CS 6817: Special Topics in Complexity Theory **Spring 2025**

Lecture 14: Mar 18, 2025

Lecturer: Eshan Chattopadhyay *Scribe: August Chen*

1 Better Bounds on the Fourier Spectrum of DNFs via Random Restrictions

Recall the definition of random restrictions from last lecture. We sample a random restriction $\rho = (J, z)$ where $J \subseteq [n]$ and $z \in \{\pm 1\}^n$ from the distribution of δ-random restrictions, denoted $\rho_{n,\delta}$. Here $\rho_{n,\delta}$ is the following distribution over (J, z): for all $1 \le i \le n$, we place $i \in J$ with probability δ, independently over all i. We sample $z \sim \{\pm 1\}^n$ uniformly. Here J represents the coordinates that are not fixed, and z represents how we fix the coordinates outside J.

Also recall from last lecture, for $f : \{\pm 1\}^n \to \mathbb{R}$ and some $\rho = (J, z)$, we consider the function f under the restriction ρ, defined as follows:

$$f_\rho(x) := f(x_J, z_{J^c}).$$

In the following, we will consider $\rho \sim \rho_{n,\delta}$ (i.e. the random restriction of f).

By direct calculations from first principles, one can readily check the following, which we did last class:

Claim 1.1. *For all $S \subseteq [n]$, $\mathbb{E}_{\rho \sim \rho_{n,\delta}}\left[\widehat{f_\rho}(S)\right] = \delta^{|S|}\hat{f}(S)$.*

Claim 1.2. *For all $S \subseteq [n]$, $\mathbb{E}_{\rho \sim \rho_{n,\delta}}\left[\widehat{f_\rho}(S)^2\right] = \sum_{W \subseteq [n]} \hat{f}(W)^2 \mathbb{P}_J(W \cap J = S)$.*

Our goal in this section is for size s DNFs f, to find a bound on k, for which we have ε-concentration up to degree k. From previous lectures, we have seen for $k = O\left(\log\left(\frac{s}{\varepsilon}\right)\frac{1}{\varepsilon}\right)$, we have $W^{\ge k}[f] \le \varepsilon$. Here, we aim to have all dependence on ε in k to be polylogarithmic. In particular we aim for the following:

Goal 1. *For size s DNFs f, for $k = O\left(\log\left(\frac{s}{\varepsilon}\right)\log\left(\frac{1}{\varepsilon}\right)\right)$, we have $W^{\ge k}[f] \le \varepsilon$.*

The above will yield improved bounds for PAC learning algorithms in the query model, as discussed in the prior lecture.

To show Goal 1, we will use the following hammer:

Theorem 1 (Hastad's Switching Lemma)**.** *Suppose f is computable by a width w DNF. Then for all $\delta \le \frac{1}{7}$, $d \ge 0$, we have*

$$\mathbb{P}_{\rho \sim \rho_{n,\delta}}\left(DT_{depth}(f_\rho) \ge d\right) \le (7\delta w)^d.$$

With Hastad's Switching Lemma, we can establish Goal 1:

Theorem 2. *Let f be computable by a size s DNF. Then*

$$W^{\ge k}[f] \le \varepsilon \text{ for } k = O\left(\log\left(\frac{s}{\varepsilon}\right)\log\left(\frac{1}{\varepsilon}\right)\right).$$

Proof. We will first do the following standard approximation argument, which is by now in the course very familiar. By Proposition 1.1, Lecture 13, we can approximate f by a narrow approximator g such that f and g are ε_1 close, where g is computable by a width $w = \log\left(\frac{s}{\varepsilon_1}\right)$ DNF. Suppose we show that $W^{\geq k}[g] \leq \varepsilon_1$ for $k = O\left(\log\left(\frac{s}{\varepsilon_1}\right)\log\left(\frac{1}{\varepsilon_1}\right)\right)$. Choosing $\varepsilon_1 = \frac{\varepsilon}{2}$ (say) and using that f, g are ε_1 close, we obtain the desired conclusion for f.

Now to establish that $W^{\geq k}[g] \leq \varepsilon_1$ for $k = O\left(\log\left(\frac{s}{\varepsilon_1}\right)\log\left(\frac{1}{\varepsilon_1}\right)\right)$ for g, fix $\delta = \frac{1}{14w}$, where $w = \log\left(\frac{s}{\varepsilon_1}\right)$. Consider for any k_1 to be set later,

$$\mathbb{E}_{\rho\sim\rho_{n,\delta}}\left[W^{\geq k_1}[g_\rho]\right] = \sum_{\rho} \mathbb{P}_{\rho\sim\rho_{n,\delta}}\left(\rho_{n,\delta} = \rho\right) W^{\geq k_1}[g_\rho]. \tag{1}$$

The motivation behind considering this quantity is that we can understand it via our Fourier formulas Claim 1.1, Claim 1.2.

