Lecture 11: Fourier Basics for Boolean functions. Linearity testing.

Lecturer: Ronitt Rubinfeld Spring 2014 6.842: Randomness and Computation

Why all the fuss about Boolean functions?

- Truth table of a function (complexity theory)
- Concept to be learned (machine learning)
- Subset of the Boolean cube (coding theory, combinatorics,...)
- Etc.

Why Fourier/Harmonic Analysis?

- Study "structural properties" of Boolean functions
 - Low complexity
 - Depends on few inputs (dictator, junta)
 - "fair" (no variable has too much influence)
 - Homomorphism
 - Spread out/concentrated

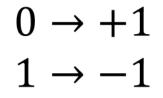
The Boolean function

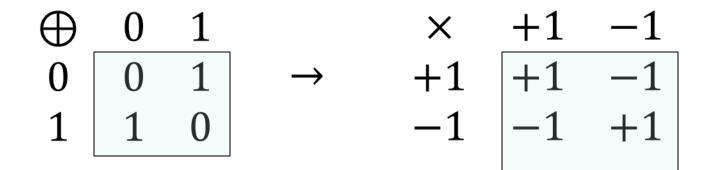
 $\begin{aligned} f: \{0,1\}^n &\to \{0,1\} \\ (x_1, x_2, \dots, x_n) \oplus (y_1, y_2, \dots, y_n) \\ &= (x_1 \oplus y_1, \dots, x_n \oplus y_n) \end{aligned}$

A "new" representation!

 $f: \{\pm 1\}^n \to \{\pm 1\}$ (x₁, x₂, ..., x_n) \bigcirc (y₁, y₂, ..., y_n) = (x₁ · y₁, ..., x_n · y_n)

The slick (notational) trick:





The set of functions and inner product

- $G = \{g | g : \{\pm 1\}^n \rightarrow \mathbb{R}\}$ (all *n*-bit fctns into Reals)
 - A vector space of dimension 2ⁿ
 - For any set of basis functions of size 2^n , every $g \in G$ is a linear combination of basis functions.
 - Which basis to use?

Which basis?

- $G = \{g | g : \{\pm 1\}^n \rightarrow \mathbb{R}\}$ (all *n*-bit fctns into Reals)
 - A "natural" basis: indicator functions

$$\bullet e_a(x) = \begin{cases} 1 & if \ x = a \\ 0 & o.w. \end{cases}$$

- Orthonormal
- Used to describe function via "truth table" $f(x) = \sum_{a} f(a)e_{a}(x)$

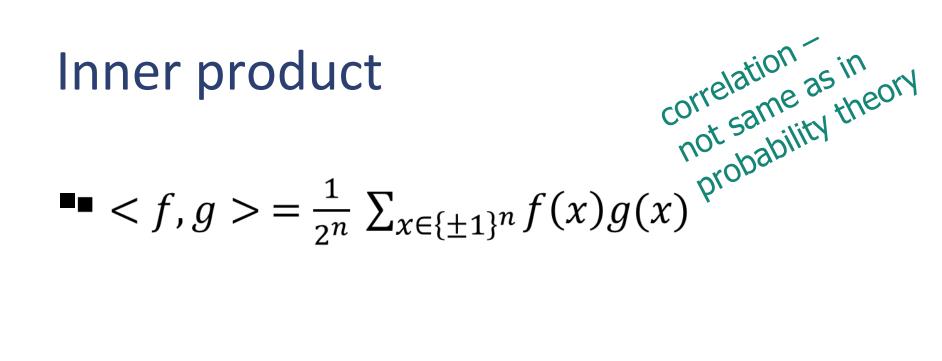
A very useful basis:

- $G = \{g | g : \{\pm 1\}^n \rightarrow \mathbb{R}\}$ (all *n*-bit fctns into Reals)
 - Parity functions
 - For $S \subseteq [n]$, $\chi_S(x) = \prod_{i \in S} x_i$
 - Let's agree that $\chi_{\emptyset}(x) = 1 \ \forall x$

