6.842 Randomness and Computation March 12, 2014

Lecture 11

Lecturer: Ronitt Rubinfeld *Scribe: Aikaterini Sotiraki*

1 Lecture overview

In this lecture, we will introduce the basic notions of Fourier analysis and linearity testing.
Boolean functions $f : \{0,1\}^n \rightarrow \{0,1\}$ are very important in many areas, such as complexity, machine learning, coding theory and combinatorics. The main tool to study them will be Fourier/harmonic analysis because this will allow us to study "structural properties" of boolean functions. For example, if a boolean function:

- has low complexity
- does not depend on all the bits of the input:
  - dictator: The function depends only on one bit of the input.
  - k- junta: The function depends only on a subset of size k of the bits of the input
- is "fair", which means that no variable has to much influence
- is a homomorphism: $f(x \bigoplus y) = f(x) \oplus f(y)$
- spreads out or is concentrated

2 Preliminaries

A boolean function is a function:

$$f : \{0,1\}^n \rightarrow \{0,1\}$$

For every $x = (x_1, ..., x_n) \in \{0,1\}^n$, $y = (y_1, ..., y_n) \in \{0,1\}^n$, we define the operation:

$$x \bigoplus y = (x_1, ..., x_n) \bigoplus (y_1, ..., y_n) = (x_1 \oplus y_1, ..., x_n \oplus y_n)$$

where $\oplus$ is the xor operation.
However, we will use an alternative representation of boolean function, where $f : \{1,-1\}^n \rightarrow \{1,-1\}$ and the operation is defined as follows:

$$x \odot y = (x_1, ..., x_n) \odot (y_1, ..., y_n) = (x_1 \cdot y_1, ..., x_n \cdot y_n)$$

where $\cdot$ is the usual multiplication in integers.
The truth table of xor, using the new notation, is represented as follows:

$\odot$	$+1$	-1
$+1$	$+1$	-1
-1	-1	$+1$

The set of functions $G = \{g | g : \{-1,+1\}^n \rightarrow \mathbb{R}\}$ is a vector space with dimension 2^n, which means that for any set of basis functions of size 2^n, every $g \in G$ can be uniquely expressed as a linear combination of the basis functions.

3 Basis functions

There are more than one sets of basis functions that we can use:

- The "natural" basis: indicator functions

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

  This basis is orthonormal and is used to describe functions via their "truth table":

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

  Example: for $n = 1$

  $$e_{+1}(x) = \begin{cases} 1 & \text{if } x = 1 \\ 0 & \text{if } x = 1 \end{cases}$$

  $$e_{-1}(x) = \begin{cases} 1 & \text{if } x = -1 \\ 0 & \text{if } x = +1 \end{cases}$$

  Now, if $f : \{-1, +1\} \to \mathbb{R}$, with $f(+1) = b_{+1}$ and $f(-1) = b_{-1}$, then

  $$f(x) = b_{+1} \cdot e_{+1}(x) + b_{-1} \cdot e_{-1}(x)$$

- Parity functions:

  $$\text{For } S \subset [n], \chi_S(x) = \prod_{i \in S} x_i$$

  We will define that $\chi_\emptyset(x) = 1, \forall x \in \{-1, +1\}^n$.

4 Useful facts and Theorems

- Fact 0: $\chi_S(x) \cdot \chi_T(x) = \chi_{S \triangle T}(x)$, where $S \triangle T = (S \setminus T) \cup (T \setminus S)$ (symmetric difference)
  Proof

  $$\begin{aligned} \chi_S(x) \cdot \chi_T(x) &= \prod_{i \in S} x_i \cdot \prod_{i \in T} x_i \\ &= \prod_{i \in S \cap T} x_i^2 \cdot \prod_{i \in S \triangle T} x_i = \prod_{i \in S \cap T} 1 \cdot \prod_{i \in S \triangle T} x_i \\ &= \prod_{i \in S \triangle T} x_i \end{aligned}$$

  ■

- The set of parity functions forms an orthonormal basis of the vector space G.
  Proof
  We will define the inner product $< f, g >= \frac{1}{2^n} \sum\limits_{x \in \{-1,+1\}^n} f(x) g(x)$.

