Discrete harmonic analysis and its applications to testing, learning, and complexity

Thesis by

Joseph Slote

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy in Computing and Mathematical Sciences

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2025
Defended May 29

Joseph Slote
ORCID: 0000-0002-6363-7821

Acknowledgements

My PhD experience has felt like a sort of arrival—vocationally and scientifically, but also socially. It has been an unexpected and deeply cherished gift to meet and work with the friends and colleagues in theoretical computer science that I have. I am so grateful to my advisor, Chris Umans, for admitting me into this world, for encouraging my blue-sky questioning, for supporting my travel, and for modeling for me such a clear-eyed approach to phrasing and attacking computer science questions. My interactions with the quantum community at Caltech have been crucial, especially with Thomas Vidick and his postdocs Atul Singh Arora, Saeed Mehreban, and Ulysse Chaubad. I am also very grateful to Urmila Mehedev and her quantum seminars, which gave me an opportunity to dive deeply into literature that ended up being a large focus of my work. I am also grateful for the constellation of friends and colleagues that I have been blessed with in the larger TCS universe: many thanks especially to Henry Yuen for the opportunity to visit Columbia multiple times, and for the friendships that formed there and at Berkeley. My life is greatly enriched by my friends Shivam Nadimpalli, Ansh Nagda, Sam Gunn, Adam Bene Watts, Fermi Ma, and so many others. I did not expect to develop enough inside jokes as to form a crossword puzzle, but here we are. Thanks Sam.

The other major vein of my PhD has been in harmonic analysis, and I cannot overstate my gratitude for that community's welcome and mentorship. When I emailed Paata Ivanisvili about attending an AIM workshop on the analysis of Boolean functions and its relation to quantum computation, little did I know that the connections I formed there would lead to the main body of work in my PhD as well as deep, yearslong friendships. I have learned so much from my collaborators Alexander Volberg, Haonan Zhang, Ohad Klein, and Lars Becker, and look forward to making many more discoveries together in the years to come. And I am especially grateful to Sasha, who embodies the spirit of mathematics with such inspiring force and who has shown me so much about how to think. And also of course to Olga Volberg—I will cherish my memories of the many rich days of math, borscht, and walks around Berkeley with the Volbergs. Finally, I am deeply grateful to Irina Holmes Fay, who originally connected me with Sasha and who has remained a colleague, mentor, and friend ever since. Without Irina's support and encouragement my PhD would have surely looked very different.

And thanks to my longterm friends Jake and Allie, whom I met many years before while we were all greenhorn masters' students. It has been deeply meaningful to stay connected as we've all found our ways through our 20s, and such a joy to discuss science and meaning when we have found ourselves in the same city.

And finally, I thank my dear family, Audrey, Ben, and Susan, who have patiently listened to oh so many breathless and often dubious explanations of theoretical computer science and math, and who have been constant bulwarks of support and care during the difficult periods. Your ability to deeply appreciate people and ideas of all stripes is a rare thing, and I hope to live and work by that code as best I can in the next chapter.

Abstract

This thesis consists of two parts. In Part I we present a new class of norm discretization inequalities suited for low-degree polynomials in many dimensions, with applications to discrete harmonic analysis and to quantum and classical learning theory.

Discretization inequalities (of Bernstein type) control the supremum norm of polynomials f by their supremum norms over certain finite subsets T of the domain. Unlike earlier multivariate Bernstein-type discretization inequalities we establish dimension-free comparisons for simple and generic T, such as product sets $T = S_1 \times \cdots \times S_n$ for S_j's consisting of well-spread points in $\mathbf{R}$ or $\mathbf{C}$, in exchange for a constant that grows with $\deg(f)$.

Our results also introduce the notion of *individual degree*—the maximum degree of f in any one variable—as a fundamental parameter for discretization inequalities: we show for the first time that dimension-free discretizations of the uniform norm are possible for T with cardinality independent of $\deg(f)$, provided f has bounded individual degree.

Our work offers a new, high-dimensional perspective on discretization inequalities and yields several new results in analysis on the hypergrid (*i.e.,* products of cyclic groups), including Bohnenblust–Hille-type inequalities, dimension-free supremum norm bounds on level-k Fourier projections, and junta theorems. These estimates in turn provide the key analytic tools for extending recent breakthroughs in learning low-degree functions to the hypergrid and to its quantum analogue, local observables on K-level qudit systems.

In Part II we apply ideas from analysis of Boolean functions to study other aspects of (quantum) computation: circuit complexity and property testing.

First, we introduce and study a deceptively simple model of constant-depth quantum circuits and begin the project of proving bounds on its capabilities, ultimately drawing on connections to nonlocal games and notions of approximate degree.

Second, we introduce a new access model for property testing, *quantum data*, which allows for ultrafast testing algorithms where classical data provably yields no fast testers—such as for monotonicity, symmetry, and triangle-freeness.

Published Content and Contributions

Part I *In order of announcement*

[SVZ24a] Joseph Slote, Alexander Volberg, and Haonan Zhang. "Bohnenblust-Hille inequality for cyclic groups". In: *Adv. Math.* 452 (2024), Paper No. 109824, 35. DOI: 10.1016/j.aim.2024.109824.

▶ Equal contribution.

[SVZ25] Joseph Slote, Alexander Volberg, and Haonan Zhang. "A dimension-free Remez-type inequality on the polytorus". In: *Discrete Analysis (to appear)* (2025). arXiv: 2305.10828 [math.CA].

▶ J.S. proved the main theorem, building on tools from [SVZ24a].

[KSVZ24] Ohad Klein, Joseph Slote, Alexander Volberg, and Haonan Zhang. "Quantum and Classical Low-Degree Learning via a Dimension-Free Remez Inequality". In: *15th Innovations in Theoretical Computer Science Conference, ITCS 2024*. Ed. by Venkatesan Guruswami. Vol. 287. LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2024, 69:1–69:22. DOI: 10.4230/LIPICS.ITCS.2024.69.

– *Also appeared at TQC 2024.*

▶ Equal contribution.

[SVZ24b] Joseph Slote, Alexander Volberg, and Haonan Zhang. *Noncommutative Bohnenblust–Hille Inequality for qudit systems*. 2024. arXiv: 2406.08509 [math.FA].

▶ Equal contribution.

[Bec+25] Lars Becker, Ohad Klein, Joseph Slote, Alexander Volberg, and Haonan Zhang. "Dimension-free discretizations of the uniform norm by small product sets". In: *Invent. Math.* 239.2 (2025), pp. 469–503. DOI: 10.1007/s00222-024-01306-9.

▶ Equal contribution on all matters except (*i*) J.S. did not participate in the proof of the "subsampling theorem" [Bec+25, Theorem 3], appearing here as Theorem 17; and (*ii*) J.S. independently proved the L^p comparison, [Bec+25, Theorem 5], appearing here as Theorem 7.

Part II

[Slo24] Joseph Slote. “Parity vs. AC^0 with Simple Quantum Preprocessing”. In: *15th Innovations in Theoretical Computer Science Conference, ITCS 2024*. Ed. by Venkatesan Guruswami. Vol. 287. LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2024, 92:1–92:21. DOI: 10.4230/LIPICS.ITCS.2024.92.

– *Also appeared at TQC 2024.*

[CNS25] Matthias Caro, Preksha Naik, and Joseph Slote. “Testing classical properties from quantum data”. In: *TQC 2025 (to appear)* (2025). arXiv: 2411.12730 [quant-ph].

► J.S. was main contributor to “Fourier sampling does not suffice” and “Hardness of 3-fold intersection detection” theorems. J.S. and P.N. made equal contribution to analysis of monotone ensembles.

Table of Contents

List of Illustrations

List of Tables

List of Symbols

$\mathbf{Z}_K$	cyclic group of order K
$\mathbf{R}$	real numbers
$\mathbf{C}$	complex numbers
$\mathbf{T}^n$	unit n-torus (or *polytorus*) $\{z \in \mathbf{C} : \lvert z\rvert = 1\}^n$
$\mathbf{D}^n$	unit n-disk (or *polydisk*) $\{z \in \mathbf{C} : \lvert z\rvert \leq 1\}^n$
Ω_K	K^{th} roots of unity, multiplicative representation of $\mathbf{Z}_K$
$\lVert \cdot \rVert_X$	supremum norm over domain X
$\mathfrak{D}$	maximum support pseudoprojection, see Definition 1

Conventions

Computer Science

- $[n] := \{1, 2, \ldots, n\}$
- $x \sim X$ for X a finite set means denotes a sample from the uniform probability measure over X.
- The "soft-$\mathcal{O}$" notation $g(n) \leq \widetilde{\mathcal{O}}(f(n))$ means there exists a $k \in \mathbf{N}$ for which
$$g(n) \leq \mathcal{O}\Big(f(n) \cdot \log^k n\Big).$$
- $\mathcal{O}$ attaches first in Landau notation: $\mathcal{O}(x)^d := \Big(\mathcal{O}(x)\Big)^d$

Analysis

- *Integral norms.* For functions $f : X \to \mathbf{C}$, X finite, the notation $\|f\|_p$ will denote L^p norm of f w.r.t. the uniform measure. The notation $\|\widehat{f}\|_p$ denotes the ℓ^p norm of the Fourier transform of f (with respect to counting measure).
- *Vinogradov notation.* For two quantities X and Y, the notation $X \lesssim Y$ means there exists a universal constant $C > 0$ such that $X \leq C \cdot Y$. The notation $X \lesssim_\ell Y$ allows the implicit constant $C = C(\ell)$ to depend on ℓ only. We do *not* allow an additive constant (*e.g.,* the D in $X \leq C\dot{Y} + D$) in Vinogradov notation, which distinguishes it from the Landau notation $X \leq \mathcal{O}(Y)$.
- Quantities X and Y are *comparable*, $X \simeq Y$, if that $X \lesssim Y$ and $Y \lesssim X$.

Part I

Dimension-free discretization inequalities

with applications to learning theory

Published as

In order of announcement:

[SVZ24a] Joseph Slote, Alexander Volberg, and Haonan Zhang. "Bohnenblust-Hille inequality for cyclic groups". In: *Adv. Math.* 452 (2024), Paper No. 109824, 35. DOI: 10.1016/j.aim.2024.109824.

[SVZ25] Joseph Slote, Alexander Volberg, and Haonan Zhang. "A dimension-free Remez-type inequality on the polytorus". In: *Discrete Analysis (to appear)* (2025). arXiv: 2305.10828 [math.CA].

[KSVZ24] Ohad Klein, Joseph Slote, Alexander Volberg, and Haonan Zhang. "Quantum and Classical Low-Degree Learning via a Dimension-Free Remez Inequality". In: *15th Innovations in Theoretical Computer Science Conference, ITCS 2024, January 30 to February 2, 2024, Berkeley, CA, USA*. Ed. by Venkatesan Guruswami. Vol. 287. LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2024, 69:1–69:22. DOI: 10.4230/LIPICS.ITCS.2024.69.

[SVZ24b] Joseph Slote, Alexander Volberg, and Haonan Zhang. *Noncommutative Bohnenblust–Hille Inequality for qudit systems*. 2024. arXiv: 2406.08509 [math.FA].

[Bec+25] Lars Becker, Ohad Klein, Joseph Slote, Alexander Volberg, and Haonan Zhang. "Dimension-free discretizations of the uniform norm by small product sets". In: *Invent. Math.* 239.2 (2025), pp. 469–503. DOI: 10.1007/s00222-024-01306-9.

Chapter 1

Mathematical overview

A summary of the mathematical contributions in Part I. For motivations from computer science, one may skip to Chapter 2.

The analysis of Boolean functions is by now a well-established theory and an essential fiber in the fabric of theoretical computer science and discrete mathematics more generally. Any function $f : \{-1,1\}^n \to \mathbf{R}$ is uniquely expressed as a multilinear (or multi-affine) polynomial

$$f(x) = \sum_{S \subseteq [n]} \widehat{f}(S) \chi_S \qquad \text{with} \qquad \chi_S := \prod_{j \in S} x_j,$$

and various measures and aspects of the Fourier coefficients $\{\widehat{f}(S)\}_{S \subseteq [n]}$ provide significant insight into the structure of f. This perspective has led to many discoveries in computational complexity theory, learning theory, voting theory, coding theory, and combinatorics, among other areas [ODo14].

But not every function encountered in theoretical computer science is on the hypercube. A key generalization of $\{\pm 1\}^n$ is the so-called *hypergrid*, $[K]^n$, conveniently represented as the product of multiplicative cyclic groups

$$\Omega_K^n := \{\exp(2\pi i k/K) : k = 0, 1, \dots, K-1\}^n.$$

Harmonic and functional analysis on Ω_K^n is important in combinatorics [ALM91; Mes95], number theory [BS80], and graph theory [ADFS04], and naturally models many-candidate social choice functions and even certain aspects of K-level qudit systems in quantum computing.

Coming from the hypercube $\{\pm 1\}^n = \Omega_2^n$, some aspects of functional and harmonic analysis on Ω_K^n for general K are familiar. For example, the basics of influence and hypercontractivity are well-understood in this setting [Wei80; JPPP17]. But the analysis of functions on Ω_K^n is not just a retelling of the Boolean story; new and formidable barriers appear already at $K = 3$.

One example is the well-known *Plurality is stablest* conjecture from hardness of approximation [KKMO07], which remains open even though the Boolean case (*Majority is stablest*) was resolved 20 years ago [MOO10] (announced in 2005). In Part I we present a sequence of works surmounting another barrier in

the analysis over Ω_K^n appearing only for $K \geq 3$, motivated by a desire to extend a recent breakthrough in learning theory by Eskenazis and Ivanisvili [EI22] to more-general product spaces and their quantum analogues. Ultimately we reach beyond this goal, obtaining results with consequences for approximation theory and discrete analysis more broadly.

A challenge

Analysis over Ω_K^n is hard in part because Fourier expansions of functions on Ω_K^n, $K \geq 3$ are no longer multilinear. As we elaborate in the next chapter, for our target applications an essential technical step is to compare the supremum norm of $f : \Omega_K^n \to \mathbf{C}$ to the supremum norm of its extension to the product of convex hulls of Ω_K, $\mathrm{conv}(\Omega_K)^n$. Specifically, for f of degree at most d, we seek a comparison independent of dimension n of the form

$$\|f\|_{\mathrm{conv}(\Omega_K)^n} \stackrel{?}{\leq} C(K,d)\|f\|_{\Omega_K^n}. \tag{1.0.1}$$

Here and throughout, $\|\cdot\|_X$ will denote the supremum norm over a domain X.

To make sense of (1.0.1), recall that any $f : \Omega_K^n \to \mathbf{C}$ admits the unique Fourier expansion

$$f(z) = \sum_{\alpha \in \{0,1,\ldots,K-1\}^n} \widehat{f}(\alpha) z^\alpha \qquad \text{with} \qquad z^\alpha := \prod_{j=1}^n z_j^{\alpha_j}.$$

In this way, f can be uniquely extended to an analytic polynomial over $\mathbf{C}^n$. Throughout Part I we will conflate $f : \Omega_K^n \to \mathbf{C}$ with its associated polynomial and use terms like the *degree* of f to mean its total degree as an analytic polynomial. Note also that all polynomials arising thusly have *individual degree* (*i.e.,* the largest degree of any coordinate) at most $K-1$. Our results will hold for domains beyond Ω_K^n so we will find it most natural to state our results directly for analytic polynomials of bounded individual degree.

Now that we can parse Equation (1.0.1), what does it *mean*? One might compare it to the maximum modulus principal, which in this context would say that

$$\|f\|_{\mathrm{conv}(\Omega_K)^n} \leq \|f\|_{P_K^n},$$

where P_K is the boundary of $\mathrm{conv}(\Omega_K)$, *i.e.,* the boundary of the regular K-gon. Equation (1.0.1) asks whether a discrete grid of points can be used in place of the entire product of boundaries, possibly at the expense of introducing a dimension-free multiplicative constant.

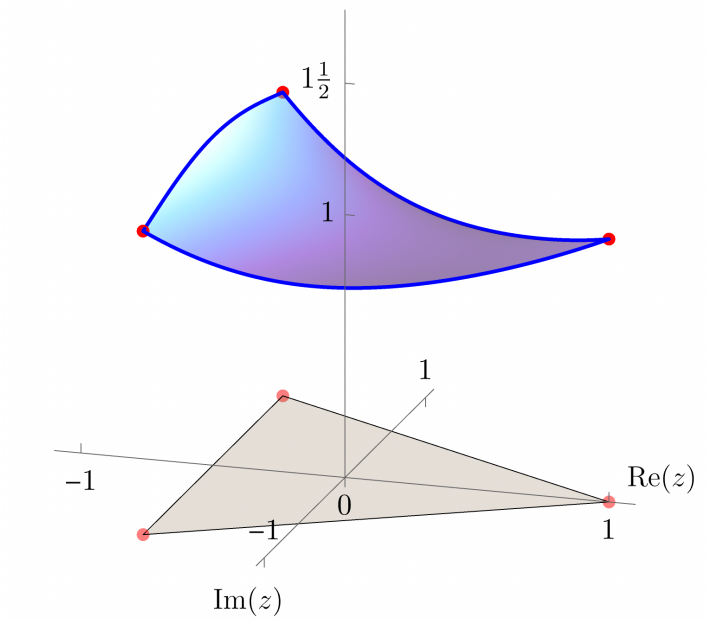

Figure 1.1: Failure of the "discrete maximum modulus principle." Here we plot the modulus of $p(z) = z^2/3 - z/3 + 1$ evaluated on $\mathrm{conv}(\Omega_3)$. Notice that while $|p(z)|$ is greatest on the boundary of $\mathrm{conv}(\Omega_3)$ (the maximum modulus principle), it is *not* maximized by a vertex from Ω_3 itself. We conclude $\|f\|_{\mathrm{conv}(\Omega_K)^n} \leq C\|f\|_{\Omega_K^n}$ cannot hold with constant 1 in general.

Let us sketch the difficulty in resolving (1.0.1) for different K. When $K = 2$, the comparison is a trivial *equality* with constant $C = 1$: because functions on the hypercube have multilinear extensions, we obtain $\|f\|_{[-1,1]^n} = \|f\|_{\{\pm 1\}^n}$ by convexity. But at $K = 3$ this doesn't work; in fact one can prove (see Figure 1.1 and [SVZ24a, Appendix B]) that even for $n = 1$ we must have $C(d, K) > 1$ in (1.0.1). As a result, naive approaches to proving (1.0.1) for $K \geq 3$ yield constants with exponential dependence on n.

On the other end of the K spectrum, if we take "$K = \infty$" and consider analytic functions on the polytorus

$$\mathbf{T}^n := \{z \in \mathbf{C} : |z| = 1\}^n,$$

the question is trivial, and again we get equality with constant 1:

$$\|f\|_{\mathrm{conv}(\mathbf{T})^n} = \|f\|_{\mathbf{D}^n} = \|f\|_{\mathbf{T}^n}$$

(this of course is just the maximum modulus principle). But there is no such easy fact for finite K because Ω_K is not the entire boundary of $\mathrm{conv}(\Omega_K)$. There seems to be a particular difficulty for finite $K \geq 3$.

Norm discretization inequalities

In approximation theory, comparisons of the kind (1.0.1) are known as *Bernstein-type discretization inequalities*—or discretizations of the uniform norm—and there is a vast body of work on the subject; see for example the surveys [DPTT19; KKLT22].[1] Bernstein-type discretization inequalities have the general shape of

$$\|f\|_X \leq C(d,n)\|f\|_T \tag{1.0.2}$$

for some domain X (usually convex) and finite *sampling set* $T \subset X$.

The typical goal for multivariate Bernstein-type inequalities is to establish a comparison of the kind (1.0.2) for C independent of degree d, but possibly depending on dimension n, while keeping T as small as possible. While this is the most natural multivariate generalization of the univariate Bernstein discretization inequality [Ber31; Ber32] (where independence of C from d was of course paramount), dependence of C on dimension n makes much of the norm discretization literature inapplicable for our hoped-for bound (1.0.1).

On the other hand, in some cases results with *universal* constants C are known, for example in the recent work of Dai and Prymak [DP24] resolving an important conjecture of Kroó [Kro11]. But these results also do not seem to apply either. The most salient reason is that the literature does not seem to distinguish between individual degree $K-1$ and (total) degree d, and as a result the cardinality of T always depends directly on $\deg(f)$. For our purposes, where d can easily grow much larger than K, we still require $T = \Omega_K^n$. And even if we restrict ourselves to the very special case of $d \leq K$, there is the issue of the *structure* of T. Ω_K^n has a very simple but rigid product structure, whereas Bernstein-type discretization inequalities in the literature with C independent of n are either proofs of existence of T of low cardinality, or they construct T's that have intricate ε-net-type structures, as is the case for the sampling sets in Dai and Prymak [DP24], which are far from being product sets.

In fact, as we explain in the sequel, it turns out that for $T = \Omega_K^n$ to be a workable sampling set at all for us (*i.e.,* yielding C independent of n), the constant C *must* depend on degree d. So our hoped-for norm discretization inequality is of a distinct flavor from the traditional results of multivariate norm discretization. Indeed, if (1.0.1) were true, it would represent a new perspective in high-dimensional norm discretization—one that demands C be

[1] *Warning:* in the norm discretization literature, d is typically used for dimension and n for degree, precisely the opposite of what is typical in analysis of Boolean functions.

independent of dimension while allowing degree dependence—and would show that new and interesting features appear in this regime, such as the role of individual degree.

1.1 Results

The first headline result of Part I is a resolution of (1.0.1). (It turns out $\|f\|_{\mathrm{conv}(\Omega_K)^n}$ and $\|f\|_{\mathbf{T}^n}$ are easily comparable independent of dimension[2] so we will state results using the latter as it is cleaner.) In fact, a broad generalization of (1.0.1) is proved, but we stick to the below for the purposes of this discussion.

Theorem 1 (Dimension-free discretization of the uniform norm)**.** *Let f be an n-variate analytic polynomial of degree d and individual degree $K-1$. Then*

$$\|f\|_{\mathbf{T}^n} \leq \mathcal{O}(\log K)^d \|f\|_{\Omega_K^n}.$$

We will give two proofs of Theorem 1 which generalize it in different ways.

Proof I: Fourier multipliers [SVZ25]**.** This is the historically-first proof and obtains only an implicit constant (still dimension-free) in place of $\mathcal{O}(\log K)^d$. Along the way to Theorem 1 the proof develops a rich class of Fourier multipliers that are L^∞-bounded independent of dimension and may be of independent interest. For prime K this class of multipliers is characterized exactly thanks to connections to transcendental number theory and Baker's celebrated theorem on the logarithms of algebraic numbers [Bak22].

Theorem 2 (Bounded Fourier projections, prime $K \geq 3$)**.** *Suppose K is an odd prime and let S be a maximal subset of $\{0,1,\ldots,K-1\}^n$ such that for all $\alpha, \beta \in S$:*

- *(Total) degrees are equal: $\sum_{j=1}^n \alpha_j = \sum_{j=1}^n \beta_j$.*
- *Individual degree symmetry: there is a bijection $\pi : [n] \to [n]$ such that for all $j \in [n]$, $\alpha_j = \beta_{\pi(j)}$ or $\alpha_j = K - \beta_{\pi(j)}$. (In particular, α and β are nonzero on the same number of indices.)*

Then for any n-variate analytic polynomial f of degree at most d and individual degree at most $K-1$, the S-part of f, $f_S(x) := \sum_{\alpha \in S} \widehat{f}(\alpha) z^\alpha$, satisfies

$$\|f_S\|_{\Omega_K^n} \lesssim_{d,K} \|f\|_{\Omega_K^n}.$$

[2] One direction is immediate. For the other, let $c_K > 0$ be such that $c_K \mathbf{T} \subset \mathrm{conv}(\Omega_K)$. Because f has bounded degree, $\|f\|_{\mathbf{T}^n} \simeq \|f\|_{(c_K \mathbf{T})^n}$, which is at most $\|f\|_{(\mathrm{conv}\Omega_K)^n}$.

Proof II: polynomial interpolation [KSVZ24; Bec+25]**.** This proof is probabilistic and obtains Theorem 1 with the quoted constant. The key technique is a novel *interpolation formula*, which is our second headline result.

Theorem 3 (Dimension-free multivariate interpolation)**.** *For any $z \in \mathbf{D}^n$ there exist explicit coefficients $\{a_x^{(z)}\}_{x \in \Omega_K^n}$ such that for any n-variate analytic polynomial f of degree d and individual degree $K-1$,*

$$f(z) = \sum_{x \in \Omega_K^n} a_x^{(z)} f(x)\,,$$

and $\sum_x |a_x^{(z)}| \le \mathcal{O}(\log K)^d$.

While interpolation of degree-d polynomials f can be accomplished by much smaller sets than Ω_K^n, Theorem 3 appears to be the first multivariate interpolation formula for which the coefficients $\{a_x^{(z)}\}_x$ have ℓ^1 norm independent of dimension n. Moreover, a point set with cardinality exponential in n, like Ω_K^n, is *necessary* to get $\|a^{(z)}\|_{\ell^1}$ independent of dimension. This dimension-independence is crucial for obtaining Theorem 1 from Theorem 3 and to downstream applications. We hope Theorem 3 finds other uses in approximation theory and analysis.

Remark. Theorems 1 and 3 may actually be generalized from Ω_K^n to a very generic class of sampling sets $S \subset \mathbf{D}^n$ (though the K-dependence of the constant becomes more complicated). This is done in the work [Bec+25]; see Chapter 3 for the general statement.

1.2 Applications in analysis

Theorems 1 and 3 have several consequences in analysis and approximation theory.

New Bohnenblust–Hille inequalities

Theorem 4 (Cyclic-group Bohnenblust–Hille inequality)**.** *Let $f : \Omega_K^n \to \mathbf{C}$ have degree at most d. Then*

$$\|\widehat{f}\|_{\frac{2d}{d+1}} \lesssim_{K,d} \|f\|_{\Omega_K^n}\,.$$

Here and throughout, $\|\widehat{f}\|_p$ denotes the ℓ^p norm of the Fourier coefficients for f. This theorem extends to Ω_K^n the classical inequality of Bohnenblust and Hille (BH), which originally appeared in the study of Dirichlet series and has a long

history in harmonic analysis. The hypercube formulation was studied more recently [Ble01; DMP18] and has found surprising applications in computer science. The BH inequality for Ω_K^n, $K \geq 3$ was not known until we gave a proof in [SVZ24a]. That publication actually predates our discretization inequality (Theorem 1), but we elected to omit [SVZ24a] from this thesis because (*i*) with the discretization inequality in hand, Theorem 4 becomes a one-line reduction to the original BH inequality [BH31]; and (*ii*) the techniques employed in [SVZ24a] are directly subsumed by those in the Fourier multiplier proof of Theorem 1, [SVZ25].

Theorem 1 is also a key ingredient in the proof of a *noncommutative* (or quantum) Bohnenblust–Hille inequality.

Theorem 5 (Qudit Bohnenblust–Hille inequality)**.** *Let A be a d-local quantum observable on n-many K-level qudits. Then*

$$\|\widehat{A}\|_{\frac{2d}{d+1}} \lesssim_{d,K} \|A\|_{\mathrm{op}}.$$

The "Fourier transform" $\widehat{A}$ here refers to the vector of coefficients obtained by decomposing A in the Heisenberg–Weyl basis, which is a unitary generalization of the familiar Pauli basis. This is discussed in detail in Section 6.3.

Theorem 5 generalizes the qu*b*it BH inequality that was first proved in [HCP22] and then improved in [VZ23]. The proof of Theorem 5 combines Theorem 1 with a careful analysis of the commutation structure of Heisenberg–Weyl matrices.

Level-ℓ Fourier projections are bounded

Theorem 6 (Boundedness of the level-ℓ Fourier projection)**.** *Let $f : \Omega_K^n \to \mathbf{C}$ have degree at most d and let ℓ be a positive integer. Then with $f_\ell = \sum_{|\alpha|=\ell} \widehat{f}(\alpha) z^\alpha$ denoting the ℓ-homogeneous part of f,*

$$\|f_\ell\|_{\Omega_K^n} \lesssim_{K,d} \|f\|_{\Omega_K^n}.$$

This theorem shows that the level-ℓ Fourier projection, when applied to low-degree polynomials, is bounded independent of dimension. This is a common Fourier multiplier-type estimate, and dimension free-ness for polynomials on Ω_K^n, $K \geq 3$ was not known until the discovery of Theorem 1. The $K = 2$ case has a short proof usually attributed to Figiel [MS86, §14.6]. When K is prime Theorem 6 can be seen as a specialization of the very fine-grained projection theorem Theorem 2, though more generally it follows easily from Theorem 1 and can be stated with the explicit constant $\mathcal{O}(\log K)^d$.

L^p discretization

Theorem 7 (Dimension-free discretization of L^p norms)**.** *Let $d, n \geq 1, K \geq 2$. Let $1 \leq p \leq \infty$. Then for any polynomial $f : \mathbf{T}^n \to \mathbf{C}$ of degree at most d and individual degree at most $K-1$, we have*

$$\|f\|_{L^p(\mathbf{T}^n)} \leq d \cdot \mathcal{O}(\log K)^d \|f\|_{L^p(\Omega_K^n)} .$$

This L^p discretization inequality could be called a Marcinkiewicz–Zygmund-type (MZ-type) discretization inequality, though it differs from the typical MZ-type inequalities in the literature in that the constant in Theorem 7 is independent of p but depends strongly on d. The proof combines the interpolation formula of Theorem 3 with invariance of the uniform (Haar) measure on Ω_K^n. One can also get a dimension-free comparison *à la* Theorem 7 via hypercontractivity [Wei80; JPPP17], but again the constant will depend on p.

1.3 Applications in learning theory

A basic task in learning theory is to learn an L^2 approximation to a degree-d function f, given access to random samples. The naive algorithm requires $\mathcal{O}(n^d)$ samples but we obtain an algorithm with exponentially-better n dependence.

Theorem 8 (Cyclic-group low degree learning)**.** *Let $f : \Omega_K^n \to \mathbf{D}$ have degree d. Then with*

$$(\log K)^{\mathcal{O}(d^2)} \cdot \log\left(\frac{n}{\delta}\right) \cdot \left(\frac{1}{\varepsilon}\right)^{d+1}$$

independent random samples of the form $(\boldsymbol{x}, f(\boldsymbol{x}))$, $\boldsymbol{x} \sim \Omega_K^n$, we may with confidence $1-\delta$ learn a function $\widetilde{f} : \Omega_K^n \to \mathbf{C}$ with $\|f - \widetilde{f}\|_2^2 \leq \varepsilon$.

This learning theorem extends a breakthrough result of Eskenazis and Ivanisvili [EI22] from functions on the hypercube to those on Ω_K^n, and is obtained by combining the cyclic-group Bohnenblust–Hille inequality, Theorem 4, with ideas in [EI22].

An analogous result holds in the context of quantum computing, where role of the function $f : \Omega_K^n \to \mathbf{C}$ is taken by a quantum observable on K-level qudits.

Theorem 9 (Low-degree Qudit Learning)**.** *Let A be a degree-d (or d-local) observable on n K-level qudits with $\|A\|_{\mathrm{op}} \leq 1$. Then there is a collection S of product states such that with a number*

$$\mathcal{O}\left(\left(K\|A\|_{\mathrm{op}}\right)^{C \cdot d^2} d^2 \varepsilon^{-d-1} \log\left(\tfrac{n}{\delta}\right)\right)$$

of samples of the form $(\rho, \mathrm{tr}[A\rho])$, $\rho \sim S$, *we may with confidence* $1-\delta$ *learn an observable* $\widetilde{A}$ *with* $\|A - \widetilde{A}\|_2^2 \leq \varepsilon$.

See Section 6.4 for the full definition and explanation.

I am very grateful to my coauthors Lars Becker, Ohad Klein, Alexander Volberg, and Haonan Zhang for their collaboration in the various papers constituting Part I.

Chapter 2

From learning theory to discretization inequalities

We chart a path from motivations in quantum and classical learning theory to our discretization inequality, Theorem 1.

Boolean functions whose multilinear expansions are *low degree* (think of $\deg(f)$ as much less than dimension n, or even constant) are ubiquitous in computer science [ODo14] and it is a fundamental task to learn them from random samples.

Task (Low-degree learning)**.** *Using R-many uniformly-random samples*

$$\left\{\left(\boldsymbol{x}^{(r)}, f(\boldsymbol{x}^{(r)})\right)\right\}_{r=1}^{R}$$

of an unknown Boolean function $f : \{\pm 1\}^n \to [-1,1]$ of degree d, produce with probability $2/3$ a function $g : \{\pm 1\}^n \to \mathbf{R}$ such that $\|f - g\|_2 \leq \varepsilon$.

Here the norm $\|\cdot\|_2$ is with respect to the uniform probability measure on $\{\pm 1\}^n$, and the figure of merit for the task is *sample complexity*, *i.e.,* the dependence of R on d, ε, and especially n.

In 1993 Linial, Mansour, and Nisan [LMN93a] gave a very natural $\widetilde{\mathcal{O}}_\varepsilon(n^d)$-sample algorithm for low degree learning: from samples one may form the *empirical Fourier coefficients*

$$\widehat{\boldsymbol{g}}(S) := \tfrac{1}{R}\textstyle\sum_{r=1}^{R} f\left(\boldsymbol{x}^{(r)}\right)\chi_S\left(\boldsymbol{x}^{(r)}\right)$$

for all $S \subseteq [n]$ with $|S| \leq d$, and from there define the estimator

$$\boldsymbol{g} := \textstyle\sum_{|S| \leq d} \widehat{\boldsymbol{g}}(S)\chi_S\,.$$

The analysis is textbook: to obtain, say, $\|f - \boldsymbol{g}\|_2 \leq 0.01 = \varepsilon$, it suffices to get

$$|\widehat{f}(S) - \widehat{\boldsymbol{g}}(S)| \leq 0.001/n^d \quad \text{for each} \quad S \subseteq [n], |S| \leq d, \tag{2.0.1}$$

for then

$$\|f - \boldsymbol{g}\|_2 \overset{\text{(Plancherel)}}{=} \|\widehat{f} - \widehat{\boldsymbol{g}}\|_2 \;\leq\; \|\widehat{f} - \widehat{\boldsymbol{g}}\|_1 \;=\; \sum_{|S| \leq d} |\widehat{f}(S) - \widehat{\boldsymbol{g}}(S)| \;\leq\; 0.01.$$

And standard concentration arguments say (2.0.1) can be accomplished with $\widetilde{\mathcal{O}}(n^d)$ samples with high probability.

For almost 30 years this sample complexity stood without improvement, and for good reason: a natural heuristic calculation suggests no improvement is available. Degree-d polynomials have roughly n^d Fourier coefficients, and each is bounded by $\|\widehat{f}\|_\infty \leq 1$, so we might expect an ε-net of such f has cardinality at least $(1/\varepsilon)^{n^d}$—which would require $\Omega(n^d \log(1/\varepsilon))$ samples to learn.

Yet, 28 years after [LMN93a], [Iye+21] found a $\widetilde{\mathcal{O}}_\varepsilon(n^{d-1})$-sample algorithm based on improved bounds on the *Fourier growth* of low degree polynomials, *i.e.,* bounds on $\sum_{|S|=k} |\widehat{f}(S)|$ for degree-d f as a function of d, k, and n. Although in some sense a modest improvement on the performance of the basic low degree learning algorithm, the work of [Iye+21] already contradicts the heuristic lower bound just described. So what is wrong with our back-of-the-envelope calculation? It turns out we failed to take into account the *interaction* of the degree constraint and the L^∞ bound constraint on f. Soon after the work of [Iye+21], Eskenazis and Ivanisvili [EI22] showed that these two constraints interact so strongly that actually an $\widetilde{\mathcal{O}}_{\varepsilon,d}(\log n)$-sample algorithm was possible.

The breakthrough of Eskenazis–Ivanisvili

Eskenazis and Ivanisvili leveraged a theorem about bounded low-degree functions coming from classical harmonic analysis, *Bohnenblust–Hille (BH) inequality*, that is at this point quite classical [BH31]. The original BH inequality applies to degree-d analytic polynomials on $\mathbf{D}^n$.

Theorem 10 (Bohnenblust–Hille inequality [BH31])**.** *Let $f : \mathbf{T}^n \to \mathbf{C}$ be an analytic polynomial of degree at most d. Then*

$$\|\widehat{f}\|_{\frac{2d}{d+1}} \leq C_d \|f\|_{\mathbf{T}^n}.$$

Here and throughout the norm $\|\widehat{g}\|_p$ denotes the ℓ^p norm of the Fourier coefficients of g. The two key aspects of the BH inequality are (i) the independence of the constant C_d from dimension n, and (ii) that the ℓ^p norm on the left-hand side is for a p strictly less than 2. For $p = 2$ the comparison of course holds with constant 1 for *all* functions (Plancherel); actually $p = 2d/(d+1)$ is the smallest p admitting a dimension-free comparison of the kind above [BH31]. Together the properties (i) and (ii) place powerful constraints on the Fourier spectrum of bounded, low-degree polynomials, as we will detail in the sequel. We also remark that in many applications the dependence of C_d on d is quite

important; the best bound we know currently is $C_d \le C^{\sqrt{d \log d}}$ due to [BPS14]. See [DGMS19, Ch. 6] for more on the classical Bohnenblust–Hille inequalities.

