COMS 6998: Unconditional Lower Bounds and Derandomization | Spring 2024

Lecture 9: March 19, 2024

Lecturer: Rocco Servedio | *Scribe: Shiv Kampani*

1 Lecture Agenda

- Finish proof of Viola's Theorem by proving $\mathsf{L2}$ lemma on fooling "balanced" functions (hard case).
- Introduce and motivate fooling AC^0 with $(\log n)^{O(1)}$ independence.
- Define and prove "point-wise" BRS polynomial approximator and "L_2" LMN polynomial approximator for AC^0 (tools for Braverman's Theorem).

2 Viola's Theorem (continued)

To complete the proof of Viola's Theorem from last time, we need to prove the following lemma (referred to as the Balanced Case lemma or $\mathsf{L2}$ in the notes):

Lemma 1 (Balanced Case)**.** *Suppose* $\mathbf{W}$ *γ-fools* DEG_{i-1}. *Let* $\mathbf{Y}$ *be independent of* $\mathbf{W}$ *and δ-biased. Then,*

$$\mathbf{W}+\mathbf{Y}\ \left(\mathrm{imbal}(f)+\sqrt{\frac{\gamma}{2}+\frac{\delta}{2}}\right)\text{-}\textit{fools}\ \mathrm{DEG}_i.$$

In order to prove Lemma 1, we will make use of the following technical lemma:

Lemma 2 (Technical)**.** *Fix even $n \in \mathbb{N}$. Let $g : \mathbb{F}_2^{n/2} \to \{\pm 1\}$ and $f(x,y) \coloneqq g(x)g(y)$. Let $\mathbf{U} \sim \mathcal{U}_{n/2}$ and $\mathbf{Y}$ be ϵ-biased over $\mathbb{F}_2^{n/2}$. Then, $(\mathbf{U}, \mathbf{U}+\mathbf{Y})$ ϵ-fools f. Explicitly:*

$$\left|\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{U}'}[f(\mathbf{U},\mathbf{U}')] - \mathop{\mathbb{E}}_{\mathbf{U},\mathbf{Y}}[f(\mathbf{U},\mathbf{U}+\mathbf{Y})]\right| \le \epsilon.$$

Proof. (Lemma 2) Let $F : \mathbb{F}_2^{n/2} \to [-1,1]$, $F(x) \coloneqq \mathbb{E}_{\mathbf{U}}[g(\mathbf{U})g(\mathbf{U}+x)]$. $\mathbf{U}', \mathbf{U}'' \sim \mathcal{U}_{n/2}$. Note that $F(x) = \mathbb{E}_{\mathbf{U}}[f(\mathbf{U}, \mathbf{U}+x)]$. Consider some random variable $\mathbf{X}$ that ϵ-fools F.

$$\begin{aligned}
&\left| \mathop{\mathbb{E}}_{\mathbf{U}''}[F(\mathbf{U}'')] - \mathop{\mathbb{E}}_{\mathbf{X}}[F(\mathbf{X})] \right| \le \epsilon \\
\implies &\left| \mathop{\mathbb{E}}_{\mathbf{U}''}[\mathop{\mathbb{E}}_{\mathbf{U}}[f(\mathbf{U}, \mathbf{U}+\mathbf{U}'')]] - \mathop{\mathbb{E}}_{\mathbf{X}}[\mathop{\mathbb{E}}_{\mathbf{U}}[f(\mathbf{U}, \mathbf{U}+\mathbf{X})]] \right| \le \epsilon \\
\implies &\left| \mathop{\mathbb{E}}_{\mathbf{U},\mathbf{U}'}[f(\mathbf{U}, \mathbf{U}')] - \mathop{\mathbb{E}}_{\mathbf{U},\mathbf{X}}[f(\mathbf{U}, \mathbf{U}+\mathbf{X})] \right| \le \epsilon \\
\implies &\mathbf{X} \ \epsilon\text{-fools } f
\end{aligned}$$

Thus, to prove Lemma 2, it is sufficient to show that $\mathbf{Y}$ ϵ-fools F. We will first show that $L_1(F) = 1$ and apply $\triangle$-inequality to complete our proof.
Fix $S \subseteq [n/2]$. We will determine $\widehat{F}(S)$. By definition:

$$\begin{aligned}
\widehat{F}(S) &= \mathop{\mathbb{E}}_{\mathbf{U}'}[\chi_S(\mathbf{U}')F(\mathbf{U}')] \\
&= \mathop{\mathbb{E}}_{\mathbf{U},\mathbf{U}'}[g(\mathbf{U})g(\mathbf{U}+\mathbf{U}')\,\chi_S(\mathbf{U}')] \\
&= \mathop{\mathbb{E}}_{\mathbf{U},\mathbf{U}'}[g(\mathbf{U})g(\mathbf{U}')\,\chi_S(\mathbf{U}+\mathbf{U}')] && \dots \quad \text{(Fact 3)} \\
&= \mathop{\mathbb{E}}_{\mathbf{U},\mathbf{U}'}[g(\mathbf{U}')g(\mathbf{U}')\,\chi_S(\mathbf{U})\chi_S(\mathbf{U}')] \\
&= \mathop{\mathbb{E}}_{\mathbf{U}}[\chi_S(\mathbf{U})g(\mathbf{U})]\ \mathop{\mathbb{E}}_{\mathbf{U}'}[\chi_S(\mathbf{U}')g(\mathbf{U}')] && \dots \quad \text{(independence)} \\
&= \left(\mathop{\mathbb{E}}_{\mathbf{U}}[\chi_S(\mathbf{U})g(\mathbf{U})] \right)^2 = (\widehat{g}(S))^2
\end{aligned}$$

