COMS 6998: Unconditional Lower Bounds and Derandomization **Spring 2024**

Lecture 8: March 5, 2024

Lecturer: Rocco Servedio *Scribe: Mark Chen*

Last Time:

- Finished the missing proof for the second correlation bound, stated as the following: for some deg-d $\mathbb{F}_2$-polynomial p and $P = e(p) = (-1)^p$, we have

  $$\mathsf{Cor}[F, P] \leq U_{d+1}(F)^{\frac{1}{2^{d+1}}}.$$

  The proof was done by using three facts that we also proved in the last lecture.

- Started derandomization part of the course by introducing basic tools for PRGs:
  - K-wise independent RV's (we did a construction with seed length $K \cdot \log n$).
  - ϵ-biased RV's (we did a construction with seed length $O(\log \frac{n}{\epsilon})$.
  - K-wise independent ϵ-biased RV's (we did a construction with seed length $O(K + \log \frac{1}{\epsilon} + \log \log n)$.

Today:

- (1) Basic Fourier analysis over Boolean functions. We apply it to better the following results (compared with what we have seen last time using more bare-minimum approaches):
  - Fooling size-s DTs.
  - Fooling K-juntas.
- (2) Sandwiching approximators & fooling.
- (3) Viola's Theorem: Sum of d ϵ-biased RVs fools DEG_d (we showed most of the proof for Viola's theorem this time, except for the second case [balanced case] of the key lemma to be shown next time).

1 Fourier Analysis over Boolean Functions

Motivation (partially): Fourier analysis can be used to better some of the results we have seen from last class (see section 1.4).

The specific kind of Fourier analysis we will look at is over Boolean functions (to see more beyond today's lecture on this topic, see Ryan O'Donnell's *Analysis of Boolean Functions* [O'D14]).

1.1 Basics

Definition 1 (Basics)**.** *All functions of the form $f : \{0,1\}^n \to \mathbb{R}$ form* ***a 2^n-dimensional vector space*** *(with one dimension for each x).*

Inner product *of this space is given by*

$$\langle f, g\rangle = \mathop{\mathbb{E}}_{x\sim\mathcal{U}}[f(x)\cdot g(x)],$$

and we define the ***norm*** *in this space as:*

$$||f|| = \sqrt{\langle f, f\rangle},$$

which is known as the ℓ_2 norm, in contrast to the L_1 norm given in definition 3.

Proposition 1 (Basis)**.** *The set of all 2^n character functions, $\boxed{(\chi_S)_{S\subseteq[n]}}$, is an* ***orthonormal basis*** *of the space as defined in definition 1. [Recall that we defined $\underline{\text{character functions}}$ last time:*

$$\chi_S(x) = (-1)^{\sum\limits_{i\in S} x_i} = e\left(\sum_{i\in S} x_i\right)].$$

Proof. WTS, by the definition of orthonormal basis, that:

- $\langle\chi_S, \chi_S\rangle = 1, \forall S.$
- $\langle\chi_S, \chi_T\rangle = 0$, if $S \neq T$.

Consider general S, T (i.e. they may not be different), we have, by definition:

$$\begin{aligned}\langle\chi_S, \chi_T\rangle &= \mathop{\mathbb{E}}_{x\sim\mathcal{U}_n}[\chi_S(x)\chi_T(x)] = \mathop{\mathbb{E}}_{x\sim\mathcal{U}_n}\left[(-1)^{\sum\limits_{i\in S} x_i + \sum\limits_{j\in T} x_j}\right] \\ &= \mathop{\mathbb{E}}_{x\sim\mathcal{U}_n}\left[(-1)^{\sum\limits_{j\in S\triangle T} x_j}\right] = \mathop{\mathbb{E}}_{x\sim\mathcal{U}_n}[\chi_{S\triangle T}(x)],\end{aligned}$$

where $\triangle$ is the "symmetric difference" (basic definition is $A \triangle B = (A \backslash B) \cup (B \backslash A)$, see wikipedia for more).

