0368.416701 Sublinear Time Algorithms December 28, 2015

Lecture 11

Lecturer: Ronitt Rubinfeld *Scribes: Regev Schweiger, Shahar Davidson, Shahaf Gonen, Roy Sheffer*

1 Lesson Overview

1. Linearity (homomorphism) testing of functions over a finite group.
2. Linearity tester correctness proof using tools from Fourier Analysis over the boolean cube.

2 Linearity Testing

Definition 1 (Linearity) *Let $(G, +)$ be a finite group. A function $f : G \to G$ is said to be linear if $\forall x, y \in G$ the following condition is satisfied: $f(x) + f(y) = f(x + y)$.*

Testing for linearity is useful in the case of program verification (for example, checking that a matrix multiplication algorithm is correct).

Examples of linear functions:

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

Definition 2 *F is ϵ-close to linear, or "ϵ-linear", if there exists a linear function g so that either of the following equivalent definitions hold:*

1. *f and g agree on at least $(1 - \epsilon)|G|$ input values;*
2. *$\Pr\limits_{x \in G} [f(x) = g(x)] \geq 1 - \epsilon$, where the elements $x \in G$ are drawn with a uniform distribution over G;*
3. *$\frac{|\{x \in G | \, f(x) = g(x)\}|}{|G|} \geq 1 - \epsilon$.*

A trivial way of checking if f is linear (or ϵ-close to linear) is by learning all the values of f. This can be very time consuming, leading us to search for a method that uses a sublinear amount of queries.

Observation 3 $\forall a, y \in G$

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

This derives from the fact that G is a finite group and therefore the only $x \in G$ that satisfies the equality is $x = y - a$, while $a + x$ is distributed uniformly in G (we will denote this as $a + x \in_R G$). This observation is also correct in the case of $G = \mathbb{Z}_2^n$ and coordinate-wise addition.

2.1 Self correcting (random self-reducibility)

Given a function f which is 1/8-linear (namely, a linear g exists so that $\Pr_{x\in G}[f(x)=g(x)] \geq 7/8$) and any $x \in G$, we wish to compute $g(x)$ using query access to f. In order to successfully compute $g(x)$ with high probability (β), we shall use the following algorithm:

Algorithm 1

for $i = 1, \ldots, C \cdot \log(\frac{1}{\beta})$ **do**
  Pick $y \in_R G$
  $Answer_i \leftarrow f(y) + f(x-y)$
end for
return the most common value for $Answer_i$

Claim 4 $\Pr[output = g(x)] \geq 1 - \beta$

Proof f is 1/8-linear and both $y \in_R G$ and $x - y \in_R G$. Therefore a linear g exists so that the following applies:

$$\Pr_{y\in G}[f(y) \neq g(y)] \leq \frac{1}{8}$$

$$\Pr_{y\in G}[f(x-y) \neq g(x-y)] \leq \frac{1}{8}$$

Then,

$$\begin{aligned}
&\Pr_{y\in G}[f(y)+f(x-y) \neq g(x)] \\
&= \Pr_{y\in G}[f(y)+f(x-y) \neq g(y)+g(x-y)] \\
&\leq \Pr_{y\in G}[f(y) \neq g(y)] + \Pr_{y\in G}[f(x-y) \neq g(x-y)] \\
&= \frac{1}{8} + \frac{1}{8} = \frac{1}{4}
\end{aligned}$$

where the inequality was obtained from the union bound. We can then use the Chernoff inequality to bound the probability of returning an incorrect value. ■

2.2 Linearity tester

We propose the following natural algorithm for testing linearing

Algorithm 2 Linearity tester

for $i = 1, \ldots, O(?)$ **do**
  Pick $x, y \in_R G$
  if $f(x) + f(y) \neq f(x+y)$ **then**
    return FAIL
  end if
end for
return PASS

We need to determine the number of examined pairs, and prove that it is indeed a tester with the correct rejection probability.