Fact 3. *When considering products of functions of random variables, renaming the random variables such that the resulting distribution is identical, gives the same result, in expectation.*

The product whose variables we are interested in renaming (above) is $g(\mathbf{U}')g(\mathbf{U} + \mathbf{U}')\,\chi_S(\mathbf{U}')$. Observe that the first variable is uniformly random, the third variable is again uniformly random and independent with respect to the first. The second variable is related to the first and third variables; specifically, the sum (over $\mathbb{F}_2$) of the first and the third variables equals the second. We can rename the product to $\mathbf{U})g(\mathbf{U}')\,\chi_S(\mathbf{U} + \mathbf{U}')$, keeping the distribution identical, giving the same result in expectation by Fact 3.

$$L_1(F) = \sum_S |\widehat{F}(S)| = \sum_S |\widehat{g}(S)|^2 = \mathop{\mathbb{E}}_{\mathbf{U}}[g(\mathbf{U})^2] = 1 \quad \dots \quad \text{(Parseval's Identity)}$$

The $\triangle$-inequality implies that for ϵ-biased $\mathbf{Y}$ over $\mathbb{F}_2^{n/2}$, $\mathbf{Y}$ $(L_1(f)\cdot\epsilon)$-fools any $F : \mathbb{F}_2^{n/2} \to \mathbb{R}$. Since $L_1(F) = 1$, $\mathbf{Y}$ ϵ-fools $F \implies (\mathbf{U}, \mathbf{U}+\mathbf{Y})$ ϵ-fools f. ■

Using technical Lemma 2, we will now prove Lemma 1.

Proof. (Lemma 1) We are interested in bounding the following quantity:

$$\left|\mathop{\mathbb{E}}_{\mathbf{W},\mathbf{Y}}[f(\mathbf{W}+\mathbf{Y})] - \mathop{\mathbb{E}}_{\mathbf{U}}[f(\mathbf{U})]\right| = \frac{1}{2}\left|\mathop{\mathbb{E}}_{\mathbf{W},\mathbf{Y}}[(-1)^{f(\mathbf{W}+\mathbf{Y})}] - \mathop{\mathbb{E}}_{\mathbf{U}}[(-1)^{f(\mathbf{U})}]\right|$$

The reason for the factor of 1/2 is because of the change to ±1 notation. In general:

Fact 4. *Let* $f : \{0,1\}^n \to \{0,1\}$. $\mathbb{E}[f] = 1 - 2\mathbb{E}[(-1)^f]$

Recall that $\mathrm{imbal}(f) \coloneqq \mathbb{E}_{\mathbf{U}}[(-1)^{f(\mathbf{U})}]$. Thus, by $\triangle$-inequality:

$$\frac{1}{2}\left|\mathop{\mathbb{E}}_{\mathbf{W},\mathbf{Y}}[(-1)^{f(\mathbf{W}+\mathbf{Y})}] - \mathop{\mathbb{E}}_{\mathbf{U}}[(-1)^{f(\mathbf{U})}]\right| \le \frac{1}{2}\left|\mathop{\mathbb{E}}_{\mathbf{W},\mathbf{Y}}[(-1)^{f(\mathbf{W}+\mathbf{Y})}]\right| + \frac{1}{2}(\mathrm{imbal}(f))$$

Our new goal is to bound $\left|\mathbb{E}_{\mathbf{W},\mathbf{Y}}[(-1)^{f(\mathbf{W}+\mathbf{Y})}]\right|$. We will use the same ideas for proving correlation bounds from previous lectures, namely squaring and Cauchy-Schwarz.

Corollary 5. *Cauchy-Schwarz inequality implies that for any RV* $\boldsymbol{A}$, $\mathbb{E}[\mathbf{A}]^2 \le \mathbb{E}[\mathbf{A}^2]$

Squaring the quantity of interest:

$$\begin{aligned}
\left(\mathop{\mathbb{E}}_{\mathbf{W},\mathbf{Y}}[(-1)^{f(\mathbf{W}+\mathbf{Y})}]\right)^2 &= \left(\mathop{\mathbb{E}}_{\mathbf{W}}[\mathop{\mathbb{E}}_{\mathbf{Y}}[(-1)^{f(\mathbf{W}+\mathbf{Y})}]]\right)^2 \\
&\le \mathop{\mathbb{E}}_{\mathbf{W}}\left[\left(\mathop{\mathbb{E}}_{\mathbf{Y}}[(-1)^{f(\mathbf{W}+\mathbf{Y})}]\right)^2\right] && \ldots \quad \text{(Corollary 5)} \\
&= \mathop{\mathbb{E}}_{\mathbf{W},\mathbf{Y},\mathbf{Y}'}[(-1)^{f(\mathbf{W}+\mathbf{Y})+f(\mathbf{W}+\mathbf{Y}')}] && \ldots \quad (\mathbf{Y}, \mathbf{Y}' \text{ are i.i.d.})
\end{aligned}$$

