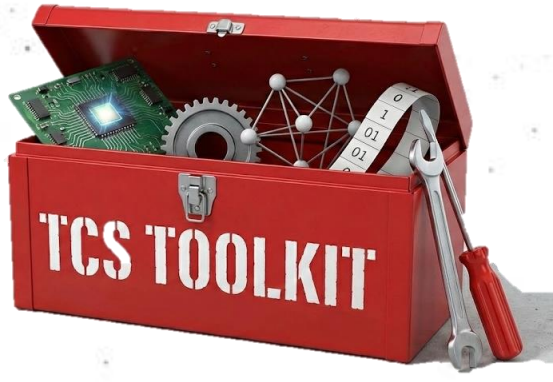
CS 58500 – Theoretical Computer Science Toolkit

Lecture 16 (04/07)

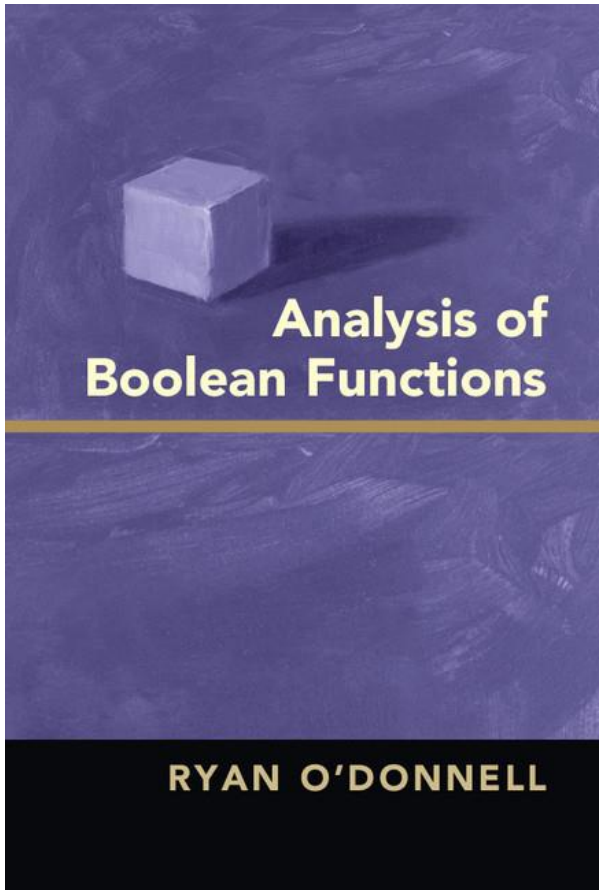
Boolean Function Analysis (I)



https://ruizhezhang.com/course_spring_2026.html

Today's Lecture

- Basic Definitions
- Examples
- Applications



Basic Definitions: Boolean hypercube

- Boolean hypercube $\{0,1\}^n$ equipped with some probability measure
- By default, we use the uniform measure over $\{0,1\}^n$
- Sometimes we need to study other measures on the Boolean hypercube, such as the p-biased measure
- Further the theory extends to finite product spaces:

$$\Omega = \Omega_1 \times \cdots \times \Omega_n, \qquad \mu = \mu_1 \times \cdots \times \mu_n$$

Basic Definitions: Boolean functions

We focus on the functions defined on the Boolean hypercube:

$$f:\{0,1\}^n \to \mathbb{R}$$

- Sometimes people use $\{-1,1\}^n$ $\quad b \in \{0,1\} \leftrightarrow (-1)^b \in \{-1,1\}$
- Let $N \coloneqq 2^n$ and $\{0,1\}^n \leftrightarrow \{0,1,\dots,N-1\}$
- $\{f:\{0,1\}^n \to \mathbb{R}\} \leftrightarrow \mathbb{R}^N$
- In the vector space $\mathbb{R}^N$, each standard basis $\boldsymbol{e}_i \in \mathbb{R}^N \leftrightarrow$ a Boolean function $\delta_i:\{0,1\}^n \to \mathbb{R}$:

$$\delta_i(x) = \begin{cases} 1 & \text{if } x = i \\ 0 & \text{if } x \neq i \end{cases} \qquad \forall x \in \{0,1\}^n$$

