Analysis of Boolean Functions **CMU 18-859S, Spring 2007**

PROBLEM SET 2

Due: Tuesday, February 20

Homework policy: I encourage you to try to solve the problems by yourself. However, you may collaborate as long as you do the writeup yourself and list the people you talked with.

Notation used:

$[n]$	:	the set $\{1, 2, \ldots, n\}$
$x^{(i)}$	:	the n-bit string x with its ith bit flipped, where $i \in [n]$
$x^{(i=b)}$	:	the n-bit string x with its ith bit set to b
$\Delta(x, y)$	:	the Hamming distance between $x, y \in \{-1,1\}^n$; i.e., $\lvert\{i : x_i \neq y_i\}\rvert$
S	:	always a subset of $[n]$, unless otherwise specified
$\Vert f\Vert_2$	:	$= \sqrt{\langle f, f\rangle} = \sqrt{\mathbf{E}[f^2]}$ when $f : \{-1,1\}^n \to \mathbb{R}$ is a function
$\Vert y\Vert_2$	:	when $y \in \mathbb{R}^n$ is a vector, the usual (Euclidean) length of y; i.e., $\sqrt{\sum_i y_i^2}$
f^{odd}	:	when $f : \{-1,1\}^n \to \mathbb{R}$, denotes the function $f^{\mathrm{odd}}(x) = (f(x) - f(-x))/2$
$\mathrm{Inf}_i^{(\rho)}(f)$	:	$= \sum_{S \ni i} \rho^{\lvert S\rvert-1} \hat{f}(S)^2$ when $f : \{-1,1\}^n \to \mathbb{R}$ is a function
$\mathrm{Inf}_i(f)$	:	$= \mathrm{Inf}_i^{(1)}(f)$
$\mathbb{I}(f)$	:	$= \sum_{i=1}^n \mathrm{Inf}_i(f)$ for any $f : \{-1,1\}^n \to \mathbb{R}$
$\deg(f)$	:	$\max\{\lvert S\rvert : \hat{f}(S) \neq 0\}$ for nonzero $f : \{-1,1\}^n \to \mathbb{R}$
$\mathbf{Pr}_{\boldsymbol{x}}$, $\mathbf{E}_{\boldsymbol{x}}$, etc.	:	always denotes Probability, Expectation, (Variance, Covariance, ...) with respect to the *uniform* probability distribution of $\boldsymbol{x}$ on its range, unless otherwise specified

1. Poincaré Inequality II. For $i \in [n]$, the *ith derivative* operator (AKA *ith annihilation operator*) D_i on functions $f : \{-1,1\}^n \to \mathbb{R}$ is defined by letting $\mathrm{D}_i f : \{-1,1\}^n \to \mathbb{R}$ be the function given by

$$(\mathrm{D}_i f)(x) = \frac{f(x^{(i=1)}) - f(x^{(i=-1)})}{2}.$$

(a) Show that D_i acts as the usual derivative with respect to x_i by showing that the Fourier expansion of $\mathrm{D}_i f$ is

$$(\mathrm{D}_i f)(x) = \sum_{S \ni i} \hat{f}(S) x_{S \setminus \{i\}}.$$

Conclude that

$$\mathrm{Inf}_i(f) = \|\mathrm{D}_i f\|_2^2.$$

(b) The *gradient* of f, written $\nabla f : \{-1,1\}^n \to \mathbb{R}^n$, is defined by $\nabla f = (\mathrm{D}_1 f, \mathrm{D}_2 f, \ldots \mathrm{D}_n f)$. Show that

$$\mathop{\mathbf{E}}_{\boldsymbol{x}} \left[\|\nabla f(\boldsymbol{x})\|_2^2 \right] = \mathbb{I}(f).$$

(c) Express $\mathbf{E}_{\boldsymbol{x}} \left[\|\nabla f(\boldsymbol{x})\|_2^2 \right]$ and $\mathbf{Var}[f]$ in terms of weighted sums of squared Fourier coefficients; then conclude that for all $f : \{-1,1\}^n \to \mathbb{R}$,

$$\mathbf{Var}[f] \leq \mathbf{E} \left[\|\nabla f\|_2^2 \right]. \tag{1}$$

(d) Show that when f's range is $\{-1,1\}$, the above result generalizes Problem 1 from Problem Set 1 (when T is treated as -1 and F is treated as 1).