For any fixed outcome of $\mathbf{Y}$, $f^{+\mathbf{Y}}(x) \coloneqq f(x+\mathbf{Y})$ is a degree-i polynomial in x. Observe that, for any fixed outcome of $\mathbf{Y}+\mathbf{Y}'$, we have $f(x+\mathbf{Y}) + f(x+\mathbf{Y}') = \partial_{\mathbf{Y}+\mathbf{Y}'} f^{+\mathbf{Y}}(x)$. Recall that ∂ is the directional derivative defined last time. Clearly, $f(x+\mathbf{Y})+f(x+\mathbf{Y}')$

has degree (at most) $i-1$. By definition, $\mathbf{W}$ γ-fools DEG_{i-1}. Thus:

$$\begin{aligned}
&\left|\mathop{\mathbb{E}}_{\mathbf{W},\mathbf{Y},\mathbf{Y}'}[f^{+\mathbf{Y}}(\mathbf{W})+f^{+\mathbf{Y}'}(\mathbf{W})]-\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{Y},\mathbf{Y}'}[f^{+\mathbf{Y}}(\mathbf{U})+f^{+\mathbf{Y}'}(\mathbf{U})]\right|\leq\gamma\\
\iff&\left|\mathop{\mathbb{E}}_{\mathbf{W},\mathbf{Y},\mathbf{Y}'}[(-1)^{f^{+\mathbf{Y}}(\mathbf{W})+f^{+\mathbf{Y}'}(\mathbf{W})}]-\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{Y},\mathbf{Y}'}[(-1)^{f^{+\mathbf{Y}}(\mathbf{U})+f^{+\mathbf{Y}'}(\mathbf{U})}]\right|\leq 2\gamma \quad \ldots \quad \text{(Fact 4)}\\
\implies&\mathop{\mathbb{E}}_{\mathbf{W},\mathbf{Y},\mathbf{Y}'}[(-1)^{f^{+\mathbf{Y}}(\mathbf{W})+f^{+\mathbf{Y}'}(\mathbf{W})}]\leq\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{Y},\mathbf{Y}'}[(-1)^{f^{+\mathbf{Y}}(\mathbf{U})+f^{+\mathbf{Y}'}(\mathbf{U})}]+2\gamma
\end{aligned}$$

Our new (and final) goal is to bound $\mathbb{E}_{\mathbf{U},\mathbf{Y},\mathbf{Y}'}[(-1)^{f^{+\mathbf{Y}}(\mathbf{U})+f^{+\mathbf{Y}'}(\mathbf{U})}]$.
Observe that the distribution of $(\mathbf{U}+\mathbf{Y}),(\mathbf{U}+\mathbf{Y}')$ is identical to the distribution of $(\mathbf{U},\mathbf{U}+\mathbf{Y}+\mathbf{Y}')$. Specifically, both distributions have the first variable as uniformly random and the second as uniformly random but shifted by $(\mathbf{Y}+\mathbf{Y}')$. By Fact 3, both distributions produce the same result for the term of interest (in expectation).

$$\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{Y},\mathbf{Y}'}[(-1)^{f^{+\mathbf{Y}}(\mathbf{U})+f^{+\mathbf{Y}'}(\mathbf{U})}]=\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{Y},\mathbf{Y}'}[(-1)^{f(\mathbf{U})+f^{+\mathbf{Y}+\mathbf{Y}'}(\mathbf{U})}]$$

Observation 6. *Let $\mathbf{Y_1},\ldots,\mathbf{Y_d}$ be independent δ-biased RV's over $\mathbb{F}_2^n$. Then, $Y\coloneqq\mathbf{Y_1}+\cdots+\mathbf{Y_d}$ is a δ^d-biased RV over $\mathbb{F}_2^n$.*

This observation was proved last time. A corollary is that $\mathbf{Y}+\mathbf{Y}'$ is a δ^2-biased RV. Let $g(x)\coloneqq(-1)^{f(x)}$ and $h(x,y)\coloneqq g(x)g(y)$.

$$\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{Y},\mathbf{Y}'}[(-1)^{f(\mathbf{U})+f^{+\mathbf{Y}+\mathbf{Y}'}(\mathbf{U})}]=\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{Y},\mathbf{Y}'}[g(\mathbf{U})g(\mathbf{U}+\mathbf{Y}+\mathbf{Y}')]=\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{Y},\mathbf{Y}'}[h(\mathbf{U},\mathbf{U}+\mathbf{Y}+\mathbf{Y}')]$$

Applying our technical Lemma 2:

$$\begin{aligned}
&\left|\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{U}'}[h(\mathbf{U},\mathbf{U}')]-\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{Y},\mathbf{Y}'}[h(\mathbf{U},\mathbf{U}+\mathbf{Y}+\mathbf{Y}')]\right|\leq\delta^2\\
&\qquad\implies\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{Y},\mathbf{Y}'}[h(\mathbf{U},\mathbf{U}+\mathbf{Y}+\mathbf{Y}')]\leq\mathop{\mathbb{E}}_{\mathbf{U},\mathbf{U}'}[h(\mathbf{U},\mathbf{U}')]+\delta^2\\
&\qquad\qquad\leq\left(\mathop{\mathbb{E}}_{\mathbf{U}}[(-1)^{f(\mathbf{U})}]\right)^2+\delta^2\\
&\qquad\qquad=\mathrm{imbal}(f)^2+\delta^2
\end{aligned}$$