- If $S = T$:

$$< \chi_S, \chi_T >=< \chi_S, \chi_S >= \frac{1}{2^n} \sum_{x \in \{-1,+1\}^n} (\chi_S(x))^2 = 1$$

- If $S \neq T$:

$$\begin{aligned} < \chi_S, \chi_T > &= \frac{1}{2^n} \sum_{x \in \{-1,+1\}^n} \chi_S(x) \cdot \chi_T(x) = \\ &= \frac{1}{2^n} \sum_{x \in \{-1,+1\}^n} \chi_{S \triangle T}(x) \\ &= \frac{1}{2^n} \sum_{x \in \{-1,+1\}^n} \prod_{i \in S \triangle T} x_i \\ &= \frac{1}{2^n} \sum_{\text{pairs } x, x^{\oplus j}} (\prod_{i \in S \triangle T} x_i + \prod_{i \in S \triangle T} (x^{\oplus j})_i) \\ &= 0 \end{aligned}$$

where $x^{\oplus j}$ is equal to x with the j-th bit flipped, namely if $x = (x_1, ..., x_n)$, then we define $x^{\oplus j}$ such that $x_i^{\oplus j} = -x_i$ if $i = j$ and $x_i^{\oplus j} = x_i$ otherwise. Also, since from hypothesis we have that $S \neq T \Rightarrow S \triangle T \neq \emptyset$, we pick a j such that $j \in S \triangle T$. The final calculation follows from the fact that

$$\prod_{i \in S \triangle T} x_i + \prod_{i \in S \triangle T} (x^{\oplus j})_i = (x_j \prod_{i \in (S \triangle T) \setminus \{j\}} x_i) - (x_j \prod_{i \in (S \triangle T) \setminus \{j\}} (x^{\oplus j})_i) = 0$$

■

- **Theorem 1** *Every $f \in G$ can be written as $f(x) = \sum_{S \subset [n]} \hat{f}(S) \chi_S(x)$, where $\hat{f}(S)$ are the Fourier Coefficient and*

$$\hat{f}(S) =< f, \chi_S >= \frac{1}{2^n} \sum_{x \in \{-1,+1\}^n} f(x) \chi_S(x)$$

Proof
Since the set of parity functions forms an orthonormal basis, for every $f \in G$, we have that $f(x) = \sum_{S \subset [n]} a_S \chi_S(x)$.
Now, we need to calculate the coefficients a_i.

$$\begin{aligned} < f, \chi_S > &= \frac{1}{2^n} \sum_{x \in \{-1,+1\}^n} f(x) \chi_S(x) \\ &= \frac{1}{2^n} \sum_{x \in \{-1,+1\}^n} (\sum_{T \subset [n]} a_T \chi_T(x)) \cdot \chi_S(x) \\ &= \frac{1}{2^n} \sum_{x \in \{-1,+1\}^n} a_S (\chi_S(x))^2 \\ &= \frac{2^n}{2^n} a_S \\ &= a_S \end{aligned}$$

■

- Fact 1: f is a parity function:
  - if and only if $f(x) = \chi_S(x)$ for an $S \subset [n]$
  - if and only if there exists an $S \subset [n]$ such that:
    1. $\hat{f}(S) = 1$ and
    2. for all $T \neq S$, $\hat{f}(T) =< \chi_S, \chi_T >= 0$

  Proof
  The first part is obvious from the definition of the parity functions.
  For the second part:

  - If $f = \chi_S$, then form the fact that the set of parity functions is an orthonormal basis: $\hat{f}(S) =< \chi_S, \chi_S >= 1$ and for all $T \neq S$, $\hat{f}(T) =< \chi_S, \chi_T >= 0$.
  - If $\hat{f}(S) = 1$ and for all $T \neq S$, $\hat{f}(T) = 0$, then $f = \sum\limits_{T \subset [n]} \hat{f}(T)\chi_T(x) = \chi_S(x)$. So, f is a parity function.

