Fourier Growth and the Coin Problem (lecture notes)

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

Previously, we showed that polynomial-size DNFs are concentrated up to degree $O(\log n)$, hence learnable from random examples in time $n^{O(\log n)}$. In these notes, we will prove Mansour's theorem, which says that polynomial-size DNFs are concentrated on $n^{O(\log \log n)}$ Fourier coefficients, hence learnable from queries in time $n^{O(\log \log n)}$.

The key is to bound the "Fourier growth" of DNFs. What this means is that we will bound the quantity $\sum_{S:|S|=k} |\widehat{f}(S)|$.

Definition 0.1. Let $f\colon \{\pm 1\}^n \to \mathbb{R}$. We define $L_{1,k}(f) = \sum_{|S|=k} |\widehat{f}(S)|$.

It turns out that Fourier growth bounds have additional applications as well, beyond Fourier concentration and learnability. For example, we will use our Fourier growth bounds to prove that AC^0 circuits do a poor job of solving the so-called *coin problem*. To further illustrate this technique, we will also prove Fourier growth bounds for *regular read-once branching programs.*

1 Fourier growth of bounded-depth circuits

Theorem 1.1. *If $f\colon \{\pm 1\}^n \to \{\pm 1\}$ is a size-s AC^0_d circuit, then $L_{1,k}(f) \le O(\log S)^{(d-1)\cdot k}$.*

Proof. Previously, we showed that if ρ is a restriction and x is a completion, then

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

Consequently, if we sample $\rho \sim R_p$ and let x be a uniform random completion, then

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

Therefore,

$$\begin{aligned}
\sum_{|S|=k} |\widehat{f}(S)| = \sum_{|S|=k} p^{-k} \left|\mathbb{E}\left[\widehat{f|_\rho}(S)\right]\right| &\le p^{-k} \cdot \mathbb{E}\left[\sum_{|S|=k} \left|\widehat{f|_\rho}(S)\right|\right] \\
&\le p^{-k} \cdot \sum_{D=k}^{\infty} 2^D \cdot \Pr[\mathsf{DTDepth}(C|_\rho) = D] \\
&\le p^{-k} \cdot \sum_{D=k}^{\infty} (2p \cdot O(\log S)^{d-1})^D.
\end{aligned}$$

(The second inequality uses the facts that $\deg(f) \le \mathsf{DTDepth}(f)$ and $\hat{\|} f \hat{\|}_1 \le 2^{\mathsf{DTDepth}(f)}$.) If we choose p small enough (e.g., there is a value $p = 1/O(\log S)^{d-1}$ that works), then the geometric sum is dominated by its first term:

$$\sum_{|S|=k} |\widehat{f}(S)| \le p^{-k} \cdot 2 \cdot (2p \cdot O(\log S)^{d-1})^k = O(\log S)^{(d-1)\cdot k}. \qquad \square$$

Corollary 1.2. *Let d be a constant, for simplicity. If $f\colon \{\pm 1\}^n \to \{\pm 1\}$ is a size-s AC^0_d circuit, then f is ε-concentrated on a set of $2^{O((\log s)^{d-1} \cdot \log \log s \cdot \log(1/\varepsilon))}$ Fourier coefficients.*

Proof. We showed previously that f is $(\varepsilon/2)$-concentrated on degree up to some $k = O(\log s)^{d-1} \cdot \log(1/\varepsilon)$. Define

$$\mathcal{F} = \{S \subseteq [n] : |S| \leq k \text{ and } |\widehat{f}(S)| \geq \theta\}$$

for a suitable value $\theta = \varepsilon/O(\log S)^{(d-1)\cdot k}$. Then f is ε-concentrated on $\mathcal{F}$, because

$$\sum_{|S|\leq k, |\widehat{f}(S)|<\theta} \widehat{f}(S)^2 \leq \theta \cdot \sum_{|S|\leq k} |\widehat{f}(S)| \leq \theta \cdot O(\log S)^{(d-1)\cdot k} = \varepsilon/2,$$

provided we choose θ appropriately. Furthermore, the cardinality of $\mathcal{F}$ is bounded by

$$|\mathcal{F}| \leq \sum_{D=0}^{k} \frac{\sum_{|S|=D} |\widehat{f}(S)|}{\theta} \leq O(\log s)^{(d-1)\cdot k}/\varepsilon = 2^{O((\log s)^{d-1} \cdot \log\log s \cdot \log(1/\varepsilon))}. \qquad \square$$

In particular, when $d = 2$, we get:

Theorem 1.3 (Mansour's theorem)**.** *If $f \colon \{\pm 1\}^n \to \{\pm 1\}$ is a size-s DNF, then f is ε-concentrated on a set of $s^{O(\log\log s \cdot \log(1/\varepsilon))}$ Fourier coefficients.*

When $s = \mathrm{poly}(n)$ and ε is a constant, the bound in Mansour's theorem is $n^{O(\log\log n)}$. The Fourier-entropy conjecture $H[\mathcal{S}_f] \leq O(\mathrm{I}[f])$ would imply that the bound can be improved to polynomial. This special case of the Fourier-entropy conjecture is known as "Mansour's conjecture."