- Thus, for every f,

$$f(x) = f(0)\delta_0(x) + f(1)\delta_1(x) + \cdots + f(N-1)\delta_{N-1}(x)$$

Basic Definitions: Fourier Basis

For any $S \in \{0,1\}^n$, the Fourier basis is defined as:

$$\chi_S(x) := (-1)^{\langle S,x \rangle} = (-1)^{S_1 x_1 + \cdots + S_n x_n}$$

Equivalently, if we are working in $\{+1,-1\}^n$, for any $S \subset [n]$, the Fourier basis is defined as

$$\chi_S(x) := x^S = \prod_{i \in S} x_i$$

Lemma. For any $S \in \{0,1\}^n$,

$$\sum_{x \in \{0,1\}^n} \chi_S(x) = N \cdot \delta_S(0)$$

Proof.

$$\mathbb{E}[\chi_S(x)] = \prod_{i=1}^{n} \mathbb{E}[(-1)^{S_i x_i}] = \prod_{i=1}^{n} \frac{(1+(-1)^{S_i})}{2} = \delta_S(0)$$

■

Basic Definitions: Fourier Basis

We define the inner product of two Boolean functions $f, g: \{0,1\}^n \to \mathbb{R}$ as:

$$\langle f, g \rangle := \frac{1}{N} \sum_{x \in \{0,1\}^n} f(x)g(x) = \mathbb{E}_{x \sim \{0,1\}^n}[f(x)g(x)]$$

Lemma. For any $S, T \in \{0,1\}^n$,

$$\langle \chi_S, \chi_T \rangle = \delta_S(T)$$

Proof.

$$\chi_S(x)\chi_T(x) = \prod_{i=1}^{n} (-1)^{S_i x_i + T_i x_i} = \prod_{i: S_i \neq T_i} (-1)^{x_i}$$

$$\mathbb{E}[\chi_S \chi_T] = \prod_{i: S_i \neq T_i} \mathbb{E}[(-1)^{x_i}] = \delta_S(T)$$

■

Basic Definitions: Fourier Basis

- We have shown that

$$\{\chi_S : S \in \{0,1\}^n\}$$

  is an orthonormal basis w.r.t. $\langle \cdot, \cdot \rangle$

- Thus, for any $f: \{0,1\}^n \to \mathbb{R}$,

$$f(x) = \sum_{S \in \{0,1\}^n} \hat{f}(S)\chi_S(x) \qquad \forall x \in \{0,1\}^n$$

- The coefficients $\hat{f}: \{0,1\}^n \to \mathbb{R}$ is another Boolean function, called the Fourier transform of f
- Since the Fourier basis is orthonormal, we have

$$\hat{f}(S) = \langle f, \chi_S \rangle = \mathbb{E}_{x \sim \{0,1\}^n}[f(x)\chi_S(x)]$$

Basic Definitions: Fourier Transformation

$$\hat{f}(S) = \langle f, \chi_S \rangle = \mathbb{E}_{x \sim \{0,1\}^n}[f(x)\chi_S(x)]$$

- Linearity: for any $f, g \in \{0,1\}^n \to \mathbb{R}$,
$$\widehat{f+g} = \hat{f} + \hat{g}$$
- Scaling: for any $f \in \{0,1\}^n \to \mathbb{R}$, $c \in \mathbb{R}$,
$$\widehat{cf} = c\hat{f}$$
- Involution: for any $f \in \{0,1\}^n \to \mathbb{R}$,
$$\left(\hat{f}\right) = \frac{1}{N} f$$
- Expectation: for any $f \in \{0,1\}^n \to \mathbb{R}$,
$$\hat{f}(0) = \mathbb{E}[f]$$

Basic Definitions: Fourier Transformation

Theorem (Plancherel). Suppose $f, g \in \{0,1\}^n \to \mathbb{R}$. Then,

$$\langle f, g \rangle = \sum_{S \in \{0,1\}^n} \hat{f}(S)\hat{g}(S) = N\langle \hat{f}, \hat{g} \rangle$$

Proof.