The analogous inequality for Boolean functions has a shorter history. The hypercube BH inequality was originally proved with an implicit constant C_d by Blei [Ble01] and then was only very recently improved to the subexponential $C_d = C^{\sqrt{d \log d}}$ [DMP18]. Again, $p = 2d/(d+1)$ is the best possible.

Theorem 11 (Hypercube BH inequality)**.** *Let $f : \{\pm 1\}^n \to \mathbf{R}$ be a Boolean function of degree d. Then*

$$\|\widehat{f}\|_{\frac{2d}{d+1}} \le C_d \|f\|_\infty . \tag{2.0.2}$$

Moreover, $C_d \le C^{\sqrt{d \log d}}$ for a universal constant C.

How can we make use of such a bound? The insight of Eskenazis–Ivanisvili is as follows. Beginning with some unknown vector $v \in \mathbf{R}^n$ (such as the vector of Fourier coefficients of f), if one has an ℓ^∞ estimate of v, *i.e.*, a w such that

$$\|v - w\|_\infty \le \varepsilon,$$

as well as the guarantee that $\|v\|_{\ell^p}$ is bounded for some $p < 2$, then w can be improved to an ℓ^2 estimate $\widetilde{w}$ of v, still controlled by ε:

$$\|v - \widetilde{w}\|_2 \lesssim \varepsilon^{1 - \frac{p}{2}} .$$

Specifically, $\widetilde{w}$ is obtained from w by replacing small entries by 0—a task that crucially does not need any knowledge of v's entries.[1]

Lemma 12. *Let $p \in [0, 2)$ and $\varepsilon, B > 0$. Suppose $v, w \in \mathbf{C}^n$ with $\|v - w\|_\infty \le \varepsilon$ and $\|v\|_p \le B$. Then for $\widetilde{w}$ defined as*

$$\widetilde{w}_j = \begin{cases} w_j & \text{if } |w_j| \ge \varepsilon\left(1 + \sqrt{\frac{2}{2-p}}\right) \\ 0 & \text{otherwise,} \end{cases} \tag{2.0.3}$$

we have the bound

$$\|\widetilde{w} - v\|_2 \le 5B\varepsilon^{1 - \frac{p}{2}}.$$

[1] With the one exception of choosing the threshold for "small," which is set in terms of the norm bounds just mentioned.

With these pieces in place, the algorithm of Eskenazis–Ivanisvili is straightforward: begin by forming the empirical vector of Fourier coefficients $\widehat{g}$ as before, using enough samples so that with high probability, $\|\widehat{f} - \widehat{g}\|_\infty \leq \varepsilon$. From Theorem 11 we know $\|\widehat{f}\|_{\frac{2d}{d+1}} \leq \mathcal{O}(1)$ always, so we are exactly in the situation of Lemma 12: replacing the small entries of $\widehat{g}$ with 0, we may obtain a new estimator $\widehat{h}$ of $\widehat{f}$ with the guarantee that $\|\widehat{f} - \widehat{h}\|_2 = \|f - h\|_2$ is controlled by a power of ε, without any n-dependence in the inequality. As a result, we only need to gather enough data so that $\|\widehat{f} - \widehat{g}\|_\infty \leq \mathrm{poly}(\varepsilon)$, which is much easier to achieve than the $\mathcal{O}(\varepsilon/n^d)$ requirement of the naive algorithm. In fact, the only source of n dependence ends up being the $\mathcal{O}_d(\log n)$ copies required for concentration of measure to overcome the loss when union-bounding over the events

$$\left\{|\widehat{f}(S) - \widehat{g}(S)| \leq \mathcal{O}(1)\right\}_{S \subset [n], |S| \leq d}.$$

We will not delve into it here, but the estimator $\widehat{h}$ for $\widehat{f}$ could be called a *hard-thresholding estimator* in statistics. The proof of Lemma 4 appears later on in Section 6.4, but for now we just point out that p being strictly less than 2 is absolutely critical: eventually one needs the ℓ^2 norm of the small entries of v (the ones zeroed-out in $\widetilde{w}$) to be bounded by a function of $\varepsilon \geq \|v - w\|_\infty$. With $t = t(\varepsilon, p)$ the threshold in (2.0.3),

$$\sum_{j : |w_j| < t} |v_j|^2 = \sum_{j : |w_j| < t} |v_j|^{2-p} |v_j|^p \leq (t + \varepsilon)^{2-p} \sum_{j : |w_j| < t} |v_j|^p \leq (t + \varepsilon)^{2-p} B^{1/p}.$$

And recalling t is linear is ε, we recognize the right-hand side of this display as $C_{B,p} \varepsilon^{2-p}$. On the other hand, if we only had the trivial bound $\|v\|_{\ell^2} \leq 1$, there is no reason the ℓ^2 norm of the small entries of v should go to 0 as $\varepsilon \to 0$.

We also remark that while the algorithm of Eskenazis–Ivanisvili achieves a $\log(n)$ *sample* complexity, the *time* complexity of the algorithm is still $\mathcal{O}(n^d)$ because each $\widehat{g}(S)$ for $|S| \leq d$ is compared with the threshold in sequence. It is an interesting open problem to determine if the runtime can be improved to $o(n^d)$, as the related problem of k-junta learning has seen a few improvements of this kind, *e.g.,* [Val15].

Quantum generalizations

Soon after the Eskenazis–Ivanisvili breakthrough, Rouzé, Wirth, and Zhang conjectured a quantum generalization [RWZ24a]. The idea is to learn *local quantum observables.* An (n-qubit) quantum observable is a Hermitian matrix

$A \in \mathbf{M}_{2^n \times 2^n}(\mathbf{C})$. Such an observable models the following process: an n-qubit quantum state ρ is subjected to an unknown transformation (quantum channel) $\mathcal{N}$, and then a numerical measurement (with outcomes in, say, $[-1,1]$) is performed on the output. The matrix A captures the statistics of this process, in the sense that $\mathrm{tr}[A\rho]$ is the expected measurement outcome for $\mathcal{N}(\rho)$. Indeed, if $\{N_j\}_j$ are the Kraus operators for $\mathcal{N}$ (so that $\mathcal{N}(\rho) = \sum_j N_j \rho N_j^\dagger$) and M is the measurement operator of the measurement device, then

$$A = \textstyle\sum_j N_j^\dagger M N_j \,.$$

It is a natural task to learn a description of the matrix A so that the expectation value map $\rho \mapsto \mathrm{tr}[A\rho]$ can be predicted for new ρ. Learning descriptions of arbitrary $2^n \times 2^n$ matrices is very difficult, so one simplifying assumption comes by way of a quantum (or noncommutative) generalization of polynomial degree bounds. Any n-qubit observable A admits a unique Fourier-like decomposition

$$A = \sum_{\alpha \in \{0,1,2,3\}^n} \widehat{A}(\alpha)\sigma_\alpha, \qquad \sigma_\alpha := \bigotimes_{j=1}^n \sigma_{\alpha_j},$$

where the σ_j's are the Pauli matrices

$$\sigma_0 := \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_1 := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 := \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 := \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The observable A is *d-local* if $\widehat{A}(\alpha) = 0$ when $|\alpha| > d$, where $|\alpha|$ is the number of entries j for which $\alpha_j \neq 0$.

This is all looking very familiar. As we explain in the Section 6.4, by obtaining samples of the form $(\rho, \mathrm{tr}[A\rho])$ for a certain distribution of states ρ, it is easy to construct an estimator for the "Pauli coefficients" of A, and one might hope for an $\mathcal{O}_{\varepsilon,d}(\log n)$-sample learning algorithm for d-local A. The missing piece was a quantum (qubit) version of the Bohnenblust–Hille inequality.

Theorem 13 (Qubit BH Inequality)**.** *Let A be a d-local observable. Then*

$$\|\widehat{A}\|_{\frac{2d}{d+1}} \leq C_d \|A\|_{\mathrm{op}} \,.$$

This theorem was conjectured in [RWZ24b] and a proof appeared very soon after in [HCP22] with constant $C_d = \mathcal{O}(d^d)$. A very different proof that improves the constant to $C_d = C^d$ was then given by Volberg and Zhang [VZ23].

This second proof is very efficient and directly reduces to the hypercube BH inequality. In brief, for any d-local observable A the authors identify a degree-d Boolean function $f_A : \{\pm 1\}^{3n} \to \mathbf{R}$ such that

$$\|\widehat{A}\|_{\frac{2d}{d+1}} \leq 3^d \|\widehat{f_A}\|_{\frac{2d}{d+1}} \qquad \text{and} \qquad \|\widehat{f}_A\|_{\{\pm 1\}^{3n}} \leq \|A\|_{\mathrm{op}} \,. \tag{2.0.4}$$

The hypercube BH inequality joins these two bounds to complete the theorem.

To larger product spaces

It is natural to ask whether this story continues to larger product spaces (quantum and classical), and this was the starting point for the work presented in Part I of this thesis. Classically, such an extension looks like learning low-degree functions on products of cyclic groups $\mathbf{Z}_K^n$, which we shall represent by the multiplicative group of K^{th} roots of unity, Ω_K^n. As mentioned in the previous chapter, the space Ω_K^n is sometimes called the "hypergrid" in property testing [CS14; BCS23], is a key setting for studying the hardness of approximation (*e.g.*, the Plurality is Stablest Conjecture of [KKMO07], see also [MOO10]), and appears frequently in coding theory and cryptography.

The argument of Eskenazis–Ivanisvili generalizes directly to such functions, provided the appropriate BH inequality can be proved:

Conjecture 1 (Cyclic-group Bohnenblust–Hille inequality)**.** *Let $f : \Omega_K^n \to \mathbf{C}$ have degree d. Then*

$$\|\widehat{f}\|_{\frac{2d}{d+1}} \leq C_{d,K} \|f\|_{\Omega_K^n} \,.$$

As will be explained soon, it is *not* possible to mimic the proof of the hypercube (or the original polytorus) BH inequality for functions on Ω_K^n. This conjecture is the main sticking point for extending classical low-degree learning to larger product spaces.

The *quantum* case of generalized low-degree learning is just as important, both for the study of fundamental physics via quantum simulation (*e.g.*, [Kur+21; Gon+22]) and in the operation and validation of quantum computers. In both contexts, gains in efficiency are possible when the underlying hardware system is composed of higher-dimensional subsystems, sometimes carrying an algorithm from theoretical fact to practical reality in the NISQ era [Gon+22]—and this benefit may very well remain as quantum computing advances. Such devices are called *multilevel system*, or *qudit*, quantum computers [WHSK20]. While the qu*b*it case gives a conceptual sense of the possibilities

for learning on qudit systems, it is of practical value to derive guarantees and algorithms that work directly in the native dimension of the quantum system.

To formulate a Bohnenblust–Hille inequality for qudit systems, we must decide on a qudit generalization of Pauli matrices. A natural choice are the Heisenberg–Weyl (HW) matrices, which are their unitary generalizations. We defer a formal definition till Section 6.3.

Conjecture 2 (Qudit BH inequality)**.** *Let A be a degree-d (or d-local) observable on n-many K-level qudits. Then*

$$\|\widehat{A}\|_{\frac{2d}{d+1}} \leq C_{d,K}\|A\|_{\mathrm{op}}\,.$$

Following the approach of Zhang and Volberg, there are two steps to proving the qudit BH inequality. The first is, given an n-qudit operator A, to identify a commutative polynomial f_A with bounds analogous to Equation (2.0.4). This is technically much more challenging than in the qubit case because the Heisenberg–Weyl matrices have a very intricate commutation structure. Partial results on this part appeared in [KSVZ24] and the full reduction finished in [SVZ24b], the proof of which is included in Section 6.3. The other part is to prove the appropriate commutative BH inequality, which turns out again to be the *cyclic-group* Bohnenblust–Hille inequality because the eigenvalues of HW matrices are roots of unity.

Therefore our goals of generalizing low-degree learning to larger product spaces, both classically and quantumly, dovetail into the task of proving a cyclic-group Bohnenblust–Hille inequality.

BH for cyclic groups: the barrier and a first resolution

Although a cyclic-group BH inequality is the natural interpolating statement between the now well-understood hypercube BH (polynomials on Ω_2^n) and the classical BH inequality, for polynomials on $\mathbf{T}^n$ (or "Ω_∞^n"), there was no proof in the literature. And this is for good reason, as the regime of $2 < K < \infty$ faces unique difficulties.

At a very coarse level, the hypercube and polytorus BH inequalities are both proved in the following steps [DS14] (using X^n to represent either product space):

1. Symmetrization: express f as the restriction of a certain symmetric d-linear form L_f to the diagonal $\Delta := \{(z,\dots,z) : z \in X^n\}$; that is, $f(z) = L_f(z,\dots,z)$.

2. BH inequality for multilinear forms: bound the $\ell_{2d/(d+1)}$ norm of the coefficients of L_f (which are directly related to the coefficients of f) by the supremum norm of L_f over $(X^n)^d$. This step is rather involved and includes several estimates, manipulations, and an application of hypercontractivity and Khintchine's inequality.

3. Polarization: estimate the supremum norm of L_f on its entire domain $(X^n)^d$ by the supremum over Δ; that is,

$$\|L_f\|_{(X^n)^d} \lesssim \|L_f\|_{\Delta} = \|f\|_{X^n},$$

where $\|\cdot\|_E$ denotes the supremum norm over some space E.

When adapting this proof structure to cyclic groups of order $2 < K < \infty$, the main point of failure is in step three, polarization, as is fully worked out in [SVZ24a, Appx. A]. In both the polytorus and hypercube cases, one uses Markov–Bernstein-type inequalities to obtain the intermediate comparison

$$\|L_f\|_{(X^n)^d} \lesssim \|f\|_{\mathrm{conv}(X)^n},$$

where $\mathrm{conv}(X)$ denotes the convex hull of X. The passage from $\mathrm{conv}(X)$ to X is then immediate for the polytorus by the maximum modulus principle ($\|f\|_{\mathbf{D}^n} = \|f\|_{\mathbf{T}^n}$) and for the hypercube by multilinearity ($\|f\|_{[-1,1]^n} = \|f\|_{\{-1,1\}^n}$). It is at this most trivial step that the BH proof breaks down for $K > 3$: there is no such easy fact in the setting of the multiplicative cyclic group $\Omega_K := \{e^{2\pi i k/K} : k = 0, \dots, K-1\}$ with $2 < K < \infty$ because Ω_K is not the entire boundary of $\mathrm{conv}(\Omega_K)$. Even for $n = 1$ and $K = 3$ one can construct example f's for which $\|f\|_{\mathrm{conv}(\Omega_K)^n} > \|f\|_{\Omega_K^n}$. It's easy to show that $\|f\|_{\mathrm{conv}(\Omega_K)^n}$ and $\|f\|_{\mathbf{D}^n} = \|f\|_{\mathbf{T}^n}$ are comparable independent of dimension, so essentially one is left wondering whether

$$\|f\|_{\mathbf{T}^n} \lesssim_{d,K} \|f\|_{\Omega_K^n}. \tag{2.0.5}$$

And in fact, if one could prove such a comparison, then there would be no need to retrace the rest of the BH proof recipe in the first place! One would immediately get the cyclic-group BH inequality via

$$\|\widehat{f}\|_{\frac{2d}{d+1}} \overset{\text{(Classical BH)}}{\lesssim_d} \|f\|_{\mathbf{T}^n} \overset{(2.0.5)}{\lesssim_{d,K}} \|f\|_{\Omega_K^n}.$$

The inequality (2.0.5), a certain "discretization of the uniform norm," and its generalizations are the main contribution of Part I. For now, let us remark that

the cyclic-group BH inequality was first proved in [SVZ24a] by a method that actually avoided proving the comparison (2.0.5) and instead made an intricate reduction to the hypercube BH inequality. That argument has two main parts. The first is to prove a cyclic-group BH-type inequality for *support-homogeneous* polynomials, *i.e.,* those whose monomials all contain the same number of distinct variables. The second is to show the comparison $\|f_{(\ell)}\|_{\Omega_K^n} \lesssim_{d,K} \|f\|_{\Omega_K^n}$ where $f_{(\ell)}$ is the ℓ-support-homogeneous part of a degree-d polynomial f. The constant in the comparison, while free of dependence on n, was quite large and never estimated explicitly. We have elected to not include the proof of [SVZ24a] in this thesis as the techniques involved were later sharpened and used again in [SVZ25], the first proof of (2.0.5).

The discretization inequality: a lingering desire

Although the cyclic-group BH inequality was established in [SVZ24a], the question of whether a comparison of the type (2.0.5) is possible still lingered. And from the point of view of the larger project of harmonic analysis on Ω_K^n, there are other motivations for proving (2.0.5). New difficulties arise when trying to get *other* standard estimates in this space too. For example, consider the boundedness of k-level Fourier projections. Concretely, given a polynomial $f : \Omega_K^n \to \mathbf{C}$ of total degree at most d and individual degree at most $K-1$, we seek to control its degree-ℓ homogeneous part f_ℓ as follows:

$$\|f_\ell\|_{\Omega_K^n} \lesssim_{d,K} \|f\|_{\Omega_K^n} . \tag{2.0.6}$$

This is a typical kind of dimension-free estimate in harmonic analysis, and something that is very easy to accomplish both on Ω_2^n and on $\mathbf{T}^n$.

When the domain is the polytorus $\mathbf{T}^n$, this comparison is a standard Cauchy estimate and bears constant 1: given an analytic function $f : \mathbf{T}^n \to \mathbf{C}$ and z^* a maximizer of $|f_\ell|$, define the univariate function $Q(t) = f(tz_1^*, \dots, tz_n^*)$ for $t \in \mathbf{T}$. Taking the ℓ^{th} derivative and appealing to the Cauchy's integral formula, we find

$$\|f_\ell\|_{\mathbf{T}^n} = \frac{|Q^{(\ell)}(0)|}{\ell!} = \frac{1}{2\pi}\left|\int_{\mathbf{T}} \frac{Q(t)}{t^{\ell+1}}\,\mathrm{d}t\right| \le \|Q\|_{\mathbf{T}} \le \|f\|_{\mathbf{T}^n} . \tag{2.0.7}$$

For the hypercube ($K = 2$), this estimate is similar and usually attributed to Figiel [MS86, §14.6]. Given $f : \Omega_2^n \to \mathbf{R}$, $\deg(f) \le d$, and with $x^* \in \Omega_2^n$ a maximizer of $|f_\ell|$, one considers the polynomial $Q(t) := f(tx_1^*, \dots, tx_n^*)$ for

$t \in [-1,1]$ (f is extended to $[-1,1]^n$ as a multilinear function). A Markov–Bernstein-type estimate gives

$$\|f_\ell\|_{\Omega_2^n} = \frac{|Q^{(\ell)}(0)|}{\ell!} \le C(d,\ell)\|Q\|_{[-1,1]} \le C(d,\ell)\|f\|_{[-1,1]^n},$$

with optimal constant $C(d,\ell) \le (1+\sqrt{2})^d$ [DMP18, Lemma 1.3 (4)]. The final step is to recognize that $\|f\|_{[-1,1]^n} = \|f\|_{\{-1,1\}^n}$ because the extension of f to $[-1,1]^n$ is affine in each coordinate.

However, as soon as $K = 3$ it is quite unclear how to proceed. For example, one could analogize the argument from above, constructing a polynomial $Q(t)$ with t now in the disk $\mathbf{D}$, to obtain via the Cauchy estimate

$$\|f_\ell\|_{\Omega_K^n} = \frac{|Q^{(\ell)}(0)|}{\ell!} \le \|Q\|_{\mathbf{D}} = \|Q\|_{\mathbf{T}}\,.$$

Unfortunately, there is no simple way to relate $\|Q\|_{\mathbf{T}}$ to $\|f\|_{\Omega_3^n}$. Certainly $\|Q\|_{\mathbf{T}} \le \|f\|_{\mathbf{T}^n}$, but then it seems we would again need the dimension-free comparison

$$\|f\|_{\mathbf{T}^n} \lesssim_{d,K} \|f\|_{\Omega_K^n}\,,$$

a reappearance of the discretization inequality.

And so we see that many roads lead to a dimension-free discretization of the uniform norm on $\mathbf{T}^n$, which would form a "bridge" from analysis on Ω_K^n to established theory on $\mathbf{T}^n$. The settling of Theorem 1 brings the state of analysis over Ω_K^n closer to matching what we understand on the hypercube, and we discuss it in detail next.

Chapter 3

The discretization inequality and its accoutrements

Aspects of optimality of Theorem 1, extensions, consequences, proof techniques, and related literature.

The discretization inequality (2.0.5) was first proved in [SVZ25] with a large (but dimension-free) constant, which was then improved by a different argument in [KSVZ24].

Theorem (Theorem 1 restated [SVZ25; KSVZ24])**.** *Let f be an n-variate analytic polynomial of individual degree at most $K-1$ and degree at most d. Then*

$$\|f\|_{\mathbf{T}^n} \le \mathcal{O}(\log K)^d \|f\|_{\Omega_K^n}.$$

One might wonder how important the specific grid of points Ω_K^n is for getting a dimension-free estimate. In [Bec+25] we extended the proof in [KSVZ24] to show the answer is actually "not so much."

Theorem 14 ([Bec+25])**.** *Let $f : \mathbf{D}^n \to \mathbf{C}$ be an analytic polynomial of degree d and individual degree $K-1$. Let $\mathbf{X} = \prod_{j=1}^n X_j \subset \mathbf{D}^n$ such that for all $j, |X_j| = K$. Put*

$$\eta = \min_j \min_{x,y \in X_j} |x-y|.$$

Then

$$\|f\|_{\mathbf{D}^n} \le (C_{\eta,K})^d \|f\|_{\mathbf{X}}, \tag{3.0.1}$$

for a universal constant $C_{\eta,K}$ independent of n. As above, when $\mathbf{X} = \Omega_K^n$ we may take $C_{\eta,K} \le \mathcal{O}(\log K)$.

With Theorem 1 in hand, both the Bohnenblust–Hille inequality and the boundedness of level-ℓ Fourier projections on Ω_K^n become one-line arguments.

Corollary 15. *Let f be a polynomial of degree d and individual degree $K-1$. Then*

$$\|\widehat{f}\|_{\frac{2d}{d+1}} \le \mathcal{O}(\log K)^d \cdot \mathrm{BH}_{\mathbf{T}^n}^{\le d} \cdot \|f\|_{\Omega_K^n},$$

where $\mathrm{BH}_{\mathbf{T}^n}^{\le d} \le C^{\sqrt{d \log d}}$ is the best constant in the polytorus BH inequality.

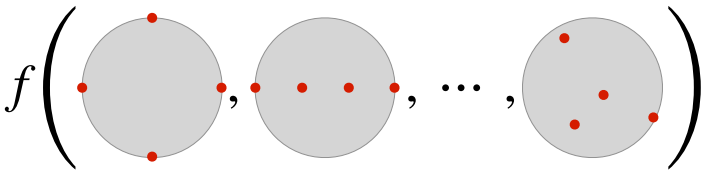

Figure 3.1: A visual depiction of Theorem 14. As long as there are K well-spaced red points in each of the coordinate discs, one may control the supremum norm of any individual-degree-$(K-1)$ polynomial f over the polydisk $\mathbf{D}^n$ by its absolute supremum over the finite grid of red points, in a dimension-free way.

Proof. Any such polynomial f has the same Fourier expansion with respect to the groups Ω_K^n and $\mathbf{T}^n$. Therefore,

$$\|\widehat{f}\|_{\frac{2d}{d+1}} \overset{\text{(Classical BH)}}{\leq} \mathrm{BH}_{\mathbf{T}^n}^{\leq d}\|f\|_{\mathbf{T}^n} \leq (\mathcal{O}(\log K))^d \mathrm{BH}_{\mathbf{T}^n}^{\leq d}\|f\|_{\Omega_K^n}. \qquad \square$$

Corollary 16. *With f as before let f_ℓ be its ℓ-homogeneous part. Then*

$$\|f_\ell\|_{\Omega_K^n} \leq \mathcal{O}(\log K)^d \|f\|_{\Omega_K^n}.$$

Proof. Again because f has the same Fourier expansion over $\mathbf{T}^n$ and Ω_K^n,

$$\|f_\ell\|_{\Omega_K^n} \leq \|f_\ell\|_{\mathbf{T}^n} \overset{(2.0.7)}{\leq} \|f\|_{\mathbf{T}^n} \leq \mathcal{O}(\log K)^d \|f\|_{\Omega_K^n}. \qquad \square$$

Remarks and refinements

We now describe in what senses Theorem 1 and its generalization Theorem 14 are optimal, what aspects we do not yet understand, and certain extensions. This discussion is from [Bec+25].

Sharp degree-dependence of the constant

The best constant in the comparison Theorem 14 has exponential dependence on d. For simplicity we will argue this with $Y_n = \Omega_K^n$. Consider the univariate inequality

$$\|f\|_{\mathbf{T}} \leq C(K)\|f\|_{\Omega_K} \tag{3.0.2}$$

for polynomials f with degree at most $K-1$, and where $C(K)$ is the best constant. A Lagrange interpolation argument shows $C(K) > 1$ for any $K \geq 3$ [SVZ24a, Appendix B]. Let g be any extremizer of this inequality and put $f(z) = \prod_{j=1}^{d/(K-1)} g(z_j)$, assuming $K-1$ divides d for simplicity. Then

$$\|f\|_{\mathbf{T}^n} = \big(C(K)\big)^{d/(K-1)}\|f\|_{Y_n} =: D(K)^d\|f\|_{Y_n}$$

which is exponential in d.

On the other hand, for this specific construction one may calculate that $D(K) > 1$ does not grow in K. It remains an interesting question to determine the optimal K-dependence of the constant $C(\eta, K)$ in (3.0.1).

Question 1. *What is the optimal dependence on K in the constant in* (3.0.1) *of Theorem 14?*

On the cardinality of the sampling set

Minimal cardinality of product sampling sets

The cardinality of Y_n in Theorem 14 is optimal in the following sense. If the sampling sets are of the *product* form $Y_n = \prod_{1 \le j \le n} Z_j$ and one expects (3.0.1) to hold at least for polynomials of individual degree at most $K-1$, then each Z_j must have cardinality at least K and so Y_n contains at least K^n points. If $|Y_n|$ were any smaller, there would exist a j such that Z_j has at most $K-1$ points, and no such inequality can hold: the polynomial $f_j(z) := \prod_{\xi \in Z_j} (z_j - \xi)$ is of degree at most $K-1$ but $\|f_j\|_{Y_n} = 0$.

Contrapositively, if the sampling set has a product set structure $Y_n = \prod_{j=1}^n Z_j$ with $|Z_j| = K$, then the individual degree constraint on f is of course necessary.

Improvements for non-product sets

On the other hand, if we remove the product constraint on our sampling set, we can do better. Indeed, in [Bec+25, §4] we show that one may take a "small" part of $\prod_{j=1}^n Z_j$ and retain a dimension-free constant.

Theorem 17. *Let $K \ge 2$. Consider $\{Z_j \subset \mathbf{D}\}_{j \ge 1}$ a sequence of sets such that for all $j \ge 1$ we have $|Z_j| = K$ and*

$$\eta := \min_{1 \le j \le n} \min_{z \ne z' \in Z_j} |z - z'| > 0.$$

Then for any $\varepsilon > 0$ one can find a subset Y_n of size at most $C_1(d, \varepsilon)(1+\varepsilon)^n$ contained in $\prod_{j=1}^n Z_j$ such that for any analytic polynomial $f : \mathbf{D}^n \to \mathbf{C}$ of degree d and individual degree $K-1$,

$$\|f\|_{\mathbf{D}^n} \le C_2(d, K, \eta, \varepsilon) \|f\|_{Y_n}, \tag{3.0.3}$$

where

$$C_1(d, \varepsilon) \le \left(\frac{d}{\varepsilon}\right)^{100d}.$$

Furthermore, if $0 < \varepsilon \leq 1/2$*, then*

$$C_2(d, K, \eta, \varepsilon) \leq \exp\left(C_3(d, K, \eta)\left(\varepsilon^{-1}\log(\varepsilon^{-1})\right)^d\right),$$

for some constant $C_3(d, K, \eta)$ *depending on* $d, K,$ *and* η*.*

We do not print the argument in this thesis as it is similar in spirit to the proof of Theorem 14 (and the [KSVZ24] proof of Theorem 1): instead of beginning with a univariate interpolation formula (as is done in the proof of Theorem 14), we begin with a multivariate interpolation on a small number of coordinates, and then apply the probabilistic tensorization ideas from Theorem 14 to get to n coordinates.

Sharp dependence of sampling set cardinality on dimension n

On the third hand, the cardinality of Y_n cannot be sub-exponential in n. It suffices to prove this for $d = 1$; the general $d \geq 1$ case follows immediately by definitions and the case $d = 1$.

Theorem 18. *Suppose that the uniform norm discretization* (3.0.1) *holds for sampling set* $V_n \subset \mathbf{D}^n$ *with* $d = 1, K = 2$*; that is,*

$$\|f\|_{\mathbf{D}^n} \leq C_0 \|f\|_{V_n}$$

holds for all multi-affine polynomials f *of degree* 1 *with* $C_0 > 1$ *being the best constant, then* $|V_n| \geq C_1 C_2^n$*, where* $C_1 > 0$ *is universal and* $C_2 > 1$ *depends on* C_0*.*

See Section 18 for the proof of Theorem 18.

Uniform separation

In Theorem 14, the constant $C(K, \eta)^d$ grows with η^{-1}, where η is the minimum pairwise distance between points in the Z_j's. In fact, this is unavoidable; uniform separation (*i.e.,* independence of η from n) is *required* to retain the dimension-freeness of the inequality of Theorem 14. This is easy to see in one dimension, nor can it be avoided in higher dimensions, as illustrated by the following example.

Suppose $Y_1, Y_2, \ldots$ is a sequence of sets with $Y_n \subset \mathbf{D}^n$ and $c(n)$ is a sequence of coordinates; that is, $1 \leq c(n) \leq n$ for all n. Let $P_n = \{z_{c(n)} : z \in Y_n\} \subset \mathbf{D}$

be the projection of Y_n onto the $c(n)$-th coordinate. Suppose $|P_n| = K$ for all n and

$$\lim_{n\to\infty} \min_{z\neq z'\in P_n} |z - z'| = 0\,.$$

For each P_n we may then choose a subset $A_n \subset P_n$ with $|A_n| = K-1$ and an excluded point ζ_n^* such that $P_n = A_n \sqcup \zeta_n^*$ and

$$\min_{\zeta\in A_n} |\zeta_n^* - \zeta| \leq \varepsilon_n,$$

where $\lim_{n\to\infty} \varepsilon_n = 0$. Now consider the sequence of polynomials

$$f_n(z) := \prod_{\zeta\in A_n}(z_{c(n)} - \zeta)\,.$$

Certainly $\|f_n\|_{\mathbf{D}^n}$ is at least as large as any of its coefficients, so we have $\|f_n\|_{\mathbf{D}^n} \geq 1$. On the other hand, $\|f_n\|_{Y_n}$ is very small: $f_n(z) = 0$ for all $z \in Y_n$ except those z with $z_{c(n)} = \zeta_n^*$, and for these z we have

$$|f_n(\ldots, \zeta_n^*, \ldots)| = \left|\prod_{\zeta\in A_n}(\zeta_n^* - \zeta)\right| \leq \varepsilon_n \cdot 2^{K-2},$$

which tends to 0 as $n \to \infty$. Therefore no dimension-free uniform norm discretization like Theorem 3 is available for such $(Y_n)_{n\geq 1}$.

3.1 Theorem 1, sketch of Proof I: via Fourier multipliers

The historically-first proof of Theorem 1 appears in [SVZ25], and is based on Fourier multiplier techniques which split f into certain special polynomials that are more amenable to direct argument. We sketch it now.

Let us begin with the following idea, which is a somewhat familiar starting place for these sorts of comparisons: suppose we could find a scaled-down copy of the circle, $\eta\mathbf{T} \subset \mathbf{D}$ for $\eta = \eta(K) > 0$, such that for any $z \in \eta\mathbf{T}$, there is a probability distribution μ_z on Ω_K with

$$z^m = \mathop{\mathbb{E}}_{\xi\sim\mu_z} \xi^m \qquad \text{for all } m = 0, 1, \ldots, K-1\,. \tag{3.1.1}$$

Then we would essentially be done: for example, for f a d-homogeneous polynomial the argument would be as simple as

$$\|f\|_{\mathbf{T}^n} = \eta^{-d}\|f\|_{\eta\mathbf{T}^n} = \eta^{-d} \max_{z\in\eta\mathbf{T}^n} \left| \mathop{\mathbb{E}}_{\xi_j\sim\mu_{z_j}:j\in[n]} f(\xi)\right| \leq \eta^{-d}\|f\|_{\Omega_K^n}\,,$$

because the μ_{z_j}'s are supported on Ω_K. (The non-homogeneous case is not much worse.)

Unfortunately, this argument does not work for us because such a μ_z does not always exist in our setting. Examining the constraints on μ_z in (3.1.1), it turns out that to satisfy them all we need about twice as many degrees of freedom in μ_z as is afforded by making $\mathrm{supp}(\mu) = \Omega_K$. We are able to recover the property (3.1.1) only by letting μ_z be supported on Ω_{2K} instead. So this approach gets us to

$$\|f\|_{\mathbf{T}^n} \lesssim_{d,K} \|f\|_{\Omega_{2K}^n}, \tag{3.1.2}$$

and it remains to compare $\|f\|_{\Omega_{2K}^n}$ with $\|f\|_{\Omega_K^n}$. Given the apparent saturation of degrees of freedom in the Ω_{2K}^n comparison (3.1.2), this second step required a fully novel argument and came as a surprise to the authors.

The second comparison is achieved as follows. Fix a z^* that maximizes $|f|$ over Ω_{2K}. For some coordinates j, we already have $z_j^* \in \Omega_K$ (the goal domain), so let us forget about those. By a change of variables of the form $z_j \mapsto \xi z_j$ for $\xi \in \Omega_K$, we can assume the remaining coordinates in z^* are all equal to $\sqrt{\omega} := \exp(\pi i/K)$.

Thus it actually would suffice to prove the two-point comparison

$$|g(\sqrt{\omega}, \dots, \sqrt{\omega})| \overset{?}{\lesssim}_{d,K} |g(1, \dots, 1)| \tag{3.1.3}$$

for all g of degree at most d and individual degree at most $K-1$ (noting that the right-hand side is of course at most $\|g\|_{\Omega_K^n}$). When g is homogeneous (3.1.3) is an equality: just multiply by the appropriate root of unity. However, it is entirely unclear how to reduce the general case of (3.1.3) to the homogeneous case. With g_k the k-homogeneous part of g, a typical approach might be

$$\begin{aligned}
|g(\sqrt{\omega}, \dots, \sqrt{\omega})| &\le \sum_{k=0}^{d} |g_k(\sqrt{\omega}, \dots, \sqrt{\omega})| \\
&= \sum_{k=0}^{d} |g_k(1, \dots, 1)| \\
&\le \sum_{k=0}^{d} \|g_k\|_{\Omega_K^n} \\
&\overset{?}{\lesssim}_{d,K} \|g\|_{\Omega_K^n},
\end{aligned}$$

which would work provided the level-k Fourier projections are bounded independent of dimension: $\|g_k\|_{\Omega_K^n} \lesssim_{d,K} \|g\|_{\Omega_K^n}$.

But before this work that question was also open and seemed just as difficult as the discretization inequality. The central technical contribution of [SVZ25]

is to introduce a special class of Fourier projections (denoted by $\mathfrak{D}$ in the proof) that allow us to write $g = \sum_j h_j$ for a small number of polynomials h_j which...

i. Are bounded by g independent of dimension: $\|h_j\|_{\Omega_K^n} \lesssim_{k,D} \|g\|_{\Omega_K^n}$, and

ii. Are *not* homogeneous but, by a very serendipitous identity for "half-roots" of unity (4.2.13), nevertheless satisfy

$$|h_j(\sqrt{\omega}, \dots, \sqrt{\omega})| = |h_j(1, \dots, 1)|.$$

The proof is completed by replacing the g_k's in the previous display with our h_j's.

A byproduct of this technique is a rich class of Fourier multipliers (actually, projections onto certain fine-grained classes of monomials) that are bounded independent of dimension. The class has only an implicit description for general K, but when K is prime we may leverage some results from transcendental number theory to get the following characterization.

For an $S \subset \{0, 1, \dots, K-1\}^n$ we denote by f_S the S-part of f:

$$f_S := \sum_{\alpha \in S} \widehat{f}(\alpha) z^\alpha \,.$$

Theorem (Theorem 2 restated: Bounded Fourier projections, prime $K \geq 3$)**.** *Suppose K is an odd prime and let S be a maximal subset of $\{0, 1, \dots, K-1\}^n$ such that for all $\alpha, \beta \in S$:*

- *Degrees are equal: $\sum_{j=1}^n \alpha_j = \sum_{j=1}^n \beta_j$.*
- *Individual degree symmetry: there is a bijection $\pi : [n] \to [n]$ such that for all $j \in [n]$, $\alpha_j = \beta_{\pi(j)}$ or $\alpha_j = K - \beta_{\pi(j)}$.*

Then for any n-variate analytic polynomial f of degree at most d and individual degree at most $K-1$, the S-part of f, $f_S(x) := \sum_{\alpha \in S} \widehat{f}(\alpha) z^\alpha$, satisfies

$$\|f_S\|_{\Omega_K^n} \lesssim_{d,K} \|f\|_{\Omega_K^n} \,.$$

Theorem 2 and related techniques do not seem to follow from the later argument in [KSVZ24; Bec+25] and can be considered as one of the main contributions of the work [SVZ25].

