0368.416701 Sublinear Time Algorithms January 12, 2015

Lecture 11

Lecturer: Ronitt Rubinfeld *Scribe: Amit Levi*

Lecture Overview

In this lecture we will cover one of the most basic algorithms for testing boolean functions - Testing *Linearity*. In order to establish the proof we will introduce some basic tools from Fourier analysis.

1 Definitions and Introduction

Definition 1 (Linearity) *Assume that we have a function $f : G \to G$ where G is a finite group. f is linear (or equivalently, homomorphism), if $\forall x, y \in G$ it holds that $f(x) + f(y) = f(x + y)$.*

For example the following functions are linear:

1. $f(x) = x$
2. $f(x) = ax \bmod p$ where $G = \mathbb{Z}_p$ and $a \in G$
3. $f(\bar{x}) = \sum_{i\in[n]} a_i x_i \bmod 2$ where $\bar{x} \in \{0, 1\}^n$

Definition 2 *We say that f is ϵ-close to linear over G if there exist a linear function g such that f and g agree on at least $1 - \epsilon$ fraction of the inputs. Equivalently,*

$$\Pr_{x\in G}[f(x) = g(x)] \geq 1 - \epsilon$$

Fact 1 $\forall a, y \in G$

$$\Pr_{x\in G}[y = a + x] = \frac{1}{|G|}$$

This fact is true since over a finite group G only $x = y - a$ satisfy the above equality. Therefore, if we pick an element x uniformly at random from the group then $a + x$ is distributed uniformly in G. Furthermore, this fact also applies for $G = \mathbb{Z}_2^n$ where $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{x} = (x_1, \ldots, x_n)$.

1.1 Self-correcting

Given f such that is $1/8$-close to linear, i.e. there exist a linear function g such that $\Pr[f(x) = g(x)] \geq 7/8$ there exist a randomized algorithm that can compute $g(x)$ using oracle calls to f. The algorithm is as follows:

1. **for** $i = 1, \ldots, c \log(1/\beta)$
   (a) Pick y uniformly at random from G
   (b) $\text{Answer}_i \leftarrow f(y) + f(x - y)$
2. **Output** the most common answer

Note that from Fact 1, $f(x - y)$ is uniformly distributed in G. Since $\Pr[f(x) \neq g(x)] \leq 1/8$ and $\Pr[f(x-y) \neq g(x-y)] \leq 1/8$ if $f(y) = g(y)$ and $f(x-y) = g(x-y)$ then the answer Answer_i is exactly equal to $g(x)$ with probability grater then $3/4$. Thus, by using Chernoff bounds the Self-Corrector outputs the corrected function with high probability.

1.2 Linearity tester

Consider the following tester:

1. **Do** $O\left(\frac{1}{\epsilon}\log\left(\frac{1}{\beta}\right)\right)$ times:
   (a) Pick x, y uniformly at random from G
   (b) If $f(x) + f(y) \neq f(x+y)$
      - **Reject**
2. **Accept**

Observe that for general group the tester might fail. Take for example the following function over $\mathbb{Z}_p$ due to Coppersmith.

$$f(x) = \begin{cases} 1 & \text{if} \quad x \equiv 1 \bmod 3 \\ 0 & \text{if} \quad x \equiv 0 \bmod 3 \\ -1 & \text{if} \quad x \equiv 2 \bmod 3 \end{cases}$$

If, for example $x = y \equiv 1 \bmod 3$ then, $f(x) = f(y) = 1$, $f(x) + f(y) = 2$ but $f(x+y) = -1$, which is a contradiction. We note that same thing happens for $x = y \equiv 2 \bmod 3$, while all other cases pass. It is easy to see the closest linear function to $f(x)$ is $g(x) = 0$ for all x. Therefore, f is $2/3$-far from g but the tester passes $7/9$ fraction of x, y choices. It turns out that it can be showed that if we pass more than $7/9$ fraction of the choices of x, y, then the function is close to linear.

2 Introduction to Fourier Analysis

In the following we will establish basic tools that will enable us to prove the correctness of the tester. Consider the function $f : \{0,1\}^n \to \{0,1\}$ and the binary operation $x \oplus y \stackrel{\text{def}}{=} \sum_{i\in[n]} x_i + y_i \mod 2$. The class of linear functions is defined as follows: $L_a(x) = ax$ for $a \in \{0,1\}^n$, or equivalently, we can define the set $A \subseteq [n]$ which contains all the indices in a that are set to 1, and get that

$$L_A(x) = \bigoplus_{i\in A} x_i$$

For technical reasons we will make the following notational switch.

