Fourier Growth of Structured $\mathbb{F}_2$-Polynomials and Applications

Jarosław Błasiok
Columbia University, New York, NY, USA

Peter Ivanov
Northeastern University, Boston, MA, USA

Yaonan Jin
Columbia University, New York, NY, USA

Chin Ho Lee
Columbia University, New York, NY, USA

Rocco A. Servedio
Columbia University, New York, NY, USA

Emanuele Viola
Northeastern University, Boston, MA, USA

Abstract

We analyze the *Fourier growth*, i.e. the L_1 Fourier weight at level k (denoted $L_{1,k}$), of various well-studied classes of "structured" $\mathbb{F}_2$-polynomials. This study is motivated by applications in pseudorandomness, in particular recent results and conjectures due to [9, 10, 8] which show that upper bounds on Fourier growth (even at level $k = 2$) give unconditional pseudorandom generators.

Our main structural results on Fourier growth are as follows:

- We show that any symmetric degree-d $\mathbb{F}_2$-polynomial p has $L_{1,k}(p) \leq \mathbf{Pr}[p = 1] \cdot O(d)^k$. This quadratically strengthens an earlier bound that was implicit in [33].
- We show that any read-Δ degree-d $\mathbb{F}_2$-polynomial p has $L_{1,k}(p) \leq \mathbf{Pr}[p = 1] \cdot (k\Delta d)^{O(k)}$.
- We establish a composition theorem which gives $L_{1,k}$ bounds on disjoint compositions of functions that are closed under restrictions and admit $L_{1,k}$ bounds.

Finally, we apply the above structural results to obtain new unconditional pseudorandom generators and new correlation bounds for various classes of $\mathbb{F}_2$-polynomials.

2012 ACM Subject Classification Theory of computation → Pseudorandomness and derandomization; Theory of computation → Circuit complexity

Keywords and phrases Fourier analysis, Pseudorandomness, Fourier growth

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2021.53

Category RANDOM

Related Version *Full Version*: http://arxiv.org/abs/2107.10797

Funding *Jarosław Błasiok*: Junior Fellowship from the Simons Society of Fellows.
Peter Ivanov: NSF awards CCF-1813930 and CCF-2114116.
Yaonan Jin: NSF IIS-1838154, NSF CCF-1703925, NSF CCF-1814873 and NSF CCF-1563155.
Chin Ho Lee: Croucher Foundation and Simons Collaboration on Algorithms and Geometry.
Rocco A. Servedio: NSF grants CCF-1814873, IIS-1838154, CCF-1563155, and Simons Collaboration on Algorithms and Geometry.
Emanuele Viola: NSF awards CCF-1813930 and CCF-2114116.

Acknowledgements We thank Shivam Nadimpalli for stimulating discussions at the early stage of the project.

1 Introduction

1.1 Background: L_1 Fourier norms and Fourier growth

Over the past several decades, Fourier analysis of Boolean functions has emerged as a fundamental tool of great utility across many different areas within theoretical computer science and mathematics. Areas of application include (but are not limited to) combinatorics, the theory of random graphs and statistical physics, social choice theory, Gaussian geometry and the study of metric spaces, cryptography, learning theory, property testing, and many branches of computational complexity such as hardness of approximation, circuit complexity, and pseudorandomness. The excellent book of O'Donnell [29] provides a broad introduction. In this paper we follow the notation of [29], and for a Boolean-valued function f on n Boolean variables and $S \subseteq [n]$, we write $\widehat{f}(S)$ to denote the Fourier coefficient of f on S.

Given the wide range of different contexts within which the Fourier analysis of Boolean functions has been pursued, it is not surprising that many different quantitative parameters of Boolean functions have been analyzed in the literature. In this work we are chiefly interested in the L_1 *Fourier norm at level* k:

▶ **Definition 1** (L_1 Fourier norm at level k)**.** *The* L_1 Fourier norm of a function $f\colon \{-1,1\}^n \to \{0,1\}$ at level k *is the quantity*

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

For a function class $\mathcal{F}$*, we write* $L_{1,k}(\mathcal{F})$ *to denote* $\max_{f \in \mathcal{F}} L_{1,k}(f)$*.*

As we explain below, strong motivation for studying the L_1 Fourier norm at level k (even for specific small values of k such as $k = 2$) is given by exciting recent results in unconditional *pseudorandomness*. More generally, the notion of *Fourier growth* is a convenient way of capturing the L_1 Fourier norm at level k for every k:

▶ **Definition 2** (Fourier growth)**.** *A function class* $\mathcal{F} \subseteq \{f\colon \{-1,1\}^n \to \{0,1\}\}$ *has* Fourier growth $L_1(a,b)$ *if there exist constants* a *and* b *such that* $L_{1,k}(\mathcal{F}) \leq a \cdot b^k$ *for every* k*.*

The notion of Fourier growth was explicitly introduced by Reingold, Steinke, and Vadhan in [33] for the purpose of constructing pseudorandom generators for space-bounded computation (though we note that the Fourier growth of DNF formulas was already analyzed in [26], motivated by applications in learning theory). In recent years there has been a surge of research interest in understanding the Fourier growth of different types of functions [38, 19, 11, 22, 17, 39, 36, 16]. One strand of motivation for this study has come from the study of quantum computing; in particular, bounds on the Fourier growth of AC^0 [38] were used in the breakthrough result of Raz and Tal [31] which gave an oracle separation between the classes BQP and PH. More recently, in order to achieve an optimal separation between quantum and randomized query complexity, several researchers [39, 1, 36] have studied the Fourier growth of decision trees, with the recent work of [36] obtaining optimal bounds. Analyzing the Fourier growth of other classes of functions has also led to separations between quantum and classical computation in other settings [16, 17, 18].

Our chief interest in the current paper arises from a different line of work which has established powerful applications of Fourier growth bounds in pseudorandomness. We describe the relevant background, which motivates a new conjecture that we propose on Fourier growth, in the next subsection.

1.2 Motivation for this work: Fourier growth, pseudorandomness, $\mathbb{F}_2$-polynomials, and the CHLT conjecture

1.2.1 Pseudorandom generators from Fourier growth bounds

Constructing explicit, unconditional pseudorandom generators (PRGs) for various classes of Boolean functions is an important goal in complexity theory. In the recent work [9], Chattopadhyay, Hatami, Hosseini, and Lovett introduced a novel framework for the design of such PRGs. Their approach provides an explicit pseudorandom generator for any class of functions that is closed under restrictions and has bounded Fourier growth:

▶ **Theorem 3** (PRGs from Fourier growth: Theorem 23 of [9]). *Let $\mathcal{F}$ be a family of n-variable Boolean functions that is closed under restrictions and has Fourier growth $L_1(a,b)$. Then there is an explicit pseudorandom generator that ϵ-fools $\mathcal{F}$ with seed length $O(b^2\log(n/\epsilon)(\log\log n + \log(a/\epsilon)))$.*

Building on Theorem 3, in [10] Chattopadhyay, Hatami, Lovett, and Tal showed that in fact it suffices to have a bound just on $L_{1,2}(\mathcal{F})$ in order to obtain an efficient PRG for $\mathcal{F}$:

▶ **Theorem 4** (PRGs from L_1 Fourier norm bounds at level $k=2$: Theorem 2.1 of [10]). *Let $\mathcal{F}$ be a family of n-variable Boolean functions that is closed under restrictions and has $L_{1,2}(\mathcal{F}) \le t$. Then there is an explicit pseudorandom generator that ϵ-fools $\mathcal{F}$ with seed length $O((t/\epsilon)^{2+o(1)} \cdot \mathrm{polylog}(n))$.*

Observe that while Theorem 4 requires a weaker structural result than Theorem 3 (a bound only on $L_{1,2}(\mathcal{F})$ as opposed to $L_{1,k}(\mathcal{F})$ for all $k \ge 1$), the resulting pseudorandom generator is quantitatively weaker since it has seed length polynomial rather than logarithmic in the error parameter $1/\epsilon$. Even more recently, in [8] Chattopadhyay, Gaitonde, Lee, Lovett, and Shetty further developed this framework by interpolating between the two results described above. They showed that a bound on $L_{1,k}$[1] for any $k \ge 3$ suffices to give a PRG, with a seed length whose ϵ-dependence scales with k:

▶ **Theorem 5** (PRGs from L_1 Fourier norm bounds up to level k for any k: Theorem 4.3 of [8]). *Let $\mathcal{F}$ be a family of n-variable Boolean functions that is closed under restrictions and has $L_{1,k}(\mathcal{F}) \le b^k$ for some $k \ge 3$. Then there exists a pseudorandom generator that ϵ-fools $\mathcal{F}$ with seed length $O\left(\frac{b^{2+\frac{4}{k-2}} \cdot k \cdot \mathrm{polylog}(\frac{n}{\epsilon})}{\epsilon^{\frac{2}{k-2}}}\right)$.*

1.2.2 $\mathbb{F}_2$-polynomials and the CHLT conjecture

The works [9] and [10] highlighted the challenge of proving $L_{1,k}$ bounds for the class of *bounded-degree $\mathbb{F}_2$-polynomials* as being of special interest. Let

$\mathsf{Poly}_{n,d} :=$ the class of all n-variate $\mathbb{F}_2$-polynomials of degree d.

It follows from Theorem 4 that even proving

$$L_{1,2}(\mathsf{Poly}_{n,\mathrm{polylog}(n)}) \le n^{0.49} \tag{1}$$

[1] In fact, they showed that a bound on the weaker quantity $M_{1,k}(f) := \max_{x\in\{-1,1\}^n} |\sum_{|S|=k} \widehat{f}(S)x^S|$ suffices.

would give nontrivial PRGs for $\mathbb{F}_2$-polynomials of polylog(n) degree, improving on [5, 24, 41]. By the classic connection (due to Razborov [32]) between such polynomials and the class $\mathsf{AC}^0[\oplus]$ of constant-depth circuits with parity gates, this would also give nontrivial PRGs, of seed length n^{1-c}, for $\mathsf{AC}^0[\oplus]$. This would be a breakthrough improvement on existing results, which are poor either in terms of seed length [15] or in terms of explicitness [12].