3.2 Theorem 1, sketch of Proof II: via interpolation

The second proof, presented in [KSVZ24] for the domain Ω_K^n and in [Bec+25] for more general domains, takes a probabilistic view of interpolation.

In one coordinate, polynomial interpolation (Lagrange interpolation) admits a probabilistic interpretation of the form

$$f(z) = D \cdot \mathbb{E}[\boldsymbol{R} \cdot f(\boldsymbol{W})], \tag{3.2.1}$$

where $D = D(K) > 1$ is a constant and $\boldsymbol{R}$ and $\boldsymbol{W}$ are correlated random variables taking values in Ω_4 and Ω_K respectively. Repeating (3.2.1) coordinatewise gives the identity

$$f(z) = D^n \, \mathbb{E}\Big[\Big(\textstyle\prod_{j=1}^n \boldsymbol{R}_j\Big) f\big(\boldsymbol{W}_1, \dots, \boldsymbol{W}_n\big)\Big], \tag{3.2.2}$$

which immediately implies a discretization inequality of the desired form, except with exponential dependence on n. The idea is to notice that (3.2.2) is an expectation over n-many independent pairs of variables $(\boldsymbol{R}_j, \boldsymbol{W}_j)$, while f is of bounded total degree d and thus is not very "aware" that $\big((\boldsymbol{R}_1, \boldsymbol{W}_1), \dots, (\boldsymbol{R}_n, \boldsymbol{W}_n)\big)$ is a product distribution.

It turns out that by introducing certain correlations among the $\boldsymbol{W}_j$'s, we can reduce the power on D at the expense of picking up an error term:

$$f(z) = D^d \, \mathbb{E}\Big[\Big(\textstyle\prod_{j=1}^d S_j\Big) f\big(\widetilde{W}_1, \dots, \widetilde{W}_n\big)\Big] \; + \; \mathrm{error}_{f,z}\,. \tag{3.2.3}$$

Here the S_j's are i.i.d. over Ω_4 and the $\widetilde{W}_j$'s are still over Ω_K, but now the joint distribution $(\widetilde{W}_1, \dots, \widetilde{W}_n)$ has an intricate dependence structure. If we only had the first term we would be done of course, and with the right d-dependence in the constant. To remove the error term, we will use special features of the error's relationship to the introduced correlations. Specifically, the correlation construction actually defines a *family* of identities similar to (3.2.3) of the form

$$f(z) = D^m \, \mathbb{E}\Big[\Big(\textstyle\prod_{j=1}^m S_j^{(m)}\Big) f\big(\widetilde{W}_1^{(m)}, \dots, \widetilde{W}_n^{(m)}\big)\Big] \; + \; \mathrm{error}_{f,z}\big(\tfrac{1}{m}\big),$$

for any integer $m > 1$, and where $\mathrm{error}_{f,z}$ is a fixed polynomial in $1/m$ of degree at most $d-1$ and with no constant term. These properties imply there is an affine combination of these identities for $m = 1, \dots, d$ that eliminates the error term:

$$f(z) \;=\; \sum_{m=d}^{2d-1} a_m f(z) \;=\; \sum_{m=d}^{2d-1} a_m D^m \, \mathbb{E}\Big[\Big(\textstyle\prod_{j=1}^m S_j^{(m)}\Big) f\big(\widetilde{W}_1^{(m)}, \dots, \widetilde{W}_n^{(m)}\big)\Big], \tag{3.2.4}$$

and where the absolute sum of the a_m's is suitably small. Placing $|\cdot|$'s in the right spots finishes the theorem.

A new interpolation formula

All the expectations in (3.2.4) are over finite probability spaces, so we actually have proved a new interpolation formula:

Theorem 19. *Let f, X, η be as in Theorem 14. Then for any $z \in \mathbf{D}^n$, there exist explicit coefficients $\{a_x^{(z)}\}_{x \in \Omega_K^n}$ such that $\sum_x |a_x^{(z)}| \le (C_{\eta,K})^d$ and*

$$f(z) = \sum_{x \in X} a_x^{(z)} f(x) \,. \tag{3.2.5}$$

Comparing (3.2.5) to classical multivariate polynomial interpolation formulas, we obtain coefficients with dimension-free absolute sum at the expense of sampling more points than strictly necessary. As a result the linear combination (3.2.5) is not unique, and it is interesting to understand whether this flexibility can lead to sharpenings of Theorem 1.

In the full proof, the identity (3.2.5) appears in detail as Equation (5.1.9). We hope this interpolation formula can have future applications and offer as a first example usage a short proof of a dimension-free discretization inequality for L^p norms, $1 \le p < \infty$, as we describe next.

L^p discretization

Let $L^p(\mathbf{T}^n)$ and $L^p(\Omega_K^n)$ denote the L^p-space with respect to the uniform probability measures on $\mathbf{T}^n$ and Ω_K^n, respectively. When $Y_n = \Omega_K^n$, one way to prove a dimension-free L^p discretization inequality for $p < \infty$ would be to use hypercontractivity over $\mathbf{T}^n$ [Jan97] and over Ω_K^n [Wei80; JPPP17]. Hypercontractivity is a workhorse of high dimensional analysis [BGL14; Hu17] and implies dimension-free L^2-L^p comparisons for bounded-degree polynomials when $1 \le p < \infty$ (see [ODo14, Chapter 9.5] and [Def+11; DGMS19, Chapter 8.4] for discussion). For example, with $2 \le p < \infty$, and f a degree-d function on Ω_K^n, the argument is

$$\|f\|_{L^p(\mathbf{T}^n)} \lesssim_{d,p} \|f\|_{L^2(\mathbf{T}^n)} = \|f\|_{L^2(\Omega_K^n)} \le \|f\|_{L^p(\Omega_K^n)} \,,$$

where hypercontractivity on the polytorus is applied in the first inequality. (*N.b.* that this hypercontractivity argument does not work for $p = \infty$.)

In Section 6.1 we show a proof that avoids hypercontractivity altogether by making use of the interpolation formula (3.2.5) (or more concretely, Equation (5.1.9)). The main result of Section 6.1 is the following.

Theorem 20. *Let $d, n \geq 1, K \geq 2$. Let $1 \leq p \leq \infty$. Then for each polynomial $f : \mathbf{T}^n \to \mathbf{C}$ of degree at most d and individual degree at most $K-1$, the following holds:*

$$\|f\|_{L^p(\mathbf{T}^n)} \leq C(d,K)\|f\|_{L^p(\Omega_K^n)}$$

with $C(d,K) \leq d(C_1 \log(K) + C_2)^d$ with universal $C_1, C_2 > 0$.

We remark that the constant in the inequality of Theorem 20 is independent from p but dependent on d, so has a different character from Marcinkiewicz–Zygmund inequalities, where the constant depends on p but is typically required to be independent from the total degree d for $1 < p < \infty$.

The proof combines the interpolation formula with group-invariance of the uniform measure on Ω_K^n.

3.3 Relationships to other literatures

Theorem 14 in the context of approximation theory

The context and history for discretization inequalities in approximation theory begins in dimension $n = 1$. For $1 < p < \infty$, the so-called *Marcinkiewicz–Zygmund inequality* [Zyg02, Chapter X, Theorem (7.5)] states that for all analytic polynomials f of degree at most $K-1$, one has

$$C_p^{-1} \cdot \tfrac{1}{K} \sum_{z \in \Omega_K} |f(z)|^p \;\leq\; \int_{\mathbf{T}} |f(z)|^p dz \;\leq\; C_p \cdot \tfrac{1}{K} \sum_{z \in \Omega_K} |f(z)|^p. \tag{3.3.1}$$

Here C_p is a constant depending only on p (independent of K), and $\mathbf{T} = \{z \in \mathbf{C} : |z| = 1\}$ denotes the unit circle.

The inequality (3.3.1) is an example of a *discretization of the L^p-norm*, and integral norm inequalities of this type are usually called *Marcinkiewicz-type theorems*. At the endpoint $p = \infty$ this type of inequality is often called a *Bernstein-type theorem* or a *discretization of the uniform norm* (see [Ber31; Ber32] and [Zyg02, Chapter X, Theorem (7.28)]). In our notation, the $p = \infty$ endpoint of (3.3.1) reads

$$\|f\|_{\Omega_K} \leq \|f\|_{\mathbf{T}} \leq C(K)\|f\|_{\Omega_K}. \tag{3.3.2}$$

In this $p = \infty$ case (and unlike $1 < p < \infty$) we emphasize the right-hand side inequality cannot have constant independent of K. See for example [OS07, Theorem 5].

We refer to surveys [DPTT19; KKLT22] and references therein for more historical background about norm discretizations. Bernstein-type discretization

theorems also have some overlap with *discrete Remez-type inequalities* as we discuss below.

Now let us return to the high-dimensional case, where Theorem 1 can be understood as a Bernstein-type discretization inequality for bounded-degree multivariate polynomials in many dimensions n. In this setting there are intricate tradeoffs between the cardinality (and structure) of the sampling set, the constant in the discretization inequality, and the function space to which the estimate applies. Recently there has been very important progress on understanding the minimum cardinality of sampling sets when one demands a universal constant (independent from any notion of degree or dimension) in the inequality.

In [KKT23], Kashin, Konyagin, and Temlyakov give a discretization of the uniform norm that applies to any N-dimensional subspace of continuous functions on a compact subset of $\mathbf{R}^n$, achieving a universal constant 2 with a sampling set of cardinality 9^N. Moreover, as the authors show, this is essentially the best possible sampling set cardinality for a Bernstein-type discretization inequality at this level of generality.

On the other hand, much smaller sampling sets—again for L^∞ norm discretizations with universal constants—can be had when one fixes the function space to be polynomials of degree at most d. A significant recent work along these lines is [DP24]. Here Dai and Prymak resolved an important problem of Kroó [Kro11] in real approximation theory by showing there are discretizations of the uniform norm for n-variate polynomials of (total) degree at most d over any convex domain in $\mathbf{R}^n$, with universal constant 2 and a sampling set of cardinality $C_n d^n$ in our notation.[1] When degree d is large in comparison to dimension n, this cardinality $C_n d^n$ matches the dimension of the set of such polynomials, and is therefore the best possible. (*N.b.* our primary interest is in the opposite of their regime, $d \ll n$.)

Our motivating application to functions on Ω_K^n—that is, to obtain a comparison

$$\|f\|_{\mathbf{T}^n} \lesssim_{d,K} \|f\|_{\Omega_K^n}$$

for analytic polynomials f of individual degree at most $K-1$ and total degree at most d—is in some ways more demanding, and in other ways more relaxed, than the parameter regime traditionally considered in approximation theory. On the

[1] *N.b.*, in the notation of [DP24] it will be $C_d n^d$ where they used d for the dimension and n for the degree.

one hand, we are restricted by the sampling set Ω_K^n, which is a fixed product set of small cardinality. Existing Bernstein-type estimates do not seem to apply in the parameter regime $K < d$, which is the setting dictated by applications to harmonic analysis in the high-dimensional realm of combinatorics, computer science, and learning theory. On the other hand, we do not require an absolute constant; indeed, as we discuss in below, dependence of the constant on degree d is unavoidable under these constraints.

Remez-type inequalities in many dimensions

Consider J a finite interval in $\mathbf{R}$ and a subset $E \subset J$ with positive Lebesgue measure $\mu(E) > 0$. Let $f : \mathbf{R} \to \mathbf{R}$ be a real polynomial of degree at most d. The classical Remez inequality [Rem36] states that

$$\max_{x \in J} |f(x)| \leq \left(\frac{4\mu(J)}{\mu(E)} \right)^d \max_{x \in E} |f(x)|. \tag{3.3.3}$$

Despite a large literature extending (3.3.3), we are not aware of any direct multi-dimensional generalizations that are dimension-free. Multi-dimensional versions of the Remez inequality are considered in the papers of Brudnyi and Ganzburg [BG73], Ganzburg [Gan17], Kroó and Schmidt [KS97] but they are not at all dimension-free: it is instructive to take a look at inequality (23) in [KS97] and see how the estimates blow up with dimension (called m in [KS97]). If one abandons the L^∞ norm on the left-hand side of (3.3.3) then something can be said; there are distribution function inequalities for volumes of level sets of polynomials that are dimension-free, see [Fra09; NSV02; NSV03]. But those are distribution function estimates, not L^∞ estimates. Some other related results include Nazarov's extension [Naz93] of Turán's inequality [Tur53], as well as more generalizations [Fon06; FY13].

The lack of a dimension-free multi-dimensional Remez inequality of the form (3.3.3) is not surprising: there is no hope for such an inequality phrased in terms of $\mu(E)$ for any positive-measure $E \subseteq J$. This can already be seen when J is a unit ball in $\mathbf{R}^n$ and $f_n(x) = 1 - \sum_{j=1}^n x_j^2$. For large n, most of the volume of the ball is concentrated in a neighborhood of the unit sphere where f_n is very small.

However, this observation does not preclude the existence of *certain* sets E giving multi-dimensional analogues of (3.3.3) that are dimension-free. Indeed, Lundin [Lun85], and later Aron–Beauzamy–Enflo [ABE93] and Klimek [Kli95], show this is possible in certain cases of (J, E) with convex E, such as for

bounded-degree polynomials over the polydisk $J = \mathbf{D}^n$ and the real cube $E = [-1, 1]^n$. As an explicit example, with the prevailing notation, Klimek [Kli95] showed that for n-variate analytic polynomials of degree d, we have the comparison $\|f\|_{\mathbf{D}^n} \leq (1 + \sqrt{2})^d \|f\|_{[-1,1]^n}$.

On the other hand, it was not at all clear when dimension-free Remez inequalities should exist in non-convex settings like $J = \mathbf{T}^n$ and $E \subset \mathbf{T}^n$. The arguments in [Lun85; ABE93; Kli95] make essential use of the convexity of the testing set E and do not seem to suitably generalize. In comparison, for our application to functions on products of cyclic groups $f : \Omega_K^n \to \mathbf{C}$, we have no choice but to use the non-convex grid Ω_K^n as our E.

That our E is discrete and indeed finite is another interesting feature. Remez-type estimates with discrete E were known before; notably, Yomdin [Yom11] (see also [BY16]) identifies a geometric invariant which directly replaces the Lebesgue measure in (3.3.3) and is positive for certain finite sets E—though the comparison is not dimension-free.

Chapter 4

Proof I: Discretization by Fourier Multipliers

Recall our goal is to prove the following.

Theorem 21 (Theorem 1, implicit constant)**.** *Let $f : \mathbf{T}^n \to \mathbf{C}$ be an analytic polynomial of degree at most d and individual degree at most $K-1$. Then*

$$\|f\|_{\mathbf{T}^n} \lesssim_{d,K} \|f\|_{\Omega_K^n}.$$

Our approach has two steps:

$$\text{Step 1.} \quad \|f\|_{\mathbf{T}^n} \lesssim_{d,K} \|f\|_{\Omega_{2K}^n}, \quad \text{and}$$

$$\text{Step 2.} \quad \|f\|_{\Omega_{2K}^n} \lesssim_{d,K} \|f\|_{\Omega_K^n}.$$

4.1 Step 1

Proposition 1 (Torus bounded by Ω_{2K})**.** *Let $d, n \geq 1, K \geq 3$. Let $f : \mathbf{T}^n \to \mathbf{C}$ be an analytic polynomial of degree at most d and individual degree at most $K-1$. Then*

$$\|f\|_{\mathbf{T}^n} \leq C_K^d \|f\|_{\Omega_{2K}^n},$$

where $C_K \geq 1$ is a universal constant depending on K only.

To prove this proposition, we need the following lemma.

Lemma 22. *Fix $K \geq 3$. There exists $\varepsilon = \varepsilon(K) \in (0,1)$ such that, for all $z \in \mathbf{C}$ with $|z| \leq \varepsilon$, one can find a probability measure μ_z on Ω_{2K} such that*

$$z^m = \mathop{\mathbb{E}}_{\xi \sim \mu_z} \xi^m, \qquad \forall \quad 0 \leq m \leq K-1. \tag{4.1.1}$$

Proof. Put $\theta = 2\pi/2K = \pi/K$ and $\omega = \omega_{2K} = e^{\mathbf{i}\theta}$. Fix a $z \in \mathbf{C}$. Finding a probability measure μ_z on Ω_{2K} satisfying (4.1.1) is equivalent to solving

$$\begin{cases} \sum_{k=0}^{2K-1} p_k = 1 \\ \sum_{k=0}^{2K-1} p_k \cos(km\theta) = \Re z^m \quad 1 \leq m \leq K-1 \\ \sum_{k=0}^{2K-1} p_k \sin(km\theta) = \Im z^m \quad 1 \leq m \leq K-1 \end{cases} \tag{4.1.2}$$

with non-negative $p_k = \mu_z(\{\omega^k\})$ for $k = 0, 1, \ldots, 2K-1$. Note that the p_k's are non-negative and thus real.

For this, it is sufficient to find a solution $\vec{p} = \vec{p}_z$ to $D_K \vec{p} = \vec{v}_z$ with each entry of $\vec{p} = (p_0, \dots, p_{2K-1})^\top$ being non-negative. Here D_K is a $2K \times 2K$ real matrix given by

$$D_K = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \cos(\theta) & \cos(2\theta) & \cdots & \cos((2K-1)\theta) \\ \vdots & \vdots & \vdots & \vdots & \\ 1 & \cos(K\theta) & \cos(2K\theta) & \cdots & \cos((2K-1)K\theta) \\ 1 & \sin(\theta) & \sin(2\theta) & \cdots & \sin((2K-1)\theta) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & \sin((K-1)\theta) & \sin(2(K-1)\theta) & \cdots & \sin((2K-1)(K-1)\theta) \end{bmatrix},$$

and $\vec{v}_z = (1, \Re z, \dots, \Re z^{K-1}, \Re z^K, \Im z, \dots, \Im z^{K-1})^\top \in \mathbf{R}^{2K}$. Note that (4.1.2) does not require the $(K+1)$-th row

$$(1, \cos(K\theta), \cos(2K\theta), \dots, \cos((2K-1)K\theta)) \tag{4.1.3}$$

of D_K.

The matrix D_K is non-singular. To see this, take any

$$\vec{x} = (x_0, x_1, \dots, x_{2K-1})^\top \in \mathbf{R}^{2K}$$

such that $D_K \vec{x} = \vec{0}$. Then

$$\sum_{k=0}^{2K-1} (\omega^k)^m x_k = 0, \qquad 0 \le m \le K. \tag{4.1.4}$$

This is immediate for $0 \le m \le K-1$ by definition, and $m = K$ case follows from the "additional" row (4.1.3) together with the fact that $\sin(kK\theta) = 0, 0 \le k \le 2K-1$. Conjugating (4.1.4), we get

$$\sum_{k=0}^{2K-1} (\omega^k)^m x_k = 0, \qquad K \le m \le 2K.$$

Altogether, we have

$$\sum_{k=0}^{2K-1} (\omega^k)^m x_k = 0, \qquad 0 \le m \le 2K-1,$$

that is, $V\vec{x} = \vec{0}$, where $V = V_K = [\omega^{jk}]_{0 \le j,k \le 2K-1}$ is a $2K \times 2K$ Vandermonde matrix given by $(1, \omega, \dots, \omega^{2K-1})$. Since V has determinant

$$\det(V) = \prod_{0 \le j < k \le 2K-1} (\omega^j - \omega^k) \neq 0\,,$$

we get $\vec{x} = \vec{0}$. So D_K is non-singular.

Therefore, for any $z \in \mathbf{C}$, the solution to (4.1.2), thus to (4.1.1), is given by

$$\vec{p}_z = \Big(p_0(z), p_1(z), \dots, p_{2K-1}(z)\Big) = D_K^{-1}\vec{v}_z \in \mathbf{R}^{2K}.$$

Notice one more thing about the rows of D_K. As

$$\sum_{k=0}^{2K-1} (\omega^k)^m = 0, \qquad m = 1, 2, \dots, 2K-1\,,$$

we have automatically that vector $\vec{p}_* := \left(\frac{1}{2K}, \dots, \frac{1}{2K}\right) \in \mathbf{R}^{2K}$ gives

$$D_K\vec{p}_* = (1, 0, 0, \dots, 0)^T =: \vec{v}_*\,.$$

For any k-by-k matrix A denote

$$\|A\|_{\infty\to\infty} := \sup_{\vec{0}\neq\vec{v}\in\mathbf{R}^k} \frac{\|A\vec{v}\|_\infty}{\|\vec{v}\|_\infty}.$$

So with $\vec{p_*} := D_K^{-1}\vec{v_*}$ we have

$$\begin{aligned}
\|\vec{p}_z - \vec{p}_*\|_\infty &\le \|D_K^{-1}\|_{\infty\to\infty}\|\vec{v}_z - \vec{v}_*\|_\infty \\
&= \|D_K^{-1}\|_{\infty\to\infty} \max\Big\{\max_{1\le k\le K} |\Re z^k|, \max_{1\le k\le K-1} |\Im z^k|\Big\} \\
&\le \|D_K^{-1}\|_{\infty\to\infty} \max\{|z|, |z|^K\}.
\end{aligned}$$

That is,

$$\max_{0\le j\le 2K-1}\left|p_j(z) - \frac{1}{2K}\right| \le \|D_K^{-1}\|_{\infty\to\infty} \max\{|z|, |z|^K\}.$$

Since $D_K^{-1}\vec{v}_* = \vec{p}_*$, we have $\|D_K^{-1}\|_{\infty\to\infty} \ge 2K$. Put

$$\varepsilon_* := \frac{1}{2K\|D_K^{-1}\|_{\infty\to\infty}} \in \left(0, \frac{1}{(2K)^2}\right].$$

Thus whenever $|z| < \varepsilon_* < 1$, we have

$$\max_{0\le j\le 2K-1}\left|p_j(z) - \frac{1}{2K}\right| \le |z|\|D_K^{-1}\|_{\infty\to\infty} \le \varepsilon_*\|D_K^{-1}\|_{\infty\to\infty} \le \frac{1}{2K},$$

so in particular $p_j(z) \ge 0$ for all $0 \le j \le 2K-1$. □

Now we are ready to prove Proposition 1.

Proof of Proposition 1. Let ε_* be as in Lemma 22. With a view towards applying the lemma we begin by relating $\sup|f|$ over the polytorus to $\sup|f|$ over a scaled copy. Recalling that the homogeneous parts f_k of f are trivially bounded by f over the torus: $\|f_k\|_{\mathbf{T}^n} \leq \|f\|_{\mathbf{T}^n}$ (a standard Cauchy estimate). Thus we have

$$\begin{aligned}
\|f\|_{\mathbf{T}^n} &\leq \sum_{k=0}^{d} \|f_k\|_{\mathbf{T}^n} \\
&= \sum_{k=0}^{d} \varepsilon_*^{-k} \sup_{z\in\mathbf{T}^n} |f_k(\varepsilon_* z)| \\
&\leq \sum_{k=0}^{d} \varepsilon_*^{-k} \sup_{z\in\mathbf{T}^n} |f(\varepsilon_* z)| \\
&\leq (d+1)\varepsilon_*^{-d} \sup_{z\in\mathbf{T}^n} |f(\varepsilon_* z)| \\
&= (d+1)\varepsilon_*^{-d} \|f\|_{(\varepsilon_*\mathbf{T})^n} \,. \qquad (4.1.5)
\end{aligned}$$

Let $z = (z_1, \ldots, z_n) \in (\varepsilon_*\mathbf{T})^n$. Then for each coordinate $j = 1, 2, \ldots, n$ there exists by Lemma 22 a probability distribution $\mu_j = \mu_j(z)$ on Ω_{2K} for which $\mathbb{E}_{\xi_j\sim\mu_j}[\xi_j^k] = z_j^k$ for all $0 \leq k \leq K-1$. With $\mu = \mu(z) := \mu_1 \times \cdots \times \mu_n$, this implies for a monomial ξ^α with multi-index $\alpha \in \{0, 1, \ldots, K-1\}^n$, $\mathbb{E}_{\xi\sim\mu(z)}[\xi^\alpha] = z^\alpha$, or more generally by linearity $\mathbb{E}_{\xi\sim\mu(z)}[f(\xi)] = f(z)$ for $z \in (\varepsilon_*\mathbf{T})^n$ and f under consideration. So

$$\sup_{z\in(\varepsilon_*\mathbf{T})^n} |f(z)| = \sup_{z\in(\varepsilon_*\mathbf{T})^n} \left| \mathop{\mathbb{E}}_{\xi\sim\mu(z)} f(\xi) \right| \leq \sup_{z\in(\varepsilon_*\mathbf{T})^n} \mathop{\mathbb{E}}_{\xi\sim\mu(z)} |f(\xi)| \leq \|f\|_{\Omega_{2K}^n}. \quad (4.1.6)$$

Combining observations (4.1.5) and (4.1.6) we conclude

$$\|f\|_{\mathbf{T}^n} \leq (d+1)\varepsilon_*^{-d}\|f\|_{(\varepsilon_*\mathbf{T})^n} \leq (d+1)\varepsilon_*^{-d}\|f\|_{\Omega_{2K}^n} \leq C_K^d \|f\|_{\Omega_{2K}^n}. \qquad \square$$

The last inequality follows from the fact that ε_* depends only on K.

4.2 Step 2

Now we turn to Step 2's estimate,

$$\|f\|_{\Omega_{2K}^n} \lesssim_{d,K} \|f\|_{\Omega_K^n} \,. \qquad (4.2.1)$$

We will find it useful to rephrase this question as one about the boundedness at the single point

$$f(\sqrt{\omega}, \ldots, \sqrt{\omega}) =: f(\sqrt{\omega}) \,.$$

Here and in what follows, $\omega := \omega_K = e^{2\pi \mathbf{i}/K}$, and $\sqrt{\omega}$ will be used as shorthand to denote the root $e^{\pi \mathbf{i}/K}$. It turns out the following proposition is enough to give (4.2.1).

Proposition 2. *Let $d, n \geq 1, K \geq 3$. Let $f : \mathbf{T}^n \to \mathbf{C}$ be an analytic polynomial of degree at most d and individual degree at most $K-1$. Then*

$$|f(\sqrt{\omega})| \lesssim_{d,K} \|f\|_{\Omega_K^n} .$$

To explain why Proposition 2 suffices for Step 2, let us finish the proof of Theorem 21 given Proposition 1 and assuming Proposition 2.

Proof of Theorem 21. Fix a $z^* \in \arg\max_{z \in \Omega_{2K}^n} |f(z)|$. Then there exist $w = (w_1, \dots, w_n) \in \Omega_K^n$ and $y^* \in \{1, \sqrt{\omega}\}^n$ such that

$$w_j y_j^* = z_j^*, \quad j \in [n] ,$$

where $[n] := \{1, 2, \dots, n\}$. Define $\widetilde{f} : \mathbf{T}^n \to \mathbf{C}$ by

$$\widetilde{f}(z) = f(w_1 z_1, w_2 z_2, \dots, w_n z_n) .$$

We therefore have

$$|\widetilde{f}(y^*)| = \|f\|_{\Omega_{2K}^n} \quad \text{and} \tag{4.2.2}$$

$$\|\widetilde{f}\|_{\Omega_K^n} = \|f\|_{\Omega_K^n} . \tag{4.2.3}$$

Equation (4.2.2) holds by the definition of y^*, and (4.2.3) holds by the group property of Ω_K (recall $w \in \Omega_K^n$).

Now let $S = \{j : y_j^* = \sqrt{\omega}\}$ and $m = |S|$. Let $\pi : S \to [m]$ be any bijection. Define the "selector" function $s_{y^*} : \mathbf{T}^m \to \mathbf{T}^n$ coordinate-wise by

$$\Big(s_{y^*}(z)\Big)_j = \begin{cases} y_j^* & \text{if } j \notin S \\ z_{\pi(j)} & \text{if } j \in S . \end{cases}$$

Finally, define $g : \mathbf{T}^m \to \mathbf{C}$ by

$$g(z) = \widetilde{f}(s_{y^*}(z)) .$$

Then we observe that g is analytic with degree at most d and individual degree at most $K-1$, and

$$|g(\sqrt{\omega}, \sqrt{\omega}, \dots, \sqrt{\omega})| = |\widetilde{f}(y^*)| \overset{(4.2.2)}{=} \|f\|_{\Omega_{2K}^n} \tag{4.2.4}$$

$$\|g\|_{\Omega_K^m} \leq \|\widetilde{f}\|_{\Omega_K^n} \overset{(4.2.3)}{=} \|f\|_{\Omega_K^n} , \tag{4.2.5}$$

with the inequality holding because we are optimizing over a subset of points. From (4.2.4) and (4.2.5) we see Theorem 21 would follow if we could prove

$$|g(\sqrt{\omega}, \sqrt{\omega}, \ldots, \sqrt{\omega})| \lesssim_{d,K} \|g\|_{\Omega_K^m},$$

independent of $m \geq 1$. This is precisely Proposition 2. □

The proof of Proposition 2 is the subject of the rest of this subsection. Our approach is to split f into parts $f = \sum_j g_j$ such that each part g_j has the properties A and B:

$$\|f\|_{\Omega_K^n} \overset{\text{Property A}}{\gtrsim_{d,K}} \|g_j\|_{\Omega_K^n} \overset{\text{Property B}}{\gtrsim_{d,K}} |g_j(\sqrt{\omega})|. \tag{4.2.6}$$

Such splitting gives

$$|f(\sqrt{\omega})| \leq \sum_j |g_j(\sqrt{\omega})| \lesssim_{d,K} \sum_j \|g_j\|_{\Omega_K^n} \lesssim_{d,K} \sum_j \|f\|_{\Omega_K^n}.$$

So as long as the number of g_j's is independent of n such a splitting with Properties A and B entails the result.

We will split f via an operator that was first employed to prove the Bohnenblust–Hille inequality for cyclic groups [SVZ24a]. We will only need the basic version of the operator here; a generalized version is considered in [SVZ24a]. Recall that any polynomial $f : \Omega_K^n \to \mathbf{C}$ has the Fourier expansion

$$f(z) = \sum_{\alpha \in \{0,1,\ldots,K-1\}^n} a_\alpha z^\alpha.$$

Recall the support of a monomial z^α is $\operatorname{supp}(\alpha) := \{j : \alpha_j \neq 0\}$, and the support size $|\operatorname{supp}(\alpha)|$ refers to the cardinality of $\operatorname{supp}(\alpha)$.

Definition 1 (Maximum support pseudoprojection)**.** *For any multi-index $\alpha \in \{0, 1, \ldots, K-1\}^n$ define the factor*

$$\tau_\alpha = \prod_{j : \alpha_j \neq 0} (1 - \omega^{\alpha_j}).$$

For any polynomial on Ω_K^n with the largest support size $\ell \geq 0$

$$f(z) = \sum_{|\operatorname{supp}(\alpha)| \leq \ell} a_\alpha z^\alpha,$$

we define $\mathfrak{D} f : \Omega_K^n \to \mathbf{C}$ via

$$\mathfrak{D} f(z) = \sum_{|\operatorname{supp}(\alpha)| = \ell} \tau_\alpha \, a_\alpha z^\alpha.$$

The operator $\mathfrak{D}$ can be considered a Fourier multiplier, and this somewhat technical definition is motivated by the following key property, the $L^\infty \to L^\infty$ boundedness when restricted to certain polynomials.

Lemma 23 (Boundedness of maximum support pseudoprojection)**.** *Let $f : \Omega_K^n \to \mathbf{C}$ be a polynomial and ℓ be the maximum support size of monomials in f. Then*

$$\|\mathfrak{D} f\|_{\Omega_K^n} \le (2+2\sqrt{2})^\ell \|f\|_{\Omega_K^n}. \tag{4.2.7}$$

The proof of Lemma 23 is given in [SVZ24a]. We repeat it here in a slightly simplified form for convenience.

Proof. Let $\omega = e^{\frac{2\pi \mathbf{i}}{K}}$. Consider the operator G:

$$G(f)(x) = f\Big(\frac{1+\omega}{2} + \frac{1-\omega}{2}x_1, \dots, \frac{1+\omega}{2} + \frac{1-\omega}{2}x_n\Big), \qquad x \in \Omega_2^n$$

that maps any function $f : \{1,\omega\}^n \subset \Omega_K^n \to \mathbf{C}$ to a function $G(f) : \Omega_2^n \to \mathbf{C}$. Then by definition

$$\|f\|_{\Omega_K^n} \ge \|f\|_{\{1,\omega\}^n} = \|G(f)\|_{\Omega_2^n}. \tag{4.2.8}$$

Fix $m \le \ell$. For any α we denote

$$m_k(\alpha) := |\{j : \alpha_j = k\}|, \qquad 0 \le k \le K-1.$$

Then for α with $|\mathrm{supp}(\alpha)| = m$, we have

$$m_1(\alpha) + \cdots + m_{K-1}(\alpha) = |\mathrm{supp}(\alpha)| = m.$$

For $z \in \{1,\omega\}^n$ with $z_j = \frac{1+\omega}{2} + \frac{1-\omega}{2}x_j, x_j = \pm 1$, note that

$$z_j^{\alpha_j} = \Big(\frac{1+\omega}{2} + \frac{1-\omega}{2}x_j\Big)^{\alpha_j} = \frac{1+\omega^{\alpha_j}}{2} + \frac{1-\omega^{\alpha_j}}{2}x_j\,.$$

So for any $A \subset [n]$ with $|A| = m$, and for each α with $\mathrm{supp}(\alpha) = A$, we have for $z \in \{1,\omega\}^n$:

$$\begin{aligned} z^\alpha &= \prod_{j:\alpha_j \neq 0} z_j^{\alpha_j} \\ &= \prod_{j:\alpha_j \neq 0} \Big(\frac{1+\omega^{\alpha_j}}{2} + \frac{1-\omega^{\alpha_j}}{2}x_j\Big) \\ &= \prod_{j:\alpha_j \neq 0} \Big(\frac{1-\omega^{\alpha_j}}{2}\Big) \cdot x^A + \cdots \\ &= 2^{-m}\tau_\alpha x^A + \cdots \end{aligned}$$

where $x^A := \prod_{j\in A} x_j$ is of degree $|A| = m$ while $\cdots$ is of degree $< m$. Then for $f(z) = \sum_{|\mathrm{supp}(\alpha)|\le \ell} a_\alpha z^\alpha$ we have

$$G(f)(x) = \sum_{m\le \ell} \frac{1}{2^m} \sum_{|A|=m} \left(\sum_{\mathrm{supp}(\alpha)=A} \tau_\alpha a_\alpha \right) x^A + \cdots, \qquad x \in \Omega_2^n.$$

Again, for each $m \le \ell$, $\cdots$ is some polynomial of degree $< m$. So $G(f)$ is of degree $\le \ell$ and the ℓ-homogeneous part is nothing but

$$\frac{1}{2^\ell} \sum_{|A|=\ell} \left(\sum_{\mathrm{supp}(\alpha)=A} \tau_\alpha a_\alpha \right) x^A.$$

Consider the projection operator Q that maps any polynomial on Ω_2^n onto its highest level homogeneous part; *i.e.*, for any polynomial $g : \Omega_2^n \to \mathbf{C}$ with $\deg(g) = m$ we denote $Q(g)$ its m-homogeneous part. Then we just showed that

$$Q(G(f))(x) = \frac{1}{2^\ell} \sum_{|A|=\ell} \left(\sum_{\mathrm{supp}(\alpha)=A} \tau_\alpha a_\alpha \right) x^A. \tag{4.2.9}$$

It is known that [DMP18, Lemma 1 (iv)] for any polynomial $g : \Omega_2^n \to \mathbf{C}$ of degree at most $d > 0$ and g_m its m-homogeneous part, $m \le d$, we have the estimate

$$\|g_m\|_{\Omega_2^n} \le (1+\sqrt{2})^d \|g\|_{\Omega_2^n}.$$

Applying this estimate to $G(f)$ and combining the result with (4.2.8), we have

$$\|Q(G(f))\|_{\Omega_2^n} \le (\sqrt{2}+1)^\ell \|G(f)\|_{\Omega_2^n} \le (1+\sqrt{2})^\ell \|f\|_{\Omega_K^n}$$

and thus by (4.2.9)

$$\left\| \sum_{|A|=\ell} \left(\sum_{\mathrm{supp}(\alpha)=A} \tau_\alpha a_\alpha \right) x^A \right\|_{\Omega_2^n} \le (2+2\sqrt{2})^\ell \|f\|_{\Omega_K^n}.$$

The function on the left-hand side is almost $\mathfrak{D} f$. Observe that Ω_K^n is a group, so we have

$$\sup_{z,\xi\in\Omega_K^n} \left| \sum_\alpha a_\alpha z^\alpha \xi^\alpha \right| = \sup_{z\in\Omega_K^n} \left| \sum_\alpha a_\alpha z^\alpha \right|.$$

Thus we have actually shown

$$\sup_{z\in\Omega_K^n, x\in\Omega_2^n} \left| \sum_{|A|=\ell} \left(\sum_{\mathrm{supp}(\alpha)=A} \tau_\alpha a_\alpha z^\alpha \right) x^A \right| \le (2+2\sqrt{2})^\ell \|f\|_{\Omega_K^n}.$$

Setting $x = \vec{1}$ gives (4.2.7). □

Note that $\mathfrak{D}f$ is exactly the part of f composed of monomials of maximum support size, except where the coefficients a_α have picked up the factor τ_α. The relationships among the τ_α's can be intricate: while in general they are different for distinct α's, this is not always true. Consider the case of $K = 3$ and the two monomials

$$z^\beta := z_1^2 z_2 z_3 z_4 z_5 z_6 z_7 z_8, \quad z^{\beta'} := z_1^2 z_2^2 z_3^2 z_4^2 z_5^2 z_6^2 z_7^2 z_8 \,.$$

Then

$$\tau_\beta \;=\; (1-\omega)^7(1-\omega^2) \;=\; (1-\omega)(1-\omega^2)^7 \;=\; \tau_{\beta'},$$

which follows from the identity $(1-\omega)^6 = (1-\omega^2)^6$ for $\omega = e^{2\pi \mathbf{i}/3}$.

