Exercise 4 – Due 2.2.2006

Purpose

Complete Håstad's analysis of the long-code based three-query PCP. Practice use of Fourier analysis of the long-code.

Definitions

- Let $M > N$ be integers, $f : \{0,1\}^N \to \{\pm 1\}$, $g : \{0,1\}^M \to \{\pm 1\}$ be Boolean functions, and $\pi : [M] \to [N]$ be a mapping. The Håstad test with completeness error $\varepsilon > 0$ and mapping π, applied to f, g is the following three query-bits, linear test:
  - Pick $x \in \{0,1\}^N, y \in \{0,1\}^M$ uniformly at random. Let $(x \circ \pi) \in \{0,1\}^M$ be defined by setting $(x \circ \pi)_j = x_{\pi(j)}$ for all $j \in [M]$.
  - Pick $\mu = (\mu_1, \ldots, \mu_M) \in \{0,1\}^M$ by setting for each $j = 1, \ldots, M$ independently at random $\mu_j = 0$ with probability $1 - \varepsilon$ and $\mu_j = 1$ with probability ε.
  - Let $z = y + (x \circ \pi) + \mu$.
  - Accept iff $g(z) \cdot f(x) \cdot g(y) = 1$.
- For $\alpha \in \{0,1\}^N$, let $|\alpha| = |\{i \mid \alpha_i = 1\}|$.
- For $\gamma \in \{0,1\}^M$, let $\pi_2(\gamma) \in \{0,1\}^N$ be defined by setting for all $i \in [N]$:

$$(\pi_2(\gamma))_i = \left(\sum_{j \in [M], \pi(j) = i} \gamma_j \right) \pmod 2.$$

- We say $f : \{0,1\}^N \to \{\pm 1\}$ is *balanced* if the expectation of $f(x)$ over uniformly chosen $x \in \{0,1\}^N$ is zero (i.e. the truth-table of f has an equal number of plus ones and minus ones).

Questions

1. Prove: If the Håstad test with completeness error ε and mapping π accepts f, g with probability at least $1/2+\eta$ for some $\eta > 0$, then

$$\sum_{\substack{\alpha \in \{0,1\}^N, \beta \in \{0,1\}^M, \\ |\alpha|, |\beta| \leq \mathcal{O}\left(\frac{1}{\varepsilon}\log\left(\frac{1}{\eta}\right)\right), \alpha \subset \pi(\beta)}} \hat{f}_\alpha^2 \cdot \hat{g}_\beta^2 \geq \eta^2. \tag{1}$$

   Some tools you may want to use along the way:

   - Fourier analysis, ortho-normality of Fourier basis and Parseval's equality ($\sum_\alpha \hat{f}_\alpha^2 = 1$) (as we did in class).
   - Definition of $\pi_2(\gamma)$ above.
   - The Cauchy-Schwartz inequality: for any pair of vectors $u, v \in \mathbb{R}^n$:

$$\left(\sum_i u_i v_i\right)^2 \leq \left(\sum_i u_i^2\right) \cdot \left(\sum_i v_i^2\right).$$

   - For all $x \in [0,1]$ we have $1 - x \leq e^{-x}$.

2. In class we used Equation (1) to randomly select elements of $[N], [M]$ as follows. First we pick $\alpha \in \{0,1\}^N$ with probability $\hat{f}_\alpha^2$ and then we pick i uniformly at random from $\{i \mid \alpha_i = 1\}$. The process for selecting $j \in [M]$ from $\hat{g}_\beta^2$ is identical. However, if $|\alpha| = 0$, (i.e. $\alpha = \vec{0}$), we cannot pick i (and Håstad's soundness analysis will fail). To circumvent this problem, prove:

   - If f is balanced then $\hat{f}_{\vec{0}} = 0$.
   - Show there exist domains $D_N \subset \{0,1\}^N, |D| = 2^{N-1}$ and $D_M \subset \{0,1\}^M, |D_M| = 2^{M-1}$ and a modified Håstad test for functions $\tilde{f} : D_N \to \{\pm 1\}, \tilde{g} : D_M \to \{\pm 1\}$ such that if $\tilde{f}, \tilde{g}$ are restrictions of dictatorship functions to D_N, D_M respectively, the modified Håstad test still accepts f, g with probability $\geq 1 - \varepsilon$.
   - Argue (informally) that using the restricted domains and modified functions instead of f, g will solve the problem of large $\hat{f}_{\vec{0}}$ (this method is called 'folding' in PCP literature).