Lecture 8: Fourier Basics for Boolean functions.
Linearity testing.

Lecturer: Ronitt Rubinfeld
Spring 2013
6.893: Sub-linear Algorithms
Why Boolean?

- 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

\[ f: \{0,1\}^n \rightarrow \{0,1\} \]
\[ (x_1, x_2, ..., x_n) \oplus (y_1, y_2, ..., y_n) = (x_1 \oplus y_1, ..., x_n \oplus y_n) \]

\[ f: \{\pm1\}^n \rightarrow \{\pm1\} \]
\[ (x_1, x_2, ..., x_n) \odot (y_1, y_2, ..., y_n) = (x_1 \cdot y_1, ..., x_n \cdot y_n) \]
The great (notational) switch:

\[
\begin{array}{c}
0 \rightarrow +1 \\
1 \rightarrow -1
\end{array}
\]

\[
\begin{array}{c|c|c}
\oplus & 0 & 1 \\
0 & 0 & 1 \\
1 & 1 & 0 \\
\end{array} 
\rightarrow 
\begin{array}{c|c|c}
\times & +1 & -1 \\
+1 & +1 & -1 \\
-1 & -1 & +1 \\
\end{array}
\]
The set of functions and inner product

• \( G = \{ g \mid g : \{\pm 1\}^n \to \mathbb{R} \} \) (all \( n \)-bit fctns into Reals)

• A vector space of dimension \( 2^n \)
  
  • 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?

- \( G = \{g \mid g: \{\pm 1\}^n \rightarrow \mathbb{R}\} \) (all \( n \)-bit fctns into Reals)
- A “natural” basis: indicator functions
  - \( e_a(x) = \begin{cases} 1 & \text{if } x = a \\ 0 & \text{o.w.} \end{cases} \)

- Orthonormal
- Used to describe function via “truth table”
  \[ f(x) = \sum_a f(a)e_a(x) \]
A very useful basis:

- \( G = \{g \mid g: \{\pm 1\}^n \to \mathbb{R}\} \) (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 \( \chi_\emptyset(x) = 1 \forall x \)
A useful property:

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

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

\[= \prod_{S \cap T} x_i^2 \prod_{i \in S \Delta T} x_i \]

\[=1 \]
Inner product

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

- Note:

$\langle \chi_S, \chi_S \rangle = \frac{1}{2^n} \sum_{x \in \{\pm 1\}^n} (\chi_S(x))^2 = 1$
Orthogonal:

- If $S \neq T$:

  \[
  \langle \chi_S, \chi_T \rangle = \frac{1}{2^n} \sum_x \chi_S(x) \chi_T(x) \\
  = \frac{1}{2^n} \sum_x \chi_{S \Delta T}(x) \\
  = \frac{1}{2^n} \sum_x \prod_{i \in S \Delta T} x_i \\
  = \frac{1}{2^n} \sum_{pairs \ \ x, x \oplus j} \left( \prod_{i \in S \Delta T} x_i + \prod_{i \in S \Delta T} \left( x \oplus j \right)_i \right) \\
  = 0 \text{ since each pair sums to } 0: \\
  x_j \left( \prod_{(i \in S \Delta T \setminus \{j\})} x_i \right) - x_j \left( \prod_{(i \in S \Delta T \setminus \{j\})} x_i \right) = 0
  \]
So we have an orthonormal basis!

