0368.4163 Randomness and Computation May 20, 2009

Lecture 9

Lecturer: Ronitt Rubinfeld *Scribe: Mateus de Oliveira Oliveira and Daniel Shahaf*

1 The Boolean Function

$$f : \{0,1\}^n \to \{0,1\}$$
$$f : \{\pm 1\}^n \to \{\pm 1\}$$

Can be viewed as: a truth table, a circuit, a 2-coloring of the n-dimensional discrete cube, an indicator of a set ($f(x) = 1 \iff x \in S$).

Some concepts that are studied: "simple" functions — k-juntas and dictatorships (only k inputs, respectively one input, affect the function's value); fairness (each input bit has the 'same' influence over the output) and noise-sensitivity (behavior under flipping of some input bits); symmetry (behavior under permutations of inputs).

1.1 Linear (homomorphic) functions

Our goal today: linearity testing (homomorphism testing): to decide whether a function f, given as a blackbox (oracle), is *linear* (*homomorphic*).

Definition 1 *A function $f : \{0,1\}^n \to \{0,1\}$ is* linear *(*homomorphic*) if*

$$f(x) + f(y) = f(x + y)$$

for every $x, y \in \{0,1\}^n$.

(The addition $x + y$ is addition modulo 2 in the vector space $(\mathbb{Z}_2)^n$; that is, $x + y = (x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 \oplus y_1, \ldots, x_n \oplus y_n)$.)

Examples

- The constant function $f(x) \equiv 0$ is homomorphic ($f(x) + f(y) = 0 + 0 = 0 = f(x + y)$).
- The constant function $f(x) \equiv 1$ is not ($1 + 1 \neq 1$).
- The projection function $f(x) = x_i$ (for some fixed i) is a homomorphism.
- As is the function $f(x) = \bigoplus_{i=1}^n x_i$.

Claim 2 *A function $f : \{0,1\}^n \to \{0,1\}$ is homomorphic iff it is one of the functions $f_S(x) = \bigoplus_{i \in S} x_i$ (for $S \subset [n]$).*

Sketch of Proof Every homomorphic function is uniquely determined by the values on the vectors $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ (with 1 at the ith coordinate, $1 \leq i \leq n$). Since there are 2^n possible settings for the values $f(e_i)$ ($1 \leq i \leq n$), there are 2^n linear functions. It is easy to see that all functions f_S are linear, and there are $|\{S : S \subset [n]\}| = 2^n$ of them. ■

Note If $S = \emptyset$ then $f_S(x) \equiv 1$, i.e., $f_\emptyset$ is a constant function.

1.2 Testing for linearity

Given a function f as a blackbox, in order to check whether or not it is linear, we have to query it on *every* possible input. For example, the function given by $f(x) = 1$ if $x = e_{17}$ and $f(x) = 0$ otherwise is not homomorphic, but agrees with the zero homomorphism everywhere except at $x = e_{17}$.

Definition 3 *A function f is* ϵ-close to linear *if there exists a linear function g that agrees with f on all but an ϵ-fraction of the domain; that is,*

$$\Pr_x[f(x) = g(x)] = \frac{|\{x : f(x) = g(x)\}|}{2^n} \geq 1 - \epsilon.$$

Otherwise, f is ϵ-far from linear.

1.2.1 Proposed tester

- Repeat $r = O(\frac{1}{\epsilon} \log \frac{1}{\delta^*})$ times:
  - Pick $x, y \in_R \{0,1\}^n$ independently and uniformly.
  - If $f(x) + f(y) \neq f(x + y)$:
    - Output fail and halt.
- Output pass.

1.2.2 Analysis

Claim 4 *f is linear if and only if* $\Pr[\mathit{pass}] = 1$.

Claim 5 *If* $\Pr_{x,y}[f(x) + f(y) \neq f(x + y)] \geq \epsilon$ *then* $\Pr[\mathit{fail}] \geq 1 - \delta^*$.

Claim 6 *If f is ϵ-close to linear, then the test fails with probability at most 3ϵ.*

Proof Idea Let A_x denote the event $f(x) \neq g(x)$. Then $\Pr_x[A_x] \leq \epsilon$ and thus $\Pr[\mathsf{fail}] \leq \Pr_{x,y}[A_x \vee A_y \vee A_{x+y}] \leq 3\epsilon$ by union bound. ■

Plan By claim 4, the test fails with probabiliy zero iff the distance of f from linear is zero. By claim 6, a similar relation also holds — in one direction — if we say 'small' instead of 'zero'. The remainder of this lecture shows the converse of claim 6.