The authors of [10] in fact conjectured the following bound, which is much stronger than Equation (1):

▶ **Conjecture 6** ([10]). *For all $d \geq 1$, it holds that $L_{1,2}(\mathsf{Poly}_{n,d}) = O(d^2)$.*

1.2.3 Extending the CHLT conjecture

Given Conjecture 6, and in light of Theorem 5, it is natural to speculate that an even stronger result than Conjecture 6 might hold. We consider the following natural generalization of the [10] conjecture, extending it from $L_{1,2}(\mathsf{Poly}_{n,d})$ to $L_{1,k}(\mathsf{Poly}_{n,d})$:

▶ **Conjecture 7.** *For all $d, k \geq 1$, it holds that $L_{1,k}(\mathsf{Poly}_{n,d}) = O(d)^k$.*

The work [10] proved that $L_{1,1}(\mathsf{Poly}_{n,d}) \leq 4d$, and already in [9] it was shown that $L_{1,k}(\mathsf{Poly}_{n,d}) \leq (2^{3d} \cdot k)^k$, but to the best of our knowledge no other results towards Conjecture 6 or Conjecture 7 are known.

Given the apparent difficulty of resolving Conjecture 6 and Conjecture 7 in the general forms stated above, it is natural to study $L_{1,2}$ and $L_{1,k}$ bounds for specific subclasses of degree-d $\mathbb{F}_2$-polynomials. This study is the subject of our main structural results, which we describe in the next subsection.

1.3 Our results: Fourier bounds for structured $\mathbb{F}_2$-polynomials

Our main results show that $L_{1,2}$ and $L_{1,k}$ bounds of the flavor of Conjecture 6 and Conjecture 7 indeed hold for several well-studied classes of $\mathbb{F}_2$-polynomials, specifically *symmetric* $\mathbb{F}_2$-polynomials and *read-*Δ $\mathbb{F}_2$-polynomials. We additionally prove a composition theorem that allows us to combine such polynomials (or, more generally, any polynomials that satisfy certain $L_{1,k}$ bounds) in a natural way and obtain $L_{1,k}$ bounds on the resulting combined polynomials.

Before describing our results in detail, we pause to briefly explain why (beyond the fact that they are natural mathematical objects) such "highly structured" polynomials are attractive targets of study given known results. It has been known for more than ten years [2, Lemma 2] that for any degree $d < (1-\epsilon)n$, a *random* $\mathbb{F}_2$-polynomial of degree d (constructed by independently including each monomial of degree at most d with probability $1/2$) is extremely unlikely to have bias larger than $\exp(-n/d)$. It follows that as long as d is not too large, a random degree-d polynomial p is overwhelmingly likely to have $L_{1,k}(p) = o_n(1)$, which is much smaller than d^k. (To verify this, consider the polynomials p_S obtained by XORing p with the parity function $\sum_{i \in S} x_i$. Note that the bias of p_S is the Fourier coefficient of $(-1)^p$ on S. Now apply [2, Lemma 2] to each polynomial p_S, and sum the terms.)

Since the conjectures hold true for random polynomials, it is natural to investigate highly structured polynomials.

1.3.1 Symmetric $\mathbb{F}_2$-polynomials

A *symmetric* $\mathbb{F}_2$-polynomial over $x_1, \ldots, x_n$ is one whose output depends only on the Hamming weight of its input x. Such a polynomial of degree d can be written in the form

$$p(x) := \sum_{k=0}^{d} c_k \sum_{|S|=k, S\subseteq[n]} \prod_{i\in S} x_i,$$

where $c_0, \ldots, c_d \in \{0,1\}$. While symmetric polynomials may seem like simple objects, their study can sometimes lead to unexpected discoveries; for example, a symmetric, low-degree $\mathbb{F}_2$-polynomial provided a counterexample to the "Inverse conjecture for the Gowers norm" [25, 20].

We prove the following upper bound on the L_1 Fourier norm at level k for any symmetric polynomial:

▶ **Theorem 8.** *Let $p(x_1, \ldots, x_n)$ be a symmetric $\mathbb{F}_2$-polynomial of degree d. Then $L_{1,k}(p) \leq \mathbf{Pr}[p = 1] \cdot O(d)^k$ for every k.*

We note that if $d = n$ and p is the AND function, then an easy computation shows that $L_{1,k}(p) = \mathbf{Pr}[p = 1] \cdot \binom{d}{k}$. Moreover, in Appendix A we show that this implies that the upper bounds conjectured in Conjecture 7 are best possible for any constant k. Theorem 8 verifies the [10] conjecture (Conjecture 6), and even the generalized version Conjecture 7, for the class of symmetric polynomials.

Theorem 8 provides a quadratic sharpening of an earlier bound that was implicit in [33] (as well as providing the "correct" dependence on $\mathbf{Pr}[p = 1]$). In [33] Reingold, Steinke and Vadhan showed that any function f computed by an oblivious, read-once, regular branching program of width w has $L_{1,k}(f) \leq (2w^2)^k$. It follows directly from a result of [3] (Lemma 15 below) that any symmetric $\mathbb{F}_2$-polynomial p of degree d can be computed by an oblivious, read-once, regular branching program of width at most $2d$, and hence the [33] result implies that $L_{1,k}(p) \leq 8^k d^{2k}$.

1.3.2 Read-Δ $\mathbb{F}_2$-polynomials

For $\Delta \geq 1$, a *read-Δ* $\mathbb{F}_2$-polynomial is one in which each input variable appears in at most Δ monomials. The case $\Delta = 1$ corresponds to the class of *read-once polynomials*, which are simply sums of monomials over disjoint sets of variables; for example, the polynomial $x_1x_2 + x_3x_4$ is read-once whereas $x_1x_2 + x_1x_4$ is read-twice. Read-once polynomials have been studied from the perspective of pseudorandomness [23, 27, 22, 14] as they capture several difficulties in improving Nisan's generators [28] for width-4 read-once branching programs.

We show that the $L_{1,k}$ Fourier norm of read-Δ polynomials is polynomial in d and Δ:

▶ **Theorem 9.** *Let $p(x_1, \ldots, x_n)$ be a read-Δ polynomial of degree d. Then $L_{1,k}(p) \leq \mathbf{Pr}[p = 1] \cdot O(k)^k \cdot (d\Delta)^{8k}$.*

[22] showed that read-once polynomials satisfy an $L_{1,k}$ bound of $O(d)^k$ for every k, but we are not aware of previous bounds on even the L_1 Fourier norm at level $k = 2$ for read-Δ polynomials, even for $\Delta = 2$.

As any monomial with degree $\Omega(\log n)$ vanishes under a random restriction with high probability, we have the following corollary which applies to polynomials of any degree.

▶ **Corollary 10.** *Let $p(x_1, \ldots, x_n)$ be a read-Δ polynomial. Then $L_{1,k}(p) \leq O(k)^{9k} \cdot (\Delta \log n)^{8k}$.*

1.3.3 A composition theorem

The upper bounds of Theorem 8 and Theorem 9 both include a factor of $\mathbf{Pr}[p = 1]$. (We observe that negating p, i.e. adding 1 to it, does not change its $L_{1,2}$ or $L_{1,k}$ and keeps p symmetric (respectively, read-Δ) if it was originally symmetric (respectively, read-Δ), and hence in the context of those theorems we can assume that this $\mathbf{Pr}[p = 1]$ factor is at most $1/2$.) Level-k bounds that include this factor have appeared in earlier works for other classes of functions [30, 4, 11, 39, 18], and have been used to obtain high-level bounds for other classes of functions [11, 39, 18] and to extend level-k bounds to more general classes of functions [22]. Having these $\mathbf{Pr}[p = 1]$ factors in Theorem 8 and Theorem 9 is important for us in the context of our composition theorem, which we now describe. We begin by defining the notion of a *disjoint composition* of functions:

▶ **Definition 11.** *Let $\mathcal{F}$ be a class of functions from $\{-1,1\}^m$ to $\{-1,1\}$ and let $\mathcal{G}$ be a class of functions from $\{-1,1\}^\ell$ to $\{-1,1\}$. Define the class $\mathcal{H} = \mathcal{F} \circ \mathcal{G}$ of* disjoint compositions *of $\mathcal{F}$ and $\mathcal{G}$ to be the class of all functions from $\{-1,1\}^{m\ell}$ to $\{-1,1\}$ of the form*

$$h(x^1, \ldots, x^m) = f(g_1(x^1), \ldots, g_m(x^m)),$$

where $g_1, \ldots, g_m \in \mathcal{G}$ are defined on m disjoint sets of variables and $f \in \mathcal{F}$.

As an example of this definition, the class of *block-symmetric* polynomials (i.e. polynomials whose variables are divided into blocks and are symmetric within each block but not overall) are a special case of disjoint compositions where $\mathcal{G}$ is taken to be the class of symmetric polynomials. We remark that block-symmetric polynomials are known to correlate better with parities than symmetric polynomials in certain settings [21].

