18.218 Topics in Combinatorics Spring 2021 – Lecture 1

Dor Minzer

In this course, we will mostly be studying Boolean functions over the hypercube, i.e. $f\colon \{0,1\}^n \to \{0,1\}$. Our primary (but not only) tool will be, what is often called, discrete Fourier analysis. Here and throughout, we will think of the Boolean hypercube as being equipped with some probability measure, which for the majority of the time will just be the uniform distribution.

Having said that, the theory we will present often generalizes to other measures on the Boolean hypercube, such as the p-biased measure. Further the theory often extends to finite product spaces, i.e. spaces of the form $(\Omega = \Omega_1 \times \ldots \times \Omega_n, \mu = \mu_1 \times \ldots \mu_n)$, and in recent years beyond that (we may discuss this towards the end of the course).

1 Course overview

Since this is the first time this course is ran, the exact material we cover is yet to be determined. The plan is to touch the following topics.

1. **Fundamentals of discrete Fourier analysis.** We will begin the course by presenting basic definitions and notions, such as the Fourier decomposition and influences of variables. Throughout the course, we will present basic tools and results in the area, such as the hypercontractive inequality, the KKL theorem, Junta theorems by Friedgut and Bourgain, sharp thresholds and the invariance principle.

2. **Applications of Fourier analysis in various areas in TCS.** We will discuss several applications of Fourier analysis in areas such as property testing and learning theory.

3. **Applications of Fourier analysis in Hardness of Approximation.** We will briefly discuss a prominent outstanding conjecture in theoretical computer science known as the Unique-Games Conjecture. We will use some basic results from Fourier analysis to show several consequences of this conjecture, such as the (conditional) optimality of the Goemans-Williamson algorithm for Max-Cut, and (conditional) hardness of the Vertex-Cover problem.

4. **Applications of Fourier analysis in Extremal Combinatorics.** Extremal Combinatorics is an area, which roughly speaking asks how large a collection of specific objects can be if it satisfies a certain constraint. For example, how large can a collection of subsets of $[n]$ be provided any two intersect non-trivially? We will see how results in analysis play an important role in giving detailed answers to some of these questions.

5. **Advanced Topics.** Towards the end of the course we will discuss more advanced topics, which are advances in the area achieved only recently. These includes the resolution of the sensitivity conjecture, an extension of the hypercontractive inequality referred to in the literature as "global hypercontractivity", and the Fourier entropy conjecture.

Without further ado, let's get down to business.

2 The basic set-up

2.1 The Fourier basis

We will think of the domain $\{0,1\}^n$ as the additive group modulo 2. Often times, it will be notationally convenient for us to turn this "addition" operation into a product, by the identification $b \to (-1)^b$ from $\{0,1\}$ to $\{1,-1\}$, and thus we will think of functions $f\colon \{-1,1\}^n \to \{-1,1\}$.

We note that the collection of functions $f\colon \{-1,1\}^n \to \mathbb{R}$ forms a linear space over $\mathbb{R}$, and we next introduce an inner product operation over it. This inner product is simply the L^2 inner product: for any $f, g\colon \{-1,1\}^n \to \mathbb{R}$ we define

$$\langle f, g\rangle = \mathop{\mathbb{E}}_{x}[f(x)g(x)].$$

Here and throughout, unless stated otherwise, the distribution over x is uniform over $\{-1,1\}^n$.

Now that we have an inner product, is makes sense to come up with an orthonormal basis with respect to it, which is often a useful tool from linear algebra that helps us understand vector spaces better. In this particular case there is a very nice basis, given by the characters of the additive group. Formally, for each $S \subseteq [n]$ we define a function $\chi_S\colon \{-1,1\}^n \to \{-1,1\}$ as

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

Claim 2.1. *The collection $\{\chi_S\}_{S\subseteq[n]}$ forms an orthonormal set (In particular, it is a linearly independent set).*

Proof. We start with two simple observations. Let $S, T \subseteq [n]$ be any two subsets.

1. $\chi_S(x)\chi_T(x) = \chi_{S\Delta T}(x)$, where $\Delta = (S \setminus T) \cup (T \setminus S)$ is the symmetric difference between S and T. Indeed,

$$\chi_S(x)\chi_T(x) = \prod_{i\in S\setminus T} x_i \prod_{i\in S\cap T} x_i \prod_{i\in T\setminus S} x_i \prod_{i\in S\cap T} x_i = \prod_{i\in S\setminus T\cup T\setminus S} x_i \left(\prod_{i\in S\cap T} x_i\right)^2 = \chi_{S\Delta T}(x).$$

2. If $S \subseteq [n]$ is non-empty, then $\mathbb{E}_x[\chi_S(x)] = 0$. Indeed, fix $i \in S$, then writing $S = Q \cup \{i\}$ and $x = (y, x_i)$ we have

$$\mathop{\mathbb{E}}_{x}[\chi_S(x)] = \mathop{\mathbb{E}}_{y,x_i}[\chi_Q(y)x_i] = \frac{1}{2}\mathop{\mathbb{E}}_{y}[\chi_Q(y) - \chi_Q(y)] = 0.$$