Definition 5 (Rejection probability of f)

$$\delta_f \equiv \Pr_{x,y \in G}[f(x) + f(y) \neq f(x+y)]$$

The above algorithm does not work in all cases. Consider the following function $f : \mathbb{Z}_p \to \mathbb{Z}_p$, when $p > 3$.

$$\forall x \in \mathbb{Z}_p \qquad f(x) = \begin{cases} 1 & x = 1 \bmod 3 \\ 0 & x = 0 \bmod 3 \\ -1 & x = 2 \bmod 3 \end{cases}$$

The above function linear in all cases but the following two:

1. if $x \equiv y \equiv 1 \bmod 3$ then
   - $f(x) + f(y) = 1 + 1 = 2$
   - $f(x+y) = f(2 \bmod 3) = -1$
2. if $x \equiv y \equiv 2 \bmod 3$ then
   - $f(x) + f(y) = -1 + -1 = -2$
   - $f(x+y) = f(1 \bmod 3) = 1$

Therefore, in this case $\delta_f = 2/9$, meaning that the tester passes $7/9$ of the availiable $x, y \in G$ combinations. On the other hand, it can be shown that the closest linear function to f is $g(x) = 0$, making f $2/3$-far from linear. It turns out that $\delta_f = 2/9$ is a threshold, and that if you know that $\delta_f < 2/9$, the function must be δ_f-close to linear.

We will prove the correctness of the tester for boolean functions, but first we will need some tools from Fourier analysis.

3 Fourier analysis over the boolean cube

Let f be a function $f\colon \{0,1\}^n \to \{0,1\}$, with the inner product: $x \cdot y \equiv x \oplus y = \sum_{i=1}^{n} x_i y_i \bmod 2$. There are 2^n unique linear functions on $\{0,1\}^n$ that can be defined in one of the following equivalent ways:

1. $L_a(x) = a \cdot x$ for fixed $a \in \{0,1\}^n$
2. $L_A(x) = \sum_{i \in A} x_i \bmod 2$ where $A \subseteq \{1 \cdots n\}$ (set notation)

3.1 Notation change

For the rest of the lecture, we will work with a less natural, but easier to work with boolean set, $\{\pm 1\}^n$ and functions $f : \{\pm 1\}^n \to \{\pm 1\}$. The following proofs are correct with respect to the original boolean notation as well.

In this notation, $0 \to +1$ and $1 \to -1$, or generally: $a \to (-1)^a$. Consequently, after changing notation addition becomes multiplication: $(a + b) \to (-1)^{a+b} = (-1)^a \cdot (-1)^b$. Therefore, linearity can now be defined for a function if $\forall x, y \in \{\pm 1\}^n$ the following condition is satisfied:

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

where $\odot \equiv$ coordinate-wise multiplication. Similarly, linear functions will be of the form:

$$\chi_s(x) = \prod_{i \in s} x_i$$

where $S \subseteq \{1 \cdots n\}$. In light of the notation change, the linear tester described previously is changed so that it halts when it encounters $x, y \in \{\pm 1\}^n$ where $f(x) \cdot f(y) \neq f(x \odot y)$.

Claim 6 $\delta_f = E\left[\frac{1-f(x)f(y)f(x\odot y)}{2}\right]$

Proof

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

$$\Updownarrow$$

$$f(x) \cdot f(y) \cdot f(x \odot y) = \begin{cases} 1 & \text{if test accepts x,y} \\ -1 & \text{if test rejects x,y} \end{cases}$$

$$\Updownarrow$$

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

I is an indicator variable for values $x, y \in \{\pm 1\}^n$ such that $f(x \odot y) \neq f(x) \cdot f(y)$, therefore:

$$\delta_f \equiv \Pr_{x,y \in \{\pm 1\}^n} [f(x) \cdot f(y) \neq f(x \odot y)] = \mathrm{E}_{x,y}[I]$$

■

3.2 Choosing a Fourier basis

Let G be defined as all the n-bit functions mapping to the real vector space:

$$G = \{g | g : \{\pm 1\}^n \to \mathbb{R}\}$$

The dimension of G is 2^n, i.e. all functions in G can be written as a linear combination of 2^n basis functions. We will attempt to find a convenient basis.