Now,

$$\mathop{\mathbb{E}}_{x \sim \mathcal{U}_n}[\chi_{S \triangle T}(x)] = \begin{cases} 1 \text{ if } S = T \text{ (because } S = T \implies S \triangle T = \emptyset \implies (-1)^0 = 1) \\ 0 \text{ if } S \neq T \text{ (we have shown last time)} \end{cases},$$

so we have exactly what we wanted to show. ■

Corollary 1. *As a corollary of proposition 1, any $f : \{0,1\}^n \to \mathbb{R}$ has a unique representation as a linear combination of $(\chi_S)_{S \subseteq [n]}$. We write $\widehat{f}(S)$ as a coefficient (these are called* ***Fourier coefficients*** *of f) of χ_S in the linear combination:*

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

Definition 2 (Fourier Coefficient)**.** *Fourier coefficient is the* ***inner product*** *of $\underline{f}$ and a $\underline{\text{basis element, } \chi_S}$:*

$$\begin{aligned} \langle f, \chi_S \rangle &= \mathop{\mathbb{E}}_{\mathcal{U}}[\chi_S(\mathcal{U}) \cdot f(\mathcal{U})] \\ &= \mathop{\mathbb{E}}_{\mathcal{U}}\left[\chi_S(\mathcal{U}) \cdot \sum_{T \subseteq [n]} \widehat{f}(T) \cdot \chi_T(\mathcal{U})\right] \\ &= \sum_{T \subseteq [n]} \widehat{f}(T) \underbrace{\mathop{\mathbb{E}}_{\mathcal{U}}[\chi_S(\mathcal{U}) \cdot \chi_T(\mathcal{U})]}_{\text{just shown } = 1 \text{ if } S = T;\ 0 \text{ o/w}} \\ &= \widehat{f}(S). \end{aligned}$$

In other words, $\widehat{f}(S)$ measures the correlation of f and χ_S.

Example 1. *A special case of Fourier coefficient says that:*

$$\widehat{f}(\emptyset) = \mathbb{E}[f(x) \cdot \overbrace{\chi_\emptyset(x)}^{=1}] = \mathbb{E}[f].$$

Remark 1. *Sometimes, it's nice to view $f : \{-1,1\}^n \to \mathbb{R}$. Then,*

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) \cdot \prod_{i \in S} x_i,$$

and the Fourier representation is exactly the same as the representation as multi-linear polynomial.

1.2 Some Identities

Proposition 2 (**Plancherel's Identity**). *For any $f, g : \{0,1\}^n \to \mathbb{R}$, we have*

$$\boxed{\langle f, g\rangle} = \mathbb{E}[f(\mathcal{U}) \cdot g(\mathcal{U})] = \mathbb{E}\left[\left(\sum_S \widehat{f}(S)\chi_S\right) \cdot \left(\sum_T \widehat{g}(T)\chi_T\right)\right]$$

$$= \sum_{S,T} \widehat{f}(S) \cdot \widehat{g}(T) \cdot \overbrace{\mathop{\mathbb{E}}_{\mathcal{U}}[\chi_S(\mathcal{U}) \cdot \chi_T(\mathcal{U})]}^{=1 \iff S=T} \boxed{= \sum_S \widehat{f}(S) \cdot \widehat{g}(S)}$$

Proposition 3 (**Parseval's Identity**). *The special case of* ***Plancherel's*** *where $f = g$ is*

$$\|f\|^2 = \mathop{\mathbb{E}}_{\mathcal{U}}[f(\mathcal{U})^2] = \sum_S \widehat{f}(S)^2.$$

Remark 2. *Proposition 3 is really a generalization of Pythagorean's theorem (square of the length of a vector is equal to the sum of the squares of the vector's length in each direction of its basis).*

Remark 3. *Sometimes, it's helpful to think of our Boolean functions as having outputs in $\{-1, 1\}$, because, with such an $f : \{0,1\}^n \to \{\pm 1\}$,*

$$\|f\|^2 = \mathbb{E}[f^2] = 1 = \sum_{S \subseteq [n]} \widehat{f}(S)^2,$$

which lets us think of the Boolean function's "energy" to spread out among its Fourier coefficients (as their squares sum up to 1).

Official Homework Problem: Let the inner product function be defined as:

$$\mathrm{IP} : \{0,1\}^n \to \{\pm 1\}, \mathrm{IP}(x_1, \ldots, x_n) = (-1)^{x_1x_2 + x_3x_4 + \ldots + x_{n-1}x_n}.$$

- Write the Fourier representation of IP.
- Infer that ∀ deg-1 $\mathbb{F}_2$ polynomial $p(x)$ (which is parity or its negation), we have:

$$\mathop{\Pr}_{\mathcal{U}}[\mathrm{IP}(\mathcal{U}) = p(\mathcal{U})] = \frac{1}{2} \pm \frac{1}{2^{n/2}}.$$

