ECCC

# TIGHT BOUNDS ON THE FOURIER GROWTH OF BOUNDED FUNCTIONS ON THE HYPERCUBE

SIDDHARTH IYER, ANUP RAO, VICTOR REIS, THOMAS ROTHVOSS,
AND AMIR YEHUDAYOFF

ABSTRACT. We give tight bounds on the degree $\ell$ homogenous parts $f_\ell$ of a bounded function $f$ on the cube. We show that if $f : \{\pm 1\}^n \to [-1, 1]$ has degree $d$, then $\|f_\ell\|_\infty$ is bounded by $d^\ell/\ell!$, and $\|\hat{f}_\ell\|_1$ is bounded by $d^\ell e^{\binom{\ell+1}{2}} n^{\frac{\ell-1}{2}}$. We describe applications to pseudorandomness and learning theory. We use similar methods to generalize the classical Pisier's inequality from convex analysis. Our analysis involves properties of real-rooted polynomials that may be useful elsewhere.

## 1. INTRODUCTION

The goal of complexity theory is to understand the space of functions that are efficiently computable. Every function $f : \{\pm 1\}^n \to \mathbb{R}$ corresponds to a multilinear polynomial in $n$ variables, and under many models of computation, efficiently computable functions correspond to bounded polynomials of low degree. This motivates an investigation of the characteristics of such functions. Our main results are tight bounds on the magnitudes of coefficients.

A set $S \subseteq [n]$ corresponds to the *monomial* or *character* $\chi_S(x) := \prod_{j \in S} x_j$. Every $f : \{\pm 1\}^n \to \mathbb{R}$ can be uniquely expressed as

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \cdot \chi_S(x),$$

where $\hat{f}(S) \in \mathbb{R}$ are the *Fourier coefficients* of $f$.

Bounds on the Fourier coefficients play a key role in computer science (see the textbook [20] and also [15, 1, 10, 7] and references within). Typical results bound the growth of the $\ell_1$ norm of the Fourier coefficients in terms of their degree. The $\ell$-th homogenous part of $f$ is

$$f_\ell(x) := \sum_{S \subseteq [n]:|S|=\ell} \hat{f}(S) \cdot \chi_S(x).$$

The main objective is to control the two norms

$$\|\hat{f}_\ell\|_1 := \sum_{S \subseteq [n]:|S|=\ell} |\hat{f}(S)|,$$

and

$$\|f_\ell\|_\infty := \max_{x \in \{\pm 1\}^n} |f_\ell(x)|.$$

1

| Class of functions $f : \{\pm 1\}^n \to [-1, 1]$ | $\|\hat{f}_\ell\|_1 \leq$ | Ref. |
|---|---|---|
| CNFs of width $w$ | $w^{O(\ell)}$ | [14] |
| Width $w$ regular oblivious read-once branching programs | $(2w^2)^\ell$ | [23] |
| Width $w$ oblivious read-once branching programs | $(O(\log n))^{w\ell}$ | [8] |
| Boolean functions of maximum sensitivity $s$ | $s^{O(\ell)}$ | [12] |
| $\mathbb{F}_2$ polynomials of degree $d$ | $\ell^\ell \cdot 2^{3d\ell}$ | [7] |
| Decision trees of depth $d$ | $(O(\sqrt{d \log n}))^\ell$ | [25, 24] |
| Parity decision trees of depth $d$ | $d^{\ell/2} \cdot O(\ell \log n)^\ell$ | [11] |

FIGURE 1. Known bounds on the Fourier growth of various classes of functions

By the triangle inequality, we must have $\|f_\ell\|_\infty \leq \|\hat{f}_\ell\|_1$. For general functions, the first quantity can be substantially smaller than the second. For symmetric functions $f$, the two quantities are the same: $\|\hat{f}_\ell\|_1 = |f_\ell(1^n)| \leq \|f_\ell\|_\infty$.

Several works have proved non-trivial bounds on $\|\hat{f}_\ell\|_1$ for functions that are efficiently computable. Figure 1 lists some known results in this direction. An important motivation for bounding these norms is that a class of functions with small Fourier growth can be efficiently *learned* [15], and admits efficient *pseudorandom generators* [6]. A pseudorandom generator for a class of functions is a function that generates a distribution that uses a small random seed to generate a distribution that is supported on a small set, yet is indistinguishable from the uniform distribution to functions from the class. Chattopadhyay, Hatami, Hosseini and Lovett [7] showed how to construct pseudorandom generators for any class of functions satisfying $\|\hat{f}_\ell\|_1 \leq t^\ell$, using $t^2 \cdot \mathsf{polylog}(n)$ bits of seed. Similarly, Chattopadhyay, Gaitonde, Lee, Lovett and Shetty [6] showed that bounds on $\|f_\ell\|_\infty$ also lead to efficient pseudorandom generators. Let $\mathcal{F}$ be a class of functions that is closed under restrictions (i.e., setting a variable to $\pm 1$ keeps the function in $\mathcal{F}$). Suppose there are parameters $k > 2$ and $t > 0$ such that every function $f \in \mathcal{F}$ satisfies $\|f_k\|_\infty \leq t^k$, then there is a pseudorandom generator of seed length $k \cdot t^{2+4/(k-2)} \cdot \mathsf{polylog}(n)$ for the class of functions $\mathcal{F}$.

Given these applications, it is interesting to ask for the most general bounds. What can we say about $\|\hat{f}_\ell\|_1$ and $\|f_\ell\|_\infty$ if $f : \{\pm 1\}^n \to [-1, 1]$ is an arbitrary function of degree $d$? Backurs and Bavarian [1] and later Filmus, Hatami, Keller and Lifshitz [10] studied bounds on the *influences* of such functions. With regards to the questions we study here, the techniques of [10] imply that $\|f_1\|_\infty = \|\hat{f}_1\|_1 \leq d$. In this work, we give tight bounds on $\|f_\ell\|_\infty$ and $\|\hat{f}_\ell\|_1$ for every $\ell$.

Our methods are intimately connected to proofs of a classical result in convex analysis called *Pisier's inequality* [21, 22]. Let $f : \{\pm 1\}^n \to \mathbb{R}^m$ be a vector valued function.

The $m$ coordinates of $f$ can be expressed as polynomials, so as before we have

$$f(x) = \sum_{S \subseteq [n]} \hat{f}(S) \cdot \chi_S(x),$$

where now $\hat{f}(S) \in \mathbb{R}^m$ is a vector. We define $f_\ell$ by projecting $f$ to its degree $\ell$ part, just as we did earlier. Pisier's inequality says that every norm $\| \cdot \|$ on $\mathbb{R}^m$ must satisfy

$$\mathbb{E}\left[\|f_1(X)\|^2\right]^{1/2} \leq O(\log(m+1)) \cdot \mathbb{E}\left[\|f(X)\|^2\right]^{1/2},$$

where $X \sim \{\pm 1\}^n$ is uniformly distributed.

This inequality has important applications in geometry. Most strikingly, combined with a result of Figiel and Tomczak-Jaegermann [9] it implies the $MM^*$-estimate, which says that in an average sense, symmetric convex bodies behave much more like ellipsoids than one could derive from John's theorem [13]. The $MM^*$-estimate is a central piece in the proofs of Milman's QS-theorem [17, 18, 19] and the construction of $M$-ellipsoids [19], some of the most consequential results in convex geometry.

In our work, we generalize Pisier's inequality to higher degrees, and make the proof more explicit (see discussion in Section 1.3 below).

1.1. **Results.** Our results and proofs are intimately connected with the Chebsyshev polynomial $T_d(z)$. This is the unique polynomial of degree $d$ so that $T_d(\cos(\theta)) = \cos(d\theta)$. Denote by $C(d, \ell)$ the coefficient of $z^\ell$ in $T_d(z)$. Our first result is that these coefficients give upper bounds on $\|f_\ell\|_\infty$.