2.1 The Great Notational Switch

Instead of working over $\mathbb{F}_2^n$ with the operation of addition we will work over $\mathbb{Z}_2^n = \{\pm 1\}^n$ with the operation of multiplication. Thus, our "new" objects of interest are of the form

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

Where 1 corresponds to **FALSE** and -1 corresponds to **TRUE**. Therefore, using the new notations a function f is linear if for every $a, b \in \{\pm 1\}^n$ it holds that $f(a \cdot b) = f(a) \cdot f(b)$. Also, for this case linear functions will be of the form

$$\chi_S(x) \stackrel{\text{def}}{=} \prod_{i\in S} x_i$$

Where $S \subseteq [n]$. Our convention is that if $S = \emptyset$ then $\chi_\emptyset(x) = 1$. Using our new notation we can rephrase our linearity tester as follows.

1. **Do** $O\left(\frac{1}{\epsilon}\log\left(\frac{1}{\beta}\right)\right)$ times:
   (a) Pick x, y uniformly at random from $\{\pm 1\}^n$
   (b) If $f(x) \cdot f(y) \neq f(x \cdot y)$
      - **Reject**
2. **Accept**

We note that $f(x) \cdot f(y) \neq f(x \cdot y)$ if and only if $f(x) \cdot f(y) \cdot f(x \cdot y) = -1$. Hence, we can define the following indicator function.

$$I^f_{\text{FAIL}}(x, y) \stackrel{\text{def}}{=} \frac{1 - f(x) \cdot f(y) \cdot f(x \cdot y)}{2} = \begin{cases} 0 & \text{if} \quad \text{Tester Pass} \\ 1 & \text{if} \quad \text{Tester Fail} \end{cases}$$

And note that,

$$\Pr_{x,y}[\text{Tester Rejects } f] = \mathbb{E}_{x,y}[I^f_{\text{FAIL}}(x, y)] = \frac{1}{2} - \frac{1}{2} \cdot \mathbb{E}_{x,y}[f(x) \cdot f(y) \cdot f(x \cdot y)]$$

Therefore, in order to analyze the tester rejection rate, it is suffices to study the term

$$\mathbb{E}_{x,y}[f(x) \cdot f(y) \cdot f(x \cdot y)]$$

2.2 The Fourier Basis

Consider the following class of functions

$$\mathcal{G} = \{g \mid g : \{\pm 1\}^n \to \{\pm 1\}\}$$

It is easy to see that $\dim(\mathcal{G}) = 2^n$ and thus, all functions of $\mathcal{G}$ are expressible as a linear combination of 2^n basis functions.
One possibility for a basis is the indicator functions:

$$e_a(x) = \begin{cases} 1 & \text{if} \quad x = a \\ 0 & \text{otherwise} \end{cases}$$

Where $a \in \{\pm 1\}^n$. Under this basis we have that each function g can be expressed as

$$g(x) = \sum_a g(a) e_a(x)$$

Where $g(a)$ is a scaler.
For our purpose we will use the following basis.

$$\chi_S(x) = \prod_{i \in S} x_i$$

In addition, we define the inner product

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

Lemma 2 *$\{\chi_S\}_S$ is orthonormal basis with respect to the inner product $\langle \cdot, \cdot \rangle$.*

Proof We first show that the basis is normal.

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

For two different subsets of the indices S and T such that $S \neq T$

$$\begin{aligned}\langle \chi_S, \chi_T \rangle =& \frac{1}{2^n} \sum_x \chi_S(x)\chi_T(x) = \frac{1}{2^n} \sum_x \prod_{i \in S} x_i \prod_{j \in T} x_j = \frac{1}{2^n} \sum_x \prod_{i \in S \setminus T} x_i \prod_{j \in T \setminus S} x_j \prod_{k \in S \cap T} x_k^2 \\ =& \frac{1}{2^n} \sum_x \prod_{i \in S \Delta T} x_i \qquad \star\end{aligned}$$

Pick $j \in S\Delta T$, and define $x^{\oplus j} \stackrel{\text{def}}{=} (x_1, \ldots, x_{j-1}, (-1) \cdot x_j, x_{j+1}, \ldots, x_n)$

$$\star = \frac{1}{2^n} \sum_{x, x^{\oplus j} \text{Pairs}} \left(\prod_{i \in S\Delta T} x_i + \prod_{i \in S\Delta T} x_i^{\oplus j} \right) = \frac{1}{2^n} \sum_{x, x^{\oplus j} \text{Pairs}} \prod_{i \in S\Delta T \setminus \{j\}} x_i \left(x_j + x_j^{\oplus j} \right) = 0$$

Which conclude the proof. ■

Definition 3 *We define the* Fourier Coefficients *of a boolean function f as follows.*

$$\hat{f}(S) = \langle f, \chi_S \rangle = \frac{1}{2^n} \sum_x f(x)\chi_S(x)$$

Theorem 3 $\forall f : \{\pm 1\}^n \to \mathbb{R}$ *there exist a unique representation of f as a multi-linear polynomial,*

$$f(x) = \sum_S \hat{f}(S)\chi_S(x)$$

In what follows we assume that $f : \{\pm 1\}^n \to \{\pm 1\}$.