Understanding precisely when $\tau_\alpha = \tau_\beta$ seems to be a formidable task in transcendental number theory. When K is prime there is a relatively simple characterization (see Section 4.3) but for composite K the situation is much less clear. Nevertheless, it turns out that for the purposes of Theorem 21 we do not need a full understanding. Indeed, our g_j's shall be defined according to the τ's.

Definition. *Two monomials z^α, z^β with associated multi-indexes*

$$\alpha, \beta \in \{0, 1 \ldots, K-1\}^n$$

are called inseparable *if $|\mathrm{supp}(\alpha)| = |\mathrm{supp}(\beta)|$ and $\tau_\alpha = \tau_\beta$. When m and m' are inseparable, we write $m \sim m'$.*

Inseparability is an equivalence relation among monomials. We may split any polynomial f into parts $f = \sum_j g_j$ according to this relation. That is, any two monomials in f are inseparable if and only if they belong to the same g_j. Call these g_j's the inseparable parts *of f.*

It is these inseparable parts that are our g_j's in (4.2.6). We shall formally check it later, but it is easy to see the number of inseparable parts is independent of n. We formulate and prove Properties A & B next.

Property A: Boundedness of inseparable parts

Repeated applications of the operator $\mathfrak{D}$ enable splitting into inseparable parts.

Proposition 3 (Property A)**.** *Fix $K \geq 3$ and $d \geq 1$. Suppose that $f : \Omega_K^n \to \mathbf{C}$ is a polynomial of degree at most d with maximum support size L. For $0 \leq \ell \leq L$*

let f_ℓ denote the part of f composed of monomials of support size ℓ, and let $g_{(\ell,1)},\dots,g_{(\ell,J_\ell)}$ be the inseparable parts of f_ℓ. Then there exists a universal constant $C_{d,K}$ independent of n and f such that for all $0 \le \ell \le L$ and $1 \le j \le J_\ell$,

$$\|g_{(\ell,j)}\|_{\Omega_K^n} \le C_{d,K}\|f\|_{\Omega_K^n}.$$

Proof. We first show the proposition for $g_{(L,j)}$, $1 \le j \le J_L$. Suppose that

$$f(z) = \sum_{\alpha:|\mathrm{supp}(\alpha)|\le L} a_\alpha z^\alpha.$$

Inductively, one obtains from Lemma 23 that for $1 \le k \le J_L$,

$$\begin{aligned} \mathfrak{D}^k f &= \sum_{|\mathrm{supp}(\alpha)|=L} \tau_\alpha^k a_\alpha z^\alpha \\ \text{with} \qquad \left\|\mathfrak{D}^k f\right\|_{\Omega_K^n} &\le (2+2\sqrt{2})^{kL}\|f\|_{\Omega_K^n}. \end{aligned} \tag{4.2.10}$$

By definition there are J_L distinct values of τ_α among the monomials of f_L; label them $c_1,\dots,c_{J_L}$. Then

$$f_L(z) = \sum_{|\mathrm{supp}(\alpha)|=L} a_\alpha z^\alpha = \sum_{1\le j\le J_L} g_{(L,j)}(z), \quad \text{and}$$

$$\mathfrak{D}^k f(z) = \sum_{|\mathrm{supp}(\alpha)|=L} \tau_\alpha^k a_\alpha z^\alpha = \sum_{1\le j\le J_L} c_j^k g_{(L,j)}(z), \qquad k \ge 1.$$

Let us confirm J_L is independent of n. Consider α with $|\mathrm{supp}(\alpha)| = L$. We may count the support size of α by binning coordinates according to their degree: $|\mathrm{supp}(\alpha)| = L$,

$$\sum_{1\le t\le K-1} |\{s \in [n] : \alpha_s = t\}| = L \le d,$$

so

$$\begin{aligned} J_L &\le |\{(m_1,\dots,m_{K-1}) \in \{0,\dots,L\}^{K-1} : m_1+\cdots+m_{K-1} = L\}| \\ &\le \binom{K-1+L-1}{L-1} \le (K+d)^d. \end{aligned} \tag{4.2.11}$$

According to (4.2.10), we have

$$\begin{pmatrix} \mathfrak{D}f \\ \mathfrak{D}^2 f \\ \vdots \\ \mathfrak{D}^{J_L} f \end{pmatrix} = \underbrace{\begin{pmatrix} c_1 & c_2 & \cdots & c_{J_L} \\ c_1^2 & c_2^2 & \cdots & c_{J_L}^2 \\ \vdots & \vdots & \ddots & \vdots \\ c_1^{J_L} & c_2^{J_L} & \cdots & c_{J_L}^{J_L} \end{pmatrix}}_{=:V_L} \begin{pmatrix} g_{(L,1)} \\ g_{(L,2)} \\ \vdots \\ g_{(L,J_L)} \end{pmatrix}.$$

The $J_L \times J_L$ modified Vandermonde matrix V_L has determinant

$$\det(V_L) = \left(\prod_{j=1}^{J_L} c_j\right)\left(\prod_{1\le s<t\le J_L}(c_s - c_t)\right).$$

Since the c_j's are distinct and nonzero we have $\det(V_L) \neq 0$. So V_L is invertible and in particular $g_{(L,j)}$ is the j^{th} entry of $V_L^{-1}(\mathfrak{D}^1 f, \dots, \mathfrak{D}^{J_L} f)^\top$. Letting $\eta^{(L,j)} = (\eta_k^{(L,j)})_{1\le k\le J_L}$ be the j^{th} row of V_L^{-1}, this means

$$g_{(L,j)} = \sum_{1\le k\le J_L} \eta_k^{(L,j)}\, \mathfrak{D}^k f.$$

As $\eta^{(L,j)}$ depends on d and K only, so for all $1 \le j \le J_L$,

$$\|g_{(L,j)}\|_{\Omega_K^n} \le \sum_{1\le k\le J_L} \left|\eta_k^{(L,j)}\right| \cdot \left\|\mathfrak{D}^k f\right\|_{\Omega_K^n} \le \|\eta^{(L,j)}\|_1 \left(2 + 2\sqrt{2}\right)^{J_L d} \|f\|_{\Omega_K^n}, \quad (4.2.12)$$

where we used (4.2.10) in the last inequality. The constant

$$\|\eta^{(L,j)}\|_1 (2 + 2\sqrt{2})^{J_L d} \le C(d,K) < \infty$$

for appropriate $C(d,K)$ that is dimension-free and depends only on d and K only. This finishes the proof for the inseparable parts in f_L.

We now repeat the argument on $f - f_L$ to obtain (4.2.12) for the inseparable parts of support size $L-1$. In particular, there are vectors $\eta^{(L-1,j)}$, $1 \le j \le J_{L-1}$ of dimension-free 1-norm with

$$\|g_{(L-1,j)}\|_{\Omega_K^n} \le C(d,K)\|\eta^{(L-1,j)}\|_1 \|f - f_L\|_{\Omega_K^n} \lesssim_{d,K} \|f - f_L\|_{\Omega_K^n}.$$

This can be further repeated to obtain for $0 \le \ell \le L$ and $1 \le j \le J_\ell$, the vectors $\eta^{(\ell,j)}$ with dimension-free 1-norm such that

$$\|g_{(\ell,j)}\|_{\Omega_K^n} \lesssim_{d,K} \left\| f - \sum_{\ell+1\le k\le L} f_k \right\|_{\Omega_K^n}.$$

It remains to relate $\|f - \sum_{\ell+1\le k\le L} f_k\|_{\Omega_K^n}$ to $\|f\|_{\Omega_K^n}$. Note that with V_L as originally defined, by considering $(1\,1\,\dots\,1)V_L^{-1}(\mathfrak{D}^1 f, \dots, \mathfrak{D}^{J_L} f)^\top$ we obtain a constant $D_L = D_L(d,K)$ independent of n for which

$$\|f_L\|_{\Omega_K^n} \le D_L \|f\|_{\Omega_K^n}.$$

This means

$$\|f - f_L\|_{\Omega_K^n} \le (1 + D_L)\|f\|_{\Omega_K^n}.$$

Notice the top-support part of $f - f_L$ is exactly f_{L-1}, so repeating the argument above on $f - f_L$ yields a constant $D_{L-1} = D_{L-1}(d, K)$ such that

$$\|f_{L-1}\|_{\Omega_K^n} \leq D_{L-1}\|f - f_L\|_{\Omega_K^n} \leq D_{L-1}(1+D_L)\|f\|_{\Omega_K^n} = (D_{L-1}+D_{L-1}D_L)\|f\|_{\Omega_K^n}\,.$$

Continuing, for $1 \leq \ell \leq L$ we find

$$\begin{aligned}\|f_{L-\ell}\|_{\Omega_K^n} &\leq D_{L-\ell}\|f - \sum_{L-\ell+1\leq k\leq L} f_k\|_{\Omega_K^n} \\ &\leq D_{L-\ell}(1 + D_{L-\ell+1})\|f - \sum_{L-\ell+2\leq k\leq L} f_k\|_{\Omega_K^n} \\ &\leq D_{L-\ell}\prod_{0\leq k\leq \ell-1}(1 + D_{L-k})\|f\|_{\Omega_K^n}.\end{aligned}$$

We have found for each ℓ-support-homogeneous part of f,

$$\|f_\ell\|_{\Omega_K^n} \lesssim_{d,K} \|f\|_{\Omega_K^n},$$

so we have $\|f - \sum_{\ell+1\leq k\leq L} f_k\|_{\Omega_K^n} \lesssim_{d,K} \|f\|_{\Omega_K^n}$ as well. □

Property B: Boundedness at $\sqrt{\omega}$ for inseparable parts

Here we argue $g(\sqrt{\omega})$ is bounded for inseparable g. Recall that $\omega = e^{\frac{2\pi\mathbf{i}}{K}}$ and $\sqrt{\omega} = e^{\frac{\pi\mathbf{i}}{K}}$.

Proposition 4 (Property B)**.** *If g is inseparable then $|g(\sqrt{\omega})| \leq \|g\|_{\Omega_K^n}$.*

Proof. We will need an identity for half-roots of unity. For $k = 1, \ldots, K-1$ we have

$$(\sqrt{\omega})^k = \mathbf{i}\frac{1-\omega^k}{|1-\omega^k|}\,, \tag{4.2.13}$$

following from the orthogonality of $(\sqrt{\omega})^k$ and $1 - \omega^k$ in the complex plane.

We claim that for two monomials m and m'

$$m \sim m' \implies m(\sqrt{\omega}) = m'(\sqrt{\omega})\,.$$

By definition $m \sim m'$ means m and m' have the same support size (call it ℓ) and

$$\textstyle\prod_{j:\alpha_j\neq 0}(1-\omega^{\alpha_j}) = \prod_{j:\beta_j\neq 0}(1-\omega^{\beta_j})\,.$$

Dividing both sides by the modulus and multiplying by $\mathbf{i}^\ell$ allows us to apply (4.2.13) to find

$$\textstyle\prod_{j:\alpha_j\neq 0}(\sqrt{\omega})^{\alpha_j} = \prod_{j:\beta_j\neq 0}(\sqrt{\omega})^{\beta_j}\,,$$

as desired.

Now let $\zeta = m(\sqrt{\omega}) \in \mathbf{T}$ for some monomial m in g. Then because ζ is independent of m, with $g = \sum_{\alpha \in S} a_\alpha z^\alpha$, we have $g(\sqrt{\omega}) = \zeta \sum_{\alpha \in S} a_\alpha$ and

$$|g(\sqrt{\omega})| = |\textstyle\sum_{\alpha \in S} a_\alpha| = |g(\vec{1})| \le \|g\|_{\Omega_K^n} . \qquad \square$$

We may now prove Proposition 2.

Proof of Proposition 2. Write $f = \sum_{0 \le \ell \le L} \sum_{1 \le j \le J_\ell} g_{(\ell,j)}$ in terms of inseparable parts, where $g_{(\ell,j)}, 1 \le j \le J_\ell, 0 \le \ell \le L$ are as in Proposition 3. Then by Propositions 3 (Property A) and 4 (Property B)

$$\begin{aligned} |f(\sqrt{\omega})| \;&\le\; \sum_{0 \le \ell \le L} \sum_{1 \le j \le J_\ell} |g_{(\ell,j)}(\sqrt{\omega})| \\ &\le\; \sum_{0 \le \ell \le L} \sum_{1 \le j \le J_\ell} \|g_{(\ell,j)}\|_{\Omega_K^n} && \text{(Property B)} \\ &\lesssim_{d,K} \|f\|_{\Omega_K^n} \sum_{0 \le \ell \le L} J_\ell \,. && \text{(Property A)} \end{aligned}$$

In view of (4.2.11) and $L \le d$, we obtain $|f(\sqrt{\omega})| \lesssim_{d,K} \|f\|_{\Omega_K^n}$. $\square$

4.3 Aside: characterizing inseparable parts for prime K

Although it is not required for the proof of Theorem 21, it is interesting to understand what are the parts g of f for which

$$\|g\|_{\Omega_K^n} \lesssim_{d,K} \|f\|_{\Omega_K^n} \tag{4.3.1}$$

via our Property A (Proposition 3)? Recall that (4.3.1) holds when g is a part of f containing all monomials in f from an equivalence class of the inseparability equivalence relation $\sim$.

Thus we are led to ask for a characterization of inseparability. It turns out that for prime K this can be done completely via connections to transcendental number theory including Baker's theorem [Bak22].

Proposition 5. *Suppose $K \ge 3$ is prime and $\alpha, \beta \in \{0, 1, \dots, K-1\}^n$. Then two monomials z^α, z^β are inseparable if and only if*

- *Support sizes are equal:* $|\mathrm{supp}(\alpha)| = |\mathrm{supp}(\beta)|$,
- *Degrees are equal mod $2K$:* $|\alpha| = |\beta| \mod 2K$,
- *Individual degree symmetry: there is a bijection $\pi : \mathrm{supp}(\alpha) \to \mathrm{supp}(\beta)$ such that for all $j \in \mathrm{supp}(\alpha)$, $\alpha_j = \beta_{\pi(j)}$ or $\alpha_j = K - \beta_{\pi(j)}$.*

Proof. Recall that by definition, two monomials z^α and z^β are inseparable if and only if they have the same support size and $\tau_\alpha = \tau_\beta$; that is,

$$\prod_{j:\alpha_j \neq 0} (1 - \omega^{\alpha_j}) = \prod_{j:\beta_j \neq 0} (1 - \omega^{\beta_j}),$$

where $\omega = e^{2\pi \mathbf{i}/K}$. For these quantities to be equal, their respective moduli and arguments must coincide.

To compare arguments, observe that for any multi-index $\sigma \in \{0, 1, \ldots, K-1\}^n$, by the identity (4.2.13) we may normalize τ_σ like so:

$$\frac{\tau_\sigma}{|\tau_\sigma|} = \mathbf{i}^{-|\mathrm{supp}(\sigma)|} \prod_{j=1}^{n} (\sqrt{\omega})^{\sigma_j} = \mathbf{i}^{-|\mathrm{supp}(\sigma)|} (\sqrt{\omega})^{|\sigma|},$$

where as before $\sqrt{\omega} = e^{\pi \mathbf{i}/K}$. It is given that $|\mathrm{supp}(\alpha)| = |\mathrm{supp}(\beta)|$, so the arguments of τ_α and τ_β are equal exactly when $(\sqrt{\omega})^{|\alpha|} = (\sqrt{\omega})^{|\beta|}$, or equivalently, $|\alpha| = |\beta| \bmod 2K$.

As for the moduli, using the identity $|1 - \omega^k| = 2\sin(k\pi/K)$ we find for any multi-index σ that

$$|\tau_\sigma| = \prod_{j:\sigma_j \neq 0} 2\sin(\sigma_j \pi/K) = \prod_{j:\sigma_j \neq 0} 2\sin(\min\{\sigma_j, K - \sigma_j\} \cdot \pi/K),$$

where the last step follows from the symmetry of sine about $\pi/2$.

So when are $|\tau_\alpha|$ and $|\tau_\beta|$ equal? By the last display, certainly they are the same if there is a bijection $\pi : \mathrm{supp}(\alpha) \to \mathrm{supp}(\beta)$ such that for all $j \in \mathrm{supp}(\alpha)$, $\alpha_j = \beta_{\pi(j)}$ or $\alpha_j = K - \beta_{\pi(j)}$. Is this the only time $|\tau_\alpha| = |\tau_\beta|$?

Returning to σ, define for $1 \leq k \leq (K-1)/2$ the quantity

$$\widehat{\sigma}(k) = |\{j : \sigma_j = k \text{ or } \sigma_j = K - k\}|.$$

Then

$$\log(|\tau_\sigma|) = \sum_{k=1}^{(K-1)/2} \widehat{\sigma}(k) \cdot \log(2\sin(k\pi/K)).$$

Therefore if the numbers

$$\{b_k := \log(2\sin(k\pi/K)), k = 1, \ldots, (K-1)/2\}$$

were linearly independent over $\mathbf{Z}$, the only way $|\tau_\alpha| = |\tau_\beta|$ is the existence of a bijection π as above.

Conveniently, the question of the linear independence of the b_k's has already appeared in a different context, concerning an approach of Livingston to resolve

a folklore conjecture of Erdös on the vanishing of certain Dirichlet L-series. It was answered in [Pat17] in the positive for $K \geq 3$ prime and in the negative for all composite $K \geq 4$ using several tools including Baker's celebrated theorem on linear forms in logarithms of algebraic numbers [Bak22]. □

Finally, recalling (*e.g.,* Corollary 16) that Theorem 21 implies $\|f_k\|_{\Omega_K^n} \lesssim_{d,K} \|f\|_{\Omega_K^n}$ for all k-homogeneous parts f_k of f, $0 \leq k \leq d$, we may conclude by Proposition 3:

Corollary 24. *Suppose K is an odd prime and let S be a maximal subset of $\{0,1,\ldots,K-1\}^n$ such that for all $\alpha, \beta \in S$:*

- *Support sizes are equal:* $|\mathrm{supp}(\alpha)| = |\mathrm{supp}(\beta)|$.
- *Degrees are equal:* $|\alpha| = |\beta|$.
- *Individual degree symmetry: there is a bijection $\pi : \mathrm{supp}(\alpha) \to \mathrm{supp}(\beta)$ such that for all $j \in \mathrm{supp}(\alpha)$, $\alpha_j = \beta_{\pi(j)}$ or $\alpha_j = K - \beta_{\pi(j)}$.*

Then for any n-variate analytic polynomial f of degree at most d and individual degree at most $K-1$, the S-part of f, i.e., *$f_S = \sum_{\alpha \in S} \widehat{f}(\alpha) z^\alpha$, satisfies:*

$$\|f_S\|_{\Omega_K^n} \lesssim_{d,K} \|f\|_{\Omega_K^n}.$$

Chapter 5

Proof II: Discretization by interpolation

Here we give a proof of Theorem 1 (or Theorem 14 for $X = \Omega_K^n$) via interpolation, follow [KSVZ24]. The more general case of Theorem 14, as proved in [Bec+25], is more complicated in notation only.

5.1 The proof

A natural approach to proving Theorem 1 is to consider a specific maximizer $z \in \mathbf{T}^n$ of $|f|$ and write $f(z)$ as a linear combination of evaluations of f at points in Ω_K^n. We might begin with this lemma for a single coordinate:

Lemma 25. *Suppose $z \in \mathbf{T}$. Then there exists $\boldsymbol{c} := (c_0, \dots, c_{K-1})$ such that for all $k = 0, 1, \dots, K-1$,*

$$z^k = \sum_{j=0}^{K-1} c_j (\omega^j)^k.$$

Moreover, $\|\boldsymbol{c}\|_1 \le B \log(K)$ for a universal constant B.

Proof. Let $\omega = \exp(2\pi i/K)$. The discrete Fourier transform (DFT) of the array $A = (1, z, \dots, z^{K-1})$ yields K complex numbers $\widetilde{c_0}, \dots, \widetilde{c_{K-1}}$ so that

$$z^k = A_k = \frac{1}{K} \sum_{j=0}^{K-1} \widetilde{c_j} \omega^{jk}$$

for all $k = 0, \dots, K-1$. Using $c_j := \frac{1}{K} \widetilde{c_j}$ we get

$$z^k = \sum_{j=0}^{K-1} c_j \omega^{jk}. \tag{5.1.1}$$

Recall the DFT coefficients are given by

$$\widetilde{c_j} = \sum_{k=0}^{K-1} A_k \omega^{-kj}.$$

Since $A_k = z^k$ we have

$$\widetilde{c_j} = \sum_{k=0}^{K-1} z^k \omega^{-kj} = \frac{1 - (z/\omega^j)^K}{1 - (z/\omega^j)}.$$

By the triangle inequality,

$$|\widetilde{c_j}| \le \min\left(K, \frac{2}{|\omega^j - z|}\right).$$

Using that the harmonic number $H_K = \sum_{k=1}^{K} 1/k$ satisfies $H_K \le \log(K) + 1$, it is elementary to see that we have

$$\sum_{j=0}^{K-1} |\widetilde{c_j}| \le BK \log K$$

for B a sufficiently large constant. That is,

$$\|c\|_1 = \sum_{j=0}^{K-1} |c_j| = \frac{1}{K} \sum_{j=0}^{K-1} |\widetilde{c_j}| \le B \log(K)\,. \qquad \square$$

In a single coordinate, Lemma 25 provides the desired inequality as follows. With $z \in \mathbb{T}$ a maximizer of $|f(z)|$ we have

$$\begin{aligned} \|f\|_{\mathbb{T}} = |f(z)| = |\sum_{k=0}^{d} a_k z^k| = |\sum_{k=0}^{d} \sum_{j=0}^{K-1} a_k c_j (\omega^j)^k| = |\sum_{j=0}^{K-1} c_j f(\omega^j)| \\ \le \|c\|_1 \|f\|_{\Omega_K} \le C \log(K) \|f\|_{\Omega_K}\,. \qquad \text{(Hölder)} \end{aligned}$$

However, in higher dimensions, repeating this argument coordinatewise introduces an exponential dependence on n. We circumvent this by taking a probabilistic view of the foregoing display: the sum over j can be interpreted as an expectation over a (complex-valued) measure on Ω_K. When it is repeated in several dimensions, this is like taking an expectation over n independent random variables. The key insight is that this independence is more than we need: by correlating the random variables, we "save on randomness" (which reduces the multiplicative constant) while retaining control of the error.

Lemma 26. *Let f be a degree-d n-variate polynomial and $z \in \mathbb{T}^n$. Then there is a univariate polynomial $p = p_{f,z}$ such that for any positive integer m there are (dependent) random variables $R, \boldsymbol{W}$ taking values in Ω_4 and Ω_K^n respectively such that*

$$f(z) = D^m \mathop{\mathbb{E}}_{R,\boldsymbol{W}} [R f(\boldsymbol{W})] + p(1/m)\,. \tag{5.1.2}$$

Moreover, p has $\deg(p) < d$ and zero constant term, and $D = D(K)$ is a universal constant depending on K only.

Lemma 26 is the crux of our argument and we are not aware of a similar statement in the literature. Theorem 1 follows quickly, though it is interesting to note that instead of clearing the error term by taking $m \to \infty$ (which would indeed make $p(1/m) \to 0$ but also send $D^m \to \infty$), we will end up using *algebraic* features of p (namely, low-degree-ness) to remove it. But first, the lemma:

Proof of Lemma 26. We will argue Lemma 26 for $f(z) = z^\alpha$, a monomial of degree at most d. The claim extends to general degree-d f by linearity.

We begin by examining a single coordinate with the aim of rewriting Lemma 25 in a probabilistic form. To that end, we first decouple the angle and magnitude information of the c_j's. Fix $z \in \mathbb{T}$ and let c_j be as in Lemma 25. We may write a decomposition

$$c_j = 1 \cdot c_j^{(0)} + i \cdot c_j^{(1)} + (-1) \cdot c_j^{(2)} + (-i) \cdot c_j^{(3)} = \sum_{s=0}^{3} i^s \cdot c_j^{(s)} \,,$$

with all $c_j^{(s)} \in \mathbf{R}^{\geq 0}$ and $c_j^{(0)} c_j^{(2)} = c_j^{(1)} c_j^{(3)} = 0$. This can be done for all j so that, with $C := B \log K$ from Lemma 25,

$$\|\boldsymbol{c}^{(s)}\|_1 \leq C \tag{5.1.3}$$

is satisfied for each $s \in \{0,1,2,3\}$, where $\boldsymbol{c}^{(s)} = (c_1^{(s)}, \ldots, c_n^{(s)})$. So we have for all $k = 0, \ldots, K-1$,

$$z^k = \sum_{j=0}^{K-1} \sum_{s=0}^{3} i^s \cdot c_j^{(s)} \cdot (\omega^j)^k \,.$$

We now rewrite the sum in Lemma 25 in probabilistic form.

Put $D = 4C + 1$ and define $r : [0, D] \to \mathbf{C}$ by

$$r(t) = \begin{cases} 1 & 0 \leq t \leq C+1, \\ i & C+1 < t \leq 2C+1, \\ -1 & 2C+1 < t \leq 3C+1, \\ -i & 3C+1 < t \leq 4C+1 = D \,. \end{cases}$$

Also define a piecewise-constant function $w : [0, D] \to \Omega_K$ as follows. Consider any collection of disjoint intervals $I_j^{(s)}, 0 \leq j \leq K-1, 0 \leq s \leq 3$ such that

$$I_j^{(s)} \subset [0, D], \qquad s \in \{0,1,2,3\}, j \in \{0, 1, \ldots, K-1\}$$

and for each s and j, $I_j^{(s)} \subseteq [sC+1, (s+1)C+1]$ and $|I_j^{(s)}| = c_j^{(s)}$. Disjointness is possible because for each s,

$$|[sC+1, (s+1)C+1]| = C \geq \textstyle\sum_{j=0}^{K-1} c_j^{(s)}$$

by (5.1.3). Now assign $w(I_j^{(s)}) = \omega^j$ and in the remaining region of $[0, D]$ (that is, $[0, D] \setminus \sqcup_{s,j} I_j^{(s)}$) let w take on each element of Ω_K with in equal amount (w.r.t. the uniform measure).

Claim 1. *Let T be sampled uniformly from $[0, D]$. Then for all $k = 0, 1, \ldots, K-1$,*

$$z^k = D \mathop{\mathbb{E}}_{T}[r(T) w(T)^k]\,. \tag{5.1.4}$$

Proof of Claim 1. Let us begin with $k = 0$, which simplifies to

$$D \mathop{\mathbb{E}}_{T}[r(T)] = 1\,. \tag{5.1.5}$$

This can be seen by direct computation:

$$\mathop{\mathbb{E}}_{T}[r(T)] = \frac{1}{D}(1 + 1 \cdot C + i \cdot C + (-1) \cdot C + (-i) \cdot C) = \frac{1}{D}\,.$$

For $k \geq 1$, consider the joint distribution of $(r(T), w(T)^k)$, whose product appears in (5.1.4). Fix $s \in \{0, 1, 2, 3\}$, and condition on $r(T) = i^s$. The conditional distribution of $w(T)$ has two parts. One part, corresponding to $\sqcup_j I_j^{(s)}$, has $w(T) = \omega^j$ over $I_j^{(s)}$ with the probability $\Pr[r(T) = i^s \wedge w(T) = \omega^j]$ equal to $c_j^{(s)}/D$, while the other has $w(T)$ uniformly distributed in Ω_K. The latter part contributes 0 to the expectation $\mathbb{E}[r(T) w(T)^k]$, since $\sum_{j=0}^{K-1} (\omega^j)^k = 0$ for $k = 1, 2, \ldots, K-1$. The former part contributes

$$i^s \cdot \sum_{j=0}^{K-1} \frac{c_j^{(s)}}{D} \omega^{jk}.$$

Summing this display over $s \in \{0, 1, 2, 3\}$ and rearranging, we get that

$$\mathbb{E}[r(T) w(T)^k] = \sum_{j=0}^{K-1} \frac{c_j}{D} (\omega^j)^k = \frac{1}{D} z^k,$$

completing proof of (5.1.4). ◊

We return to the multivariate setting. Fix $\boldsymbol{z} := (z_1, \ldots, z_n) \in \mathbb{T}^n$ and define the functions $w_1, \ldots, w_n$ corresponding to the above construction applied to

each coordinate $z_1, \dots, z_n$. If each coordinate were to receive a fresh copy of T this would lead to an identity with exponential constant:

$$z^\alpha = D^n \mathop{\mathbb{E}}_{\substack{T_\ell \overset{\text{iid}}{\sim} T, \\ 1 \le \ell \le n}} \left[\textstyle\prod_{\ell=1}^n r(T_\ell) w_\ell(T_\ell)^{\alpha_\ell} \right].$$

Instead, we consider only m independent copies of T: $T_1, \dots, T_m \overset{\text{iid}}{\sim} \mathcal{U}[0, D]$. The decision of which coordinates are integrated with respect to which T_ℓ is also made randomly, via a uniformly random function $P : [n] \to [m]$. We finally arrive at the definitions of R and $\boldsymbol{W}$:

$$\begin{aligned} R &:= \prod_{\ell=1}^m R_\ell \quad \text{with} \quad R_\ell := r(T_\ell), 1 \le \ell \le m \\ \boldsymbol{W} &:= \left(w_1\left(T_{P(1)}\right), w_2\left(T_{P(2)}\right), \dots, w_n\left(T_{P(n)}\right) \right) =: (W_1, \dots, W_n) \,. \end{aligned}$$

When P is injective on $\operatorname{supp}(\alpha)$, we easily achieve the smaller constant.

Claim 2. *Consider $m \ge |\operatorname{supp}(\alpha)|$. Then*

$$\mathop{\mathbb{E}}_{R, \boldsymbol{W}} [R \cdot \boldsymbol{W}^\alpha \mid P \textit{ is injective on } \operatorname{supp}(\alpha)] = D^{-m} z^\alpha \,.$$

Proof of Claim 2. It suffices to prove this for an arbitrary projection $\widetilde{P}$ that is injective on $\operatorname{supp}(\alpha)$. Consider the partition of $[n]$ given by $\widetilde{P}^{-1}([m])$ and write $S_\ell = \widetilde{P}^{-1}(\ell)$ for $\ell \in [m]$. By independence, the expectation splits over these S_ℓ's:

$$\mathop{\mathbb{E}}_{R, \boldsymbol{W}} [R \cdot \boldsymbol{W}^\alpha \mid P = \widetilde{P}] = \textstyle\prod_{\ell=1}^m \mathbb{E}\left[R_\ell \prod_{k \in S_\ell} W_k^{\alpha_k} \right]. \tag{5.1.6}$$

Because $\widetilde{P}$ is injective on $\operatorname{supp}(\alpha)$, every S_ℓ contains one or zero elements of $\operatorname{supp}(\alpha)$. By Claim 1, in the latter case we have

$$\mathbb{E}\left[R_\ell \textstyle\prod_{k \in S_\ell} W_k^{\alpha_k} \right] = \mathbb{E}[R_\ell] = \frac{1}{D},$$

and in the former case we have

$$\mathbb{E}\left[R_\ell \textstyle\prod_{k \in S_\ell} W_k^{\alpha_k} \right] = \mathbb{E}[R_\ell W_j^{\alpha_j}] = \frac{1}{D} z_j^{\alpha_j},$$

for the specific j for which $\{j\} = S_\ell \cap \operatorname{supp}(\alpha)$. Substituting these observations into (5.1.6) completes the argument. ◊

When P is not injective, we still have some control. Let $\mathcal{S} = \{S_j\}$ be a partition of $\operatorname{supp}(\alpha)$. We say P *induces* $\mathcal{S}$ if

$$\{P^{-1}(j) \cap \operatorname{supp}(\alpha) : j \in [m]\} = \mathcal{S} \,.$$

Claim 3. *For each partition $\mathcal{S}$ of* $\mathrm{supp}(\alpha)$ *there is a number $E(\mathcal{S})$ such that for all $m \geq |\mathcal{S}|$,*

$$\mathop{\mathbb{E}}_{R,\boldsymbol{W}}[R \cdot \boldsymbol{W}^{\alpha} \mid P \text{ induces } \mathcal{S}] = D^{-m}E(\mathcal{S}).$$

Proof of Claim 3. Condition again on a specific $\widetilde{P}$ that induces $\mathcal{S}$. There are two types of $\ell \in [m]$: those that $\boldsymbol{W}^{\alpha}$ depends on (that is, $\widetilde{P}(\mathrm{supp}(\alpha))$), and those that only R depends on. Call these sets $L = \widetilde{P}(\mathrm{supp}(\alpha))$ and L^c respectively. Then by independence of the T_ℓ's,

$$\begin{aligned}\mathop{\mathbb{E}}_{R,\boldsymbol{W}}[R \cdot \boldsymbol{W}^{\alpha} \mid P = \widetilde{P}] &= \mathop{\mathbb{E}}_{R,\boldsymbol{W}}[(\textstyle\prod_{\ell \in L^c} R_\ell)(\prod_{\ell \in L} R_\ell) \cdot \boldsymbol{W}^{\alpha} \mid P = \widetilde{P}] \\ &= D^{-m+|\mathcal{S}|} \underbrace{\mathop{\mathbb{E}}_{R,\boldsymbol{W}}[(\textstyle\prod_{\ell \in L} R_\ell) \cdot \boldsymbol{W}^{\alpha} \mid P = \widetilde{P}]}_{*}.\end{aligned}$$

We observe that the expectation $(*)$ does not depend on the specific $\widetilde{P}$ inducing $\mathcal{S}$, nor on m. Thus we may define $E(\mathcal{S})$ by setting $D^{-|\mathcal{S}|}E(\mathcal{S})$ equal to $(*)$. $\lozenge$

To summarize claims 2 and 3, we have that for all partitions $\mathcal{S}$ of $\mathrm{supp}(\alpha)$ there is a number $E(\mathcal{S})$ such that for all $m \geq |\mathcal{S}|$,

$$\mathbb{E}[R \cdot \boldsymbol{W}^{\alpha} | P \text{ induces } \mathcal{S}] = D^{-m}E(\mathcal{S}).$$

And using $\mathcal{S}^*$ to denote the singleton partition $\left\{\{j\}\right\}_{j \in \mathrm{supp}(\alpha)}$, we additionally have $E(\mathcal{S}^*) = z^{\alpha}$.

Now we consider the unconditional expectation $\mathbb{E}[R \cdot \boldsymbol{W}^{\alpha}]$ with $P \sim \mathcal{U}([m]^{[n]})$. Elementary combinatorics give that for all partitions $\mathcal{S}$ and all $m \geq 1$, with $s = |\mathcal{S}|$,

$$\begin{aligned}\Pr[P \text{ induces } \mathcal{S}] &= \frac{m(m-1)\cdots(m-s+1)}{m^{|\mathrm{supp}(\alpha)|}} \\ &=: \begin{cases} 1 + q_s\left(\frac{1}{m}\right) & \text{if } s = |\mathrm{supp}(\alpha)| \\ q_s\left(\frac{1}{m}\right) & \text{if } s < |\mathrm{supp}(\alpha)| \end{cases},\end{aligned}$$

for polynomials q_s with zero constant term and $\deg(q_s) < d$.

