CAS CS 599 B: Mathematical Methods for TCS

Lecturer: Mark Bun

Spring 2022

Lecture Notes 1:

Basics of Boolean Fourier Analysis

Reading.

- O'Donnell, Analysis of Boolean Functions §1.1-1.4

1 Course information

The purpose of the class is to study a handful of mathematical techniques that appear all the time in research in algorithm design and complexity theory. The tentative list of course units is:

1. Analysis of Boolean functions
2. Pseudorandomness
3. Spectral graph theory
4. Codes, information, and communication
5. Ramsey theory, extremal and additive combinatorics
6. Linear and semidefinite programming

This class is meant to satisfy the algorithms/theory depth requirement, so there's a lower bound on your deliverables. The course components are 1) Weekly exercise sets. We'll try to find a convenient time to reserve a room for you to break into small groups to work on these. 2) More thinking-intensive problem sets (about 4-5 throughout the semester). 3) A course project. 4) Class participation.

Things you should do are: Sign up for Piazza using the code [redacted]. Be on the lookout for a poll for office hours and for the weekly exercise session.

2 Boolean functions

A Boolean function is a mapping $f : \{-1, 1\}^n \to \{-1, 1\}$. A typical convention is to interpret $+1$ as logical false and -1 as logical true. This isn't essential, but as we'll see it's generally useful for doing and interpreting calculations. It's also often helpful to consider the more general class of functions with codomain $\mathbb{R}$.

A simple example of a Boolean function is the 3-bit majority function $\mathrm{MAJ}_3 : \{-1, 1\}^3 \to \{-1, 1\}$ defined by $\mathrm{MAJ}_3(x_1, x_2, x_3) = 1$ iff at least two of the input bits are equal to 1. Our study of Boolean functions is guided by two principles.

Principle 1: Boolean functions are everywhere.

- In circuit complexity, f captures the truth table of a circuit with n inputs and 1 output.
- In machine learning, f is a classification rule that takes n binary features and outputs a true/false label.
- In social choice theory, f is a voting rule taking n votes for two possible candidates to an election outcome.
- In combinatorics, f is the indicator of a subset $S = \{x \in \{-1,1\}^n \mid f(x) = -1\}$ of the Boolean hypercube. It can also be thought of as the indicator for a "set system" $\mathcal{F}$ of subsets of $[n] = \{1, \dots, n\}$, defined by $X \in \mathcal{F} \iff f(x) = -1$, where x is the indicator vector for X.

Principle 2: Every Boolean function is a polynomial. We'll start with an example to convince ourselves why this is true. Write

$$\mathrm{MAJ}_3(x) = \sum_{v \in \{-1,1\}^3} \mathrm{MAJ}_3(v) \cdot \mathbf{1}_{\{v\}}(x)$$

where $\mathbf{1}_{\{v\}}(x) = 1$ if $x = v$ and $\mathbf{1}_{\{v\}}(x) = 0$ if $x \neq v$. For every v, the indicator function $\mathbf{1}_{\{v\}}(x)$ is a polynomial in x:

$$\mathbf{1}_{\{v\}}(x) = \left(\frac{1+v_1x_1}{2}\right) \cdot \left(\frac{1+v_2x_2}{2}\right) \cdot \left(\frac{1+v_3x_3}{2}\right).$$

So MAJ_3 is a sum of polynomials, hence itself a polynomial. Somewhat more explicitly:

$$\begin{aligned}\mathrm{MAJ}_3(x) = (1) &\cdot \left(\frac{1+x_1}{2}\right) \cdot \left(\frac{1+x_2}{2}\right) \cdot \left(\frac{1+x_3}{2}\right) \\ + (1) &\cdot \left(\frac{1+x_1}{2}\right) \cdot \left(\frac{1+x_2}{2}\right) \cdot \left(\frac{1-x_3}{2}\right) \\ + (-1) &\cdot \left(\frac{1+x_1}{2}\right) \cdot \left(\frac{1-x_2}{2}\right) \cdot \left(\frac{1-x_3}{2}\right) \\ + \dots\end{aligned}$$