1.3 Notational switch

We now consider boolean functions as $f : \{\pm 1\}^n \to \{\pm 1\}$ rather than $f : \{0,1\}^n \to \{0,1\}$: we map $0 \mapsto +1$ and $1 \mapsto -1$, and write the operation as multiplication ($x \cdot y = (x_1 y_1, \ldots, x_n y_n)$ for $x, y \in \{\pm 1\}^n$) rather than addition ($x + y = (x_1 \oplus y_1, \ldots, x_n \oplus y_n)$ for $x, y \in \{0,1\}^n$). (In other words, we switch our representation from the group $\mathbb{Z}_2$ of integers modulo 2 to the group μ_2 of square roots of unity.)

Example The homomorphic functions are now written as $f_S(x) = \prod_{i \in S} x_i$. The rejection condition of the proposed linearity tester is $f(x) \cdot f(y) \neq f(x \cdot y)$, where $x \cdot y$ is as defined in the previous paragraph (and *not* an inner product).

1.4 Rejection probability

It will be more convenient to represent the rejection condition of the proposed tester in terms of equality rather than inequality. We have the equivalence:

$$f(x)f(y) \neq f(xy) \iff \big(f(x)f(y)\big)f(xy) = -1$$

which suggests to consider the following indicator:

$$\frac{1 - f(x)f(y)f(xy)}{2} = \begin{cases} 0, & \text{if the test accepts;} \\ 1, & \text{if the test rejects.} \end{cases}$$

We also define

$$\delta = \mathrm{Exp}_{x,y}\left[\frac{1 - f(x)f(y)f(xy)}{2}\right]$$

as the *rejection probability* of one loop-iteration of the proposed tester. This gives the *acceptance probability* of one loop-iteration of that tester as:

$$1 - \delta = \mathrm{Exp}_{x,y}\left[\frac{1 + f(x)f(y)f(xy)}{2}\right].$$

2 Basics of Fourier analysis of parity functions

$\mathcal{G} = \{g : \{\pm 1\}^n \to \mathbb{R}\}$ is a 2^n-dimensional vector space (over the field $\mathbb{R}$, i.e., linear combinations are to be taken with real coefficients). This space is equipped with the inner product

$$\langle f, g \rangle = \frac{1}{2^n} \sum_{x \in \{\pm 1\}^n} f(x)g(x).$$

2.1 Looking for a basis

We look for a convenient basis of $\mathcal{G}$.

- The first idea is the *indicator functions*: the functions e_a (for $a \in \{\pm 1\}^n$) given by $e_a(x) = 1$ if $a = x$ and 0 otherwise.

  It is easy to see that $\{e_a : a \in \{\pm 1\}^n\}$ is a basis, and that $g = \sum_a g(a) \cdot e_a$ (i.e., $g(x) = \sum_a g(a) \cdot e_a(x)$) for any function g.

- However, the basis of *character functions* $\chi_S(x) = \prod_{i \in S} x_i$ will be more convenient. (These are the functions we used to call f_S.)

Lemma 7 *$\{\chi_S : S \subset [n]\}$ is an orthonormal basis.*

Proof Let $S \neq T$ be two distinct subsets of $[n]$, and let $j \in S \triangle T = \{x : (x \in S) \neq (x \in T)\}$. Denote "$x$ with the jth bit flipped" by '$x^{\oplus j}$'. Then

$$\langle \chi_S, \chi_S \rangle = \frac{1}{2^n} \sum_x \underbrace{\chi_S(x)^2}_{=1} = 1$$

and

$$\begin{aligned}
\langle \chi_S, \chi_T \rangle &= \frac{1}{2^n} \sum_x \chi_S(x)\chi_T(x) = \frac{1}{2^n} \sum_x \left(\prod_{i \in S} x_j \cdot \prod_{j \in T} x_j \right) \\
&= \frac{1}{2^n} \sum_x \prod_{i \in S \triangle T} x_i \qquad \text{(because } \{x_i : i \in S \cap T\} \text{ cancel out)} \\
&= \frac{1}{2^n} \sum_{\{x, x^{\oplus j}\}} \left(\prod_{i \in S \triangle T} x_i + \prod_{i \in S \triangle T} (x^{\oplus j})_i \right) \\
&= \frac{1}{2^n} \sum_{\{x, x^{\oplus j}\}} \left(x_j \cdot \prod_{j \neq i \in S \triangle T} x_i + \overline{x_j} \cdot \prod_{j \neq i \in S \triangle T} (x^{\oplus j})_i \right) \\
&= \frac{1}{2^n} \sum_{\{x, x^{\oplus j}\}} \left(x_j \cdot \prod_{j \neq i \in S \triangle T} x_i + \overline{x_j} \cdot \prod_{j \neq i \in S \triangle T} x_i \right) \\
&= \frac{1}{2^n} \sum_{\{x, x^{\oplus j}\}} (x_j + \overline{x_j}) \left(\prod_{i \in S \triangle T, i \neq j} x_i \right) = \frac{1}{2^n} \sum 0 = 0.
\end{aligned}$$