We prove a composition theorem for upper-bounding the L_1 Fourier norm at level k of the disjoint composition of any classes of functions that are closed under restriction and admit a $L_{1,k}$ bound of the form $\mathbf{Pr}[f = 1] \cdot a \cdot b^k$:

▶ **Theorem 12.** *Let $g_1, \ldots, g_m \in \mathcal{G}$ and let $f \in \mathcal{F}$, where $\mathcal{F}$ is closed under restrictions. Suppose that for every $1 \le k \le K$, we have*
1. *$L_{1,k}(f) \le \mathbf{Pr}[f = 1] \cdot a_{\mathsf{out}} \cdot b_{\mathsf{out}}^k$ for every $f \in \mathcal{F}$, and*
2. *$L_{1,k}(g) \le \mathbf{Pr}[g = 1] \cdot a_{\mathsf{in}} \cdot b_{\mathsf{in}}^k$ for every $g \in \mathcal{G}$.*

Then for every ± 1-valued function $h \in \mathcal{H} = \mathcal{F} \circ \mathcal{G}$, we have that

$$L_{1,K}(h) \le \mathbf{Pr}[h = 1] \cdot a_{\mathsf{out}} \cdot (a_{\mathsf{in}} b_{\mathsf{in}} b_{\mathsf{out}})^K.$$

See the full version of this paper for a slightly sharper bound. We remark that Theorem 12 does not assume any $\mathbb{F}_2$-polynomial structure for the functions in $\mathcal{F}$ or $\mathcal{G}$ and thus may be of broader utility.

1.4 Applications of our results

Our structural results imply new pseudorandom generators and correlation bounds.

1.4.1 Pseudorandom generators

Combining our Fourier bounds with the polarizing framework, we obtain new PRGs for read-few polynomials. The following theorem follows from applying Theorem 5 with some $k = \Theta(\log n)$ and the $L_{1,k}$ bound in Corollary 10.

▶ **Theorem 13.** *There is an explicit pseudorandom generator that ϵ-fools read-Δ $\mathbb{F}_2$-polynomials with seed length* $\mathrm{poly}(\Delta, \log n, \log(1/\epsilon))$.

For constant ϵ, this improves on a PRG by Servedio and Tan [34], which has a seed length of $2^{O(\sqrt{\log(\Delta n)})}$. (Note that read-$\Delta$ polynomials are also (Δn)-sparse.) We are not aware of any previous PRG for read-2 polynomials with polylog(n) seed length.

Note that the OR function has L_1 Fourier norm $O(1)$. By expressing a DNF in the Fourier expansion of OR in its terms, it is not hard to see that the same PRG also fools the class of read-Δ DNFs (and read-Δ CNFs similarly) [35].

1.4.2 Correlation bounds

Exhibiting explicit Boolean functions that do not *correlate* with low-degree polynomials is a fundamental challenge in complexity. Perhaps surprisingly, this challenge stands in the way of progress on a striking variety of frontiers in complexity, including circuits, rigidity, and multiparty communication complexity. For a survey of correlation bounds and discussions of these connections we refer the reader to [40, 42, 44].

For polynomials of degree larger than $\log_2 n$, the state-of-the-art remains the lower bound proved by Razborov and Smolensky in the 1980s' [32, 37], showing that for any degree-d polynomial p and an explicit function h (in fact, majority) we have:

$$\mathbf{Pr}[p(x) = h(x)] \leq 1/2 + O(d/\sqrt{n}).$$

Viola [43] recently showed that upper bounds on $L_{1,k}(\mathcal{F})$ imply correlation bounds between $\mathcal{F}$ and an explicit function h_k that is related to majority and is defined as

$$h_k(x) := \mathrm{sgn}\left(\sum_{|S|=k} x^S\right).$$

In particular, proving Conjecture 6 or related conjectures implies new correlation bounds beating Razborov–Smolensky. The formal statement of the connection is given by the following theorem.

▶ **Theorem 14** (Theorem 1 in [43]). *For every $k \in [n]$ and $\mathcal{F} \subseteq \{f\colon \{0,1\}^n \to \{-1,1\}\}$, there is a distribution D_k on $\{0,1\}^n$ such that for any $f \in \mathcal{F}$,*

$$\Pr_{x \sim D_k}[f(x) = h_k(x)] \leq \frac{1}{2} + \frac{e^k}{2\sqrt{\binom{n}{k}}} L_{1,k}(\mathcal{F}).$$

For example, if $k = 2$ and we assume that the answer to Conjecture 6 is positive, then the right-hand side above becomes $1/2 + O(d^2/n)$, which is a quadratic improvement over the bound by Razborov and Smolensky.

Therefore, Theorems 8 and 9 imply correlation bounds between these polynomials and an explicit function that are better than $O(d/\sqrt{n})$ given in [32, 37]. We note that via a connection in [41], existing PRGs for these polynomials already imply strong correlation bounds between these polynomials and the class of NP. Our results apply to more general classes via the composition theorem, where it is not clear if previous techniques applied. For a concrete example, consider the composition of a degree-(n^α) symmetric polynomial with degree-(n^α) read-(n^α) polynomials. Theorem 12 shows that such polynomial has $L_{1,2} \leq n^{O(\alpha)}$. For a sufficiently small $\alpha = \Omega(1)$, we again obtain correlation bounds improving on Razborov–Smolensky.

1.5 Related work

We close this introduction by discussing a recent work of Girish, Tal and Wu [18] on parity decision trees that is related to our results.

Parity decision trees are a generalization of decision trees in which each node queries a parity of some input bits rather than a single input bit. The class of depth-d parity decision trees is a subclass of $\mathbb{F}_2$ degree-d polynomials, as such a parity decision tree can be expressed as a sum of products of sums over $\mathbb{F}_2$, where each product corresponds to a path in the tree (and hence gives rise to $\mathbb{F}_2$-monomials of degree at most d). The Fourier spectrum of parity decision trees was first studied in [4], which obtained a level-1 bound of $O(\sqrt{d})$. This bound was recently extended to higher levels in [18], showing that any depth-d parity decision tree T over n variables has $L_{1,k}(T) \leq d^{k/2} \cdot O(k \log n)^k$.

2 Our techniques

We now briefly explain the approaches used to prove our results. We note that each of these results is obtained using very different ingredients, and hence the results can be read independently of each other.

2.1 Symmetric polynomials (Theorem 8, Section 4)

The starting point of our proof is a result from [3], which says that degree-d symmetric $\mathbb{F}_2$-polynomials only depend on the Hamming weight of their input modulo m for some m (a power of two) which is $\Theta(d)$. Given this, since $p(x)$ takes the same value for all strings x with the same weights $\ell \bmod m$, to analyze $L_{1,k}(p)$ it suffices to analyze $\mathbf{E}[(-1)^{x_1+\cdots+x_k}]$ conditioned on x having Hamming weight exactly $\ell \bmod m$.

We bound this conditional expectation by considering separately two cases depending on whether or not $k \leq n/m^2$. For the case that $k \leq n/m^2$, we use a (slight sharpening of a) result from [6], which gives a bound of $m^{-k}e^{-\Omega(n/m^2)}$. In the other case, that $k \geq n/m^2$, in Lemma 17 we prove a bound of $O(km/n)^k$. This is established via a careful argument that gives a new bound on the Kravchuk polynomial in certain ranges (see the full version of the paper for more details), extending and sharpening similar bounds that were recently established in [13] (the bounds of [13] would not suffice for our purposes).

In each of the above two cases, summing over all the $\binom{n}{k}$ coefficients gives the desired bound of $O(m)^k = O(d)^k$.

2.2 Read-Δ polynomials (Theorem 9, Section 5)

Writing $f := (-1)^p$ for an $\mathbb{F}_2$-polynomial p, we observe that the coefficient $\widehat{f}(S)$ is simply the bias of $p_S(x) := p(x) + \sum_{i \in S} x_i$. Our high-level approach is to decompose the read-few polynomial p_S into many disjoint components, then show that each component has small bias. Since the components are disjoint, the product of these biases gives an upper bound on the bias of p_S.

In more detail, we first partition the variables according to the minimum degree t_i of the monomials containing each variable x_i. Then we start decomposing p_S by collecting all the monomials in p containing x_i to form the polynomial p_i. We observe that the larger t_i is, the more likely p_i is to vanish on a random input, and therefore the closer $p_i + x_i$ is to being unbiased. For most S, we can pick many such p_i's ($i \in S$) from p so that they are disjoint. For the remaining polynomial r, because Δ and d are small, we can further decompose r into

many disjoint polynomials r_i. Finally, our upper bound on $|\widehat{f}(S)|$ will be the magnitude of the product of the biases of the p_i's and r_i's. We note that our decomposition of p uses the structure of S; and so the upper bound on $\widehat{f}(S)$ depends on S (see Lemma 18). Summing over each $|\widehat{f}(S)|$ gives our upper bound.

2.3 Composition theorem (Theorem 12, Section 6)

As a warmup, let us first consider directly computing a degree-1 Fourier coefficient $\widehat{h}(\{(i,j)\})$ of the composition. Since the inner functions g_i depend on disjoint variables, by writing the outer function f in its Fourier expansion, it is not hard to see that

$$\widehat{h}(\{(i,j)\}) = \sum_{S \ni i} \widehat{f}(S) \prod_{\ell \in S \setminus \{i\}} \mathbf{E}[g_\ell] \cdot \widehat{g}_i(\{j\}).$$

When the g_i's are balanced, i.e. $\mathbf{E}[g_i] = 0$, we have $\widehat{f}(\{(i,j)\}) = \widehat{f}(\{i\})\widehat{g}_i(\{j\})$, and it follows that $L_{1,1}(h) \le L_{1,1}(\mathcal{F})L_{1,1}(\mathcal{G})$. To handle the unbalanced case, we apply an idea from [9] that lets us relate $\sum_{S \ni i} \widehat{f}(S) \prod_{\ell \in S \setminus \{i\}} \mathbf{E}[g_\ell]$ to the average of $\widehat{f_R}(\{i\})$, for some suitably chosen random restriction R on f (see Claim 20). As $\mathcal{F}$ is closed under restrictions, we can apply the $L_{1,1}(\mathcal{F})$ bound on f_R, which in turns gives a bound on $\sum_{S \ni i} \widehat{f}(S) \prod_{\ell \in S \setminus \{i\}} \mathbf{E}[g_\ell]$ in terms of $L_{1,1}(\mathcal{F})$ and $\mathbf{E}[g_i]$.