Putting all the pieces together, we have shown the following:

$$\begin{aligned}\left(\mathop{\mathbb{E}}_{\mathbf{W},\mathbf{Y}}[(-1)^{f(\mathbf{W}+\mathbf{Y})}]\right)^2 &\leq \mathrm{imbal}(f)^2 + \delta^2 + 2\gamma \\ \implies \mathop{\mathbb{E}}_{\mathbf{W},\mathbf{Y}}[(-1)^{f(\mathbf{W}+\mathbf{Y})}] &\leq \mathrm{imbal}(f) + \delta + \sqrt{2\gamma} \\ \therefore \left|\mathop{\mathbb{E}}_{\mathbf{W},\mathbf{Y}}[f(\mathbf{W}+\mathbf{Y})] - \mathop{\mathbb{E}}_{\mathbf{U}}[f(\mathbf{U})]\right| &\leq \frac{1}{2}\left(\mathrm{imbal}(f) + \delta + \sqrt{2\gamma}\right) + \frac{1}{2}(\mathrm{imbal}(f)) \\ &= \left(\mathrm{imbal}(f) + \frac{\delta}{2} + \sqrt{\frac{\gamma}{2}}\right)\end{aligned}$$

This completes our proof of Lemma 1. ■

3 Bounded Independence Fools AC^0

In this section, we will discuss Braverman's Theorem which states that any k-wise independent RV fools AC^0. Before formally stating and proving this theorem, we will motivate PRGs for AC^0.

Motivation. In short, we ♡ AC^0. In previous lectures, we have proved both worst-case and average-case lower bounds for AC^0, constructing a pseudorandom generator for the class is a natural extension. Here is a **brief history** of this problem:

- 1990 • Linial & Nisan made conjecture that polynomial-size constant depth circuits (AC^0) is fooled by any k-wise independent distribution for $k \sim \mathrm{polylog}(n)$. [LN90]
- 2007 • Bazzi proved the conjecture for $d = 2$ (50 page paper). [Baz07]
- 2008 • Razborov presented a simplified proof for $d = 2$ (4 page paper). [Raz08]
- 2009 • Braverman proved the conjecture for all d [Bra09]

It is worth mentioning that this PRG is neither the simplest, nor the first, nor with the best parameters. The current state-of-the-art for an ϵ-PRG for $\mathsf{AC}^0_{s,d}$ has seed-length $k = (\log s)^{d+O(1)} \cdot \log\frac{1}{\epsilon}$ [ST19] (to achieve this result, we need to make use of the **multi-switching lemma**). Nevertheless, k-wise independence is "general" machinery (we do not need to make contrived constructions for random variables).

Intuition (for conjecture). Why is k-wise independence "good enough" to fool AC^0? There are several reasons for this:

1. k-wise independence fools $\mathcal{DT}_k$, k-juntas ($\mathcal{J}_k$), etc.
2. k-wise independence fools real polynomials of degree-k (can be shown by definition and application of $\triangle$-inequality).
3. Since the 90's it has been known that circuits in AC^0 are well-approximated by low-degree real polynomials. There are two such approximators of interest:
   (a) Point-wise approximator (Beigel, Reingold, Spielman) [BRS91]
   (b) L_2 approximator (Linial, Mansour, Nisan) [LMN89]

Neither of the approximators above, however, is a sandwiching approximator. The idea behind Braverman's proof was to combine these approximators in a clever way to construct a sandwiching polynomial approximator.

Theorem 7 (Braverman's Theorem)**.** *Let* $k = \left(\log \frac{s}{\epsilon}\right)^{O(d)}$. *Let* $\mathbf{D}$ *be any* k*-wise independent RV over* $\{0,1\}^n$. *Then* $\mathbf{D}$ ϵ*-fools* $\mathsf{AC}^0_{s,d}$.

Remark. Braverman's original proof was worse; it had a $O(d^2)$ exponent for k. Before proving Braverman's theorem, we will prove both polynomial approximators.

3.1 Point-wise approximator

Lemma 8 ("point-wise" BRS approximator)**.** *Let* $f \in \mathsf{AC}^0_{s,d}$. *Let* $\mathcal{D}$ *be any distribution over* $\{0,1\}^n$. $\exists$ *a real polynomial* $p:$ $\Pr_{\boldsymbol{x}\sim\mathcal{D}}[p(\boldsymbol{x}) = f(\boldsymbol{x})] \geq 1-\epsilon$, *where:*

(i) $deg(p) \leq \left(\log \frac{s}{\epsilon}\right)^{O(d)}$

(ii) $\forall x \in \{0,1\}^n$ $|p(x)| \leq \exp\left(\log \frac{s}{\epsilon}\right)^{O(d)}$

Proof. (Lemma 8) We will begin by proving the existence of polynomial "point-wise" approximators for OR gates. An OR gate with fan-in t computes $f(x) = x_1 \vee \cdots \vee x_t$. We will perform a probabilistic construction by describing a distribution over polynomials, making random draws from the distribution, and showing that they "do well" with high probability.