■

Remark The technique of separating out x_j and its complement is an example of a *pairing argument*. It considers together all pairs of words that differ only on a specific coordinate; for instance, $(+1, +1, -1, +1)$ with $(+1, +1, +1, +1)$, $(+1, +1, -1, -1)$ with $(+1, +1, +1, -1)$, $(-1, -1, -1, +1)$ with $(-1, -1, +1, +1)$, etc.

Corollary 8 *We can write every function f as $f = \sum_{S \subset [n]} \hat{f}(S)\chi_S$, where $\hat{f}(S) = \langle f, \chi_S \rangle$.*

For example, if $f : x \mapsto x_i$ is the projection function, we have that $f = \chi_i$, thus the Fourier coefficients of f are $\hat{f}(S) = \langle \chi_i, \chi_S \rangle$ which is equal to 1 if $S = \{i\}$ and 0 otherwise. Similarly, if $f : x \mapsto 1$ is the constant function, then $f = \chi_\emptyset$ and $\hat{f}(S)$ will be equal to 1 if $S = \emptyset$ and 0 otherwise.

2.2 Some useful facts about the Fourier Transform

Lemma 9 $\chi_S \cdot \chi_T = \chi_{S \triangle T}$

Lemma 10 *Fourier Coefficient of any parity function*

$$f(x) = \chi_S(x) \Leftrightarrow \forall Z \subseteq [n], \quad \hat{f}(Z) = \begin{cases} 1 & \textit{when } S = Z \\ 0 & \textit{Otherwise} \end{cases}$$

Lemma 11 *Agreement with linear functions vs max Fourier coefficient*

$$\hat{f}(S) = 1 - 2\Pr[f(x) \neq \chi_S(x)] \Leftrightarrow \mathrm{DIST}(f, \chi_S) = \frac{1 - \hat{f}(S)}{2}$$

or equivalently

$$\hat{f}(S) = -1 + 2\Pr[f(x) \neq \chi_S(x)] \Leftrightarrow \mathrm{DIST}(f, \chi_S) = \frac{1 - \hat{f}(S)}{2}$$

Proof

Its enough to prove that

$$\mathrm{DIST}(f, \chi_S) = \Pr_{x \in \{\pm 1\}^n}[f(x) - \chi_S(x)].$$

The proof of this fact proceeds as follows:

$$\begin{aligned}
\hat{f}(S) &= \tfrac{1}{2^n} \textstyle\sum_x f(x)\chi_S(x) \\
&= \tfrac{1}{2^n} \left[\textstyle\sum_{x, f(x) = \chi_S(x)} 1 + \sum_{x, f(x) \neq \chi_S(x)} -1\right] \\
&= (1 - \mathrm{DIST}(f, \chi_S)) \cdot 1 + \mathrm{DIST}(f, \chi_S) \cdot (-1) \\
&= 1 - 2\,\mathrm{DIST}(f, \chi_S)
\end{aligned} \tag{1}$$

■

Lemma 12 *If* $S \neq T$ *then* $\mathrm{DIST}(\chi_S, \chi_T) = \frac{1}{2}$.

Proof Let $f = \chi_T$. Then

$$\begin{aligned}
\hat{f}(S) &= 0 \quad (\text{ by lemma } 10) \\
&= 1 - 2\,\mathrm{DIST}(f, \chi_S) \quad (\text{by lemma } 11) \\
&\Rightarrow \mathrm{DIST}(f, \chi_S) = \tfrac{1}{2} \\
&\Rightarrow \mathrm{DIST}(\chi_T, \chi_S) = \tfrac{1}{2}
\end{aligned} \tag{2}$$

■

A very important theorem in Fourier Analysis is the following:

Theorem 13 (Plancherel's theorem) *Let* $f, g : \{\pm 1\} \to \mathbb{R}$. *Then*

$$\langle f, g \rangle = \mathrm{Exp}_{x \in \{\pm 1\}^n}[f(x)g(x)] = \sum_{S \subseteq [n]} \hat{f}(S)\hat{g}(S).$$

Proof

$$\begin{aligned}
\langle f, g \rangle &= \langle \textstyle\sum_S \hat{f}(S)\chi_S, \sum_T \hat{g}(T)\chi_T \rangle \\
&= \textstyle\sum_S \sum_T \hat{f}(S)\hat{g}(T)\langle \chi_S, \chi_T \rangle \quad \text{by bilinearity of } \langle , \rangle \\
&= \textstyle\sum_S \hat{f}(S)\hat{g}(S) \quad (\text{because } \langle \chi_S, \chi_T \rangle = 1 \text{ if } S = T \text{ and } 0 \text{ if } S \neq T)
\end{aligned}$$

■

We call special attention to the following corollary of Plancherel's theorem:

Corollary 14 (Parseval's Theorem) *If* $f : \{\pm 1\}^n \to \mathbb{R}$ *then* $\langle f, f \rangle = \mathrm{Exp}[f(x)^2] = \sum_S \hat{f}(S)^2$.