Of course P can only induce $\mathcal{S}$ for $|\mathcal{S}| \leq m$, so by the law of total probability,

$$\mathop{\mathbb{E}}_{R,\boldsymbol{W}}[R \cdot \boldsymbol{W}^{\alpha}] = \sum_{\mathcal{S}, |\mathcal{S}| \leq \min(m, |\mathrm{supp}(\alpha)|)} \mathbb{E}[R \cdot \boldsymbol{W}^{\alpha} \mid P \text{ induces } \mathcal{S}] \Pr[P \text{ induces } \mathcal{S}].$$

Consider first the case $m \geq |\mathrm{supp}(\alpha)|$. We obtain

$$\begin{aligned}
\mathop{\mathbb{E}}_{R,\boldsymbol{W}}[R \cdot \boldsymbol{W}^\alpha] &= \sum_{\mathcal{S}} \mathbb{E}[R \cdot \boldsymbol{W}^\alpha \mid P \text{ induces } \mathcal{S}] \Pr[P \text{ induces } \mathcal{S}] \\
&= D^{-m} E(\mathcal{S}^*) \left(1 + q_{|\mathrm{supp}(\alpha)|}\left(\tfrac{1}{m}\right)\right) \\
&\qquad + \sum_{\mathcal{S}, |\mathcal{S}| < |\mathrm{supp}(\alpha)|} D^{-m} E(\mathcal{S}) \cdot q_{|\mathcal{S}|}\left(\tfrac{1}{m}\right) \\
&= D^{-m}\left[z^\alpha + \sum_{\mathcal{S}} E(\mathcal{S}) \cdot q_{|\mathcal{S}|}\left(\tfrac{1}{m}\right)\right]. \qquad (5.1.7)
\end{aligned}$$

Now when $m < |\mathrm{supp}(\alpha)|$, we combine the fact that $\Pr[P \text{ induces } \mathcal{S}] = 0$ for $|\mathcal{S}| > m$ with the definition of q_s to see

$$\begin{aligned}
\mathop{\mathbb{E}}_{R,\boldsymbol{W}}[R \cdot \boldsymbol{W}^\alpha] &= 0 + \sum_{\mathcal{S}, |\mathcal{S}| \leq m} \mathbb{E}[R \cdot \boldsymbol{W}^\alpha \mid P \text{ induces } \mathcal{S}] \Pr[P \text{ induces } \mathcal{S}] \\
&= \sum_{\mathcal{S}, |\mathcal{S}| > m} D^{-m} E(\mathcal{S}) \Pr[P \text{ induces } \mathcal{S}] + \sum_{\mathcal{S}, |\mathcal{S}| \leq m} D^{-m} E(\mathcal{S}) \Pr[P \text{ induces } \mathcal{S}] \\
&= D^{-m} E(\mathcal{S}^*) \left(1 + q_{|\mathrm{supp}(\alpha)|}\left(\tfrac{1}{m}\right)\right) \\
&\qquad + \sum_{|\mathrm{supp}(\alpha)| > |\mathcal{S}| > m} D^{-m} E(\mathcal{S}) \cdot q_{|\mathcal{S}|}\left(\tfrac{1}{m}\right) \\
&\qquad + \sum_{m \geq |\mathcal{S}|} D^{-m} E(\mathcal{S}) \cdot q_{|\mathcal{S}|}\left(\tfrac{1}{m}\right) \\
&= D^{-m}\left[z^\alpha + \sum_{\mathcal{S}} E(\mathcal{S}) \cdot q_{|\mathcal{S}|}\left(\tfrac{1}{m}\right)\right]. \qquad (5.1.8)
\end{aligned}$$

Noting that (5.1.8) and (5.1.7) are identical, we rearrange to find

$$z^\alpha = D^m \, \mathbb{E}[R \cdot \boldsymbol{W}^\alpha] - \sum_{\mathcal{S}} E(\mathcal{S}) \cdot q_{|\mathcal{S}|}\left(\tfrac{1}{m}\right),$$

and the second part is in total a polynomial in $\frac{1}{m}$ with no constant term and degree $< d$. □

Finally, the error term $p\left(\frac{1}{m}\right)$ is removed by considering several values of m.

Proof of Theorem 1. Suppose there were some coefficients $a_m \in \mathbf{C}$ with $\sum_{m=1}^{d} a_m = 1$, so that for any polynomial p of degree $< d$ and $p(0) = 0$ we would have

$$\textstyle\sum_{m=1}^{d} a_m p(\frac{1}{m}) = 0.$$

We could then sum (5.1.2) for $m = 1, \dots, d$, weighted by a_m, and get

$$\begin{aligned} f(z) &= \sum_{m=1}^{d} a_m f(z) \\ &= \sum_{m=1}^{d} a_m D^m \, \mathbb{E}[R_m f(W_m)] + \sum_{m=1}^{d} a_m p\Big(\tfrac{1}{m}\Big) \\ &= \sum_{m=1}^{d} a_m D^m \, \mathbb{E}[R_m f(W_m)], \end{aligned} \tag{5.1.9}$$

where R_m, W_m are those $R, \boldsymbol{W}$ from (5.1.2) marked with explicit dependence on m.

Well, these coefficients a_m can be arranged, since the monomial vectors $(1/m^t)_{m=1,\dots,d}$ for $t = 0, \dots, d-1$ are linearly independent (Vandermonde). Since always $|R_m| \le 1$, we deduce

$$|f(z)| \le \sum_{m=1}^{d} |a_m D^m| \cdot \|f\|_{\Omega_K^n} \le \frac{\max_{m=1}^{d} |a_m|}{1 - 1/D} \cdot D^d \|f\|_{\Omega_K^n}.$$

An explicit formula for the a_m's is given by

$$a_m = (-1)^{d-m} \frac{m^d}{m!(d-m)!},$$

and it is evident that $\max_{m=1}^{d} |a_m| \le \exp(O(d))$, and specifically $\max_{m=1}^{d} |a_m| \le \exp(1.28d)$.

Without loss of generality, we may assume $D \ge 11$ and so $1/(1-1/D) \le 1.1$. We conclude

$$|f(z)| \le (4D)^d \|f\|_{\Omega_K^n} = (4B \log(K) + 4)^d \|f\|_{\Omega_K^n}. \qquad \square$$

5.2 Nonexistence of subexponential-cardinality meshes

Proof of Theorem 18. For any $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n) \in \{-1, 1\}^n$, consider the polynomials $f_\varepsilon(\boldsymbol{x}) = \sum_{j=1}^{n} \varepsilon_j x_j$ on $\{-1, 1\}^n$ of degree at most 1. Then by definition,

$$n = \|f_\varepsilon\|_{\{\pm 1\}^n} \le C \|f_\varepsilon\|_{V_n}.$$

In other words, we have for all $\varepsilon \in \{-1, 1\}^n$ that

$$\max_{v = (v_1, \dots, v_n) \in V_n} \left| \sum_{j=1}^{n} v_j \varepsilon_j \right| \ge 2\delta n \qquad \text{with} \qquad \delta = \frac{1}{2C} \in (0, \infty).$$

So we have the inclusion $\{-1, 1\}^n \subset \bigcup_{v \in V_n} \Lambda_v$, where

$$\Lambda_v := \left\{ \varepsilon \in \{-1, 1\}^n : |f_v(\varepsilon)| = \left| \sum_{j=1}^{n} v_j \varepsilon_j \right| \ge 2\delta n \right\}, \qquad v \in V_n.$$

For each $\boldsymbol{v} \in V_n$ and i.i.d. Bernoulli random variables $\varepsilon_1, \dots, \varepsilon_n$, we have by Hoeffding's inequality that

$$\begin{aligned}
|\Lambda_v| &= 2^n \Pr\left[\left|\sum_{j=1}^n v_j \varepsilon_j\right| \geq 2\delta n\right] \\
&\leq 2^n \Pr\left[\left|\sum_{j=1}^n \Re(v_j) \varepsilon_j\right| \geq \delta n\right] + 2^n \Pr\left[\left|\sum_{j=1}^n \Im(v_j) \varepsilon_j\right| \geq \delta n\right] \\
&\leq 2^{n+1} \exp\left(-\frac{\delta^2 n^2}{2\|\boldsymbol{a}\|_2^2}\right) + 2^{n+1} \exp\left(-\frac{\delta^2 n^2}{2\|\boldsymbol{b}\|_2^2}\right),
\end{aligned}$$

where $\boldsymbol{a} = \Re \boldsymbol{v}$ and $\boldsymbol{b} = \Im \boldsymbol{v}$ are real vectors. Recalling that $\boldsymbol{v} \in V_n \subset \mathbf{D}^n$, we have $\|\boldsymbol{a}\|_2^2 \leq n$ and $\|\boldsymbol{b}\|_2^2 \leq n$. Therefore,

$$|\Lambda_v| \leq 2^{n+2} \exp\left(-\delta^2 n/2\right) = 4\left(2e^{-\delta^2/2}\right)^n.$$

All combined, we just proved

$$2^n = |\{-1, 1\}^n| \leq \sum_{v \in V_n} |\Lambda_v| \leq 4|V_n|\left(2e^{-\delta^2/2}\right)^n.$$

This gives the bound

$$|V_n| \geq \frac{1}{4} e^{\frac{\delta^2 n}{2}},$$

as desired. □

Chapter 6

Applications

6.1 L^p discretization inequalities

Theorem 27. *Let $1 \le p \le \infty$ and $f : \Omega_K^n \to \mathbf{C}$ of degree at most d. Then*

$$\|f\|_{L^p(\mathbf{T}^n)} \le d\mathcal{O}(\log K)^d \|f\|_{L^p(\Omega_K^n)}.$$

Proof. Beginning with the interpolation formula (5.1.9), it is evident that for any $\boldsymbol{\xi} \in \Omega_N^n$,

$$f(\boldsymbol{\xi} \odot \boldsymbol{z}) = \sum_{m=1}^{d} a_m D^m \, \mathbb{E}[R_m f(\boldsymbol{\xi} \odot W_m)],$$

where $\odot$ denotes coordinatewise multiplication. Thus by Jensen we have

$$\begin{aligned} |f(\boldsymbol{\xi} \odot \boldsymbol{z})|^p &\le d^p \cdot \frac{1}{d} \sum_{m=1}^{d} |a_m D^m|^p \, \mathbb{E}[|f(\boldsymbol{\xi} \odot W_m)|^p] \\ &\le d^p \mathcal{O}(\log K)^{dp} \cdot \tfrac{1}{d} \sum_{m=1}^{d} \mathbb{E}[|f(\boldsymbol{\xi} \odot W_m)|^p]. \end{aligned}$$

Now consider $\boldsymbol{\xi}$ sampled uniformly from Ω_N^n. For any fixed value of W_m, the distribution $\boldsymbol{\xi} \odot W_m$ is also uniform on Ω_N^n, so we have

$$\mathop{\mathbb{E}}_{\boldsymbol{\xi} \sim \Omega_N^n} [|f(\boldsymbol{\xi} \odot \boldsymbol{z})|^p] \le d^p \mathcal{O}(\log K)^{dp} \mathop{\mathbb{E}}_{\boldsymbol{\xi} \sim \Omega_N^n} [|f(\boldsymbol{\xi})|^p].$$

Let $A = e^{[0,1) \cdot 2\pi i/N}$ be the arc on $\mathbf{T}$ from the first to the second N^{th} root of unity. Observe that the random variable

$$\xi z \quad \text{with} \quad z \sim A, \xi \sim \Omega_N$$

is distributed uniformly on $\mathbf{T}$, and accordingly the random variable

$$\boldsymbol{\xi} \odot \boldsymbol{z} \quad \text{with} \quad \boldsymbol{z} \sim A^n, \boldsymbol{\xi} \sim \Omega_N^n$$

is distributed uniformly on $\mathbf{T}^n$. Therefore,

$$\begin{aligned} \mathop{\mathbb{E}}_{\boldsymbol{z} \sim \mathbf{T}^n} [|f(\boldsymbol{z})|^p] &= \mathop{\mathbb{E}}_{\boldsymbol{z} \sim A^n, \boldsymbol{\xi} \sim \Omega_N^n} [|f(\boldsymbol{\xi} \odot \boldsymbol{z})|^p] \\ &\le d^p \mathcal{O}(\log K)^{dp} \mathop{\mathbb{E}}_{\boldsymbol{\xi} \sim \Omega_N^n} [|f(\boldsymbol{\xi})|^p], \end{aligned}$$

which finishes the argument. □

6.2 Junta theorem for functions on the hypergrid

We say a function $f : \Omega_K^n \to \mathbf{D}$ is a *k-junta* if it depends on only k coordinates. The concept of juntas is a central tool in the analysis of Boolean functions [ODo14].

So-called "junta theorems" show that functions which are simple in some way are close to juntas. For functions from the hypercube to the range $\{-1, 1\}$ these include Friedgut's theorem (which constrains the functions' total influence) and the FKN Theorem and Bourgain's junta theorem, both of which constrain the weight of the Fourier tail $\sum_{|S|\geq k} \widehat{f}(S)^2$. When the image of f is allowed to lie in $[-1, 1]$, we have the following result [DFKO07].

Theorem 28 ([DFKO07])**.** *Let $f : \Omega_2^n \to [-1, 1]$, ε and k be such that*

$$\sum_{|S|>k} \widehat{f}(S)^2 \leq \exp\Big(-\mathcal{O}(k^2 \log k)/\varepsilon\Big).$$

Then f is ε-close to a $(2^{\mathcal{O}(k)}/\varepsilon^2)$-junta.

In the case of functions on Ω_K^n, there is a statement along the lines of Friedgut's theorem for functions whose image is $\{0, 1\}$ [Ben+16]. Here we obtain a junta theorem for general $f : \Omega_K^n \to \mathbf{D}$ (in loose analogy to [DFKO07]) using the cyclic-group Bohnenblust–Hille inequality.

Theorem 29. *If $f : \Omega_K^n \to \mathbf{D}$ has degree at most d, then there exists another function $h : \Omega_K^n \to \mathbf{D}$ such that $\|f - h\|_2 \leq \varepsilon$ and h is a k-junta for*

$$k \leq d\left(\frac{\mathrm{BH}_{\Omega_K}^{2d}}{\varepsilon}\right)^{2d} \leq d\left(\frac{\mathcal{O}(\log K)^d}{\varepsilon}\right)^{2d}.$$

This argument is similar to the junta theorem for qubits in [VZ23], which is credited to Eskenazis. It is also a good warmup for the learning results in the sequel.

Proof. Denote the heavy Fourier coefficients of f by

$$S = \{\alpha : |\widehat{f}(\alpha)| \geq t\}$$

and define the function

$$f_t(z) = \sum_{\alpha \in \{0,1,\ldots,K-1\}^n} \mathbf{1}_{\alpha \in S} \cdot \widehat{f}(\alpha) z^\alpha.$$

Using the definition of S and the cyclic group BH inequality we have

$$|S| = \sum_{\alpha \in S} \left(\frac{|\alpha|}{|\alpha|} \right)^{\frac{2d}{d+1}} = t^{-\frac{2d}{d+1}} \sum_{\text{all } \alpha} |\alpha|^{\frac{2d}{d+1}} \leq t^{-\frac{2d}{d+1}} \left(\mathrm{BH}_{\widehat{\Omega}_K}^{\leq d} \right)^{\frac{2d}{d+1}},$$

so f_t is a k-junta with

$$k \leq d \left(\tfrac{1}{t} \cdot \mathrm{BH}_{\widehat{\Omega}_K}^{\leq d} \right)^{\frac{2d}{d+1}}.$$

Let $\mathrm{cl}(x) = x / \max\{1, |x|\}$ be a clamp function. Then $\mathrm{cl}(f_t)$ has image in $\mathbf{D}$ and is a k-junta for the same k. Because $|\mathrm{cl}(f_t)(x) - f(x)| \leq |f_t(x) - f(x)|$ pointwise, we also have

$$\|f - \mathrm{cl}(f_t)\|_2^2 \leq \|f - f_t\|_2^2 = \sum_{\alpha \notin S} |\widehat{f}(\alpha)|^2 \leq t^{\frac{2}{d+1}} \sum_{\text{all } \alpha} |\alpha|^{\frac{2d}{d+1}} \leq t^{\frac{2}{d+1}} \left(\mathrm{BH}_{\widehat{\Omega}_K}^{\leq d} \right)^{\frac{2d}{d+1}}.$$

Now with the choice $t = \varepsilon^{d+1} \left(\mathrm{BH}_{\widehat{\Omega}_K}^{\leq d} \right)^{-d}$, we have $\|f - \mathrm{cl}(f_t)\|_2 \leq \varepsilon$ with $\mathrm{cl}(f_t)$ a k-junta for

$$k = d \left(\frac{\mathrm{BH}_{\widehat{\Omega}_K}^{\leq d}}{\varepsilon} \right)^{2d}. \qquad \square$$

6.3 Qudit Bohnenblust–Hille in the Heisenberg–Weyl basis

Definition 2 (Heisenberg–Weyl Basis)**.** *Fix $K \geq 2$ and let $\omega = \omega_K = \exp(2\pi i / K)$. Define the K-dimensional* clock *and* shift *matrices respectively via*

$$Z |j\rangle = \omega^j |j\rangle, \qquad X |j\rangle = |j+1\rangle \qquad \text{for all} \qquad j \in \mathbb{Z}_K.$$

Here $\mathbf{Z}_K := \{0, 1, \ldots, K-1\}$ denotes the additive cyclic group of order K. Note that $X^K = Z^K = \mathbf{I}$. See more in [AEHK16]. Then the Heisenberg–Weyl basis for $M_K(\mathbf{C})$ is

$$\mathrm{HW}(K) := \{X^\ell Z^m\}_{\ell, m \in \mathbf{Z}_K}.$$

Any observable $A \in M_K(\mathbf{C})^{\otimes n}$ has a unique Fourier expansion with respect to $\mathrm{HW}(K)$ as well:

$$A = \sum_{\vec{\ell}, \vec{m} \in \mathbb{Z}_K^n} \widehat{A}(\vec{\ell}, \vec{m}) X^{\ell_1} Z^{m_1} \otimes \cdots \otimes X^{\ell_n} Z^{m_n}, \tag{6.3.1}$$

where $\widehat{A}(\vec{\ell}, \vec{m}) \in \mathbf{C}$ is the Fourier coefficient at $(\vec{\ell}, \vec{m})$. We say that A is *of degree at most d* if $\widehat{A}(\vec{\ell}, \vec{m}) = 0$ whenever

$$|(\vec{\ell}, \vec{m})| := \sum_{j=1}^{n} (\ell_j + m_j) > d.$$

Here, $0 \leq \ell_j, m_j \leq K-1$.

Noting that the eigenvalues of Heisenberg–Weyl matrices are the roots of unity, it is natural to pursue a reduction to a scalar BH inequality over Ω_K^n, the multiplicative cyclic group of order K—precisely the inequality needed for classical learning on functions on Ω_K^n. This reduction works well when K is prime.

Theorem 30 (Qudit Bohnenblust–Hille, Heisenberg–Weyl Basis: prime case)**.** *Fix a prime number $K \geq 2$ and suppose $d \geq 1$. Consider an observable $A \in M_K(\mathbf{C})^{\otimes n}$ of degree at most d. Then we have*

$$\|\widehat{A}\|_{\frac{2d}{d+1}} \leq C(d,K)\|A\|_{\mathrm{op}}, \tag{6.3.2}$$

with $C(d,K) \leq (K+1)^d \mathrm{BH}_{\Omega_K}^{\leq d}$.

When K is non-prime, the reduction still works under modifications, namely the degree may jump from d up to $(K-1)d$.

Theorem 31 (Qudit Bohnenblust–Hille, Heisenberg–Weyl Basis: non-prime case)**.** *Fix a non-prime number $K \geq 4$ and suppose $d \geq 1$. Consider an observable $A \in M_K(\mathbf{C})^{\otimes n}$ of degree at most d. Then we have*

$$\|\widehat{A}\|_{\frac{2(K-1)d}{(K-1)d+1}} \leq C(d,K)\|A\|_{\mathrm{op}}, \tag{6.3.3}$$

with $C(d,K) \leq K^{2d}\mathrm{BH}_{\Omega_K}^{\leq (K-1)d}$. In fact, the constant K^{2d} can be replaced by $|\Sigma_K|^d$ with $|\Sigma_K|$ being the cardinality of $\Sigma_K = \{(\ell, m) \in \mathbf{Z}_K \times \mathbf{Z}_K : \ell \text{ and } m \text{ are coprime}\}$.

The proofs of Theorems 30 and 31 are contained in Section 6.3. The full strength of Theorems 30 and 31 relies on the BH inequality for the cyclic groups Ω_K^n, *i.e.*, the finiteness of the Bohnenblust–Hille constant $\mathrm{BH}_{\Omega_K}^{\leq d}$ for cyclic groups which we quote here for reference.

Fix $K \geq 3$ and denote $\omega := e^{2\pi i/K}$. Let $\Omega_K := \{1, \omega, \omega^2, \ldots, \omega^{K-1}\}$. Then any function $f : \Omega_K^n \to \mathbf{C}$ admits the unique Fourier expansion

$$f(z) = \sum_{\alpha} \widehat{f}(\alpha) z^{\alpha}, \tag{6.3.4}$$

where $\alpha = (\alpha_1, \ldots, \alpha_n)$ are vectors of non-negative integers and each $\alpha_j \leq K-1$. We say f is *of degree at most d* if $\widehat{f}(\alpha) = 0$ whenever $|\alpha| > d$. The following result was proved in [SVZ24a; SVZ25; KSVZ24; Bec+25].

Theorem 32 (Cyclic Bohnenblust–Hille)**.** *Fix $K \geq 3$ and $d \geq 1$. There exists $C(d,K) > 0$ such that for all $n \geq 1$ and for all $f : \Omega_K^n \to \mathbf{C}$ of degree at most d, we have*

$$\|\widehat{f}\|_{\frac{2d}{d+1}} \leq C(d,K) \sup_{z \in \Omega_K^n} |f(z)|. \tag{6.3.5}$$

Denote by $\mathrm{BH}_{\Omega_K}^{\leq d}$ the best constant $C(d,K)$ in (6.3.5). An upper bound $\mathrm{BH}_{\Omega_K}^{\leq d} \leq O(\log K)^d$ was obtained in [KSVZ24; Bec+25].

* * *

In this section we prove Theorems 30 and 31 via reduction to the BH inequality for cyclic groups, Theorem 32. We collect first a few facts about the Heisenberg–Weyl basis $\{X^\ell Z^m\}$.

Fix $K \geq 3$. Recall that $\gcd(a,b)$ denotes the greatest common divisor of a and b. For $(\ell, m) \in \mathbf{Z}_K \times \mathbf{Z}_K$, $\gcd(\ell, m)$ is understood as when $\ell, m \in \{1, 2, \ldots, K\}$, i.e. we do not mod K freely here. For example if $K = 6$, then $\gcd(0,2)$ is understood as $\gcd(6,2) = 2$. For a group G, we use the convention that $\langle g \rangle$ is the abelian subgroup generated by $g \in G$. So for any $(\ell, m) \in \mathbf{Z}_K \times \mathbf{Z}_K$, we have

$$\langle (\ell, m) \rangle = \{(k\ell, km) : k \in \mathbf{Z}_K\}. \tag{6.3.6}$$

In the sequel, we denote $\omega = \omega_K = e^{2\pi i/K}$ and use the notation $\omega^{1/2} := \omega_{2K} = e^{\pi i/K}$.

Lemma 33. *We have the following:*

1. *$\{X^\ell Z^m : \ell, m \in \mathbb{Z}_K\}$ form a basis of $M_K(\mathbf{C})$.*
2. *For all $k, \ell, m \in \mathbb{Z}_K$:*

$$(X^\ell Z^m)^k = \omega^{\frac{1}{2}k(k-1)\ell m} X^{k\ell} Z^{km}$$

and for all $\ell_1, \ell_2, m_1, m_2 \in \mathbb{Z}_K$:

$$X^{\ell_1} Z^{m_1} \cdot X^{\ell_2} Z^{m_2} = \omega^{\ell_2 m_1 - \ell_1 m_2} X^{\ell_2} Z^{m_2} \cdot X^{\ell_1} Z^{m_1}.$$

3. *If $\gcd(\ell_1, m_1) = 1$ and $(\ell, m) \notin \langle (\ell_1, m_1) \rangle$, then*

$$X^{\ell_1} Z^{m_1} \cdot X^{\ell} Z^{m} = \omega^{\ell m_1 - \ell_1 m} X^{\ell} Z^{m} \cdot X^{\ell_1} Z^{m_1} \tag{6.3.7}$$

with $\omega^{\ell m_1 - \ell_1 m} \neq 1$.

4. *If* $\gcd(\ell, m) = 1$*, then the set of eigenvalues of* $X^\ell Z^m$ *is either* Ω_K *or* $\Omega_{2K} \setminus \Omega_K$.

Proof. 1. Suppose that $\sum_{\ell,m} a_{\ell,m} X^\ell Z^m = 0$. For any $j, k \in \mathbb{Z}_K$, we have

$$\sum_{\ell,m} a_{\ell,m} \langle X^\ell Z^m e_j, e_{j+k} \rangle = \sum_m a_{k,m} \omega^{jm} = 0.$$

Since the Vandermonde matrix associated to $(1, \omega, \ldots, \omega^{K-1})$ is invertible, we have $a_{k,m} = 0$ for all $k, m \in \mathbb{Z}_K$.

2. It follows immediately from the identity $ZX = \omega XZ$ which can be verified directly: for all $j \in \mathbb{Z}_K$

$$ZXe_j = Ze_{j+1} = \omega^{j+1} e_{j+1} = \omega^{j+1} Xe_j = \omega XZe_j.$$

3. It is a direct consequence of (2) and the following fact: for $(\ell_1, m_1) \in \mathbf{Z}_K \times \mathbf{Z}_K$ such that $\gcd(\ell_1, m_1) = 1$ and $(\ell_2, m_2) \in \mathbf{Z}_K \times \mathbf{Z}_K$, we have $\ell_1 m_2 - \ell_2 m_1 \equiv 0 \mod K$ if and only if $(\ell_2, m_2) \equiv (k\ell_1, k m_1) \mod K$ for some $k \in \mathbf{Z}_K$.

The "if" direction is obvious. To show the "only if" part, recall that by Bézout's lemma, there exist integers α and β such that $\alpha\ell_1 + \beta m_1 = \gcd(\ell_1, m_1) = 1$. Take $k \equiv \alpha\ell_2 + \beta m_2 \mod K$. Then

$$\ell_2 = \ell_2(\alpha\ell_1 + \beta m_1) \equiv \alpha\ell_1\ell_2 + \beta\ell_1 m_2 \equiv k\ell_1 \mod K, \tag{6.3.8}$$

where we used $\ell_1 m_2 \equiv \ell_2 m_1 \mod K$. Similarly,

$$m_2 = m_2(\alpha\ell_1 + \beta m_1) \equiv \alpha\ell_2 m_1 + \beta m_1 m_2 \equiv k m_1 \mod K, \tag{6.3.9}$$

as desired. This finishes the proof of the fact.

4. By (2), we have

$$(X^\ell Z^m)^{2K} = \omega^{K(2K-1)\ell m} X^{2\ell K} Z^{2mK} = \mathbf{I}.$$

So the eigenvalues of $X^\ell Z^m$ must be roots of unit of order $2K$. Then the proof is finished as soon as we prove the following claim: for $\gcd(\ell, m) = 1$, if λ is an eigenvalue of $X^\ell Z^m$, then so is $\omega\lambda$. To prove the claim, recall that by Bézout's lemma, there exist integers α and β such that $\alpha\ell + \beta m = \gcd(\ell, m) = 1$. By (2), we get

$$X^\ell Z^m X^\beta Z^{-\alpha} = \omega^{\alpha\ell + \beta m} X^\beta Z^{-\alpha} X^\ell Z^m = \omega X^\beta Z^{-\alpha} X^\ell Z^m. \tag{6.3.10}$$

Suppose $\vec{0} \neq \xi$ is an eigenvector of $X^\ell Z^m$ with eigenvalue λ. Then

$$X^\ell Z^m X^\beta Z^{-\alpha} \xi = \omega^{\alpha\ell+\beta m} X^\beta Z^{-\alpha} X^\ell Z^m \xi = \omega\lambda X^\beta Z^{-\alpha} \xi, \tag{6.3.11}$$

implying that $X^\beta Z^{-\alpha}\xi$ (non-zero since $X^\beta Z^{-\alpha}$ is invertible) is an eigenvector of $X^\ell Z^m$ with eigenvalue $\omega\lambda$. This finishes the proof of the claim. □

Let us record the following observation as a lemma.

Lemma 34. *Suppose that $k \geq 1$, A, B are two unitary matrices such that $B^k = \mathbf{I}$, $AB = \lambda BA$ with $\lambda \in \mathbf{C}$ and $\lambda \neq 1$. If $\xi \neq \vec{0}$ is an eigenvector of B with eigenvalue μ ($\mu \neq 0$ since $\mu^k = 1$), then*

$$\langle \xi, A\xi \rangle = 0.$$

Proof. By assumption

$$\mu\langle \xi, A\xi \rangle = \langle \xi, AB\xi \rangle = \lambda \langle \xi, BA\xi \rangle.$$

Since $B^\dagger = B^{k-1}$, $B^\dagger \xi = B^{k-1}\xi = \mu^{k-1}\xi = \overline{\mu}\xi$. Thus

$$\mu\langle \xi, A\xi \rangle = \lambda\langle \xi, BA\xi \rangle = \lambda\langle B^\dagger \xi, A\xi \rangle = \lambda\mu\langle \xi, A\xi \rangle.$$

Hence, $\mu(\lambda - 1)\langle \xi, A\xi \rangle = 0$. This gives $\langle \xi, A\xi \rangle = 0$ as $\mu(\lambda - 1) \neq 0$. □

The prime K case

In this subsection we prove Theorem 30. When K is prime, the basis $\{X^\ell Z^m\}$ has nicer properties.

Lemma 35. *Fix $K \geq 3$ a prime number. Consider the set of generators*

$$\Sigma_K := \{(1,0), (1,1), \ldots, (1, K-1), (0,1)\}. \tag{6.3.12}$$

Then the group $\mathbf{Z}_K \times \mathbf{Z}_K$ is the union of subgroups

$$\mathbf{Z}_K \times \mathbf{Z}_K = \bigcup_{(\ell,m)\in\Sigma_K} \langle (\ell, m) \rangle, \tag{6.3.13}$$

where each two subgroups intersects with the unit $(0,0)$ only. Moreover, for any $(\ell, m) \in \Sigma_K$, the set of eigenvalues of each $X^\ell Z^m$ is exactly Ω_K.

Proof. To prove the first statement, take any $(\ell, m) \in \mathbf{Z}_K \times \mathbf{Z}_K$. If $\ell = 0$, then $(\ell, m) = (0, m) \in \langle (0,1) \rangle$. If $\ell \neq 0$, then $\gcd(\ell, K) = 1$. So by Bézout's lemma, there exists ℓ' such that $\ell\ell' \equiv 1 \mod K$. Thus $(\ell, m) = (\ell, \ell\ell' m) \in \langle (1, \ell' m) \rangle$. The statement about the intersection is clear since otherwise, the cardinality of the union of these subgroups is strictly smaller than $1 + (K+1)(K-1) = K^2$ which leads to a contradiction.

The second statement follows from the proof of Lemma 33. In fact, when K is odd, $(K-1)/2$ is an integer and we have by Lemma 33 (2) that

$$(X^\ell Z^m)^K = \omega^{\frac{1}{2}K(K-1)\ell m} X^{\ell K} Z^{mK} = \mathbf{I}.$$

So the eigenvalues of $X^\ell Z^m$ must be roots of unity of order K. This, together with the claim in the proof of Lemma 33 (4) and the fact that $\gcd(\ell, m) = 1, (\ell, m) \in \Sigma_K$, completes the proof of the lemma. □

Now we are ready to prove Theorem 30:

Proof of Theorem 30. Fix a prime number $K \geq 2$. Recall that $\omega = e^{\frac{2\pi i}{K}}$. Consider Σ_K defined in (6.3.12). For any $(\ell, m) \in \Sigma_K$, by Lemma 35 any $z \in \Omega_K$ is an eigenvalue of $X^\ell Z^m$ and we denote by $e_z^{\ell,m}$ the corresponding unit eigenvector. For any vector $\vec{\omega} \in \Omega_K^{(K+1)n}$ of the form (noting that $|\Sigma_K| = K+1$)

$$\vec{\omega} = (\vec{\omega}^{\ell,m})_{(\ell,m)\in\Sigma_K}, \qquad \vec{\omega}^{\ell,m} = (\omega_1^{\ell,m}, \ldots, \omega_n^{\ell,m}) \in \Omega_K^n, \tag{6.3.14}$$

we consider the matrix

$$\rho(\vec{\omega}) := \rho_1(\vec{\omega}) \otimes \cdots \otimes \rho_n(\vec{\omega}),$$

where

$$\rho_k(\vec{\omega}) := \frac{1}{K+1} \sum_{(\ell,m)\in\Sigma_K} |e^{\ell,m}_{\omega_k^{\ell,m}}\rangle\langle e^{\ell,m}_{\omega_k^{\ell,m}}|.$$

Then each $\rho_k(\vec{\omega})$ is a density matrix and so is $\rho(\vec{\omega})$.

Fix $(\ell, m) \in \Sigma_K$ and $1 \leq k \leq K-1$. We have by Lemma 33

$$\begin{aligned} \mathrm{tr}[X^{k\ell} Z^{km} |e_z^{\ell,m}\rangle\langle e_z^{\ell,m}|] &= \omega^{-\frac{1}{2}k(k-1)\ell m} \langle e_z^{\ell,m}, (X^\ell Z^m)^k e_z^{\ell,m}\rangle \\ &= \omega^{-\frac{1}{2}k(k-1)\ell m} z^k, \qquad z \in \Omega_K. \end{aligned}$$

On the other hand, for any $(\ell, m) \neq (\ell', m') \in \Sigma_K$, we have $(k\ell, km) \notin \langle (\ell', m') \rangle$ by Lemma 35. From our choice $\gcd(\ell', m') = 1$. So Lemma 33 gives

$$X^{k\ell} Z^{km} X^{\ell'} Z^{m'} = \omega^{k\ell' m - k\ell m'} X^{\ell'} Z^{m'} X^{k\ell} Z^{km}$$

with $\omega^{k\ell' m - k\ell m'} \neq 1$. This, together with Lemma 34, implies

$$\mathrm{tr}[X^{k\ell} Z^{km} |e_z^{\ell',m'}\rangle\langle e_z^{\ell',m'}|] = \langle e_z^{\ell',m'}, X^{k\ell} Z^{km} e_z^{\ell',m'}\rangle = 0, \qquad z \in \Omega_K.$$

for any $1 \leq k \leq K-1$. All combined, for all $1 \leq k \leq K-1, (\ell, m) \in \Sigma_K$ and $1 \leq j \leq n$ we get

$$\begin{aligned}
\mathrm{tr}[X^{k\ell} Z^{km} \rho_j(\vec{\omega})] &= \frac{1}{K+1} \sum_{(\ell',m') \in \Sigma_K} \left\langle e^{\ell',m'}_{\omega_j^{\ell',m'}}, X^{k\ell} Z^{km} e^{\ell',m'}_{\omega_j^{\ell',m'}} \right\rangle \\
&= \frac{1}{K+1} \left\langle e^{\ell,m}_{\omega_j^{\ell,m}}, X^{k\ell} Z^{km} e^{\ell,m}_{\omega_j^{\ell,m}} \right\rangle \\
&= \frac{1}{K+1} \omega^{-\frac{1}{2}k(k-1)\ell m} (\omega_j^{\ell,m})^k.
\end{aligned}$$

Note that by Lemma 35 any polynomial in $M_K(\mathbf{C})^{\otimes n}$ of degree at most d is a linear combination of monomials

$$A(\vec{k}, \vec{\ell}, \vec{m}; \vec{i}) := \cdots \otimes X^{k_1 \ell_1} Z^{k_1 m_1} \otimes \cdots \otimes X^{k_\kappa \ell_\kappa} Z^{k_\kappa m_\kappa} \otimes \cdots,$$

where

- $\vec{k} = (k_1, \ldots, k_\kappa) \in \{1, \ldots, K-1\}^\kappa$ with $0 \leq \sum_{j=1}^\kappa k_j \leq d$;
- $\vec{\ell} = (\ell_1, \ldots, \ell_\kappa), \vec{m} = (m_1, \ldots, m_\kappa)$ with each $(\ell_j, m_j) \in \Sigma_K$;
- $\vec{i} = (i_1, \ldots, i_\kappa)$ with $1 \leq i_1 < \cdots < i_\kappa \leq n$;
- $X^{k_j \ell_j} Z^{k_j m_j}$ appears in the i_j-th place, $1 \leq j \leq \kappa$, and all the other $n - \kappa$ elements in the tensor product are the identity matrices $\mathbf{I}$.

So for any $\vec{\omega} \in \Omega_K^{(K+1)n}$ of the form (6.3.14) we have from the above discussion that

$$\begin{aligned}
\mathrm{tr}[A(\vec{k}, \vec{\ell}, \vec{m}; \vec{i}) \rho(\vec{\omega})] &= \prod_{j=1}^\kappa \mathrm{tr}[X^{k_j \ell_j} Z^{k_j m_j} \rho_{i_j}(\vec{\omega})] \\
&= \frac{\omega^{-\frac{1}{2}\sum_{j=1}^\kappa k_j(k_j-1)\ell_j m_j}}{(K+1)^\kappa} (\omega_{i_1}^{\ell_1,m_1})^{k_1} \cdots (\omega_{i_\kappa}^{\ell_\kappa,m_\kappa})^{k_\kappa}.
\end{aligned}$$

Thus $\vec{\omega} \mapsto \mathrm{tr}[A(\vec{k}, \vec{\ell}, \vec{m}; \vec{i}) \rho(\vec{\omega})]$ is a monomial on $(\Omega_K)^{(K+1)n}$ of degree at most $\sum_{j=1}^\kappa k_j \leq d$.