Bounding $L_{1,k}(h)$ for $k \ge 2$ is more complicated, as each $\widehat{h}(S)$ involves $\widehat{f}(J)$ and $\widehat{g}_i(T)$'s, where the sets J and T have different sizes. We provide more details in Section 6.

3 Preliminaries

Notation. For a string $x \in \{0,1\}^n$ we write $|x|$ to denote its Hamming weight $\sum_{i=1}^n x_i$. We use $\mathcal{X}_w$ to denote $\{x : |x| = w\}$, the set of n-bit strings with Hamming weight w, and $\mathcal{X}_{\ell \bmod m} = \bigcup_{w : w \equiv \ell \bmod m} \mathcal{X}_w = \{x : |x| \equiv \ell \bmod m\}$.

We recall that for an n-variable Boolean function f, the *level-k Fourier L_1 norm of f* is

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

We note that a function f and its negation have the same $L_{1,k}$ for $k \ge 1$. Hence we can often assume that $\mathbf{Pr}[f = 1] \le 1/2$, or replace the occurrence of $\mathbf{Pr}[f = 1]$ in a bound by $\min\{\mathbf{Pr}[f = 1], \mathbf{Pr}[f = 0]\}$ for a $\{0,1\}$-valued function f (or by $\min\{\mathbf{Pr}[f = 1], \mathbf{Pr}[f = -1]\}$ for a $\{-1,1\}$-valued function). If f is a $\{-1,1\}$-valued function then $\frac{1-|\mathbf{E}[f]|}{2}$ is equal to $\min\{\mathbf{Pr}[f = 1], \mathbf{Pr}[f = -1]\}$, and we will often write $\frac{1-|\mathbf{E}[f]|}{2}$ for convenience.

Unless otherwise indicated, we will use the letters p, q, r, etc. to denote $\mathbb{F}_2$-polynomials (with inputs in $\{0,1\}^n$ and outputs in $\{0,1\}$) and the letters f, g, h, etc. to denote general Boolean functions (where the inputs may be $\{0,1\}^n$ or $\{-1,1\}^n$ and the outputs may be $\{0,1\}$ or $\{-1,1\}$ depending on convenience). We note that changing from $\{0,1\}$ outputs to $\{-1,1\}$ outputs only changes $L_{1,k}$ by a factor of 2.

We use standard multilinear monomial notation as follows: given a vector $\beta = (\beta_1, \ldots, \beta_n)$ and a subset $T \subseteq [n]$, we write β^T to denote $\prod_{j \in T} \beta_j$.

4 $L_{1,k}$ bounds for symmetric polynomials

In this section we prove Theorem 8, which gives an upper bound on $L_{1,k}(p)$ for any symmetric $\mathbb{F}_2$-polynomial p of degree d, covering the entire range of parameters $1 \le k, d \le n$.

4.1 Proof idea

As the polynomial p is symmetric, its Fourier coefficient $\widehat{p}(S)$ only depends on $|S|$, the size of S. Hence to bound $L_{1,k}$ it suffices to analyze the coefficient $\widehat{p}(\{1,\dots,k\}) = \mathbf{E}_{x\sim\{0,1\}^n}[p(x)(-1)^{x_1+\cdots+x_k}]$.

Our proof uses a result from [3] (Lemma 15 below), which says that degree-d symmetric $\mathbb{F}_2$-polynomials only depend on the Hamming weight of their input modulo m for some $m = O(d)$. Given this, since $p(x)$ takes the same value for strings x with the same weights $\ell \bmod m$, we can in turn bound each $\mathbf{E}[(-1)^{x_1+\cdots+x_k}]$ conditioned on x having Hamming weight exactly $\ell \bmod m$, i.e. $x \in \mathcal{X}_{\ell \bmod m}$. We consider two cases depending on whether or not $k \le n/m^2$. If $k \le n/m^2$, we can apply a (slight sharpening of a) result from [6], which gives a bound of $m^{-k}e^{-\Omega(n/m^2)}$. If $k \ge n/m^2$, in Lemma 17 we prove a bound of $O(km/n)^k$. In each case, summing over all the $\binom{n}{k}$ coefficients gives the desired bound of $O(m)^k = O(d)^k$.

We now give some intuition for Lemma 17, which upper bounds the magnitude of the ratio

$$\mathop{\mathbf{E}}_{x\sim\mathcal{X}_{\ell \bmod m}}[(-1)^{x_1+\cdots+x_k}] = \frac{\sum_{x\in\mathcal{X}_{\ell \bmod m}}(-1)^{x_1+\cdots+x_k}}{|\mathcal{X}_{\ell \bmod m}|} \tag{2}$$

by $O(km/n)^k$. Let us first consider $k = 1$ and $m = \Theta(\sqrt{n})$. As most strings x have Hamming weight within $[n/2 - \Theta(\sqrt{n}), n/2 + \Theta(\sqrt{n})]$, it is natural to think about the weight $|x|$ in the form of $n/2 + m\mathbb{Z} + \ell'$. It is easy to see that the denominator is at least $\Omega(2^n/\sqrt{n})$, so we focus on bounding the numerator. Consider the quantity $\sum_{x\in\mathcal{X}_{n/2+s}}\mathbf{E}[(-1)^{x_1}]$ for some s. As we are summing over all strings of the same Hamming weight, we can instead consider $\sum_{x\in\mathcal{X}_{n/2+s}}\mathbf{E}_{i\sim[n]}[(-1)^{x_i}]$. For any string of weight $n/2+s$, it is easy to see that

$$\mathop{\mathbf{E}}_{i\sim[n]}[(-1)^{x_i}] = (1/2 - s/n) - (1/2 + s/n) = -2s/n. \tag{3}$$

Therefore, in the $k = 1$ case we get that

$$\left|\mathop{\mathbf{E}}_{x\sim\mathcal{X}_{\ell \bmod m}}[(-1)^{x_1+\cdots+x_k}]\right| \le 2\sum_c \binom{n}{n/2 + cm + \ell'}\frac{|cm+\ell'|}{n}.$$

Using the fact that $\binom{n}{n/2+cm+\ell'}$ is exponentially decreasing in $|c|$, in the full version of the paper we show that this is at most $O(2^n/n)$. So the ratio in (2) is at most $O(1/\sqrt{n})$, as desired, when $k = 1$.

However, already for $k = 2$, a direct (but tedious) calculation shows that

$$\mathop{\mathbf{E}}_{i<j}[(-1)^{x_i+x_j}] = \frac{4s^2 - 2ns + n}{n(n-1)}, \tag{4}$$

which no longer decreases in s like in (3). Nevertheless, we observe that this is bounded by $O(1/n + (|s|/n)^2)$, which is sufficient for bounding the ratio by $O(1/n)$. Building on this, for any k we obtain a bound of $2^{O(k)}((k/n)^{k/2} + (|s|/n))^k$ in the full version of the paper, and by a more careful calculation we are able to obtain the desired bound of $O(km/n)^k$ on Equation (2).

4.2 Proof of Theorem 8

We now prove the theorem. We will use the following result from [3], which says that degree-d symmetric $\mathbb{F}_2$-polynomials only depend on their input's Hamming weight modulo $O(d)$.

▶ **Lemma 15** (Theorem 2.4 in [3], $p = 2$)**.** *Let $p\colon \{0,1\}^n \to \{0,1\}$ be a symmetric $\mathbb{F}_2$-polynomial of degree d, where $m/2 \le d < m$ and m is a power of two. Then $p(x)$ only depends on $|x| \bmod m$.*

We will also use two bounds on the biases of parities under the uniform distribution over $\mathcal{X}_{\ell \bmod m}$, one holds for $k \le n/(2d)^2 \le n/m^2$ (Claim 16) and the other for $k \ge n/(2d)^2 \ge n/(4m^2)$ (Lemma 17). Claim 16 is essentially taken from [6]. However, the statement in [6] has a slightly worse bound; so in the full version of the paper we explain the changes required to give the bound of Claim 16. The proof of Lemma 17 involves bounding the magnitude of Kravchuk polynomials. As it is somewhat technical we defer its proof to the full version of the paper.

▷ Claim 16 (Lemma 10 in [6]). For every $1 \le k \le n/m^2$ and every integer ℓ,

$$2^{-n}\left|\sum_{x\in\mathcal{X}_{\ell \bmod m}} (-1)^{x_1+\cdots+x_k}\right| \le m^{-(k+1)}e^{-\Omega(n/m^2)},$$

while for $k = 0$,

$$\left|2^{-n}|\mathcal{X}_{\ell \bmod m}| - 1/m\right| \le m^{-1}e^{-\Omega(n/m^2)}.$$

▶ **Lemma 17.** *For $k \ge n/(4m^2)$, we have*

$$\binom{n}{k} \cdot \max_{\ell}\left|\frac{\sum_{x\in\mathcal{X}_{\ell \bmod m}}(-1)^{x_1+\cdots+x_k}}{|\mathcal{X}_{\ell \bmod m}|}\right| \le O(m)^k.$$

We now use Claim 16 and Lemma 17 to prove Theorem 8.

