Computer Science 294 Lecture 10 Notes

Daniel Raban

February 16, 2023

1 Fourier Concentration of DNFs

1.1 Recap: DNFs and random restrictions

Recall that a DNF looks like

$$(x_{i_1} \wedge x_{i_2} \wedge \cdots \wedge x_{i_w}) \vee (x_{j_1} \wedge \cdots) \vee \cdots,$$

where the width is the maximum number of literals per term and the size is the number of terms. Last time, we proved the following proposition.

Proposition 1.1. *Suppose f is computable by a width w DNF. Then $\mathbb{I}(f) \leq 2w$.*

The connection between size and width is given by the fact that every size s DNF is ε-close to a width $\log(s/\varepsilon)$ DNF.

Our goal is to prove Mansour's theorem.

Theorem 1.1 (Mansour). *Width w DNFs are ε-concentrated on at most $w^{O(w\log(1/\varepsilon))}$ coefficients. All these coefficients are up to degree $O(w\log(1/\varepsilon))$.*

We will prove it next time. Today, we will prove an important ingredient, the LMN lemma.

Lemma 1.1 (LMN). *Width w DNFs are ε-concentrated up to degree $O(w\log(1/\varepsilon))$.*

To prove Mansour's theorem, we introduced the method of random restrictions: Instead of analyzing $f : \{\pm 1\}^N \to \{\pm 1\}$ directly, consider the rest of f for variables $J \subseteq [n]$ by assigning $\overline{J}$ with $z \in \{\pm 1\}^{\overline{J}}$. This gives $f_{J,z} : \{\pm 1\}^J \to \{\pm 1\}$ with

$$f_{J,z}(y) = f(\underbrace{y}_{J}, \underbrace{z}_{\overline{J}}).$$

We also think of this function as $f_{J,z} : \{\pm 1\}^J \to \{\pm 1\}$ as

$$f_{J,z}(x) = f(x_J, z).$$

We considered p-random restrictions with the distribution $\mathcal{R}_p$ given by picking each $i \in [n]$ to be in J independently with probability p and picking $z \in \{\pm 1\}^{\overline{J}}$ uniformly at random.

We proved the following facts about the Fourier coefficients of random restrictions.

Lemma 1.2. *For $0 < p < 1$ and $S \subseteq [n]$,*

$$\mathbb{E}_{(J,Z)\sim\mathcal{R}_p}[\widehat{f_{J,Z}}(S)] = \widehat{f}(S)p^{|S|},$$

$$\mathbb{E}_{(J,Z)\sim\mathcal{R}_p}[\widehat{f_{J,Z}}(S)^2] = \sum_{U\subseteq[n]} \widehat{f}(U)^2 \mathbb{P}_J(U \cap J = S) = \sum_{U \supseteq S} \widehat{f}(U)^2 p^{|S|}(1-p)^{|U\setminus S|}.$$

1.2 Influence of DNFs

Let's give another interpretation of this last equality. Recall that $\mathscr{S}_f$ denotes the Fourier distribution of f:

$$\mathbb{P}(U) = \widehat{f}(U)^2.$$

Then $\mathscr{S}_{f,p}$ is the distribution over the set $S \subseteq [n]$ given by

$$\mathbb{P}(S) = \mathbb{E}_{(J,Z)\sim\mathcal{R}_p}[\widehat{f_{J,Z}}(S)^2].$$

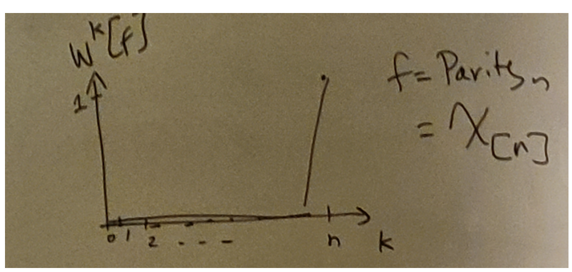
We can obtain the distribution $\mathscr{S}_{f,p}$ by the following steps:

1. Sample $U \sim \mathscr{S}_f$
2. Sample $J \subseteq_p [n]$ (each $i \in J$ independently with probability p).
3. Output $U \cap J$.