We first find an upper bound on (1), aiming to relate it to $DT_{\text{depth}}(g_\rho)$ and then apply Hastad's Switching Lemma. Consider $DT_{\text{depth}}(g_\rho)$. If $DT_{\text{depth}}(g_\rho) < k_1$, then $W^{\geq k_1}[g_\rho] = 0$. Else, since g is computable by a DNF it is Boolean, so Parseval's gives $W^{\geq k_1}[g_\rho] \leq 1$. This observation, together with applying Hastad's Switching Lemma, yields

$$\mathbb{E}_{\rho\sim\rho_{n,\delta}}\left[W^{\geq k_1}[g_\rho]\right] = \sum_{\rho} \mathbb{P}_{\rho\sim\rho_{n,\delta}}\left(\rho_{n,\delta} = \rho\right) W^{\geq k_1}[g_\rho] \leq \mathbb{P}_{\rho\sim\rho_{n,\delta}}\left(DT_{\text{depth}}(g_\rho) \geq k_1\right) \cdot 1 \leq (7\delta w)^{k_1} = \left(\frac{1}{2}\right)^{k_1}, \tag{2}$$

where the last step uses the definition of δ.

Now choose $k_1 = \log\left(\frac{1}{\varepsilon_1}\right) + 1$. With this choice of k_1, we aim to relate $\mathbb{E}_{\rho\sim\rho_{n,\delta}}\left[W^{\geq k_1}[g_\rho]\right]$ to concentration on g. Repeatedly swapping the order of summation and applying Claim 1.2 in the third equality, we obtain

$$\begin{aligned}
\mathbb{E}_{\rho}\left[W^{\geq k_1}[g_\rho]\right] &= \mathbb{E}_{\rho}\left[\sum_{S:|S|\geq k_1} \widehat{g_\rho}(S)^2\right] \\
&= \sum_{S:|S|\geq k_1} \mathbb{E}_{\rho}\left[\widehat{g_\rho}(S)^2\right] \\
&= \sum_{S:|S|\geq k_1} \sum_{W\subseteq[n]} \hat{g}(W)^2 \mathbb{P}_J(W \cap J = S) \\
&= \sum_{W\subseteq[n]} \hat{g}(W)^2 \mathbb{P}_J(|W \cap J| \geq k_1).
\end{aligned} \tag{3}$$

Now consider

$$k_2 = C_1 \cdot \frac{k_1}{\delta},$$

where C_1 is a sufficiently large universal constant. Note for any W with $|W| \geq k_2$, we have

$$\mathbb{E}_J[|W \cap J|] = |W| \cdot \delta \geq C_1 k_1.$$

A straightforward application of Chernoff bounds (this is where we use that C_1 is a large enough universal constant; C_1 being a *universal* constant independent of δ follows from the fact that the variance of a Bernoulli with parameter $\delta \in [0,1]$ is at most $\delta - \delta^2 \leq \frac{1}{4}$ and that $k_1 \geq 1$) now gives for any W with $|W| \geq k_2$,

$$\mathbb{P}_J(|W \cap J| \geq k_1) \geq \frac{1}{2}.$$

Combining with (3), (2) yields

$$\left(\frac{1}{2}\right)^{k_1} \geq \sum_{W \subseteq [n]} \hat{g}(W)^2 \mathbb{P}_J(|W \cap J| \geq k_1) \geq \frac{1}{2} \sum_{W \subseteq [n]: |W| \geq k_2} \hat{g}(W)^2 = \frac{1}{2} W^{\geq k_2}[g].$$

Recalling our choice of k_1, we obtain

$$\frac{1}{2}\varepsilon_1 \geq \frac{1}{2} W^{\geq k_2}[g] \implies W^{\geq k_2}[g] \leq \varepsilon_1.$$

Recalling the definition of k_2, δ, w, and ε_1, it is evident that

$$k_2 = O\left(\log\left(\frac{s}{\varepsilon_1}\right) \log\left(\frac{1}{\varepsilon_1}\right)\right).$$

Recalling our earlier remarks completes the proof. □

The aforementioned *technique* of random restrictions is quite natural, and we just saw its usefulness. The spectrum of g_ρ becomes nice, after 'going up' by a factor of $\frac{1}{\delta} = O\left(\log\left(\frac{s}{\varepsilon_1}\right)\right)$ from k_1.