The proof is now concluded by noting that if $S \neq T$, then $\langle \chi_S, \chi_T\rangle = \mathbb{E}_x[\chi_S(x)\chi_T(x)] = \mathbb{E}_x[\chi_{S\Delta T}(x)] = 0$, and if $S = T$ then this is 1. $\square$

Now from dimension considerations, it follows that the collection $\{\chi_S\}_{S\subseteq[n]}$ is in fact a basis for the space of real-valued functions over $\{-1,1\}^n$, and thus any function $f\colon \{-1,1\}^n \to \mathbb{R}$ can be written as a linear combination of it. The standard notation for this is

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

where $\widehat{f}(S)$ are called the Fourier coefficients of f. Moreover, since the basis we used is orthonormal, there is a simple formula for each Fourier coefficient, namely $\widehat{f}(S) = \langle f, \chi_S\rangle$.

Claim 2.2. *The following holds for any $f, g\colon \{-1,1\}^n \to \mathbb{R}$:*

1. *Plancherel's equality:* $\langle f, g\rangle = \sum_{S\subseteq[n]} \widehat{f}(S)\widehat{g}(S)$.
2. *Parseval's equality:* $\|f\|_2^2 = \sum_{S\subseteq[n]} \widehat{f}(S)^2$.

Proof. For the first item, we use the bi-linearity of the inner product

$$\langle f, g\rangle = \left\langle \sum_{S\subseteq[n]} \widehat{f}(S)\chi_S, \sum_{T\subseteq[n]} \widehat{g}(T)\chi_T \right\rangle = \sum_{S,T\subseteq[n]} \widehat{f}(S)\widehat{g}(T)\left\langle \chi_S, \chi_T\right\rangle = \sum_{S,T\subseteq[n]} \widehat{f}(S)\widehat{g}(T)1_{S=T}$$
$$= \sum_{S\subseteq[n]} \widehat{f}(S)\widehat{g}(S).$$

The second item is just an instantiation of the first one with $f = g$, using $\|f\|_2^2 = \langle f, f\rangle$. $\square$

Next, we define the mean and the variance of a Boolean function. To do that, we think of $f(x)$ as a random variable, which is sampled by taking $x \in_R \{-1,1\}^n$, and then evaluating $f(x)$. The mean of f is the mean of this random variable, and it denoted by

$$\mathbb{E}\left[f\right] = \mathop{\mathbb{E}}_{x}\left[f(x)\right] = \widehat{f}(\emptyset).$$

The variance of f, denoted by $\mathsf{var}(f)$, is the variance of this random variable, i.e.

$$\mathsf{var}(f) = \mathop{\mathbb{E}}_{x}\left[\left(f(x) - \mathbb{E}\left[f\right]\right)^2\right].$$

Claim 2.3. $\mathsf{var}(f) = \sum_{S\neq\emptyset} \widehat{f}(S)^2$.

Proof. Consider the function $g(x) = f(x) - \mathbb{E}\left[f\right]$, and note that for any $S \neq \emptyset$, $\widehat{g}(S) = \widehat{f}(S)$, and $\widehat{g}(\emptyset) = 0$. Thus, by Parseval

$$\mathsf{var}(f) = \|g\|_2^2 = \sum_{S\neq\emptyset} \widehat{f}(S)^2.$$ $\square$

2.2 Property testing

We mention some applications of the material we have seen so far that we may see in the subsequent lectures.

2.2.1 Linearity testing

A function $f\colon \{-1,1\}^n \to \{-1,1\}$ is said to be linear if for any $x, y \in \{-1,1\}^n$ it holds that $f(x)f(y) = f(xy)$, where $(xy)_i = x_i y_i$. We have already seen a good collection of linear functions, namely the Fourier characters (check that!). Are there any "inherently different linear functions"?

Another related question is the following. Suppose $f\colon \{-1,1\}^n \to \{-1,1\}$ is approximately satisfies the definition of a linear function, i.e. $f(x)f(y) = f(xy)$ holds for $1-\varepsilon$ 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.

Next time, we will discuss a more challenging version of this question, and see how the power of Fourier analysis gives a very elegant solution to this problem.

2.2.2 Sparse functions

A function $f\colon \{-1,1\}^n \to \mathbb{R}$ is said to be t-Fourier sparse if the support size of its Fourier spectrum has size at most t. A function $g\colon \{-1,1\}^n \to \{-1,1\}$ is said to be (t,ε)-Fourier sparse if there is a t-Fourier sparse function f such that $\|f-g\|_2 \leqslant \varepsilon$. How can one test whether a function is Fourier sparse? Can one *learn* such functions (i.e. find an approximator) for such functions efficiently, given query access to them?

2.2.3 Junta testing

An important subclass of sparse functions is the class of juntas. A function $f\colon \{-1,1\}^n \to \mathbb{R}$ is said to be a t-junta if there is $T \subseteq [n]$ of size at most t, and $g\colon \{-1,1\}^T \to \mathbb{R}$, such that $f(x) = g(x_T)$. Can one test juntas more efficiently? Learn?