6.5420 Randomness and Computation March 19, 2026

Lecture 12

Lecturer: Ronitt Rubinfeld *Scribe: Vasawat Rawangwong*

This lecture includes the topics:

- Linearity Testing
- Self-Correction
- Linear Functions Over the Boolean Cube (Introduction)

1 Linearity

First we define linearity.

Definition 1 (Linear Functions)**.** *For any finite groups G and H, a function $f : G \to H$ is linear if and only if*

$$f(x) +_H f(y) = f(x +_G y)\ \forall x, y \in G$$

Where $+_G$ and $+_H$ denote the addition operations for groups G and H respectively.

The following are some examples of linear functions:

- $f : G \to G$, $f(x) = x$.
  f is linear because $f(x) + f(y) = x + y = f(x + y)$.
- $f : \mathbb{Z}_p \to \mathbb{Z}_p$, $f(x) = ax \pmod p$.
  f is linear because $f(x) + f(y) = ax + ay \pmod p = a(x + y) \pmod p = f(x + y)$
- $f : \mathbb{F}_2^n \to \mathbb{F}_2$, $f_{\mathbf{a}}(\mathbf{x}) = \mathbf{a}^T\mathbf{x} \pmod 2 = \sum_{i=1}^n a_i x_i \pmod 2$.
  $f_{\mathbf{a}}$ is linear because $f_{\mathbf{a}}(\mathbf{x}) + f_{\mathbf{a}}(\mathbf{y}) = \mathbf{a}^T\mathbf{x} + \mathbf{a}^T\mathbf{y} \pmod 2 = \mathbf{a}^T(\mathbf{x} + \mathbf{y}) \pmod 2 = f_{\mathbf{a}}(\mathbf{x} + \mathbf{y})$

Note that by definition 1, some affine linear functions such as $f : \mathbb{Z}_p \to \mathbb{Z}_p$, $f(x) = ax + b \pmod p$ for $b \neq 0$ aren't linear.
We also define functions which are close to being linear.

Definition 2 (ε-Linear Functions)**.** *A function $f : G \to H$ is 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. Or equivalently, $\Pr_{x\in G}[f(x) = g(x)] \geq 1 - \varepsilon$.*

Remark. For sufficiently small ε, the linear function g chosen by the above definition is unique. Here is a proof of this fact for $\varepsilon < \frac{1}{4}$:
Proof: Suppose f is ε-linear and there exists distinct linear g_1, g_2 such that

$$\Pr_{x\in G}[f(x) = g_1(x)], \Pr_{x\in G}[f(x) = g_2(x)] \geq 1 - \varepsilon$$

Let $h(x) = g_1(x) - g_2(x)$, which must also be a linear function. Then by the union bound:

$$\begin{aligned}\Pr_{x\in G}[h(x) = 0] &\geq \Pr_{x\in G}[f(x) = g_1(x) = g_2(x)] \\ &\geq 1 - \Pr_{x\in G}[f(x) \neq g_1(x)] - \Pr_{x\in G}[f(x) \neq g_2(x)] \\ &\geq 1 - 2\varepsilon > \frac{1}{2}\end{aligned}$$

As g_1 and g_2 are distinct, there must exist some $y \neq 0$ such that $h(y) \neq 0$. (Note that $h(0)+h(0) = h(0+0)$, so $h(0) = 0$ is forced.) By linearity, $h(x) + h(y-x) = h(y) \neq 0$, so either $h(x) \neq 0$ or $h(y-x) \neq 0$. This pairing argument implies that $\Pr_{x \in G}[h(x) = 0] \leq \frac{1}{2}$, which is a contradiction. Therefore, g is unique.

From the definition of ε-linearity, two questions naturally arise:

- Given a function f which is ε-linear, how do we correctly recover the value of true linear function $g(x)$? (Self-Correction)
- Given a function f, how do we test whether it is ε-linear? (Linearity Testing)

2 Self-Correction

Given a function $f : G \to G$ such that $\exists$ linear $g : G \to G$, $\Pr[f(x) = g(x)] \geq \frac{7}{8}$. Consider the following algorithm to output $g(x)$:

1. For $i = 1, 2, \ldots, c\log(1/\beta)$:
   - Pick $y \in_R G$
   - Set answer$_i \leftarrow f(y) + f(x-y)$
2. Output most common answer$_i$.