Let $v_0 \coloneqq \{x_1, \ldots, x_t\}$. Let $p_0(x) \coloneqq x_1 + \cdots + x_t$. For $i \in 1, \ldots, \log_2(t)+1$, let $\boldsymbol{v}_i$ be constructed from $\boldsymbol{v}_{i-1}$ by independently and randomly discarding each variable in $\boldsymbol{v}_{i-1}$ with probability 1/2. Let $\boldsymbol{p}_i(x) \coloneqq \sum_{x_j \in \boldsymbol{v}_i} x_j$.

Fact 9. $p_0, \boldsymbol{p}_1, \ldots, \boldsymbol{p}_{\log_2(t)+1}$ *are all deg-*1 *polynomials mapping* $\{0,1\}^t \to [0,t]$

The idea behind these polynomials is a process of **random sifting** to isolate a single 1 from the variables $x_1, \ldots, x_t$. $p_0(x)$ "fails" at approximation when the input x contains multiple 1's, since it leads to a sum > 1. But, among the "sifted" polynomials, there is a good chance that some polynomial is exactly 1. Note that in the special case that the input is 0^t, all polynomials are 0 (correct approximators).

Fix any input assignment $z \in \{0,1\}^t : z \neq 0^t \implies z_1 \vee \cdots \vee z_t = 1$.

Claim 10. $\Pr[$*at least one of* $p_0(z), \boldsymbol{p}_1(z), \ldots, \boldsymbol{p}_{\log_2(t)+1}(z) = 1] \geq 1/3$

Proof. We will consider three mutually-exclusive cases:

(a) **Case:** $\forall i = 1, \ldots, \log_2(t)+1 \;\; \boldsymbol{p}_i(z) > 1$.

Each fixed variable has a $1/2t$ chance of surviving all stages of sifting. Thus, the probability that any variable z_j survives all the stages of sifting $\leq 1/2$ (union bound). Thus, as a coarse upper bound, $\Pr[(a)] \leq 1/2$.

(b) **Case:** $p_0(z) = 1$.

(c) **Case:** for some stage $i \in [1, \log_2(t)]$, $\boldsymbol{p}_i(z) > 1$ and $\boldsymbol{p}_{i+1}(z) \leq 1$.

For any j, given the value of $\boldsymbol{p}_j(z)$, we have:

- $\Pr[\boldsymbol{p}_{j+1}(z) = 0] = \left(\frac{1}{2}\right)^{\boldsymbol{p}_j(z)}$ (all the 1-variables in z need to be discarded in iteration $j+1$ for $\boldsymbol{p}_{j+1}(z) = 0$).
- $\Pr[\boldsymbol{p}_{j+1}(z) = 1] = \boldsymbol{p}_j(z) \cdot \left(\frac{1}{2}\right)^{\boldsymbol{p}_j(z)}$ (all but **exactly one** of the 1-variables in z need to be discarded in iteration $j+1$ for $\boldsymbol{p}_{j+1}(z) = 0$).

For stage i, in the worst-case, $\boldsymbol{p}_i(z) = 2$. In this specific case, we are given that $\boldsymbol{p}_i(z) > 1$ and $\boldsymbol{p}_{i+1}(z) \leq 1$. Thus, $\Pr[\boldsymbol{p}_{i+1}(z) = 1 | \boldsymbol{p}_i(z) > 1 \wedge \boldsymbol{p}_{i+1}(z) \leq 1] = \frac{\boldsymbol{p}_j(z)}{1+\boldsymbol{p}_j(z)} \geq 2/3$ (worst-case probability).

Consider the cases in which our claim event (at least one of the polynomials equals 1) is true. As we have shown, in case (a) our claim event is always false, in case (b) our claim event is always true, and in case (c) our claim event is true with probability $\geq 2/3$. Since we showed $\Pr[\text{(a)}] \leq 1/2$, either (b) or (c) occurs with probability $\geq 1/2$. Thus, the probability that our claim event is true $= \frac{1}{2}(\Pr[(b)|\text{not } (a)] + \frac{2}{3}\Pr[(c)|\text{not } (a)]) \geq \frac{1}{2} \cdot \frac{2}{3} = \frac{1}{3}$, which proves our claim. ■

To prove Lemma 8, we need a **single** polynomial approximator with amplified success probability. Define $\boldsymbol{r}(x) \coloneqq \prod_{i=0}^{\log_2 t+1}(1 - \boldsymbol{p}_i(x))$. $\boldsymbol{r}(x)$ is a polynomial of degree $\leq \log_2(t) + 1$ that maps $\{0,1\}^t \to [-t^{O(\log t)}, t^{O(\log t)}]$. When $x = 0^t, \boldsymbol{r}(x) = 1$ since $\forall i, \boldsymbol{p}_i(x) = 0$. For any $x \neq 0^t, \boldsymbol{r}(x) = 0$ with probability $\geq 1/3$ (from Claim 10).