**Theorem 1.** *If $f : \{\pm 1\}^n \to [-1, 1]$ has degree $d$, then*

$$\|f_\ell\|_\infty \leq \begin{cases} |C(d, \ell)| & \text{if } d = \ell \bmod 2, \\ |C(d-1, \ell)| & \text{otherwise.} \end{cases}$$

To understand the theorem better, recall the known formula [16]:

$$(1) \qquad C(d, \ell) = \begin{cases} (-1)^{(d-\ell)/2} \cdot 2^\ell \cdot \frac{d}{d+\ell} \cdot \binom{\frac{d+\ell}{2}}{\ell} & \text{if } d = \ell \bmod 2, \\ 0 & \text{otherwise.} \end{cases}$$

We can use the arithmetic-mean-geometric-mean inequality and (1) to show

$$(2) \qquad |C(d, \ell)| = 2^\ell \cdot \frac{d}{d+\ell} \cdot \frac{1}{2^\ell \ell!} \prod_{k=0}^{\ell-1} (d + \ell - 2k) \leq \frac{d^\ell}{\ell!}.$$

In particular, the theorem states that $\|f_\ell\|_\infty \leq \frac{d^\ell}{\ell!}$. The following proposition shows that the bound cannot be significantly improved when $n \gg d$:

**Proposition 2.** *For every $n, d, \ell$ such that $d = \ell \bmod 2$, the bounded function $f(x) = T_d((x_1 + \cdots + x_n)/n)$ satisfies $\|f_\ell\|_\infty \geq |C(d, \ell)| - 2e^d(d+1)!/n$.*

We also provide general bounds on the larger $\|\hat{f}_\ell\|_1$:

**Theorem 3.** *If $f : \{\pm 1\}^n \to [-1, 1]$ has degree $d$, then for $\ell \geq 1$, $\|\hat{f}_\ell\|_1 \leq n^{\frac{\ell-1}{2}} \cdot d^\ell \cdot e^{\binom{\ell+1}{2}}$.*

Once again, we give an example matching this bound when $d = \ell \ll n$:

**Proposition 4.** *There is a homogenous degree d polynomial $f : \{\pm 1\}^n \to [-1, 1]$ so that*

$$\|\hat{f}_d\|_1 = \frac{1}{2} \cdot \sqrt{\frac{1}{n} \cdot \binom{n}{d}}.$$

Our methods allow to prove the following generalization of Pisier's inequality.

**Theorem 5.** *Let $\ell, m, n$ be positive integers and $\| \cdot \|$ be a norm on $\mathbb{R}^m$. Let $X$ be uniformly distributed in $\{\pm 1\}^n$. Then for any function $f : \{\pm 1\}^n \to \mathbb{R}^m$,*

$$\mathbb{E}\left[\|f_\ell(X)\|^2\right]^{1/2} \leq \left(4 + \frac{6\log(m+1)}{\ell}\right)^\ell \cdot \mathbb{E}\left[\|f(X)\|^2\right]^{1/2}.$$

Bourgain showed that Pisier's inequality is sharp [3]. An adaptation of his construction shows that Theorem 5 is also sharp, though we omit the details here.

We conclude this section with an application to learning theory. Suppose we want to approximate an unknown function $f$. Access to $f$ is given by random queries of the form $(X, f(X))$ where $X \sim \{\pm 1\}^n$ is uniformly distributed. The goal is to efficiently compute $g$ so that $\mathbb{E}\left[|f(X) - g(X)|^2\right] \leq \varepsilon$. This problem was studied in several works (see e.g. [15] and references within). The theorems above lead to improving the sample complexity from polynomial in $n^d$ to $o(n^d)$, for $d$ fixed and $n \to \infty$.

**Theorem 6.** *There is a constant $c > 1$ so that the following holds. Let $f : \{\pm 1\}^n \to [-1, 1]$ be of degree $d \geq 1$ and let $\varepsilon > 0$. From $N \leq 2^{cd^2} \frac{n^{d-1}\log(n)}{\varepsilon^3}$ random queries to $f$, we can efficiently construct a function $g : \{\pm 1\}^n \to \mathbb{R}$ with $\mathbb{E}\left[|f(x) - g(x)|^2\right] \leq \varepsilon$.*

The function $g$ is obtained by estimating the large Fourier coefficients of $f$. The analysis closely follows classical arguments that can be found for example in [15, 20].

1.2. **Outline.** Theorem 1 is proved in Section 3, Theorem 3 is proved in Section 4, and Theorem 5 is proved in Section 5. Propositions 2 and 4 are proved in Section 9. Theorem 6 is proved in Section 8.

1.3. **Techniques.** Here we give a high-level sketch of some of our proofs, omitting many details that are explained later. Our techniques are inspired by proofs of Pisier's inequality. Pisier's original proof used complex analysis and interpolation. Bourgain and Milman found a different and more direct proof [4]. Their proof relies on the Hahn-Banach theorem, the Riesz representation theorem, and Bernstein's theorem from approximation theory. These deep results are used to show that there is a function that is close to the linear function $L(x) = x_1 + \ldots + x_n$, yet has much smaller $\ell_1$ norm than $L(x)$. The existence of this linear proxy is proved in a clever but non-constructive way.

In our work, we give an explicit formula for a (more general) proxy with the properties alluded to above. Our key technical contribution is an explicit *filter* that can be used to project a polynomial in $\cos(\theta)$ to its degree $\ell$ part. The filter is a central component of the proxy, and can potentially be useful elsewhere.

**Theorem 7.** *For every $d \geq \ell$ with $d = \ell \bmod 2$, there is a function $\phi : [0, 2\pi) \to \mathbb{R}$ and a distribution on $\theta$ such that $\mathbb{E}[|\phi(\theta)|] = |C(d, \ell)|$ and*

(3) $$\mathbb{E}[\phi(\theta) \cos^k(\theta)] = \begin{cases} 1 & \text{if } k = \ell, \\ 0 & \text{if } k \neq \ell, \ k \leq d + 1. \end{cases}$$

The theorem cannot be improved, in the sense that any function $\phi$ satisfying (3) must also satisfy

$$\mathbb{E}[|\phi(\theta)|] \geq |\mathbb{E}[\phi(\theta) \cos(d\theta)]| = |C(d, \ell)|.$$

The proof of Theorem 7 is based on properties of Chebyshev polynomials and some non-trivial facts about real-rooted polynomials that may be of independent interest. If $p(z) = \sum_{j=0}^{d} c_j z^j$ is a polynomial, we write $p_{>k}(z) = \sum_{j=k+1}^{d} c_j z^j$. The proof of Theorem 7 relies on the following theorem:

**Theorem 8.** *Let $p(z)$ be a real-rooted degree-$d$ polynomial whose roots are all positive. Then for every root $r$ of $p(z)$ and every $k \in \{0, \ldots, d\}$, we have $(-1)^{d-k-1} \cdot p_{>k}(r) \geq 0$.*

We prove Theorem 7 in Section 6 and Theorem 8 in Section 7. Theorem 1 is proved using the filter as follows. We construct a proxy $P : \{\pm 1\}^n \to \mathbb{R}$ using the formula[1]:

$$P(x) = \mathbb{E}_{\theta} \left[ \phi(\theta) \cdot \prod_{j=1}^{n} (1 + \cos(\theta) \cdot x_j) \right].$$

When $X \sim \{\pm 1\}^n$ is uniformly distributed, we can bound

$$\mathbb{E}[|P(X)|] \leq \mathbb{E}[|\phi(\theta)|] \leq |C(d, \ell)|.$$

By construction, we have

(4) $$\hat{P}(S) = \begin{cases} 1 & \text{if } |S| = \ell, \\ 0 & \text{if } |S| \neq \ell, \ |S| \leq d, \end{cases}$$

so we can compute $f_\ell$ via convolution as $f_\ell = f * P$. The properties of $P$ imply that the convolution with it cannot be large at any point.

Theorem 3 is proved by induction. In the proof, we apply a random restriction to $f_\ell$. We set each variable of $f_\ell$ randomly with probability $\frac{1}{\ell}$ and leave it unset with probability $1 - \frac{1}{\ell}$. This gives a degree $\ell$ function $g$. We use Khintchine's inequality to bound $\|\hat{f}_\ell\|_1$ in terms of $\|\hat{g}_{\ell-1}\|_1$. Since $g$ is bounded by $\|f_\ell\|_\infty$, induction combined with Theorem 1 can be used to prove Theorem 3.

Theorem 5 is proved by setting $d \approx \log(m + 1)$ and using the proxy:

$$P(x) := 2^\ell \cdot \mathbb{E}_{\theta} \left[ \phi(\theta) \cdot \prod_{j=1}^{n} (1 + \tfrac{\cos(\theta) \cdot x_j}{2}) \right]$$

Once again, the construction ensures that (4) holds. Because $|\hat{P}(S)| \leq 2^{\ell-d} \cdot |C(d, \ell)|$ for $|S| > d$, the proxy $P$ is close to the symmetric homogenous polynomial of degree $\ell$ whose coefficients are all 1. We can use the bound on $\mathbb{E}[|\phi(\theta)|]$ to bound $\mathbb{E}[|P(X)|] \leq 2^\ell |C(d, \ell)|$. Theorem 5 is again proved by analyzing the convolution $f * P$.

---

[1]This formula is inspired by earlier proofs of Pisier's inequality.