Remark: In analysis, (1) is known as the *Poincaré Inequality* for the discrete cube.

2. Distortion lower bounds for $\boldsymbol{\ell_1 \to \ell_2}$. Without going into what all the words mean, the discrete cube $\{-1,1\}^n$ with the Hamming distance $\Delta(\cdot,\cdot)$ is an example of an ℓ_1 *metric space*. For $D \geq 1$, we say that the discrete cube can be *embedded into* ℓ_2 *with distortion* D if there is a mapping $F : \{-1,1\}^n \to \mathbb{R}^m$ for some $m \in \mathbb{N}$ such that:

- (a) ["no contraction"] $\|F(x) - F(y)\|_2 \geq \Delta(x,y)$ for all x, y, and
- (b) ["expansion at most D"] $\|F(x) - F(y)\|_2 \leq D \cdot \Delta(x,y)$ for all x, y.

In this exercise we will show that every embedding has distortion at least $\sqrt{n}$.

(a) Show that for any $f : \{-1,1\}^n \to \mathbb{R}$,

$$\mathop{\mathbf{E}}_{\boldsymbol{x}}[(f(\boldsymbol{x}) - f(-\boldsymbol{x}))^2] \leq \sum_{i=1}^{n} \mathop{\mathbf{E}}_{\boldsymbol{x}} \left[\left(f(\boldsymbol{x}^{(i=1)}) - f(\boldsymbol{x}^{(i=-1)}) \right)^2 \right], \tag{2}$$

by showing that $\|f^{\mathrm{odd}}\|_2^2 \leq \mathbb{I}(f)$ and then expanding definitions. (In fact, $\|f^{\mathrm{odd}}\|_2^2 \leq \mathbf{Var}[f]$.)

(b) Suppose $F : \{-1,1\}^n \to \mathbb{R}^m$, and write $F(x) = (f_1(x), f_2(x), \dots, f_m(x))$ for functions $f_i : \{-1,1\}^n \to \mathbb{R}$. By summing (2) over $i = 1 \dots m$, show that any F with no contraction must have expansion at least $\sqrt{n}$.

Remark: Congratulations, you've proven the famous Enflo Bound[1]: Embedding ℓ_1 metrics on N points into ℓ_2 can require distortion $\sqrt{\log N}$. A nearly matching upper bound of $O(\sqrt{\log N} \log\log N)$ was proven only two years ago, via CS-theory methods.

3. Noise stability. For $\rho \in [-1,1]$, the *noise operator* T_ρ (AKA *Bonami-Beckner operator*) on functions $f : \{-1,1\}^n \to \mathbb{R}$ is defined by letting $T_\rho f : \{-1,1\}^n \to \mathbb{R}$ be the function given by

$$(T_\rho f)(x) = \mathop{\mathbf{E}}_{\boldsymbol{y} \sim_\rho x}[f(\boldsymbol{y})],$$

where the notation $\boldsymbol{y} \sim_\rho x$ means that $\boldsymbol{y}$ is a *ρ-correlated copy of x*: Each $\boldsymbol{y}_i$ is chosen independently via

$$\boldsymbol{y}_i = \begin{cases} x_i & \text{with probability } \frac{1}{2} + \frac{1}{2}\rho, \\ -x_i & \text{with probability } \frac{1}{2} - \frac{1}{2}\rho. \end{cases}$$

Note that when $\rho \geq 0$ this is the same as saying $\boldsymbol{y}_i$ is set to x_i with probability ρ and is set uniformly at random with probability $1 - \rho$. We further define the *noise stability of f and g at ρ* to be

$$\mathbb{S}_\rho(f, g) = \langle f, T_\rho g \rangle = \mathop{\mathbf{E}}_{\boldsymbol{x}}[f(\boldsymbol{x})(T_\rho g)(\boldsymbol{x})],$$

and define $\mathbb{S}_\rho(f) = \mathbb{S}_\rho(f, f)$ to be the *noise stability of f at ρ*.