We claim that $\Pr[\text{output} = g(x)] \geq 1 - \beta$ for a suitable constant c.
Proof: If $f(y) = g(y)$ and $f(x-y) = g(x-y)$, we must have $f(y)+f(x-y) = g(y)+g(x-y) = g(x)$. Therefore by the union bound:

$$\begin{aligned}\Pr[\text{answer}_i = g(x)] &\geq 1 - \Pr[f(y) \neq g(y)] - \Pr[f(x-y) \neq g(x-y)] \\ &\geq \frac{3}{4}\end{aligned}$$

Then we can bound the probability that the majority answer is incorrect using the Chernoff Bound:

$$\begin{aligned}\Pr[\text{Majority answer} \neq g(x)] &= \Pr\left[\text{Number of incorrect answer}_i \geq \frac{c\log(1/\beta)}{2}\right] \\ &\leq \exp\left(-\frac{c\log(1/\beta)}{4}\left(2\log 2 - 2 + 1\right)\right) \\ &= \beta^{\frac{c(2\log 2 - 1)}{4}}\end{aligned}$$

Setting $c = \frac{4}{2\log 2 - 1}$, we have $\Pr[\text{output} \neq g(x)] \leq \beta$, so $\Pr[\text{output} = g(x)] \geq 1 - \beta$.

3 Linearity Testing

Given a function $f : G \to G$ and parameter ε, consider the following algorithm to test for ε-linearity:

1. Repeat $k = c\log(\varepsilon^{-1})$ times:
   - Pick random $x, y \in_R G$
   - Test whether $f(x) + f(y) = f(x+y)$.

We claim that for a suitable constant c, the above tester satisfies the following properties:

- If f is linear, then $\Pr[\text{tester passes}] = 1$

- If f is not ε-linear, then Pr[tester fails] $\geq \frac{3}{4}$

Note that we don't impose any constraints on the case where f is ε-linear but not exactly linear. Define $\delta_f = \Pr_{x,y}[f(x)+f(y) \neq f(x+y)]$. The problem is that the relation between ε and δ_f is not clear. In fact, it is possible for a function to be far from linear but is likely to pass an iteration of the linearity test.

3.1 Coppersmith's Example

Define the function f as follows:

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

Note that the best linear function approximating $f(x)$ is $g(x) = 0$, so f is $\frac{2}{3}$-linear. However, running one iteration of the linearity test on f, we find that failure occurs only when $x \equiv y \equiv 1, 2 \pmod 3$, which happens with probability $\frac{2}{9}$. For this example, f is far from being a linear function, but does well on the Linearity Test. The good news is that this is the worst case scenario.

Theorem 3. *If $\delta_f < \frac{2}{9}$, then f is δ_f-linear.*

The correctness of the linearity test will be proved in a later lecture.

4 The Boolean Cube (Introduction)

We are interested in linear functions $f : \{0,1\}^n \to \{0,1\}$. Define the inner product $\mathbf{x} \cdot \mathbf{y} = \sum_{i=1}^{n} x_i y_i \pmod 2$. Then all such linear functions can be written in the following form:

$$L_{\mathbf{a}}(\mathbf{x}) = \mathbf{a} \cdot \mathbf{x}$$

for some fixed $\mathbf{a} \in \{0,1\}^n$. Alternatively, we can also write linear functions in the form:

$$L_A(\mathbf{x}) = \sum_{i \in A} x_i \pmod 2$$

for some fixed $A \subseteq \{1, 2, \ldots, n\}$.

4.1 Notational Switch

We will make a change in notation that keeps the properties of linear functions mentioned above, but will be easier to work with later on. The notational switch is essentially mapping from $(0, 1, +)$ to $(1, -1, \times)$. This mapping preserves correctness as the underlying group structure is preserved. Under this new notation, the definition of linearity becomes the following:

$+$	0	1
0	0	1
1	1	0

$\times$	1	-1
1	1	-1
-1	-1	1

Definition 4. *A function $\{\pm 1\}^n \to \{\pm 1\}$ is linear if and only if $f(x)f(y) = f(x \odot y)$, where $\odot$ denotes element-wise multiplication. All linear functions can be written in the form $f_S(x) = \prod_{i \in S} x_i$ for some $S \subseteq \{1, 2, \ldots, n\}$.*

Recall that one iteration of the linearity test accepts if and only if $f(x)f(y)f(x \odot y) = 1$ under this new notation, so we can write:

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

Using this fact, we can construct an indicator variable for whether the linearity test rejects or not:

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

Then the rejection probability can be written as the expectation of this indicator variable:

$$\begin{aligned} \delta_f &= \Pr_{x,y}[f(x)f(y) \neq f(x \odot y)] \\ &= \mathbf{E}_{x,y}\left[\frac{1 - f(x)f(y)f(x \odot y)}{2}\right] \end{aligned}$$

This will be useful later.
We would also like a nice basis to describe all functions $f : \{\pm 1\}^n \to \{\pm 1\}$. A simple example is to use indicator functions:

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

Then $f(x) = \sum_a f(a) \cdot e_a(x)$.
Next time: All linear functions form a good orthonormal basis for all such functions f.