1.3 Fourier L_1 Norm

Definition 3 (**Fourier L_1 Norm**)**.** *Let $f : \{0,1\}^n \to \mathbb{R}$, the Fourier L_1 Norm is*

$$\boxed{L_1(f) = \sum_{S\subseteq[n]} \left|\widehat{f}(S)\right|}.$$

Proposition 4. *If $f : \{0,1\}^n \to [-1,1]$ (i.e. an interval instead of two discrete points), then we have*

$$L_1(f) \leq 2^{n/2}$$

Proof. Use Cauchy-Schwarz:

$$\boxed{L_1(f)} = \sum_{S\subseteq[n]} |\widehat{f}(S)| \overset{c\text{-}s}{\leq} \sqrt{2^n \cdot \sum_{S\subseteq[n]} \widehat{f}(S)^2} = \sqrt{2^n \cdot \mathbb{E}[f(\mathcal{U})^2]}\boxed{\leq \sqrt{2^n}},$$

since $\mathbb{E}[f(\mathcal{U})^2] \leq 1$. ■

1.4 Applications

Here, equipped with Fourier, we can better three things we already showed last time:

- Use △-inequality to come up with an explicit error bound for δ-biased RV: corollary 2.
- Fool DTs "Better" (not exactly comparable): proposition 5.
- Fool K-juntas better: proposition 7.

Lemma 1 (△-inequality)**.** *Recall the △-inequality for PRGs which we proved in the last lecture. Let $f_1, \ldots, f_t : \{0,1\}^n \to \mathbb{R}$. Let $\lambda_0, \lambda_1, \ldots, \lambda_t \in \mathbb{R}$. Let*

$$f(x) = \lambda_0 + \sum_{i=1}^{t} \lambda_i f_i(x).$$

If $\mathbf{X}$ ϵ_i-fools each f_i, for $i \in [t]$, then $\mathbf{X}$ ϵ-fools f, where

$$\epsilon = \sum_{i=1}^{t} |\lambda_i| \cdot \epsilon_i.$$

Now that we can think of Boolean functions as linear combination of character functions, but we also constructed a PRG that fools character functions last time (for ϵ-biased PRG)! So,

Corollary 2. *If $\mathbf{X}$ is δ-biased RV over $\{0,1\}^n$, then $\mathbf{X}$ $(\delta \cdot L_1(f))$-fools $f : \{0,1\}^n \to \mathbb{R}$.*

Official Homework Problem: Let $f : \{0,1\}^n \to \{0,1\}$ be a conjunction of literals over distinct variables,

$$e.g., x_1 \wedge \overline{x_3} \wedge \overline{x_4} \wedge x_6.$$

Show that

$$L_1(f) = 1.$$

Proposition 5. *Also, recall from last time that we were able to 0-fool depth-d DTs, using d-wise independence (with seed length of $d \cdot \log n$). Now, we can fool size-s DTs!*

Formally (modified this time): If $\mathbf{X}$ is a δ-biased RV over $\{0,1\}^n$, then $\mathbf{X}$ fools size-s DTs with error $\delta \cdot s$ (so, get ϵ-PRG for $\mathcal{DT}_s$ with seed length $O(\log s + \log n + \log \frac{1}{\epsilon})$).

Proof. Let

- T = size-s DT.
- L = the set of 1-leaves (leaves that evaluate to 1) on T.
- f_l = conjunctions that correspond to the 1-leaf, $l \in L$.

We have $f(x) = \sum\limits_{l \in L} f_l$, so

$$L_1(f) \leq \sum_{l \in L} L_1(f_l) \overset{\text{last OHP}}{\leq} |L| \leq s.$$

Finally, by corollary 2, the error that $\mathbf{X}$ fools f with is

$$\delta \cdot L_1(f) \leq \delta \cdot s, \text{ as desired.}$$

■

The PRG gives us an "input-oblivious" algorithm to do deterministic approximate counting, but, if we are willing to look at the input representation as a size-s decision tree (DT), T, then it's easy to do exact counting for decision trees:

Proposition 6. *Given a size-s DT, T, we can compute exactly*

$$\Pr_{\mathcal{U}}[T(\mathcal{U})=1]=\sum_{1\text{-}leaf, l\in L} 2^{-depth(l)}.$$

Proposition 7. *Now, let's fool K-juntas better (last time, we needed a seed length of $K\cdot\log n$). We do so by using K-wise, ϵ-biased.*

Formally (modified this time): If $\mathbf{X}$ is K-wise, ϵ-biased RV over $\{0,1\}^n$, then $\mathbf{X}$ fools $\{\pm 1\}$-valued K-juntas with error $\delta\cdot 2^{K/2}=\epsilon$, with seed length of $O(K+\log\frac{1}{\epsilon}+\log\log n)$.