## 2. Preliminaries

### 2.1. Fourier analysis.

**Fact 9** (Parseval's identity). *If $f : \{\pm 1\}^n \to \mathbb{R}^m$ and $X \sim \{\pm 1\}^n$ is uniformly distributed then*

$$\mathbb{E}\left[\|f(X)\|_2^2\right] = \sum_{S \subseteq [n]} \|\hat{f}(S)\|_2^2.$$

*Proof.* The proof is based on the orthonormality of the characters:

$$\mathbb{E}\left[\|f(X)\|_2^2\right] = \mathbb{E}\left[\left\langle \sum_{S \subseteq [n]} \hat{f}(S) \cdot \chi_S(X), \sum_{T \subseteq [n]} \hat{f}(T) \cdot \chi_T(X) \right\rangle\right]$$

$$= \sum_{S,T \subseteq [n]} \left\langle \hat{f}(S), \hat{f}(T) \right\rangle \mathbb{E}\left[\chi_S(X)\chi_T(X)\right]$$

$$= \sum_{S \subseteq [n]} \|\hat{f}(S)\|_2^2. \qquad \square$$

Convolution is a powerful tool when there is an underlying group structure. Here the group is the cube $\{\pm 1\}^n$ with the operation $x \odot z = (x_1 z_1, \ldots, x_n z_n)$. The convolution of a (vector-valued) function $f : \{\pm 1\}^n \to \mathbb{R}^m$ and a (scalar-valued) function $g : \{\pm 1\}^n \to \mathbb{R}$ is the function $f * g : \{\pm 1\}^n \to \mathbb{R}^m$ defined by

$$f * g(x) = \mathbb{E}_{Z}\left[g(Z) \cdot f(x \odot Z)\right]$$

where $Z$ is uniformly random in $\{\pm 1\}^n$. We list some basic properties of convolution.

**Fact 10.** *If $L : \mathbb{R}^m \to \mathbb{R}^m$ is a linear map then $L(f * g) = L(f) * g$.*

**Fact 11.** *One has $\widehat{f * g}(S) = \hat{g}(S) \cdot \hat{f}(S)$ for every $S \subseteq [n]$.*

### 2.2. Norms and convexity.

**Fact 12** (Jensen's inequality). *Given a convex function $f$ and a random variable $X$, we have $f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$.*

**Fact 13.** *For any norm $\|\cdot\| : \mathbb{R}^m \to \mathbb{R}_{\geq 0}$ and functions $f : \{\pm 1\}^n \to \mathbb{R}^m$ and $g : \{\pm 1\}^n \to \mathbb{R}$ one has*

$$\mathbb{E}\left[\|f * g(X)\|^2\right]^{1/2} \leq \mathbb{E}\left[|g(X)|\right] \cdot \mathbb{E}\left[\|f(X)\|^2\right]^{1/2}.$$

*where $X \sim \{\pm 1\}^n$ uniformly.*

*Proof.* We bound

$$\mathbb{E}\left[\|f * g(X)\|^2\right] = \mathbb{E}_{X}\left[\left\|\mathbb{E}_{Z}\left[g(Z) \cdot f(X \odot Z)\right]\right\|^2\right]$$

$$\leq \mathbb{E}_{X}\left[\left(\mathbb{E}_{Z}\left[|g(Z)| \cdot \|f(X \odot Z)\|\right]\right)^2\right],$$

where the inequality follows from the convexity of the norm $\|\cdot\|$. By the Cauchy-Schwarz inequality, we can continue

$$
\begin{aligned}
&\leq \mathop{\mathbb{E}}_{X}\left[\mathop{\mathbb{E}}_{Z}\left[|g(Z)|\right]\cdot\mathop{\mathbb{E}}_{Z'}\left[|g(Z')|\cdot\|f(X\odot Z')\|^2\right]\right]\\
&= \mathop{\mathbb{E}}_{Z}\left[|g(Z)|\right]\cdot\mathop{\mathbb{E}}_{Z'}\left[|g(Z')|\cdot\mathop{\mathbb{E}}_{X}\left[\|f(X)\|^2\right]\right]\\
&= \left(\mathop{\mathbb{E}}_{Z}\left[|g(Z)|\right]\right)^2\cdot\mathop{\mathbb{E}}_{X}\left[\|f(X)\|^2\right]. \qquad\qquad \square
\end{aligned}
$$

It is convenient to replace a norm with the Euclidean norm. For this, we use the following standard result in convex geometry.

**Fact 14** (John's Theorem [13]). *For any norm $\|\cdot\|$ on $\mathbb{R}^m$, there is an invertible linear map $J:\mathbb{R}^m\to\mathbb{R}^m$ such that for every $x\in\mathbb{R}^m$,*

$$
\|J(x)\|_2 \leq \|x\| \leq \sqrt{m}\cdot\|J(x)\|_2.
$$

## 2.3. Some useful inequalities.

**Fact 15** (Stirling's approximation). *For every $n\in\mathbb{N}$, we have*

$$
\sqrt{2\pi}\cdot n^{n+1/2}e^{-n} \leq n! \leq e\cdot n^{n+1/2}e^{-n}.
$$

**Fact 16** (Khintchine's inequality). *Let $Y\in\{\pm 1\}^n$ be uniformly random. For every integer $k>0$, there exist constants $A_k,B_k>0$ such that for every $x\in\mathbb{R}^n$,*

$$
A_k\|x\|_2 \leq \mathbb{E}\left[\Big|\sum_{i=1}^{n}Y_i x_i\Big|^k\right]^{1/k} \leq B_k\|x\|_2.
$$

*We can take $A_1=\frac{1}{\sqrt{2}}$ and $B_k=k!$.*

**Lemma 17** (Chernoff bound). *Let $X_1,\ldots,X_n\in[-1,1]$ be independent random variables with $\mathbb{E}[X_i]=\mu$ for all $i\in[n]$. For every $\lambda\geq 0$,*

$$
\Pr\left[\Big|\mu-\tfrac{1}{n}\sum_{i=1}^{n}X_i\Big|\geq\lambda\right] \leq 2\exp(-\lambda^2 n/2).
$$

**Fact 18** (Bernstein's inequality [2]). *Let $X_1,\ldots,X_n$ be independent zero-mean random variables with $|X_i|\leq M$ for all $i\in[n]$. For every $t>0$,*

$$
\Pr\left[\Big|\sum_{i=1}^{n}X_i\Big|\geq t\right] \leq 2\exp\left(-\frac{t^2}{2\sum_{i=1}^{n}\mathbb{E}[X_i^2]+\frac{2}{3}Mt}\right).
$$

**Fact 19** (Minkowski's inequality). *Let $1\leq p<\infty$, let $\|\cdot\|:\mathbb{R}^m\to\mathbb{R}_{\geq 0}$ be a norm and let $X,Y$ be jointly distributed random variables on $\mathbb{R}^m$ so that $\mathbb{E}[\|X\|^p],\mathbb{E}[\|Y\|^p]<\infty$. Then*

$$
\mathbb{E}[\|X+Y\|^p]^{1/p} \leq \mathbb{E}[(\|X\|+\|Y\|)^p]^{1/p} \leq \mathbb{E}[\|X\|^p]^{1/p}+\mathbb{E}[\|Y\|^p]^{1/p}.
$$

2.4. **Real-rooted polynomials.** A univariate polynomial $p(z)$ over $\mathbb{R}$ is *real-rooted* if for all $w \in \mathbb{C}$, the equality $p(w) = 0$ implies that $w \in \mathbb{R}$. Newton's inequality implies that the coefficients of real-rooted polynomials are log-concave. A sequence $c_0, \ldots, c_d$ is *log-concave* if $c_j^2 \geq c_{j-1} \cdot c_{j+1}$ for $j \in [d-1]$.

