Random Restrictions and Bounded-Depth Circuits (lecture notes) [Edited 2025-11-26]

Course: Analysis of Boolean Functions, Autumn 2025, University of Chicago
Instructor: William Hoza (`williamhoza@uchicago.edu`)

In these notes, we will use *random restrictions* to prove total influence and Fourier concentration bounds for DNFs and, more generally, for bounded-depth circuits.

Definition 0.1 (Restrictions)**.** A *restriction* is a string $\rho \in \{+1, -1, \star\}^n$. If $f\colon \{\pm 1\}^n \to \mathbb{R}$, then $f|_\rho\colon \{\pm 1\}^n \to \mathbb{R}$ is defined by $f|_\rho(x) = f(y)$, where

$$y_i = \begin{cases} \rho_i & \text{if } \rho_i \in \{\pm 1\} \\ x_i & \text{if } \rho_i = \star. \end{cases}$$

Definition 0.2 (Random Restrictions)**.** We define R_p to be the distribution over $\rho \in \{+1, -1, \star\}^n$ in which the coordinates are independent and

$$\rho_i = \begin{cases} \star & \text{with probability } p \\ +1 & \text{with probability } (1-p)/2 \\ -1 & \text{with probability } (1-p)/2. \end{cases}$$

1 Influence of size-s DNFs

Previously, we proved that width-w DNFs have total influence $O(w)$. In this section, we will prove that size-s DNFs have total influence $O(\log s)$. The first step is to show that if we apply a random restriction to a size-s DNF, the restricted function tends to have low total influence.

Lemma 1.1. *Let f be a size-s DNF. Then*

$$\mathop{\mathbb{E}}_{\rho \sim R_{1/2}}[\mathsf{DNFWidth}(f|_\rho)] \le O(\log s).$$

Proof. The probability that a fixed term has width at least w after the restriction is at most $(3/4)^w$. (If it had width less than w initially, this is trivial; if it had width at least w initially, then it collapses to 0 except with this probability.) Therefore, by the union bound, $\Pr[\mathsf{DNFWidth}(f|_\rho) \ge w] \le s \cdot (3/4)^w$. Therefore,

$$\mathbb{E}[\mathsf{DNFWidth}(f|_\rho)] = \sum_{w=1}^{\infty} \Pr[\mathsf{DNFWidth}(f|_\rho) \ge w] \le \log_{4/3} s + s \cdot \sum_{w \ge \log_{4/3} s} (3/4)^w = \log_{4/3} s + O(1). \qquad \square$$

The second step is to analyze the effect of random restrictions on total influence.

Lemma 1.2. *Let $f\colon \{\pm 1\}^n \to \{\pm 1\}$ and $p \in [0, 1]$. Then*

$$\mathrm{I}[f] = \frac{1}{p} \cdot \mathop{\mathbb{E}}_{\rho \sim R_p}[\mathrm{I}[f|_\rho]].$$

Proof.

$$\mathop{\mathbb{E}}_{\rho \sim R_p}[\mathrm{Inf}_i[f|_\rho]] = \Pr_{\rho \sim R_p, x \in \{\pm 1\}^n}[f|_\rho(x) \ne f|_\rho(x^{\oplus i})] = p \cdot \mathrm{Inf}_i[f]. \qquad \square$$

It follows that size-s DNFs have total influence $O(\log s)$. Note that this implies that DNFs computing the parity function must have size $2^{\Omega(n)}$.

2 Exponentially small Fourier tails

Previously, we proved that width-w DNFs have total influence $O(w)$, which implies that they are ε-concentrated up to degree $O(w/\varepsilon)$, hence learnable from random examples in time $n^{O(w/\varepsilon)}$. In this section, we will prove that they are ε-concentrated up to degree $O(w \cdot \log(1/\varepsilon))$, hence learnable from random examples in time $n^{O(w\cdot\log(1/\varepsilon))}$, which is much better when ε is small. E.g., think $\varepsilon = 1/n$.

The first step is to show that width-w DNFs become low-degree functions under random restrictions. This follows from the famous Switching Lemma:

Lemma 2.1 (Switching Lemma)**.** *Let f be a width-w DNF. Then*

$$\Pr_{\rho\sim R_p}[\mathsf{DTDepth}(f|_\rho) \geq k] \leq O(pw)^k.$$

We omit the proof of the switching lemma. The next step is a formula for the Fourier coefficients of a restricted function.

