18.218 Topics in Combinatorics Spring 2021 – Lecture 2

Dor Minzer

1 The BLR lineairty test

Recall that a function $f\colon \{-1,1\}^n \to \{-1,1\}$ is called linear if for every $x, y \in \{-1,1\}^n$ it holds that $f(xy) = f(x)f(y)$ where $(xy)_i = x_i y_i$. How does test, given a query access to a function f, if f is a linear function or is far from any linear function?

One idea that makes sense is to consider the following problem. Suppose that $f(xy) = f(x)f(y)$ holds for $\frac{1}{2} + \delta$ fraction of the pairs $x, y \in \{-1,1\}^n$; what can we say about f? Does it have to be close, in some way, to a linear function? This question makes sense both for large δ, i.e. $\delta = \frac{1}{2} - \varepsilon$, as well as for small $\delta > 0$.

To answer this question, it is convenient to consider the convolution operation that is defined as follows. Given $f, g\colon \{-1,1\}^n \to \mathbb{R}$, we define $f * g\colon \{-1,1\}^n \to \mathbb{R}$ by:

$$(f * g)(x) = \mathop{\mathbb{E}}_{y}\left[f(y)g(xy)\right].$$

The most significant property of convolutions is the effect they have in the Fourier domain, as given in the following claim.

Claim 1.1. *For all $S \subseteq [n]$ it holds that $\widehat{f * g}(S) = \widehat{f}(S)\widehat{g}(S)$.*

Proof. By definition,

$$\begin{aligned}\widehat{f * g}(S) = \mathop{\mathbb{E}}_{x}\left[(f * g)(x)\chi_S(x)\right] = \mathop{\mathbb{E}}_{x,y}\left[f(y)g(xy)\chi_S(x)\right] &= \mathop{\mathbb{E}}_{x,y}\left[f(y)\chi_S(y)g(xy)\chi_S(xy)\right] \\ &= \mathop{\mathbb{E}}_{y,z}\left[f(y)\chi_S(y)g(z)\chi_S(z)\right],\end{aligned}$$

and as y, z are independent, the last expectation is equal to $\widehat{f}(S)\widehat{g}(S)$. □

Armed with the above claim, we are ready to analyze the linearity test proposed above.

Theorem 1.2. *Suppose $f\colon \{-1,1\}^n \to \{-1,1\}$ is a function such that $\Pr_{x,y}\left[f(x)f(y) = f(xy)\right] \geqslant \frac{1}{2} + \delta$. Then there exists $S \subseteq [n]$ such that $\widehat{f}(S) \geqslant 2\delta$.*

Proof. Note that whenever the test passes, the value of $f(x)f(y)f(xy)$ is 1, and otherwise the value is -1, so

$$\begin{aligned}\mathop{\mathbb{E}}_{x,y}\left[f(x)f(y)f(xy)\right] = \Pr_{x,y}\left[f(x)f(y) = f(xy)\right] - \Pr_{x,y}\left[f(x)f(y) \neq f(xy)\right] &= 2\Pr_{x,y}\left[f(x)f(y) = f(xy)\right] - 1 \\ &\geqslant 2\delta.\end{aligned}$$

Next, we relate the left hand side to the Fourier coefficients of f. By definition of the convolution,

$$\mathop{\mathbb{E}}_{x,y}\left[f(x)f(y)f(xy)\right] = \mathop{\mathbb{E}}_{x}\left[f(x)\mathop{\mathbb{E}}_{y}\left[f(y)f(xy)\right]\right] = \mathop{\mathbb{E}}_{x}\left[f(x)(f*f)(x)\right] = \langle f, f*f\rangle.$$

Next, using Plancherel and Claim 1.1, we have

$$\langle f, f*f\rangle = \sum_{S}\widehat{f}(S)\widehat{f*f}(S) = \sum_{S}\widehat{f}(S)^3 \leqslant \max_{S}\widehat{f}(S)\sum_{S}\widehat{f}(S)^2 \leqslant \max_{S}\widehat{f}(S)\|f\|_2^2 = \max_{S}\widehat{f}(S).$$

Combining the two inequalities yields the result. □

Recalling that $\widehat{f}(S) = \langle f, \chi_S\rangle = 2\Pr_x\left[f(x) = \chi_S(x)\right] - 1$, we get that a function that passes the linearity test with probability $1/2 + \delta$ must have correlation with a Fourier character. This very nice result exemplifies the power of the basic machinery we have set up so far; proving it without appealing to Fourier analysis is highly challenging.

Note that when $\delta = \frac{1}{2} - \varepsilon$, we even get that $\Pr_x\left[f(x) = \chi_S(x)\right] \geqslant 1 - 2\varepsilon$, so in this case f is *close* to a linear function. This is one of the earliest and basic results in the field of property testing, and later on in the course we will use it in the context of hardness of approximation.

Remark 1.3. *Those of you that are familiar with Roth's theorem regarding the appearance of* 3*-term arithmetic progression in dense subsets of* $[N]$ *may notice that similarity between the argument. The case here is much simpler since we are working with a group.*

2 Random restrictions

Another basic and useful tool we will want to add to our toolbox is the notion of restrictions and random restrictions.

