from these sets satisfies edge (u, v) with probability $|L(u)|^{-1}|L(v)|^{-1}$. Then argue that for such pairs u, v, the sets $L(u)$ and $L(v)$ are small.

(d) Show that, assuming the UGC, it is **NP**-hard to approximate the above GCSP to better than a s/c factor.

[1]This strong version is known to be equivalent to the other version we've used in class.

2. **Amortized query complexity of linearity tests.** In class we described a PCP which made (the optimal) 3 queries into bits of the proof, and had soundness error close to $1/2$. To reduce the soundness error to 2^{-k} we could run the verifier k times, making $3k$ queries; i.e, each 3 queries reduce the soundness error by an additional $1/2$ factor. However certain reductions require something much stronger: one can hope that *each additional* query reduces the soundness error by $1/2$. Put another way, one can define the *amortized query complexity* of a test with soundness error s and query complexity q to be $q/\log_2(1/s)$; the goal is to reduce this from 3 to $1+\delta$ for every $\delta > 0$. This was achieved by Samorodnitsky and Trevisan, and the main idea is present in the simpler setting of linearity testing.

   In the presentation below, we view Boolean functions as functions $f : \{-1,+1\}^n \to \{-1,+1\}$. We are interested in a test that makes few queries to f, and accepts if f is linear (i.e. if $f = \chi_S$ for some subset $S \subseteq [n]$) and rejects with high probability if f is far from linear. We say that f is *ϵ-far from linear* if $\max_S \hat{f}(S) \leq \epsilon$.

   (a) Consider the following test: chose random $x, y \in \{-1,+1\}^n$, and accept iff $f(x)f(y)f(xy) = 1$. Show that the acceptance probability is 1 when f is linear and at most $1/2+\epsilon/2$ when f is ϵ-far from linear. Hint: Start with $\Pr[\text{accept}] = \mathbf{E}_{x,y}[1/2+(1/2)f(x)f(y)f(xy)]$ and replace each occurrence of f with its Fourier expansion. You will probably need to use Cauchy-Schwarz: $(\sum_i a_i b_i)^2 \leq (\sum_i a_i^2)\cdot(\sum_i b_i^2)$.

   (b) Show that for *every* pair of function $g, h : \{-1,+1\}^n \to \{-1,+1\}$, the soundness analysis holds even for when the test is modified to check $g(x)h(y)f(xy)$. In other words, show that if f is ϵ-far from linear, $\mathbf{E}_{x,y}[1/2+(1/2)g(x)h(y)f(xy)] \leq 1/2+\epsilon/2$.

   (c) Consider the following k-query test: chose random $x_1, x_2, \ldots, x_k \in \{-1,+1\}^n$ and for *every pair* $i, j \in [k]$ with $i \neq j$, check whether $f(x)f(y)f(xy) = 1$; accept iff *all* of these tests pass. Show that the acceptance probability is 1 when f is linear, and at most $2^{-\binom{k}{2}} + \epsilon$ when f is ϵ-far from linear. Hint: the acceptance probability can be written as

   $$\begin{aligned}\Pr[\text{accept}] &= \mathbf{E}_{x_1,x_2,\ldots,x_k}\left[\prod_{i\neq j}(1/2+(1/2)f(x_i)f(x_j)f(x_ix_j))\right] \\ &= 2^{-\binom{k}{2}} \cdot \sum_{P\subseteq\binom{[k]}{2}} \mathbf{E}_{x_1,x_2,\ldots,x_k}\left[\prod_{(i,j)\in P} f(x_i)f(x_j)f(x_ix_j)\right]\end{aligned}$$

   where $\binom{[k]}{2}$ denotes the set of all pairs $i, j \in [k]$ with $i \neq j$. Argue that for each $P \neq \emptyset$, for any fixed $(i^*, j^*) \in P$, there exist Boolean functions g, h for which

   $$\mathbf{E}_{x_1,x_2,\ldots,x_k}\left[\prod_{(i,j)\in P} f(x_i)f(x_j)f(x_ix_j)\right] \leq \mathbf{E}_{x_{i^*},x_{j^*}}[g(x_{i^*})h(x_{j^*})f(x_{i^*}x_{j^*})].$$

   To obtain g, h you will fix all variables except x_{i^*}, x_{j^*}.

   (d) Conclude that the linearity test described in the previous part has soundness error $s+\epsilon$, and that the amortized query complexity $q/\log_2(1/s)$ can be made arbitrarily close to 1 by taking k sufficiently large.