Now for general polynomial $A \in M_K(\mathbf{C})^{\otimes n}$ of degree at most d:

$$A = \sum_{\vec{k}, \vec{\ell}, \vec{m}, \vec{i}} c(\vec{k}, \vec{\ell}, \vec{m}; \vec{i}) A(\vec{k}, \vec{\ell}, \vec{m}; \vec{i}),$$

where the sum runs over the above $(\vec{k},\vec{\ell},\vec{m};\vec{i})$. This is the Fourier expansion of A and each $c(\vec{k},\vec{\ell},\vec{m};\vec{i})\in\mathbf{C}$ is the Fourier coefficient. So

$$\|\widehat{A}\|_p=\left(\sum_{\vec{k},\vec{\ell},\vec{m},\vec{i}}|c(\vec{k},\vec{\ell},\vec{m};\vec{i})|^p\right)^{1/p}.$$

To each A we assign the function f_A on $\Omega_K^{(K+1)n}$ given by

$$\begin{aligned}f_A(\vec{\omega})&=\mathrm{tr}[A\rho(\vec{\omega})]\\&=\sum_{\vec{k},\vec{\ell},\vec{m},\vec{i}}\frac{\omega^{-\frac{1}{2}\sum_{j=1}^{\kappa}k_j(k_j-1)\ell_jm_j}c(\vec{k},\vec{\ell},\vec{m};\vec{i})}{(K+1)^{\kappa}}(\omega_{i_1}^{\ell_1,m_1})^{k_1}\cdots(\omega_{i_\kappa}^{\ell_\kappa,m_\kappa})^{k_\kappa}.\end{aligned}$$

Note that this is the Fourier expansion of f_A since the monomials $(\omega_{i_1}^{\ell_1,m_1})^{k_1}\cdots(\omega_{i_\kappa}^{\ell_\kappa,m_\kappa})^{k_\kappa}$ differ for distinct $(\vec{k},\vec{\ell},\vec{m};\vec{i})$'s. Therefore, for $p>0$

$$\begin{aligned}\|\widehat{f_A}\|_p&=\left(\sum_{\vec{k},\vec{\ell},\vec{m},\vec{i}}\left|\frac{c(\vec{k},\vec{\ell},\vec{m};\vec{i})}{(K+1)^{\kappa}}\right|^p\right)^{1/p}\\&\geq\frac{1}{(K+1)^d}\left(\sum_{\vec{k},\vec{\ell},\vec{m},\vec{i}}|c(\vec{k},\vec{\ell},\vec{m};\vec{i})|^p\right)^{1/p}\\&=\frac{1}{(K+1)^d}\|\widehat{A}\|_p.\end{aligned}$$

According to Theorem 32, one has

$$\|\widehat{f_A}\|_{\frac{2d}{d+1}}\leq\mathrm{BH}_{\Omega_K}^{\leq d}\|f_A\|_{\Omega_K^{(K+1)n}}$$

for some $\mathrm{BH}_{\Omega_K}^{\leq d}<\infty$. Since each $\rho(\vec{\omega})$ is a density matrix, we have by duality that

$$\|f_A\|_{\Omega_K^{(K+1)n}}=\sup_{\vec{\omega}\in(\Omega_K)^{(K+1)n}}|\,\mathrm{tr}[A\rho(\vec{\omega})]|\leq\|A\|_{\mathrm{op}}.$$

All combined, we obtain

$$\|\widehat{A}\|_{\frac{2d}{d+1}}\leq(K+1)^d\|\widehat{f_A}\|_{\frac{2d}{d+1}}\leq(K+1)^d\mathrm{BH}_{\Omega_K}^{\leq d}\|f_A\|_{\Omega_K^{(K+1)n}}\leq(K+1)^d\mathrm{BH}_{\Omega_K}^{\leq d}\|A\|_{\mathrm{op}}.$$

□

The non-prime K case

This subsection is devoted to the proof of Theorem 31. Throughout this part, $K\geq 4$ is a non-prime integer.

We start with a substitute of Σ_K in (6.3.12) for non-prime K.

Lemma 36. *Fix non-prime $K \geq 4$. Consider*

$$\Sigma_K := \{(\ell, m) \in \mathbf{Z}_K \times \mathbf{Z}_K : \gcd(\ell, m) = 1\}. \tag{6.3.15}$$

Then $\mathbf{Z}_K \times \mathbf{Z}_K$ is the union of subgroups generated by elements in Σ_K:

$$\mathbf{Z}_K \times \mathbf{Z}_K = \bigcup_{(\ell,m)\in\Sigma_K} \langle (\ell, m) \rangle. \tag{6.3.16}$$

Proof. The proof is direct: any $(\ell, m) \in \mathbf{Z}_K \times \mathbf{Z}_K$ belongs to $\langle (\ell_1, m_1) \rangle$ for $(\ell_1, m_1) = (\ell/\gcd(\ell, m), m/\gcd(\ell, m)) \in \Sigma_K$. □

Recall that when K is prime, for two different subgroups $\langle (\ell_1, m_1) \rangle \neq \langle (\ell_2, m_2) \rangle$ one has the singleton set $\{(0, 0)\}$ as their intersection. However, this is no longer the case when K is not prime. For example, for $K = 6$, we have $\langle (1, 0) \rangle \neq \langle (2, 3) \rangle$ while $\langle (1, 0) \rangle \cap \langle (2, 3) \rangle = \{(0, 0), (2, 0), (4, 0)\}$. This difference will make the proof of Theorem 31 more involved.

Proof of Theorem 31. Fix non-prime $K \geq 4$. Consider Σ_K defined in (6.3.15). Then we know by Lemma 36 that the set of eigenvalues of $X^\ell Z^m$ is either Ω_K or $\Omega_{2K} \setminus \Omega_K$. In either case, suppose that z is an eigenvalue of $X^\ell Z^m$. We denote by $e_z^{\ell,m}$ the unit eigenvector of $X^\ell Z^m$ corresponding to z.

Write $\Sigma_K = \Sigma_K^+ \cup \Sigma_K^-$ with

$$\Sigma_K^+ := \{(\ell, m) \in \Sigma_K : \text{ the set of eigenvalues of } X^\ell Z^m \text{ is } \Omega_K\} \tag{6.3.17}$$

and

$$\Sigma_K^- := \{(\ell, m) \in \Sigma_K : \text{ the set of eigenvalues of } X^\ell Z^m \text{ is } \Omega_{2K} \setminus \Omega_K\}. \tag{6.3.18}$$

As before, for any $\vec{\omega} \in \Omega_K^{|\Sigma_K| n}$ of the form

$$\vec{\omega} = (\vec{\omega}^{\ell,m})_{(\ell,m)\in\Sigma_K}, \qquad \vec{\omega}^{\ell,m} = (\omega_1^{\ell,m}, \ldots, \omega_n^{\ell,m}) \in \Omega_K^n, \tag{6.3.19}$$

we shall consider

$$\rho(\vec{\omega}) := \rho_1(\vec{\omega}) \otimes \cdots \otimes \rho_n(\vec{\omega}) \tag{6.3.20}$$

where each $\rho_j(\vec{\omega})$ is the average of some eigen-projections of $X^\ell Z^m, (\ell, m) \in \Sigma_K$. If $(\ell, m) \in \Sigma_K^+$, then $X^\ell Z^m$ has $\omega_j^{\ell,m} \in \Omega_K$ as an eigenvalue with $e^{\ell,m}_{\omega_j^{\ell,m}}$ being the unit eigenvector. If $(\ell, m) \in \Sigma_K^-$, then $\omega_j^{\ell,m} \in \Omega_K$ is not an eigenvalue of $X^\ell Z^m$. In this case, $X^\ell Z^m$ has $\omega^{1/2}\omega_j^{\ell,m} \in \Omega_{2K} \setminus \Omega_K$ as an eigenvalue with $e^{\ell,m}_{\omega^{1/2}\omega_j^{\ell,m}}$ being the unit eigenvector.

For each $1 \le j \le n$, consider

$$\rho_j(\vec{\omega}) := \frac{1}{|\Sigma_K|} \sum_{(\ell,m)\in\Sigma_K^+} |e^{\ell,m}_{\omega_j^{\ell,m}}\rangle \langle e^{\ell,m}_{\omega_j^{\ell,m}}| + \frac{1}{|\Sigma_K|} \sum_{(\ell,m)\in\Sigma_K^-} |e^{\ell,m}_{\omega^{1/2}\omega_j^{\ell,m}}\rangle \langle e^{\ell,m}_{\omega^{1/2}\omega_j^{\ell,m}}| . \tag{6.3.21}$$

By definition, each $\rho_j(\vec{\omega})$ is a density matrix and so is $\rho(\vec{\omega})$.

For any $(0,0) \neq (\ell',m') \in \mathbf{Z}_K \times \mathbf{Z}_K$ and any $(\ell,m) \in \Sigma_K$, either $(\ell',m') \notin \langle(\ell,m)\rangle$ or $(\ell',m') = (k\ell, km)$ for some $k \in \mathbf{Z}_K$. If $(\ell',m') \notin \langle(\ell,m)\rangle$, then by Lemma 33

$$X^{\ell'}Z^{m'} \cdot X^{\ell}Z^{m} = \omega^{\ell m' - \ell' m} X^{\ell}Z^{m} \cdot X^{\ell'}Z^{m'} \tag{6.3.22}$$

with $\omega^{\ell m' - \ell' m} \neq 1$. So Lemma 34 gives

$$\mathrm{tr}[X^{\ell'}Z^{m'} |e^{\ell,m}_z\rangle \langle e^{\ell,m}_z|] = 0, \tag{6.3.23}$$

for any eigenvalue z of $X^{\ell}Z^{m}$.

If $(\ell',m') = (k\ell, km)$ for some $k \in \mathbf{Z}_K$, then by Lemma 33

$$X^{\ell'}Z^{m'} = X^{k\ell}Z^{km} = \omega^{-\frac{1}{2}k(k-1)\ell m}(X^{\ell}Z^{m})^k. \tag{6.3.24}$$

So for any eigenvalue z of $X^{\ell}Z^{m}$.

$$\mathrm{tr}[X^{\ell'}Z^{m'} |e^{\ell,m}_z\rangle \langle e^{\ell,m}_z|] = \omega^{-\frac{1}{2}k(k-1)\ell m} z^k. \tag{6.3.25}$$

All combined, we have for any $\vec{\omega} \in (\Omega_K)^{|\Sigma_K|n}$ that

$$\begin{aligned}
\mathrm{tr}\Big[X^{\ell'}Z^{m'}\rho_j(\vec{\omega})\Big] &= \frac{1}{|\Sigma_K|} \sum_{(\ell,m)\in\Sigma_K^+ : (\ell',m') = (k_{\ell,m}\ell, k_{\ell,m}m)} \omega^{-\frac{1}{2}k_{\ell,m}(k_{\ell,m}-1)\ell m} (\omega_j^{\ell,m})^{k_{\ell,m}} \\
&\quad + \frac{1}{|\Sigma_K|} \sum_{(\ell,m)\in\Sigma_K^- : (\ell',m') = (k_{\ell,m}\ell, k_{\ell,m}m)} \omega^{-\frac{1}{2}k_{\ell,m}(k_{\ell,m}-1)\ell m} (\omega^{1/2}\omega_j^{\ell,m})^{k_{\ell,m}} \\
&= \frac{1}{|\Sigma_K|} \sum_{(\ell,m)\in\Sigma_K^+ : (\ell',m') = (k_{\ell,m}\ell, k_{\ell,m}m)} \omega^{-\frac{1}{2}k_{\ell,m}(k_{\ell,m}-1)\ell m} (\omega_j^{\ell,m})^{k_{\ell,m}} \\
&\quad + \frac{1}{|\Sigma_K|} \sum_{(\ell,m)\in\Sigma_K^- : (\ell',m') = (k_{\ell,m}\ell, k_{\ell,m}m)} \omega^{-\frac{1}{2}k_{\ell,m}(k_{\ell,m}-1)\ell m + \frac{1}{2}k_{\ell,m}} (\omega_j^{\ell,m})^{k_{\ell,m}}.
\end{aligned}$$

Here in the summation, when $(\ell',m') \in \langle(\ell,m)\rangle$ we write $(\ell',m') = (k_{\ell,m}\ell, k_{\ell,m}m)$ with $1 \le k_{\ell,m} \le K-1$. So $\vec{\omega} \mapsto \mathrm{tr}\Big[X^{\ell'}Z^{m'}\rho_j(\vec{\omega})\Big]$ is a polynomial on $\Omega_K^{|\Sigma_K|n}$ of degree at most $K-1$, and all the (non-zero) coefficients are of modulus $|\Sigma_K|^{-1}$. To compare, in the prime K case, we have only one non-zero term in the above summation. In the non-prime K case, there might be more than one term.

That is, we may have different (ℓ_1, m_1) and (ℓ_2, m_2) in Σ_K such that $(\ell', m') = (k_1\ell_1, k_1 m_1) = (k_2\ell_2, k_2 m_2)$ with $1 \le k_1, k_2 \le K-1$. For example for $K = 6$, we have $(0,3) = (3\cdot 0, 3\cdot 1) = (3\cdot 2, 3\cdot 3) = (3\cdot 4, 3\cdot 5) = (3\cdot 2, 3\cdot 1) = (3\cdot 4, 3\cdot 3)$. Though $\mathrm{tr}[X^{\ell'} Z^{m'} \rho_j(\vec{\omega})]$ is no longer a monomial of degree $\deg(X^{\ell'} Z^{m'})$, it is still a non-zero polynomial of degree at most $K-1$.

Now for any monomial in $M_K(\mathbf{C})^{\otimes n}$ of degree at most d admitting the form

$$A(\vec{\ell}, \vec{m}; \vec{i}) := \cdots \otimes X^{\ell_1} Z^{m_1} \otimes \cdots \otimes X^{\ell_\kappa} Z^{m_\kappa} \otimes \cdots, \tag{6.3.26}$$

where $\kappa \le d$ and

- $\vec{\ell} = (\ell_1, \ldots, \ell_\kappa), \vec{m} = (m_1, \ldots, m_\kappa)$ with each $(0,0) \neq (\ell_j, m_j) \in \mathbf{Z}_K \times \mathbf{Z}_K$;
- $\vec{i} = (i_1, \ldots, i_\kappa)$ with $1 \le i_1 < \cdots < i_\kappa \le n$;
- and each $X^{\ell_j} Z^{m_j}$ appears in the i_j-th place, and all the other $n - \kappa$ elements in the tensor product are the identity matrices $\mathbf{I}$.

According to our previous discussion,

$$\mathrm{tr}[A(\vec{\ell}, \vec{m}; \vec{i}) \rho(\vec{\omega})] = \prod_{1 \le j \le \kappa} \mathrm{tr}[X^{\ell_j} Z^{m_j} \rho_{i_j}(\vec{\omega})] \tag{6.3.27}$$

is a linear combination of monomials

$$(\omega_{i_1}^{a_1, b_1})^{c_1} \cdots (\omega_{i_\kappa}^{a_\kappa, b_\kappa})^{c_\kappa}$$

of degree at most $(K-1)\kappa \le (K-1)d$, with $(a_j, b_j) \in \Sigma_K$ and $1 \le c_j \le K-1$ such that $(c_j a_j, c_j b_j) \equiv (\ell_j, m_j) \mod K$. This implies that these monomials remember the profile $(\vec{\ell}, \vec{m}; \vec{i})$ well; *i.e.*, for distinct $(\vec{\ell}, \vec{m}; \vec{i}) \neq (\vec{\ell'}, \vec{m'}; \vec{i'})$, the corresponding polynomials $\mathrm{tr}[A(\vec{\ell}, \vec{m}; \vec{i}) \rho(\vec{\omega})]$ and $\mathrm{tr}[A(\vec{\ell'}, \vec{m'}; \vec{i'}) \rho(\vec{\omega})]$ do not admit common monomials. Moreover, the coefficients of the those monomials are all of the modulus $|\Sigma_K|^{-\kappa} \ge |\Sigma_K|^{-d}$.

For general $A \in M_K(\mathbf{C})^{\otimes n}$ of degree at most d admitting

$$A = \sum_{\vec{\ell}, \vec{m}, \vec{i}} c(\vec{\ell}, \vec{m}; \vec{i}) A(\vec{\ell}, \vec{m}; \vec{i}) \tag{6.3.28}$$

as the Fourier expansion, consider the polynomial

$$f_A(\vec{\omega}) = \mathrm{tr}[A \rho(\vec{\omega})] = \sum_{\vec{\ell}, \vec{m}, \vec{i}} c(\vec{\ell}, \vec{m}; \vec{i}) \, \mathrm{tr}[A(\vec{\ell}, \vec{m}; \vec{i}) \rho(\vec{\omega})] \tag{6.3.29}$$

on $\Omega_K^{|\Sigma_K|^n}$. From the above discussion, the ℓ^p-norm $\|\widehat{f_A}\|_p$ of Fourier coefficients of f_A satisfies

$$\|\widehat{f_A}\|_p \geq |\Sigma_K|^{-d}\left(\sum_{\vec{\ell},\vec{m},\vec{i}} |c(\vec{\ell},\vec{m};\vec{i})|^p\right)^{1/p} = |\Sigma_K|^{-d}\|\widehat{A}\|_p, \qquad p > 0.$$

Moreover, f_A is of degree at most $(K-1)d$. So Theorem 32 implies

$$\|\widehat{f_A}\|_{\frac{2(K-1)d}{(K-1)d+1}} \leq \mathrm{BH}_{\Omega_K}^{\leq (K+1)d}\|f_A\|_\infty.$$

Recall that each $\rho(\vec{\omega})$ is a density matrix, so by duality

$$\|f_A\|_\infty = \sup_{\vec{\omega}\in(\Omega_K)^{(K+1)n}} |\operatorname{tr}[A\rho(\vec{\omega})]| \leq \|A\|_{\mathrm{op}}.$$

All combined, we prove that

$$\|\widehat{A}\|_{\frac{2(K-1)d}{(K-1)d+1}} \leq |\Sigma_K|^d\|\widehat{f_A}\|_{\frac{2(K-1)d}{(K-1)d+1}} \leq |\Sigma_K|^d\mathrm{BH}_{\Omega_K}^{\leq (K+1)d}\|f_A\|_\infty \leq |\Sigma_K|^d\mathrm{BH}_{\Omega_K}^{\leq (K+1)d}\|A\|_{\mathrm{op}}.$$

□

6.4 Learning

Here we give learning algorithms for low-degree functions on Ω_K^n and local qudit observables. This work is published in [KSVZ24]. We begin by extracting the estimation lemma implicit in [EI22] that will allow us to use our new Bohnenblust–Hille-type inequalities.

Theorem 37 (Generic Eskenazis–Ivanisvili)**.** *Let $d \in \mathbb{N}$ and $\eta, B > 0$. Suppose $v, w \in \mathbf{C}^n$ with $\|v-w\|_\infty \leq \eta$ and $\|v\|_{\frac{2d}{d+1}} \leq B$. Then for $\widetilde{w}$ defined as $\widetilde{w}_j = w_j\mathbf{1}_{[|w_j|\geq\eta(1+\sqrt{d+1})]}$ we have the bound*

$$\|\widetilde{w}-v\|_2^2 \leq (e^5\eta^2 dB^{2d})^{\frac{1}{d+1}}.$$

Proof. Let $t > 0$ be a threshold parameter to be chosen later. Define $S_t = \{j : |w_j| \geq t\}$ and note from the triangle inequality in $\mathbf{C}$ that

$$|v_j| \geq |w_j| - |v_j - w_j| = t - \eta \text{ for } j \in S_t \tag{6.4.1}$$

$$|v_j| \leq |w_j| + |v_j - w_j| = t + \eta \text{ for } j \notin S_t. \tag{6.4.2}$$

We may also estimate $|S_t|$ as

$$|S_t| = \sum_{j\in S}\frac{|v_j|}{|v_j|} \overset{(6.4.1)}{\leq} (t-\eta)^{-\frac{2d}{d+1}}\sum_{j\in[n]}|v_j|^{\frac{2d}{d+1}} \leq (t-\eta)^{-\frac{2d}{d+1}}\|v\|_{\frac{2d}{d+1}}^{\frac{2d}{d+1}} \leq (t-\eta)^{-\frac{2d}{d+1}}B^{\frac{2d}{d+1}}. \tag{6.4.3}$$

With $\widetilde{w}^{(t)} := (w_j \mathbf{1}_{[w_j \geq t]})_{j=1}^n$, we find

$$\|\widetilde{w}^{(t)} - v\|_2^2 = \sum_{j \in S_t} |w_j - v_j|^2 + \sum_{j \notin S_t} |v_j|^2 \overset{(6.4.2)}{\leq} |S_t|\eta^2 + (t+\eta)^{\frac{2}{d+1}} \sum_{j \in [n]} |v_j|^{\frac{2d}{d+1}}$$
$$\overset{(6.4.3)}{\leq} B^{\frac{2d}{d+1}} \left(\eta^2 (t-\eta)^{-\frac{2d}{d+1}} + (t+\eta)^{\frac{2}{d+1}} \right).$$

Choosing $t = \eta(1+\sqrt{d+1})$ then yields $\widetilde{w}$ with error, after some careful scalar estimates,

$$\|\widetilde{w} - v\|_2^2 \leq (e^5 \eta^2 d B^{2d})^{\frac{1}{d+1}}.$$

See [EI22, Eqs. 18 & 19] for details on the scalar estimates. □

In the context of low-degree learning, v is the true vector of Fourier coefficients, and w is the vector of empirical coefficients obtained through Fourier sampling.

Cyclic group learning

Theorem 38. *Let $f : \mathbf{Z}_K^n \to \mathbf{D}$ be a degree-d function. Then with*

$$(\log K)^{\mathcal{O}(d^2)} \log(n/\delta) \varepsilon^{-d-1}$$

independent random samples $(x, f(x))$, $x \sim \mathcal{U}(\mathbf{Z}_K^n)$, we may with confidence $1-\delta$ learn a function $\widetilde{f} : \mathbf{Z}_K^n \to \mathbf{C}$ with $\|f - \widetilde{f}\|_2^2 \leq \varepsilon$.

Proof. Let $f = \sum_\alpha \widehat{f}(\alpha) z^\alpha$ be the Fourier expansion. For a number of samples s to be specified later, sample $x^{(1)}, \ldots, x^{(s)} \overset{\text{iid}}{\sim} \mathcal{U}(\{0, 1, \ldots, K-1\}^n)$ and for each $\alpha \in \mathbf{Z}_K^n$ with $|\alpha| \leq d$, form the empirical Fourier coefficient

$$w_\alpha := \frac{1}{s} \sum_{j=1}^s f(x^{(j)}) \omega_K^{-\sum_{\ell=1}^n \alpha_\ell x_\ell^{(j)}},$$

where $\omega_K = e^{\frac{2\pi i}{K}}$ and $x^{(j)} = (x_1^{(j)}, \ldots, x_n^{(j)})$. Then w_α is a sum of bounded i.i.d. random variables with expected value $\widehat{f}(\alpha)$, so Chernoff gives

$$\begin{aligned}
\Pr[|\widehat{f}(\alpha) - w_\alpha| \geq \eta] &= \Pr\left[\Re\left(\widehat{f}(\alpha) - w_\alpha\right)^2 + \Im\left(\widehat{f}(\alpha) - w_\alpha\right)^2 \geq \eta^2 \right] \\
&\leq \Pr\left[|\Re\left(\widehat{f}(\alpha) - w_\alpha\right)| \geq \eta/\sqrt{2} \right] + \Pr\left[|\Im\left(\widehat{f}(\alpha) - w_\alpha\right)| \geq \eta/\sqrt{2} \right] \\
&\leq 4 \exp\left(-s\eta^2/4\right).
\end{aligned}$$

So the probability we simultaneously estimate all nonzero Fourier coefficients of f to within η is

$$\Pr\left[|\widehat{f}(\alpha) - w_\alpha| < \eta \text{ for all } \alpha \text{ with } |\alpha| \leq d \right] \geq 1 - 4 \sum_{k=0}^d \binom{n}{k} \exp\left(\frac{-s\eta^2}{4} \right),$$

which in turn we will require to be $\geq 1-\delta$.

Now applying Theorem 37 to obtain $\widetilde{w}$ and recalling $\|\widehat{f}\|_{\frac{2d}{d+1}} \leq \mathrm{BH}_{\mathbf{Z}_K}^{\leq d}\|f\|_\infty = \mathrm{BH}_{\mathbf{Z}_K}^{\leq d}$ we have that with probability $1-\delta$, the function $\widetilde{f}(x) := \sum_\alpha \widetilde{w}_\alpha \prod_{j=1}^n \omega_K^{\alpha_j x_j}$ has L_2 error

$$\|\widetilde{f}-f\|_2^2 \overset{\text{(Parseval)}}{=} \sum_\alpha |\widehat{f}(\alpha)-\widetilde{\omega}_\alpha|^2 \leq \left(e^5\eta^2 d(\mathrm{BH}_{\mathbf{Z}_K}^{\leq d})^{2d}\right)^{\frac{1}{d+1}}. \tag{6.4.4}$$

So in order to achieve $\|\widetilde{f}-f\|_2^2 \leq \varepsilon$ it is enough to pick $\eta^2 = \varepsilon^{d+1}e^{-5}d^{-1}(\mathrm{BH}_{\mathbf{Z}_K}^{\leq d})^{-2d}$, which entails by standard estimates that taking a number of samples s with

$$s \geq \frac{4e^5 d^2 (\mathrm{BH}_{\mathbf{Z}_K}^{\leq d})^{2d}}{\varepsilon^{d+1}} \log\left(\frac{4en}{\delta}\right)$$

suffices. □

Qudit learning

We will find it more convenient to use a different orthonormal basis for qudit learning, the so-called Gell–Mann matrices.

Definition 3 (Gell-Mann Basis)**.** *Let $K \geq 2$. Put $E_{jk} = |j\rangle\langle k|, 1 \leq j,k \leq K$. The generalized Gell-Mann Matrices are a basis of $M_K(\mathbf{C})$ and are comprised of the identity matrix $\mathbf{I}$ along with the following generalizations of the Pauli matrices:*

$$\begin{aligned}
&\textit{symmetric:} && \mathbf{A}_{jk} = \sqrt{\tfrac{K}{2}}\left(E_{jk}+E_{kj}\right) && \textit{for } 1 \leq j < k \leq K \\
&\textit{antisymmetric:} && \mathbf{B}_{jk} = \sqrt{\tfrac{K}{2}}\left(-iE_{jk}+iE_{kj}\right) && \textit{for } 1 \leq j < k \leq K \\
&\textit{diagonal:} && \mathbf{C}_m = \Gamma_m\left(\textstyle\sum_{k=1}^m E_{kk} - mE_{m+1,m+1}\right) && \textit{for } 1 \leq m \leq K-1,
\end{aligned}$$

where $\Gamma_m := \sqrt{\frac{K}{m^2+m}}$. We denote

$$\mathrm{GM(K)} := \{\mathbf{I}, \mathbf{A}_{jk}, \mathbf{B}_{jk}, \mathbf{C}_m\}_{1\leq j<k\leq K, 1\leq m\leq K-1}.$$

An observable $\mathcal{A} \in M_K(\mathbf{C})^{\otimes n}$ has expansion in the GM basis as

$$\mathcal{A} = \sum_{\alpha\in\Lambda_K^n} \widehat{\mathcal{A}}(\alpha) M_\alpha = \sum_{\alpha\in\Lambda_K^n} \widehat{\mathcal{A}}(\alpha) \bigotimes_{j=1}^n M_{\alpha_j}$$

for some index set Λ_K (so $\{M_\alpha\}_{\alpha\in\Lambda_K} = \mathrm{GM}(K)$). Letting $|\alpha| = |\{j : M_{\alpha_j} \neq \mathbf{I}\}|$, we say $\mathcal{A}$ is *of degree at most d* if $\widehat{\mathcal{A}}(\alpha) = 0$ for all α with $|\alpha| > d$.

In [SVZ24b] we find the Gell-Mann BH inequality enjoys a reduction to the Boolean cube BH inequality on $\{-1,1\}^{n(K^2-1)}$ and obtain the following.

Theorem 39 (Qudit Bohnenblust–Hille, Gell-Mann Basis)**.** *Fix any $K \geq 2$ and $d \geq 1$. There exists $C(d,K) > 0$ such that for all $n \geq 1$ and GM observable $\mathcal{A} \in M_K(\mathbf{C})^{\otimes n}$ of degree at most d, we have*

$$\|\widehat{\mathcal{A}}\|_{\frac{2d}{d+1}} \leq C(d,K)\|\mathcal{A}\|_{\mathrm{op}}. \tag{6.4.5}$$

Moreover, we have $C(d,K) \leq \left(\frac{3}{2}(K^2-K)\right)^d \mathrm{BH}^{\leq d}_{\{\pm 1\}}$.

In particular, for $K = 2$ we recover the main result of [VZ23] exactly.

Theorem (Low-degree Qudit Learning, restatement of Theorem 9)**.** *Let $\mathcal{A}$ be a degree-d observable on n qudits with $\|\mathcal{A}\|_{\mathrm{op}} \leq 1$. Then there is a collection S of product states such that with a number*

$$\mathcal{O}\Big(\Big(K\|\mathcal{A}\|_{\mathrm{op}}\Big)^{C\cdot d^2} d^2 \varepsilon^{-d-1} \log\Big(\tfrac{n}{\delta}\Big)\Big)$$

of samples of the form $(\rho, \mathrm{tr}[\mathcal{A}\rho])$, $\rho \sim \mathcal{U}(S)$, we may with confidence $1-\delta$ learn an observable $\widetilde{\mathcal{A}}$ with $\|\mathcal{A} - \widetilde{\mathcal{A}}\|_2^2 \leq \varepsilon$.

Here $\|\mathcal{A}\|_2$ denotes the normalized L_2 norm induced by the inner product $\langle A, B\rangle := K^{-n}\,\mathrm{tr}[A^\dagger B]$. Also, we choose to include explicit mention of $\|\mathcal{A}\|_{\mathrm{op}}$ here as it will be useful later. For applications it is natural to assume $\|\mathcal{A}\|_{\mathrm{op}}$ is bounded independent of n.

Proof. We will first pursue an L_∞ estimate of the Fourier coefficients in the Gell-Mann basis. To that end, sample $\vec{\boldsymbol{x}}_1, \ldots, \vec{\boldsymbol{x}}_s \overset{\text{iid}}{\sim} \{-1,1\}^{n(K^2-1)}$. As in the proof of Theorem 39, for any such $\vec{\boldsymbol{x}}$ we partition indices as $\vec{\boldsymbol{x}} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \in (\{-1,1\}^{K^2-1})^n$ with each $\boldsymbol{x}_\ell$, $1 \leq \ell \leq n$, corresponding to a qudit. Each $\boldsymbol{x}_\ell$ is further partitioned as

$$\boldsymbol{x}_\ell = (x^{(\ell)}, y^{(\ell)}, z^{(\ell)}) \in \{-1,1\}^{\binom{K}{2}} \times \{-1,1\}^{\binom{K}{2}} \times \{-1,1\}^{K-1},$$

with each sub-coordinate associated with a specific Gell-Mann basis element for that qudit.

Again for each $\vec{\boldsymbol{x}}$, for each qudit $\ell \in [n]$ form the mixed state

$$r(x^{(\ell)}, y^{(\ell)}, z^{(\ell)}) = \frac{1}{3\binom{K}{2}}\left(\sum_{1\leq j<k\leq K} A_{jk}^{(x_{jk}^{(\ell)})} + \sum_{1\leq j<k\leq K} B_{jk}^{(y_{jk}^{(\ell)})} + \sum_{m=1}^{K-1} z_m^{(\ell)} \tfrac{1}{\sqrt{2K}}\mathbf{C}_m + \tfrac{K-1}{2}\cdot\mathbf{I}\right).$$

Then we may define for $\vec{\boldsymbol{x}}$ the n qudit mixed state

$$r(\vec{\boldsymbol{x}}) = \bigotimes_{\ell=1}^{n} r(x^{(\ell)}, y^{(\ell)}, z^{(\ell)})$$

and consider the function

$$f_{\mathcal{A}}(\vec{x}) := \operatorname{tr}[\mathcal{A} \cdot r(\vec{x})].$$

Let $S(\alpha)$ denote the index map from the GM basis to subsets of $[n(K^2-1)]$. With these states in hand and in view of the identity

$$\widehat{f_{\mathcal{A}}}\big(S(\alpha)\big) = c^{|\alpha|}\widehat{\mathcal{A}}(\alpha) \quad \text{with} \quad c := \frac{\sqrt{K/2}}{3\binom{K}{2}} < 1,$$

we may now define the empirical Fourier coefficients

$$\mathcal{W}(\alpha) = c^{-|\alpha|} \cdot \frac{1}{s}\sum_{t=1}^{s} f_{\mathcal{A}}(\vec{x}_t) \prod_{j \in S(\alpha)} x_j \;=\; c^{-|\alpha|}\frac{1}{s}\sum_{t=1}^{s} \operatorname{tr}[\mathcal{A} \cdot r(\vec{x})] \prod_{j \in S(\alpha)} x_j.$$

The coefficient $\mathcal{W}(\alpha)$ is a sum of bounded i.i.d. random variables each with expectation $\widehat{\mathcal{A}}(\alpha)$, so by Chernoff we have

$$\Pr\left[|\mathcal{W}(\alpha) - \widehat{\mathcal{A}}(\alpha)| \geq \eta\right] \leq 2\exp(-s\eta^2 c^{|\alpha|}).$$

Taking the union bound, we find as before the chance of achieving ℓ_∞ error η is

$$\Pr\left[|\mathcal{W}(\alpha) - \widehat{\mathcal{A}}(\alpha)| < \eta \text{ for all } \alpha \text{ with } |\alpha| \leq d\right] \geq 1 - 2\sum_{k=0}^{d}\binom{n}{k}\exp(-s\eta^2 c^d),$$

which again we shall require to be $\geq 1-\delta$.

Applying Theorem 37 to obtain $\widetilde{\mathcal{W}}$ and recalling $\|\widehat{\mathcal{A}}\|_{\frac{2d}{d+1}} \leq \mathrm{BH}^{\leq d}_{\mathrm{GM}(K)}\|\mathcal{A}\|_{\mathrm{op}}$ we find the estimated operator

$$\widetilde{\mathcal{A}} := \sum_{\alpha} \widetilde{\mathcal{W}}(\alpha) M_\alpha$$

has L_2-squared error

$$\|\widetilde{\mathcal{A}} - \mathcal{A}\|_2^2 \overset{\text{(Parseval)}}{=} \sum_{\alpha}\left|\widetilde{\mathcal{W}}(\alpha) - \widehat{\mathcal{A}}(\alpha)\right|^2 \leq \left(e^5\eta^2 d(\mathrm{BH}^{\leq d}_{\mathrm{GM}(K)}\|\mathcal{A}\|_{\mathrm{op}})^{2d}\right)^{\frac{1}{d+1}}.$$

Thus to obtain error $\leq \varepsilon$ it suffices to pick $\eta^2 = \varepsilon^{d+1}e^{-5}d^{-1}(\mathrm{BH}^{\leq d}_{\mathrm{GM}(K)}\|\mathcal{A}\|_{\mathrm{op}})^{-2d}$, which entails by standard estimates that the algorithm will meet the requirements with a sample count of

$$s \geq e^6 K^{3/2} d^2 (\mathrm{BH}^{\leq d}_{\mathrm{GM}(K)}\|\mathcal{A}\|_{\mathrm{op}})^{2d} \log\left(\tfrac{2en}{\delta}\right)\varepsilon^{-d-1}. \qquad \square$$

Part II

Other applications

Published as

[Slo24] Joseph Slote. "Parity vs. AC^0 with Simple Quantum Preprocessing". In: *15th Innovations in Theoretical Computer Science Conference, ITCS 2024*. Ed. by Venkatesan Guruswami. Vol. 287. LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2024, 92:1–92:21. DOI: 10.4230/LIPICS.ITCS.2024.92.

[CNS25] Matthias Caro, Preksha Naik, and Joseph Slote. "Testing classical properties from quantum data". In: *TQC 2025 (to appear)* (2025). arXiv: 2411.12730 [quant-ph].

Chapter 7

Introduction

This part concerns two other applications of discrete harmonic analysis in (quantum) complexity theory.

The first, "Parity vs. AC^0 with simple quantum preprocessing" concerns concrete complexity theory of quantum circuits. Compared to the community's understanding of shallow *classical* circuits, small quantum circuits are still very mysterious. This paper introduces a circuit model, termed $\mathsf{AC}^0 \circ \mathsf{QNC}^0$, which is arguably the simplest possible model of constant-depth quantum computation that is still capable of solving nontrivial decision problems and makes initial progress in proving bounds on its capabilities. From a technical perspective, the work identifies certain basic mysteries related to nonlocal games that appear in quantum circuits under restriction. From a conceptual standpoint, the work suggests that while constant-depth quantum circuits have dramatic advantage over their classical counterparts for total *search* (or multi-output) problems, they may not have much advantage for decision problems—matching a dichotomy that was recently identified in the context of query complexity [YZ22]. This project also provided inspiration for later influential works [NPVY23; ADOY25] on a related model of constant-depth quantum computation known as QAC^0 (*Warning*: it is not known whether classical $\mathsf{AC}^0 \subset \mathsf{QAC}^0$ and is widely expected to be false).

The second application, "Testing classical properties with quantum data" introduces a novel mode of property testing and opens the door to a new category of quantum advantage. In classical complexity theory, property testing is traditionally the home of super-fast algorithms which use queries to determine whether a black-box Boolean function f has a certain property (for example, whether f is monotone, which has an $\mathcal{O}(\sqrt{n})$-query algorithm). Unfortunately, for applications to data analysis and machine-learning, when a testing algorithm only has access to random samples $\Big(x, f(x)\Big)$, it becomes much harder to test properties of f, often requiring just as much data as learning the whole function. (For the example of monotonicity, we have the lower bound of $2^{\Omega(\sqrt{n})}$ samples [Bla24]).

This work shows that for a wide range of properties, if a tester is provided

with data in certain *quantum* encodings, ultrafast testers are again available. And this quantum data shares "physical" properties with classical data, in that it can be collected far in advance and is independent of the property to be tested. There is much that remains to be understood about the power of quantum data for testing, and the work lays out several directions for further research.