Lemma 2.2. *Let $f\colon \{\pm 1\}^n \to \{\pm 1\}$, let ρ be a restriction, let x be a completion of ρ, and let $S \subseteq [n]$. Then*

$$\widehat{f|_\rho}(S) = \sum_{U\subseteq[n]} \widehat{f}(S\cup U)\cdot \chi_U(x)\cdot 1[S \subseteq \rho^{-1}(\star) \text{ and } U \subseteq \rho^{-1}(\{0,1\})].$$

Proof. We have

$$f(x) = \sum_{T\subseteq[n]} \widehat{f}(T)\cdot\chi_T(x) = \sum_{T\subseteq[n]} \widehat{f}(T)\cdot \chi_{T\cap\rho^{-1}(\{0,1\})}(x)\cdot\chi_{T\cap\rho^{-1}(\star)}(x),$$

so the coefficient on $\chi_S(x)$ is

$$\sum_{T:T\cap\rho^{-1}(\star)=S} \widehat{f}(T)\cdot\chi_{T\cap\rho^{-1}(\{0,1\})}(x).$$

The change of variables $U = T\cap\rho^{-1}(\{0,1\})$ completes the proof. □

The next step is to show that the spectral sample distribution is not affected much by random restrictions, hence the width-w DNF must have had good spectral concentration even before the restriction. Specifically, the following lemma [based on my circuit complexity lecture notes from Autumn 2024] says that the operation of drawing a spectral sample "commutes with" the operation of applying a random restriction.

Lemma 2.3 (Spectral sample after a random restriction)**.** *Let $f\colon\{\pm1\}^n\to\{\pm1\}$. The following two distributions over subsets of $[n]$ are identical.*

1. *Sample $\rho\sim R_p$, then sample $S\sim\mathcal{S}_{f|_\rho}$, then output S.*
2. *Sample $T\sim\mathcal{S}_f$, then sample $\rho\sim R_p$, then output $T\cap\rho^{-1}(\star)$.*

Proof. By squaring the previous lemma, we find that for any restriction ρ and any completion x of ρ, we have

$$\widehat{f|_\rho}(S)^2 = \sum_{U,U'\subseteq[n]} \widehat{f}(S\cup U)\cdot\widehat{f}(S\cup U')\cdot\chi_{U\Delta U'}(x)\cdot 1[S\subseteq\rho^{-1}(\star)\text{ and } U,U'\subseteq\rho^{-1}(\{0,1\})],$$

where $U\Delta U'$ is the symmetric difference between U and U'. If ρ is a random restriction sampled from R_p and x is a uniform random completion of ρ, then in expectation, we have

$$\mathbb{E}\left[\widehat{f|_\rho}(S)^2\right] = \sum_{U,U'\subseteq[n]} \widehat{f}(S\cup U)\cdot\widehat{f}(S\cup U')\cdot\mathbb{E}\left[\chi_{U\Delta U'}(x)\cdot 1[S\subseteq\rho^{-1}(\star)\text{ and } U,U'\subseteq\rho^{-1}(\{0,1\})]\right].$$

The completion x and the star-set $\rho^{-1}(\star)$ are independent, so we can exchange the expectation with the product:

$$\mathbb{E}\left[\widehat{f|_\rho}(S)^2\right] = \sum_{U,U' \subseteq [n]} \widehat{f}(S \cup U) \cdot \widehat{f}(S \cup U') \cdot \mathbb{E}[\chi_{U \Delta U'}(x)] \cdot \Pr[S \subseteq \rho^{-1}(\star) \text{ and } U, U' \subseteq \rho^{-1}(\{0,1\})].$$

Nontrivial character functions have expectation zero, so the equation above simplifies to

$$\begin{aligned}\mathbb{E}\left[\widehat{f|_\rho}(S)^2\right] &= \sum_{U \subseteq [n]} \widehat{f}(S \cup U)^2 \cdot \Pr[S \subseteq \rho^{-1}(\star) \text{ and } U \subseteq \rho^{-1}(\{0,1\})] \\ &= \sum_{T \subseteq [n]} \widehat{f}(T)^2 \cdot \Pr[S = T \cap \rho^{-1}(\star)].\end{aligned}$$

The left-hand side in the equation above is the probability of getting S under distribution 1 in the lemma statement. The right-hand side is the probability of getting S under distribution 2 in the lemma statement. □

Theorem 2.4. *If $f\colon \{\pm 1\}^n \to \{\pm 1\}$ is a width-w DNF, then $W^{\geq k}[f] \leq 2 \cdot 2^{-\Omega(k/w)}$, and hence f is ε-concentrated on degree at most $O(w \cdot \log(1/\varepsilon))$.*

