6.842 Randomness and Computation March 30, 2022

Lectures on Linearity Testing

Lecturer: Ronitt Rubinfield *Scribe: Yuchong Pan*

1 Linearity Testing

Definition 1. Let G and H be finite groups. Let $f : G \to H$. Then f is said to be *linear* (i.e., is a *homomorphism*) if for all $x, y \in G$,

$$f(x) +_H f(y) =_H f(x +_G y).$$

For all $\varepsilon > 0$, f is said to be ε-*linear* if there exists a linear function $g : G \to H$ such that f and g agree on at least $1 - \varepsilon$ fraction of inputs in G, i.e.,

$$\mathop{\mathbb{P}}_{x \in G}[f(x) = g(x)] \geq 1 - \varepsilon,$$

or equivalently,

$$\frac{|\{x \in G : f(x) = g(x)\}|}{|G|} \geq 1 - \varepsilon.$$

Algorithm 1 is a natural test for the linearity of a function $f : G \to H$, where G and H are finite groups.

repeat *?* **times**
  pick random $x, y \in G$
  if $f(x) + f(y) \neq f(x + y)$ **then**
    return *"fail"*
return *"pass"*

Algorithm 1: A proposed test for the linearity of a function $f : G \to H$, where G and H are finite groups.

Observation 2. *Let G be a finite group. For all $a, y \in G$, $\mathbb{P}_{x \in G}[y = a + x] = 1/|G|$. In other words, if x is chosen uniformly from G, then $a + x$ is also uniformly distributed in G.*

Proof. Since only $x = y - a$ satisfies $y = a + x$, then $\mathbb{P}_{x \in G}[y = a + x] = \mathbb{P}_{x \in G}[x = y - a] = 1/|G|$. □

2 Self-Correcting (Random Self-Reducibility)

Theorem 3. *Let G be a fintie group. Let $f : G \to G$ be a function such that there exists a linear function $g : G \to G$ and that $\mathbb{P}_{x \in G}[f(x) = g(x)] \geq 7/8$. Then for all $x \in G$, $g(x)$ can be computed with only $O(\log(1/\beta))$ calls to f (with at most β probability of error).*

Given input $x \in G$ and black box access to f, we define a *self corrector* in Algorithm 2.

Proposition 4. $\mathbb{P}[\mathit{output} = g(x)] \geq 1 - \beta$.

```
for i ← 1, ..., C · log(1/β) do
    pick y uniformly in G
    answer_i ← f(y) + f(x − y)
output the most common answer
```

Algorithm 2: A self corrector for a 1/8-linear function $f : G \to G$ on input x, where G is a finite group.

Proof. Let y be chosen uniformly in G. By Observation 2, $x - y$ is also uniformly distributed in G. Therefore,

$$\mathbb{P}[f(y) \neq g(y)] \leq \frac{1}{8}, \qquad \mathbb{P}[f(x-y) \neq g(x-y)] \leq \frac{1}{8}.$$

By the union bound,

$$\begin{aligned} \mathbb{P}[f(y) + f(x-y)] = g(x)] &= \mathbb{P}[f(y) + f(x-y)] = g(y) + g(x-y)] \\ &\geq \mathbb{P}[f(y) = g(y), f(x-y) = g(x-y)] \\ &\geq 1 - \left(\frac{1}{8} + \frac{1}{8}\right) = \frac{3}{4}. \end{aligned}$$

This implies that $\mathbb{P}[answer_i = g(x)] \geq 3/4$ for all i. The proof is hence complete. $\square$

3 Coppersmith's Example

Let $m \in \mathbb{N}$. Let $f : \mathbb{Z}_m \to \mathbb{Z}_m$ be defined by

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

The graph of f is plotted in Figure 1.

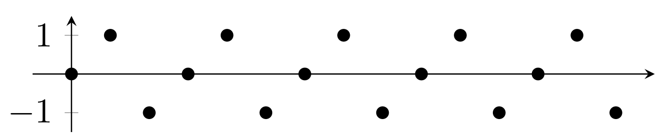

Figure 1: The graph of Coppersmith's example.

