0368.416701 Sublinear Time Algorithms	December 14, 2009
Lecture 9	
Lecturer: Ronitt Rubinfeld	*Scribe: Arie Zilberstein*

1 Lecture Outline

- Testing linearity of boolean functions.
- Fourier analysis basics.

2 Definitions

A function f is a *boolean function* if

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

Boolean functions are "all things to all people": They have numerous uses in all fields of computer sciences.

Definition 1 *A function $f : \{0,1\}^n \to \{0,1\}$ is* linear *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, $x+y = (x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 \oplus y_1, \ldots, x_n \oplus y_n)$.)

Important example. The constant function $f(x) \equiv 0$ is linear ($f(x)+f(y) = 0+0 = 0 = f(x+y)$).

Another example. For every constant y, $\chi_y(x) \stackrel{\text{def}}{=} \sum_{i=1\ldots n} x_i y_i \pmod 2 = x \cdot y$ is a linear function.

Claim 2 *A function $f : \{0,1\}^n \to \{0,1\}$ is linear $\Leftrightarrow$ it is one of the functions $\chi_y(x)$.*

Proof ($\Leftarrow$) We verify that $\chi_y(x)$ is linear:

$$\chi_y(x_1) + \chi_y(x_2) = x_1 y + x_2 y = (x_1 + x_2)y = \chi_y(x_1 + x_2).$$

($\Rightarrow$) As we just proved, all $\chi_y(x)$ functions are linear. There are exactly 2^n such functions (choices of y). Now, let f be a linear function: f is uniquely determined by the values on the vectors $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$ (with 1 at the ith coordinate, $1 \le i \le n$). Since there are 2^n possible settings for the values $f(e_i)$ $(1 \leq i \leq n)$, there are at most 2^n linear functions. It follows that the only possible linear functions are $\chi_y(x)$. ■

It will be useful to reason about sets S of indices instead of vectors y. We therefore introduce another notation. Let $S \subseteq \{1, \ldots, n\}$ be a set of indices in y that are 1. Then, let $\chi_S(x) \stackrel{\text{def}}{=} \sum_{i \in S} x_i \pmod 2$.

2.1 Notational shift

From now on we 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 $x, y \in \{0, 1\}^n)$. This notational shift will turn out to be more convenient for us.

Definition 3 *A function f is a* boolean function *if*

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

Definition 4 *A function $f : \{\pm 1\}^n \to \{\pm 1\}$ is* linear *if*

$$f(x) \cdot f(y) = f(x \cdot y)$$

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

Let $S \subseteq \{1, \ldots, n\}$ be a set of indices. Then, let $\chi_S(x) \stackrel{\text{def}}{=} \prod_{i \in S} x_i$.

3 Testing for linearity

Given a boolean function f, we wish to determine, in time sub-linear in n, whether f is linear or whether f is "far" from linear. We now precisely define what "far" from linear means.

Definition 5 *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*.*

Testing linearity by learning. How many queries should we issue to f in order to check its linearity? Here's a method based on learning theory. We make an initial guess $S \subseteq [n]$ by the following algorithm: First, pick a random $x \in \{\pm 1\}^n$. Then, for every bit $i, 1 \leq i \leq n$, we compare $f(x)$ and $f(x^{\oplus i})$ (where $x^{\oplus i}$ means x with the i'th bit inverted): if the resulting values are different, we add i to S, otherwise we keep S unchanged. After n bit-switching queries, our guess of S is complete, and we can start examining whether $f(x) = \chi_S(x)$.

Such an algorithm makes $O(n)$ queries, a quantity sublinear in the size of the input function f; we would like to improve upon this algorithm by presenting a method whose runtime is independent of n. Our method's query complexity will depend only on ϵ.

3.1 Proposed tester

- Repeat $r = O(\frac{1}{\epsilon} \log \frac{1}{\beta})$ times:
  - Pick $x, y \in_R \{\pm 1\}^n$ independently and uniformly.
  - If $f(x) \cdot f(y) \neq f(x \cdot y)$:
    - Output 'test fails' and halt.
- Output 'test passes'.

Claim 6 *If f is linear,* $\Pr[$*tester outputs 'test passes'*$] = 1$. ∎

Claim 7 *If f is ϵ-close to linear,* $\Pr[$*tester outputs 'test fails'*$] \leq 3\epsilon$.

Proof Let f be ϵ-close to linear, and let g be the function as defined in 5. Let A_x denote the event $f(x) \neq g(x)$. Then $\Pr_x[A_x] \leq \epsilon$ and thus $\Pr[\text{tester outputs 'test fails'}] \leq \Pr_{x,y}[A_x \vee A_y \vee A_{x+y}] \leq 3\epsilon$ by union bound. ∎

Indeed, if f is ϵ-close to linear, most tests will pass, because x and y are chosen uniformly at random. We would like to show the opposite direction: that an f which passes most tests is ϵ-close to linear.