Proof of Theorem 8. As p is symmetric, all the level-k coefficients are the same, so it suffices to give a bound on $\widehat{p}(\{1,2,\ldots,k\})$. Let $\tilde{p}\colon \{0,\ldots,n\} \to \{0,1\}$ be the function defined by $\tilde{p}(|x|) := p(x_1,\ldots,x_n)$. By Lemma 15, we have $\tilde{p}(\ell) = \tilde{p}(\ell \bmod m)$ for some $d < m \le 2d$ where m is a power of 2. Using the definition of $\widehat{p}(\{1,\ldots,k\})$, we have

$$\begin{aligned}
|\widehat{p}(\{1,\ldots,k\})| &= \left|\mathop{\mathbf{E}}_{x\sim\{0,1\}^n}\left[p(x)(-1)^{x_1+\cdots+x_k}\right]\right| \\
&= \left|\sum_{\ell=0}^{m-1}\tilde{p}(\ell)\frac{|\mathcal{X}_{\ell \bmod m}|}{2^n}\cdot\frac{\sum_{x\in\mathcal{X}_{\ell \bmod m}}(-1)^{x_1+\cdots+x_k}}{|\mathcal{X}_{\ell \bmod m}|}\right| \\
&\le \mathbf{E}[p]\cdot\max_{0\le\ell\le m-1}\left|\frac{\sum_{x\in\mathcal{X}_{\ell \bmod m}}(-1)^{x_1+\cdots+x_k}}{|\mathcal{X}_{\ell \bmod m}|}\right|,
\end{aligned}$$

where we use the shorthand $\mathbf{E}[p] = \mathbf{E}_{x\sim\{0,1\}^n}[p(x)]$ in the last step.

When $k \le n/(2d)^2 \le n/m^2$, by Claim 16 (using the first bound for the numerator and the second $k = 0$ bound for the denominator) we have

$$\max_{0\le\ell\le m-1}\left|\frac{\sum_{x\in\mathcal{X}_{\ell \bmod m}}(-1)^{x_1+\cdots+x_k}}{|\mathcal{X}_{\ell \bmod m}|}\right| \le \frac{m^{-(k+1)}e^{-\Omega(n/m^2)}}{m^{-1}(1-e^{-\Omega(n/m^2)})} \le O(1)\cdot m^{-k}e^{-\Omega(n/m^2)},$$

where the last inequality holds because $1 \le k \le n/m^2$ and hence the $(1 - e^{-\Omega(n/m^2)})$ factor in the denominator of the left-hand side is $\Omega(1)$. Hence, summing over all the $\binom{n}{k}$ level-k coefficients, we get that

$$L_{1,k}(p) \le \mathbf{E}[p] \cdot \binom{n}{k} \cdot O(1) \cdot m^{-k} e^{-\Omega(n/m^2)} \le \mathbf{E}[p] \cdot O(1) \cdot m^k \left(\frac{ne}{km^2}\right)^k e^{-\Omega(n/m^2)} \le \mathbf{E}[p] \cdot O(m)^k,$$

where the last inequality is because for constant c, the function $(x/k)^k e^{-cx}$ is maximized when $x = k/c$, and is $O(1)^k$.

When $k \ge n/(2d)^2 \ge n/(4m^2)$, by Lemma 17 we have

$$L_{1,k}(p) \le \mathbf{E}[p] \cdot \binom{n}{k} \max_{0 \le \ell \le m-1} \left| \frac{\sum_{x \in \mathcal{X}_{\ell \bmod m}} (-1)^{x_1 + \cdots + x_k}}{|\mathcal{X}_{\ell \bmod m}|} \right| \le \mathbf{E}[p] \cdot O(m)^k. \qquad \blacktriangleleft$$

5 $L_{1,k}$ bounds for read-Δ polynomials

In this section we prove our $L_{1,k}$ bounds for read-few polynomials, proving Theorem 9.

5.1 Proof idea

We first observe that for $f = (-1)^p$, the Fourier coefficient $\widehat{f}(S)$ is simply the bias of the $\mathbb{F}_2$-polynomial $p_S(x) := p(x) + \sum_{i \in S} x_i$. Assuming that p_S depends on all n variables, by a simple greedy argument we can collect $n/\mathrm{poly}(\Delta, d)$ polynomials in p_S so that each of them depends on disjoint variables, and it is not hard to show that the product of the biases of these polynomials upper bounds the bias of p_S. From this is easy to see that any read-Δ degree-d polynomial has bias $\exp(2^{-d} n/\mathrm{poly}(\Delta, d))$. However, this quantity is too large to sum over $\binom{n}{k}$ coefficients.

Our next idea (Lemma 18) is to give a more refined decomposition of the polynomial p by inspecting the variables $x_i : i \in S$ more closely. Suppose the variables $x_i : i \in S$ are far apart in their dependency graph (see the definition of G_p below), as must indeed be the case for most of the $\binom{n}{k}$ size-k sets S. Then we can collect all the monomials containing each x_i to form a polynomial p_i, and these p_i's will depend on disjoint variables. Moreover, if every monomial in p_i has high degree (see the definition of $V_t(p)$ below), then $p_i = 0$ with high probability and therefore $p_i + x_i$ is almost unbiased. Therefore, we can first collect these p_i and x_i from p_S; then, for the remaining $m \ge |S| \cdot \mathrm{poly}(\Delta, d)$ monomials in p_S, as before we collect $m/\mathrm{poly}(\Delta, d)$ polynomials r_i so that they depend on disjoint variables, but this time we collect these monomials using the variables in $V_t(p)$, and give an upper bound in terms of the size $|V_t(p)|$. Multiplying the biases of the $p_i + x_i$'s and the bias of r gives our refined upper bound on $\widehat{f}(S)$ in Lemma 18.

5.2 Proof of Theorem 9

We now proceed to the actual proof. We first define some notions that will be used throughout our arguments. For a read-Δ degree-d polynomial p, we define $V_t(p) : t \in [d]$ and G_p as follows.

For every $t \in [d]$, define

$$V_t(p) := \{i \in [n] : \text{the minimum degree of the monomials in } p \text{ containing } x_i \text{ is } t\}.$$

Note that the sets $V_1(p), \ldots, V_d(p)$ form a partition of the input variables p depends on.

Define the undirected graph G_p on $[n]$, where $i, j \in [n]$ are adjacent if x_i and x_j both appear in the same monomial in q. Note that G_p has degree at most Δd. For $S \subseteq [n]$, we use $N_{=d}(S)$ to denote the indices that are at distance exactly i to S in G_p, and use $N_{\le d}(S)$ to denote $\bigcup_{j=0}^{d} N_{=j}(S)$.

We first state our key lemma, which gives a refined bound on each $\widehat{f}(S)$ stronger than the naive bound sketched in the first paragraph of the "Proof Idea" above, and use it to prove Theorem 9. Due to lack of space, we defer its proof to the full version of the paper.

▶ **Lemma 18** (Main lemma for read-Δ polynomials). *Let $p(x_1, \dots, x_n)$ be a read-Δ degree-d polynomial. Let $S \subseteq [n], |S| \ge \ell$ be a subset containing some ℓ indices $i_1, \dots, i_\ell \in S$ whose pairwise distances in G_p are at least 4, and let $t_1, \dots, t_\ell \in [d]$ be such that each $i_j \in V_{t_j}(p)$. Let $f = (-1)^p$. Then*

$$|\widehat{f}(S)| \le O(1)^{|S|} \cdot \Delta^\ell \prod_{j \in [\ell]} \left(2^{-t_j} \exp\left(-\frac{2^{-t_j} |V_{t_j}(p)|}{\ell \cdot (\Delta d)^4} \right) \right).$$

Proof of Theorem 9. Using a reduction given in the proof of [7, Lemma 2.2], it suffices to prove the same bound without the acceptance probability factor, i.e. to prove that for every $1 \le k \le n$,

$$L_{1,k}(p) \le O(k)^k \cdot (\Delta d)^{8k}.$$

As [7] did not provide an explicit statement of the reduction, for completeness we provide a self-contained statement and proof in Lemma 22 in Appendix A.

For every subset $S \subseteq [n]$ of size k, there exists an $\ell \le k$ and $i_1, \dots, i_\ell \in S$ such that their pairwise distances in G_p are at least 4, each $i_j \in V_{t_j}(p)$ for some $t_j \in [d]$, and each of the remaining $k - \ell$ indices in S is within distance at most 3 to some i_j.

Fix any $i_1, \dots, i_\ell$, and let us bound the number of subsets $S \subseteq [n]$ of size k that can contain $i_1, \dots, i_\ell$. Because $|N_{\le 3}(j)| \le (\Delta d)^3 + (\Delta d)^2 + \Delta d + 1 \le 4(\Delta d)^3$ for every $j \in [n]$, the remaining $k - \ell$ indices of S can appear in at most

$$\begin{aligned}
\sum_{j_1 + \dots + j_\ell = k - \ell} \prod_{b \in [\ell]} \binom{4(\Delta d)^3}{j_b} &= \binom{4\ell(\Delta d)^3}{k - \ell} \\
&\le (4(\Delta d)^3)^k \cdot e^{k-\ell} \left(\frac{\ell}{k - \ell} \right)^{k-\ell} \\
&\le (e \Delta d)^{3k}
\end{aligned}$$

different ways, where the equality uses the Vandermonde identity, the first inequality uses $\binom{n}{k} \le (en/k)^k$, and the last one uses $(\frac{\ell}{k-\ell})^{k-\ell} \le (1 + \frac{\ell}{k-\ell})^{k-\ell} \le e^\ell$ and $4e < e^3$. Therefore, by Lemma 18,

$$\begin{aligned}
\sum_{S : |S| = k} |\widehat{f}(S)| &\le \sum_{\ell=1}^{k} \sum_{t \subseteq [d]^\ell} \left[\left(\prod_{j \in [\ell]} |V_{t_j}(p)| \right) \cdot (e\Delta d)^{3k} \cdot O(1)^k \Delta^\ell \prod_{j' \in [\ell]} \left(2^{-t_{j'}} \exp\left(-\frac{2^{-t_{j'}} |V_{t_{j'}}(p)|}{\ell (\Delta d)^4} \right) \right) \right] \\
&\le O(1)^k \cdot (\Delta d)^{3k} \sum_{\ell=1}^{k} \Delta^\ell \sum_{t \subseteq [d]^\ell} \prod_{j \in [\ell]} \left(2^{-t_j} |V_{t_j}(p)| \exp\left(-\frac{2^{-t_j} |V_{t_j}(p)|}{\ell (\Delta d)^4} \right) \right) \\
&\le O(1)^k \cdot (\Delta d)^{3k} \sum_{\ell=1}^{k} \Delta^\ell \cdot d^\ell \cdot (\ell (\Delta d)^4)^\ell \\
&\le O(k)^k \cdot (\Delta d)^{3k} \cdot (\Delta d)^{5k} \\
&= O(k)^k \cdot (\Delta d)^{8k},
\end{aligned}$$

where the third inequality is because the function $x \mapsto x e^{-x/c}$ is maximized when $x = c$. This completes the proof. ◀

6 $L_{1,k}$ bounds for disjoint compositions

In this section we give $L_{1,k}$ bounds on *disjoint compositions* of functions, proving Theorem 12.