Note that the closest linear function $g : \mathbb{Z}_m \to \mathbb{Z}_m$ to f is given by $g(x) = 0$ for all $x \in \mathbb{Z}_m$, so f is 2/3-far from being linear. Note that f fails for $x, y \in \mathbb{Z}_m$ with $x \equiv y \equiv 1 \pmod 3$ or $x \equiv y \equiv 2 \pmod 3$, and passes for all other $x, y \in \mathbb{Z}_m$. Therefore, the *rejection probability of the linearity test* for f, denoted by δ_f, is given by

$$\delta_f = \mathbb{P}_{x,y \in \mathbb{Z}_m}[f(x) + f(y) \neq f(x+y)] = \frac{2}{9}.$$

Fortunately, 2/9 is the threshold; in other words, Coppersmith's example is the worst example. If $\delta_f < 2/9$ for some function $f : G \to G$ and finite group G, then f must be δ_f-close to being linear.

4 Fourier Analysis for Boolean Functions

The *n-dimensional Boolean hypercube* $\{0,1\}^n$ can be interpreted as having $n+1$ layers, where the i^{th} layer consists of n-bit Boolean strings with i ones for each $i \in \{0,\ldots,n\}$, and where two n-bit Boolean strings in consecutive layers are joined by an edge if they differ at exactly one bit. What are linear maps $\{0,1\}^n \to \{0,1\}$?

Definition 5. Given $x, y \in \{0,1\}^n$, the *inner product* of x and y is defined to be

$$x \cdot y = \sum_{i=1}^{n} x_i y_i \pmod 2.$$

Note that addition modulo 2 is the XOR operation. Linear functions on $\{0,1\}^n$ are of the form

$$L_a(x) = a \cdot x, \qquad \text{for fixed } a \in \{0,1\}^n,$$

or, alternatively,

$$L_A(x) = \sum_{i \in A} x_i \pmod 2, \qquad \text{for fixed } A \subset [n].$$

Therefore, there are exactly 2^n linear functions on $\{0,1\}^n$.

To simplify the presentation, we change the notation by letting $a \mapsto (-1)^a$ for $a \in \{0,1\}$ and by changing addition $a+b$ to multiplication $(-1)^a(-1)^b = (-1)^{a+b}$. Hence, the condition of linearity $f(a)+f(b) = f(a \oplus b)$ for all $a, b \in \{0,1\}^n$ is changed to $f(a) \cdot f(b) = f(a \odot b)$ for all $a, b \in \{1,-1\}^n$, where $(x_1,\ldots,x_n) \oplus (y_1,\ldots,y_n) = (x_1+y_1,\ldots,x_n+y_n)$ denotes the bitwise XOR (i.e., addition modulo 2) of two n-bit Boolean strings, and $(x_1,\ldots,x_n) \odot (y_1,\ldots,y_n) = (x_1 \cdot y_1,\ldots,x_n \cdot y_n)$ denotes the bitwise multiplication of two n-bit $\{1,-1\}$-valued strings. Moreover, linear functions on $\{1,-1\}^n$ are of the form

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

We want to find a basis to describe all functions $f : \{\pm 1\}^n \to \{\pm 1\}$. The first idea is to use the "input-output table"; in other words, the basis consists of all indicator functions

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

for all $a \in \{\pm 1\}^n$. Then for any function $f : \{\pm 1\}^n \to \{\pm 1\}$,

$$f(x) = \sum_{a \in \{\pm 1\}^n} f(a) e_a(x).$$

For the purpose of linearity testing, we introduce the second idea, i.e., to use linear functions (a.k.a. *parity functions*) $\chi_S(x) = \prod_{i \in S} x_i$ for all $S \subset [n]$.