Fact 0:
$$\chi_S(x) \cdot \chi_T(x) = \chi_{S \triangle T}(x)$$

Proof:
$$\chi_S(x) \cdot \chi_T(x) = \prod_{i \in S} x_i \prod_{j \in T} x_j$$

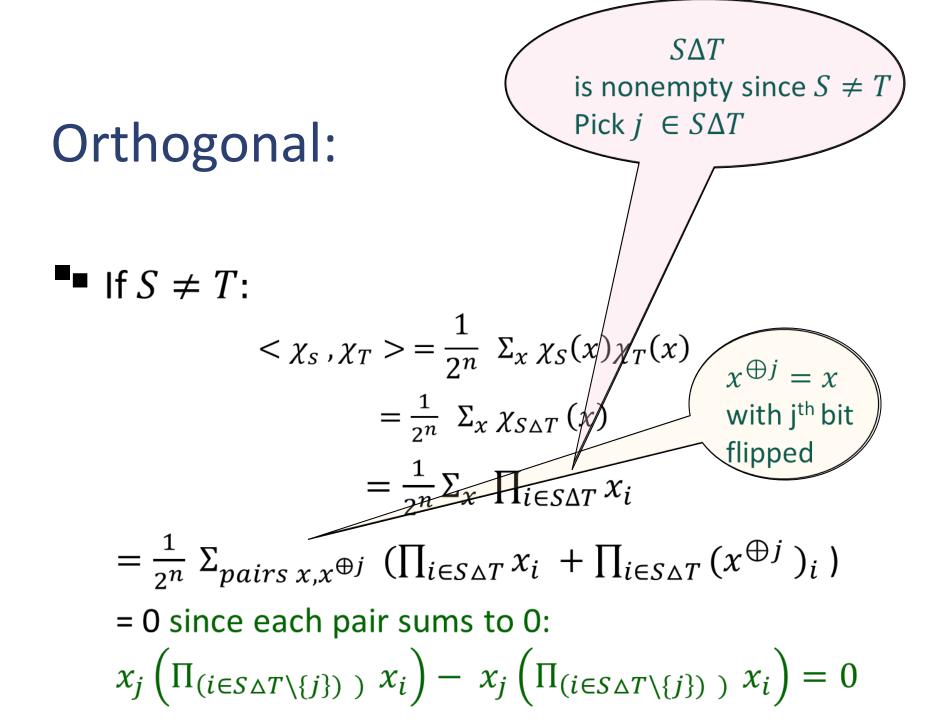
= $\prod_{S \cap T} x_i^2 \prod_{i \in S \Delta T} x_i$
= 1



• Note:

$$<\chi_S, \chi_S > = \frac{1}{2^n} \sum_{x \in \{\pm 1\}^n} (\chi_{S(x)})^2 = 1$$

Always 1



So we have an orthonormal basis!

Every function can be written as a linear combination of these χ_S 's

Theorem: $\forall f, f(x) = \Sigma_S \hat{f}(S)\chi_S(x) \text{ where}$ $\hat{f}(S) = \langle f, \chi_S \rangle = \frac{1}{2^n} \Sigma_{x \in \{\pm 1\}^n} f(x)\chi_S(x)$

Some examples:

Function Fourier Representation

 $f(x)=1 = \chi(\emptyset)$ $f(x)=x_{i} = \chi(\{i\})$ $f(x)=AND(x_{1}, x_{2})$ 1 χ_{i} $\chi_{2} + \chi_{2} + \chi_{1} + \chi_{2} + \chi_{2} - \chi_{1} + \chi_{2} + \chi_{2} - \chi_{2} + \chi_{2}$

Fourier coefficients of parity functions:

Fact 1: f is a parity function iff $f = \chi_S(x)$ iff (1) $\hat{f}(S) = 1$ and (2) for all $T \neq S$, $\hat{f}(T) = \langle \chi_S, \chi_T \rangle = 0$ By orthogonality Agreement with parity function vs. max Fourier coefficient

Fact 2:
$$\hat{f}(S) = 1 - 2 \Pr_{x \in \pm 1^n} [f(x) \neq \chi_S(x)]$$

Draaf

$$\hat{f}(S) = \frac{1}{2^n} \Sigma_x f(x) \chi_S(x)$$

= $\frac{1}{2^n} \Sigma_{x \text{ s.t. } f(x) = \chi_{S(x)}} (+1) + \frac{1}{2^n} \Sigma_{x \text{ s.t. } f(x) \neq \chi_{S(x)}} (-1)$
= $(1 - \Pr_{x \in \pm 1^n} [f(x) \neq \chi_S(x)]) - \Pr_{x \in \pm 1^n} [f(x) \neq \chi_S(x)]$

Distance between parity functions

Fact 3: if $S \neq T$ then $\Pr_{x \in \{\pm 1\}^n} [\chi_S(x) = \chi_T(x)] = 1/2$

Proof: Let
$$f = \chi_T$$
, then
 $\hat{f}(S) = 0$ (fact 1)
 $= 1 - 2 \Pr[\chi_T(x) \neq \chi_S(x)]$ (fact 2)