**Fact 20** (e.g. [5]). *Let $p(z) = \sum_{j=0}^{d} c_j z^j$ be a real-rooted polynomial with real coefficients. Then the sequence $c_0, \ldots, c_d$ is log-concave.*

An important consequence is that the magnitudes of the coefficients of such polynomials are *unimodal*. A sequence $a_0, \ldots, a_d$ is unimodal if there is an index $m$ such that

$$a_0 \leq a_1 \leq \ldots \leq a_m \geq a_{m+1} \geq \ldots \geq a_d.$$

**Fact 21.** *If $c_0, \ldots, c_d$ is a log-concave sequence of positive numbers, then it is unimodal.*

*Proof.* Since $c_j > 0$, it follows that $c_j/c_{j-1} \geq c_{j+1}/c_j$ for $j \in \{1, \ldots, d-1\}$, that is, the sequence of consecutive ratios is non-increasing. Thus if $m$ is the largest index with $c_m/c_{m-1} \geq 1$, it follows that $c_0 \leq c_1 \leq \cdots \leq c_m \geq c_{m+1} \geq \cdots \geq c_d$.           $\square$

## 3. Bounds on $\|f_\ell\|_\infty$

In this section, we prove Theorem 1 assuming Theorem 7. We start by constructing a proxy $P : \{\pm 1\}^n \to \mathbb{R}$ that filters out $f_\ell$ from $f$ by convolution. If $d = \ell \bmod 2$, we use the parameters $(d, \ell)$ to obtain $\phi$ as in Theorem 7. If $d \neq \ell \bmod 2$, we use the parameters $(d-1, \ell)$ to obtain $\phi$. The proxy is defined as

$$P(x) := \mathbb{E}_\theta \left[ \phi(\theta) \cdot \prod_{j=1}^{n} \left( 1 + x_j \cos \theta \right) \right].$$

Property (3) implies that for any $S \subseteq [n]$,

$$\hat{P}(S) = \begin{cases} 1 & \text{if } |S| = \ell, \\ 0 & \text{if } |S| \neq \ell, \, |S| \leq d. \end{cases}$$

Since $f$ has degree $d$, Fact 11 implies that $f_\ell = f * P$. Because $f$ is bounded, for every $x \in \{\pm 1\}^n$,

$$|f_\ell(x)| = \left| \mathbb{E}\left[ P(Z) \cdot f(x \odot Z) \right] \right| \leq \mathbb{E}\left[ |P(Z)| \right],$$

where $Z \sim \{\pm 1\}^n$ is uniform. Finally,

$$\mathbb{E}\left[ |P(Z)| \right] \leq \mathbb{E}\left[ |\phi(\theta)| \cdot \left| \prod_{j=1}^{n}(1 + Z_j \cos \theta) \right| \right] \qquad \text{triangle inequality}$$

$$= \mathbb{E}\left[ |\phi(\theta)| \cdot \prod_{j=1}^{n}(1 + Z_j \cos(\theta)) \right] \qquad (1 + Z_j \cos(\theta)) \geq 0$$

$$= \mathbb{E}[|\phi(\theta)|] \qquad \mathbb{E}[Z_j] = 0$$

$$\leq \begin{cases} |C(d, \ell)| & \text{when } d = \ell \bmod 2, \\ |C(d-1, \ell)| & \text{otherwise.} \end{cases} \qquad \text{Theorem 7}$$

## 4. Bounds on $\|\hat{f}_\ell\|_1$

Here we prove Theorem 3 assuming Theorem 1. The proof is by induction on $\ell$. When $\ell = 1$, we have $\|\hat{f}_1\|_1 = \|f_1\|_\infty$, because we can pick an input $x$ for which $f_1(x) = \sum_{i=1}^n \hat{f}(\{i\})x_i = \|\hat{f}_1\|_1$. And Theorem 1 implies $\|f_1\|_\infty \leq d$.

For the induction step, let $\ell > 1$. We apply a random restriction to $f_\ell$, and use induction on the degree-$(\ell-1)$ homogenous part of the restricted function. Let $Q \subseteq [n]$ be a subset of the variables sampled by including each variable independently with probability $\frac{1}{\ell}$, and let $Z \sim \{\pm 1\}^n$ be uniformly random and independent of $Q$. The random restriction of $f_\ell$ is

$$g(x) := \sum_{S \subseteq [n]} \hat{f}_\ell(S) \cdot \chi_{S \setminus Q}(x) \cdot \chi_{S \cap Q}(Z).$$

The main idea is to relate $\|\hat{f}_\ell\|_1$ and $\|\hat{g}_{\ell-1}\|_1$.

**Lemma 22.** $\|\hat{f}_\ell\|_1 \leq e \cdot \sqrt{\frac{2n}{\ell}} \cdot \mathbb{E}[\|\hat{g}_{\ell-1}\|_1]$.

Before proving the lemma, we use it to complete the proof. Since $g/\|f_\ell\|_\infty$ is bounded and of degree at most $\ell$,

$$
\begin{aligned}
\|\hat{f}_\ell\|_1 &\leq e \cdot \sqrt{\frac{2n}{\ell}} \cdot \|f_\ell\|_\infty \cdot n^{\frac{\ell-2}{2}} \ell^{\ell-1} e^{\binom{\ell}{2}} && \text{induction \& Lemma 22} \\
&\leq e \cdot \sqrt{\frac{2n}{\ell}} \cdot \frac{d^\ell}{\ell!} \cdot n^{\frac{\ell-2}{2}} \ell^{\ell-1} e^{\binom{\ell}{2}} && \text{Theorem 1} \\
&\leq e \cdot \sqrt{\frac{2}{\ell}} \cdot \frac{d^\ell \cdot e^\ell}{\sqrt{2\pi\ell} \cdot \ell^\ell} \cdot n^{\frac{\ell-1}{2}} \ell^{\ell-1} e^{\binom{\ell}{2}} && \text{Fact 15} \\
&\leq \frac{e}{\sqrt{\pi} \cdot \ell^2} \cdot d^\ell \cdot n^{\frac{\ell-1}{2}} e^{\binom{\ell}{2}+\ell} \\
&\leq d^\ell \cdot n^{\frac{\ell-1}{2}} \cdot e^{\binom{\ell+1}{2}}.
\end{aligned}
$$

*Proof of Lemma 22.* Start by fixing a set $U \subseteq [n]$ of size $\ell - 1$. Denote by $1_{Q \cap U = \emptyset}$ the indicator random variable for the event that $Q \cap U = \emptyset$. The corresponding coefficient in $g$ is

$$\hat{g}(U) = 1_{Q \cap U = \emptyset} \cdot \sum_{j \in Q \setminus U} Z_j \cdot \hat{f}_\ell(U \cup \{j\}).$$

We first fix $Q$ and take the expectation over $Z$. Denote by $S$ the random variable that is zero if $Q = \emptyset$ and is equal to $\frac{1}{\sqrt{|Q|}}$ when $Q$ is not empty. For $j \notin U$, let $S_j$ be the random variable that is zero if $j \notin Q$ and is $\frac{1}{\sqrt{|Q|}}$ when $j \in Q$. For every $Q$, we can

bound

$$\mathbb{E}_Z[|\hat{g}(U)|] = 1_{Q \cap U = \emptyset} \cdot \mathbb{E}_Z\left[\Big| \sum_{j \in Q \setminus U} Z_j \cdot \hat{f}_\ell(U \cup \{j\})\Big|\right]$$

$$\geq 1_{Q \cap U = \emptyset} \cdot \sqrt{\frac{1}{2} \sum_{j \in Q \setminus U} \hat{f}_\ell(U \cup \{j\})^2} \qquad\qquad \text{Fact 16}$$

$$\geq \frac{1}{\sqrt{2}} \cdot 1_{Q \cap U = \emptyset} \cdot S \cdot \sum_{j \in Q \setminus U} |\hat{f}_\ell(U \cup \{j\})| \qquad \text{Cauchy-Schwarz}$$

$$\geq \frac{1}{\sqrt{2}} \cdot 1_{Q \cap U = \emptyset} \cdot \sum_{j \notin U} S_j \cdot |\hat{f}_\ell(U \cup \{j\})|. \qquad\qquad S \geq S_j$$

We now take the expectation over $Q$ as well:

$$\mathbb{E}_{Z,Q}\left[|\hat{g}(U)|\right] \geq \frac{1}{\sqrt{2}} \sum_{j \notin U} |\hat{f}_\ell(U \cup \{j\})| \cdot \mathbb{E}_Q[1_{Q \cap U = \emptyset} \cdot S_j].$$

Because $\xi \mapsto \frac{1}{\sqrt{\xi}}$ is convex, for each $j \notin U$ we use Jensen's inequality to bound

$$\mathbb{E}_Q[1_{Q \cap U = \emptyset} \cdot S_j] = \mathbb{P}[Q \cap U = \emptyset] \cdot \mathbb{P}[j \in Q] \cdot \mathbb{E}_{Q | Q \cap U = \emptyset, j \in Q}\left[\frac{1}{\sqrt{|Q|}}\right]$$

$$\geq \mathbb{P}[Q \cap U = \emptyset] \cdot \mathbb{P}[j \in Q] \cdot \frac{1}{\sqrt{\mathbb{E}_{Q | Q \cap U = \emptyset, j \in Q}[|Q|]}}.$$