Definition 6. Given $f, g : \{\pm 1\}^n \to \{\pm 1\}$, the *(normalized) inner product* of f, g is defined to be

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

Proposition 7. *The set of parity functions $\{\chi_S : S \subset [n]\}$ is an orthonormal basis with respect to the inner product.*

Proof. For $S \subset [n]$,

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

Let $S, T \subset [n]$ be such that $S \neq T$. Then

$$\begin{aligned}
\langle \chi_S, \chi_T \rangle &= \frac{1}{2^n} \sum_{x \in \{\pm 1\}^n} \chi_S(x)\chi_T(x) = \frac{1}{2^n} \sum_{x \in \{\pm 1\}^n} \left(\prod_{i \in S} x_i \right) \left(\prod_{i \in T} x_i \right) \\
&= \frac{1}{2^n} \sum_{x \in \{\pm 1\}^n} \left(\prod_{i \in S \cap T} x_i^2 \right) \left(\prod_{i \in S \triangle T} x_i \right) = \frac{1}{2^n} \sum_{x \in \{\pm 1\}^n} 1 \cdot \prod_{i \in S \triangle T} x_i \\
&= \frac{1}{2^n} \sum_{x \in \{\pm 1\}^n} \chi_{S \triangle T}(x).
\end{aligned}$$

Note $S \triangle T \neq \emptyset$ since $S \neq T$. Let $j \in S \triangle T$. Let $x^{\oplus j}$ be obtained by flipping the j^{th} bit in x. Then

$$\begin{aligned}
\langle \chi_S, \chi_T \rangle &= \frac{1}{2^n} \sum_{\text{pairs } x, x^{\oplus j}} \left(\chi_{S \triangle T}(x) + \chi_{S \triangle T}\left(x^{\oplus j}\right) \right) \\
&= \frac{1}{2^n} \sum_{\text{pairs } x, x^{\oplus j}} \left(x_j \prod_{i \in (S \triangle T) \setminus \{j\}} x_i + (-x_j) \prod_{i \in (S \triangle T) \setminus \{j\}} x_i \right) \\
&= \frac{1}{2^n} \sum_{\text{pairs } x, x^{\oplus j}} 0 = 0. \qquad (1)
\end{aligned}$$

This completes the proof. $\square$

Corollary 8. *Any function $f : \{\pm 1\}^n \to \{\pm 1\}$ is uniquely expressible as a linear combination of the parity functions χ_S for $S \subset [n]$.*

Definition 9. For any function $f : \{\pm 1\}^n \to \{\pm 1\}$ and any $S \subset [n]$, the *Fourier coefficient* of f at S is defined to be

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

Theorem 10. *For any function $f : \{\pm 1\}^n \to \{\pm 1\}$,*

$$f(x) = \sum_{S \subset [n]} \hat{f}(S)\chi_S(x).$$

Proposition 11 (Fourier coefficients of linear functions)**.** *Any function $f : \{\pm 1\}^n \to \{\pm 1\}$ is linear if and only if there exists $S \subset [n]$ such that $\hat{f}(S) = 1$ and $\hat{f}(T) = 0$ for all $T \subset [n]$ with $T \neq S$.*

Proposition 12. *For any $S \subset [n]$,*

$$\hat{f}(S) = 1 - 2\,\mathrm{dist}\left(f, \chi_S\right),$$

where

$$\mathrm{dist}\left(f, \chi_S\right) := \mathop{\mathbb{P}}_{x \in \{\pm 1\}^n} \left[f(x) \neq \chi_S(x)\right] = \frac{|\{x \in \{\pm 1\}^n : f(x) \neq \chi_S(x)\}|}{2^n}.$$

Proof. We have

$$\begin{aligned} 2^n \hat{f}(S) &= \sum_{x\in\{\pm1\}^n} f(x)\chi_S(x) = \sum_{\substack{x\in\{\pm1\}^n \\ f(x)=\chi_S(x)}} 1 + \sum_{\substack{x\in\{\pm1\}^n \\ f(x)\neq\chi_S(x)}} (-1) \\ &= (1-\operatorname{dist}(f,\chi_S))\cdot 2^n \cdot 1 + \operatorname{dist}(f,\chi_S)\cdot 2^n\cdot(-1) \\ &= 2^n(1-2\operatorname{dist}(f,\chi_S)). \end{aligned}$$

This completes the proof. □

Lemma 13. *Any two distinct linear functions differ on exactly half of the inputs.*