Proof. By definition of K-juntas, we know that $f(x)=g(x_{i_1},\ldots,x_{i_k})$ for some $g:\{0,1\}^K\to\{\pm 1\}$. Now,

$$L_1(g)\leq 2^{K/2},$$

so $\mathbf{X}$ $\delta\cdot 2^{K/2}$-fools g (as well as f). ■

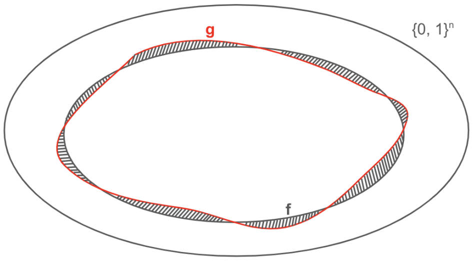
2 Sandwiching & Approximation

Suppose we have $f:\{0,1\}^n\to\{0,1\}$ which we want to fool. Suppose we have $g:\{0,1\}^n\to\{0,1\}$ that approximates f as well as

$$f(x)\neq g(x) \text{ for } \leq\epsilon\cdot 2^n \text{ inputs.}$$

Finally, suppose $\mathbf{X}$ as an RV ϵ-fools g. The problem is: **WE CANNOT SAY THAT X THUS ϵ-FOOLS f!**

To see why not, let's visualize the set-up: in the following illustration, the area enclosed by the red line represents the space for function g; the area enclosed by the grey line represents the space for function f. The grey-shaded areas are the total errors when g approximates f. Note that such error is said to be $\leq\epsilon\cdot 2^n$ points.



Think about the following situation to understand the issue better: say $\mathbf{X}$ is supported on $2^{\text{seed length}}$ number of points. Say, the seed length is $\frac{n}{10}$, then:

- $2^{n/10}$ points are a lot of points, but
- It is still very likely for $2^{n/10} < \epsilon \cdot 2^n$, since, in this case, ϵ actually needs to be a very very small value for the inequality to not be true.

In other words, it's very possible for all of $\mathbf{X}$'s support to be in the grey-shaded error region. So, in order for the above attempt to make $\mathbf{X}$ fooling g, an approximation, sufficient for $\mathbf{X}$ to fool f, we need a stronger notion of approximation, in the sandwiching sense:

Definition 4 (δ-sandwiched)**.** *Let $f, f_l, f_u : \{0,1\}^n \to \mathbb{R}$ ("l" for lower and "u" for upper), then f is δ-sandwiched by (f_l, f_u) if*

1). $f_l(x) \leq f(x) \leq f_u(x), \forall x \in \{0,1\}^n$, and

2). $\underset{\mathbf{x}\sim\mathcal{U}}{\mathbb{E}}[f_u(\mathbf{x}) - f_l(\mathbf{x})] \leq \delta$.

Lemma 2 (Sandwiching Lemma)**.** *Suppose $f : \{0,1\}^n \to \mathbb{R}$ is*

- *δ-sandwiched by (f_l, f_u), and*
- *suppose $\mathbf{X}$ (RV over $\{0,1\}^n$) ϵ-fools f_l and ϵ-fools f_u.*

Then, $\mathbf{X}$ $(\epsilon + \delta)$-fools f.

Proof. For the upper inequality, we have:

$$\begin{aligned}\underset{\mathbf{X}}{\mathbb{E}}[f(\mathbf{X})] &\leq \underset{\mathbf{X}}{\mathbb{E}}[f_u(\mathbf{X})] \\ &\leq \underset{\mathcal{U}}{\mathbb{E}}[f_u(\mathcal{U})] + \epsilon, \text{ by fooling condition} \\ &\leq \underset{\mathcal{U}}{\mathbb{E}}[f(\mathcal{U})] + \delta + \epsilon, \text{ by sandwiching condition.}\end{aligned}$$