2 Hypercontractivity and Applications

We now start a new topic, *hypercontractivity and its applications*, a topic originally from functional analysis.

Recall the T_ρ operator on any $f : \{\pm 1\}^n \to \mathbb{R}$, defined as follows. We first define the noisy distribution $N_\rho(x)$ for a given $x \in \{\pm 1\}^n$ and parameter $\rho \in [-1, 1]$ as follows: to sample $y \sim N_\rho(x)$, independently for each i, we let

$$y_i = \begin{cases} x_i & : \text{ with probability } \frac{1}{2} + \frac{\rho}{2} \\ -x_i & : \text{ with probability } \frac{1}{2} - \frac{\rho}{2}. \end{cases}$$

We then define the *noise operator* T_ρ (applied on functions) as follows. When the noise operator is applied to f, we define its output by the function $T_\rho f$, which is defined as follows:

$$T_\rho f(x) := \mathbb{E}_{y \sim N_\rho(x)}[f(y)].$$

Our intuition is that '$T_\rho f$ is smoother than f', and we make this quantitative in the following.

To do so, we introduce the following standard definition:

Definition 1 (L_p norms)**.** *For $p \geq 1$, we define the L_p norm of $f : \{\pm 1\}^n \to \mathbb{R}$ by*

$$\|f\|_p = \mathbb{E}_{x \sim \{\pm 1\}^n}[|f(x)|^p]^{1/p}.$$

Considering any $f, g : \{\pm 1\}^n \to \mathbb{R}$ and $p \geq q \geq 1$, we have the following standard properties:

- Triangle Inequality:
$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$
- Monotonicity:
$$\|f\|_p \geq \|f\|_q.$$

- Hölder: For p, q conjugate exponents (that is, $\frac{1}{p} + \frac{1}{q} = 1$; note this allows $p = \infty, q = 1$):

$$\langle f, g \rangle \leq \|f\|_p \|g\|_q.$$

We can easily prove the following smoothening effect of the noise operator T_ρ:

Claim 2.1. *For any $f : \{\pm 1\}^n \to \mathbb{R}$ and $\rho \in [-1, 1]$,*

$$\|T_\rho f\|_2 \leq \|f\|_2.$$

That is, T_ρ is a contraction in the L_2 norm.

Proof. As established in Lecture 6, we have $\widehat{T_\rho f}(S) = \hat{f}(S)\rho^{|S|}$. Now notice as $|\rho| \leq 1$, irrespective of its sign,

$$\begin{aligned}
\|T_\rho f\|_2^2 &= \sum_{S \subseteq [n]} \widehat{T_\rho f}(S)^2 \\
&= \sum_{S \subseteq [n]} \hat{f}(S)^2 \rho^{2|S|} \\
&\leq \sum_{S \subseteq [n]} \hat{f}(S)^2 = \|f\|_2^2.
\end{aligned}$$

□

However, we want to establish more, that T_ρ is *hypercontractive*: colloquially, $T_\rho f$ is *much* smoother than f. Here, we will quantify this by showing $T_\rho f$ is smaller than f even when $T_\rho f$ is measured with a larger norm. We will see proofs in the following lecture.

Theorem 3 ((2, 4) Hypercontractivity Theorem)**.** *For all $f : \{\pm 1\}^n \to \mathbb{R}$ and for $\rho = \frac{1}{\sqrt{3}}$, we have*

$$\|T_\rho f\|_4 \leq \|f\|_2.$$

The above can be proven using first principles.
We also have:

Theorem 4 (($\frac{4}{3}$, 2) Hypercontractivity Theorem)**.** *For all $f : \{\pm 1\}^n \to \mathbb{R}$ and for $\rho = \frac{1}{\sqrt{3}}$, we have*

$$\|T_\rho f\|_2 \leq \|f\|_{4/3}.$$

Notice the above two Theorems are defined as the (p, p') Hypercontractivity Theorem for $p < p'$. The typically smaller p-norm $\|\cdot\|_p$ is used to measure f. The smoothening effect of T_ρ is strongly asserted, as the p-norm in fact upper bounds the typically larger p'-norm $\|\cdot\|_{p'}$, which we use to measure $T_\rho f$.

The $(\frac{4}{3}, 2)$ Hypercontractivity Theorem in fact follows from the $(2, 4)$ Hypercontractivity Theorem by Hölder's Inequality. Moreover, it turns out the $(\frac{4}{3}, 2)$ Hypercontractivity Theorem is enough to prove the KKL Theorem.