We have $\mathbb{P}[j \in Q] = \frac{1}{\ell}$, and $\mathbb{P}[Q \cap U = \emptyset] = \left(1 - \frac{1}{\ell}\right)^{\ell-1} \geq 1/e$, since $(1 - 1/\ell)^{\ell-1}$ is decreasing in $\ell$ and converges to $1/e$. We can compute:

$$\mathbb{E}_{Q | Q \cap U = \emptyset, j \in Q}[|Q|] = 1 + (n - (\ell - 1) - 1)\frac{1}{\ell} = \frac{n}{\ell}.$$

So, we can bound

$$\mathbb{E}_Q[1_{Q \cap U = \emptyset} \cdot S_j] \geq \frac{1}{e\ell} \cdot \sqrt{\frac{\ell}{n}} = \frac{1}{e\sqrt{n\ell}}.$$

Overall, for every $U$ of size $\ell - 1$,

$$\mathbb{E}_{Q,Z}\left[|\hat{g}(U)|\right] \geq \frac{1}{e\sqrt{2n\ell}} \cdot \sum_{j \notin U} |\hat{f}_\ell(U \cup \{j\})|.$$

Summing over $U$,

$$\mathbb{E}_{Q,Z}\left[\|\hat{g}_{\ell-1}\|_1\right] \geq \frac{1}{e\sqrt{2n\ell}} \cdot \sum_{U : |U| = \ell - 1} \sum_{j \notin U} |\hat{f}_\ell(U \cup \{j\})|$$

$$= \frac{1}{e\sqrt{2n\ell}} \cdot \sum_{S : |S| = \ell} \ell \cdot |\hat{f}_\ell(S)|$$

$$= \frac{1}{e} \cdot \sqrt{\frac{\ell}{2n}} \cdot \|\hat{f}_\ell\|_1. \qquad\qquad \square$$

## 5. A Higher level Pisier inequality

In this section, we prove Theorem 5 assuming Theorem 7. We can express $f_\ell$ as the convolution of $f$ with the level function

$$L_\ell(x) := \sum_{S \subseteq [n]: |S| = \ell} \chi_S(x);$$

see Fact 11. In order to analyze the norm of $f_\ell$, we construct a proxy $P$ that is close to $L_\ell$. Let $d$ be a parameter with $d = \ell \bmod 2$ to be determined. Let $\phi$ be as in Theorem 7. Define

$$P(x) := 2^\ell \cdot \mathbb{E}_\theta \left[ \phi(\theta) \cdot \prod_{j=1}^n \left( 1 + \frac{\cos(\theta) \cdot x_j}{2} \right) \right]$$

$$= \sum_{S \subseteq [n]} 2^\ell \mathbb{E}_\theta \left[ \phi(\theta) \cdot \frac{\cos^{|S|}(\theta)}{2^{|S|}} \right] \cdot \chi_S(x).$$

We think of $P$ as a "good" proxy for $L_\ell$:

$$\mathbb{E} \left[ \|f_\ell(X)\|^2 \right]^{1/2} = \mathbb{E} \left[ \|f * L_\ell(X)\|^2 \right]^{1/2}$$

$$= \mathbb{E} \left[ \|f * P(X) + f * (L_\ell - P)(X)\|^2 \right]^{1/2}$$

$$\leq \mathbb{E} \left[ \|f * P(X)\|^2 \right]^{1/2} + \mathbb{E} \left[ \|f * (L_\ell - P)(X)\|^2 \right]^{1/2}. \qquad \text{Fact 19}$$

Next, we bound each of the two terms separately.

To bound the first term, apply Fact 13,

$$\mathbb{E} \left[ \|f * P(X)\|^2 \right]^{1/2} \leq \mathbb{E} \left[ |P(X)| \right] \cdot \mathbb{E} \left[ \|f(X)\|^2 \right]^{1/2}.$$

Similarly to the end of Section 3, we may bound the $\ell_1$ norm of $P$ by

$$\mathbb{E} \left[ |P(X)| \right] \leq 2^\ell \cdot \mathbb{E} \left[ |\phi(\theta)| \cdot \prod_{j=1}^n \left( 1 + \frac{\cos(\theta) \cdot X_j}{2} \right) \right]$$

$$= 2^\ell \cdot \mathbb{E} \left[ |\phi(\theta)| \right] \leq 2^\ell \cdot |C(d, \ell)|.$$

To bound the second term, use John's theorem (Fact 14). There is an invertible linear map $J : \mathbb{R}^m \to \mathbb{R}^m$ so that for every $x \in \mathbb{R}^m$,

$$\|J(x)\|_2 \leq \|x\| \leq \sqrt{m} \cdot \|J(x)\|_2.$$

Using $J$ we can switch between $\|\cdot\|$ and $\|\cdot\|_2$:

$$\mathbb{E} \left[ \|f * (L_\ell - P)(X)\|^2 \right]^{1/2} \leq \sqrt{m} \cdot \mathbb{E} \left[ \|J(f * (L_\ell - P)(X))\|_2^2 \right]^{1/2}$$

$$= \sqrt{m} \cdot \mathbb{E} \left[ \|J(f) * (L_\ell - P)(X)\|_2^2 \right]^{1/2} \qquad \text{Fact 10}$$

$$= \sqrt{m} \cdot \sqrt{\sum_S \|\widehat{J(f)}(S)\|_2^2 \cdot (\widehat{L_\ell - P}(S))^2}. \qquad \text{Facts 9 and 11}$$

We now claim that

$$\hat{P}(S) = \begin{cases} 1 & \text{if } |S| = \ell, \\ 0 & \text{if } |S| \neq \ell, |S| \leq d, \\ \leq 2^{\ell-d} \cdot |C(d,\ell)| & \text{otherwise.} \end{cases}$$

The first two cases follow directly from the properties of $\phi$ and the formula for $P$. The last case follows from

$$|\hat{P}(S)| \leq \frac{2^\ell}{2^{|S|}} \cdot \mathbb{E}\left[|\phi(\theta)|\right] \leq 2^{\ell-d} \cdot |C(d,\ell)|.$$

We can continue to bound

$$\mathbb{E}\left[\|f * (L_\ell - P)(X)\|^2\right]^{1/2} \leq \sqrt{m} \cdot 2^{\ell-d} \cdot |C(d,\ell)| \cdot \sqrt{\sum_S \|\widehat{J(f)}(S)\|_2^2}$$

$$= \sqrt{m} \cdot 2^{\ell-d} \cdot |C(d,\ell)| \cdot \mathbb{E}\left[\|J(f(X))\|_2^2\right]^{1/2} \qquad \text{Fact 9}$$

$$\leq \sqrt{m} \cdot 2^{\ell-d} \cdot |C(d,\ell)| \cdot \mathbb{E}\left[\|f(X)\|^2\right]^{1/2}.$$

Putting the two parts together,

$$\mathbb{E}\left[\|f_\ell(X)\|^2\right]^{1/2} \leq 2^\ell \cdot |C(d,\ell)| \cdot \left(1 + \frac{\sqrt{m}}{2^d}\right) \mathbb{E}\left[\|f(X)\|^2\right]^{1/2}.$$

For $\ell < \frac{1}{2}\log(m+1)$, we can set $d$ to be the smallest integer that is larger than $\frac{1}{2}\log(m+1)$ and has the same parity as $\ell$, so that

$$2^\ell \cdot |C(d,\ell)| \cdot \left(1 + \frac{\sqrt{m}}{2^d}\right) \leq 2^\ell \cdot \frac{d^\ell}{\ell!} \cdot 2 \qquad \text{Fact 15}$$

$$\leq \left(\frac{6\log(m+1)}{\ell}\right)^\ell.$$

For $\ell \geq \frac{1}{2}\log(m+1)$, we can set $d := \ell$ so that

$$2^\ell \cdot |C(d,\ell)| \cdot \left(1 + \frac{\sqrt{m}}{2^d}\right) = 2^{2\ell-1} \cdot \left(1 + \frac{\sqrt{m}}{2^\ell}\right) \leq 4^\ell.$$

**Remark.** *There is a slightly more general version of Theorem 5. The Banach-Mazur distance of the norm $\|\cdot\|$ from the Euclidean norm $\|\cdot\|_2$ is*

$$D = \inf\{d \in \mathbb{R} : \exists T \in \mathsf{GL}_m \ \forall x \in \mathbb{R}^m \ \|T(x)\|_2 \leq \|x\| \leq d \cdot \|T(x)\|_2\},$$

*where $\mathsf{GL}_m$ is the group of invertible linear transformations from $\mathbb{R}^m$ to itself. John's theorem states that always $D \leq \sqrt{m}$. The above argument proves that, more generally, we can replace the $C\log(m+1)$ term by $C\log(D+1)$.*

## 6. CONSTRUCTING THE FILTER

Here we construct the filter $\phi$ and prove Theorem 7. Let $\theta$ be uniformly distributed over the $2d$ equally spaced angles

$$\mathcal{D} = \left\{0, \frac{\pi}{d}, \ldots, \frac{(2d-1)\pi}{d}\right\}.$$

An important property of this distribution is that for integer $a$, we have

(5)
$$\mathbb{E}[e^{ia\theta}] = \begin{cases} 1 & \text{if } a = 0 \bmod 2d \\ 0 & \text{otherwise.} \end{cases}$$

Define

$$Q(z) := \prod_{j=0}^{d}\left(z - \cos\left(\frac{j\pi}{d}\right)\right) = \sum_{j=0}^{d+1} c_j z^j,$$

for some $c_j \in \mathbb{R}$. Let $Q_{>\ell}$ denote the suffix of $Q$:

$$Q_{>\ell}(z) := \sum_{j=\ell+1}^{d+1} c_j z^j.$$

The rational function $\frac{Q_{>\ell}(z)}{z^{\ell+1}}$ is a polynomial. Finally, define

$$\phi(\theta) := 2^{d-1} \cdot \cos(d\theta) \cdot \frac{Q_{>\ell}(\cos(\theta))}{\cos^{\ell+1}(\theta)}.$$

It remains to prove that the filter $\phi$ has the desired properties.