To improve success probability, let $\boldsymbol{r}'(x) \coloneqq$ product of $O(\log \frac{1}{\epsilon})$ independent $\boldsymbol{r}(x)$'s. By definition, $\boldsymbol{r}'(x)$ is a polynomial of degree $\leq O(\log \frac{1}{\epsilon} \cdot \log t)$ that maps $\{0,1\}^t \to [-t^{O(\log \frac{1}{\epsilon} \cdot \log t)}, t^{O(\log \frac{1}{\epsilon} \cdot \log t)}]$. Again, we have $\boldsymbol{r}'(0^t) = 1$. We also have:

$$\text{Let } x \neq 0^t,\ \Pr[\boldsymbol{r}'(x) \neq 0] \leq \left(\frac{2}{3}\right)^{O(\log \frac{1}{\epsilon})} \leq \epsilon \quad \ldots \quad \text{(independence)}$$

Finally, define $\boldsymbol{a}(x) \coloneqq 1 - \boldsymbol{r}'(x)$. Based on the properties of $\boldsymbol{r}'(x)$, we have:

$$\forall x \ \Pr_{\boldsymbol{a}}[\boldsymbol{a}(x) = x_1 \vee \cdots \vee x_t] \geq 1 - \epsilon \quad \ldots \quad (*)$$

Consider Table 1. Each column corresponds to an $x \in \{0,1\}^t$. Each row corresponds to a polynomial outcome of a with respect to the randomness used to generate $\boldsymbol{a}$. A check-mark (✓) in cell (i,j) indicates that $a_i(x_j) = x_{j_1} \vee x_{j_2} \vee \cdots \vee x_{j_t}$. This is an illustration of the **minimax theorem**.

Clearly, there are 2^t columns in the check-mark matrix. By statement $(*)$, we have that, in every column (corresponding to some x_j), the probability of any cell having ✓is $\geq 1 - \epsilon \implies$ ✓-density in every column $\geq 1 - \epsilon$. For any distribution $\mathcal{D}$ over columns, the matrix ✓-density according to this distribution $\mathcal{D}$ is $\geq 1 - \epsilon$. Finally, this

	$x_0 = 0^t$	$x_1 = 0^{t-1}1$	$\dots$
a_1	$\checkmark$	$\dots$	$\ddots$
a_2	$\checkmark$	$\dots$	$\ddots$
$\vdots$	$\vdots$	$\ddots$	$\ddots$

Table 1: Check-mark ($\checkmark$) Matrix (Minimax Theorem)

implies that there must exist a row in the matrix (corresponding to some polynomial outcome a_i) such that the $\checkmark$-density in that row is $\geq 1-\epsilon$. Formally:

$$\forall \text{ distributions } \mathcal{D}, \exists \text{ polynomial outcome of } a : \Pr[a(\boldsymbol{x}) = \boldsymbol{x}_1 \vee \cdots \vee \boldsymbol{x}_t] \geq 1-\epsilon \quad (**)$$

The degree of a is equal to the degree of $\boldsymbol{r}' \leq O(\log \frac{1}{\epsilon} \cdot \log t)$.

So far, we have shown the existence of polynomial "point-wise" approximators for OR gates. In general, Boolean circuits (in the De Morgan basis) may make use of OR, AND, NOT gates. Thus, we need to show the existence of approximators for AND, NOT gates with guarantee $(**)$:

($\neg$) NOT gates take only a single bit as input and flip its value. Thus, $a(x) = 1 - x$ is a degree-1 polynomial that exactly computes $\neg x$.

($\wedge$) We can show that there exist polynomial approximators for AND gates with the same degree and guarantee $(**)$ as the OR gate approximator using the NOT gate polynomial.

Proof sketch: Observe that AND, OR are dual-functions: $\text{AND}(x_1, \dots, x_t) = \text{NOT}(\text{OR}(\text{NOT}(x_1), \dots, \text{NOT}(x_t)))$. For any distribution $\mathcal{D}$, take the polynomial approximator for OR, i.e., $a(x)$. Replace every literal x_i in a with $(1 - x_i)$. Let the updated polynomial be $a'(x)$. Define $a''(x) \coloneqq 1 - a'(x)$. Clearly, from the check-mark matrix, $a''(x)$ is a polynomial approximator for AND on distribution $\mathcal{D}$ with same degree. ■

Any circuit C computing $f \in \mathsf{AC}^0_{s,d}$ will have $\leq s$ gates by definition. Consider any gate g (either OR, AND, NOT) in circuit C. g takes in at most s inputs. Let $\mathcal{D}'$ be the distribution of inputs into gate g when the circuit input $\boldsymbol{x} \sim \mathcal{D}$. By our proofs for OR, AND, NOT gates, for any $\mathcal{D}'$, there exists a polynomial a that "point-wise" approximates gate g with error $1-\epsilon/s$. a has degree at-most $O(\log\frac{s}{\epsilon}\cdot\log s)$.

We can replace every gate in C with the corresponding polynomial approximator and call the result $p(x)$. By union bound $p(x)$ correctly computes every gate in C, and therefore correctly computes f, with probability $\geq 1-\frac{s}{\epsilon}\cdot s = 1-\epsilon$.

$p(x)$ has degree $O(\log\frac{s}{\epsilon}\cdot\log s)^d \leq \left(\log\frac{s}{\epsilon}\right)^{O(d)}$ (satisfies (i)).
In our proof, we showed that $|\boldsymbol{r}'(x)| \leq t^{O(\log\frac{1}{\epsilon}\cdot\log t)} \implies |a(x)| \leq t^{O(\log\frac{1}{\epsilon}\cdot\log t)}$. For any gate g, let a_g be the approximator. $|a_g(x)| \leq s^{O(\log\frac{s}{\epsilon}\cdot\log s)} = \exp\left(O(\log\frac{s}{\epsilon}\cdot\log^2 s)\right)$. Thus, for the polynomial $p(x)$, $|p(x)| \leq |a_g(x)|^d \leq \exp\left(\log\frac{s}{\epsilon}\right)^{O(d)}$ (satisfies (ii)).