We now discuss the specific contributions of these works in more detail.

Application I: Parity vs. AC^0 with simple quantum preprocessing

A recent line of work [BGK18; WKST19; GS20; BGKT20; WP23] has shown the unconditional advantage of constant-depth quantum computation, or QNC^0, over NC^0, AC^0, and related models of classical computation. Problems exhibiting this advantage include search and sampling tasks related to the parity function, and it is natural to ask whether QNC^0 can be used to help compute parity itself. Namely, we study $\mathsf{AC}^0 \circ \mathsf{QNC}^0$—a hybrid circuit model where AC^0 operates on measurement outcomes of a QNC^0 circuit—and we ask whether $\textsc{Par} \in \mathsf{AC}^0 \circ \mathsf{QNC}^0$.

We believe the answer is negative. In fact, we conjecture $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ cannot even achieve $\Omega(1)$ correlation with parity. As evidence for this conjecture, we prove:

- When the QNC^0 circuit is ancilla-free, this model can achieve only negligible correlation with parity, even when AC^0 is replaced with any function having LMN-like decay in its Fourier spectrum.

- For the general (non-ancilla-free) case, we show via a connection to nonlocal games that the conjecture holds for any class of postprocessing functions that has approximate degree $o(n)$ and is closed under restrictions. Moreover, this is true even when the QNC^0 circuit is given arbitrary quantum advice. By known results [BKT19], this confirms the conjecture for linear-size AC^0 circuits.

- Another approach to proving the conjecture is to show a switching lemma for $\mathsf{AC}^0 \circ \mathsf{QNC}^0$. Towards this goal, we study the effect of quantum preprocessing on the decision tree complexity of Boolean functions. We find that from the point of view of decision tree complexity, nonlocal channels are no better than randomness: a Boolean function f precomposed with

an n-party nonlocal channel is together *equal* to a randomized decision tree with worst-case depth at most $\mathrm{DT}_{\mathrm{depth}}[f]$.

Taken together, our results suggest that while QNC^0 is surprisingly powerful for search and sampling tasks, that power is "locked away" in the global correlations of its output, inaccessible to simple classical computation for solving decision problems.

Application II: Testing classical properties from quantum data

Many classes of Boolean functions can be tested much faster than they can be learned. However, this speedup tends to rely on *query* access to the function f. When access is limited to random samples $(x, f(x))$—the *passive* testing model and a natural setting for data science—testing can become much harder. Here we introduce *quantum passive testing* as a quantum version of this "data science scenario": quantum algorithms that test properties of a function f solely from quantum data in the form of copies of the function state $|f\rangle \propto \sum_x |x, f(x)\rangle$. Just like classical samples, function states are independent of the property of interest and can be collected well in advance.

Quantum advantage in testing from data: an emerging theme.

For three well-established properties—monotonicity, symmetry, and triangle-freeness—we show passive quantum testers are unboundedly- or super-polynomially better than their classical passive testing counterparts, and in fact are competitive with classic *query*-based testers in each case. Existing quantum testers for k-juntas and linearity can be interpreted as passive quantum testers too and exhibit the same phenomena.

Inadequacy of Fourier sampling.

Our new testers use techniques beyond quantum Fourier sampling, and it turns out this is necessary: we show a certain class of bent functions can be tested from $\mathcal{O}(1)$ function states but has a sample complexity lower bound of $2^{\Omega(\sqrt{n})}$ for any tester relying exclusively on Fourier and classical samples.

Classical queries vs. quantum data.

Our passive quantum testers are competitive with classical *query*-based testers, but this isn't universal: we exhibit a testing problem that can be solved from $\mathcal{O}(1)$ classical queries but requires $\Omega(2^{n/2})$ function state copies. The

FORRELATION problem provides a separation of the same magnitude in the opposite direction, so we conclude that quantum data and classical queries are "maximally incomparable" resources for testing.

Towards lower bounds.

We also begin the study of *lower bounds* for testing from quantum data. For quantum monotonicity testing, we prove that the ensembles of [Gol+00; Bla24], which give exponential lower bounds for classical sample-based testing, do not yield any nontrivial lower bounds for testing from quantum data. New insights specific to quantum data will be required for proving copy complexity lower bounds for testing in this model.

Chapter 8

Parity vs. AC^0 with simple quantum preprocessing

In 2017, Bravyi, Gosset, and König [BGK18] proved a breakthrough unconditional separation between constant-depth quantum circuits, or QNC^0, and constant-depth bounded fan-in classical circuits, or NC^0. The authors showed that for a certain search problem solvable by QNC^0 circuits, any randomized NC^0 circuit solving the same problem with high probability must have logarithmic depth. The realization that unconditional proofs of quantum advantage were possible—albeit over weak models of classical computation—inspired an exciting series of results strengthening and generalizing the work of Bravyi, Gosset, and König. There are now separations against stronger classical circuit models such as constant depth circuits with unbounded fan-in, or AC^0 [WKST19], average-case separations [Gal20], separations between more intricate interactive models [GS20], separations that remain even for quantum circuits subject to noise (*e.g.,* [BGKT20]), and separations for sampling problems with no input [WP23], among others.

Although these separations are for comparatively weak models of computation, they are concrete non-oracle, non-query separations, and are free from complexity-theoretic assumptions, making them important companions to the query complexity and conditional separations studied since the founding of quantum computer science. One notable feature of these QNC^0 separations, however, is that they are all for search or sampling problems; decision separations appear to be absent from this list.

On the surface, there is a somewhat trivial reason for this: QNC^0 cannot solve interesting decision problems alone. Indeed, any single output qubit in a constant-depth quantum circuit can only depend on constantly-many input qubits, so any QNC^0 circuit with one output bit may be simulated by randomized NC^0. However, this "lightcone barrier" may be removed by instead measuring all qubits in the quantum circuit and then applying a classical Boolean function f to the result. As long as f depends on all of its inputs, it might be possible for f to leverage QNC^0's search and sampling prowess for decision-making ends. Given Bene Watts et al.'s search separation between QNC^0 and AC^0 [WKST19], a natural class of Boolean functions to choose for

this postprocessing is AC^0 itself. This gives rise to the following definition, which does not appear to have been studied before.

Definition. *Let* $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ *denote the model of computation composed of a* QNC^0 *circuit* $\mathcal{C}$*, followed by a computational basis measurement, and then an* AC^0 *function* f *applied to the result. This process defines the randomized Boolean function* $f \circ \mathcal{C} : \{0,1\}^n \to \mathcal{M}(\{-1,1\})$ *from the hypercube to the set* $\mathcal{M}(\{-1,1\})$ *of probability measures on* $\{-1,1\}$*.*

In this chapter we take a QNC^0 circuit to be a polynomial-size constant-depth quantum circuit composed of arbitrary 2-qubit unitary gates. Ancilla qubits are allowed and are initialized in the state $|0^m\rangle$ for $m \in \mathrm{poly}(n)$. No geometric locality or clean computation constraints are assumed. A formal definition appears later as Definition 8.2.

Certainly $\mathsf{QNC}^0 \subseteq \mathsf{AC}^0 \circ \mathsf{QNC}^0$, so the search separation between QNC^0 and AC^0 in Bene Watts et al. is also a search separation between $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ and AC^0. Moreover, this modification obviates the lightcone barrier mentioned above and allows us to ask meaningful questions about decision separations between concrete models of quantum and classical computation.

Specifically, Bene Watts et al. [WKST19] show exponential advantage of QNC^0 over AC^0 for (a variant of) the "parity halving problem":

> *Parity halving.* Given $x \in \{0,1\}^n$ with the promise $|x| \equiv 0 \mod 2$, output any even string if $|x| \equiv 0 \mod 4$ and any odd string otherwise.

Given the form of this problem, it is natural to ask whether parity is itself computable by a hybrid model such as $\mathsf{AC}^0 \circ \mathsf{QNC}^0$.

Before summarizing our progress on this question, we pause to note another reason to study $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ coming from the rich subject of quantum-classical interactive proofs. A central project in this area is the classical verification of quantum computations [GKK18]. In a landmark 2018 work, Mahadev gave a cryptographic protocol for this task [Mah18]; however, whether or not this task may be accomplished without cryptographic hardness assumptions remains open despite many efforts [GKK18]. It therefore makes sense to consider the question in simpler contexts, such as where the prover and verifier are replaced with QNC^0 and AC^0 respectively and interact for constantly-many rounds to

establish the correctness of a QNC^0 computation. With this perspective we see that $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ models the first round of interaction in such a proof system.

Parity vs. $\mathbf{AC^0 \circ QNC^0}$: Overview and organization

We conjecture that $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ cannot approximate parity ($\textsc{Par}_n$) on average, over both choice of uniformly random input $x \sim \mathcal{U}(\{0,1\}^n)$ and the randomness in $f \circ \mathcal{C}$. It is convenient to take $\textsc{Par}$ and $f \circ \mathcal{C}$ to be (± 1)-valued and phrase this in terms of the correlation

$$\mathop{\mathbb{E}}_{x}[(f \circ \mathcal{C})(x) \cdot \textsc{Par}(x)],$$

proportional to the advantage of $f \circ \mathcal{C}$ over random guessing for computing parity.

Conjecture 3. *$\mathsf{AC}^0 \circ \mathsf{QNC}^0$ cannot achieve correlation $\Omega(1)$ with the parity function. That is, fix a polynomial size bound $p(n)$ and constant depth d. Then for all sequences $\{(f_n, \mathcal{C}_n)\}_n$ of circuits such that $\mathrm{size}(f_n), \mathrm{size}(\mathcal{C}_n) \leq p(n)$ and $\mathrm{depth}(f_n), \mathrm{depth}(\mathcal{C}_n) \leq d$, we have*

$$\mathop{\mathbb{E}}_{x}[(f_n \circ \mathcal{C}_n)(x) \cdot \textit{PAR}_n(x)] \to 0 \quad \textit{as} \quad n \to \infty\,.$$

Although proving correlation bounds against AC^0 is a well-understood topic with many techniques (among them Håstad's switching lemma [Hås86] and Razborov-Smolensky [Raz87; Smo87a]), when QNC^0 precomputation is added these approaches cannot be used directly. The pursuit of new techniques leads us to connections with many-player nonlocal games, approximate degree bounds, and new directions for generalizing Håstad's switching lemma. Evidence for Conjecture 3 is laid out as follows.

The ancilla-free case

In Section 8.1 we prove Conjecture 3 when QNC^0 is restricted to be ancilla-free. A key feature of such QNC^0 circuits is that they correspond to unitary transformations, and we find in this case the correlation of $f \circ \mathcal{C}$ with $\textsc{Par}$ is controlled by the Fourier tail of f. Recall the k^{th} *Fourier tail* of a Boolean function f is given by

$$\mathbf{W}^{\geq k}[f] := \textstyle\sum_{|S| \geq k} \widehat{f}(S)^2\,.$$

Appealing to the Linial-Mansour-Nisan-type (LMN-type) estimates of the Fourier tail of AC^0 [LMN93b], we obtain the following strong correlation bound.

Theorem 40 (Ancilla-free QNC^0, general AC^0 case)**.** *If $\mathcal{C}$ is an ancilla-free QNC^0 circuit and f is an AC^0 function then*

$$\mathop{\mathbb{E}}_{x}[(f \circ \mathcal{C})(x) \cdot P_{AR_n}(x)] \leq 2^{-n/\mathrm{polylog}(n)}.$$

This is proved as Corollary 44 in Section 8.1. The full statement holds for any Boolean function f with sufficient decay in the tail of the Fourier spectrum, including those outside of AC^0.

However, as we explain in the end of Section 8.1, the proof technique of Theorem 40 cannot extend to the case of general QNC^0 and we must find a different approach.

Reducing to nonlocal games

To move beyond ancilla-free QNC^0, in Section 8.2 we reduce Conjecture 3 to a question about the value of a certain class of nonlocal games, which we call *n-player parity games* and which are parameterized by a postprocessing Boolean function f. Through a connection to the notion of k-wise indistinguishability introduced in [BIVW16], we show the quantum value of a parity game is controlled by the approximate degree of the associated f.

Recall for $\varepsilon > 0$ the *ε-approximate degree* of a $(0,1)$-valued[1] Boolean function f is given by

$$\widetilde{\deg}_{\varepsilon}[f] = \min\{\deg(g) \mid g : \{0,1\}^n \to \mathbb{R} \text{ a polynomial with } \|f - g\|_{\infty} \leq \varepsilon\}.$$

Of course, $\widetilde{\deg}_{\varepsilon}[f] \leq n$ for any n-variate f and $\varepsilon > 0$. By convention $\widetilde{\deg}[f] := \widetilde{\deg}_{1/3}[f]$. A *function class* $\mathcal{F} = (\mathcal{F}_n)_{n \geq 1}$ is a sequence of sets $\mathcal{F}_n$ of n-variate Boolean functions, and we extend approximate degree to function classes via $\widetilde{\deg}[\mathcal{F}](n) := \max_{f \in \mathcal{F}_n} \widetilde{\deg}[f]$. With this notation, we have the following theorem.

Theorem 41 (Corollary 51, Section 8.2)**.** *Suppose function class $\mathcal{F}$ is closed under inverse-polynomial-sized restrictions. Then if $\widetilde{\deg}[\mathcal{F}] \in o(n)$, $\mathcal{F} \circ \mathsf{QNC}^0$ cannot achieve $\Omega(1)$ correlation with P_{AR_n}, even if QNC^0 is given arbitrary quantum advice.*

It follows from Theorem 41 that Conjecture 1 would be confirmed in full generality if $\widetilde{\deg}[\mathsf{AC}^0] \in o(n)$, a notorious open problem [BT22]. Such a bound

[1] For (± 1)-valued f, we use the same definition after making the standard identification $+1 \mapsto 0, -1 \mapsto 1$.

is already known for large subclasses of AC^0, however: for example, for AC^0 circuits of size $\mathcal{O}(n)$ (termed LC^0), we may appeal to the recent bounds of [BKT19] to conclude:

Theorem 42 (General QNC^0, linear-size AC^0 case)**.** *Suppose $f \in \mathsf{AC}^0$ has size $\mathcal{O}(n)$. Then $f \circ \mathsf{QNC}^0$ achieves correlation at most $1/\operatorname{poly}(n)$ with $\textsc{Par}_n$. This holds even if QNC^0 is given arbitrary quantum advice. That is,*

$$\mathbb{E}[(\mathsf{LC}^0 \circ \mathsf{QNC}^0/\mathsf{qpoly}) \cdot \textsc{Par}_n] \in \operatorname{negl}(n)\,.$$

(This is proved as Corollary 52 in Section 8.2).

Is the difficulty of proving approximate degree bounds for AC^0 a barrier for resolving Conjecture 3? It seems unlikely: the reduction to approximate degree bounds is via a series of substantial relaxations and it would be surprising if all the required converses held. In fact, we conclude Section 8.2 with a self-contained approximation theory question (Question 2) concerning a notion of blockwise approximate degree which may be easier to solve than $\widetilde{\deg}[\mathsf{AC}^0]$ but would still imply Conjecture 3.

Towards an $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ switching lemma

In Section 8.3 we chart a different route to resolving Conjecture 3, aiming to prove a switching lemma for our hybrid $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ circuits. Recall that Håstad's original switching lemma is used to argue that (very roughly) randomly fixing a large fraction of inputs to an AC^0 circuit with high probability yields a function that can be computed by a shallow decision tree. At the same time, $\textsc{Par}$ retains maximum decision tree complexity under the same restrictions, so this leads to AC^0 correlation bounds.

In comparison to Håstad's switching lemma and its descendants, a challenge with $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ circuits is that QNC^0 can correlate, spread out, and bias random restrictions before they reach the bottom layer of DNFs or CNFs in the AC^0 circuit. If QNC^0 were replaced with randomized NC^0 this problem could be readily addressed by considering each deterministic circuit in the distribution, applying standard arguments there, and computing the expected correlation with parity across circuits in the distribution. But unlike randomized computation, and as discussed *e.g.,* in [AIK22], a recurring theme in quantum complexity theory is the impossibility of "pulling out the quantumness" from a quantum circuit.

Contrary to this theme, however, we show that when QNC^0 is replaced by an n-party nonlocal channel $\mathcal{N}$, it *is* possible to pull out the quantumness in a particular sense:

Theorem (Theorem 53, restated)**.** *Let $f : \{0,1\}^m \to \{0,1\}$ be any Boolean function and consider an n-party nonlocal channel $\mathcal{N}$, where the i^{th} party receives one bit and responds with $m_i \geq 0$ bits, such that $\sum_i m_i = m$. Then the random function $f \circ \mathcal{N}$ is equal to a randomized decision tree Γ such that $\mathrm{depth}(T) \leq \mathrm{DT}_{\mathrm{depth}}[f]$ for all $T \in \mathrm{Supp}(\Gamma)$.*

(This theorem is proved in Section 8.3 as Theorem 53.) By an *n-party nonlocal channel* we mean the channel corresponding to a quantum strategy in an n-player nonlocal game: parties receive one bit of input each and may measure disjoint systems of a shared quantum state as part of their responses, but they are not allowed to communicate. A formal definition appears as Definition 8.2. In fact, Theorem 53 is true not only for nonlocal channels, but for any channel where parties obey the no-signaling property; that is, the output of any subset $S \subset [n]$ of the parties is a function only of the inputs to those parties in S. A formal definition of no-signaling channels appears as Definition 8.3.

The regime where Theorem 53 is truly interesting is when $\mathrm{DT}_{\mathrm{depth}}[f] \geq \log(n)$. Then f may depend on all the input coordinates and (potentially) make great use of the processing power afforded by no-signaling channels. Theorem 53 says that to the contrary, precomposition of f by any no-signaling channel has no effect on the (randomized) decision tree complexity of f.

How does Theorem 53 connect to Conjecture 3? As we detail in Section 8.2, the replacement of QNC^0 by the channel $\mathcal{N}$ is essentially without loss of generality from the point of view of Conjecture 3. Unfortunately, however, AC^0 circuits can easily have maximum decision tree complexity, so Theorem 53 cannot be immediately applied. Instead, we believe Theorem 53 stands as a striking example of the inability of classical postprocessing to make use of the search and sampling power of quantum and super-quantum models of computation. Additionally, we hope that this theorem's proof technique, which involves tracking the interplay between a decision tree for f and the no-signaling channel $\mathcal{N}$, represents the style of argument that could eventually lead to a switching lemma for $\mathsf{AC}^0 \circ \mathsf{QNC}^0$.

Outlook

Taken together, these results suggest QNC^0 cannot render its power in a way AC^0 or other simple models of classical computation can access for the purpose of making decisions. Several questions for further research are posed in Section 8.4.

Related work

Unlike the quantum-classical separations surveyed in the introduction, which show quantum upper bounds and classical lower bounds, this chapter aims to prove a lower bound against a concrete model of quantum computation. The pursuit of lower bounds against quantum circuits for computational problems is a nascent area and very little is known.

One quantum circuit model where lower bounds have received some concerted study is QAC^0 [Moo99; H03; PFGT20; Ros21; NPVY23]. A superset of QNC^0 circuits, QAC^0 additionally allows for arbitrarily-large Toffoli gates,

$$|x_1, \ldots, x_k, x_{k+1}\rangle \mapsto \left|x_1, \ldots, x_k,\, x_{k+1} \oplus (\wedge_{i=1}^{k} x_i)\right\rangle ,$$

which are quantum analogues of classical AND gates with unbounded fan-in. In this setting correlation with parity is also a central open question, and there is growing evidence that QAC^0 cannot achieve $\Omega(1)$ correlation with parity either. Recent work has shown negligible correlation bounds between QAC^0 and parity when a) the QAC^0 circuit is restricted to depth 2 [Ros21], and b) when the QAC^0 circuit is of any depth d and is restricted to $\mathcal{O}(n^{1/d})$-many ancillas [NPVY23]. In fact, the second result is a corollary to a Pauli-basis analogue of the LMN theorem for the same subclass of QAC^0 [NPVY23].

The relationship between QAC^0 and $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ is rather unclear, and they are likely incomparable as decision classes. In fact, as far as we know, it is even open whether $\mathsf{AC}^0 \subseteq \mathsf{QAC}^0$, let alone whether $\mathsf{AC}^0 \circ \mathsf{QNC}^0 \subseteq \mathsf{QAC}^0$ (noting the trivial containment $\mathsf{AC}^0 \subseteq \mathsf{AC}^0 \circ \mathsf{QNC}^0$).

The difficulty in comparing these models stems from a subtlety concerning the difference between unbounded fan-in and unbounded fan-out when implemented coherently. AC^0 circuits have no restriction on the *fan-out* of their gates, while the definition of QAC^0 appears to strongly limit outward propagation of information. If one augments QAC^0 with the so-called *fan-out gate*—which is a CNOT gate with any number of target qubits,

$$|x_1, \ldots, x_k\rangle \mapsto |x_1, x_1 \oplus x_2, \ldots, x_1 \oplus x_k\rangle ,$$

one obtains the circuit model QAC^0_f, and it is known QAC^0_f can compute parity exactly in depth 3 [Moo99]. In view of existing lower bounds against QAC^0, it is expected that QAC^0 is strictly contained in QAC^0_f, and assuming this holds we immediately have that the function version of $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ is not in the function version of QAC^0. This follows, for example, from the fact that multi-output AC^0 circuits easily implement the classical reversible fan-out gate, $(x_1, \ldots, x_k) \mapsto (x_1, x_1 \oplus x_2, \ldots, x_1 \oplus x_k)$. It is safe to say the interaction of nonlocal gates with QNC^0—whether that interaction is coherent as in QAC^0 and QAC^0_f, or preceded by measurement as in $\mathsf{AC}^0 \circ \mathsf{QNC}^0$—is only beginning to be understood.

A separate area where concrete quantum circuit lower bounds have been very successfully developed is for *state preparation* problems. We do not attempt a survey here, but just mention they were crucial to the resolution of the NLTS conjecture [ABN23] and make use of ideas from error correction, which partially originate in sampling lower bounds from classical complexity [LV11]. However, it is not clear how to transfer these methods to quantum circuit lower bounds for computational problems in the $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ model.

8.1 Lower bounds when QNC^0 is ancilla-free

Here we show any Boolean function f with small Fourier tail retains a small top-degree coefficient when composed with ancilla-free QNC^0. By the celebrated work of Håstad [Hås86] and Linial, Mansour, and Nisan [LMN93b], any $f \in \mathsf{AC}^0$ is an example—but this theorem addresses a broader set of functions. On the other hand, as we discuss at the end of the section, once ancillas are allowed, the theorem no longer holds for such a general class of functions.

Recall a function $f : \{-1,1\}^n \to \mathbf{R}$ admits a unique *Fourier decomposition*

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

where $\chi_S(x) := \prod_{i \in S} x_i$ is the S^{th} Fourier character (see *e.g.,* [ODo14] for more). We will later make use of the familiar Plancherel theorem, which states for any $f, g : \{-1,1\}^n \to \mathbb{R}$ that

$$\mathop{\mathbb{E}}_x[f(x)g(x)] = \sum_{S \subseteq [n]} \widehat{f}(S)\widehat{g}(S)\,.$$

Let us briefly connect this perspective to quantum observables. Given a Boolean function $f : \{\pm 1\}^n \to \mathbf{R}$ we define its *Von Neumann observable* as

$$M_f := \sum_x f(x) \, |x\rangle\langle x| \, .$$

An identity we will use is

$$M_{\chi_S} = Z^S \, ,$$

where the operator Z here is the Pauli operator $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$, and generally for any 1-qubit operator A we use the notation

$$A^S := \bigotimes_i \begin{cases} A & \text{if } i \in S \\ \mathbf{1} & \text{otherwise.} \end{cases}$$

Any Von Neumann observable M (that is, any Hermitian operator) has expectation value on state ρ given by

$$\langle M \rangle_\rho := \mathrm{tr}[M\rho] \, ,$$

and when $M = M_f$ and $x \in \{0,1\}^n$ we note the identity

$$\langle M_f \rangle_x := \langle M_f \rangle_{|x\rangle\langle x|} = f(x) \, .$$

With this notation, we prove the following.

Theorem 43 (Correlation bound for ancilla-free QNC^0)**.** *Let $f : \{\pm 1\}^n \to \mathbf{R}$ and U an ancilla-free QNC^0 circuit of depth t. Then the correlation of $f \circ U$ and $\textsc{Par}$ is bounded as*

$$\mathop{\mathbb{E}}_x [\langle U^\dagger M_f U \rangle_x \cdot \textit{PAR}_n(x)] \;\leq\; \left(\mathbf{W}^{\geq 2^{-t} n}[f] \right)^{1/2} .$$

For example, when f is an AC^0 circuit, we may use an LMN-type Fourier concentration bound, such as from [Tal17], to a obtain:

Corollary 44. *If U is an ancilla-free QNC^0 n-qubit circuit of depth t, and $f : \{\pm 1\}^n \to \{\pm 1\}$ is implemented by an AC^0 circuit of depth d and size s, we have*

$$\mathop{\mathbb{E}}_x [\langle U^\dagger M_f U \rangle_x \cdot \textit{PAR}_n(x)] \leq \sqrt{2} \cdot \exp\left(\frac{-n}{2^{t+1} \mathcal{O}(\log s)^{d-1}} \right).$$

The proof of Theorem 43 relies on two brief lemmas. The first says that when measuring correlations, we could just as well have compared the correlation of f alone to the random function $\textsc{Par}_n \circ U^\dagger$, defined by applying $\textsc{Par}_n$ to the output of $U^\dagger |x\rangle$.

Lemma 45 (Symmetry of correlation)**.** *Let $f, g : \{\pm 1\}^n \to \{\pm 1\}$ and U any n-qubit unitary. Then*

$$\begin{aligned}\mathop{\mathbb{E}}_x[\langle U^\dagger M_f U\rangle_x \cdot g(x)] &= \mathop{\mathbb{E}}_x[f(x) \cdot \langle U M_g U^\dagger\rangle_x] \\ &= 2^{-n} \operatorname{tr}[M_f U M_g U^\dagger]\,.\end{aligned}$$

Proof. Expanding the trace we have

$$\begin{aligned}\operatorname{tr}[M_f U M_g U^\dagger] &= \sum_z \langle z| \Big(\textstyle\sum_y f(y)|y\rangle\langle y|\Big) U \Big(\textstyle\sum_x g(x)|x\rangle\langle x|\Big) U^\dagger |z\rangle \\ &= \sum_{x,y,z} f(y)g(x)\, \langle z|y\rangle\, \langle y| U^\dagger |x\rangle\, \langle x| U |z\rangle \\ &= \sum_{x,y} f(y)g(x)\, \langle y| U |x\rangle\, \langle x| U^\dagger |y\rangle\,, \qquad (8.1.1)\end{aligned}$$

while expanding the expectations we see

$$\mathop{\mathbb{E}}_x[\langle U^\dagger M_f U\rangle_x \cdot g(x)] = \frac{1}{2^n}\sum_{x,y} f(y)g(x)\, \langle x| U^\dagger |y\rangle\, \langle y| U |x\rangle = \mathop{\mathbb{E}}_y[f(y) \cdot \langle U M_g U^\dagger\rangle_y]\,.$$

Identifying the center expression with (a multiple of) (8.1.1) and changing variables completes the lemma. □

The second lemma roughly says when Fourier characters Z_S and Z_T correspond to sets S, T of very different cardinality, they remain orthogonal (with respect to the inner product $\langle A, B\rangle = \operatorname{tr}[A^\dagger B]$) after an application of U.

Lemma 46 (Lightcone lemma)**.** *Suppose U is a depth-t ancilla-free quantum circuit and $|S|2^t < n$. Then*

$$\operatorname{tr}[Z_{[n]} U Z_S U^\dagger] = 0.$$

Proof. The number of qubits on which Z_S acts nontrivially at most doubles upon conjugation by each layer in U. Therefore the number of non-identity coordinates in $UZ_SU^\dagger$ is at most $|S|2^t$. Now if $|S|2^t < n$, then there is at least one coordinate j such that $UZ_SU^\dagger = V_{[n]\setminus j} \otimes \mathbf{1}_j$ for some $(n-1)$-qubit unitary $V_{[n]\setminus j}$, so

$$\operatorname{tr}[Z_{[n]} U Z_S U^\dagger] = \operatorname{tr}[Z_{[n]}(V_{[n]\setminus j} \otimes \mathbf{1}_j)] = \operatorname{tr}[Z_{[n]\setminus j} V_{[n]\setminus j}] \cdot \operatorname{tr}[Z] = 0$$

because Z is traceless. □

With these lemmas in hand, we can give the proof of Theorem 1 in a single display:

Proof of Theorem 43.

$$\begin{aligned}
\mathop{\mathbb{E}}_{x}[\langle U^\dagger M_f U\rangle_x \cdot \chi_{[n]}(x)] &= \mathop{\mathbb{E}}_{x}[f(x)\cdot\langle UZ_{[n]}U^\dagger\rangle_x] && \text{(Lemma 45)}\\
&= \sum_{S\subseteq[n]} \widehat{f}(S)\cdot\widehat{\langle UZ_{[n]}U^\dagger\rangle}(S) && \text{(Plancherel)}\\
&= \sum_{S\subseteq[n]} \widehat{f}(S)\underbrace{\mathop{\mathbb{E}}_{x}[\langle UZ_{[n]}U^\dagger\rangle_x\cdot\chi_S(x)]}\\
&\qquad\qquad = 2^{-n}\operatorname{tr}[Z_{[n]}U^\dagger Z_S U] && \text{(Lemma 45)}\\
&\qquad\qquad = 0 \quad \text{if } |S|2^t < n && \text{(Lemma 46)}\\
&= \sum_{\substack{S\subseteq[n]\\ |S|\geq 2^{-t}n}} \widehat{f}(S)\cdot\widehat{\langle U^\dagger Z_{[n]}U\rangle}(S)\\
&\leq \left(\sum_{|S|\geq 2^{-t}n}\widehat{f}(S)^2\right)^{1/2}\left(\sum_{|S|\geq 2^{-t}n}\widehat{\langle U^\dagger Z_{[n]}U\rangle}(S)^2\right)^{1/2} && \text{(Cauchy-Schwarz)}\\
&\leq \left(\mathbf{W}^{\geq 2^{-t}n}[f]\right)^{1/2}. && \square
\end{aligned}$$

One may ask whether this proof approach extends to QNC^0 circuits with ancillas. Although it might be possible to prove slight generalizations, we present an example demonstrating that any proof approach using an LMN-type theorem as a black box will fail for general QNC^0 circuits. This is essentially because functions with Fourier decay are not closed under composition.

Example 47. *Consider the following "Trojan horse" function on an even number of bits $n = 2m$:*

$$\begin{aligned}
h : \{\pm 1\}^{2m} &\to \{\pm 1\}\\
x &\mapsto \begin{cases} \chi_{[m]}(x) & \text{if } x_{[m+1,2m]} = 11\cdots 1\\ 1 & \text{otherwise}. \end{cases}
\end{aligned}$$

By direct computation one finds the Fourier coefficients of h are given by

$$\widehat{h}(S) = \begin{cases} 1-2^{-m} & S = \emptyset,\\ -2^{-m} & S\subseteq[m], S\neq\emptyset\\ 2^{-m} & [m+1,2m]\subseteq S\\ 0 & \text{otherwise}. \end{cases}$$

This means for any $t \geq 1$, the t^{th} Fourier tail of h is $\mathbf{W}^{\geq t}[h] \in \mathcal{O}(2^{-n/2})$. Thus by Theorem 43, for any ancilla-free QNC^0 circuit $\mathcal{C}$, $h \circ \mathcal{C}$ has negligible correlation with parity.

On the other hand, consider the (deterministic) function $C : \{\pm 1\}^m \to \{\pm 1\}^{2m}$ given by $x \mapsto x11\cdots 1$. Certainly C can be implemented in QNC^0, and we have $h \circ C = \chi_{[m]} = PAR_m$.

This example shows that exponential Fourier decay of f is not sufficient to entail Conjecture 3 for general $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ circuits. We must take a different approach that exploits finer structural properties of AC^0 and QNC^0.

8.2 Lower bounds against $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ via nonlocal games

Here we pass from QNC^0 to nonlocal games to make an argument that works for general QNC^0. First let us fix ideas about QNC^0.

Definition (QNC^0). *An n-input, depth-d QNC^0 circuit $\mathcal{C}$ is a quantum circuit composed of d layers of arbitrary 2-qubit gates, acting on an* input register *of n qubits and an* ancilla register *of $m \in \mathrm{poly}(n)$ qubits initialized to $|0^m\rangle$. Via measurement of the entire output of $\mathcal{C}$ in the computational basis, the circuit $\mathcal{C}$ effects a randomized mapping from n bits of input to $n + m \in \mathrm{poly}(n)$ bits of output. A* QNC^0 circuit with v qubits of quantum advice, *has v out of m ancilla qubits initialized to a v-qubit state, not necessarily a product state. For general $v \in \mathrm{poly}(n)$, this is denoted by the class $\mathsf{QNC}^0/\mathsf{qpoly}$.*

We will show a reduction from QNC^0 circuits to nonlocal channels.

Definition. *(Nonlocal channel) Let $n, k \geq 1$ and $m \geq 0$. An (n, k, m) nonlocal channel is the randomized mapping defined by a quantum strategy in a nonlocal game where n parties receive one bit of input each and respond with k bits each, along with a referee response of m bits.*

Concretely, each party $i \in [n]$ is assigned a local Hilbert space $\mathcal{H}_i$ and for each $b \in \{0,1\}^n$, a POVM

$$M_{(i,b)} = \left\{ M_{(i,b)}^y : y \in \{0,1\}^k \right\}$$

on $\mathcal{H}_i$. There is also a referee Hilbert space $\mathcal{H}_{\mathrm{ref}}$ with a fixed POVM

$$M_{\mathrm{ref}} = \left\{ M_{\mathrm{ref}}^y : y \in \{0,1\}^m \right\}.$$

The definition of the nonlocal channel is completed by a choice of shared state $|\psi\rangle \in \left(\bigotimes_{i=1}^n \mathcal{H}_i \right) \otimes \mathcal{H}_{\mathrm{ref}}$ and works as follows. Upon receipt of an input string

$x \in \{0,1\}^n$, the n players and one referee perform the joint measurement $\left(M_{(1,x_1)}, \ldots, M_{(n,x_n)}, M_{\text{ref}}\right)$ on $|\psi\rangle$, resulting in the outcomes $y_1, \ldots, y_n$, and y_{ref}. The output of the channel is the $(nk+m)$-bit string $y = y_1 || \cdots || y_n || y_{\text{ref}}$.

Definition (No-signaling channel)**.** *An (n,k,m) no-signaling channel is defined analogously, except the correlations among parties may be general no-signaling correlations. (A very detailed definition of such channels is given in Definition 8.3.)*

Definition (Parity games)**.** *Let n, k, m be fixed and consider $f : \{0,1\}^{kn+m} \to \{0,1\}$. The (n,k,f) parity game is played by n entangled and non-communicating players, with the i^{th} player receiving input bit x_i from x drawn uniformly from $\{0,1\}^n$. A (quantum) parity game* strategy *is an (n,k,m) nonlocal channel with output string y. Players win when $f(y) = \textsc{Par}(x)$. We say a parity game strategy has* advantage ε *if its winning probability is at least $1/2 + \varepsilon$.*

As a final piece of notation, for Boolean f let $\neg f$ denote its negation. We are prepared to give our reduction to parity games.

Lemma 48. *Fix $n \geq 1, m \in \text{poly}(n)$, let $\mathcal{C}$ be a n-qubit, depth-d QNC^0 circuit with m ancilla and arbitrary quantum advice, and let $f : \{0,1\}^{n+m} \to \{0,1\}$ be any Boolean function. Suppose $f \circ \mathcal{C}$ has correlation ε with $\textsc{Par}_n$. Then for some $n' \geq n/(2^d+1)$ there is a quantum strategy for the $(n', 2^d, f)$ or $(n', 2^d, \neg f)$ parity game with advantage $\varepsilon/2$.*

Proof. Suppose $f \circ \mathcal{C}$ has correlation ε with $\textsc{Par}$. For each input qubit j denote by L_j the set of output qubits in the forward lightcone of j. Consider the graph with vertices the input qubits $[n]$ and edges drawn between qubits j and k when L_j and L_k have nonempty intersection. Then G has degree at most 2^d, so there exists an independent set $S \subseteq [n]$ of size at least $n/(2^d+1)$.

For each $y \in \{0,1\}^{S^c}$, define the circuit $\mathcal{C}_y$ to be $\mathcal{C}$ but where for $j \in S^c$, the j^{th} input is hardcoded to y_j. Then $\mathcal{C}_y$ is a circuit on at least $n/(2^d+1)$ variables such that the forward lightcones of input qubits are pairwise disjoint. Such a circuit defines an $(n', 2^d, m')$ nonlocal channel for some $n' \geq 2^{-d} + 1$ and $m' = n + m - n'2^d$. (Note this m' is without loss of generality because we may freely assign a player some output bits of the referee if their lightcone is smaller than 2^d.)