The following claim helps to understand the correlation of $\phi$ with powers of cos.

**Claim 23.** *For integers $k, d \geq 0$, we have*

$$\mathbb{E}[\cos(d\theta)\cos^k(\theta)] = \begin{cases} 0 & \text{if } k \neq d \bmod 2 \\ 0 & \text{if } k < d \\ 2^{-(d-1)} & \text{if } k = d. \end{cases}$$

*Proof.* If $k \neq d \bmod 2$, the symmetry $\cos(d\theta) = (-1)^d \cos(d(\pi + \theta))$ and the symmetry of the distribution of $\theta$ complete the proof.

For $k < d$, we use the identity $\cos(\theta) = \frac{e^{i\theta} + e^{-i\theta}}{2}$. Property (5) implies

$$\mathbb{E}[\cos(d\theta)\cos^k(\theta)] = \mathbb{E}\left[\frac{e^{id\theta} + e^{-id\theta}}{2} \cdot \left(\frac{e^{i\theta} + e^{-i\theta}}{2}\right)^k\right] = 0.$$

For $k = d$, the expectation reduces via (5) to

$$\mathbb{E}\left[\frac{e^{id\theta} + e^{-id\theta}}{2} \cdot \left(\frac{e^{i\theta} + e^{-i\theta}}{2}\right)^d\right] = \mathbb{E}\left[\frac{e^{id\theta} + e^{-id\theta}}{2} \cdot \left(\frac{e^{id\theta} + e^{-id\theta}}{2^d}\right)\right] = 2^{-(d-1)}.$$
$\square$

Now, we can prove (3). The argument is based on Claim 23. We use the following terminology. The expressions we consider are sums of terms of the form $\cos(d\theta)\cos^k(\theta)$. The *degree* of such a term is $k$.

For $k \leq \ell - 1$, all terms in $\phi(\theta)\cos^k(\theta)$ have degree at most $d - 1$. Claim 23 implies that $\mathbb{E}[\phi(\theta)\cos^k(\theta)] = 0$.

For $k = \ell$, we have a single term of degree $d$, so that

$$\mathbb{E}[\phi(\theta)\cos^\ell(\theta)] = \mathbb{E}[2^{d-1}\cos(d\theta)\cos^d(\theta)] = 1.$$

For $\ell + 1 \leq k \leq d$,

$$\mathbb{E}[\phi(\theta)\cos^k(\theta)] = \mathbb{E}[2^{d-1}\cos(d\theta) \cdot Q_{>\ell}(\cos(\theta))\cos^{k-(\ell+1)}(\theta)]$$
$$= \mathbb{E}[2^{d-1}\cos(d\theta) \cdot \underbrace{Q(\cos(\theta))}_{=0}\cos^{k-(\ell+1)}(\theta)] = 0;$$

the second equality holds because we added terms in $Q$ of degree at most $\ell$, and $\ell + k - (\ell + 1) \leq d - 1$.

Finally, for $k = d + 1$, we need one more observation. Since $\cos(\theta) = -\cos(\pi - \theta)$, the distinct roots of the real-rooted polynomial $Q$ come in pairs of the form $r, -r$. So, there is a polynomial $q$ so that

$$Q(z) = \begin{cases} z \cdot q(z^2) & \text{if } d = 0 \bmod 2, \\ q(z^2) & \text{otherwise,} \end{cases}$$

where

$$q(z) = \prod_{j=0}^{\lfloor (d-1)/2 \rfloor} \left( z - \cos^2\left(\frac{j\pi}{d}\right) \right).$$

Because $\ell = d \bmod 2$, the coefficient $c_\ell$ in $Q$ is zero. Similarly to the previous case, we can bound

$$\mathbb{E}[\phi(\theta)\cos^k(\theta)] = \mathbb{E}[2^{d-1}\cos(d\theta) \cdot Q_{>\ell}(\cos(\theta))\cos^{k-(\ell+1)}(\theta)]$$
$$= \mathbb{E}[2^{d-1}\cos(d\theta) \cdot \underbrace{Q(\cos(\theta))}_{=0}\cos^{k-(\ell+1)}(\theta)] = 0;$$

here we additionally used that $c_\ell = 0$.

Next, we turn to computing $\mathbb{E}[|\phi(\theta)|]$. The key claim is the following:

**Claim 24.** *For all $\theta \in \mathcal{D}$, the sign of $\frac{Q_{>\ell}(\cos(\theta))}{\cos^{\ell+1}(\theta)}$ is the same.*

*Proof.* The polynomial $q$ has positive roots corresponding to nonzero $\cos^2(\theta)$ for $\theta \in \mathcal{D}$. Because $d = \ell \bmod 2$, the sign of $\frac{Q_{>\ell}(\cos(\theta))}{\cos^{\ell+1}(\theta)}$ is the same as the sign of $q_{>k}(\cos^2(\theta))$ for $k = \lfloor \frac{\ell-1}{2} \rfloor$. Theorem 8 completes the proof. $\qquad\square$

Claim 24 implies that the sign of $\phi(\theta)$ is determined by the sign of $\cos(d\theta)$. We can finally compute

$$\mathbb{E}[|\phi(\theta)|] = |\mathbb{E}[\phi(\theta)\cos(d\theta)]| \qquad\qquad |\cos(d\theta)| = 1 \text{ for } \theta \in \mathcal{D}$$
$$= |C(d, \ell)|. \qquad\qquad\qquad \text{property (3) \& definition of } T_d$$

## 7. On Real Rooted Polynomials

In this section, we prove Theorem 8. First, we need a useful property of unimodal sequences.

**Claim 25.** *Let $a_0, \ldots, a_d$ be a unimodal sequence of positive numbers so that $\sum_{j=0}^d (-1)^j a_j = 0$. Then for all $k \in \{0, \ldots, d\}$, we have $(-1)^k \sum_{j=0}^k (-1)^j a_j \geq 0$.*

*Proof.* Let $m$ be the position of a maximum of the unimodal sequence. For $k \leq m$ even,

$$\sum_{j=0}^{k} (-1)^j a_j = a_0 + \sum_{j=1}^{\frac{k}{2}} (a_{2j} - a_{2j-1}) \geq 0.$$

For $k \leq m$ odd,

$$\sum_{j=0}^{k} (-1)^j a_j = \sum_{j=0}^{\frac{k-1}{2}} (a_{2j+1} - a_{2j}) \leq 0.$$

This proves the claim when $k \leq m$. A symmetric argument can be applied to the suffix sums to conclude that for $m \leq k < d$,

$$(-1)^{d-k+1} \sum_{j=k+1}^{d} (-1)^{d-j} a_j \geq 0.$$

Together with the condition $\sum_{j=0}^{d} (-1)^j a_j = 0$, this implies that when $k > m$,

$$(-1)^k \cdot \sum_{j=0}^{k} (-1)^j a_j = (-1)^{k+1} \cdot \sum_{j=k+1}^{d} (-1)^j a_j$$

$$= (-1)^{d-k+1} \sum_{j=k+1}^{d} (-1)^{d-j} a_j \geq 0.$$

$\square$

Now we turn to proving Theorem 8.

*Proof of Theorem 8.* Write $p$ as

$$p(z) = \prod_{j=1}^{d} (z - r_j) = \sum_{j=0}^{d} c_j z^j,$$

with $r_1, \ldots, r_d > 0$. For every $j \in \{0, \ldots, d\}$, we have $(-1)^{d+j} \cdot c_j > 0$. So, by Fact 20, the sequence $|c_0|, \ldots, |c_d|$ is log-concave.