Proof. On the one hand, by the Switching Lemma, there is a value $p = \Theta(1/w)$ such that for every $d \in \mathbb{N}$, we have

$$\Pr_{\substack{\rho \sim R_p \\ S \sim \mathcal{S}_{f|_\rho}}}[|S| \geq d] \leq \Pr_{\rho \sim R_p}[\mathsf{DTDepth}(f|_\rho) \geq d] \leq 2^{-d}.$$

On the other hand, by Lemma 2.3, we have

$$\Pr_{\substack{\rho \sim R_p \\ S \sim \mathcal{S}_{f|_\rho}}}[|S| \geq d] = \mathop{\mathbb{E}}_{T \sim \mathcal{S}_f}\left[\Pr_{\rho \sim R_p}[|T \cap \rho^{-1}(\star)| \geq d]\right].$$

For any fixed set $T \subseteq [n]$, we expect $|T \cap \rho^{-1}(\star)| \approx p \cdot |T|$. Indeed, one can show that

$$\Pr\left[|T \cap \rho^{-1}(\star)| \geq \lfloor\, p \cdot |T|\, \rfloor\right] \geq 1/2.$$

(Note that such a statement amounts to bounding the median of the binomial distribution.[1]) Therefore,

$$\mathop{\mathbb{E}}_{T \sim \mathcal{S}_f}\left[\Pr_{\rho \sim R_p}\left[|T \cap \rho^{-1}(\star)| \geq \lfloor pk \rfloor\right]\right] \geq \Pr_{T \sim \mathcal{S}_f}[|T| \geq k] \cdot \frac{1}{2}.$$

Rearranging, we get $\Pr_{T \sim \mathcal{S}_f}[|T| \geq k] \leq 2 \cdot 2^{-\lfloor pk \rfloor}$. If $pk \geq 2$, then this is at most $2 \cdot 2^{-pk/2}$, and if $pk \leq 2$, then trivially $\Pr_{T \sim \mathcal{S}_C}[|T| \geq k] \leq 2 \cdot 2^{-pk/2}$. □

3 Deeper circuits

An AC^0_d *circuit* is a depth-d circuit consisting of alternating layers of AND gates and OR gates with unbounded fan-in, ultimately applied to variables and negated variables. The *size* of the circuit is the total number of gates. Let's analyze the total influence of such a circuit. Once again, the first step is to analyze the effect of a random restriction on such a circuit. The "AC^0 Criticality Theorem" is analogous to the Switching Lemma.

Theorem 3.1 (AC^0 Criticality Theorem)**.** *Let f be a size-s AC^0_d circuit, let $p \in (0,1)$, and let $k \in \mathbb{N}$. Then*

$$\Pr_{p \sim R_p}[\mathsf{DTDepth}(f|_\rho) \geq k] \leq (p \cdot O(\log s)^{d-1})^k.$$

[1] An alternative and more elementary approach is to use Cantelli's inequality to prove $\Pr[|T \cap \rho^{-1}(\star)| \geq \lfloor pk/2 \rfloor] \geq 1/3$.

We omit the proof of Theorem 3.1. Let's take Theorem 3.1 for granted and use it to bound the total influence of AC^0 circuits.

Corollary 3.2. *Let f be a size-s AC^0_d circuit. Then*

$$\mathrm{I}[f] \le O(\log s)^{d-1}.$$

Proof. For a suitable value $p = 1/O(\log s)^{d-1}$, Theorem 3.1 implies

$$\mathop{\mathbb{E}}_{p \sim R_p}[\mathsf{DTDepth}(f|_\rho)] \le \sum_{k=1}^{\infty} 2^{-k} = 1.$$

If f is a depth-k decision tree, then $\mathrm{sens}_f(x) \le k$ for every x, hence $\mathrm{I}[f] \le k$. Applying Lemma 1.2 completes the proof. $\square$

Similarly, one can prove that AC^0 circuits have exponentially small Fourier tails:

Theorem 3.3 (Linial-Mansour-Nisan)**.** *If $f\colon \{\pm 1\}^n \to \{\pm 1\}$ is a size-s AC^0_d circuit, then $W^{\ge k}[f] \le 2 \cdot 2^{-k/O(\log s)^{d-1}}$, hence f is ε-concentrated on degree at most $O(\log s)^{d-1} \cdot \log(1/\varepsilon)$.*