Computer Science 294 Lecture 12 Notes

Daniel Raban

February 23, 2023

1 Mansour's Theorem and $\mathbf{AC}^0$ Circuits

1.1 Recap: concentration of DNFs and the Fourier entropy influence conjecture

Recall that a width w DNF looks like

$$(x_{i_1}\overline{x_{i_2}} \wedge x_{i_3} \wedge \cdots \wedge \overline{x_{i_w}}) \vee \cdots ,$$

where there are at most w literals per term. The size is the total number of terms. Previously, we proved the LMN lemma:

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

Today, we will use the LMN lemma to prove Mansour's theorem, which tells us that even within these lower levels, the coefficients are concentrated.

Theorem 1.1 (Mansour)**.** *Width w DNFs are ε-concentrated on at most $w^{O(w\log(1/\varepsilon))}$ many coeficients.*

In the same paper that he proved this theorem Mansour stated the following conjecture.

Conjecture 1.1 (Mansour)**.** *Width w DNFs are 0.01-concentrated on at most $2^{O(w)}$ many coefficients.*

This is related to the Fourier Entropy Influence Conjecture.

Definition 1.1. The **Fourier entropy** of f is the entropy of the probability distribution $\mathscr{S}_f$. That is,

$$H(f) = \sum_{S\subseteq[n]} \widehat{f}(s)^2 \cdot \log_2(1/\widehat{f}(S)^2).$$

Here is a fact.

Proposition 1.1. *The Fourier spectrum of any function is 0.01-concentrated on* $2^{O(H(f))}$ *many coefficients.*

Conjecture 1.2 (FEI)**.** *For all boolean functions* f*,*

$$H(f) \leq O(I(f)).$$

This tells us that the Fourier entropy influence conjecture would imply Mansour's theorem, as DNFs have influence $O(w)$.

To prove the LMN lemma, we proved Håstad's switching lemma.

Lemma 1.2. *Suppose* f *is a width* w *DNF. Then for all* k*,*

$$\mathbb{P}_{(J,Z)\sim\mathcal{R}_p}(\textit{Decision Tree depth}(f_{J,Z}) \geq k) \leq (5pw)^k.$$

Here $\mathcal{R}_p$ is the distribution of p-random restrictions given by first picking $J \subseteq_p [n]$ variables to still be alive, picking $z \in \{\pm 1\}^{\overline{J}}$ uniformly at random, and assigning $\overline{J}$ according to z.

1.2 Proof of Mansour's theorem

Denote

$$L_{1,k}(f) := \sum_{S:|S|=k} |\widehat{f}(S)|.$$

This measures sparsity. If the coefficients are spread out, then by Cauchy-Schwarz, this will be close to $\sqrt{\binom{n}{k}}$; of the coefficients are concentrated, this will be much smaller. When $k = 1$, this is the total effect.

Lemma 1.3.

$$\mathbb{E}_{(J,Z)\sim\mathcal{R}_p}[L_{1,k}(f)] \geq p^k L_{1,k}(f).$$

Proof.

$$\begin{aligned}\mathbb{E}_{(J,Z)\sim\mathcal{R}_p}\left[\sum_{S:|S|=k} |\widehat{f_{J,Z}}(S)|\right] &= \sum_{S:|S|=k} \mathbb{E}_{(J,Z)\sim\mathcal{R}_p}[|\widehat{f_{J,Z}}(S)|] \\ &\geq \sum_{S:|S|=k} |\,\mathbb{E}_{(J,Z)\sim\mathcal{R}_p}[\widehat{f_{J,Z}}(S)]|\end{aligned}$$

Using our calculation from before about the expectation of Fourier coefficients for a p-random restriction,

$$= \sum_{S:|S|=k} p^k |\widehat{f}(S)|. \qquad \square$$

Corollary 1.1. *If f is a width w DNF, then for all k, $L_{1,k}(f) \le 2(20w)^k$.*

Proof. Consider a p-random restriction with $p = 1/(20w)$. Then

$$\begin{aligned}
L_{1,k}(f) &\le \frac{1}{p^k}\,\mathbb{E}_{(J,Z)\sim\mathcal{R}_p}[L_{1,k}(f_{J,Z})] \\
&= \frac{1}{p^k}\sum_{d=0}^{\infty}\mathbb{P}(\text{DT depth}(f_{J,Z}) = d)\cdot\mathbb{E}[L_{1,k}(f_{J,Z}) \mid \text{DT depth}(f_{J,Z}) = d]
\end{aligned}$$

The sum of the absolute values of all the coefficients for a decision tree of depth d is 2^d. Using Håstad's switching lemma,

$$\begin{aligned}
&\le \frac{1}{p^k}\sum_{d=0}^{\infty}(5pw)^d\cdot 2^d \\
&= \frac{1}{p^k}\sum_{d=0}^{\infty}\left(\frac{1}{2}\right)^d \\
&= 2p^k \\
&= 2\cdot(20w)^k.
\end{aligned}$$

□

Now let's prove Mansour's theorem.