6.1 Proof idea

Before proving Theorem 12, we briefly describe the main ideas of the proof. For a subset $J \subseteq [m]$, let $\partial_J f$ denote the *J-th derivative of f*, which can be expressed as

$$\partial_J f(x_1, \ldots, x_m) := \sum_{T \supseteq J} \widehat{f}(T) x^{T \setminus J}.$$

Note that $\widehat{f}(J) = \partial_J f(\vec{0})$.

Let us begin by considering the task of bounding $L_{1,1}(h) = \sum_{(i,j)\in[m]\times[\ell]} |\widehat{h}\{(i,j)\}|$. Let $\beta = (\beta_1, \ldots, \beta_m)$, where $\beta_i := \mathbf{E}[g_i]$. Using the Fourier expansion of f, we have

$$\widehat{h}\{(i,j)\} = \sum_{S \subseteq [m]} \widehat{f}(S)\, \mathbf{E}\left[\prod_{k \in S} g_k(x_k) \cdot x_{i,j}\right].$$

If $S \not\ni i$, then the expectation is zero, because $\prod_{k\in S} g_k(x_k)$ and $x_{i,j}$ are independent and $\mathbf{E}[x_{i,j}] = 0$. So, we have

$$\widehat{h}\{(i,j)\} = \sum_{S \ni i} \widehat{f}(S)\beta^{S\setminus\{i\}} \cdot \widehat{g}_i(\{j\}) = \partial_i f(\beta) \cdot \widehat{g}_i(\{j\}).$$

If the functions g_i are balanced, i.e. $\mathbf{E}[g_i] = 0$ for all i, then we would have $\beta = \vec{0}$, and

$$\widehat{h}\{(i,j)\} = \partial_i f(\vec{0}) \cdot \widehat{g}_i(\{j\}) = \widehat{f}(\{i\})\widehat{g}_i(\{j\}).$$

So in this case we have

$$L_{1,1}(h) = \sum_{i\in[m], j\in[\ell]} \left|\widehat{h}(\{(i,j)\})\right| = \sum_{i\in[m]} \sum_{j\in[\ell]} |\widehat{f}(\{i\})\widehat{g}_i(\{j\})| = \sum_{i\in[m]} |\widehat{f}(\{i\})| \sum_{j\in[\ell]} |\widehat{g}_i(\{j\})|$$

and we can apply our bounds on $L_{1,1}(\mathcal{F})$ and $L_{1,1}(\mathcal{G})$ to $\sum_{i\in[m]} \widehat{f}\{i\}$ and $\sum_{j\in[\ell]} \widehat{g}_i\{j\}$ respectively. Specializing to the case $g_1 = \cdots = g_m$, we have

▷ Claim 19. Suppose $g_1 = g_2 = \cdots = g_m =: g$ and $\mathbf{E}[g] = 0$. Then $L_{1,1}(h) = L_{1,1}(f)L_{1,1}(g)$.

In general the g_i's may not all be the same and may not be balanced, and so it seems unclear how we can apply our $L_{1,1}(\mathcal{F})$ bound on $\sum_{i\in[m]} \partial_i f(\beta_1, \ldots, \beta_m)$ when $\beta \neq \vec{0}$. To deal with this, in Claim 20 below we apply a clever idea introduced in [9] that lets us relate $f(\beta)$ at a nonzero point β to the average of $f_{R_\beta}(\vec{0})$, where f_{R_β} is f with some of its inputs fixed by a random restriction R_β. As $\mathcal{F}$ is closed under restrictions, we have that $f_{R_\beta} \in \mathcal{F}$ and we can apply the $L_{1,1}(\mathcal{F})$ bound on $\sum_i \partial_i f_{R_\beta}(\vec{0})$, which in turn gives a bound on $\sum_{i\in[m]} \partial_i f(\beta_1, \ldots, \beta_m)$.

Bounding $L_{1,K}(h)$ for $K \geq 2$ is more complicated, as now each $\widehat{h}(S)$ involves many $\widehat{f}(J)$ and $\widehat{g}_i(T)$'s, where the sets J and T have different sizes. So one has to group the coefficients carefully.

6.2 Useful notation

For a set $S \subseteq [m] \times [\ell]$, let $S|_f := \{i \in [m] : (i,j) \in S \text{ for some } j \in [\ell]\}$ be the "set of first coordinates" that occur in S, and let $S|_i := \{j \in [\ell] : (i,j) \in S\}$. Note that if $(i,j) \in S$, then $i \in S|_f$ and $j \in S|_i$. Let β denote the vector $(\beta_1, \ldots, \beta_m)$, where $\beta_i := \mathbf{E}[g_i]$ for each $i \in [m]$. For a set $J = \{i_1, \ldots, i_{|J|}\} \subseteq [m]$ and $f = f(y_1, \ldots, y_m)$, we write $\partial_J f$ to denote $\frac{\partial^{|J|} f}{\partial y_{i_1} \cdots \partial y_{i_{|J|}}}$. Since $\partial_J y^T = \mathbb{1}(T \supseteq J) y^{T \setminus J}$, by the multilinearity of f we have that

$$\partial_J f(\beta) = \sum_{T \supseteq J} \widehat{f}(T) \beta^{T \setminus J}. \tag{5}$$

6.3 The random restriction R_β

Given $\beta \in [-1,1]^m$, let R_β be the random restriction which is the randomized function from $\{-1,1\}^m$ to $\{-1,1\}^m$ whose i-th coordinate is (independently) defined by

$$R_\beta(y)_i := \begin{cases} \operatorname{sgn}(\beta_i) & \text{with probability } |\beta_i| \\ y_i & \text{with probability } 1 - |\beta_i|. \end{cases}$$

Note that we have

$$\mathop{\mathbf{E}}_{R_\beta, y}[R_\beta(y)_i] = \mathop{\mathbf{E}}_{R_\beta}[R_\beta(\vec{0})_i] = \beta_i.$$

Define $f_{R_\beta}(y)$ to be the (randomized) function $f(R_\beta(y))$. By the multilinearity of f and independence of the $R_\beta(y)_i$ we have

$$\mathop{\mathbf{E}}_{R_\beta, y}[f_{R_\beta}(y)] = \mathop{\mathbf{E}}_{R_\beta}[f_{R_\beta}(\vec{0})] = f(\beta).$$

The following claim relates the two derivatives $\partial_S f(\beta)$ and $\partial_S f_{R_\beta}(\vec{0}) = \widehat{f_{R_\beta}}(S)$.

▷ Claim 20.

$$\partial_S f(\beta) = \prod_{i \in S} \frac{1}{1 - |\beta_i|} \cdot \mathop{\mathbf{E}}_{R_\beta}[\partial_S f_{R_\beta}(\vec{0})] = \prod_{i \in S} \frac{1}{1 - |\beta_i|} \cdot \mathop{\mathbf{E}}_{R_\beta}[\widehat{f_{R_\beta}}(S)].$$

Proof. Due to lack of space, we defer the proof to the full version of the paper. ◁

We can use Claim 20 to express each coefficient of h in terms of the coefficients of f and g_i.

▶ **Lemma 21.** *For $S \subseteq [m] \times [\ell]$, we have $\widehat{h}(S) = \prod_{i \in S|_f} \widehat{g}_i(S|_i) \cdot \prod_{i \in S|_f} \frac{1}{1-|\beta_i|} \cdot \mathbf{E}_{R_\beta}[\widehat{f_{R_\beta}}(S|_f)]$.*

Proof. Due to lack of space, we defer the proof to the full version of the paper. ◀

6.4 Proof of Theorem 12

By Lemma 21, $L_{1,K}(h)$ is equal to

$$\sum_{S \subseteq [m] \times [\ell] : |S| = K} |\widehat{h}(S)| = \sum_{S \subseteq [m] \times [\ell] : |S| = K} \left| \prod_{i \in S|_f} \widehat{g}_i(S|_i) \cdot \prod_{i \in S|_f} \frac{1}{1 - |\beta_i|} \cdot \mathop{\mathbf{E}}_{R_\beta}[\widehat{f_{R_\beta}}(S|_f)] \right|.$$