  ■

- Fact 2: $\hat{f}(S) = 1 - 2Pr_{x\in\{-1,+1\}^n}(f(x) \neq \chi_S(x))$
  Proof

$$\begin{aligned}
\hat{f}(S) &= \frac{1}{2^n} \sum_{x\in\{-1,+1\}^n} f(x)\chi_S(x) \\
&= \frac{1}{2^n} \sum_{x \text{ s.t. } f(x)=\chi_S(x)} (+1) + \frac{1}{2^n} \sum_{x \text{ s.t. } f(x)\neq\chi_S(x)} (-1) \\
&= (1 - Pr_{x\in\{-1,+1\}^n}(f(x) \neq \chi_S(x))) - Pr_{x\in\{-1,+1\}^n}(f(x) \neq \chi_S(x)) \\
&= 1 - 2Pr_{x\in\{-1,+1\}^n}(f(x) \neq \chi_S(x))
\end{aligned}$$

  ■

- Fact 3: If $S \neq T$, then $Pr_{x\in\{-1,+1\}^n}(\chi_S(x) = \chi_T(x)) = 1/2$
  Proof
  Let $f = \chi_T$, then using fact 0, we have

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

  and using fact 2 we get that

$$\hat{f}(S) = 1 - 2Pr_{x\in\{-1,+1\}^n}(\chi_T(x) \neq \chi_S(x))$$

  So,

$$\begin{aligned}
&1 - 2Pr_{x\in\{-1,+1\}^n}(\chi_T(x) \neq \chi_S(x)) = 0 \\
&\Rightarrow Pr_{x\in\{-1,+1\}^n}(\chi_T(x) \neq \chi_S(x)) = 1/2 \\
&\Rightarrow Pr_{x\in\{-1,+1\}^n}(\chi_T(x) = \chi_S(x)) = 1/2
\end{aligned}$$

  ■

- Plancherel's Theorem

  Theorem 2 *For $f, g : \{-1,+1\}^n \to \mathbb{R}$, we have*

  $$< f, g > \equiv E_{\{-1,+1\}^n}[f(x)g(x)] = \sum_{S\subset[n]} \hat{f}(S)\cdot\hat{g}(S)$$

  Proof

  $$\begin{aligned} < f, g > &= < \sum_{S\subset[n]} \hat{f}(S)\chi_S(x), \sum_{T\subset[n]} \hat{g}(T)\chi_T(x) > \\ &= \sum_{S\subset[n]}\sum_{T\subset[n]} \hat{f}(S)\hat{g}(S)\cdot < \chi_S, \chi_T > \\ &= \sum_{S\subset[n]} \hat{f}(S)\hat{g}(S) \end{aligned}$$

  where the first equality is immediate from the definition, the second is due to the bilinearity of the inner product and the third due to the orthogonality of the parity functions.
  ■

- Parseval's Theorem

  Corollary 3 *For $f : \{-1,+1\}^n \to \mathbb{R}$, we have*

  $$< f, f > \equiv E_{\{-1,+1\}^n}[f^2(x)] = \sum_{S\subset[n]} \hat{f}(S)^2$$

  Proof
  We use the Plancherel's theorem for $g = f$.
  ■

  A special case of the above corollary is when $f : \{-1,+1\}^n \to \{-1,+1\}$, then

  $$< f, f > \equiv E_{\{-1,+1\}^n}[f^2(x)] = \sum_{S\subset[n]} \hat{f}(S)^2 = 1$$

- Fact 4: $E[f] = \hat{f}(\emptyset)$
  Proof

  $$\begin{aligned} E[f] &= E[f(x)\cdot 1] = E[f(x)\cdot\chi_\emptyset(x)] \\ &= \sum_{S\subset[n]} \hat{f}(S)\cdot\hat{\chi}_\emptyset(S) \\ &= \hat{f}(\emptyset)\cdot\hat{\chi}_\emptyset(\emptyset) = \hat{f}(\emptyset) \end{aligned}$$

  The second equality is true because of Parseval's theorem and the third because

  $$\hat{\chi}_\emptyset(S) = \begin{cases} 1 & \text{if } S = \emptyset \\ 0 & \text{otherwise} \end{cases}$$

  ■

- Fact 5:

  $$E[\chi_S(x)] = \begin{cases} 1 & \text{if } S = \emptyset \\ 0 & \text{otherwise} \end{cases}$$

  Proof Using fact 4, we have that $E[\chi_S(x)] = \hat{\chi}_S(\emptyset)$ and from fact 1

  $$\hat{\chi}_S(\emptyset)] = \begin{cases} 1 & \text{if } S = \emptyset \\ 0 & \text{otherwise} \end{cases}$$

  ■

5 Linearity (homomorphism) testing

Our goal is to quickly test if a function over a group is linear, that is

$$\forall x, y, f(x) + f(y) = f(x + y)$$

In many cases, it is enough to check whether f is *close* to homomorphism. These cases include:

- Checking correctness of programs computing matrix, algebraic, trigonometric functions
- Probabilistically Checkable Proofs

Definition 4 *Let f over domain of size N, then f is ϵ-close to linear if we can change at most ϵN values to turn it into linear. Otherwise, f is ϵ-far from linear.*

Our goal is that the *query complexity*, measured in terms of the domain size N, is *constant* and independent of N. The tester should behave as follows:

- If f is *linear*, then the test should pass with probability more than 2/3.
- If f is ϵ-far from linear, then the test should fail with probability more than 2/3.
- If f is not linear, but is ϵ-*close* to linear, then either output is ok.