Similarly, for the lower inequality:

$$\begin{aligned}\underset{\mathbf{X}}{\mathbb{E}}[f(\mathbf{X})] &\geq \underset{\mathbf{X}}{\mathbb{E}}[f_l(\mathbf{X})] \\ &\geq \underset{\mathcal{U}}{\mathbb{E}}[f_l(\mathcal{U})] - \epsilon, \text{ by fooling condition} \\ &\geq \underset{\mathcal{U}}{\mathbb{E}}[f(\mathcal{U})] - \delta - \epsilon, \text{ by sandwiching condition.}\end{aligned}$$

■

Corollary 3 (Sandwiching polynomials $\implies$ PRG)**.** *A function $f : \{0,1\}^n \to \mathbb{R}$ is ϵ-fooled by any K-wise independent distribution $\mathbf{X}$ if $\exists$ "ϵ-sandwiching" real polynomials $q_l, q_u : \{0,1\}^n \to \mathbb{R}$ of degree K s.t. the two requirements for definition 4 hold:*

1). $f_l(x) \leq f(x) \leq f_u(x), \forall x \in \{0,1\}^n$, and

2). $\mathop{\mathbb{E}}_{\mathbf{x}\sim\mathcal{U}}[f_u(\mathbf{x}) - f_l(\mathbf{x})] \leq \delta$.

3 Viola's Theorem

Firstly, let's motivate.

3.1 Fooling $\mathbb{F}_2$ polynomials with DEG$_d$

Note that real deg-d polynomials can be 0-fooled by d-wise independent RVs with a seed length of $d \cdot \log n$, which is easy. Fooling DEG$_d$ ($\mathbb{F}_2$ and deg-d) polynomials is very different! Here's why:

Example 2. *Let $\mathbf{X} = (\mathbf{X_1}, \ldots, \mathbf{X_n})$ and $\mathbf{X}$ is assigned in such a way:*

$$\begin{aligned}
\mathbf{X_1} &= \mathcal{U}_1\\
\mathbf{X_2} &= \mathcal{U}_2\\
&\vdots\\
\mathbf{X_{n-1}} &= \mathcal{U}_{n-1}\\
\mathbf{X_n} &= \mathcal{U}_1 \oplus \ldots \oplus \mathcal{U}_{n-1}.
\end{aligned}$$

Then, with seed length of $\mathbf{X} = n-1$, $\mathbf{X}$ is $(n-1)$-wise independent. We know that $\mathbf{X}$ cannot fool $\mathrm{PAR}_{[n]} \in \mathrm{DEG}_1$, because $\mathrm{PAR}_{[n]} = 0$ on every $\mathbf{X}$. Therefore, $\mathbf{X}$ fails to fool even DEG_1.

3.1.1 Why is it interesting to fool DEG$_d$ after all (now that we have shown it's different)?

It is because fooling DEG$_d$ gives us insights about other things that we don't know how to show.

Definition 5 ($AC^0(\oplus)$)**.** *$AC^0(\oplus)$ is the set of all poly(n)-size, $O(1)$-depth circuits with this set of gates: $\{\wedge, \vee, \neg, \oplus\}$ ($\oplus$ is parity gates).*

Proposition 8. *$\exists\epsilon$-PRG against DEG_d for $d = (\log n)^{\omega(1)} \implies \exists 3\epsilon$-PRG against $AC^0(\oplus)$.*

Proof. Let $G : \{0,1\}^s \to \{0,1\}^n$ be ϵ-PRG against DEG_d (where $d = (\log n)^{\omega(1)}$ as we started with). Let's first see a fact (to be proved later):

Fact 1. *For any $O(1)$-depth, $\mathrm{poly}(n)$-size circuits C (with $\{\wedge, \vee, \neg, \oplus\}$ gates), there's a distribution $\mathcal{P}$ over $\text{deg-}(\mathrm{poly}\log(n))$ $\mathbb{F}_2$ polynomials, s.t.*

$$\forall z \in \mathbb{F}_2^n, \Pr_{p\sim\mathcal{P}}[p(z) = C(z)] \geq 1 - \epsilon.$$

In other words: distribution of polynomials fools C on every fixed input.