- By the orthonormality of $\{\chi_S\}$,

$$\langle f, g \rangle = \left\langle \sum_S \hat{f}(S)\chi_S, \sum_T \hat{g}(S)\chi_T \right\rangle = \sum_{S,T} \hat{f}(S)\hat{g}(T)\langle \chi_S, \chi_T \rangle = \sum_S \hat{f}(S)\hat{g}(S)$$

■

Theorem (Parseval's Identity). Suppose $f \in \{0,1\}^n \to \mathbb{R}$. Then,

$$\mathbb{E}[f^2] = \langle f, f \rangle = \sum_{S \in \{0,1\}^n} \hat{f}(S)^2 = N\mathbb{E}\left[\hat{f}^2\right]$$

Basic Definitions: Fourier Transformation

Theorem (Plancherel). Suppose $f, g \in \{0,1\}^n \to \mathbb{R}$. Then,

$$\langle f, g \rangle = \sum_{S \in \{0,1\}^n} \hat{f}(S)\hat{g}(S) = N\langle \hat{f}, \hat{g} \rangle$$

Theorem (Parseval's Identity). Suppose $f \in \{0,1\}^n \to \mathbb{R}$. Then,

$$\mathbb{E}[f^2] = \langle f, f \rangle = \sum_{S \in \{0,1\}^n} \hat{f}(S)^2 = N\mathbb{E}[\hat{f}^2]$$

Corollary (Parseval's Identity). Suppose $f \in \{0,1\}^n \to \{+1, -1\}$. Then,

$$\sum_{S \in \{0,1\}^n} \hat{f}(S)^2 = 1$$

Basic Definitions: Fourier Transformation

Theorem (Plancherel**).** Suppose $f, g \in \{0,1\}^n \to \mathbb{R}$. Then,

$$\langle f, g \rangle = \sum_{S \in \{0,1\}^n} \hat{f}(S)\hat{g}(S) = N\langle \hat{f}, \hat{g} \rangle$$

Theorem (Parseval's Identity**).** Suppose $f \in \{0,1\}^n \to \mathbb{R}$. Then,

$$\mathbb{E}[f^2] = \langle f, f \rangle = \sum_{S \in \{0,1\}^n} \hat{f}(S)^2 = N\mathbb{E}[\hat{f}^2]$$

- Variance:

$$\mathrm{Var}[f] = \mathbb{E}[f^2] - \mathbb{E}[f]^2 = \sum_{S \neq 0} \hat{f}(S)^2$$

Basic Definitions: Additive Homomorphism

Let $f: \{0,1\}^n \to \mathbb{R}$. We say that f exhibits additive homomorphism if, for all $x, y \in \{0,1\}^n$,

$$f(x+y) = f(x)f(y)$$

Here, $x \pm y$ is over $\mathbb{F}_2^n$

- The Fourier basis χ_S satisfies additive homomorphism:

$$\chi_S(x+y) = \prod_{i=1}^{n}(-1)^{S_i(x_i+y_i)} = \prod_{i=1}^{n}(-1)^{S_i x_i}(-1)^{S_i y_i} = \chi_S(x)\chi_S(y)$$

- Shifting: for any $f \in \{0,1\}^n \to \mathbb{R}$, let $f_y(x) \coloneqq f(x-y)$ for any $y \in \{0,1\}^n$. Then

$$\widehat{f_y}(S) = \hat{f}(S)\chi_S(y) = \hat{f}(S)\chi_y(S)$$

$$\begin{aligned}\widehat{f_y}(S) &= \mathbb{E}[f_y\chi_S] = \mathbb{E}_x[f(x-y)\chi_S(x)] = \mathbb{E}_z[f(z)\chi_S(z+y)] \\ &= \mathbb{E}_z[f(z)\chi_S(z)\chi_S(y)] = \hat{f}(S)\chi_S(y)\end{aligned}$$

Basic Definitions: Convolution

Let $f, g: \{0,1\}^n \to \mathbb{R}$. The convolution of f and g is the function $(f \star g): \{0,1\}^n \to \mathbb{R}$ defined as:

$$(f \star g)(x) \coloneqq \frac{1}{N} \sum_{y \in \{0,1\}^n} f(y) g(x-y) = \frac{1}{N} \sum_{y \in \{0,1\}^n} f(y) g(x+y)$$

Lemma. For any $f, g \in \{0,1\}^n \to \mathbb{R}$, we have

$$\widehat{(f \star g)}(S) = \hat{f}(S)\hat{g}(S) \qquad \forall S \in \{0,1\}^n$$

Proof.