Now, let $r$ by any root of $p$, and set $a_j = |c_j| r^j$. Because the product of log-concave sequences is log-concave, the sequence $a_0, \ldots, a_d$ is log-concave and positive.

By Fact 21, the sequence $a_0, \ldots, a_d$ is unimodal. Because $r$ is a root of $p$, we know $p(r) = \sum_{j=0}^{d}(-1)^j a_j = 0$. And Claim 25 implies

$$
\begin{aligned}
(-1)^{d-k-1} \cdot p_{>k}(r) &= (-1)^{d-k-1} \cdot \sum_{j=k+1}^{d} c_j r^j \\
&= (-1)^{-k-1} \cdot \sum_{j=k+1}^{d} (-1)^j a_j \\
&= (-1)^k \cdot \sum_{j=0}^{k} (-1)^j a_j \\
&\geq 0. \qquad \qquad \square
\end{aligned}
$$

## 8. Consequences for learning

Here we describe an application of our Fourier bounds to learning theory; we prove Theorem 6. The learning algorithm is based on standard techniques (see e.g. [14] or Chapter 3 in [20]).

First, we can estimate one specific Fourier coefficient by sampling and averaging.

**Lemma 26.** *Let $f : \{-1, 1\}^n \to [-1, 1]$ and fix a set $S \subseteq [n]$. Sample $X_1, \ldots, X_N \sim \{\pm 1\}^n$ uniformly and independently and set $\alpha_S := \frac{1}{N} \sum_{i=1}^{N} f(X_i) \cdot \chi_S(X_i)$. Then for any $\lambda \geq 0$,*

$$
\Pr\left[|\hat{f}(S) - \alpha_S| \geq \lambda\right] \leq 2 \exp(-\lambda^2 N / 2).
$$

*Proof.* Consider the random variable $Y_i := f(X_i) \cdot \chi_S(X_i)$ and note that $|Y_i| \leq 1$ and $\mathbb{E}[Y_i] = \hat{f}(S)$. The lemma follows from the Chernoff bound (Lemma 17). $\square$

The learning algorithm operates as follows. Its sample complexity is

$$
N = \left\lceil 2 \cdot 16^2 \cdot \frac{1}{\epsilon^3} \cdot L(n,d)^2 \log\left(2 \cdot \frac{\sum_{\ell=0}^{d} \binom{n}{d}}{\delta}\right)\right\rceil
$$

where

$$
L(n,d) := (d+1)d^d e^{\binom{d+2}{2}} \cdot n^{\frac{d-1}{2}}.
$$

The algorithm samples $X_1, \ldots, X_N \sim \{\pm 1\}^n$ uniformly at random and independently. It computes $\alpha_S$ for all $S \subseteq [n]$ of size $|S| \leq d$ as in the lemma above. It then computes the set

$$
B = \left\{S \subseteq [n] : |\alpha_S| \geq \frac{\epsilon}{4 \cdot L(n,d)}\right\}.
$$

The output is the function

$$
g = \sum_{S \in B} \alpha_S \chi_S.
$$

It remains to prove that, except with probability $\delta$, the algorithm above produces a function $g$ satisfying $\mathbb{E}\left[|f(X) - g(X)|^2\right] \leq \epsilon$, for uniformly random $X$.

Denote by $G$ the event that for every $S$ of size at most $d$ we have $|\hat{f}(S) - \alpha_S| \le \frac{\epsilon^{3/2}}{16 \cdot L(n,d)}$. The union bound and Lemma 26 imply that $\mathbb{P}[G] \ge 1 - \delta$. For the rest of the proof, assume that $G$ holds. For $S \in B$, we have

$$\hat{f}(S)^2 \ge \left( \frac{\epsilon}{4 \cdot L(n,d)} - \frac{\epsilon^{3/2}}{16 \cdot L(n,d)} \right)^2 \ge \frac{\epsilon^2}{64 \cdot L(n,d)^2}.$$

So, by Parseval's identity, we must have that

$$|B| \le 64 \cdot \frac{L(n,d)^2}{\epsilon^2}.$$

For $S \notin B$,

$$|\hat{f}(S)| \le \frac{\epsilon}{4 \cdot L(n,d)} + \frac{\epsilon^{3/2}}{16 \cdot L(n,d)} \le \frac{\epsilon}{2 \cdot L(n,d)}.$$

The last ingredient is Theorem 3. It implies that

$$\|\hat{f}\|_1 \le L(n,d).$$

Putting it all together,

$$\mathbb{E}\left[ |f(X) - g(X)|^2 \right] = \sum_{S \subseteq [n]} (\hat{f}(S) - \hat{g}(S))^2$$

$$= \sum_{S \in B} (\hat{f}(S) - \hat{g}(S))^2 + \sum_{S \notin B} \hat{f}(S)^2$$

$$\le |B| \cdot \frac{\epsilon^3}{16^2 \cdot L(n,d)^2} + \frac{\epsilon}{2 \cdot L(n,d)} \cdot \sum_{S \notin B} |\hat{f}(S)|$$

$$\le \epsilon.$$

## 9. Examples of bounded functions

In this section we provide a couple of examples showing that our bounds are sharp for some range of parameters.

9.1. **Lower bound for** $\|\hat{f}_\ell\|_1$**.** Here we prove Proposition 4. Let $\epsilon_S$, for $S \subseteq [n]$ of size $\ell$, be sampled uniformly and independently from $\{\pm 1\}$. Define $G(x) := \sum_S \epsilon_S \chi_S(x)$ where the sum is over $S \subseteq [n]$ of size $\ell$. By Fact 18, for each $x \in \{\pm 1\}^n$ we may bound $\Pr\left[ |G(x)| \ge 2\sqrt{n \cdot \binom{n}{\ell}} \right] < 2^{-n}$. By the union bound, there is a choice for $\epsilon_S$ so that the map

$$f := \frac{G}{2\sqrt{n \cdot \binom{n}{\ell}}}$$

satisfies $\|f\|_\infty \le 1$ and

$$\|\hat{f}_\ell\|_1 = \frac{\binom{n}{\ell}}{2 \cdot \sqrt{n \cdot \binom{n}{\ell}}} = \frac{1}{2} \cdot \sqrt{\frac{1}{n} \cdot \binom{n}{\ell}}.$$

9.2. **Lower bound for** $\|f_\ell\|_\infty$. Here we prove Proposition 2. Every coefficient of the Chebyshev polynomial $T_d$ is bounded by $\frac{d^d}{d!} \le e^d$. The theorem follows from the following more general lemma.

**Lemma 27.** *Given positive integers $\ell \le d$ and a degree $d$ polynomial $T(x) = \sum_{j=0}^{d} c_j x^j$, define $g(x) := T((x_1 + \cdots + x_n)/n)$. Then $|g_\ell(1^n)| \ge |c_\ell| - \frac{2(d+1)! \cdot \max_{j \ge \ell} |c_j|}{n}$.*

To prove the lemma, we first show:

**Claim 28.** *Let $S \subseteq [n]$ be of size $\ell$, and $h_j(x) = (x_1 + \cdots + x_n)^j$. Then:*

$$
\hat{h}_j(S) \begin{cases} = 0 & \text{if } j < \ell \text{ or } j \ne \ell \bmod 2, \\ = \ell! & \text{if } j = \ell, \\ \le j! \cdot n^{\frac{j-\ell}{2}} & \text{if } j > \ell. \end{cases}
$$

*Proof.* Let $X \sim \{\pm 1\}^n$ be uniformly distributed. We have

$$
\hat{h}_j(S) = \mathbb{E}\left[\chi_S(X) \cdot h_j(X)\right].
$$

Each $y \in [n]^j$ corresponds to the term $\prod_{i=1}^{j} x_{y_i}$ in the expansion of $h_j(x)$. This term contributes either 1 or 0 to the expectation, and it contributes 1 exactly when every variable of $S$ has odd degree, and all other variables have even degree. Thus, we must have $\hat{h}_j(S) = 0$ when $j < |S|$ or $j \ne |S| \bmod 2$, since no term can contribute 1 in those cases. Moreover, when $j = |S|$, we see that there are exactly $\ell!$ terms that can contribute 1.