We call $p(x)$ the BRS approximator. This concludes our proof of Lemma 8. ■

3.2 L_2 approximator

Theorem 11 ("L_2" LMN approximator)**.** *Let $f \in \mathsf{AC}^0_{s,d}$. $\exists$ a real polynomial p_2 of degree $O(\left(\log\frac{s}{\epsilon}\right)^d)$:* $\mathop{\mathbb{E}}_{\mathbf{U}\sim\mathcal{U}}\left[(f(\mathbf{U})-p_2(\mathbf{U}))^2\right] \leq \epsilon$.

Remark. Like the BRS approximator, the LMN approximator does not necessarily sandwich $f(x)$ since $|f(x)-p_2(x)|$ may (rarely) be large.

Notation. Let $W^k(f) \coloneqq \sum_{S\subseteq[n],|S|=k} \widehat{f}(S)^2$ and $W^{\geq k}(f) \coloneqq \sum_{S\subseteq[n],|S|\geq k} \widehat{f}(S)^2$.

Fact 12. *Let $f:\{0,1\}^n \to \{\pm1\}$. $W^{\geq 0}(f) = \sum_{S\subseteq[n]} \widehat{f}(S)^2 = 1$ (Parseval's Identity).*

Proof. (Theorem 11) The idea behind our proof is to show that for $f \in \mathsf{AC}^0_{s,d}$, the Fourier "tail weight" is very small. Consider the following theorem:

Theorem 13. *If $f \in AC^0_{s,d}$, then: $\forall r \quad W^{>r}(f) \leq 2s\cdot 2^{\frac{-r^{1/d}}{20}}$*

Intuitively, Theorem 13 states that a property of all functions $f \in \mathsf{AC}^0_{s,d}$ is that as one goes "higher up" in the Boolean hypercube, the Fourier tail weight drops exponentially.

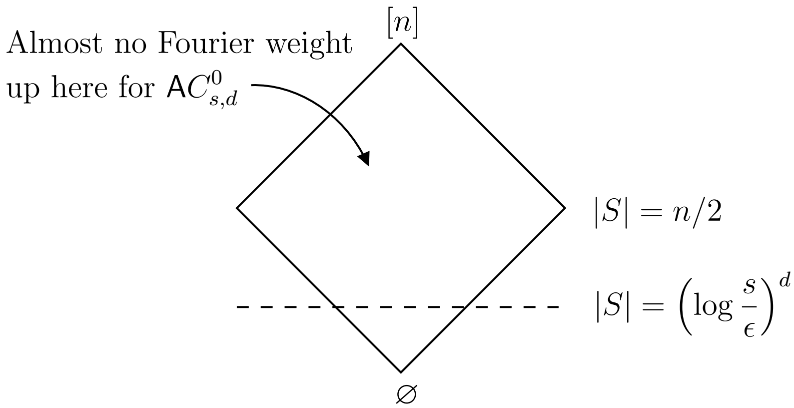

Figure 1: Theorem 13 on Boolean Hypercube

We will use Theorem 13 to prove Theorem 11. Take $r = \left(20 \log \frac{2s}{\epsilon}\right)^d$. For this value of r, $2s \cdot 2^{\frac{-r^{1/d}}{20}} \leq \epsilon$. We define the polynomial approximator p_2 as the truncated (at level r) part of the Fourier representation of f. Explicitly:

$$p_2(x) := \sum_{|S| \leq r} \widehat{f}(S)\, \chi_S(x) = \sum_{|S| \leq r} \widehat{f}(S) \prod_{i \in S} x_i$$

Clearly, the degree of $p \leq r = O(\left(\log \frac{s}{\epsilon}\right)^d)$ as required. All that remains is to prove that p_2 is a valid L_2 polynomial approximator for f. By Parseval's Identity:

$$\begin{aligned}
\mathop{\mathbb{E}}_{\mathbf{U} \sim \mathcal{U}} \left[(f(\mathbf{U}) - p_2(\mathbf{U}))^2 \right] &= \sum_{S \subseteq [n]} \left(\widehat{f - p_2}(S) \right)^2 \\
&= \sum_{S \subseteq [n]} \left(\widehat{f}(S) - \widehat{p_2}(S) \right)^2 \\
&= \sum_{|S| > r} \left(\widehat{f}(S) \right)^2 = W^{>r}(f) \leq \epsilon \quad \text{(Theorem 13)}
\end{aligned}$$

This completes our proof of Theorem 11. All that remains is to prove Theorem 13. ■

Proof. (Theorem 13) To prove this Fourier concentration theorem, we will make use of two helpful lemmas:

Lemma 14 (Håstad's Switching Lemma). *Let $f \in AC^0_{s,d}$. Fix t and let $\boldsymbol{\rho} \sim R_p$ (random restriction with $*$ probability p) with $p \leq \frac{1}{10^d \cdot t^{d-1}}$. Then:*

$$\Pr_{\boldsymbol{\rho} \sim R_p}[\,\text{DT-depth}(f \restriction \boldsymbol{\rho}) \geq t\,] \leq s \cdot 2^{-d}$$

We have already proved Lemma 14 in a previous lecture.