Conjecture 1.4 (Mansour's conjecture)**.** *If $f \colon \{\pm 1\}^n \to \{\pm 1\}$ is a polynomial-size DNF and ε is a constant, then f is ε-concentrated on a set of* $\mathrm{poly}(n)$ *Fourier coefficients.*

Without proving Mansour's conjecture, Jackson used different techniques to prove that polynomial-size DNFs are learnable from queries in polynomial time [Jac97]. We will not prove Jackson's result in this course.

2 The coin problem

In the *coin problem*, we are given a coin that lands one way with probability $1/2 + \varepsilon$ and lands the other way with probability $1/2 - \varepsilon$. The goal is to figure out which side is more likely. The optimal strategy is to toss the coin a number of times and take the majority vote of the observed outcomes. By the Chernoff bound, if we make some $n = O(1/\varepsilon^2)$ tosses, this strategy succeeds with high probability.

In this section, as an application of the Fourier growth bound from the previous section, we will show that AC_d^0 circuits do a very poor job of solving the coin problem. (In particular, this implies that small AC_d^0 circuits cannot compute the majority function.) We use the following notation.

Definition 2.1. For $\mu \in [-1, 1]$, let X_μ denote the distribution over $\{\pm 1\}^n$ in which the coordinates are independent and each has expectation μ.

Theorem 2.2. *For every $s, d \in \mathbb{N}$, there exists $\mu = 1/O(\log s)^{d-1}$ such that if $f \colon \{\pm 1\}^n \to \{\pm 1\}$ is an AC_d^0 circuit of size s, then*

$$|\mathbb{E}[f(X_\mu)] - \mathbb{E}[f(X_{-\mu})]| \leq 0.01.$$

Proof.

$$
\begin{aligned}
|\mathbb{E}[f(X_\mu)] - \mathbb{E}[f(X_0)]| = \left|\sum_{S\subseteq[n]} \widehat{f}(S)\cdot(\mathbb{E}[\chi_S(X_\mu)] - \mathbb{E}[\chi_S(X_0)])\right| &= \left|\sum_{k=1}^{n}\sum_{|S|=k}\widehat{f}(S)\cdot\mu^k\right| \\
&\le \sum_{k=1}^{n}\mu^k\cdot\sum_{|S|=k}|\widehat{f}(S)| \\
&\le \sum_{k=1}^{\infty}(\mu\cdot O(\log S)^{d-1})^k \\
&\le \sum_{k=1}^{\infty}0.001^k \\
&\le 0.005,
\end{aligned}
$$

provided we choose a small enough value $\mu = 1/O(\log S)^{d-1}$. ☐

Corollary 2.3. *If f is an AC^0_d circuit that computes majority, then f has size at least $2^{n^{\Omega(1/d)}}$.*

3 Fourier growth of regular read-once branching programs

In this section, as another example of Fourier growth bounds, we study *read-once branching programs* (ROBPs).

Definition 3.1 (Oblivious ROBPs)**.** An *oblivious ROBP* is a layered digraph with layers $V_0, V_1, \dots, V_n$. For every $i \in [n]$, each vertex $v \in V_{i-1}$ is labeled $x_{\pi(i)}$ for some permutation $\pi\colon [n] \to [n]$. Furthermore, v has two outgoing edges labeled 0 and 1 pointing to V_i. There is a designated "start vertex" $v_{\text{start}} \in V_0$. Given an input $x \in \{0,1\}^n$, we start at v_{start}, and in step $i \in [n]$, we query $x_{\pi(i)}$ to determine which outgoing edge to traverse. We arrive at a vertex $v \in V_n$. There is a designated set of "accept vertices" $V_{\text{acc}} \subseteq V_n$. We set $f(x) = 1$ if $v \in V_{\text{acc}}$ and $f(x) = 0$ otherwise. Thus, the program computes $f\colon \{0,1\}^n \to \{0,1\}$. The *width* of the program is $\max_i |V_i|$.