Proof. Let $S, T \subset [n]$ be such that $S \neq T$. Then

$$\begin{aligned} 0 &= \langle \chi_S, \chi_T \rangle && \text{(orthonormality)} \\ &= \widehat{\chi_S}(T) \\ &= 1 - 2\operatorname{dist}(\chi_S, \chi_T) && \text{(Proposition 12)}. \end{aligned}$$

This implies that $\operatorname{dist}(\chi_S, \chi_T) = 1/2$, completing the proof. □

Lemma 14 (Pancherel's identity)**.** *For functions* $f, g : \{\pm1\}^n \to \{\pm1\}$,

$$\langle f, g\rangle = \sum_{S\subset[n]} \hat{f}(S)\hat{g}(S).$$

Proof. For functions $f, g : \{\pm1\}^n \to \{\pm1\}$,

$$\begin{aligned} \langle f, g\rangle &= \left\langle \sum_{S\subset[n]} \hat{f}(S)\chi_S, \sum_{T\subset[n]} \hat{g}(T)\chi_T \right\rangle \\ &= \sum_{S,T\subset[n]} \hat{f}(S)\hat{g}(T)\langle \chi_S, \chi_T\rangle && \text{(bilinearity)} \\ &= \sum_{S\subset[n]} \hat{f}(S)\hat{g}(S). && \text{(orthonormality)} \end{aligned}$$

This completes the proof. □

Corollary 15 (Boolean Parseval's identity)**.** *For any function* $f : \{\pm1\}^n \to \{\pm1\}$,

$$\sum_{S\subset[n]} \hat{f}(S)^2 = \langle f, f\rangle = \frac{1}{2^n}\sum_{x\in\{\pm1\}^n} f(x)^2 = \frac{1}{2^n}\sum_{x\in\{\pm1\}^n} 1 = 1.$$

5 Linearity Testing for Boolean Functions

Now we apply Fourier analysis for Boolean functions developed in Section 4 to linearity testing for Boolean functions. By Propositions 11 and 12, a function $f : \{\pm1\}^n \to \{\pm1\}$ is ε-linear if and only if there exists $S \subset [n]$ such that $\hat{f}(S) \geq 1 - 2\varepsilon$.

Now, we define a linearity test for a given Boolean function in Algorithm 3. Then

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

```
1 pick random x, y ∈ {±1}^n
2 test f(x) · f(y) = f(x ⊙ y)
```

Algorithm 3: A linearity test for a given Boolean function $f : \{\pm 1\}^n \to \{\pm 1\}$

Therefore, the indicator variable for the event that the test rejects is given by

$$\mathbb{1}_{f(x)\cdot f(y)\neq f(x\odot y)} = \frac{1 - f(x)f(y)f(x\odot y)}{2} = \begin{cases} 0, & \text{if the test accepts,} \\ 1, & \text{if the test rejects.} \end{cases}$$

This allows us to express the *rejection probability* in terms of the indicator variable:

$$\delta_f := \mathop{\mathbb{P}}_{x,y\in\{1,-1\}^n}[f(x)\cdot f(y) \neq f(x\odot y)] = \mathop{\mathbb{E}}_{x,y\in\{1,-1\}^n}\left[\frac{1 - f(x)f(y)f(x\odot y)}{2}\right]. \tag{2}$$

Theorem 16. *Any function $f : \{\pm 1\}^n \to \{\pm 1\}$ is δ_f-close to some linear function.*

Proof. We have

$$\begin{aligned}
&\mathop{\mathbb{E}}_{x,y\in\{\pm 1\}^n}[f(x)f(y)f(x\odot y)] \\
&= \mathop{\mathbb{E}}_{x,y\in\{\pm 1\}^n}\left[\left(\sum_{S\subset[n]}\hat{f}(S)\chi_S(x)\right)\left(\sum_{T\subset[n]}\hat{f}(T)\chi_T(y)\right)\left(\sum_{U\subset[n]}\hat{f}(U)\chi_U(x\odot y)\right)\right] \\
&= \mathop{\mathbb{E}}_{x,y\in\{\pm 1\}^n}\left[\sum_{S,T,U\subset[n]}\hat{f}(S)\hat{f}(T)\hat{f}(U)\chi_S(x)\chi_T(y)\chi_U(x\odot y)\right] \\
&= \sum_{S,T,U\subset[n]}\hat{f}(S)\hat{f}(T)\hat{f}(U)\mathop{\mathbb{E}}_{x,y\in\{\pm 1\}^n}\left[\chi_S(x)\chi_T(y)\chi_U(x\odot y)\right].
\end{aligned}$$