When $j > \ell$, observe that if $\prod_{i=1}^{j} x_{y_i}$ contributes 1, there must be a set $W \subseteq [j]$ of size $\ell$, such that $\prod_{i \in W} x_{y_i} = \prod_{i \in S} x_i$, and every variable of $\prod_{i \notin W} x_{y_i}$ has even degree. The number of choices for $W$ is $\binom{j}{\ell}$, and the number of ways in which $\prod_{i \in W} x_{y_i} = \prod_{i \in S} x_i$ can hold is $\ell!$. For a fixed value of $W$, the number of ways in which $\prod_{i \notin W} x_{y_i}$ can have even degrees is $\mathbb{E}\left[(X_1 + \cdots + X_n)^{j-\ell}\right]$. Putting these observations together:

$$
\hat{h}_j(S) \le \binom{j}{\ell} \cdot \ell! \cdot \mathbb{E}\left[(X_1 + \cdots + X_n)^{j-\ell}\right] \le \binom{j}{\ell} \cdot \ell! \cdot (j-\ell)! \cdot n^{\frac{j-\ell}{2}} = j! \cdot n^{\frac{j-\ell}{2}};
$$

the second inequality follows from Khintchine's inequality (Fact 16). $\qquad \square$

Now we can use the claim to prove the lemma:

*Proof of Lemma 27.* The lemma trivially holds when $n < 4$, so we assume $n \ge 4$. Let $h_j$ be as in Claim 28. Note that $\hat{h}_j(S) \ge 0$. We can bound

$$
g_\ell(1^n) = \|\hat{g}_\ell\|_1 = \sum_{S \subseteq [n], |S| = \ell} \sum_{j=0}^{n} \hat{h}_j(S)/n^j
$$

$$
\ge c_\ell \binom{n}{\ell} \cdot \frac{\ell!}{n^\ell} - \binom{n}{\ell} \cdot \sum_{j=\ell+1}^{n} |c_j| \cdot j! \cdot n^{-(j+\ell)/2}.
$$

To bound the first term, observe that

$$c_\ell \binom{n}{\ell} \cdot \frac{\ell!}{n^\ell} = c_\ell \Big(1 - \frac{1}{n}\Big) \cdots \Big(1 - \frac{\ell-1}{n}\Big)$$

$$\geq c_\ell \Big(1 - \frac{1}{n} \cdot \sum_{j=1}^{\ell-1} j\Big) \qquad\qquad (*)$$

$$= c_\ell \Big(1 - \frac{1}{n} \binom{\ell}{2}\Big);$$

the inequality $(*)$ follows by induction from $(1-\alpha)(1-\beta) > 1 - \alpha - \beta$ for $\alpha, \beta > 0$. To bound the contribution of the second term, observe:

$$\binom{n}{\ell} \cdot \sum_{j=\ell+1}^{n} |c_j| \cdot j! \cdot n^{-(j+\ell)/2} = \frac{\binom{n}{\ell}}{n^\ell} \cdot \sum_{j=\ell+2}^{d} j! \cdot n^{-(j-\ell)/2} \cdot |c_j|$$

$$\leq \sum_{k=0}^{d-\ell-2} j! \cdot \frac{1}{n} \cdot n^{-k/2} \cdot |c_j|$$

$$\leq \frac{1}{n} \cdot (d+1)! \cdot \max_{k>\ell} |c_k|.$$

Finally, since $\binom{\ell}{2} \leq (d+1)!$, we get $g_\ell(1^n) \geq c_\ell - \frac{2(d+1)! \max_{k \geq \ell} |c_k|}{n}$. $\qquad\square$

## ACKNOWLEDGEMENTS

## REFERENCES

[1] A. Backurs and M. Bavarian. On the sum of L1 influences. In *IEEE 29th Conference on Computational Complexity, CCC 2014, Vancouver, BC, Canada, June 11-13, 2014*, pages 132–143. IEEE Computer Society, 2014.

[2] S. N. Bernstein. Gastehizdat Publishing House, 1946.

[3] J. Bourgain. On martingales transforms in finite dimensional lattices with an appendix on the k-convexity constant. *Mathematische Nachrichten*, 119(1):41–53, 1984.

[4] J. Bourgain and V. D. Milman. New volume ratio properties for convex symmetric bodies in $r^n$. *Inventiones mathematicae*, 88(2):319–340, 1987.

[5] P. Brändén. Unimodality, log-concavity, real-rootedness and beyond. *Handbook of Enumerative Combinatorics*, 10 2014.

[6] E. Chattopadhyay, J. Gaitonde, C. H. Lee, S. Lovett, and A. Shetty. Fractional pseudorandom generators from any fourier level. *CoRR*, abs/2008.01316, 2020.

[7] E. Chattopadhyay, P. Hatami, K. Hosseini, and S. Lovett. Pseudorandom generators from polarizing random walks. *Theory Comput.*, 15:1–26, 2019.

[8] E. Chattopadhyay, P. Hatami, O. Reingold, and A. Tal. Improved pseudorandomness for unordered branching programs through local monotonicity. In I. Diakonikolas, D. Kempe, and M. Henzinger, editors, *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 363–375. ACM, 2018.

[9] T. Figiel and N. Tomczak-Jaegermann. Projections onto hilbertian subspaces of banach spaces. *Israel Journal of Mathematics*, 33(2):155–171, 1979.

[10] Y. Filmus, H. Hatami, N. Keller, and N. Lifshitz. Bounds on the sum of L1 influences. *Israel Journal of Matematics*, 214(1):167–192, 2016.

[11] U. Girish, A. Tal, and K. Wu. Fourier growth of parity decision trees. *CoRR*, abs/2103.11604, 2021.

[12] P. Gopalan, R. A. Servedio, and A. Wigderson. Degree and sensitivity: Tails of two distributions. In R. Raz, editor, *31st Conference on Computational Complexity, CCC 2016, May 29 to June 1, 2016, Tokyo, Japan*, volume 50 of *LIPIcs*, pages 13:1–13:23. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2016.

[13] F. John. Extremum problems with inequalities as subsidiary conditions. In *Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948*, pages 187–204. Interscience Publishers, Inc., New York, N. Y., 1948.

[14] Y. Mansour. An o(n$^{log\ log\ n}$) learning algorithm for DNF under the uniform distribution. In D. Haussler, editor, *Proceedings of the Fifth Annual ACM Conference on Computational Learning Theory, COLT 1992, Pittsburgh, PA, USA, July 27-29, 1992*, pages 53–61. ACM, 1992.

[15] Y. Mansour. Learning boolean functions via the fourier transform. In *Theoretical advances in neural computation and learning*, pages 391–424. Springer, 1994.

[16] J. C. Mason. *Chebyshev polynomials*. Chapman and Hall CRC, Boca Raton, Fla, 2003.

[17] V. Milman. Almost euclidean quotient spaces of subspaces of a finite-dimensional normed space. *Proceedings of the American Mathematical Society*, 94(3):445–449, 1985.

[18] V. D. Milman. Inégalité de Brunn-Minkowski inverse et applications à la théorie locale des espaces normés. *C. R. Acad. Sci. Paris Sér. I Math.*, 302(1):25–28, 1986.

[19] V. D. Milman. Isomorphic symmetrization and geometric inequalities. In J. Lindenstrauss and V. D. Milman, editors, *Geometric Aspects of Functional Analysis*, pages 107–131, Berlin, Heidelberg, 1988. Springer Berlin Heidelberg.

[20] R. O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.

[21] G. Pisier. Sur les espaces de banach $k$-convexes. *Séminaire Analyse fonctionnelle (dit" Maurey-Schwartz")*, pages 1–15, 1979.

[22] G. Pisier. Un théorème sur les opérateurs linéaires entre espaces de banach qui se factorisent par un espace de hilbert. In *Annales scientifiques de l'École Normale Supérieure*, volume 13, pages 23–43, 1980.

[23] O. Reingold, T. Steinke, and S. P. Vadhan. Pseudorandomness for regular branching programs via fourier analysis. In P. Raghavendra, S. Raskhodnikova, K. Jansen, and J. D. P. Rolim, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques - 16th International Workshop, APPROX 2013, and 17th International Workshop, RANDOM 2013, Berkeley, CA, USA, August 21-23, 2013. Proceedings*, volume 8096 of *Lecture Notes in Computer Science*, pages 655–670. Springer, 2013.

[24] A. A. Sherstov, A. A. Storozhenko, and P. Wu. An optimal separation of randomized and quantum query complexity. *CoRR*, abs/2008.10223, 2020.

[25] A. Tal. Towards optimal separations between quantum and randomized query complexities. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 228–239. IEEE, 2020.

School of Computer Science, University of Washington
*E-mail address*: sviyer97@gmail.com

School of Computer Science, University of Washington
*E-mail address*: anuprao@cs.washington.edu

School of Computer Science, University of Washington
*E-mail address*: voreis@cs.washington.edu

School of Computer Science, University of Washington
*E-mail address*: rothvoss@uw.edu

Department of Mathematics, Technion-IIT
*E-mail address*: amir.yehudayoff@gmail.com