It's a bit tedious, but if you expand this out and collect terms, you get $\mathrm{MAJ}_3(x) = \frac{1}{2}(x_1 + x_2 + x_3 - x_1x_2x_3)$. We can generalize this construction to any Boolean function by writing

$$f(x) = \sum_{v \in \{-1,1\}^n} f(v) \cdot \mathbf{1}_{\{v\}}(x).$$

Some remarks are in order. First, this polynomial is *multilinear*, meaning each individual appears with degree at most 1, i.e., there are no terms like x_1^2, x_1^3, etc. This implies that the total degree of the polynomial is n. This property is actually without loss of generality, since on domain $\{-1, 1\}$ we can always replace any factor x_i^2 with 1.

Second, there is nothing special about taking the codomain of f to be $\{-1, 1\}$. This construction works just as well for any function $f : \{-1,1\}^n \to \mathbb{R}$. We are thus most of the way toward proving the following:

Theorem 1. *Every function $f : \{-1,1\}^n \to \mathbb{R}$ has a unique representation as a multilinear polynomial*

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

where $\chi_S(x) = \prod_{i \in S} x_i$.

Here we are using the convention that $\chi_\emptyset(x) = 1$. The quantities $\hat{f}(S)$ are called the "Fourier coefficients" of f, and the collection $(\hat{f}(S))_{S\subseteq[n]}$ is called the "Fourier transform" of f.

Example 2. The Fourier coefficients of MAJ_3 are $\widehat{\mathrm{MAJ}_3}(\{1\}) = \widehat{\mathrm{MAJ}_3}(\{2\}) = \widehat{\mathrm{MAJ}_3}(\{3\}) = 1/2, \widehat{\mathrm{MAJ}_3}(\{1,2,3\}) = -1/2$, and $\widehat{\mathrm{MAJ}_3}(S) = 0$ otherwise.

Example 3. Define the function $\mathrm{XOR}_n(x) = -1$ iff an odd number of the inputs $x_1, \ldots, x_n$ are equal to -1. The Fourier coefficients of XOR_n are $\widehat{\mathrm{XOR}_n}([n]) = 1$ and $\widehat{\mathrm{XOR}_n}(S) = 0$ otherwise.

Thus, the way to think about each monomial $\chi_S(x)$ is as the parity of the subset of bits indexed by S.

3 Linear algebra of the Fourier representation

Some basic questions to ask about the Fourier representation are: What does it mean? How do we compute it? What is it good for? We'll start developing some tools for understanding it now. The basic perspective we'll take is that the set of functions $\mathcal{F}_n = \{f : \{-1,1\}^n \to \mathbb{R}\}$ is a 2^n-dimensional real vector space. We can equip this vector space with the following inner product.

Definition 4. For functions $f, g : \{-1,1\}^n \to \mathbb{R}$, define the *inner product*

$$\langle f, g\rangle = 2^{-n} \sum_{x\in\{-1,1\}^n} f(x)g(x) = \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n} [f(x)g(x)] .$$

You can check that this satisfies the definition of a real inner product (symmetry, bilinearity, positive definiteness). The inner product has a natural interpretation as the *correlation* or *average agreement* between f and g. If f and g are Boolean functions that are perfectly correlated ($f = g$), then $\langle f, g\rangle = 1$. If they are perfectly anti-correlated ($f = -g$), then $\langle f, g\rangle = -1$.

Lemma 5. *The parity functions satisfy*

$$\langle \chi_S, \chi_T\rangle = \begin{cases} 1 & \text{if } S = T \\ 0 & \text{if } S \neq T. \end{cases}$$

Thus, the set of 2^n parity functions $\{\chi_S(x) \mid S \subseteq [n]\}$ form an *orthonormal basis* for the vector space $\mathcal{F}_n$. In particular, that means they are linearly independent. This implies the "uniqueness" part of Theorem 1.