3.2.1 First attempt - The basis of indicator functions

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

This is equivalent to viewing $g \in G$ as 2^n vector coordinates $g(a)$ for all $a \in \{\pm 1\}^n$. This basis yields the following representation of g:

$$g(x) = \sum_{a \in G} g(a) e_a(x)$$

3.2.2 Second attempt - basis of parity functions

We will use the following basis functions:

$$\chi_s(x) = \prod_{i \in s} x_i$$

and the inner product:

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

Lemma 7 *The functions $\{\chi_s(x)\}$ are orthonormal with respect to the inner product $\langle \cdot, \cdot \rangle$.*

Proof

1. $\langle \chi_s, \chi_s \rangle = \frac{1}{2^n} \sum_{x \in G} \chi_s(x) \cdot \chi_s(x) = \frac{2^n}{2^n} = 1$

2. When $s \neq t$,

$$\langle \chi_s, \chi_t \rangle = \frac{1}{2^n} \sum_{x \in G} \chi_s(x) \cdot \chi_t(x) = \frac{1}{2^n} \sum_{x \in G} \prod_{i \in s} x_i \prod_{i \in t} x_i = \frac{1}{2^n} \sum_{x \in G} \prod_{i \in s \Delta t} x_i$$

where the last equality stems from the fact that if $i \in s \cap t$ than $x_i^2 = 1$. Let $i \in s \Delta t$ (since $s \neq t$ we know that one exists) and define $x^{\bigoplus j}$ to be x with the j bit flipped. Then, instead of summing over all elements of G separately, we can enumerate over pairs of elements which differ in the j-th coordinate. Then, the above equals

$$= \frac{1}{2^n} \sum_{x, x^{\bigoplus j} \in G} \left(\prod_{i \in s \Delta t} x_i + \prod_{i \in s \Delta t} x_i^{\bigoplus j} \right) = \frac{1}{2^n} \sum_{x, x^{\bigoplus j} \in G} \left(x_j + x_j^{\bigoplus j} \right) \prod_{i \in s \Delta t, i \neq j} (x_i) = 0.$$

■

Observation 8 *The functions $\{\chi_s(x)\}$ form an orthonormal basis as there are 2^n unique orthonormal vectors χ_s.*

This concludes the proof that f can be uniquely expressed as a linear combination of χ_s.

Definition 9 *The Fourier coefficients of f are defined as*

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

for all $s \in G$.

Lemma 10 $\forall f \in G = \{g | g : \{\pm 1\}^n \to \mathbb{R}\}$ *, the unique representation of f as a linear combination of the parity function basis is as follows:*

$$f(x) = \sum_s \hat{f}(s) \chi_s(x)$$

This follows from the fact that the functions $\{\chi_s(x)\}$ form an orthonormal basis.

Fact 11 *f is linear $\Leftrightarrow \exists s \subseteq [n]$ so that $\hat{f}(s) = \langle f, \chi_s \rangle = \frac{1}{2^n} \sum 1 = 1$ and $\forall t \neq s$, $\hat{f}(s) = \langle f, \chi_t \rangle = \langle \chi_s, \chi_t \rangle = 0$*

Lemma 12 *The Fourier coefficients of a function characterize the distance of the function from linearity. Namely, it can be shown that $\forall s \subseteq [n]$,*

$$\hat{f}(s) = 1 - 2 \cdot dist(f, \chi_s) = 1 - 2 \cdot \Pr_{x \in \{\pm 1\}^n} [f(x) \neq \chi_s(x)].$$

Proof

$$\begin{aligned}
2^n \cdot \hat{f}(s) &= \sum_{x \in G} f(x)\chi_s(x) \\
&= \sum_{\{x | f(x) = \chi_s(x)\}} f(x)\chi_s(x) + \sum_{\{x | f(x) \neq \chi_s(x)\}} f(x)\chi_s(x) \\
&= \sum_{\{x | f(x) = \chi_s(x)\}} 1 + \sum_{\{x | f(x) \neq \chi_s(x)\}} (-1) \\
&= 2^n \cdot \Pr_{x \in G}[f(x) = \chi_s(x)] - 2^n \cdot \Pr_{x \in G}[f(x) \neq \chi_s(x)] \\
&= 2^n \cdot \left(1 - 2 \cdot \Pr_{x \in G}[f(x) \neq \chi_s(x)]\right)
\end{aligned}$$

The lemma immediately follows. ■

Observation 13 *Given any χ_s, χ_t such that $s \neq t$, $dist(\chi_s, \chi_t) = \frac{1}{2}$.*

Proof Consider $f = \chi_s$.

