Lecture 22: Linearity Testing
In general, we consider boolean functions to be functions that map \( \{0, 1\}^n \rightarrow \{0, 1\} \).

For ease of analysis we will interpret boolean functions to be \( \{0, 1\}^n \rightarrow \{+1, -1\} \).

In particular, the boolean output 0 will correspond to the output +1, and the boolean output 1 will correspond to the output −1.

For example, \( S \cdot x \) is a linear function that takes \( x \) as input and outputs 0/1 as output. In our context, the function \( \chi_S \) corresponds to this linear function.
Blum-Luby-Rubinfeld Linearity Testing

Given oracle access to a function $f$ ascertain whether the function is close to linear or not

$\text{BLR}^f$:  
- Sample $a, b \sim \mathbb{U}_{\{0,1\}^n}$
- Evaluate $x = f(a)$, $y = f(b)$, and $z = f(a + b)$
- Return $(x \cdot y == z)$

We want to prove: “The probability that this algorithm outputs true is high” if and only if “$f$ is close to linear”
Let \( f, g : \{0, 1\}^n \rightarrow \{+1, -1\} \) be two boolean functions

The probability that \( f \) and \( g \) agree is

\[
P \left[ f(\mathbb{U}_{\{0,1\}^n}) = g(\mathbb{U}_{\{0,1\}^n}) \right]
\]

Note that

\[
P \left[ f(\mathbb{U}_{\{0,1\}^n}) = g(\mathbb{U}_{\{0,1\}^n}) \right] = 1 - \varepsilon \text{ if and only if } \langle f, g \rangle = 1 - 2\varepsilon
\]

And recall that \( \langle f, g \rangle = \mathbb{E} [ f(\mathbb{U}) g(\mathbb{U}) ] \)
The claim

“The probability that this algorithm outputs true is high” if and only if “\( f \) is close to linear”

Is equivalent to proving

\[
\mathbb{E} \left[ f(A)f(B)f(A+B) \right] \text{ is high” if and only if “} f \text{ is close to linear”}
\]

Is equivalent to proving

\[
\mathbb{E} \left[ f(A)f(B)f(A+B) \right] \text{ is high” if and only if “There exists } S \in \{0, 1\}^n \text{ such that } \langle f, \chi_S \rangle \text{ is high”}
\]

Is equivalent to proving

\[
\mathbb{E} \left[ f(A)f(B)f(A+B) \right] \text{ is high” if and only if “} \max \hat{f}(S) \text{ is high”}
\]
A Simplification

\[
E \left[ f(A)f(B)f(A+B) \right] = \frac{1}{N^2} \sum_{a,b} f(a)f(b)f(a+b)
\]

Recall that \( f = \sum_S \hat{f}(S) \chi_S \). So, we have

\[
\frac{1}{N^2} \sum_{a,b} f(a)f(b)f(a+b)
\]

\[
= \frac{1}{N^2} \sum_{a,b} \left( \sum_R \hat{f}(R) \chi_R(a) \right) \left( \sum_S \hat{f}(S) \chi_S(b) \right) \left( \sum_T \hat{f}(T) \chi_T(a+b) \right)
\]

\[
= \frac{1}{N^2} \sum_{a,b,R,S,T} \hat{f}(R)\hat{f}(S)\hat{f}(T) \chi_R(a)\chi_S(b)\chi_T(a+b)
\]
A Simplification

\[
\frac{1}{N^2} \sum_{a,b,R,S,T} \hat{f}(R)\hat{f}(S)\hat{f}(T)\chi_R(a)\chi_S(b)\chi_T(a)\chi_T(b)
\]

\[
= \frac{1}{N^2} \sum_{a,R,S} \hat{f}(R)\hat{f}(S)\chi_R(a) \sum_T \hat{f}(T)\chi_T(a) \sum_b \chi_S(b)\chi_T(b)
\]

\[
= \frac{1}{N} \sum_{a,R,S} \hat{f}(R)\hat{f}(S)\chi_R(a) \sum_T \hat{f}(T)\chi_T(a) \langle \chi_S, \chi_T \rangle
\]

\[
= \frac{1}{N} \sum_{a,R,S} \hat{f}(R)\hat{f}(S)\chi_R(a) \hat{f}(S)\chi_S(a)
\]

\[
= \frac{1}{N} \sum_S \hat{f}(S)^2 \sum_R \hat{f}(R) \sum_a \chi_R(a)\chi_S(a)
\]

\[
= \sum_S \hat{f}(S)^2 \sum_R \hat{f}(R) \langle \chi_R, \chi_S \rangle = \sum_S \hat{f}(S)^3
\]
“$\mathbb{E} [f(A)f(B)f(A+B)]$ is high” if and only if “$\max \hat{f}(S)$ is high”

Is equivalent to proving

“$\sum_S \hat{f}(S)^3$ is high” if and only if “$\max \hat{f}(S)$ is high”

This result is false unless we add the constraint that $f$ is boolean
(Think: Why is this false?)
Since $f$ is boolean, we have $\langle f, f \rangle = 1$

By Parseval’s identity we get

$$1 = \sum_S \hat{f}(S)^2$$
Under the constraint
\[ \sum_S \hat{f}(S)^2 = 1 \]

We have
\[ \sum_s \hat{f}(S)^3 \approx 1 \] if and only if \[ \max \hat{f}(S) \approx 1 \]

Formalize and Prove this inequality
Given the promise that $f$ is close to a linear function, can we recover $S$ such that $f$ is close to $\chi_S$?