Example: Linearity testing for $f : GF(2)^n \to GF(2)$
If $x = (x_1, ..., x_n), y = (y_1, ..., y_n) \in \{0, 1\}^n$, then

$$x + y = (x_1 \oplus y_1, ..., x_n \oplus y_n)$$

So, f is linear if and only if for all $x, y \in \{0, 1\}^n$, $f(x) \oplus f(y) = f(x+y)$. It can be shown that f is linear if and only if

$$f \in \{f_a | f_a(x) = \sum_{i=1}^{n} a_i \cdot x_i \mod 2 \text{ for } (a_1, ..., a_n) \in \{0, 1\}^n\}$$

Example: Linearity testing for $f : \{-1, +1\}^n \to \{-1, +1\}$
If $x = (x_1, ..., x_n), y = (y_1, ..., y_n) \in \{-1, +1\}^n$, then

$$x \odot y = (x_1 \cdot y_1, ..., x_n \cdot y_n)$$

So, f is linear if and only if for all $x, y \in \{-1, +1\}^n, f(x) \cdot f(y) = f(x \odot y)$. It can be shown that f is linear if and only if f is a parity function.

Proposed Tester:
Repeat $r = O(\frac{1}{\rho})$ times:

- Pick $x, y \in_R \{0, 1\}^n$
- If $f(x) \cdot f(y) \neq f(x \odot y)$ output "FAIL" and halt

Output "PASS"

Analysis:

- If f is linear, then the tester passes with probability 1.
- If f is such that $Pr_{x,y}(f(x) \cdot f(y) \neq f(x \odot y)) \geq \rho$, then we can choose the constant in $O(\frac{1}{\rho})$ so that the tester fails with probability at least 2/3.

We want to show that if $Pr_{x,y}(f(x) \cdot f(y) \neq f(x \odot y)) < \epsilon$, then f is ϵ-*close* to linear. However, this is non trivial.
Example:[Coppersmith]
Let $f : \mathbb{Z}_{3^k} \to \mathbb{Z}_{3^{k-1}}$, such that

$$f(3h + d) = h, \text{ for } h < 3^k, d \in \{-1, 0, 1\}$$

Then, f satisfies $f(x) + f(y) \neq f(x + y)$ for only 2/9 of the choices of $x, y \in \mathbb{Z}_{3^k}$ (i.e. $\delta_f = 2/9$). However, f is $2/3$-far from linear!

So, we need to prove that if $Pr_{x,y}(f(x) \cdot f(y) \neq f(x \odot y)) < \epsilon$, then f is ϵ-*close* to linear. First, we need to prove the following lemma.

Lemma 5 $1 - \delta \equiv Pr_{x,y}(f(x) \cdot f(y) \cdot f(x \odot y) = 1) = \frac{1}{2} + \frac{1}{2} \sum_{S \subset [n]} \hat{f}(S)^3$

Proof
Let us define $c(x, y) = \frac{1 + f(x) \cdot f(y) \cdot f(x \odot y)}{2}$, then

$$c(x, y) = \begin{cases} 1 & \text{if } x, y \text{ PASS} \\ 0 & \text{if } x, y \text{ FAIL} \end{cases}$$

So, $c(x, y)$ is an indicator variable describing the result of the test.
We know that:

$$\begin{aligned} 1 - \delta &= E_{x,y}[\frac{1 + f(x) \cdot f(y) \cdot f(x \odot y)}{2}] \\ &= \frac{1}{2} + \frac{1}{2} E_{x,y}[f(x) \cdot f(y) \neq f(x \odot y)] \end{aligned}$$