Definition 2.1. *Suppose we have a function* $f\colon \{-1,1\}^n \to \mathbb{R}$*, a set of coordinates* $J \subseteq [n]$ *and an assignment to them* $z \in \{-1,1\}^{\bar{J}}$*. The restricted function* $f_{\bar{J}\to z}\colon \{-1,1\}^J \to \mathbb{R}$ *is defined by*

$$f_{\bar{J}\to z}(y) = f(x_{\bar{J}} = z, x_J = y).$$

Definition 2.2. *Given* $f\colon \{-1,1\}^n \to \mathbb{R}$ *and* $J \subseteq [n]$*, a random restriction of* f *on* J *is a function* $f_{\bar{J}\to z}$ *wherein* $z \in \{-1,1\}^{\bar{J}}$ *is sampled uniformly at random.*

Restrictions and random restrictions are a very powerful tool we will see some uses for throughout the course. In this lecture, we will focus on seeing some basic properties of it and intuition to where it is useful in. In the next lecture we will see a very cool application of them in the problem of learning Fourier sparse functions.

For now, we will begin by investigating several basic and useful properties of it. First, we give a formula for the Fourier coefficients of the restricted function.

Claim 2.3. *Let* $f\colon \{-1,1\}^n \to \mathbb{R}$*,* $J \subseteq [n]$*,* $z \in \{-1,1\}^{\bar{J}}$ *and* $S \subseteq J$*. We have*

$$\widehat{f_{\bar{J}\to z}}(S) = \sum_{T\subseteq \bar{J}}\widehat{f}(S\cup T)\chi_T(z).$$

Proof. We write f according to its Fourier transform, decomposing a character into its J and $\bar{J}$ parts

$$f(x) = \sum_{S\subseteq J, T\subseteq \bar{J}} \widehat{f}(S\cup T)\chi_{S\cup T}(x) = \sum_{S\subseteq J, T\subseteq \bar{J}} \widehat{f}(S\cup T)\chi_S(x_J)\chi_T(x_{\bar{J}}).$$

Plugging in the value y to x_J and z to $x_{\bar{J}}$, we get that

$$f_{\bar{J}\to z}(y) = f(y,z) = \sum_{S\subseteq J}\left(\sum_{T\subseteq \bar{J}} \widehat{f}(S\cup T)\chi_T(z)\right)\chi_S(y).$$

The claim now follows from the uniqueness of the Fourier decomposition. ☐

Using the last claim, we have the following corollary.

Claim 2.4. *Let $f\colon \{-1,1\}^n \to \mathbb{R}$, $J \subseteq [n]$ and $S \subseteq J$. We have*

$$\mathop{\mathbb{E}}_{z}\left[\widehat{f_{\bar{J}\to z}}(S)^2\right] = \sum_{T\subseteq \bar{J}} \widehat{f}(S\cup T)^2.$$

Proof. Defining $g(z) = \widehat{f_{\bar{J}\to z}}(S)$, the left hand side is $\|g\|_2^2$, and the claim follows from the last claim and Parseval. ☐

In some applications, it is useful to consider p-random restrictions, which are random restrictions in which the set J of live variables is also chosen randomly.

Definition 2.5. *Given a function $f\colon \{-1,1\}^n \to \mathbb{R}$ and a parameter $p \in [0,1]$, a p-random restriction is sampled by: taking $J \subseteq [n]$ randomly by including each $i \in [n]$ in J with probability p, and then taking $z \in \{-1,1\}^{\bar{J}}$.*

What is the effect of a random restriction on a function? Let us consider a few examples.

1. Monomials: suppose $f(x) = \chi_S(x) = \prod_{i\in S} x_i$. Then if we take (J,z) a p-random restriction, we expect the restricted function $f_{J\to z}$ to be a (signed) monomial of degree $\approx p\,|S|$. That is, random restriction "reduce" the degree of monomials; we will later see a more general statement along these lines.

2. An OR function, i.e. function of the form $\bigvee_{i\in I} x_i$. Under random restriction (J,z), the function either trivializes to 1, if there is a variable I_i in $\bar{J}$ receiving the value 1, and otherwise the function reduces to an OR on roughly $p\,|I_i|$ variables.

3. CNF formulas: i.e. a function $f\colon \{0,1\}^n \to \{0,1\}$ of the form $f(x) = \bigwedge_{i=1}^m \bigvee_{j\in I_i} x_j$. Analyzing the effect of random restrictions on such functions is significantly more difficult (the Håstad switching lemma). For now it will be enough for us to understand that intuitively, random restrictions significantly simplify them: if there is a term that becomes completely 0, the function trivializes to 1; terms that become 1 disappear, and the rest considerably shrink in width.

We will now establish more rigorously several more properties that align and express some of the above intuition. We will try to capture the sense in which random restrictions reduce degrees, and for that we define the Fourier weight of a function on degrees.

