CS 351: Problem Set 6

Assigned 11/21/2002, Due 12/05/2002, Usual rules apply

Linearity Testing and Fourier Transforms

Let $f : \mathbf{Z}_2^n \longrightarrow \mathbf{Z}_2$. We proved in class that there exists a constant $\delta_0 > 0$ such that for any $\delta < \delta_0$, if $\Pr[f(x) + f(y) \neq f(x+y)] \leq \delta$, then there is a linear function $g : \mathbf{Z}_2^n \longrightarrow \mathbf{Z}_2$ such that f and g disagree on $\leq 2\delta$ fraction of $\mathbf{Z}_2^n$. Specifically, we proved this with $\delta_0 = 1/6$.

Blum, Luby, and Rubinfeld (Journal of Computer and System Sciences, vol. 47, no. 3, pages 549–595, 1993) showed a somewhat stronger and a more general form of the bound.

Let $(G, +)$ and $(H, \star)$ be two groups, let $f : G \longrightarrow H$, and let $\mathbf{Fail}(f) = \Pr[f(x) \star f(y) \neq f(x+y)]$. For $f, g : G \longrightarrow H$, let $\mathbf{Dist}(f,g) = \Pr_{x \in G}[f(x) \neq g(x)]$, when x is uniformly disributed in G. Let $\mathbf{Lindist}(f)$ denote the distance of f to the nearest homomorphism, that is $\mathbf{Lindist}(f) = \min\{\mathbf{Dist}(f,g) \mid g \text{ is a homomorphism from } G \text{ into } H\}$. Recall that a function $g : G \longrightarrow H$ is said to be a homomorphism if for all $x, y \in G$, $g(x+y) = g(x) \star g(y)$.

(1) Work through the proof from class to show that if $d = \mathbf{Lindist}(f)$, then $\mathbf{Fail}(f) \geq (2/9)d$ for all $d \leq 1/2$.

(2) Now assume that $f : \mathbf{Z}_2^n \longrightarrow \mathbf{Z}_2$, and show that if $d = \mathbf{Lindist}(f)$, then $\mathbf{Fail}(f) \geq 3d - 6d^2$ for all $d \leq 1/2$. Again, do this by working through the steps of the proof with a view toward optimizing it for the special case $(G, +) = (\mathbf{Z}_2^n, +)$ and $(H, \star) = (\mathbf{Z}_2, +)$. (Note that addition in $\mathbf{Z}_2^n$ is technically different from addition in $\mathbf{Z}_2$—addition in $\mathbf{Z}_2^n$ is really just component-wise $\mathbf{Z}_2$ addition. To simplify notation, we will simply write "+" for both.)

In fact, it is known that for the case of $f : \mathbf{Z}_2^n \longrightarrow \mathbf{Z}_2$, if $d = \mathbf{Lindist}(f)$, then we have $\mathbf{Fail}(f) \geq d$ for all $d \leq 1/2$. To prove this, we will develop the concept of the *Fourier tranform* of Boolean functions, analogous to the Fourier transform from analysis.

Consider the space of functions from $\mathbf{Z}_2^n$ into $\mathbf{R}$, that is, the set $\mathcal{F}_n = \{f : \mathbf{Z}_2^n \longrightarrow \mathbf{R}\}$. Each $f \in \mathcal{F}_n$ can be thought of as a vector of 2^n real values. Thus it is easy to see that $\mathcal{F}_n$ is a vector space of dimension 2^n, whose standard basis consists of the so-called "delta" functions: for each $x \in \mathbf{Z}_2^n$, $e_x : \mathbf{Z}_2^n \longrightarrow \mathbf{Z}_2$ is defined by

$$e_x(y) = \begin{cases} 1 & \text{if } y = x \\ 0 & \text{otherwise} \end{cases}$$

We will endow this space with the inner product defined by $(f,g) = 2^{-n} \sum_{x \in \mathbf{Z}_2^n} f(x)g(x)$. For convenience, we will view a Boolean function on $\mathbf{Z}_2^n$ as taking on the values $\{-1, +1\}$.

The *Fourier basis* for $\mathcal{F}_n$ is given by the functions χ_α, for $\alpha \in \mathbf{Z}_2^n$, defined as follows. Let $\langle \cdot, \cdot \rangle$ denote the inner product in $\mathbf{Z}_2^n$, that is, for $x, y \in \mathbf{Z}_2^n$, $\langle x, y \rangle \in \mathbf{Z}_2$, and is given by $\langle x, y \rangle = \sum_i x_i y_i$ (where the sum and product are modulo 2 operations). For $\alpha \in \mathbf{Z}_2^n$, the function $\chi_\alpha : \mathbf{Z}_2^n \longrightarrow \mathbf{R}$ is defined by $\chi_\alpha(x) = (-1)^{\langle x, \alpha \rangle}$.