Plancherel's Theorem

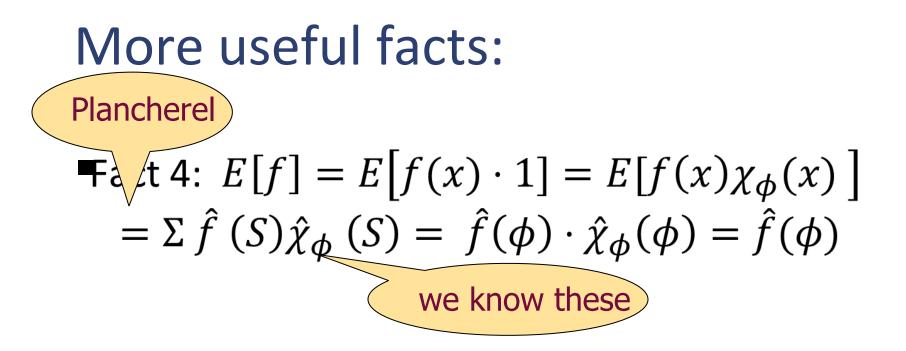
Theorem: For $f, g: \{\pm 1\}^n \to \Re$ we have $\langle f, g \rangle \equiv E_{\{\pm 1\}^n}[f(x)g(x)] = \sum_{S \subseteq [n]} \hat{f}(S) \cdot \hat{g}(S)$

Proof:

Parseval's Theorem

Corollary: For
$$f: \{\pm 1\}^n \to \Re$$
 we have
 $\langle f, f \rangle \equiv E_{\{\pm 1\}^n}[f^2(x)] = \Sigma_{S \subseteq [n]} \hat{f}(S)^2$

Boolean Parseval's: For $f: \{\pm 1\}^n \rightarrow \{\pm 1\}$ $\Sigma_{S \subseteq [n]} \hat{f}(S)^2 = E_{\{\pm 1\}^n} [f^2(x)] = 1$ =1 for all x



Fact 5: (corollary to fact 4 and to fact 1) $E[\chi_{S}(x)] = \begin{cases} 1 & \text{if } S = \phi \\ 0 & 0.W. \end{cases}$

Linearity (homomorphism) testing

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

Linearity Property

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

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

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

Is the proof of the right format?

In these cases, enough for f to be close to homomorphism

What do we mean by ``close''?

Definition: *f*, over domain of size *N*, is ε-close to linear if can change at most ε*N* values to turn it into one.

Otherwise, *ɛ-far*.

What do we mean by ``quick''?

- query complexity measured in terms of domain size N
- Our goal (if possible):
 - constant independent of N?

Linearity Testing

- If f is linear (i.e., $\forall x,y f(x) + f(y) = f(x+y)$) then test should PASS with probability >2/3
- If f is ε-far from linear then test should FAIL with probability >2/3
- Note: If f not linear, but ε-close, then either output is ok

Linearity Testing for f: $GF(2)^n \rightarrow GF(2)$

•
$$x = (x_1, ..., x_n), y = (y_1, ..., y_n) \in \{0, 1\}^n$$

• $x + y = (x_1 \oplus y_1, \dots, x_n \oplus y_n) (\oplus \text{ is "xor"})$

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

• Linear functions are exactly $\{f_a | f_a(x) = \Sigma a_i \cdot x_i \mod 2 \text{ for } a \in \{0,1\}^n \}$ Linearity Testing for f: $\{\pm 1\}^n \rightarrow \{\pm 1\}$

•
$$x = (x_1, ..., x_n), y = (y_1, ..., y_n) \in \{\pm 1\}^n$$

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

• $\forall x, y \quad f(x) \cdot f(y) = f(x \odot y)$

Linear functions are exactly the parity functions {χ_S}

Proposed Tester:

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

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

Easy to see:

- If f is linear, then tester passes with probability 1
- If f is such that $\Pr_{x,y}[f(x)f(y) \neq f(x \odot y)] \ge \rho$ then (constant in O notation can be chosen so that) tester fails with probability at least 2/3

Characterizing "close" to linear

Suppose $\Pr[f(x)f(y) \neq f(x \odot y)]$ is small... is f close to linear?

Nontriviality [Coppersmith]:

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

- f satisfies $f(x)+f(y) \neq f(x+y)$ for only 2/9 of choices of x,y (i.e. $\delta_f = 2/9$)
- *f* is 2/3-far from a linear!