But,

$$E_{x,y}[f(x) \cdot f(y) \neq f(x \odot y)] = E_{x,y}[(\sum_{S \subset [n]} \hat{f}(S)\chi_S(x)) \cdot (\sum_{T \subset [n]} \hat{f}(T)\chi_T(y)) \cdot (\sum_{U \subset [n]} \hat{f}(U)\chi_U(x \odot y))] =$$

$$= \sum_{S,T,U \subset [n]} \hat{f}(S)\hat{f}(T)\hat{f}(U) E_{x,y}[\chi_S(x) \cdot \chi_T(y) \cdot \chi_U(x \odot y)]$$

Also,

$$\begin{aligned} E_{x,y}[\chi_S(x) \cdot \chi_T(y) \cdot \chi_U(x \odot y)] &= E_{x,y}[\prod_{i \in S} x_i \prod_{t \in T} y_j \prod_{k \in U} (x_k \cdot y_k)] \\ &= E_{x,y}[\prod_{i \in S \triangle U} x_i \prod_{j \in T \triangle U} y_j] \\ &= E_{x,y}[\prod_{i \in S \triangle U} x_i] E_{x,y}[\prod_{j \in T \triangle U} y_j] \end{aligned}$$

But,

$$E_{x,y}[\prod_{i \in S \triangle U} x_i] = \begin{cases} 1 & \text{if } S \triangle U = \emptyset \\ 0 & \text{otherwise} \end{cases}$$

and

$$E_{x,y}[\prod_{j \in T \triangle U} y_j] = \begin{cases} 1 & \text{if } T \triangle U = \emptyset \\ 0 & \text{otherwise} \end{cases}$$

So,

$$E_{x,y}[\chi_S(x) \cdot \chi_T(y) \cdot \chi_U(x \odot y)] = \begin{cases} 1 & \text{if } S = T = U \\ 0 & \text{otherwise} \end{cases}$$

As a result,

$$E_{x,y}[f(x) \cdot f(y) \neq f(x \odot y)] = \sum_{S,T,U \subset [n]} \hat{f}(S)\hat{f}(T)\hat{f}(U) E_{x,y}[\chi_S(x) \cdot \chi_T(y) \cdot \chi_U(x \odot y)] = \sum_{S \subset [n]} \hat{f}(S)^3$$

Consequently,

$$1 - \delta = \frac{1}{2} + \frac{1}{2} \sum_{S \subset [n]} \hat{f}(S)^3$$

■

Now, using the above lemma we can prove the following theorem:

Theorem 6 *If f is ϵ-far from linear, then*

$$\delta \equiv Pr_{x,y}(f(x) \cdot f(y) \neq f(x \odot y)) \geq \epsilon$$

Proof
From the lemma we have that:

$$1 - \delta = \frac{1}{2} + \frac{1}{2} \sum_{S \subset [n]} \hat{f}(S)^3 \Rightarrow 1 - 2\delta = \sum_{S \subset [n]} \hat{f}(S)^3$$

So,

$$\begin{aligned}1-2\delta &= \sum_{S\subset[n]} \hat{f}(S)^3\\ &\le \max_S(\hat{f}(S)) \sum_{S\subset[n]} \hat{f}(S)^2\\ &\le \max_S(\hat{f}(S))\\ &\le \hat{f}(T)\\ &\le 1-2Pr_x(f(x)\neq \chi_T(x))\end{aligned}$$

For the third equation we used Parseval's Theorem, for the forth we picked a T such that $\hat{f}(T) = \max_S(\hat{f}(S))$ and for the final computation the fact 2.
As a result,

$$\delta \ge Pr_x(f(x)\neq \chi_T(x)) \ge \epsilon$$

■

Linearity testing is possible over other domains with constant query complexity, even for general non-abelian groups. However, in this case, there is a slightly weaker relationship between the parameters.

6 Self-correction

Given a program P computing linear f that is correct on at least 7/8 of the inputs, even though we do not know which ones, can we correctly compute f on each input?
Self-Corrector:
Repeat $r = O(\frac{1}{\rho})$ times:

- Pick $y \in_R \{0,1\}^n$
- Let $guess(x) \leftarrow P(y)\cdot f(x\odot y)$

Output most common guess

- If P is correct in both calls, then the guess is correct.
- Since y is uniformly distributed, then $x\odot y$ is also uniformly distributed, so:

$$Pr(P \text{ wrong in either } y \text{ or } x\odot y\ ) \le 1/4$$