- Every function can be written as a linear combination of these $\chi_S$'s

- **Theorem:**
  $$\forall f, f(x) = \sum_S \hat{f}(S) \chi_S(x)$$
  where
  $$\hat{f}(S) = \langle f, \chi_S \rangle = \frac{1}{2^n} \sum_{x \in \{-1,1\}^n} f(x) \chi_S(x)$$
Some examples:

<table>
<thead>
<tr>
<th>Function</th>
<th>Fourier Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$f(x)=1$</td>
<td>$1$</td>
</tr>
<tr>
<td>$f(x)= x_i$</td>
<td>$x_i$</td>
</tr>
<tr>
<td>$f(x)= \text{AND}(x_1, x_2)$</td>
<td>$\frac{1}{2} + \frac{1}{2} x_1 + \frac{1}{2} x_2 - \frac{1}{2} x_1 x_2$</td>
</tr>
</tbody>
</table>
Fourier coefficients of parity functions:

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

Proof:

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

$= \frac{1}{2^n} \sum_{x \text{ s.t. } f(x) = \chi_S(x)} (+1) + \frac{1}{2^n} \sum_{x \text{ s.t. } f(x) \neq \chi_S(x)} (-1)$

$= (1 - \Pr_{x \in \pm 1^n} [f(x) \neq \chi_S(x)]) - \Pr_{x \in \pm 1^n} [f(x) \neq \chi_S(x)]$
Distance between parity functions

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

Proof: Let \( f = \chi_T \), then
\[
\hat{f}(S) = 0 \quad \text{(fact 1)}
\]
\[
= 1 - 2 \Pr[\chi_T(x) \neq \chi_S(x)] \quad \text{(fact 2)}
\]
Plancherel’s Theorem

Theorem: For \( f, g : \{\pm 1\}^n \to \mathbb{R} \) 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:
\[
\langle f, g \rangle = \langle \sum_S \hat{f}(S) \chi_S, \sum_T \hat{g}(T) \chi_T \rangle \quad \text{(def)}
\]
\[
= \sum_S \sum_T \hat{f}(S) \hat{g}(T) \langle \chi_S, \chi_T \rangle \quad \text{(bilinearity)}
\]
\[
= \sum_S \hat{f}(S) \hat{g}(S) \quad \text{(orthogonality)}
\]
Parseval’s Theorem

Corollary: For $f: \{\pm 1\}^n \rightarrow \mathbb{R}$ we have
$$<f, f> \equiv E_{\{\pm 1\}^n}[f^2(x)] = \sum_{S \subseteq [n]} \hat{f}(S)^2$$

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

=1 for all x
More useful facts:

Fact 4: \[ E[f] = E[f(x) \cdot 1] = E[f(x)\chi_\phi(x)] \]
\[ = \sum \hat{f}(S)\hat{\chi}_\phi(S) = \hat{f}(\phi) \cdot \hat{\chi}_\phi(\phi) = \hat{f}(\phi) \]

Fact 5: (corollary to fact 4 and to fact 1)
\[ E[\chi_S(x)] = \begin{cases} 1 & \text{if } S = \phi \\ 0 & \text{otherwise} \end{cases} \]
Linearity (homomorphism) testing

∀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 $\varepsilon$-close to linear if can change at most $\varepsilon N$ values to turn it into one.

Otherwise, $\varepsilon$-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 $>\frac{2}{3}$

- If $f$ is $\epsilon$-far from linear then test should FAIL with probability $>\frac{2}{3}$

- Note: If $f$ not linear, but $\epsilon$-close, then either output is ok
Linearity Testing for 
\( f: GF(2)^n \rightarrow GF(2) \)

- \( x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in \{0,1\}^n \)
  - \( x + y = (x_1 \oplus y_1, \ldots, x_n \oplus y_n) \) (\( \oplus \) is "xor")
- \( \forall x, y f(x) \oplus f(y) = f(x + y) \)
- Linear functions are exactly
  \( \{f_a \mid f_a(x) = \Sigma a_i \cdot x_i \mod 2 \text{ for } a \in \{0,1\}^n \} \)
Linearity Testing for 

\[ f: \{\pm 1\}^n \rightarrow \{\pm 1\} \]

- \[ x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in \{\pm 1\}^n \]
- \[ x \odot y = (x_1 \cdot y_1, \ldots, x_n \cdot y_n) \]
- \[ \forall x, y \quad f(x) \cdot f(y) = f(x \odot y) \]

- Linear functions are exactly the parity functions \( \{\chi_S\} \)
Proposed Tester:

- Repeat $r = O\left(\frac{1}{\rho}\right)$ times:
  - Pick $x, y \in \mathbb{R}_0 \{0,1\}^n$
  - 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 $\Pr_{x,y}[f(x)f(y) \neq f(x \odot y)] \geq \rho$ then tester fails with probability at least $2/3$
Characterizing “close” to linear

- Suppose $\Pr_{x,y} [f(x)f(y) \neq f(x \odot y)]$ is small... is $f$ close to linear?
Nontriviality [Coppersmith]:

- $f: \mathbb{Z}_{3^k} \rightarrow \mathbb{Z}_{3^{k-1}}$
- $f(3h+d)=h$, 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

\[
\Pr_{x,y}[f(x)f(y) \neq f(x \odot y)] \geq \epsilon
\]

Main Lemma:

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

Call this \( \delta \)
Lemma ➔ Theorem

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

Proof:
Main Lemma implies
\[
1 - \delta \leq \frac{1}{2} + \frac{1}{2} \sum_{S \subseteq [n]} \hat{f}(S)^3
\]
So
\[
1 - 2\delta \leq \sum \hat{f}(S)^3
\]
\[
\leq \max_S (\hat{f}(S)) \sum \hat{f}(S)^2
\]
\[
\leq \max_S (\hat{f}(S)) \leq \hat{f}(T)
\]
\[
\leq 1 - 2 \Pr[f(x) \neq \chi_T(x)]
\]

So $\delta \geq \epsilon$

$\equiv \delta$

$= 1$ by Boolean Parseval

Pick $T$ to maximize

Fact 2
Before the main lemma:

\[
\frac{1 + f(x)f(y)f(x \odot y)}{2} \begin{cases} 
1 & \text{if } x, y \text{ PASS} \\
0 & \text{if } x, y \text{ FAIL}
\end{cases}
\]

Indicator variable describing result of test!
Main Lemma:

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

**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

\[ E_{x,y}[f(x)f(y)f(x \odot y)] \]

\[ = E[(\sum_S \hat{f}(S)\chi_S(x))(\sum_T \hat{f}(T)\chi_T(y))(\sum_U \hat{f}(U)\chi_U(x \odot y))] \]

\[ = \sum_{S,T,U} \hat{f}(S)\hat{f}(T)\hat{f}(U) E[\chi_S(x) \chi_T(y)\chi_U(x \odot y)] \]

What is this?
A final calculation:

\[ E[\chi_S(x) \chi_T(y) \chi_U(x \ominus y)] \]

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

\[ = E[\prod_{i \in S \Delta U} x_i \prod_{j \in T \Delta U} y_j] \]

\[ = E[\prod_{i \in S \Delta U} x_i] E[\prod_{j \in T \Delta U} y_j] \]

1 if \( S \Delta U = \emptyset \) 1 if \( T \Delta U = \emptyset \)

0 o.w. 0 o.w.

\[ = 1 \text{ if } S=T=U \text{ and } 0 \text{ otherwise} \]
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

\[ E_{x,y} [f(x)f(y)f(x \odot y)] \]

\[ = E[(\sum_S \hat{f}(S)\chi_S(x))(\sum_T \hat{f}(T)\chi_T(y))(\sum_U \hat{f}(U)\chi_U(x \odot y))] \]

\[ = \sum_{S,T,U} \hat{f}(S)\hat{f}(T)\hat{f}(U)E[\chi_S (x) \chi_T (y)\chi_U (x \odot y)] \]

\[ = \sum_S \hat{f}(S)^3 \]

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 $r = O\left(\frac{1}{\rho}\right)$ times:
  - Pick $y \in \{0,1\}^n$
  - Let $\text{guess}(x) \leftarrow P(y) \cdot P(x \odot y)$
- 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 $x \odot y$
  - $\Pr[P \text{ wrong on either } y \text{ or } x \odot y] \leq \frac{1}{4}$