Proof of Mansour's theorem. We already know by the LMN lemma that f is $\varepsilon/2$-concentrated up to degree $t = O(w\log(1/\varepsilon))$. Let $\mathcal{F} = \{S \subseteq [n] : |S| \le t, |\widehat{f}(S)| \ge \varepsilon/(100w)^t\}$. Then

$$\begin{aligned}
\sum_{S\notin\mathcal{F}}\widehat{f}(S)^2 &= \sum_{|S|>k}\widehat{f}(S)^2 + \sum_{\substack{|S|\le t\\ |\widehat{f}(S)|<\varepsilon/(100w)^t}}\widehat{f}(S)^2 \\
&\le \frac{\varepsilon}{2} + \frac{\varepsilon}{(100w)^t}\underbrace{\sum_{|S|\le t}|\widehat{f}(S)|}_{O(20w)^t} \\
&\le \varepsilon.
\end{aligned}$$

By Parseval's identity,

$$|\mathcal{F}| \le \left(\frac{(100w)^t}{\varepsilon}\right)^2 = O(w)^t\cdot\frac{1}{\varepsilon^2}.$$

□

Exercise 1.1. Show that the last step in the theorem can be improved to give

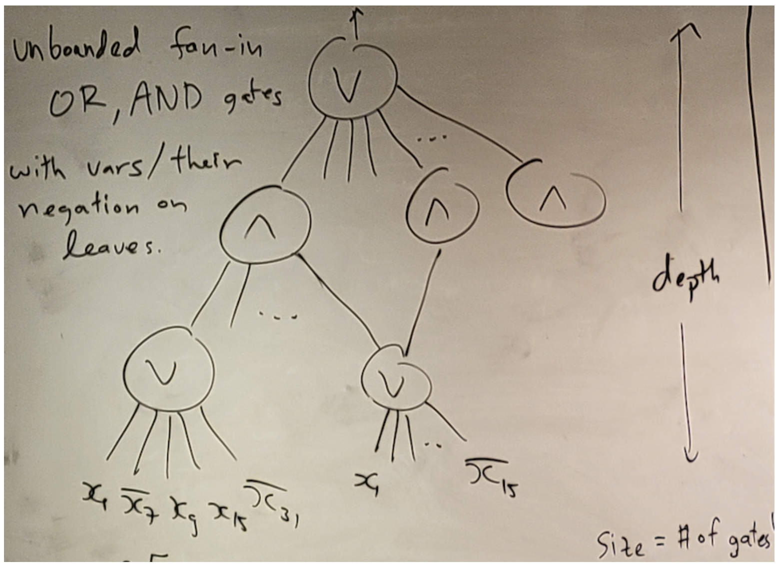
$$|\mathcal{F}| \le \frac{(100w)^{2t}}{\varepsilon} = w^{O(w\log(1/\varepsilon))}.$$

Remark 1.1. Mansour's original proof was about the size of the DNF, rather than its width.

1.3 AC^0 circuits

More generally than just and OR of AND functions, we could have some composition of ORs and ANDs.

Definition 1.2. **AC^0 circuits** are unbounded fan-in OR and AND gates with variables/their negation on the leaves.



Definition 1.3. The **depth** of the circuit is the number of levels of gates, while the size is the number of gates.

We can just assume we're switching between AND and OR because two ANDs or two ORs in a row could just be consolidated into a single AND or OR.

Definition 1.4. We denote by AC$^0[s,d]$ the set of all Boolean functions that can be implemented by such circuits of size $\leq s$ and depth $\leq d$.

Theorem 1.2 (FSS, Ajtai, Yao, Håstad)**.**

$$\text{PARITY}_n \notin \text{AC}^0[\text{poly}(n), O(1)].$$

In fact, any AC^0 circuit computing PARITY of depth d, must have size $\geq 2^{\Omega(n^{1/(d-1)})}$. For PARITY, there are circuits of depth d and size $2^{O(n^{1/(d-1)})}$, so this is the correct rate.

Theorem 1.3 (LMN)**.** *If f is computable by a size s, depth d AC^o circuit, then f is ε-concentrated up to degree $O(\log(s/\varepsilon))^{d-1}\log(1/\varepsilon)$.*

Remark 1.2. The dependence can be improved to $O(\log s)^{d-1}\log(1/\varepsilon)$, which is tight.

Corollary 1.2. *If $f \in \mathrm{AC}^0[s,d]$, then*

$$|\langle f, \mathrm{PARITY}_n\rangle| = |\widehat{f}([n])| = \sqrt{W^n(f)} \leq 2^{-(n/(\log s)^{d-1})^{1/d}}.$$

Remark 1.3. In 2014, Håstad improved this to $2^{-n/O(\log s)^{d-1}}$.

Proof sketch of LMN theorem. Let $w = \log(s/\varepsilon)$, and apply the following sequence of random restrictions:

1. A $1/10$-random restriction.
2. $d-2$ iterations of $1/(10w)$-random restrictions.
3. A $1/(10w)$-random restriction.

We claim that:

1. After Step 1, with high probability, the bottom fan-in will be $\leq w$.
2. In each iteration in Step 2, we decrease the depth by 1 (because we turn an AND of ORs into an OR of ANDs via Håstad's switching lemma and collapse and OR of ORs into just one layer of ORs).
3. In the last step, with high probability, we get a depth $\log(1/\varepsilon)$ decision tree.

Overall, this is just a random restriction with parameter $p = \frac{1}{10} \cdot (\frac{1}{10w})^{d-1} = \frac{1}{O(\log(s/\varepsilon))^{d-1}}$. In the last step, we get

$$\mathbb{P}_{(J,Z)\sim\mathcal{R}_p}(\text{DT depth}(f_{J,Z}) \geq \log(1/\varepsilon)) \leq \varepsilon,$$

so

$$\mathbb{E}[W^{\geq \log(1/\varepsilon)}(f_{J,Z})] \leq \varepsilon.$$

We know that

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

so

$$W^{\geq \log(1/\varepsilon)\cdot 1/p}(f) \leq 2\varepsilon. \qquad \square$$

Remark 1.4. To get the improvement, we need a better switching lemma.