Claim 8 *If f is ϵ-far from linear,* $\Pr[f(x) \cdot f(y) \neq f(x \cdot y)] \geq \epsilon$.

Our main focus from here forward will be to prove Claim 8, but first we note that if f is a function such that

$$\Pr[f(x) \cdot f(y) \neq f(x \cdot y)] \geq \epsilon,$$

then

$$\Pr[\text{tester outputs 'test fails'}] \geq 1 - \beta;$$

so we can fail the test with an arbitrarily high probability by choice of r.

4 Basics of Fourier analysis of boolean functions

$\mathcal{G} = \{g : \{\pm 1\}^n \to \mathbb{R}\}$ is a 2^n-dimensional vector space over the field $\mathbb{R}$; all functions in $\mathcal{G}$ are linear combinations of 2^n basis functions with real coefficients. This space is equipped with the inner product

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

We describe two bases of $\mathcal{G}$.

A natural basis. For $a \in \{\pm 1\}^n$, let $e_a(x) \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } x = a \\ 0 & \text{otherwise} \end{cases}$; the set $E \equiv \{e_a : a \in \{\pm 1\}^n\}$ is a basis of $\mathcal{G}$.

Proof There are 2^n methods in E, and every method $g \in \mathcal{G}$ can be written as $g = \sum_{a \in \{\pm 1\}^n} g(a) \cdot e_a$. ■

Note that E is an orthogonal basis but not orthonormal; this is because our definition of the inner product involves a factor of $\frac{1}{2^n}$.

Fourier basis. Recall that $\chi_S(x) \stackrel{\text{def}}{=} \prod_{i \in S} x_i$. Let $\Gamma \stackrel{\text{def}}{=} \{\chi_S : S \subseteq [n]\}$.

Lemma 9 *Γ is an orthonormal basis.*

Proof Let $S \neq T$ be two distinct subsets of $[n]$; then

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

Let $S \triangle T \stackrel{\text{def}}{=} (S \cup T) \setminus (S \cap T)$. Pick $j \in S \triangle T$, and denote "x with the jth bit inverted" by '$x^{\oplus j}$'. Then

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

■

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 10 *Knowing that $\Gamma = \{\chi_S : S \subseteq [n]\}$ is an orthonormal basis for $\mathcal{G}$, we can write every function $f \in \mathcal{G}$ as $f(x) = \sum_{S \subset [n]} \hat{f}(S)\chi_S(x)$, where $\hat{f}(S) = \langle f, \chi_S \rangle$.* ■

Definition 11 *Given $f \in \mathcal{G}$, the* Fourier coefficient *$\hat{f}$ is given by $\hat{f} \stackrel{\text{def}}{=} \langle f, \chi_S \rangle$ for every $S \subseteq [n]$.*

4.1 Useful lemmas about the Fourier Transform

Let f be a linear function. By claim 2, $f \equiv \chi_T$ for some T. The Fourier coefficients of f are $\hat{f}(Z) = \langle \chi_T, \chi_Z \rangle = \begin{cases} 1 & \text{if } T = Z \\ 0 & \text{otherwise} \end{cases}$ (by orthonornality). We see that linear functions exhibit a single large coefficient 1, and all the rest of the coefficients are 0.

Lemma 12

$$\hat{f}(S) = 1 - 2\Pr[f(x) \neq \chi_S(x)]$$

Intuitively, this means that the Fourier coefficients of f give an indication of how close f is to a linear function.
Proof

$$
\begin{aligned}
\hat{f} &= \langle f, \chi_S \rangle \\
&= \frac{1}{2^n} \sum_x f(x)\chi_S(x) \\
&= \frac{1}{2^n} \sum_{x: f(x)=\chi_S(x)} \underbrace{f(x)\chi_S(x)}_{=+1} + \frac{1}{2^n} \sum_{x: f(x)\neq\chi_S(x)} \underbrace{f(x)\chi_S(x)}_{=-1} \\
&= \frac{1}{2^n}(1 - \Pr[f(x) \neq \chi_S(x)]) \cdot 1 + \frac{1}{2^n}(\Pr[f(x) \neq \chi_S(x)]) \cdot (-1) \\
&= 1 - 2\Pr[f(x) \neq \chi_S(x)]
\end{aligned}
$$

■

Lemma 13

$$S \neq T \Rightarrow \Pr[\chi_S(x) = \chi_T(x)] = 1/2$$

Proof Assume $f = \chi_T$, and let $S \neq T$. By Lemma 12,

$$\hat{f}(S) = 1 - 2\Pr[f(x) \neq \chi_S(x)]$$

and from orthonormality, we have

$$\hat{f}(S) = 0$$

By equating and rearranging the two equations, we get

$$\Pr[f(x) \neq \chi_S(x)] = \Pr[\chi_T(x) \neq \chi_S(x)] = 1/2$$

which proves the lemma. ■

A very important theorem in Fourier Analysis is the following:

Theorem 14 (Plancherel's theorem) *Let $f, g : \{\pm 1\} \to \mathbb{R}$. Then*

$$\langle f, g \rangle = \mathrm{Exp}_{x\in\{\pm1\}^n}[f(x)g(x)] = \sum_{S\subseteq[n]} \hat{f}(S)\hat{g}(S).$$

Proof The first (left) equality is by definition of Exp and $\langle , \rangle$. To prove the rest of the theorem, we employ the Fourier representation of f:

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

■

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

Corollary 15 (Parseval's Theorem) *If* $f : \{\pm 1\}^n \to \mathbb{R}$ *then* $\langle f, f \rangle = \mathrm{Exp}[f(x)^2] = \sum_S \hat{f}(S)^2$. ■

Specifically, for boolean $f : \{\pm 1\}^n \to \{\pm 1\}$ we have:

Corollary 16 (Boolean Parseval's Theorem) *If* $f : \{\pm 1\}^n \to \{\pm 1\}$ *then* $\langle f, f \rangle = \mathrm{Exp}[f(x)^2] = \sum_S \hat{f}(S)^2$. ■

We conclude this section with two lemmas.

We define $\chi_\emptyset = 1$. This is essentially the multiplication of zero elements.

Lemma 17 $\mathrm{Exp}[f] = \mathrm{Exp}[f(x) \cdot 1] = \hat{f}(\emptyset)\chi_\emptyset(\emptyset) = \hat{f}(\emptyset)$. ■

Lemma 18 $\mathrm{Exp}[\chi_S(x)] = \begin{cases} 1 & \textit{if } S = \emptyset \\ 0 & \textit{otherwise} \end{cases}$. ■

5 Applying Fourier analysis for linearity testing

Consider a single tester step, which samples a random x, y and tests whether $f(x)f(y) = f(xy)$. Since the range of f is $\{\pm 1\}$, we turn to look at the quantity $f(x)f(y)f(xy)$, which is 1 if the test accepts, and -1 if the test rejects. We can convert this quantity into an indicator variable:

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

Definition 19 *Let the* rejection probability *be*

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

Let the acceptable probability *be*

$$1 - \delta \stackrel{\text{def}}{=} \mathrm{Exp}_{x,y}[\frac{1 + f(x)f(y)f(xy)}{2}]$$

We now turn to stating and proving the main lemma which will assist us in proving Claim 8.

Lemma 20 (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

$$\begin{aligned}
&\mathrm{Exp}_{xy}[\chi_S(x)\chi_T(y)\chi_U(xy)] \\
&= \mathrm{Exp}[\prod_{i\in S} x_i \prod_{j\in T} y_j \prod_{k\in U} x_k y_k] \\
&= \mathrm{Exp}[\prod_{i\in S\Delta U} x_i \prod_{j\in T\Delta U} y_j] \\
&= \mathrm{Exp}[\prod_{i\in S\Delta U} x_i]\mathrm{Exp}[\prod_{j\in T\Delta U} y_j] \\
&= \mathrm{Exp}[\chi_S(x)\chi_U(x)]\cdot\mathrm{Exp}[\chi_T(y)\chi_U(y)] = \begin{cases} 1 & \text{if } S=U \text{ and } T=U \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

■

Proof (of Claim 8) Assume f is ϵ-far from linear, but $\Pr[f(x)\cdot f(y) \neq f(x\cdot y)] < \epsilon$. Rearranging and substituting the expression for test acceptable, we get

$$1-\epsilon < \Pr[f(x)f(y)f(xy)=1].$$

By Main Lemma, we have

$$1-\epsilon < \frac{1}{2}+\frac{1}{2}\sum_{S\in[n]}\hat{f}(S)^3$$

Rearranging, we have

$$1-2\epsilon < \sum_S \hat{f}(S)^3 = \sum_S \hat{f}(S)^2\hat{f}(S).$$

Let T be such that $\hat{f}(T)$ maximizes $\hat{f}(S)$ over all $S \in [n]$.

$$1 - 2\epsilon < \hat{f}(T) \sum_S \hat{f}(S)^2 = \hat{f}(T) \text{ (by Corollary 16).}$$

Using Lemma 12, we have:

$$1 - 2\epsilon < 1 - 2\Pr[f(x) \neq \chi_T(x)] \Rightarrow \epsilon > \Pr[f(x) \neq \chi_T(x)]$$

Therefore f cannot be ϵ-far from linear; a contradiction. ■