Proof. We calculate

$$\begin{aligned} \langle \chi_S, \chi_T\rangle &= \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n} [\chi_S(x)\chi_T(x)] \\ &= \mathop{\mathbb{E}}_{x\sim\{-1,1\}^n} [\chi_{S\Delta T}(x)] \end{aligned}$$

We now consider two cases. If $S = T$, so $S\Delta T = \emptyset$, the quantity under the expectation is the constant 1. If $S \neq T$, then by independence we have

$$\mathop{\mathbb{E}}_{x\sim\{-1,1\}^n} [\chi_{S\Delta T}(x)] = \prod_{i\in S\Delta T} \mathop{\mathbb{E}}_{x_i\sim\{-1,1\}} [x_i] = 0.$$

□

If you like to do linear algebra using matrices, then you can think of the Fourier representation of a function f as

$$\begin{pmatrix} f(+,+,\dots,+) \\ f(-,+,\dots,+) \\ \vdots \\ f(-,-\dots,-) \end{pmatrix} = \underbrace{\begin{pmatrix} \chi_\emptyset(+,+,\dots,+) & \chi_{\{1\}}(+,+,\dots,+) & \dots & \chi_{[n]}(+,+,\dots,+) \\ \chi_\emptyset(-,+,\dots,+) & \chi_{\{1\}}(-,+,\dots,+) & \dots & \chi_{[n]}(-,+,\dots,+) \\ & \vdots & & \\ \chi_\emptyset(-,-,\dots,-) & \chi_{\{1\}}(-,-,\dots,-) & \dots & \chi_{[n]}(-,-,\dots,-) \end{pmatrix}}_{H[x,S]} \begin{pmatrix} \hat{f}(\emptyset) \\ \hat{f}(\{1\}) \\ \vdots \\ \hat{f}([n]) \end{pmatrix}$$

The matrix $H[x,S]$ is the "Hadamard matrix" of order 2^n. The Fourier transform over the Boolean hypercube is sometimes also called the "Walsh-Hadamard transform."

Taking this perspective has the useful consequence that it tells us how to compute Fourier coefficients. Lemma 5 implies that $HH^T = 2^n I_{2^n}$, and so $H^{-1} = 2^{-n}H^T$. Unpacking this, we get

Lemma 6. *For every function $f : \{-1,1\}^n \to \mathbb{R}$ and every $S \subseteq [n]$,*

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

You can also just directly verify this expression by using the Fourier expansion of f and orthonormality to compute $\langle f, \chi_S \rangle$ and show that it's equal to $\hat{f}(S)$.

Orthonormality of the parity functions lets us prove several other elegant properties.

Proposition 7 (Plancharel's Identity). *For any $f, g : \{-1,1\}^n \to \mathbb{R}$,*

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

Proof.

$$\begin{aligned} \langle f, g \rangle &= \Big\langle \sum_{S \subseteq [n]} \hat{f}(S)\chi_S, \sum_{T \subseteq [n]} \hat{g}(T)\chi_T \Big\rangle \\ &= \sum_{S,T \subseteq [n]} \hat{f}(S)\hat{g}(T)\langle \chi_S, \chi_T \rangle && \text{by linearity} \\ &= \sum_{S \subseteq [n]} \hat{f}(S)\hat{g}(S) && \text{by orthonormality.} \end{aligned}$$

□

The special case of Plancharel where $g = f$ gives us

Proposition 8 (Parseval's Theorem). *For any $f : \{-1,1\}^n \to \mathbb{R}$,*

$$\mathop{\mathbb{E}}_{x \sim \{-1,1\}^n} \left[f(x)^2\right] = \langle f, f \rangle = \sum_{S \subseteq [n]} \hat{f}(S)^2.$$

In the case where f is Boolean, this shows that the sum of squared Fourier coefficients is equal to 1.