We enumerate all the subsets $S \subseteq [m] \times [\ell]$ of size K in the following order: For every $|J| = k \in [K]$ out of the m blocks of ℓ coordinates, we enumerate all possible combinations

of the (disjoint) nonempty subsets $\{S_i : i \in J\}$ in those k blocks whose sizes sum to K. Rewriting the summation above in this order, we obtain

$$\sum_{S\subseteq[m]\times[\ell]:|S|=K}|\widehat{h}(S)| = \sum_{k=1}^{K}\sum_{\substack{J\subseteq[m]\\|J|=k}}\sum_{\substack{w\subseteq[\ell]^J\\\sum_{i\in J}w_i=K}}\sum_{\substack{\{S_i\}_{i\in J}\subseteq[\ell]^J:\\\forall i\in J:|S_i|=w_i}}\left|\prod_{i\in J}\widehat{g}_i(S_i)\prod_{i\in J}\frac{1}{1-|\beta_i|}\mathop{\mathbf{E}}_{R_\beta}\left[\widehat{f_{R_\beta}}(J)\right]\right|$$
$$\leq \sum_{k=1}^{K}\sum_{\substack{J\subseteq[m]\\|J|=k}}\sum_{\substack{w\subseteq[\ell]^J\\\sum_{i\in J}w_i=K}}\sum_{\substack{\{S_i\}_{i\in J}\subseteq[\ell]^J:\\\forall i\in J:|S_i|=w_i}}\prod_{i\in J}|\widehat{g}_i(S_i)|\prod_{i\in J}\frac{1}{1-|\beta_i|}\left|\mathop{\mathbf{E}}_{R_\beta}\left[\widehat{f_{R_\beta}}(J)\right]\right|. \tag{6}$$

Since $L_{1,w_i}(g_i) \leq \frac{1-|\beta_i|}{2}\cdot a_{\mathsf{in}}\cdot b_{\mathsf{in}}^{w_i}$, for every $\{w_i\}_{i\in J}$ such that $\sum_{i\in J}w_i = K$, we have

$$\sum_{\substack{\{S_i\}_{i\in J}\subseteq[\ell]^J:\\\forall i\in J:|S_i|=w_i}}\prod_{i\in J}|\widehat{g}_i(S_i)| = \prod_{i\in J}L_{1,w_i}(g_i) \leq \prod_{i\in J}\Big(\frac{1-|\beta_i|}{2}a_{\mathsf{in}}b_{\mathsf{in}}^{w_i}\Big) = b_{\mathsf{in}}^{K}a_{\mathsf{in}}^{|J|}\prod_{i\in J}\frac{1-|\beta_i|}{2}.$$

Plugging the above into (6), we get that

$$\sum_{S\subseteq[m]\times[\ell]:|S|=K}|\widehat{h}(S)| \leq b_{\mathsf{in}}^{K}\sum_{k=1}^{K}a_{\mathsf{in}}^{k}\sum_{\substack{J\subseteq[m]\\|J|=k}}\sum_{\substack{w\subseteq[\ell]^J\\\sum_{i\in J}w_i=K}}\prod_{i\in J}\left(\frac{1-|\beta_i|}{2}\cdot\frac{1}{1-|\beta_i|}\cdot\left|\mathop{\mathbf{E}}_{R_\beta}\left[\widehat{f_{R_\beta}}(J)\right]\right|\right)$$
$$= b_{\mathsf{in}}^{K}\sum_{k=1}^{K}\left(\frac{a_{\mathsf{in}}}{2}\right)^{k}\sum_{\substack{J\subseteq[m]\\|J|=k}}\left|\mathop{\mathbf{E}}_{R_\beta}\left[\widehat{f_{R_\beta}}(J)\right]\right|\sum_{\substack{w\subseteq[\ell]^J\\\sum_{i\in J}w_i=K}}1$$
$$\leq b_{\mathsf{in}}^{K}\sum_{k=1}^{K}\left(\frac{a_{\mathsf{in}}}{2}\right)^{k}\binom{K-1}{k-1}\sum_{\substack{J\subseteq[m]\\|J|=k}}\left|\mathop{\mathbf{E}}_{R_\beta}\left[\widehat{f_{R_\beta}}(J)\right]\right|, \tag{7}$$

where the last inequality is because for every subset $J\subseteq[m]$, the set $\{w\subseteq[\ell]^J : \sum_{i\in J}w_i = K\}$ has size at most $\binom{K-1}{|J|-1}$. We now bound $|\mathbf{E}_{R_\beta}[\widehat{f_{R_\beta}}(J)]|$. Since for every restriction R_β, we have $f_{R_\beta}\in\mathcal{F}$ (by the assumption that $\mathcal{F}$ is closed under restrictions), it follows that

$$L_{1,k}(f_{R_\beta}) \leq \frac{1-|\mathbf{E}_y[f_{R_\beta}(y)]|}{2}a_{\mathsf{out}}b_{\mathsf{out}}^{k} \leq \frac{1-\mathbf{E}_y[f_{R_\beta}(y)]}{2}a_{\mathsf{out}}b_{\mathsf{out}}^{k}.$$

So

$$\sum_{J\subseteq[m],|J|=k}\left|\mathop{\mathbf{E}}_{R_\beta}\left[\widehat{f_{R_\beta}}(J)\right]\right| \leq \mathop{\mathbf{E}}_{R_\beta}[L_{1,k}(f_{R_\beta})]$$
$$\leq \frac{1-\mathbf{E}_{R_\beta,y}[f_{R_\beta}(y)]}{2}a_{\mathsf{out}}b_{\mathsf{out}}^{k}$$
$$= \frac{1-\mathbf{E}[h]}{2}a_{\mathsf{out}}b_{\mathsf{out}}^{k}.$$

Continuing from (7), we get

$$\begin{aligned}\sum_{S\subseteq[m]\times[\ell]:|S|=K}|\widehat{h}(S)| &\le \frac{1-\mathbf{E}[h]}{2}\cdot b_{\mathsf{in}}^K\cdot\sum_{k=1}^{K}\left(\frac{a_{\mathsf{in}}}{2}\right)^k\cdot\binom{K-1}{k-1}\cdot a_{\mathsf{out}}b_{\mathsf{out}}^k\\ &= \frac{1-\mathbf{E}[h]}{2}\cdot a_{\mathsf{out}}\cdot b_{\mathsf{in}}^K\cdot\frac{a_{\mathsf{in}}b_{\mathsf{out}}}{2}\left(1+\frac{a_{\mathsf{in}}b_{\mathsf{out}}}{2}\right)^{K-1}\\ &\le \frac{1-\mathbf{E}[h]}{2}\cdot a_{\mathsf{out}}\cdot(a_{\mathsf{in}}b_{\mathsf{in}}b_{\mathsf{out}})^K.\end{aligned}$$

where the last equality used the binomial theorem. Applying the same argument to $-h$ lets us replace $\frac{1-\mathbf{E}[h]}{2}$ with $\frac{1-|\mathbf{E}[h]|}{2}$, concluding the proof of Theorem 12. ◀

References

1. Nikhil Bansal and Makrand Sinha. K-forrelation optimally separates quantum and classical query complexity. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, New York, NY, USA, 2021. doi:10.1145/3406325.3451040.
2. Ido Ben-Eliezer, Rani Hod, and Shachar Lovett. Random low-degree polynomials are hard to approximate. *Comput. Complexity*, 21(1):63–81, 2012. doi:10.1007/s00037-011-0020-6.
3. Nayantara Bhatnagar, Parikshit Gopalan, and Richard J. Lipton. Symmetric polynomials over $\mathbb{Z}_m$ and simultaneous communication protocols. *J. Comput. System Sci.*, 72(2):252–285, 2006. doi:10.1016/j.jcss.2005.06.007.
4. Eric Blais, Li-Yang Tan, and Andrew Wan. An inequality for the fourier spectrum of parity decision trees, 2015. arXiv:1506.01055.
5. Andrej Bogdanov and Emanuele Viola. Pseudorandom bits for polynomials. *SIAM J. Comput.*, 39(6):2464–2486, 2010. doi:10.1137/070712109.
6. Ravi Boppana, Johan Håstad, Chin Ho Lee, and Emanuele Viola. Bounded independence versus symmetric tests. *ACM Trans. Comput. Theory*, 11(4):Art. 21, 27, 2019. doi:10.1145/3337783.
7. Sourav Chakraborty, Nikhil S. Mande, Rajat Mittal, Tulasimohan Molli, Manaswi Paraashar, and Swagato Sanyal. Tight chang's-lemma-type bounds for boolean functions. *CoRR*, abs/2012.02335, 2020. arXiv:2012.02335.
8. Eshan Chattopadhyay, Jason Gaitonde, Chin Ho Lee, Shachar Lovett, and Abhishek Shetty. Fractional Pseudorandom Generators from Any Fourier Level, 2020. arXiv:2008.01316.
9. Eshan Chattopadhyay, Pooya Hatami, Kaave Hosseini, and Shachar Lovett. Pseudorandom generators from polarizing random walks. *Theory Comput.*, 15:Paper No. 10, 26, 2019. doi:10.4086/toc.2019.v015a010.
10. Eshan Chattopadhyay, Pooya Hatami, Shachar Lovett, and Avishay Tal. Pseudorandom generators from the second Fourier level and applications to AC0 with parity gates. In *10th Innovations in Theoretical Computer Science*, volume 124 of *LIPIcs*, 2019. doi:10.4230/LIPIcs.ITCS.2019.22.
11. Eshan Chattopadhyay, Pooya Hatami, Omer Reingold, and Avishay Tal. Improved pseudorandomness for unordered branching programs through local monotonicity. In *STOC'18—Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 363–375. ACM, New York, 2018. doi:10.1145/3188745.3188800.
12. Lijie Chen, Xin Lyu, and R. Ryan Williams. Almost-everywhere circuit lower bounds from non-trivial derandomization. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1–12, 2020. doi:10.1109/FOCS46700.2020.00009.
13. Gil Cohen, Noam Peri, and Amnon Ta-Shma. Expander random walks: A fourier-analytic approach. *Electron. Colloquium Comput. Complex.*, 27:163, 2020. URL: https://eccc.weizmann.ac.il/report/2020/163.

14 Dean Doron, Pooya Hatami, and William M. Hoza. Log-seed pseudorandom generators via iterated restrictions. In *35th Computational Complexity Conference*, volume 169 of *LIPIcs*, 2020. doi:10.4230/LIPIcs.CCC.2020.6.