We say that the program is *regular* if every vertex in $V_1 \cup \cdots \cup V_n$ has two incoming edges.

If $u \in V_i$, then we write $f_{u\to}$ to denote the ROBP on layers $V_i, V_{i+1}, \dots, V_n$ in which u is the start vertex. Similarly, if $S \subseteq V_i$, then $f_{\to S}$ is the ROBP on vertices $V_0, \dots, V_i$ in which S is the set of accepting vertices. We write $f_{\to v}$ as a shorthand for $f_{\to\{v\}}$.

Regular oblivious ROBPs are in many ways very different from AC^0 circuits. For example, we proved that AC^0 circuits are concentrated at low degree, whereas in contrast, there is a trivial width-2 regular oblivious ROBP that computes the parity function, hence regular oblivious ROBPs are *not* concentrated at low degree. In fact, one can check that the inner product function can be computed by a regular oblivious ROBP of width 4. Recall that the inner product function is "maximally non-concentrated:" every Fourier coefficient has absolute value precisely $2^{-n/2}$.

Nevertheless, we will prove that regular oblivious ROBPs satisfy a strong Fourier growth bound, similar to AC^0 circuits. The proof is completely different from the AC^0 proof. We begin by bounding the level-1 Fourier coefficients.

Lemma 3.2 (Level-1 Fourier coefficients of regular oblivious ROBPs)**.** *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a width-w regular oblivious ROBP. Then $L_{1,1}(f) \le \mathbb{E}[f]\cdot w$.*

Proof. Let m be the number of rejecting vertices in the final layer. We will prove a bound of $\mathbb{E}[f]\cdot m$ by induction on n. The base case $n = 0$ is trivial, so assume $n > 0$. Let $V_0, V_1, \dots, V_n$ be the layers of f.

Partition the penultimate layer V_{n-1} into three sets, $V_{n-1} = R \cup S \cup T$, based on the number of accepting edges from each vertex:

$$\begin{aligned} R &= \{v \in V_{n-1} : \mathbb{E}[f_{v\to}] = 0\} \\ S &= \{v \in V_{n-1} : \mathbb{E}[f_{v\to}] = 1/2\} \\ T &= \{v \in V_{n-1} : \mathbb{E}[f_{v\to}] = 1\}. \end{aligned}$$

Because f is regular, we have $m = |R| + \frac{1}{2}|S|$. Assume without loss of generality that x_n is the variable that the program reads in step n. Then for each $i < n$, we have

$$\begin{aligned} \widehat{f}(i) = \mathop{\mathbb{E}}_x[f(x) \cdot (-1)^{x_i}] &= \mathop{\mathbb{E}}_{x_1,\ldots,x_{n-1}}\left[(-1)^{x_i} \cdot \left(f_{\to T}(x) + \frac{1}{2} f_{\to S}(x)\right)\right] \\ &= \mathop{\mathbb{E}}_{x_1,\ldots,x_{n-1}}\left[(-1)^{x_i} \cdot \left(\frac{1}{2} f_{\to T}(x) + \frac{1}{2} f_{\to S\cup T}(x)\right)\right] \\ &= \frac{1}{2}\widehat{f_{\to T}}(i) + \frac{1}{2}\widehat{f_{\to S\cup T}}(i). \end{aligned}$$

Therefore, by induction, we have

$$\begin{aligned} \sum_{i=1}^{n-1} |\widehat{f}(i)| &\le \frac{1}{2}\,\mathbb{E}[f_{\to T}] \cdot |R \cup S| + \frac{1}{2}\,\mathbb{E}[f_{\to S\cup T}] \cdot |R| \\ &= \frac{1}{2}\,\mathbb{E}[f_{\to S}] \cdot |R| + \mathbb{E}[f_{\to T}] \cdot m. \end{aligned}$$