Lemma 15. *For any $f : \{\pm 1\}^n \to \{\pm 1\}$, any $p \leq 1/10$, we have:*

$$W^{\geq t/p}(f) \leq 2 \cdot \mathop{\mathbb{E}}_{(\boldsymbol{J},\boldsymbol{z}) \sim R_p}[W^{\geq t}(f_{\boldsymbol{J},\boldsymbol{z}})]$$

Notation. $(\boldsymbol{J}, \boldsymbol{z}) \sim R_p$ defines a random restriction and is equivalent to $\boldsymbol{\rho}$. $\boldsymbol{J} \subseteq [n]$ is the subset of variables that survived the restriction ($*$'s) and $\boldsymbol{z} \in \{\pm 1\}^{[n] \setminus \boldsymbol{J}}$ is the vector of assignments to the remaining non-$*$ variables.

We will prove Lemma 15 in the next lecture. We will now show that, given both lemmas above, we can prove Theorem 13.

Let $p = \frac{1}{10 r^{\frac{d-1}{d}}} \implies pr = \frac{r^{1/d}}{10}$. Let $t = \frac{pr}{2} = \frac{r^{1/d}}{20}$. Thus: $\frac{1}{10^d \cdot t^{d-1}} = \frac{2^{d-1}}{10 r^{\frac{d-1}{d}}} = 2^{d-1} p \geq p$.

From Lemma 14, $\Pr_{\boldsymbol{\rho} \sim R_p}[\,\text{DT-depth}(f \restriction \boldsymbol{\rho}) \geq t\,] \leq s \cdot 2^{-t}$. Now, we will compute the expected Fourier weight above $t = \frac{r^{1/d}}{20}$ of $f \restriction \boldsymbol{\rho} = f_{\boldsymbol{J},\boldsymbol{z}}$ (restricted function):

$$\begin{aligned}
\mathop{\mathbb{E}}_{(\boldsymbol{J},\boldsymbol{z}) \sim R_p}[W^{\geq t}(f_{\boldsymbol{J},\boldsymbol{z}})] &= W^{\geq t}(f_{\boldsymbol{J},\boldsymbol{z}}) \cdot \Pr_{\boldsymbol{\rho} \sim R_p}[\,\text{DT-depth}(f \restriction \boldsymbol{\rho}) \geq t\,] \\
&\quad + 0 \cdot \Pr_{\boldsymbol{\rho} \sim R_p}[\,\text{DT-depth}(f \restriction \boldsymbol{\rho}) < t\,] \\
&\leq W^{\geq 0}(f_{\boldsymbol{J},\boldsymbol{z}}) \cdot \Pr_{\boldsymbol{\rho} \sim R_p}[\,\text{DT-depth}(f \restriction \boldsymbol{\rho}) \geq t\,] \\
&= \Pr_{\boldsymbol{\rho} \sim R_p}[\,\text{DT-depth}(f \restriction \boldsymbol{\rho}) \geq t\,] \qquad \ldots \quad \text{(Fact 12)} \\
&\leq s \cdot 2^{-t}
\end{aligned}$$

Since $t = \frac{r^{1/d}}{20} \implies t/p = r/2$. Finally, from Lemma 15, $W^{\geq r/2}(f) \leq 2s \cdot 2^{-t} \implies W^{\geq r}(f) \leq 2s \cdot 2^{-t} = 2s \cdot 2^{\frac{-r^{1/d}}{20}}$. This completes our proof of Theorem 13. ■

4 Next Time

- Complete proof of LMN polynomial approximator by proving Lemma 15.
- Prove Braverman's Theorem (Theorem 7).
- PRGs for linear threshold functions.

References

[Baz07] L.M.J. Bazzi. Polylogarithmic independence can fool DNF formulas. *IEEE Computer Society Press*, pages 63–73, 2007. Prelim. ver. in *Proc. of FOCS'07*. 3

[Bra09] M. Braverman. Poly-logarithmic independence fools AC^0 circuits. *Journal of the ACM*, 57:1–10, 2009. Prelim. ver. in *Proc. of CCC'09*. 3

[BRS91] R. Beigel, N. Reingold, and D. Spielman. The perceptron strikes back. *IEEE*, 1991. Prelim. ver. in *Proc. of SCT'91*. 3a

[LMN89] N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, fourier transform, and learnability. *IEEE*, 1989. Prelim. ver. in *Proc. of FOCS'89*. 3b

[LN90] N. Linial and N. Nisan. Approximate inclusion-exclusion. *Combinatorica*, 10:349–365, 1990. Prelim. ver. in *Proc. of STOC'90*. 3

[Raz08] A. Razborov. A simple proof of bazzi's theorem. *Electronic Colloquium on Computational Complexity (ECCC)*, 2008. 3

[ST19] R. Servedio and L. Tan. Improved pseudorandom generators from pseudorandom multi-switching lemmas. *Theory of Computing*, 2019. Prelim. ver. in *Proc. of RANDOM'19*. 3