(Back to the proof of the proposition). *This means that*

$$\Pr_{\mathcal{U}_s}[C(G(\mathcal{U}_s)) = 1] \approx_\epsilon \Pr_{\mathcal{U}_s, p\sim\mathcal{P}}[p(G(\mathcal{U}_s)) = 1].$$

G is an ϵ-PRG for DEG_d, and $d > \deg$ of polynomials in $\mathcal{P}$, so

$$\Pr_{\mathcal{U}_s, p\sim\mathcal{P}}[p(G(\mathcal{U}_s)) = 1] \approx_\epsilon \Pr_{\mathcal{U}_n, p\sim\mathcal{P}}[p(\mathcal{U}_n) = 1].$$

Since $\mathcal{P}$ fools C on every fixed input, we have

$$\Pr_{\mathcal{U}_n, p\sim\mathcal{P}}[p(\mathcal{U}_n) = 1] \approx_\epsilon \Pr_{\mathcal{U}_n}[C(\mathcal{U}_n) = 1].$$

Connecting all these together:

$$\Pr_{\mathcal{U}_s}[C(G(\mathcal{U}_s)) = 1] \approx_{3\epsilon} \Pr_{\mathcal{U}_n}[C(\mathcal{U}_n) = 1].$$

■

3.2 Viola's Theorem [Vio09]

Now that we are fully motivated about **wanting PRGs for DEG$_d$**. The line of works in this direction culminated in the Viola's theorem.

Theorem 1 (Viola's Theorem). *The $\mathbb{F}_2$ sum of d independent δ-biased RVs fools DEG_d.*

Formally: Let $\mathbf{Y_1}, \ldots, \mathbf{Y_d}$ be independent δ-biased RVs over $\mathbb{F}_2^n$, where $\delta \leq \frac{1}{2}$. Then,

$$\boxed{\mathbf{Y} := \mathbf{Y_1} + \ldots + \mathbf{Y_d} \text{ (over } \mathbb{F}_2\text{) } \left[4\left(\delta/2\right)^{2^{\frac{1}{d-1}}}\right] \text{-fools } \mathrm{DEG}_d}.$$

Note that, when we take $\boxed{\epsilon = 4\left(\delta/2\right)^{2^{\frac{1}{d-1}}}}$*, we get a PRG with seed length*

$$\boxed{O\left(d \cdot \log\left(\frac{n}{\delta}\right)\right) = O\left(d \cdot \log n + d \cdot 2^d \cdot \log\frac{1}{\epsilon}\right)},$$

which trivializes when $d > \log(n)$ [which means that ***it breaks exactly where the correlation bound broke****].*

Observation 1. *Let $\mathbf{Y_1}, \ldots, \mathbf{Y_d}$ be independent δ-biased RVs over $\mathbb{F}_2^n$. Then, $\mathbf{Y} := \mathbf{Y_1} + \ldots + \mathbf{Y_d}$ is δ^d-biased.*

Proof. Let $\emptyset \neq S \subseteq [n]$. Then,

$$\begin{aligned} \left|\mathop{\mathbb{E}}_{\mathbf{Y}}\left[\chi_S\left(\sum_{i=1}^{d}\mathbf{Y_i}\right)\right]\right| &= \left|\mathop{\mathbb{E}}_{\mathbf{Y}}\left[(-1)^{\sum_{j\in S}\sum_{i=1}^{d}\mathbf{Y_{i,j}}}\right]\right| \\ &= \left|\mathop{\mathbb{E}}_{\mathbf{Y}}\left[(-1)^{\sum_{i=1}^{d}\sum_{j\in S}\mathbf{Y_{i,j}}}\right]\right| \\ &= \left|\prod_{i=1}^{d}\mathop{\mathbb{E}}_{\mathbf{Y}}[\chi_S(\mathbf{Y_i})]\right| \leq \delta^d, \text{ because they are independent.} \end{aligned}$$

■

Remark 4. *Viola's theorem also tells us that the sum, $\mathbf{Y_1}+\ldots+\mathbf{Y_d}$, fools higher-degree polynomials.*

3.2.1 The Main Result: High-Level Strategy

Show that, if we can fool DEG_{d-1}, then we can add one more biased RV and trade in for some worse parameters to fool DEG_d (a colloquial equivalent of key lemma 3). In particular, if we can prove the following key lemma, then we can prove the theorem quite easily:

Lemma 3 (**Key Lemma**)**.** *Suppose* $\mathbf{W}$ *fools* DEG_{i-1} *with error* γ *and suppose* $\mathbf{Y}$ *is a* δ*-biased RV independent of* $\mathbf{W}$*. Then,* $\mathbf{W}+\mathbf{Y}$ $\left(\sqrt{2\gamma}+\frac{\delta}{2}\right)$*-fools* DEG_i *[note that* $\left(\sqrt{2\gamma}+\frac{\delta}{2}\right)$ *is where we traded in for a worse parameter than* γ*].*