15 Bill Fefferman, Ronen Shaltiel, Christopher Umans, and Emanuele Viola. On beating the hybrid argument. *Theory Comput.*, 9:809–843, 2013. doi:10.4086/toc.2013.v009a026.

16 Uma Girish, Ran Raz, and Avishay Tal. Quantum Versus Randomized Communication Complexity, with Efficient Players. In James R. Lee, editor, *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*, volume 185 of *LIPIcs*, 2021. doi:10.4230/LIPIcs.ITCS.2021.54.

17 Uma Girish, Ran Raz, and Wei Zhan. Lower bounds for XOR of forrelations, 2020. arXiv:2007.03631.

18 Uma Girish, Avishay Tal, and Kewen Wu. Fourier Growth of Parity Decision Trees. In Valentine Kabanets, editor, *36th Computational Complexity Conference (CCC 2021)*, volume 200 of *LIPIcs*, 2021. doi:10.4230/LIPIcs.CCC.2021.39.

19 Parikshit Gopalan, Rocco A. Servedio, Avishay Tal, and Avi Wigderson. Degree and sensitivity: tails of two distributions. *Electron. Colloquium Comput. Complex.*, 23:69, 2016. URL: http://eccc.hpi-web.de/report/2016/069.

20 Ben Green and Terence Tao. The distribution of polynomials over finite fields, with applications to the Gowers norms. *Contrib. Discrete Math.*, 4(2):1–36, 2009.

21 Frederic Green, Daniel Kreymer, and Emanuele Viola. Block-symmetric polynomials correlate with parity better than symmetric. *Comput. Complexity*, 26(2):323–364, 2017. doi:10.1007/s00037-017-0153-3.

22 Chin Ho Lee. Fourier bounds and pseudorandom generators for product tests. In *34th Computational Complexity Conference*, volume 137 of *LIPIcs*, 2019. doi:10.4230/LIPIcs.CCC.2019.7.

23 Chin Ho Lee and Emanuele Viola. More on bounded independence plus noise: pseudorandom generators for read-once polynomials. *Theory Comput.*, 16:Paper No. 7, 50, 2020. doi:10.4086/toc.2020.v016a007.

24 Shachar Lovett. Unconditional pseudorandom generators for low-degree polynomials. *Theory Comput.*, 5:69–82, 2009. doi:10.4086/toc.2009.v005a003.

25 Shachar Lovett, Roy Meshulam, and Alex Samorodnitsky. Inverse conjecture for the Gowers norm is false. *Theory Comput.*, 7:131–145, 2011. doi:10.4086/toc.2011.v007a009.

26 Yishay Mansour. An $O(n^{\log\log n})$ learning algorithm for DNF under the uniform distribution. *J. Comput. System Sci.*, 50(3, part 3):543–550, 1995. Fifth Annual Workshop on Computational Learning Theory (COLT) (Pittsburgh, PA, 1992). doi:10.1006/jcss.1995.1043.

27 Raghu Meka, Omer Reingold, and Avishay Tal. Pseudorandom generators for width-3 branching programs. In *STOC'19—Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 626–637. ACM, New York, 2019. doi:10.1145/3313276.3316319.

28 Noam Nisan. Pseudorandom generators for space-bounded computation. *Combinatorica*, 12(4):449–461, 1992. doi:10.1007/BF01305237.

29 Ryan O'Donnell. *Analysis of Boolean Functions*. Cambridge University Press, 2014. doi:10.1017/CBO9781139814782.

30 Ryan O'Donnell and Rocco A. Servedio. Learning monotone decision trees in polynomial time. *SIAM J. Comput.*, 37(3):827–844, 2007. doi:10.1137/060669309.

31 Ran Raz and Avishay Tal. Oracle separation of BQP and PH. In *STOC'19—Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 13–23. ACM, New York, 2019. doi:10.1145/3313276.3316315.

32 Alexander A. Razborov. Lower bounds on the dimension of schemes of bounded depth in a complete basis containing the logical addition function. *Mat. Zametki*, 41(4):598–607, 623, 1987.

33 Omer Reingold, Thomas Steinke, and Salil P. Vadhan. Pseudorandomness for regular branching programs via fourier analysis. In *RANDOM 2013*, volume 8096 of *Lecture Notes in Computer Science*, pages 655–670, 2013.

34 Rocco A. Servedio and Li-Yang Tan. Improved pseudorandom generators from pseudorandom multi-switching lemmas. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, volume 145 of *LIPIcs*, 2019. doi:10.4230/LIPIcs.APPROX-RANDOM.2019.45.

35 Rocco A. Servedio and Li-Yang Tan. Pseudorandomness for read-k DNF formulas. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 621–638. SIAM, Philadelphia, PA, 2019. doi:10.1137/1.9781611975482.39.

36 Alexander A. Sherstov, Andrey A. Storozhenko, and Pei Wu. An Optimal Separation of Randomized and Quantum Query Complexity, 2020. arXiv:2008.10223.

37 Roman Smolensky. Algebraic methods in the theory of lower bounds for boolean circuit complexity. In *Proceedings of the Nineteenth Annual ACM Symposium on Theory of Computing*, STOC '87, page 77–82, New York, NY, USA, 1987. Association for Computing Machinery. doi:10.1145/28395.28404.

38 Avishay Tal. Tight bounds on the Fourier spectrum of AC^0. In *32nd Computational Complexity Conference*, volume 79 of *LIPIcs*, 2017. doi:10.4230/LIPIcs.CCC.2017.15.

39 Avishay Tal. Towards optimal separations between quantum and randomized query complexities. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 228–239, 2020. doi:10.1109/FOCS46700.2020.00030.

40 Emanuele Viola. On the power of small-depth computation. *Foundations and Trends in Theoretical Computer Science*, 5(1):1–72, 2009.

41 Emanuele Viola. The sum of d small-bias generators fools polynomials of degree d. *Comput. Complexity*, 18(2):209–217, 2009. doi:10.1007/s00037-009-0273-5.

42 Emanuele Viola. Challenges in computational lower bounds. *SIGACT News, Open Problems Column*, 48(1), 2017.

43 Emanuele Viola. Fourier conjectures, correlation bounds, and majority. *Electron. Colloquium Comput. Complex.*, 27:175, 2020. URL: https://eccc.weizmann.ac.il/report/2020/175.

44 Emanuele Viola. New lower bounds for probabilistic degree and AC0 with parity gates. *Theory of Computing*, 2021. Available at http://www.ccs.neu.edu/home/viola/.

A Reduction to bound without acceptance probability

In this section, we show that given any $L_{1,k}$ Fourier norm bound on a class of functions that is closed under XOR on disjoint variables, such a bound can be automatically "upgraded" to a refined bound that depends on the acceptance probability:

▶ **Lemma 22.** *Let $\mathcal{F}$ be a class of $\{-1,1\}$-valued functions such that for every $f \in \mathcal{F}$, the XOR of disjoint copies of f (over disjoint sets of variables) also belongs to $\mathcal{F}$. If $L_{1,k}(\mathcal{F}) \le b^k$, then for every $f \in \mathcal{F}$ it holds that $L_{1,k}(f) \le 2e \cdot \frac{1-|\mathbf{E}[f]|}{2} \cdot b^k$.*

Proof. Suppose not, and let $f \in \mathcal{F}$ be such that $L_{1,k}(f) > 2e \cdot \frac{1-|\mathbf{E}[f]|}{2} \cdot b^k$. We first observe that since $L_{1,k}(\mathcal{F}) \le b^k$, it must be the case that $1 - |\mathbf{E}[f]| \le 1/e$. Let $\alpha := \frac{1-|\mathbf{E}[f]|}{2} \in [0, \frac{1}{2e}]$ so that $|\mathbf{E}[f]| = 1-2\alpha \ge 1-1/e$. Let $f^{\oplus t}$ be the XOR of t disjoint copies of f on tn variables, where the integer t is to be determined below. By our assumption, we have $f^{\oplus t} \in \mathcal{F}$ and thus

$$\begin{aligned} L_{1,k}(f^{\oplus t}) &\ge \binom{t}{1} \cdot L_{1,0}(f)^{t-1} \cdot L_{1,k}(f) && \text{(by disjointness)} \\ &= t \cdot (1-2\alpha)^{t-1} \cdot L_{1,k}(f) && (L_{1,0}(f) = \mathbf{E}[f]) \\ &> t \cdot (1-2\alpha)^{t-1} \cdot 2e \cdot \alpha \cdot b^k =: \Lambda(t). \end{aligned}$$

We note that if $\alpha = 0$ then $|\mathbf{E}[f]| = 1$, so all the Fourier weight of f is on the constant coefficient, and hence the claimed inequality holds trivially. So we subsequently assume that $0 < \alpha \leq \frac{1}{2e}$. Let $t^* := \frac{1}{-\ln(1-2\alpha)} > 0$. It is easy to verify that $\Lambda(t)$ is increasing when $t \leq t^*$, and is decreasing when $t \geq t^*$.

We choose $t = \lceil t^* \rceil$. Since $\alpha \leq \frac{1}{2e} < \frac{e-1}{2e} \approx 0.3161$, we have $t^* > 1$ and thus

$$\begin{aligned} L_{1,k}(f^{\oplus t}) > \Lambda(\lceil t^* \rceil) \geq \Lambda(t^* + 1) &= \left(\frac{1}{-\ln(1-2\alpha)} + 1 \right) \cdot (1-2\alpha)^{\frac{1}{-\ln(1-2\alpha)}} \cdot 2e \cdot \alpha \cdot b^k \\ &= \left(\frac{2\alpha}{-\ln(1-2\alpha)} + 2\alpha \right) \cdot b^k \geq b^k, \end{aligned}$$

where the last inequality holds for every $\alpha \in (0, \frac{e-1}{2e}]$ and can be checked via elementary calculations. This contradicts $L_{1,k}(\mathcal{F}) \leq b^k$, and the lemma is proved. ◀