Meanwhile, at $i = n$, we have

$$|\widehat{f}(n)| = \left|\mathop{\mathbb{E}}_x[f(x) \cdot (-1)^{x_n}]\right| \le \mathop{\mathbb{E}}_{x_1,\ldots,x_{n-1}}\left[\left|\mathop{\mathbb{E}}_{x_n}[(-1)^{x_n} \cdot f(x)]\right|\right] = \frac{1}{2}\,\mathbb{E}[f_{\to S}] \le \frac{|S| \cdot \mathbb{E}[f_{\to S}]}{4},$$

because $|S|$ is even (recall $m = |R| + \frac{1}{2}|S|$). Combining the bounds, we get

$$\begin{aligned} \sum_{i=1}^{n} |\widehat{f}(i)| &\le \mathbb{E}[f_{\to S}] \cdot \left(\frac{|R|}{2} + \frac{|S|}{4}\right) + \mathbb{E}[f_{\to T}] \cdot m \\ &= \mathbb{E}[f_{\to S}] \cdot m/2 + \mathbb{E}[f_{\to T}] \cdot m \\ &= m \cdot (\mathbb{E}[f_{\to S}]/2 + \mathbb{E}[f_{\to T}]) \\ &= m \cdot \mathbb{E}[f]. \end{aligned}$$

□

Now we move on to the higher-order Fourier coefficients. There is a convenient formula for the Fourier coefficients of an ROBP in terms of the Fourier coefficients of its subprograms. A *standard-order ROBP* is an oblivious ROBP that reads the variables in the order $x_1, \ldots, x_n$.

Lemma 3.3. *Let f be a standard-order ROBP with layers $V_0, V_1, \ldots, V_n$. Let $i \in \{0, 1, \ldots, n\}$, let $S \subseteq [i]$, and let $T \subseteq [n] \setminus [i]$. Then*

$$\widehat{f}(S \cup T) = \sum_{v \in V_i} \widehat{f_{\to v}}(S) \cdot \widehat{f_{v\to}}(T).$$

Proof. Sample $(x, y) \in \{0,1\}^n$ uniformly at random, where $|x| = i$ and $|y| = n - i$. By the Fourier coefficient formula,

$$\begin{aligned} \widehat{f}(S \cup T) = \mathbb{E}[f(x, y) \cdot \chi_{S\cup T}(x, y)] &= \mathbb{E}\left[\left(\sum_{v \in V_i} f_{\to v}(x) f_{v\to}(y)\right) \cdot \chi_S(x) \cdot \chi_T(y)\right] \\ &= \sum_{v \in V_i} \mathbb{E}[f_{\to v}(x) \cdot \chi_S(x) \cdot f_{v\to}(y) \cdot \chi_T(y)] \\ &= \sum_{v \in V_i} \widehat{f_{\to v}}(S) \cdot \widehat{f_{v\to}}(T). \end{aligned}$$

□

Our plan is to bound $L_{1,k}(f)$ by induction on k. Indeed, using Lemma 3.3, we can bound the level-$(k+1)$ Fourier coefficients in terms of the level-k Fourier coefficients as follows:

Lemma 3.4. *Let f be a standard-order oblivious ROBP with layers $V_0, V_1, \ldots, V_n$. Then*

$$L_{1,k+1}(f) \leq \sum_{i=1}^{n} \sum_{v \in V_{i-1}} L_{1,k}(f_{\to v}) \cdot |\widehat{f_{v\to}}(i)|.$$

Proof.

$$\begin{aligned} L_{1,k+1}(f) = \sum_{|S|=k+1} |\widehat{f}(S)| = \sum_{i=1}^{n} \sum_{T\subseteq[i-1],|T|=k} |\widehat{f}(T \cup \{i\})| &= \sum_{i=1}^{n} \sum_{T\subseteq[i-1],|T|=k} \left| \sum_{v\in V_{i-1}} \widehat{f_{\to v}}(T) \cdot \widehat{f_{v\to}}(i) \right| \\ &\leq \sum_{i=1}^{n} \sum_{v\in V_{i-1}} \left(\sum_{T\subseteq[i-1],|T|=k} |\widehat{f_{\to v}}(T)| \right) \cdot |\widehat{f_{v\to}}(i)| \\ &= \sum_{i=1}^{n} \sum_{v\in V_{i-1}} L_{1,k}(f_{\to v}) \cdot |\widehat{f_{v\to}}(i)|. \end{aligned}$$

$\square$

In the bound above, the absolute value signs around $\widehat{f_{v\to}}(i)$ are annoying. Recall that if $f\colon \{0,1\}^n \to \{0,1\}$ is *monotone*, and $F = (-1)^f$, then

$$\widehat{f}(i) = -\frac{1}{2}\widehat{F}(i) = -\frac{1}{2}\operatorname{Inf}_i[f] \leq 0,$$

so we can remove the absolute value signs and say $|\widehat{f}(i)| = -\widehat{f}(i)$. More generally, the same conclusion holds if f is *locally monotone*:[1]