(a) Show that the Fourier expansion of $T_\rho f$ is

$$(T_\rho f)(x) = \sum_{S \subseteq [n]} \rho^{|S|} \hat{f}(S) x_S.$$

Conclude that

$$\mathbb{S}_\rho(f, g) = \sum_{S \subseteq [n]} \rho^{|S|} \hat{f}(S) \hat{g}(S).$$

(b) Let $f : \{-1,1\}^n \to \{-1,1\}$ and let $\epsilon \in [0,1]$. Define the *noise sensitivity of f at ϵ* to be

$$\mathbb{NS}_\epsilon(f) = \mathop{\mathbf{Pr}}_{\boldsymbol{x}, \boldsymbol{y}}[f(\boldsymbol{x}) \neq f(\boldsymbol{y})],$$

where $\boldsymbol{x}$ and $\boldsymbol{y}$ are chosen by first choosing $\boldsymbol{x}$ uniformly at random and then forming $\boldsymbol{y}$ by flipping each bit of $\boldsymbol{x}$ with probability ϵ. Show that

$$\mathbb{NS}_\epsilon(f) = \tfrac{1}{2} - \tfrac{1}{2}\mathbb{S}_{1-2\epsilon}(f).$$

(c) Show that for any $f : \{-1,1\}^n \to \mathbb{R}$ we have $\mathrm{Inf}_i^{(\rho)}(f) = \mathbb{S}_\rho(\mathrm{D}_i f)$, and also $\sum_{i=1}^n \mathrm{Inf}_i^{(\rho)}(f) = \frac{\partial}{\partial\rho}\mathbb{S}_\rho(f)$.

[1] Per Enflo, *On the nonexistence of uniform homeomorphisms between L_p-spaces*, Arkiv för matematik 8, 1969.

4. Noam and Mario's bound. Let $f : \{-1,1\}^n \to \mathbb{R}$ be a nonzero function with $\deg(f) \leq d$.

(a) Generalize Problem 3(b) from Problem Set 1 by showing that $\mathbf{Pr}_{\boldsymbol{x}}[f(\boldsymbol{x}) \neq 0] \geq 2^{-d}$. (Hint: same.)

(b) Show that if in addition f maps into $[-1,1]$ then $\mathbb{I}(f) \leq d$.

(c) Show that if in addition f is boolean-valued (maps into $\{-1,1\}$) then f is a $d2^{d-1}$-junta.

(d) The *address function with k address bits* is $\mathrm{Addr}_k : \{-1,1\}^{k+2^k} \to \{-1,1\}$ defined by

$$\mathrm{Addr}_k(x_1, \dots, x_k, y_1, \dots, y_{2^k}) = y_x,$$

where $x = (x_1, \dots, x_k)$ is identified with a number in $[2^k]$. Show that $\deg(\mathrm{Addr}_k) = k + 1$. Conclude that the junta size in (c) must be at least $2^{d-1} + d - 1$.

5. Quasirandomness implies low correlation with juntas.

(a) Recall that for $f, g : \{-1,1\}^n \to \mathbb{R}$, $\mathbf{Cov}[f,g] = \mathbf{E}_{\boldsymbol{x}}[f(\boldsymbol{x})g(\boldsymbol{x})] - \mathbf{E}_{\boldsymbol{x}}[f(\boldsymbol{x})]\,\mathbf{E}_{\boldsymbol{x}}[g(\boldsymbol{x})]$. Give a formula for $\mathbf{Cov}[f,g]$ in terms of the Fourier coefficients of f and g.