Definition 2.6. *Let $f\colon \{-1,1\}^n \to \mathbb{R}$ be a function, and $d \in \mathbb{N}$. The level d Fourier weight of a function f is defined as*

$$W^{=d}[f] = \sum_{|S|=d} \widehat{f}(S)^2.$$

We also define $W^{\leqslant d}[f] = \sum_{i\leqslant d} W^{=i}[f]$ and $W^{\geqslant d}[f] = \sum_{i\geqslant d} W^{=i}[f]$.

Claim 2.7. *Let $f\colon \{-1,1\}^n \to \mathbb{R}$, $d \in \mathbb{N}$, and let (J,z) be a p-random restriction. Then*

$$\mathop{\mathbb{E}}_{J,z}\left[W^{=d}[f_{\bar{J}\to z}]\right] = \sum_{Q} \widehat{f}(Q)^2 \Pr\left[\mathsf{Bin}(|Q|, p) = d\right].$$

Proof. Expanding,

$$\mathop{\mathbb{E}}_{J,z}\left[W^{=d}[f_{\bar{J}\to z}]\right] = \mathop{\mathbb{E}}_{J,z}\left[\sum_{S\subseteq J, |S|=d} \widehat{f_{\bar{J}\to z}}(S)^2\right] = \mathop{\mathbb{E}}_{J}\left[\sum_{|S|=d} 1_{S\subseteq J} \mathop{\mathbb{E}}_{z}\left[\widehat{f_{\bar{J}\to z}}(S)^2\right]\right].$$

Using Claim 2.4 we calculate the innermost expectation and hence get that

$$\begin{aligned}\mathop{\mathbb{E}}_{J,z}\left[W^{=d}[f_{\bar{J}\to z}]\right] = \mathop{\mathbb{E}}_{J}\left[\sum_{|S|=d} 1_{S\subseteq J} \sum_{T\subseteq \bar{J}} \widehat{f}(S\cup T)^2\right] &= \sum_{Q} \mathop{\mathbb{E}}_{J}\left[1_{|Q\cap J|=d}\right] \widehat{f}(Q)^2 \\ &= \sum_{Q} \widehat{f}(Q)^2 \Pr\left[\mathsf{Bin}(|Q|, p) = d\right]. \square\end{aligned}$$

There are two immediate corollaries one may derive from the above claim. The first one is that if f has most of its Fourier mass below level d, then $f_{\bar{J}\to z}$ has most of its Fourier mass below level $\approx pd$.

Corollary 2.8. *Suppose that $f\colon \{-1,1\}^n \to \{-1,1\}$ satisfies $W_{\geqslant d}[f] \leqslant \varepsilon$, and let (J,z) be a p-random restriction. Then*

$$\mathop{\mathbb{E}}_{J,z}\left[W^{\geqslant 2pd}[f_{\bar{J}\to z}]\right] \leqslant \varepsilon + \mathsf{exp}(-\Theta(pd)).$$

Proof. Summing the previous claim, we have that

$$\mathop{\mathbb{E}}_{J,z}\left[W^{\geqslant 2pd}[f_{\bar{J}\to z}]\right] = \sum_{Q} \widehat{f}(Q)^2 \Pr\left[\mathsf{Bin}(|Q|, p) \geqslant 2pd\right] = \sum_{k\geqslant 0} W^{=k}[f] \Pr\left[\mathsf{Bin}(k, p) \geqslant 2pd\right].$$

We break the last sum into two. For $k \geqslant d$, we bound it by $W^{\geqslant d}[f]$, which by the premise of the statement is at most ε. For $k < d$, we have that

$$\Pr\left[\mathsf{Bin}(k,p) \geqslant 2pd\right] \leqslant \Pr\left[\mathsf{Bin}(d,p) \geqslant 2pd\right] \leqslant \mathsf{exp}(-\Theta(pd)),$$

so the total contribution from these summands is at most $\sum_{k\geqslant 0} W^{=k}[f]\mathsf{exp}(-\Theta(pd)) = \|f\|_2^2 \mathsf{exp}(-\Theta(pd)) = \mathsf{exp}(-\Theta(pd))$. $\square$

The second corollary asserts that if f has sizable mass around level d, then $f_{\bar{J}\to z}$ has sizable weight around level pd.

Definition 2.9. *We define the weight around level* d *to be* $W^{\approx d}[f] = \sum_{d \leqslant k \leqslant 2d} W^{=k}[f]$.

Corollary 2.10. *Let* $d \in \mathbb{N}$ *and* $p \in [0, 1]$ *be such that* $pd \geqslant 10$. *Suppose that* $f\colon \{-1, 1\}^n \to \{-1, 1\}$ *satisfies* $W_{\geqslant d}[f] \leqslant \varepsilon$, *and let* (J, z) *be a* p*-random restriction. Then*

$$\mathop{\mathbb{E}}_{J,z}\left[W^{\approx pd}[f_{\bar{J} \to z}]\right] \geqslant \Omega(W^{\approx d}[f]).$$

Proof. The proof is similar to the proof of the last statement and is left to the reader. □

18.218 Topics in Combinatorics: Analysis of Boolean Functions
Spring 2021