(3a) Prove that for every $\alpha \in \mathbf{Z}_2^n$, $(\chi_\alpha, \chi_\alpha) = 1$, that is, $\sum_x \chi_\alpha(x)\chi_\alpha(x) = 1$.

(3b) Prove that for $\alpha \neq \beta \in \mathbf{Z}_2^n$, $(\chi_\alpha, \chi_\beta) = 0$, that is, $\sum_x \chi_\alpha(x)\chi_\beta(x) = 0$.

Hints for (3a) and (3b). You might find the following facts helpful.

Fact (1) For any $\alpha, \gamma, x, y \in \mathbf{Z}_2^n$, $\chi_\alpha(x)\chi_\gamma(x) = \chi_{\alpha+\gamma}(x)$, and similarly $\chi_\alpha(x)\chi_\alpha(y) = \chi_\alpha(x+y)$. Both of these are trivial consequences of the fact that the inner product that maps $\mathbf{Z}_2^n \times \mathbf{Z}_2^n$ into $\mathbf{Z}_2$ is a linear function.

Fact (2) For $\alpha \neq 0^n$, $\sum_x \chi_\alpha(x) = 0$.

(3c) Prove fact (2) mentioned above, that is, for $\alpha \neq 0^n$, $\sum_x \chi_\alpha(x) = 0$.

Thus there are 2^n $(= \dim(\mathcal{F}_n))$ distinct functions χ_α, which form an orthonormal basis of $\mathcal{F}_n$. This is called the *Fourier basis* of $\mathcal{F}_n$. Now any function $f : \mathbf{Z}_2^n \longrightarrow \mathbf{R}$ may be expressed in this basis as $f = \sum_{\alpha \in \mathbf{Z}_2^n} \widehat{f}_\alpha \chi_\alpha$. The orthonormality of the Fourier basis implies that for each α, $\widehat{f}_\alpha = (f, \chi_\alpha) = 2^{-n} \sum_x f(x)\chi_\alpha(x)$. The coefficients $\{\widehat{f}_\alpha\}_{\alpha \in \mathbf{Z}_2^n}$ are called the *Fourier coefficients* of f.

(3d) Prove the Plancherel theorem: for $f, g \in \mathcal{F}_n$, $\sum_\alpha \widehat{f}_\alpha \widehat{g}_\alpha = (f, g)$.

A consequence of the Plancherel theorem is *Parseval's identity,* which, for the special case of Boolean functions $f : \mathbf{Z}_2^n \longrightarrow \{-1, +1\}$, implies that $\sum_\alpha \widehat{f}_\alpha^2 = 1$.

There is a natural reason for employing the Fourier transform in studying the linearity test for a function $f : \mathbf{Z}_2^n \longrightarrow \{-1, +1\}$. The linear functions from $\mathbf{Z}_2^n$ into $\{-1, +1\}$ are precisely the basis functions in the Fourier basis, so it is natural to hope that the Fourier coefficients reveal information about how much f agrees with each linear function χ_α. Indeed, the following facts are obvious, keeping in mind that with the ± 1 convention for Boolean values, the test "$f(x) + f(y) = f(x+y)$" can be rewritten as "$f(x)f(y)f(x+y) = 1$."

$\mathbf{Fail}(f) = 2^{-2n} \sum_{x,y} \left(\frac{1 - f(x)f(y)f(x+y)}{2}\right)$, since the quantity inside the summation is 1 if the test fails at x, y, and zero otherwise.

For $f, g : \mathbf{Z}_2^n \longrightarrow \{-1, +1\}$,

$\mathbf{Agreement}(f, g) = 2^{-n} \sum_x (1 + f(x)g(x))/2$,

$\mathbf{Disagreement}(f, g) = 2^{-n} \sum_x (1 - f(x)g(x))/2$, so

$\mathbf{Agreement}(f, g) - \mathbf{Disagreement}(f, g) = 2^{-n} \sum_x f(x)g(x) = (f, g)$.

In particular,

$\mathbf{Agreement}(f, \chi_\alpha) - \mathbf{Disagreement}(f, \chi_\alpha) = 1 - 2 \cdot \mathbf{Disagreement}(f, \chi_\alpha) = \widehat{f}_\alpha$.

Consequently, $1 - 2 \cdot \mathbf{Lindist}(f) = \max_\alpha \widehat{f}_\alpha$.

(3e) Show that $\mathbf{Fail}(f) \geq \mathbf{Lindist}(f)$. (Hint: Write $\mathbf{Fail}(f)$ as given above, expand each of $f(x)$, $f(y)$, and $f(x+y)$ in the Fourier basis, e.g., as $f(x) = \sum_\alpha \widehat{f}_\alpha \chi_\alpha(x)$. Then exchange the order of summations, and simplify. You will find Parseval's identity, plus Facts (1) and (2) from above, helpful.