Now, we can prove the theorem assuming the "key lemma" is correct:

Proof. (Viola's Theorem 1, using Key Lemma). By definition of δ-biased, we have that $\mathbf{Y_1}$ $\frac{\delta}{2}$-fools DEG_1. Let $\epsilon_1 = \frac{\delta}{2}$, and $\epsilon_{i+1} = \sqrt{2\epsilon_i} + \frac{\delta}{2}$. Then, by key lemma 3, we have

$$\mathbf{Y_1} + \ldots + \mathbf{Y_d} \ \epsilon_d\text{-fools } \mathrm{DEG}_d.$$

Since $\delta \leq \frac{1}{2}$, we further have

$$\epsilon_{i+1} \leq \sqrt{2\epsilon_i} + \frac{\sqrt{\delta/2}}{2} \underset{\delta/2\leq\epsilon_i}{\leq} \left(\sqrt{2}+\frac{1}{2}\right)\sqrt{\epsilon_i} \leq 2\sqrt{\epsilon_i}.$$

So,

$$\begin{aligned}
\epsilon_2 &\leq 2\cdot(\delta/2)^{\frac{1}{2}}\\
\epsilon_3 &\leq 2^{1+\frac{1}{2}}\cdot(\delta/2)^{\frac{1}{4}}\\
&\vdots\\
\boxed{\epsilon_d} &\leq 2^{1+\frac{1}{2}+\ldots+\frac{1}{2^{d-2}}}\cdot(\delta/2)^{\frac{1}{2^{d-1}}} \boxed{< 4\cdot(\delta/2)^{\frac{1}{2^{d-1}}}},
\end{aligned}$$

as desired. ■

3.2.2 Prove Key Lemma 3: Idea by Case Analysis

We show using case analysis depending on the "**imbalance**" (see definition 6) of the $f \in \mathrm{DEG}_i$ function we are trying to fool.

Definition 6 (imbal(f))**.**

$$\mathit{imbal}(f) = \left|\underset{\mathcal{U}}{\mathbb{E}}\left[(-1)^{f(\mathcal{U})}\right]\right| = 2\left|\mathbb{E}[f] - \frac{1}{2}\right|.$$

This value is always in $\in [0,1]$ *because:*

- *(*f *always* 0*) You get* $2\left|0-\frac{1}{2}\right| = 1$.
- *(*f *always* 1*) You get* $2\left|1-\frac{1}{2}\right| = 1$.

- *($f \in \{0,1\}$ uniformly) You get* $2\left|\frac{1}{2}-\frac{1}{2}\right| = 0$.

Intuition 1. *We can sense hardness from the value of* imbal *(hard functions should be more balanced than easier functions). Here are some intuitions about* imbal*, for* $f \in \mathrm{DEG}_i$*:*

- *If f has large* imbal(f)*, it turns out* $\mathbf{W}$ *already fools f quite well (because f is biased / not too hard).*

- *If f has small* imbal(f)*, we will do an analysis like the correlation bound analysis squaring of correlation, derivatives, etc.*

So, we do the two cases for the case analysis based on intuition 1 next.

3.2.3 Prove Key Lemma 3: The Imbalanced Case

Lemma 4 (Imbalanced Case Lemma)**.** *Suppose* $\mathbf{W}$ γ*-fools* DEG_{i-1}*. Then,* $\mathbf{W}$ *fools any* $f \in \mathrm{DEG}_i$ *with error* $\frac{\gamma}{\mathsf{imbal}(f)}$.

Notation 1. $f^{+y}(x) = f(x+y)$.

Definition 7 (Directional Derivative)**.** *Recall, for* $f : \mathbb{F}_2^n \to \mathbb{F}_2$, $y \in \mathbb{F}_2^n$*, the* ***directional derivative*** $\partial_y f$ *is*

$$\partial_y f : \mathbb{F}_2^n \to \mathbb{F}_2, \partial_y f(x) = f^{+y}(x) + f(x).$$

It's easy to see that, if f is deg-i*, then* $\forall y, \partial_y f$ *has* deg *at most* $(i-1)$.

Now, the proof for lemma 4

Proof. (Imbalanced Case Lemma). Need to show that

$$\mathsf{imbal}(f) \cdot \left|\mathbb{E}\left[(-1)^{f(\mathbf{W})}\right] - \mathbb{E}\left[(-1)^{f(\mathcal{U}')}\right]\right| \leq 2\gamma,$$

where $\mathbf{W}$ is as defined and $\mathcal{U}'$ is truly uniform.