For any $S \subset [n]$,

$$\chi_S(x)\chi_S(y)\chi_S(x\odot y) = \left(\prod_{i\in S}x_i\right)\left(\prod_{i\in S}y_i\right)\left(\prod_{i\in S}x_iy_i\right) = \left(\prod_{i\in S}x_i^2\right)\left(\prod_{i\in S}y_i^2\right) = 1.$$

For any $S, T, U \subset [n]$ such that it is not the case that $S = T = U$,

$$\begin{aligned}
\mathop{\mathbb{E}}_{x,y\in\{\pm 1\}^n}\left[\chi_S(x)\chi_T(y)\chi_U(x\odot y)\right] &= \mathop{\mathbb{E}}_{x,y\in\{\pm 1\}^n}\left[\left(\prod_{i\in S}x_i\right)\left(\prod_{i\in T}y_i\right)\left(\prod_{i\in U}x_iy_i\right)\right] \\
&= \mathop{\mathbb{E}}_{x,y\in\{\pm 1\}^n}\left[\left(\prod_{i\in S\triangle U}x_i\right)\left(\prod_{i\in T\triangle U}y_i\right)\right] \\
&= \mathop{\mathbb{E}}_{x,y\in\{\pm 1\}^n}\left[\prod_{i\in S\triangle U}x_i\right]\mathop{\mathbb{E}}_{x,y\in\{\pm 1\}^n}\left[\prod_{i\in T\triangle U}y_i\right] \qquad (3) \\
&= \mathop{\mathbb{E}}_{x\in\{\pm 1\}^n}\left[\prod_{i\in S\triangle U}x_i\right]\mathop{\mathbb{E}}_{x\in\{\pm 1\}^n}\left[\prod_{i\in T\triangle U}x_i\right].
\end{aligned}$$

Note that (3) follows from the independence of x and y. Since it is not the case that $S = T = U$, then either $S \neq U$ or $T \neq U$. WLOG, assume $S \neq U$. By (1), $\mathbb{E}_{x\in\{\pm1\}^n}[\prod_{i\in S\triangle U} x_i] = 0$. Therefore, $\mathbb{E}_{x,y\in\{\pm1\}^n}[\chi_S(x)\chi_T(y)\chi_U(x\odot y)] = 0 \cdot 0 = 0$. It follows that

$$\begin{aligned}
\mathop{\mathbb{E}}_{x,y\in\{\pm1\}^n}[f(x)f(y)f(x\odot y)] &= \sum_{S\subset[n]} \hat{f}(S)^3 \\
&\leq \left(\max_{S\subset[n]} \hat{f}(S)\right) \sum_{S\subset[n]} \hat{f}(S)^2 \\
&= \left(\max_{S\subset[n]} \hat{f}(S)\right) \cdot 1 && \text{(Corollary 15)} \\
&= \max_{S\subset[n]} \hat{f}(S) \\
&= \max_{S\subset[n]} \left(1 - 2\,\mathrm{dist}\left(f, \chi_S\right)\right) && \text{(Proposition 12)} \\
&= 1 - 2 \min_{S\subset[n]} \mathrm{dist}\left(f, \chi_S\right).
\end{aligned}$$

By (2),

$$\delta_f \geq \frac{1 - \left(1 - 2\min_{S\subset[n]} \mathrm{dist}\left(f, \chi_S\right)\right)}{2} = \min_{S\subset[n]} \mathrm{dist}\left(f, \chi_S\right).$$

This completes the proof. $\square$