Our goal:

Theorem: If f is
$$\epsilon$$
 – far from linear, then

$$\Pr[f(x)f(y) \neq f(x \odot y)] \ge \epsilon$$

$$\Pr[f(x)f(y)f(x \odot y) \neq 1]$$
Call this δ

Main Lemma:

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

Lemma \rightarrow Theorem

Theorem: If f is ϵ – far from linear, then $\Pr_{x,y}[f(x)f(y)f(x \odot y) \neq 1] \ge \epsilon$

 $\equiv \delta$

Proof:

Main Lemma implies $1 - \delta \leq \frac{1}{2} + \frac{1}{2} \sum_{S \subseteq [n]} \hat{f}(S)^3$

So
$$1 - 2\delta \leq \Sigma \hat{f}(S)^{3}$$

 $\leq \max_{S} (\hat{f}(S)) \Sigma \hat{f}(S)^{2}$
 $= 1$ by Boolean
Parseval
 $\leq \max_{S} (\hat{f}(S))$
 $\leq \hat{f}(T)$
 $\leq 1 - 2 \Pr[f(x) \neq \chi_{T}(x)]$
Fact 2
Fact 2

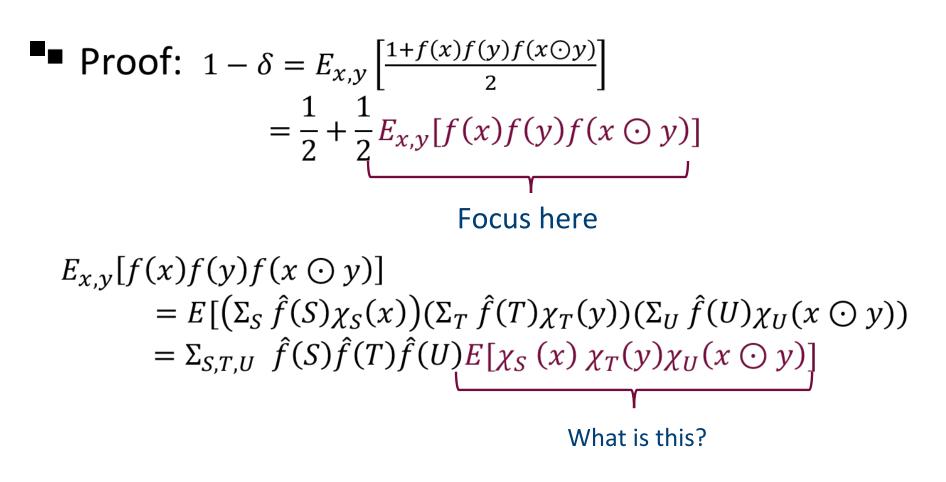
Before the main lemma:

$$= \frac{1+f(x)f(y)f(x \odot y)}{2} \begin{cases} = 1 \ if \ x, y \ PASS \\ = 0 \ if \ x, y \ FAIL \end{cases}$$

Indicator variable describing result of test!

Main Lemma:

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



A final calculation:

 $E[\chi_S(x) \chi_T(y) \chi_U(x \odot y)]$

 $= E[\Pi_{i \in S} x_i \Pi_{j \in T} y_j \Pi_{k \in U} (x_k \cdot y_k)]$ $= E[\Pi_{i \in S \land U} x_i \Pi_{j \in T \land U} y_j]$ $= E[\Pi_{i \in S \land U} x_i] E[\Pi_{i \in T \land U} y_i]$

$$\begin{array}{c} & & \\ 1 \text{ if } S \bigtriangleup U = \phi \\ 0 \text{ o.w.} \end{array} \begin{array}{c} 1 \text{ if } T \bigtriangleup U = \phi \\ 0 \text{ o.w.} \end{array}$$

= 1 if S=T=U and 0 otherwise

Main Lemma:

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

Proof:
$$1 - \delta = E_{x,y} \left[\frac{1 + f(x)f(y)f(x \odot y)}{2} \right]$$

$$= \frac{1}{2} + \frac{1}{2} E_{x,y} [f(x)f(y)f(x \odot y)]$$
Focus here

 $E_{x,y}[f(x)f(y)f(x \odot y)]$ $= E[(\Sigma_S \hat{f}(S)\chi_S(x))(\Sigma_T \hat{f}(T)\chi_T(y))(\Sigma_U \hat{f}(U)\chi_U(x \odot y))$ $= \Sigma_{S,T,U} \hat{f}(S)\hat{f}(T)\hat{f}(U)E[\chi_S(x)\chi_T(y)\chi_U(x \odot y)]$ $= \Sigma_S \hat{f}(S)^3$ 1 if S=T=U

0 otherwise

Linearity tests over other domains

- Still constant, even for general nonabelian groups
- Slightly weaker relationship between parameters

Self-correction

 Given program P computing linear f that is correct on at least 7/8 of the inputs (BUT YOU DON'T KNOW WHICH ONES!)

- Can you correctly compute f on each input?
 - To compute f(x), can't just call P on x...

Self-corrector:

Repeat
$$r = O(\frac{1}{o})$$
 times:

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

Output most common guess

- If P correct on both calls, then guess is correct
- What is probability of this?
 - Observe: Since y uniformly distributed, so is $x \odot y$
 - $\Pr[P \text{ wrong on either } y \text{ or } x \odot y] \leq \frac{1}{4}$