- Note that $(f \star g)(x) = \mathbb{E}[f g_x]$

$$\widehat{(f \star g)}(S) = \mathbb{E}_x[(f \star g)(x)\chi_S(x)] = \mathbb{E}_x\left[\mathbb{E}_y[f(y)g_x(y)]\chi_S(x)\right] = \mathbb{E}_y\left[f(y)\mathbb{E}_x[g_x(y)\chi_S(x)]\right]$$

- $\mathbb{E}_x[g_x(y)\chi_S(x)] = \mathbb{E}_x\left[g_y(x)\chi_S(x)\right] = \widehat{g_y}(S) = \hat{g}(S)\chi_S(y)$

$$\widehat{(f \star g)}(S) = \mathbb{E}_y[f(y)\hat{g}(S)\chi_S(y)] = \hat{f}(S)\hat{g}(S)$$

■

Basic Definitions: Convolution

Let $f, g: \{0,1\}^n \to \mathbb{R}$. The convolution of f and g is the function $(f \star g): \{0,1\}^n \to \mathbb{R}$ defined as:

$$(f \star g)(x) \coloneqq \frac{1}{N} \sum_{y \in \{0,1\}^n} f(y) g(x-y) = \frac{1}{N} \sum_{y \in \{0,1\}^n} f(y) g(x+y)$$

Lemma. For any $f, g \in \{0,1\}^n \to \mathbb{R}$, we have

$$\widehat{(f \star g)}(S) = \hat{f}(S)\hat{g}(S) \qquad \forall S \in \{0,1\}^n$$

- Convolution $\longleftrightarrow$ Multiplication:

$$\widehat{(f \star g)} = \hat{f}\hat{g}$$

- $\widehat{f_y} = \hat{f}\chi_y = \hat{f}\left(N\widehat{\chi_y}\right) = N\left(\widehat{f \star \chi_y}\right) \implies f_y = N\left(f \star \widehat{\chi_y}\right)$

Basic Definitions: Convolution

Let $f, g: \{0,1\}^n \to \mathbb{R}$. The convolution of f and g is the function $(f \star g): \{0,1\}^n \to \mathbb{R}$ defined as:

$$(f \star g)(x) \coloneqq \frac{1}{N} \sum_{y \in \{0,1\}^n} f(y) g(x-y) = \frac{1}{N} \sum_{y \in \{0,1\}^n} f(y) g(x+y)$$

Probability meaning:

- Let X be a random variable over $\{0,1\}^n$ with probability density ϕ w.r.t. the uniform measure, i.e.,

$$\mathbb{E}[\phi] = 1, \qquad \Pr[X = x] = \frac{\phi(x)}{N}$$

- Let Y be a random variable over $\{0,1\}^n$ with probability density μ
- Then, the density of $X \pm Y$ is $(\phi \star \mu)$

Today’s Lecture

- Basic Definitions
- Examples
- Applications

Examples

Indicator function

For any $a \in \{0,1\}^n$,

$$\delta_a(x) = \begin{cases} 1 & \text{if } x = a \\ 0 & \text{if } x \neq a \end{cases} \qquad \forall x \in \{0,1\}^n$$

- For any $S \subseteq \{0,1\}^n$,

$$\widehat{\delta_a}(S) = \langle \delta_a, \chi_S \rangle = \frac{1}{N}\chi_S(a) = \frac{1}{N}(-1)^{\langle S,a \rangle}$$

- Uncertainty principle: For any nonzero function f,

$$|\mathrm{supp}(f)| \cdot \left|\mathrm{supp}\left(\hat{f}\right)\right| \geq N$$