Fact 4 *f is linear, i.e. $f(x) = \chi_S(x)$ for some S, if and only if there exists $S \subseteq [n]$ such that $\hat{f}(S) = 1$ and for all $T \neq S$ it holds that $\hat{f}(T) = 0$.*

Lemma 5 $\forall S \in [n]$ *it holds that* $\hat{f}(S) = 1 - 2 \cdot \mathrm{dist}(f, \chi_S) = 1 - 2 \cdot \Pr_x[f(x) \neq \chi_S(x)]$.

Proof

$$\begin{aligned}2^n \hat{f}(S) =& 2^n \langle f, \chi_S \rangle = \sum_x f(x)\chi_S(x) = \sum_{x: f(x) = \chi_S(x)} f(x)\chi_S(x) + \sum_{x: f(x) \neq \chi_S(x)} f(x)\chi_S(x) \\ =& (1 - \mathrm{dist}(f, \chi_S)) \cdot 2^n + \mathrm{dist}(f, \chi_S) \cdot (-1) \cdot 2^n = (1 - 2 \cdot \mathrm{dist}(f, \chi_S)) \cdot 2^n\end{aligned}$$

And we are done. ■

Observation 6 $\forall S \neq T$ *it holds that* $\mathrm{dist}(\chi_S, \chi_T) = 1/2$.

Proof

$$0 = \langle \chi_S, \chi_T \rangle = 1 - 2\mathrm{dist}(\chi_S, \chi_T) \implies \mathrm{dist}(\chi_S, \chi_T) = 1/2$$

Which conclude the proof. ■

Theorem 7 (Plancherel's Theorem)

$$\langle f, g\rangle = \sum_S \hat{f}(S)\hat{g}(S)$$

Proof

$$\langle f, g\rangle = \langle \sum_S \hat{f}(S)\chi_S, \sum_T \hat{g}(T)\chi_T \rangle = \sum_S \sum_T \hat{f}(S)\hat{g}(T)\langle \chi_S, \chi_T\rangle = \sum_S \hat{f}(S)\hat{g}(S)$$

And we are done. ■

Corollary 8 (Parseval's Theorem)

$$\langle f, f\rangle = \sum_S \hat{f}^2(S)$$

Note that for a boolean function

$$\langle f, f\rangle = \frac{1}{2^n}\sum_x f^2(x) = 1 \implies \sum_S \hat{f}^2(S) = 1$$

3 Putting It All Together

Let δ_f denote the rejection probability of f. Namely,

$$\delta_f = \frac{1}{2} - \frac{1}{2}\cdot \mathbb{E}_{x,y}[f(x)\cdot f(y)\cdot f(x\cdot y)]$$

We will show that δ_f is quite big.

Theorem 9 *f is δ_f-close to some linear function.*

Proof

$$\begin{aligned}\mathbb{E}_{x,y}[f(x)\cdot f(y)\cdot f(x\cdot y)] =& \mathbb{E}_{x,y}\left[\sum_S \hat{f}(S)\chi_S(x)\sum_T \hat{f}(T)\chi_T(y)\sum_U \hat{f}(U)\chi_U(xy)\right]\\ =& \sum_S\sum_T\sum_U \hat{f}(S)\hat{f}(T)\hat{f}(U)\mathbb{E}_{x,y}\left[\chi_S(x)\chi_T(y)\chi_U(xy)\right]\end{aligned}$$

If $S = T = U$ then $\chi_S(x)\chi_T(y)\chi_U(xy) = \prod_{i\in S} x_i y_i(x_i y_i) = 1$. Otherwise, if $S \neq U$ or $T \neq U$,

$$\mathbb{E}_{x,y}\left[\chi_S(x)\chi_T(y)\chi_U(xy)\right] = \mathbb{E}_{x,y}\left[\prod_{i\in S} x_i \prod_{j\in T} y_j \prod_{k\in U} x_k y_k\right] =$$

$$\mathbb{E}_{x,y}\left[\prod_{i\in S\setminus U} x_i \prod_{i\in U\setminus S} x_i \prod_{i\in U\cap S} x_i^2 \prod_{j\in T\setminus U} y_j \prod_{j\in U\setminus T} y_j \prod_{j\in T\cap U} y_j^2\right] =$$

$$\mathbb{E}_x\left[\prod_{i\in S\Delta T} x_i\right]\mathbb{E}_y\left[\prod_{j\in T\Delta U} y_j\right] = 0$$

Therefore we get that

$$\mathbb{E}_{x,y}[f(x)\cdot f(y)\cdot f(x\cdot y)] = \sum_S \hat{f}^3(S) \le \max_S \hat{f}(S)\sum_S \hat{f}^2(S) = \max_S \hat{f}(S) = 1 - 2\cdot \mathrm{dist}(f, \chi_{S^*})$$

Hence, $\delta_f \ge \min_S\ \mathrm{dist}(f, \chi_S)$ and we are done. ■