Definition 3.5 (Local monotonicity)**.** Let $f\colon \{0,1\}^n \to \{0,1\}$. We say that f is *locally monotone* if for every $i \in [n]$ and every $x \in \{0,1\}^{i-1}$, we have

$$\mathop{\mathbb{E}}_{y\in\{0,1\}^{n-i}}[f(x0y)] \leq \mathop{\mathbb{E}}_{y\in\{0,1\}^{n-i}}[f(x1y)].$$

Locally monotone functions are not necessarily monotone.[2] However, locally monotone function always have non-positive degree-1 Fourier coefficients, just like monotone functions:

Lemma 3.6. *If $f\colon \{0,1\}^n \to \{0,1\}$ is locally monotone, then $\widehat{f}(i) \leq 0$ for every $i \in [n]$.*

Proof.

$$\begin{aligned} \widehat{f}(i) = \mathop{\mathbb{E}}_{xby\in\{0,1\}^n}[f(xby)\cdot(-1)^b] &= \mathop{\mathbb{E}}_{x\in\{0,1\}^{i-1}}\left[\frac{1}{2}\sum_{b\in\{0,1\}} \mathop{\mathbb{E}}_{y\in\{0,1\}^{n-i}}[f(xby)\cdot(-1)^b]\right] \\ &= \mathop{\mathbb{E}}_{x\in\{0,1\}^{i-1}}\left[\frac{1}{2}\left(\mathop{\mathbb{E}}_{y\in\{0,1\}^{n-i}}[f(x0y)] - \mathop{\mathbb{E}}_{y\in\{0,1\}^{n-i}}[f(x1y)]\right)\right] \\ &\leq 0. \end{aligned}$$

$\square$

To bound $L_{1,k+1}(f)$, our approach is to reduce to the locally monotone case, via the following construction.

Lemma 3.7 (Local monotonization)**.** *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a standard-order ROBP. By only relabeling the edges of f, it is possible construct another standard-order ROBP f' such that for every vertex v, the function $f_{v\to}$ is locally monotone.*

[1] Warning: This definition is not standard.

[2] For example, let $f(x, y, z) = (x \wedge y) \vee (\overline{x} \wedge z)$.

Proof. At each vertex v, swap the labels of the outgoing edges if necessary in order to ensure that $\mathbb{E}_x[f_{v\to}(1x)] \geq \mathbb{E}_x[f_{v\to}(0x)]$. The order in which we visit the vertices doesn't matter, because relabeling edges does not affect acceptance probabilities. $\square$

Theorem 3.8. *Let $f\colon \{0,1\}^n \to \{0,1\}$ be a width-w regular oblivious ROBP. Then $L_{1,k}(f) \leq w^k$.*

Proof. We will prove that $L_{1,k}(f) \leq w^k \cdot \mathbb{E}[f]$ by induction on k. The base case $k = 0$ is trivial. For the inductive step, assume without loss of generality that f is a standard-order ROBP. Let f' be the local monotonization from Lemma 3.7. Then

$$
\begin{aligned}
L_{1,k+1}(f) &\leq \sum_{i=1}^{n} \sum_{v \in V_{i-1}} L_{1,k}(f_{\to v}) \cdot |\widehat{f_{v\to}}(i)| && \text{(Lemma 3.4)} \\
&\leq w^k \cdot \sum_{i=1}^{n} \sum_{v \in V_{i-1}} \mathbb{E}[f_{\to v}] \cdot |\widehat{f_{v\to}}(i)| \\
&= w^k \cdot \sum_{i=1}^{n} \left| \sum_{v \in V_{i-1}} \widehat{f'_{\to v}}(\varnothing) \cdot \widehat{f'_{v\to}}(i) \right| \\
&= w^k \cdot \sum_{i=1}^{n} |\widehat{f'}(i)| \\
&\leq w^{k+1} \cdot \mathbb{E}[f'] && \text{(Lemma 3.2)} \\
&= w^{k+1} \cdot \mathbb{E}[f].
\end{aligned}
$$

$\square$

Corollary 3.9 (Coin problem bound)**.** *If $f\colon \{\pm 1\}^n \to \{0,1\}$ is a width-w regular oblivious ROBP, then $|\mathbb{E}[f(X_\mu)] - \mathbb{E}[f]| \leq O(\mu \cdot w)$.*

References

[Jac97] Jeffrey C Jackson. "An Efficient Membership-Query Algorithm for Learning DNF with Respect to the Uniform Distribution". In: *Journal of Computer and System Sciences* 55.3 (1997), pp. 414–440. ISSN: 0022-0000. DOI: https://doi.org/10.1006/jcss.1997.1533.