Examples

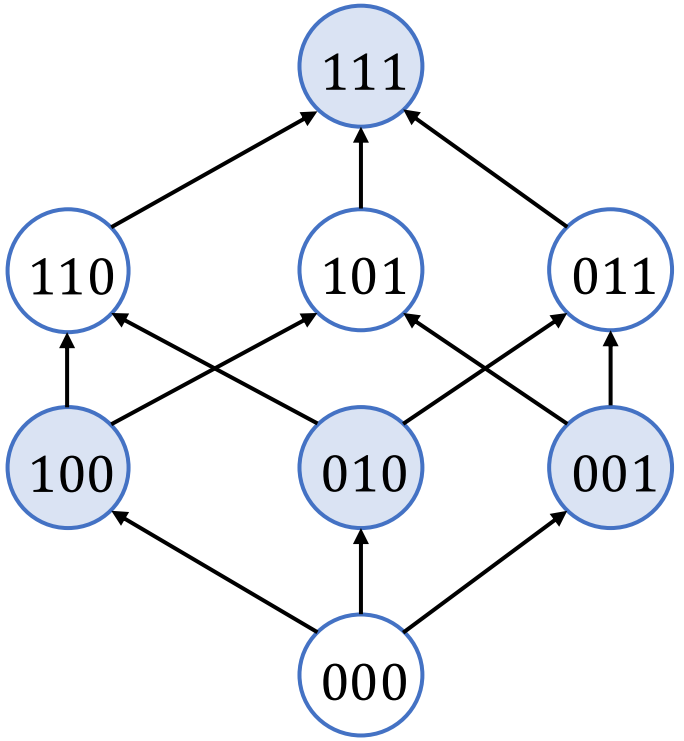
Majority function

x	Maj_3
000	1
001	1
010	1
011	−1
100	1
101	−1
110	−1
111	−1

S	$\widehat{\mathrm{Maj}_3}$
000	0
001	1/2
010	1/2
011	0
100	1/2
101	0
110	0
111	−1/2

Fourier weight of degree k:

$$\mathbf{W}^k[f] \coloneqq \sum_{|S|=k} \hat{f}(S)^2$$



$\mathbf{W}^3[f] = 1/4$

$\mathbf{W}^1[f] = 3/4$

Today’s Lecture

- Basic Definitions
- Examples
- Applications

Applications: Linearity Testing

A function $f:\{0,1\}^n \to \{-1,1\}$ is said to be linear if for any $x, y \in \{0,1\}^n$, it holds that

$$f(x)f(y) = f(x+y)$$

- This is equivalent to the additive homomorphism
- More intuitively, if we define $f:\mathbb{F}_2^n \to \mathbb{F}_2$, then f if linear $\Leftrightarrow f(x)+f(y)=f(x+y)$
- We have already shown that Fourier basis χ_S is linear
- Are there any "inherently different linear functions"?

Another related question is the following. Suppose $f:\{0,1\}^n \to \{-1,1\}$ is approximately linear, i.e., $f(x)f(y) = f(x+y)$ holds for $1-\epsilon$ fraction of the pairs x, y

- Does that tell us anything about the structure of the function f? (You are encouraged to think of this question at home)

Applications: Sparse functions

A function $f: \{0,1\}^n \to \mathbb{R}$ is said to be k-Fourier sparse if $\left|\mathrm{supp}\left(\hat{f}\right)\right| \leq k$

A function $g: \{0,1\}^n \to \{-1,1\}$ is said to be (k, ε)-Fourier sparse if there is a k-Fourier sparse function f such that $\|f - g\|_2 \leq \varepsilon$

- How can one test whether a function is Fourier sparse?
  - (Ghosh-Roy, NeurIPS '25; Ghosh-Maitra-Roy, ICLR '26)
- Can one learn such functions (i.e., find an approximator) for such functions efficiently, given query access to them?
  - Goldreich-Levin, Kushilevitz-Mansour
  - More generally, compressed sensing, sparse Fourier transform

Applications: Junta testing

An important subclass of sparse functions is the class of juntas. A function $f:\{0,1\}^n \to \mathbb{R}$ is said to be a t-junta if there is $T \subseteq [n]$ of size at most t, and $g:\{0,1\}^T \to \mathbb{R}$, such that

$$f(x) = g(x_T)$$

- Can one test juntas more efficiently?
- Learn?