Which for boolean functions $f : \{0,1\}^n \to \{\pm 1\}$ reduces to the next corollary, by observing that in this case $f(x)^2 = 1$ for every x.

Corollary 15 (Boolean Parseval's Theorem) *If* $f : \{\pm 1\}^n \to \{\pm 1\}$ *then* $\sum_S \hat{f}(S)^2 = 1$.

Lemma 16 $\mathrm{Exp}[f] = \mathrm{Exp}[f(x) \cdot 1] = \hat{f}(\emptyset)\chi_\emptyset(\emptyset) = \hat{f}(\emptyset)$.

Lemma 17 $\mathrm{Exp}[\chi_S(x)] = \begin{cases} 1 & \textit{if } S = \emptyset \\ 0 & \textit{Otherwise} \end{cases}$

3 Linearity Testing

The goal of this section is to prove the converse of claim 6, i.e, to show that if f is ϵ-far from linear, then the probability that the algorithm described in subsection 1.2.1 finds two x, y for which $f(x+y) \neq f(x) + f(y)$ is high. More precisely,

$$\Pr[f(x)f(y)f(x \cdot y) = -1] \geq \epsilon$$

Lemma 18 (Main Lemma)

$$1 - \delta = \Pr[f(x)f(y)f(xy) = 1] = \frac{1}{2} + \frac{1}{2}\sum_{S \in [n]} \hat{f}(s)^3$$

Proof

$$1 - \delta = \mathrm{Exp}_{xy}\left[\frac{1 + f(x)f(y)f(xy)}{2}\right] = \frac{1}{2} + \frac{1}{2}\mathrm{Exp}_{xy}[f(x)f(y)f(xy)]$$

and

$$\begin{aligned}
\mathrm{Exp}_{xy}[f(x)f(y)f(xy)] &= \mathrm{Exp}_{xy}[(\textstyle\sum_S \hat{f}(S)\chi_S(x))(\sum_T \hat{f}(T)\chi_T(y))(\sum_U \hat{f}(U)\chi_T(xy))] \\
&= \mathrm{Exp}_{xy}[\textstyle\sum_{STU} \hat{f}(S)\hat{f}(T)\hat{f}(U)\chi_S(x)\chi_T(y)\chi_U(xy)] \\
&= \textstyle\sum_{STU} \hat{f}(S)\hat{f}(T)\hat{f}(U)\mathrm{Exp}[\chi_S(x)\chi_T(y)\chi_U(xy)] \\
&= \textstyle\sum_{S=T=U} \hat{f}(S)^3.
\end{aligned}$$

The last equality follows from the fact that
$\mathrm{Exp}_{xy}[\chi_S(x)\chi_T(y)\chi_U(xy)] = \mathrm{Exp}[\chi_S(x)\chi_U(x)] \cdot \mathrm{Exp}[\chi_T(y)\chi_U(y)] = \begin{cases} 1 & \text{if } S = U \text{ and } T = U \\ 0 & \text{otherwise} \end{cases}$
■

Now we are ready to prove the goal stated in the beginning of this section.
Proof Assume $\Pr[f(x)f(y)f(xy) = -1] < \epsilon$. Then we show that f is ϵ-close to linear.

$$1 - \epsilon = \Pr[f(x)f(y)f(xy) = 1] = \frac{1}{2} + \frac{1}{2}\sum_{S \subseteq [n]} \hat{f}(S)^3$$

then

$$
\begin{aligned}
1 - 2\epsilon &\leq \textstyle\sum_{S \subseteq [n]} \hat{f}(S)^3 \\
&\leq \max_S \hat{f}(S) \textstyle\sum_{S \subseteq [n]} \hat{f}(S)^2 \\
&\leq \max_S \hat{f}(S)
\end{aligned}
$$

Now let T be such that $\hat{f}(T) = \max_S \hat{f}(S)$. Then $1 - 2\epsilon \leq \hat{f}(T)$. By lemma 11 $\mathrm{DIST}(f, \chi_T) \leq \epsilon$. ■