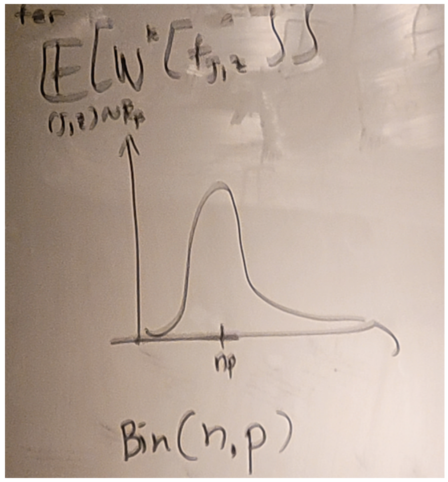
Now recall the weight

$$W^k(f) = \sum_{\substack{U\subseteq[n],\\ |U|=k}} \widehat{f}(U)^2.$$

The weight of the parity function looks like



The expected weight under a random restriction will behave like a Binomial(n, p) distribution



In general, the Fourier spectrum will look like an average, over the original weights, of binomials with parameters (k, p).

Theorem 1.2. *If f is a size s DNF, then $\mathbb{I}(f) \le O(\log S)$.*

Lemma 1.3.

$$\mathbb{E}_{(J,Z)\sim\mathcal{R}_p}[\mathbb{I}(f_{J,Z})] = p\mathbb{I}(f).$$

Proof.

$$\begin{aligned}
\mathbb{E}_{(J,Z)\sim\mathcal{R}_p}[\mathbb{I}(f_{J,Z})] &= \mathbb{E}_{J,Z}\left[\sum_{S\subseteq[n]} |S| \cdot \widehat{f_{J,Z}}(S)^2\right] \\
&= \sum_{S\subseteq[n]} |S|\,\mathbb{E}_{J,Z}\,\widehat{f_{J,Z}}(S)^2] \\
&= \sum_{S\subseteq[n]} |S| \sum_{U\subseteq[n]} \widehat{f}(U)^2 \cdot \mathbb{P}_J(J\cap U = S) \\
&= \sum_{U\subseteq[n]} \widehat{f}(U)^2 \sum_{S\subseteq[n]} |S| \cdot \mathbb{P}_J(J\cap U = S) \\
&= \sum_{U\subseteq[n]} \widehat{f}(U)^2\,\mathbb{E}_J[|J\cap U|] \\
&= \sum_{U\subseteq[n]} \widehat{f}(U)^2 \cdot p \cdot |U| \\
&= p\cdot\mathbb{I}(f).
\end{aligned}$$

□

Remark 1.1. Another way to think of this proof is via the distribution $\mathscr{S}_{f,p}$:

$$\mathbb{E}_{S\sim\mathscr{S}_{f,p}}[|S|] = \mathbb{E}_{\substack{U\sim\mathscr{S}_f, \\ J\subseteq_p[n]}}[|U\cap J|] = \mathbb{E}_{U\sim\mathscr{S}_f}[p\cdot|U|] = p\mathbb{I}(f).$$

Now we want to understand the width of a restriction.

Lemma 1.4. *If $(J,Z)\sim\mathcal{R}_{1/2}$ and f is a size s DNF, then for all w,*

$$\mathbb{P}(\text{width}(f_{J,Z})\geq w)\leq s\cdot\left(\frac{3}{4}\right)^w.$$

Proof. Write

$$f = T_1\vee T_2\vee T_3\vee\cdots\vee T_s,$$

where s is the size of the DNF and each T_i is an AND of literals. By a union bound, it suffices to show that for each term T_i,

$$\mathbb{P}(\text{width}((T_i)_{J,z})\geq w)\leq\left(\frac{3}{4}\right)^w.$$

If width($T_i < w$), this is true automatically. If width(T_i) $\geq w$, then express

$$T_i = (x_{i_1}\wedge x_{i_2}\wedge\cdots\wedge x_{i_{w'}}),$$

where $w'\geq w$. For $(T_i)_{J,Z}$ to not equal False, all literals must be either alive or assigned true. So

$$\mathbb{P}((T_i)_{J,Z}\not\equiv\text{False}) = \left(\frac{3}{4}\right)^{w'}\leq\left(\frac{3}{4}\right)^w.$$

Since width($f_{J,Z}$) $\geq w$ implies that $(T_i)_{J,Z}\not\equiv$ False, we get the bound. □

Now we will use all these tools to prove the theorem:

Proof. Pick $p = 1/2$. Then

$$\begin{aligned}
\mathbb{I}(f) &= \frac{1}{1/2}\,\mathbb{E}_{(J,Z)\sim\mathcal{R}_{1/2}}[\mathbb{I}(f_{J,Z})] \\
&\leq 2\,\mathbb{E}_{J,Z\sim\mathcal{R}_{1/2}}[\text{width}(f_{J,Z})] \\
&= 2\sum_{w=1}^{\infty}\mathbb{P}(\text{width}(f_{J,Z})\geq w) \\
&\leq 2\left(\sum_{w\leq 3\log s} 1 + \sum_{w\geq 3\log s} s\left(\frac{3}{4}\right)^w\right) \\
&\leq 6\log s + 2s\left(\frac{3}{4}\right)^{3\log s}\cdot 4 \\
&\leq 6\log s + o(1).
\end{aligned}$$

□

1.3 The Håstad switching lemma and the LMN lemma

We can reduce the dependence on the size s with the following remarkable lemma.

Lemma 1.5 (Håstad switching lemma)**.** *Suppose $f : \{\pm 1\}^n \to \{\pm 1\}$ is computable by a width w DNF, nd let $(J, Z) \sim \mathcal{R}_p$. Then for any k,*

$$\mathbb{P}_{J,Z}(\textit{decision tree depth}(f_{J,Z}) \geq k) = (5pw)^k.$$

Think of p as being something like $1/(10w)$. This lemma says that DNFs with very high probability can be represented by a shallow decision tree. For $p = 1/(10w)$ and $k = 1$, we get that with probability $1/2$, the DNF becomes a constant function! We will prove the switching lemma next time. For now, we want to prove the lemma of LMN.

Claim 1: For all k,

$$\mathbb{E}_{(J,Z)\sim\mathcal{R}_p}[W^{\geq k}(f_{J,Z})] \leq (5pw)^k.$$

Claim 2: For all k, p,

$$W^{\geq \lceil k/p \rceil}(f) \leq 2\,\mathbb{E}_{(J,Z)\sim\mathcal{R}_p}[W^{\geq k}(f_{J,Z})].$$

Assuming both claims, here is how we prove the LMN lemma:

Proof of LMN lemma. Pick $p = 1/(10w)$. Then for all k,

$$W^{\geq k\cdot 10w}(f) \leq 2 \cdot (spw)^k \leq 2\left(\frac{1}{2}\right)^k.$$

Now pick $k = \log(2/\varepsilon)$ to get $W^{\geq 10wk}(f) \leq \varepsilon$. □

Now we prove Claim 1:

Proof of Claim 1.

$$\begin{aligned}
&\mathbb{E}_{(J,Z)\sim\mathcal{R}_p}[W^{\geq k}(f_{J,Z})] \\
&\quad = \mathbb{E}_{(J,Z)\sim\mathcal{R}_p}[W^{\geq k}(f_{J,Z}) \mid \text{DT depth}(f_{J,Z}) \geq k] \cdot \mathbb{P}(\text{DT depth}(f_{J,Z}) \geq k) \\
&\qquad + \underbrace{\mathbb{E}_{(J,Z)\sim\mathcal{R}_p}[W^{\geq k}(f_{J,Z}) \mid \text{DT depth}(f_{J,Z}) < k]}_{=0} \cdot \mathbb{P}(\text{DT depth}(f_{J,Z}) < k) \\
&\quad \leq 1 \cdot \mathbb{P}(\text{DT depth}(f_{J,Z}) \geq k)
\end{aligned}$$

Since this is a DNF, we can use Håstad's switching lemma.

$$\leq (5pw)^k.$$

□

Finally, we prove Claim 2:

Proof of Claim 2.

$$\mathbb{E}_{(J,Z)\sim\mathcal{R}_p}[W^{\geq k}(f_{J,Z})] = \sum_{U\subseteq[n]} \widehat{f}(U)^2 \mathbb{P}(|U\cap J| \geq k)$$

The random variable $|U \cap J|$ has $\mathrm{Bin}(|U|, p)$ distribution

$$= \sum_{U\subseteq[n]} \widehat{f}(U)^2 \mathbb{P}(\mathrm{Bin}(|U|,p) \geq k)$$

For this to be a small event, we want $|U|p \geq k$.

$$\geq \sum_{|U|\geq k/p} \widehat{f}(U)^2 \underbrace{\mathbb{P}(\mathrm{Bin}(|U|,p) \geq k)}_{\geq 1/2}$$

$$= \frac{1}{2} W^{\geq k/p}(f). \qquad \square$$

Mansour's theorem tells us that even within the lower levels, the Fourier spectrum is concentrated on relatively few coefficients. Generally, any boolean function with behavior according to the LMN lemma will actually also have behavior according to Mansour's theorem.