Applications: Social Choice and Arrow's Theorem

Majority is a 2-candidate social choice functions

	✓		✓	✓	
		✓			✓

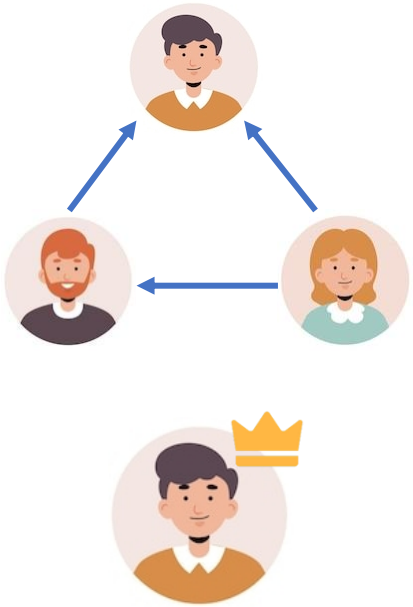
$$\mathrm{Maj}_5(01001) = 0$$

- What if there are 3 candidates?
- Condorcet suggested using the voters' preferences to conduct the three possible pairwise elections, a vs. b, b vs. c, and c vs. a

Nicolas de Condorcet (1743-1794)

Applications: Social Choice and Arrow's Theorem

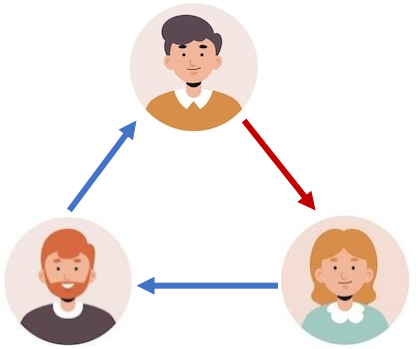
		Societal Aggregation
 (0) vs.  (1)	$0 \quad 0 \quad 1 \quad \cdots = x$	$f(x)$ 
 (0) vs.  (1)	$0 \quad 1 \quad 0 \quad \cdots = y$	$f(y)$ 
 (0) vs.  (1)	$1 \quad 1 \quad 0 \quad \cdots = z$	$f(z)$ 



- There are $3! = 6$ possible rankings $\{001,010,100,011,110,101\} = \{x : \mathrm{NAE}_3(x)\}$
- In an election employing Condorcet's method with voting rule $f: \{0,1\}^n \to \{0,1\}$, we say that a candidate is a Condorcet winner if it wins all of the pairwise elections in which it participates

Applications: Social Choice and Arrow's Theorem

		Societal Aggregation
(0) vs. (1)	$0 \quad \boxed{0} \quad 1 \quad \cdots = x$	$f(x)$
(0) vs. (1)	$0 \quad \boxed{1} \quad 0 \quad \cdots = y$	$f(y)$
(0) vs. (1)	$1 \quad \boxed{0} \quad 0 \quad \cdots = z$	$f(z)$



Condorcet's Paradox

- There are $3! = 6$ possible rankings $\{001,010,100,011,110,101\} = \{x : \mathrm{NAE}_3(x)\}$
- In an election employing Condorcet's method with voting rule $f: \{0,1\}^n \to \{0,1\}$, we say that a candidate is a Condorcet winner if it wins all of the pairwise elections in which it participates

Applications: Social Choice and Arrow's Theorem

Theorem (Arrow's Impossibility Theorem**).** Suppose $f\colon \{0,1\}^n \to \{0,1\}$ is a unanimous voting rule used in a 3-candidate Condorcet election. If there is always a Condorcet winner, then f must be a dictatorship (1-junta)

- Unanimous: $f(0,\cdots,0) = 0$ and $f(1,\cdots,1) = 1$
- In 2002, Gil Kalai gave a Fourier-based proof

Kenneth Arrow, Nobel Laureate (1921-2017)