Let $\mathcal{U}$ be independent, uniform. We have that

$$\begin{aligned}
&\mathsf{imbal}(f) \cdot \left|\mathbb{E}\left[(-1)^{f(\mathbf{W})}\right] - \mathbb{E}\left[(-1)^{f(\mathcal{U}')}\right]\right| \\
&= \left|\mathbb{E}\left[(-1)^{f(\mathbf{W})+f(\mathcal{U})}\right] - \mathbb{E}\left[(-1)^{f(\mathcal{U}')+f(\mathcal{U})}\right]\right|, \text{ by definition of } \mathsf{imbal} \\
&= \left|\mathbb{E}\left[(-1)^{f(\mathbf{W})+f(\mathbf{W}+\mathcal{U})}\right] - \mathbb{E}\left[(-1)^{f(\mathcal{U}')+f(\mathcal{U}'+\mathcal{U})}\right]\right|, \text{ true uniformity shifted still uniform} \\
&= \left|\mathbb{E}\left[(-1)^{f(\mathbf{W})+f^{+\mathcal{U}}(\mathbf{W})}\right] - \mathbb{E}\left[(-1)^{f(\mathcal{U}')+f^{+\mathcal{U}}(\mathcal{U}')}\right]\right| \\
&= \left|\mathbb{E}\left[(-1)^{\partial_{\mathcal{U}} f(\mathbf{W})}\right] - \mathbb{E}\left[(-1)^{\partial_{\mathcal{U}} f(\mathcal{U}')}\right]\right|.
\end{aligned}$$

But, both of the derivatives on the RHS would have deg-$(i-1)$, so they must respectively be γ-fooled by $\mathbf{W}$ because of condition for DEG_{i-1}. So,

$$\mathsf{imbal}(f)\cdot\left|\mathbb{E}\left[(-1)^{f(\mathbf{W})}\right]-\mathbb{E}\left[(-1)^{f(\mathcal{U}')}\right]\right| = \left|\mathbb{E}\left[(-1)^{\partial_{\mathcal{U}}f(\mathbf{W})}\right]-\mathbb{E}\left[(-1)^{\partial_{\mathcal{U}}f(\mathcal{U}')}\right]\right| \leq 2\gamma.$$

■

3.2.4 Prove Key Lemma 3: The Balanced Case

Lemma 5 (Balanced Case Lemma)**.** *Suppose* $\mathbf{W}$ *γ-fools* DEG_{i-1}*. Let* $\mathbf{Y}$ *be independent of* $\mathbf{W}$*, and* $\mathbf{Y}$ *is δ-biased. Then,*

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

Proof. (Balanced Case Lemma). To be shown next time! ■

3.2.5 Prove Key Lemma 3 by Putting Lemma 4 and Lemma 5 together!

Finally, we use lemma 4 and lemma 5 to prove the key lemma 3:

Proof. (Key Lemma 3). Fix any $f\in\mathrm{DEG}_i$. Fix any outcome of $\mathbf{Y}$. The function $f^{+\mathbf{Y}}(x)=f(\mathbf{Y}+x)$ is a deg-i polynomial in $x_1,\ldots,x_n$ and $\mathsf{imbal}(f^{+\mathbf{Y}})=\mathsf{imbal}(f)$.

- So, by lemma 4, $\mathbf{W}+\mathbf{Y}$ $\mathsf{L1}:=\left(\frac{\gamma}{\mathsf{imbal}(f)}\right)$-fools f.
- By lemma 5, $\mathbf{W}+\mathbf{Y}$ $\mathsf{L2}:=\left(\mathsf{imbal}(f)+\sqrt{\frac{\gamma}{2}}+\frac{\delta}{2}\right)$-fools f.

Hence, $\mathbf{W}+\mathbf{Y}$ $\min\{L1,L2\}$-fools f, and

$$\min\{L1,L2\}\leq\sqrt{2\gamma}+\frac{\delta}{2}.$$

■

4 Next Time:

- Prove lemma 5, which is the hard case (which finishes the Viola's theorem).
- Start the proof that PRG fools AC^0.

References

[O'D14] Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, New York, 2014. 1

[Vio09] Emanuele Viola. The sum of D small-bias generators fools polynomials of degree D. *Computational Complexity*, 18(2):209–217, 2009. 3.2