(b) Show that if $h : \{-1,1\}^n \to [-1,1]$ is (ϵ, δ)-quasirandom, then $\mathbf{Cov}[h,f] < \sqrt{r/(1-\delta)^r}\sqrt{\epsilon}$ for every r-junta $f : \{-1,1\}^n \to \{-1,1\}$. Explain why this result is trivial if $r \geq \ln(1/\epsilon)/\delta$.

(Hints: You may need: (a) Cauchy-Schwarz, $\sum a_i b_i \leq \sqrt{\sum a_i^2}\sqrt{\sum b_i^2}$; (b) the inequality $(1-x)^y \leq \exp(-xy)$ for $0 \leq x < 1$, $y \geq 0$.)

6. PCPPs may as well use Or$_3$. Suppose a property $\mathcal{P}$ of m-bit strings has PCPPs of length $\ell(m)$. Show that it has PCPPs of length $\mathrm{poly}(\ell(m))$ in which the tester makes 3 queries and then uses one of the 8 possible Or$_3$ predicates: $v_{i_1} \vee v_{i_2} \vee v_{i_3}, \quad \overline{v}_{i_1} \vee v_{i_2} \vee v_{i_3}, \quad \dots, \quad \overline{v}_{i_1} \vee \overline{v}_{i_2} \vee \overline{v}_{i_3}$.

7. A hardness reduction that doesn't work. Suppose we try the following alternate reduction in attempt to prove that the Unique Games Conjecture implies $1 - \eta$ vs. $\frac{1}{2} + \eta$ hardness for Max-3Lin for every $\eta > 0$. Given a CSP $\mathcal{G} = (V, E)$ over $[k]$ with unique constraints, we reduce it to a "3Lin" tester over $(f_v : \{-1,1\}^n \to \{-1,1\})_{v \in V}$ as follows: the tester picks an edge $(v,w) \in E$ uniformly at random and then does the Håst-Odd$_\delta$ test on the collection $\{f_v^{\mathrm{odd}}, f_w^{\mathrm{odd}} \circ \sigma_{v \to w}\}$ where $\sigma_{v \to w}$ is the edge constraint on (v,w). (Recall that $\sigma_{v \to w}$ acts on strings $x \in \{-1,1\}^k$ by $\sigma_{v \to w}(x) = y$, where $y_j = x_{\sigma^{-1}_{v \to w}(i)}$.)

(a) Show that the first part of the proof works out even better: The reduction maps $\mathcal{G}$ instances with $\mathrm{val}(\mathcal{G}) \geq 1 - \lambda$ into 3Lin CSPs with value at least $1 - \delta - \lambda$.

(b) Show that the second part of the proof can never work. Specifically, show that *regardless* of what $\mathcal{G}$ is, the resulting 3Lin system has an assignment with value at least $5/8 - \delta/4$.

Bonus Problem: Let $p(x_1, \dots, x_n)$ be a multilinear polynomial over the reals of degree at most d. Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ be independent real random variables satisfying $\mathbf{E}[\boldsymbol{X}_i] = 0$, $\mathbf{E}[\boldsymbol{X}_i^2] = 1$, $\mathbf{E}[\boldsymbol{X}_i^3] = 0$, $\mathbf{E}[\boldsymbol{X}_i^4] \leq 9$ for each $i \in [n]$. (For example, $(\boldsymbol{X}_1, \dots, \boldsymbol{X}_n)$ chosen uniformly at random from $\{-1, 1\}^n$ would be fine.) Show that $\mathbf{E}[p(\boldsymbol{X}_1, \dots, \boldsymbol{X}_n)^4] \leq 9^d \mathbf{E}[p(\boldsymbol{X}_1, \dots, \boldsymbol{X}_n)^2]^2$. (Hint: induction.)