$$\hat{f}(t) = \langle \chi_s, \chi_t \rangle = 0$$

and according to the previous lemma

$$\hat{f}(t) = 1 - 2 \cdot \mathrm{dist}(f, \chi_t)$$

which concludes the proof. ■

3.3 Useful tools

Theorem 14 (Plancherel's Theorem)

$$\langle f, g \rangle = \sum_{s \subseteq n} \hat{f}(s)\hat{g}(t)$$

Proof

$$\begin{aligned}
\langle f, g \rangle = \left\langle \sum_s \hat{f}(s)\chi_s, \sum_t \hat{g}(t)\chi_t \right\rangle &= \sum_{s,t} \hat{f}(s)\hat{g}(t) \langle \chi_s, \chi_t \rangle = \\
\text{since } \langle \chi_s, \chi_t \rangle &= \begin{cases} 0 & \text{if } s \neq t \\ 1 & \text{if } s = t \end{cases} \\
&= \sum_{s \subseteq [n]} \hat{f}(s)\hat{g}(t)
\end{aligned}$$

■

Theorem 15 (Parseval's Theorem)

$$\langle f, f \rangle = \sum_s \hat{f}^2(s)$$

Boolean Parseval:

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

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

4 Proof of linearity tester

Previously, we have shown that:

$$\delta_f \equiv \Pr_{x,y\in\{\pm1\}^n}[f(x)\cdot f(y)\neq f(x\odot y)] = \frac{1-E\left[f(x)f(y)f(x\odot y)\right]}{2}$$

Theorem 16 *If δ_f is the rejection probability of the linearity tester above, f is δ_f-close to some linear function.*

Proof

$$\begin{aligned}
&\mathbb{E}_{x,y}\left[f(x)f(y)f(x\odot y)\right]\\
=&\mathbb{E}_{x,y}\left[\sum_s \hat{f}(s)\chi_s(x)\cdot\sum_t \hat{f}(t)\chi_t(y)\cdot\sum_u \hat{f}(u)\chi_u(x\odot y)\right]\\
=&\mathbb{E}_{x,y}\left[\sum_{s,t,u}\hat{f}(s)\hat{f}(t)\hat{f}(u)\chi_s(x)\chi_t(y)\chi_u(x\odot y)\right]\\
=&\sum_{s,t,u}\hat{f}(s)\hat{f}(t)\hat{f}(u)\,\mathbb{E}_{x,y}\left[\chi_s(x)\chi_t(y)\chi_u(x\odot y)\right]=(\star)
\end{aligned}$$

If $s=t=u$:

$$\chi_s(x)\chi_t(y)\chi_u(x\odot y)=\prod_{i\in s}x_i\cdot y_i\cdot(x_i\cdot y_i)=1$$

If $\neg(s=t=u)$ then:

$$\begin{aligned}
&\mathbb{E}_{x,y}\left[\chi_s(x)\chi_t(y)\chi_u(x\odot y)\right]\\
=&\mathbb{E}_{x,y}\left[\prod_{i\in s}x_i\prod_{j\in t}y_j\prod_{k\in u}(x_k\cdot y_k)\right]\\
=&\mathbb{E}_{x,y}\left[\prod_{i\in s\Delta u}x_i\prod_{j\in t\Delta u}y_j\right]
\end{aligned}$$

Since x,y are independent,

$$=\mathbb{E}_x\left[\prod_{i\in s\Delta u}x_i\right]\cdot\mathbb{E}_y\left[\prod_{j\in t\Delta u}y_j\right]=0.$$

The last equality follows from the fact that either $s\neq u$:

$$\mathbb{E}_x\left[\prod_{i\in s\Delta u}x_i\right]=0$$

or $t\neq u$:

$$\mathbb{E}_y\left[\prod_{j\in t\Delta u}y_i\right]=0$$

Therefore,

$$(\star)=\sum_s\left[\hat{f}(s)\right]^3\leq \max_s(\hat{f}(s))\cdot\sum_s(\hat{f}(s))^2)=\max_s(\hat{f}(s))=1-2\cdot\min_s(\mathrm{dist}(f,\chi_s))$$

where the last two equalities come from Parseval's theorem and the previous lemma, respectively.

By plugging our result into the definition of δ_f, we arrive at the following conclusion:

$$\delta_f \geq \frac{1}{2} - \frac{1}{2}\left[1 - 2 \cdot \min_s(\text{dist}(f, \chi_s))\right] = \min_s(\text{dist}(f, \chi_s))$$

■