As a result, this restriction represents a strategy for the $(n', 2^d, f)$ parity game. Moreover, we have

$$\begin{aligned}\mathop{\mathbb{E}}_{x}[(f \circ \mathcal{C})(x) \cdot \textsc{Par}(x)] &= \mathop{\mathbb{E}}_{y \sim \{0,1\}^{S^c}} \mathop{\mathbb{E}}_{z \sim \{0,1\}^{S}} [f \circ \mathcal{C}_y(z) \cdot \textsc{Par}(y||z)] \\ &= \mathop{\mathbb{E}}_{y \sim \{0,1\}^{S^c}} \textsc{Par}(y) \mathop{\mathbb{E}}_{z \sim \{0,1\}^{S}} [f \circ \mathcal{C}_y(z) \cdot \textsc{Par}(z)] \,.\end{aligned}$$

Therefore since $f \circ \mathcal{C}$ has ε correlation with parity on n bits, for at least one y, $f \circ \mathcal{C}_y$ or $\neg f \circ \mathcal{C}_y$ must have at least ε correlation (in magnitude) with parity on n/d bits. This is exactly half the advantage of the strategy defined by $\mathcal{C}_y$. □

Lemma 48 shows that bounds on the value of parity games translate into correlation bounds for $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ with $\textsc{Par}$. How might we analyze parity games? They are in some sense "flipped" versions of XOR games, where parity is computed on the inputs to the players, rather than the outputs. However, it is not clear whether the rich collection of techniques developed to analyze XOR games is applicable here. Instead, we bound the no-signaling value of the game by taking the perspective of distinguishability.

For any $(n, k, 0)$ no-signaling channel $\mathcal{N}$, begin by rewriting the correlation as

$$\mathbb{E}[(f \circ \mathcal{N})(x) \cdot \textsc{Par}(x)] = \frac{\mathbb{E}[(f \circ \mathcal{N})(x) \mid x \text{ even}] - \mathbb{E}[(f \circ \mathcal{N})(x) \mid x \text{ odd}]}{2} \,.$$

Let $\mathcal{U}_{\text{even}}$ and $\mathcal{U}_{\text{odd}}$ denote the uniform distribution on even and odd bitstrings of length n respectively, and consider the pushforwards of $\mathcal{U}_{\text{even}}$ and $\mathcal{U}_{\text{odd}}$ through $\mathcal{N}$:

$$\mu := \mathcal{N}\big(\mathcal{U}_{\text{even}}\big) \qquad \text{and} \qquad \nu := \mathcal{N}\big(\mathcal{U}_{\text{odd}}\big) \,.$$

So μ and ν are distributions on strings of length $N := nk$, and

$$\mathbb{E}[(f \circ \mathcal{N})(x) \cdot \textsc{Par}(x)] = \frac{\mathbb{E}[f(\mu)] - \mathbb{E}[f(\nu)]}{2} = \Pr[f(\mu) = 1] - \Pr[f(\nu) = 1] \,.$$

Therefore the correlation of $f \circ \mathcal{N}$ with parity can be phrased in terms of f's ability to distinguish the distributions μ and ν.

What can be said about μ and ν? We claim that on every set $S \subset [N]$ of size at most $N/k - 1 = n - 1$, we must have

$$\mu_S = \nu_S. \tag{8.2.1}$$

Here the notation μ_S denotes the marginal distribution of μ on the coordinates in S. To see (8.2.1), let $T \subset [n]$ be the set of players whose outputs overlap S.

Then by the no-signaling property of $\mathcal{N}$, the marginal μ_S (resp. ν_S) is entirely determined by the marginal input distribution on T; that is, $(\mathcal{U}_{\text{even}})_T$ (resp. $(\mathcal{U}_{\text{odd}})_T$). And for any T a strict subset of $[n]$, $(U_{\text{even}})_T = (\mathcal{U}_{\text{odd}})_T = \mathcal{U}(\{0,1\}^{|T|})$, so we must have $\mu_S = \nu_S$.

So all small marginals of μ and ν are information-theoretically indistinguishable. This is exactly k-wise indistinguishability, a generalization of k-wise independence introduced by Bogdanov et al. [BIVW16] and first used in the context of secret sharing.

Definition ($\boldsymbol{k}$-wise indistinguishability [BIVW16])**.** *Two distributions μ and ν on $\{\pm 1\}^N$ are k-wise indistinguishable if for all $S \subset [N]$ with $|S| \leq k$, $\mu_S = \nu_S$.*

Additionally, for $f : \{0,1\}^n \to \{0,1\}$, we say f is ε-fooled by k-wise indistinguishability if for any pair μ, ν of k-wise indistinguishable distributions,

$$|\Pr[f(\mu) = 1] - \Pr[f(\nu) = 1]| \leq \varepsilon\,.$$

It turns out k-wise indistinguishability over the hypercube is intimately connected to approximate degree. By a linear programming duality argument, Bogdanov et al. proved the following.

Theorem 49 ([BIVW16, Theorem 1.2])**.** *Let $f : \{0,1\}^n \to \{0,1\}$ and $\varepsilon > 0$. Then f is ε-fooled by k-wise indistinguishability if and only if $\widetilde{\deg}_{\varepsilon/2}[f] \leq k$.*

With this fact, Lemma 48, and the above discussion, we are ready prove the main theorem in this section.

We say a class of Boolean functions $\mathcal{F} = (\mathcal{F}_n)_{n \geq 1}$ is *closed under inverse-polynomial restrictions* if for all $f \in \mathcal{F}_n$ and all $S \subseteq [n]$ with $n \in \text{poly}(|S|)$, fixing the bits in S^c yields a function still in $\mathcal{F}$:

$$f\restriction_{S^c \leftarrow x} \in \mathcal{F}_{|S|} \quad \forall x \in \{0,1\}^{|S^c|}\,.$$

Note that AC^0 is closed under inverse-polynomial restrictions.

Theorem 50. *Suppose $\mathcal{F}$ is a class of Boolean functions closed under negations and inverse-polynomial restrictions. Let m be fixed and suppose there is an $f \in \mathcal{F}$ on $N = \text{poly}(m)$ variables and an m-input QNC^0 circuit $\mathcal{C}$ of depth d, with $N - m$ ancilla qubits, and receiving arbitrary quantum advice, such that $f \circ \mathcal{C}$ achieves correlation ε with $\textsc{Par}_m$. Then there is a $g \in \mathcal{F}$ on $n \geq m/2$ variables with $\widetilde{\deg}_{\varepsilon/2}[g] \geq n/2^d - 1$.*

Proof. By Lemma 48, there is an $m' \geq m/(2^d+1)$ and an $(m', 2^d, N - 2^d m')$ nonlocal channel $\mathcal{N}$ such that $f \circ \mathcal{N}$ or $\neg f \circ \mathcal{N}$ achieves correlation ε with $\textsc{Par}_{m'}$.

Suppose the referee measures their system and obtains outcome string r. This event leads to an updated state shared among the parties in $\mathcal{N}$ and thereby defines an $(m', 2^d, 0)$ nonlocal channel $\mathcal{N}^{R \leftarrow r}$. By a similar averaging argument to the one used in Lemma 48, there is at least one outcome r of the referee register such that $\mathcal{N}^{R \leftarrow r}$ still yields correlation ε with $\textsc{Par}$. Define $g := f\restriction_{R \leftarrow r}$ or $g := \neg f \restriction_{R \leftarrow r}$ as appropriate and put $\mathcal{E} := \mathcal{N}^{R \leftarrow r}$. Then $g \in \mathcal{F}$ is a function on $n := 2^d m'$ bits and

$$\mathop{\mathbb{E}}_{x}[(g \circ \mathcal{E})(x) \cdot \textsc{Par}(x)] \geq \varepsilon \,.$$

Therefore, by the discussion above, we see g can ε-distinguish $(n/2^d - 1)$-wise indistinguishable distributions. Applying Theorem 49 we conclude that

$$\widetilde{\deg}_{\varepsilon/2}[g] \geq \frac{n}{2^d} - 1 \,. \qquad \square$$

Corollary 51. *Suppose function class $\mathcal{F}$ is closed under inverse-polynomial-sized restrictions. Then if $\widetilde{\deg}[\mathcal{F}] \in o(n)$, $\mathcal{F} \circ \mathsf{QNC}^0$ cannot achieve $\Omega(1)$ correlation with $\textsc{Par}$.*

The burning question, then, is whether $\widetilde{\deg}[\mathsf{AC}^0] \in o(n)$. In fact, the approximate degree of AC^0 is a longstanding open problem and its resolution would lead to several consequences in complexity theory [BT22]. To get a sense of the difficulty of this question, consider that on one hand, a sublinear upper bound is known for a large subclass of AC^0.

Theorem ([BKT19, Theorem 5])**.** *Let $p(n) \in \mathrm{poly}(n)$. Then the class of AC^0 circuits of linear size, denoted by LC^0, has*

$$\widetilde{\deg}_{1/p(n)}[\mathsf{LC}^0] \in o(n).$$

Yet on the other hand, a series of works, most recently [She22], show the following:

Theorem. *For any $\delta > 0$, there is a function $f \in \mathsf{AC}^0$ with $\widetilde{\deg}[f] \in \Omega(n^{1-\delta})$.*

The lower bound of $\Omega(n^{1-\delta})$-for-any-δ is tantalizingly close to the trivial upper bound of n for the approximate degree of any Boolean function, but

as it stands it is not unreasonable to guess that $\widetilde{\deg}[\mathsf{AC}^0] \in \Theta(n/\log n)$ either. Several questions—including now Conjecture 3—could be settled if the gap between $\Omega(n^{1-\delta})$-for-any-δ and n for $\widetilde{\deg}[\mathsf{AC}^0]$ were closed.

We may combine the sublinear lower bound on LC^0 from [BKT19] with Theorem 50 to obtain:

Corollary 52. *Let $\mathcal{C}$ be an n-input, m-ancilla QNC^0 circuit with arbitrary advice. Suppose $f : \{0,1\}^{n+m} \to \{-1,1\}$ is defined by an AC^0 circuit of size $\mathcal{O}(n)$. Then $f \circ \mathcal{C}$ achieves negligible correlation with $\textsc{Par}_n$.*

Blockwise approximate degree

We conclude this section by laying out a self-contained question concerning the approximate degree of AC^0 with respect to a modified, "blockwise" notion of approximate degree. This question is sufficient to imply Conjecture 3 in full generality and may be easier to resolve than $\widetilde{\deg}[\mathsf{AC}^0]$.

Fix $k \geq 1$ (assuming k divides n for simplicity) and let P be the partition of $[n]$ into "blocks" of size k:

$$P := \big\{\{1,\dots,k\}, \{k+1,\dots,2k\}, \dots, \{n-k+1,n\}\big\}.$$

For a monomial $\chi_S = \prod_{i \in S} x_i$ define the (k-)*block degree* $\mathrm{bdeg}_k[\chi_S]$ to be the number of distinct blocks $B \in P$ having nonempty intersection with S. This definition extends naturally to the k-block degree $\mathrm{bdeg}_k[f]$ of a Boolean function $f : \{0,1\}^n \to \{-1,1\}$ and to the *approximate k-block degree* $\widetilde{\mathrm{bdeg}}_k[f]$ of f:

$$\widetilde{\mathrm{bdeg}}_k[f] = \min\{\mathrm{bdeg}_k[g] \mid g : \{0,1\}^n \to \mathbb{R} \text{ a polynomial with } \|f-g\|_\infty \leq 1/3\}.$$

Of course $\widetilde{\mathrm{bdeg}}_k[f] \leq n/k$ for any function.

Question 2. *For all constants k, does the following hold?*

$$\widetilde{\mathrm{bdeg}}_k[\mathsf{AC}^0] \stackrel{?}{\leq} n/k - 1.$$

As we explain below, this would be enough to prove Conjecture 3. Note the following, which are immediate and hold for all f:

$$\widetilde{\deg}[f] < \frac{n}{k} \implies \widetilde{\mathrm{bdeg}}_k[f] < \frac{n}{k} \implies \widetilde{\deg}[f] < n-k.$$

Moreover, these implications are sharp in that each one cannot generically imply anything stronger, as witnessed by a parity function on an appropriate subset

of $[n]$. Regarding $f \in \mathsf{AC}^0$, the left-hand side holding for arbitrary constant k is equivalent to $\widetilde{\deg}[\mathsf{AC}^0] \in o(n)$, while the far right-hand side follows directly from LMN-type Fourier tail bounds for AC^0.

Proposition 6. *If the resolution to Question 2 is "yes", then Conjecture 3 is true.*

Proof sketch. Consider the referee-free nonlocal channel $\mathcal{E}$ from the proof of Theorem 50, with n/k players responding with k bits each. Defining μ and ν as the pushforwards of uniform distributions over even and odd bitstrings as before, it is true that μ and ν are $(n/k-1)$-wise indistinguishable when viewed as distributions on $\{0,1\}^n$. However, they may also be viewed as distributions on the hypergrid $[2^k]^m$ for $m = n/k$.

With this perspective, μ and ν are $m-1$ indistinguishable. Repeating the proof of [BIVW16, Theorem 1.2] over this larger alphabet, we recover exactly the notion of blockwise degree. The rest of the argument is as before. □

It is unclear to us whether Question 2 is easier than $\widetilde{\deg}[\mathsf{AC}^0] \stackrel{?}{\in} o(n)$. Because AC^0 is closed under permutations of input coordinates $[n]$, we can compare the two questions head-to-head as follows. Let $\mathcal{P}_k$ be all the relabelings of P:

$$\mathcal{P}_k := \left\{ \left\{ \{\pi(1), \ldots, \pi(k)\}, \{\pi(k{+}1), \ldots, \pi(2k)\}, \ldots, \{\pi(n{-}k{+}1), \ldots, \pi(n)\} \right\} \right\}_{\pi \in S_n}.$$

For any $P \in \mathcal{P}_k$, let $\mathrm{bdeg}_P[f]$ be the maximum number of blocks in P overlapped by some monomial in f. Then we have the following characterization, where g ranges over real-valued multilinear polynomials on the hypercube as usual:

$$\widetilde{\deg}[\mathsf{AC}^0] < n/k \iff \forall f \in \mathsf{AC}^0, \exists g, \forall P \in \mathcal{P}_k, \mathrm{bdeg}_P[g] \leq n/k \text{ and } \|f-g\|_\infty \leq 1/3$$
$$\widetilde{\mathrm{bdeg}_k}[\mathsf{AC}^0] < n/k \iff \forall f \in \mathsf{AC}^0, \forall P \in \mathcal{P}_k, \exists g, \mathrm{bdeg}_P[g] \leq n/k \text{ and } \|f-g\|_\infty \leq 1/3\,.$$

8.3 Towards a switching lemma for $\mathsf{AC}^0 \circ \mathsf{QNC}^0$

Recall that our approach in Section 8.1 fails because circuits with LMN-style Fourier decay are not suitably closed under precomposition by QNC^0. In fact this is true even under precomposition by NC^0, and the proof of the LMN theorem elegantly avoids an induction assumption phrased in terms of Fourier decay. Instead, the proof relies on a structural theorem about the effect of random restrictions on DNFs and CNFs—Håstad's celebrated switching lemma:

Theorem (Håstad [Hås86])**.** *Suppose f is a width-w DNF. Then for any $0 \leq \delta \leq 1$,*

$$\Pr_{\rho \sim \mathbf{R}_\delta}[\mathrm{DT}_{\mathrm{depth}}(f\restriction_\rho) > t] \leq (C\delta w)^t\,,$$

where C is a universal constant.

Here $\mathbf{R}_\delta$ is the distribution of random restrictions with star probability δ (see *e.g.,* [ODo14, §4.3] for more). This theorem has received several proofs over time, but each rely on the well-controlled structure of random restrictions. To naively repeat the switching lemma argument directly on $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ would mean to track the passage of random restrictions through QNC^0—a tall order given that QNC^0 can destroy the independence and unbiasedness of random restrictions that switching arguments tend to rely on.

The situation may be slightly improved by instead considering a switching lemma for the model studied in Section 8.2. Recalling that $f \circ \mathcal{N}$ is a randomized function, we may hope for a switching lemma of the following form:

> *An imagined switching lemma for nonlocal channels.* Let $m \geq 0$ and $k, w, n \geq 1$ and suppose $f : \{0,1\}^{kn+m} \to \{0,1\}$ is a DNF of width w and $\mathcal{N}$ is an (n,k,m) nonlocal channel. Then for each restriction ρ there exists a distribution Γ_ρ over decision trees such that $(f \circ \mathcal{N})\restriction_\rho = \{T\}_{T \sim \Gamma_\rho}$ and
>
> $$\Pr_{\rho \sim \mathbf{R}_\delta} \Pr_{T \sim \Gamma_\rho}[\mathrm{depth}(T) > t] \leq (C\delta w)^t\,.$$

By Lemma 48 such a switching lemma would be sufficient to show correlations bounds between $f \circ \mathsf{QNC}^0$ and parity for any DNF (or CNF) f, which in turn are direct prerequisites to proving Conjecture 3. While this imagined switching lemma is currently out of reach, we contend it presents a useful challenge to existing switching lemma proof techniques. As a first step in this direction, we devote this section to a proof of a simpler but related structural result.

Theorem. (Informal) *Any no-signaling channel $\mathcal{N}$ composed with a decision tree τ is equal to a probability distribution Γ of decision trees with* $\mathrm{depth}(\tau') \leq \mathrm{depth}(\tau)$ *for all $\tau' \in \mathrm{Supp}(\Gamma)$.*

Let us fix some notation. For a finite set X let $\mathcal{M}(X)$ denote the set of probability measures on X. The set $\mathcal{M}(X)$ is convex, so for ν a probability measure on $\mathcal{M}(X)$ we may define the expected distribution

$$\mathop{\mathbb{E}}_{\mu \sim \nu}[\mu] := \left\{ x \text{ w.p. } \Pr_{\mu \sim \nu} \Pr_{z \sim \mu}[z = x] \right\}_{x \in X}. \tag{8.3.1}$$

Here we study Boolean channels, or functions of the form

$$\mathcal{N} : \{\pm 1\}^n \to \mathcal{M}(\{\pm 1\}^N).$$

For a probability measure μ on the set of channels from n to N bits, we use $\mathbb{E}_{\mathcal{N}\sim\mu}\mathcal{N}$ to denote the channel defined pointwise as

$$\left(\mathop{\mathbb{E}}_{\mathcal{N}\sim\mu}\mathcal{N}\right)(x) := \mathop{\mathbb{E}}_{\mathcal{N}\sim\mu}[\mathcal{N}(x)]. \tag{8.3.2}$$

To be clear, $\mathcal{N}(x)$ is a probability measure on $\{\pm 1\}^N$, so in the right-hand side of (8.3.2) we are computing the expected distribution according to (8.3.1). Also, for $T \subseteq [N]$ define the *reduced channel*

$$\mathcal{N}^T(x) := \left\{ y \quad \text{w.p.} \sum_{\substack{z\in\{\pm1\}^N \\ z_T = y}} \Pr[\mathcal{N}(x) = z] \right\}_{y\in\{\pm1\}^{|T|}}.$$

Definition (No-signaling channel)**.** *Consider a map $\mathcal{N} : \{\pm 1\}^n \to \mathcal{M}(\{\pm 1\}^N)$ and a 'backwards lightcone' function $B : [N] \to [n] \cup \{\perp\}$. The pair $(\mathcal{N}, B)$ is a* no-signaling channel *(NSC) if for all $S \subseteq [n]$, for all $x, x' \in \{\pm 1\}^n$ with $x_S = x'_S$, we have $\mathcal{N}^{B^{-1}(S\cup\perp)}(x) = \mathcal{N}^{B^{-1}(S\cup\perp)}(x')$.*

That is, a channel is an NSC if for any collection of output indices T, $\mathcal{N}^T(x)$ is a function of $x_{B(T)\setminus\{\perp\}}$ only. Note also $\mathcal{N}^{B^{-1}(\perp)}$ is oblivious to the value of x entirely—the outputs $B^{-1}(\perp)$ could be called the referee outputs.

Recall that for a Boolean function $f : \{\pm 1\}^N \to \{\pm 1\}$, $f \circ \mathcal{N}$ denotes the channel

$$f \circ \mathcal{N}(x) = \left\{ b \quad \text{w.p.} \quad \Pr_{y\sim\mathcal{N}(x)}[f(y) = b] \right\}_{b\in\{\pm1\}}.$$

The restriction structure on NSCs interacts nicely with decision trees:

Theorem 53. *Given $f : \{\pm 1\}^N \to \{\pm 1\}$ and $\mathcal{N} : \{\pm 1\}^n \to \mathcal{M}(\{\pm 1\}^N)$ an NSC, there exists a distribution Γ over decision trees such that*

i. For all x the composition $f \circ \mathcal{N}(x) = \{\tau(x)\}_{\tau\sim\Gamma}$, so $\mathbb{E}[f \circ \mathcal{N}(x)] = \mathbb{E}_{\tau\sim\Gamma}[\tau(x)]$; and

ii. For all $\tau \in \mathrm{Supp}(\Gamma)$, $\mathrm{DT}_{depth}(\tau) \le \mathrm{DT}_{depth}(f)$.

Recall that $f \circ \mathcal{N}$ is an $\mathcal{M}(\{\pm 1\})$-valued function on the hypercube, so $x \mapsto \mathbb{E}[f \circ \mathcal{N}(x)]$ is a $[-1, 1]$-valued function on the hypercube, and accordingly has a multilinear Fourier expansion

$$\mathbb{E}[f \circ \mathcal{N}] = \sum_{S\subset[n]} a_S \chi_S \qquad \text{with} \qquad a_S := \mathop{\mathbb{E}}_{x}\left[\mathbb{E}[(f \circ \mathcal{N})(x)] \cdot \chi_S(x)\right].$$

We pause to note the related fact that in terms of the expected output $\mathbb{E}[f \circ \mathcal{N}]$, the degree of any function f does not increase under composition with an NSC: $\deg(f) \geq \deg(\mathbb{E}[f \circ \mathcal{N}])$. This claim has a very simple direct proof[2] and we emphasize that it is not equivalent to Theorem 53. For example, there are Boolean functions g with $\deg(g) = n^{2/3}$ but $\mathrm{DT}_{\mathrm{depth}}(g) = n$ (see Example 3 in [BW02]). One could imagine a Boolean function h with $\deg(h) \approx \mathrm{DT}_{\mathrm{depth}}(h) \in o(n)$ but where $\mathbb{E}[h \circ \mathcal{N}]$ is "g-like": any decision tree decomposition of $\mathbb{E}[h \circ \mathcal{N}]$ contains a tree of depth n despite having $\deg(\mathbb{E}[h \circ \mathcal{N}]) \in o(n)$. Theorem 53 says such an $h, \mathcal{N}$ pair does not exist; precomposition by an NSC cannot increase the decision tree complexity of a function.

The proof of Theorem 53 requiries some bookkeeping. The idea is to begin with τ's root vertex variable y_i and locally decompose the univariate channel $\mathcal{N}^{\{i\}}(x) \rightsquigarrow y_i$ into a distribution of deterministic functions $\{y_{i,\omega}(x_i)\}_{\omega \sim \mu}$. This decomposition of the root vertex induces a probabilistic decomposition $\{\tau'_\omega \circ \mathcal{N}'_\omega\}_{\omega \sim \mu}$ of the entire hybrid computation where the root variable y_i in τ' has been replaced with an $x_{B(i)}$ and the left and right subtrees of τ become compositions not with $\mathcal{N}$, but with conditional versions of $\mathcal{N}$ where $x_{B(i)}$ and y_i have been fixed to certain values. This conditioning preserves the NSC-ness of the new $\mathcal{N}'$s, and the decomposition recurses down the tree.

We now introduce a notion of conditioning. For any n-to-N bit Boolean channel $\mathcal{N}$, $x \in \{\pm 1\}^n$, $J \subseteq [N]$ and $Y \in \mathrm{Supp}(\mathcal{N}^J(x))$ define the *conditional channel* as

$$\mathcal{N}(x \mid y_J = Y) := \Big\{ y \quad \text{w.p.} \;\; \Pr[\mathcal{N}(x) = y \mid y_J = Y] \Big\}_{y \in \{\pm 1\}^N},$$

and for $T \subseteq [N]$ the *reduced conditional channel*

$$\mathcal{N}^T(x \mid y_J = Y) := \Big\{ y \quad \text{w.p.} \sum_{\substack{z \in \{\pm 1\}^N \\ z_T = y}} \Pr[\mathcal{N}(x) = z \mid z_J = Y] \Big\}_{y \in \{\pm 1\}^{|T|}}.$$

Note that T-reduced conditional no-signaling channels can depend on inputs outside $B(T)$. Consider for example the n-to-n-bit NSC

$$\mathcal{G}(x) = \begin{cases} \mathcal{U}\{\text{even strings}\} & x \text{ even} \\ \mathcal{U}\{\text{odd strings}\} & x \text{ odd}. \end{cases}$$

[2]Consider the Fourier expansion $f = \sum_{S \subseteq [N]} \widehat{f}(S) \chi_S$. Then $\mathbb{E}[(f \circ \mathcal{N})(x)] = \mathbb{E}\left[\sum_{S \subseteq [N]} \widehat{f}(S) \chi_S \circ \mathcal{N}(x)\right] = \sum_S \widehat{f}(S)\, \mathbb{E}[\chi_S \circ \mathcal{N}(x)] = \sum_S \widehat{f}(S)\, \mathbb{E}[\chi_S \circ \mathcal{N}^S(x)]$, a linear combination of functions of at most $|S|$ variables each for $|S| \leq \deg(f)$.

Now $\mathcal{G}^{\{i\}}(x)$ is identically a Rademacher random variable (oblivious to x entirely), but

$$\mathcal{G}^{\{i\}}(x \mid y_{[n]\setminus i} = 00\cdots 0) = \Big\{ \prod_j x_j \quad \text{w.p. } 1 \Big\},$$

the parity of all n bits of x. All the same, some structure remains after conditioning:

Proposition 7. *For $T, J \subseteq [N]$, let $x, x' \in \{\pm 1\}^n$ be such that $x_{B(J\cup T)} = x'_{B(J\cup T)}$. Then for all $Y \in \mathrm{Supp}(\mathcal{N}^J(x))$,*

$$\mathcal{N}^T(x \mid y_J = Y) \;=\; \mathcal{N}^T(x' \mid y_J = Y).$$

Proof. Let x, x' be as in the proposition statement. We have from the definition of NSCs that $\mathcal{N}^{J\cup T}(x) = \mathcal{N}^{J\cup T}(x')$. Certainly then $\mathcal{N}^{J\cup T}(x \mid y_J = Y) = \mathcal{N}^{J\cup T}(x' \mid y_J = Y)$ (we have taken the marginal of two equal distributions). The conclusion then follows from noticing that for any $U \subseteq V$, $\mathcal{N}^U = (\mathcal{N}^V)^U$. □

This proposition says $\mathcal{N}^T(x \mid y_J = Y)$ is a function of $x_{B(J\cup T)}$ only. Thus if we fix variables $x_{B(J)}$ we recover a smaller NSC:

Corollary 54. *Consider an n-to-N NSC $(\mathcal{N}, B)$, an $i \in [N]$, and $X, Y \in \{\pm 1\}$. If $B(i) = \perp$ let $\mathcal{N}'$ be the n-to-$(N-1)$ NSC*

$$\mathcal{N}' = \mathcal{N}^{[N]\setminus\{i\}}(x \mid y_i = Y)$$

and otherwise let $\mathcal{N}'$ be the $(n-1)$-to-$(N-1)$ NSC

$$\mathcal{N}' = \mathcal{N}^{[N]\setminus\{i\}}(x_{\{B(i)\}^c} \mid x_{B(i)} = X, y_i = Y).$$

Define a new lightcone function B' from B as follows. Put $B(j) = \perp$ for all $j \in B^{-1}(B(i))$ and then remove i from the domain of B. Then $(\mathcal{N}', B')$ is an NSC.

Finally we introduce an object used internally in the proof of Theorem 53.

Definition (Hybrid Decision Tree)**.** *A* hybrid decision tree $\mathcal{T}$ *on n variables with ℓ leaves consists of the data $(\tau, \mathcal{G}_1, \ldots, \mathcal{G}_\ell)$, where*

i. The first argument τ is a rooted binary tree with ℓ leaves labeled as follows. Each internal node is assigned x_i for some $i \in [n]$, the edge to its left child is labeled 1, and the edge to its right child is labeled -1.

ii. Each leaf ι of τ is associated with an n-to-1 channel $\mathcal{G}_\iota : \{\pm 1\}^n \to \mathcal{M}\{\pm 1\}$.

A hybrid tree defines a channel $\mathcal{T}_\tau(\mathcal{G}_1, \dots, \mathcal{G}_\ell) : \{\pm 1\}^n \to \mathcal{M}\{\pm 1\}$ as follows. Computation on input $x \in \{\pm 1\}^n$ proceeds just as with standard decision trees until a leaf ι is reached, at which point the distribution $\mathcal{G}_\iota(x)$ is returned.

Theorem 53 follows from these three claims. Proofs of the first two are immediate from the definitions.

Claim 4. *For any hybrid decision tree $\mathcal{T}$,*

$$\mathcal{T}\Big(\mathcal{G}_1, \dots, \mathcal{G}_{\iota-1}, \mathop{\mathbb{E}}_{\omega\sim\mu}[\mathcal{G}_\omega], G_{\iota+1}, \dots, \mathcal{G}_\ell\Big) = \mathop{\mathbb{E}}_{\omega\sim\mu}\Big[\mathcal{T}(\mathcal{G}_1, \dots, \mathcal{G}_{\iota-1}, \mathcal{G}_\omega, \mathcal{G}_{\iota+1}, \dots, \mathcal{G}_\ell)\Big]$$

Claim 5. *For any hybrid decision trees $\mathcal{T}_\tau(\mathcal{G}_1, \dots, \mathcal{G}_\ell)$ and $\mathcal{T}_{\tau'}(\mathcal{G}_{\iota 1}, \dots, \mathcal{G}_{\iota\ell'})$,*

$$\begin{aligned}\mathcal{T}_\tau\Big(\mathcal{G}_1, \dots, \mathcal{G}_{\iota-1}, \mathcal{T}_{\tau'}(\mathcal{G}_{\iota 1}, \dots, \mathcal{G}_{\iota\ell'}), G_{\iota+1}, \dots, \mathcal{G}_\ell\Big)& \\ = \mathcal{T}_{\tau \circ_\iota \tau'}(\mathcal{G}_1, \dots, \mathcal{G}_{\iota-1}, \mathcal{G}_{\iota 1}, \dots, \mathcal{G}_{\iota\ell'}, \mathcal{G}_{\iota+1}, \dots, \mathcal{G}_\ell),&\end{aligned}$$

where $\tau \circ_\iota \tau'$ is τ with the ι^{th} leaf replaced with τ'.

Claim 6. *Suppose τ is a decision tree and $(\mathcal{N}, B)$ is an NSC. Then either:*

i. $\tau \circ \mathcal{N} = \mathbb{E}_{\omega\sim\mu}[\tau_\omega \circ \mathcal{N}_\omega]$ where $\mathrm{depth}(\tau_\omega) \le$ depth$(\tau) - 1$, $|\mathrm{Supp}(\mu)| \le 2$, and each $\mathcal{N}_\omega$ is an NSC, or

ii. $\tau \circ \mathcal{N} = \mathbb{E}_{\omega\sim\mu}\Big[\mathcal{T}_{\tau^*}\Big(\tau_{\omega_L} \circ \mathcal{N}_{\omega_L}, \tau_{\omega_R} \circ \mathcal{N}_{\omega_R}\Big)\Big]$, where $|\mathrm{Supp}(\mu)| \le 3$, τ^* has one internal node, $\mathrm{depth}(\tau_{\omega_L}), \mathrm{depth}(\tau_{\omega_R}) \le \mathrm{depth}(\tau) - 1$, and each $\mathcal{N}_{\omega_L}, \mathcal{N}_{\omega_R}$ is an NSC; or

iii. (Base case) $\tau \circ \mathcal{N}(x) = \{b \ \ w.p.\ 1\}$ for all x, for some fixed $b \in \{\pm 1\}$.

Proof. If τ is the trivial decision tree with no internal nodes, clearly we satisfy case *iii.* Otherwise, let y_i be the variable at the root of τ. There are two cases depending on the value of $B(i)$.

Case i), $B(i) = \perp$. Observe that $\mathcal{N}^{\{i\}}(x)$ is the same distribution μ over $\{\pm 1\}$, independent of x. For $\omega \in \{\pm 1\}$ let τ_ω be the subtree of τ attached to the ω-valued edge of y_i. Put $\mathcal{N}_\omega = \mathcal{N}^{T\setminus\{i\}}(x \mid y_i = \omega)$. Then we have for

$z \in \{\pm 1\}$,

$$\begin{aligned}\Pr[\tau \circ \mathcal{N}(x) = z] &= \textstyle\sum_{\omega \in \{\pm 1\}} \Pr[\tau \circ \mathcal{N}(x) = z \mid D^i(x) = \omega] \Pr[\mathcal{N}^i(x) = \omega] \\ &= \textstyle\sum_{\omega \in \{\pm 1\}} \Pr[\tau_\omega \circ \mathcal{N}(x \mid y_i = \omega) = z] \Pr[\mathcal{N}^i(x) = \omega] \\ &= \textstyle\sum_{\omega \in \{\pm 1\}} \Pr[\tau_\omega \circ \mathcal{N}_\omega(x) = z] \Pr[\mathcal{N}^i(x) = \omega] \\ &= \Pr\Big[\mathop{\mathbb{E}}_{\omega \sim \mu}[\tau_\omega \circ \mathcal{N}_\omega](x) = z\Big]\end{aligned}$$

as desired. Clearly τ_ω is strictly shorter than τ, and $\mathcal{N}_\omega$ is an NSC by Corollary 54.

Case ii), $B(i) \neq \perp$. Let τ^* be the one-vertex tree consisting of the root vertex of τ relabeled with $x_{B(i)}$ and let τ_1, τ_{-1} be the left and right subtrees of τ respectively. Observe that $\mathcal{N}^{\{i\}}(x) = \mathcal{N}^{\{i\}}(x_{B(i)})$ is a univariate channel. Hence it can be decomposed as a convex combination

$$\mathcal{N}^{\{i\}}(x_{B(i)}) = a_{(1,1)}\begin{bmatrix}1 & 1\\ 0 & 0\end{bmatrix} + a_{(-1,-1)}\begin{bmatrix}0 & 0\\ 1 & 1\end{bmatrix} + a_{(1,-1)}\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix} + a_{(-1,1)}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix},$$

where only three of $a_{(L,R)}$ are nonzero. Let $\mu = \{(L,R) \text{ w.p. } a_{(L,R)}\}$. Then we claim

$$\tau \circ \mathcal{N} = \mathop{\mathbb{E}}_{(L,R)\sim\mu}\Big[\mathcal{T}_{\tau^*}\Big(\tau_L \circ \mathcal{N}_L^{(1)}, \tau_R \circ \mathcal{N}_R^{(-1)}\Big)\Big], \tag{8.3.3}$$

where for $b, c \in \{\pm 1\}^2$,

$$\mathcal{N}_c^{(b)}(x) = \mathcal{N}(x | x_{B(i)} = b, y_i = c).$$

We check Eq. (8.3.3) pointwise. First consider an x with $x_{B(i)} = 1$. We condition on the value of y_i, rearrange, and then "complete the tree":

$$\begin{aligned}\Pr[\tau \circ \mathcal{N}(x) = z] &= \sum_{L \in \{\pm 1\}} \Pr[\tau \circ \mathcal{N}(x) = z \mid \mathcal{N}_i(x) = L] \Pr[\mathcal{N}_i(x) = L] \\ &= \sum_{L \in \{\pm 1\}} \Pr[\tau \circ \mathcal{N}(x \mid y_i = L) = z](a_{(L,1)} + a_{(L,-1)}) \\ &= \sum_{L \in \{\pm 1\}} \Pr[\tau_L \circ \mathcal{N}(x \mid x_{B(i)} = 1, y_i = L) = z]\Big(\textstyle\sum_{R \in \{\pm 1\}} a_{(L,R)}\Big) \\ &= \sum_{L,R \in \{\pm 1\}} a_{(L,R)} \Pr[\tau_L \circ \mathcal{N}_L^{(1)}(x) = z] \\ &= \sum_{L,R \in \{\pm 1\}} a_{(L,R)} \Pr[\mathcal{T}_{\tau^*}(\tau_L \circ \mathcal{N}_L^{(1)}, \tau_R \circ \mathcal{N}_R^{(-1)})(x) = z] \\ &= \Pr\Big[\mathop{\mathbb{E}}_{(L,R)\sim\mu}[\mathcal{T}_{\tau^*}(\tau_L \circ \mathcal{N}_L^{(1)}, \tau_R \circ \mathcal{N}_R^{(-1)})](x) = z\Big],\end{aligned}$$

as desired. A similar argument goes through for $x_{B(i)} = -1$ by expanding over R instead of L. □

Proof of Theorem 53. Let τ be a depth-optimal decision tree for f. Construct the trivial hybrid tree $\mathcal{T}$ with no internal nodes and a single leaf with label $\tau \circ \mathcal{N}$. Put $\Gamma = \{\mathcal{T} \text{ w.p. } 1\}$. We will recursively break apart leaves of $\mathcal{T}$ into distributions of hybrid trees, which are then combined with the parent tree to become distributions over hybrid trees of greater depth.

This is done by repeated application of the following sequence of steps. Suppose $\mathcal{T}_\tau(\mathcal{G}_1, \ldots, \mathcal{G}_\ell)$ is some hybrid tree and $\mathcal{G}_\iota = \tau' \circ \mathcal{N}$ for some nontrivial DT τ' and (potentially conditioned) NSC $\mathcal{N}$. Then depending on the case in Claim 3 we either have

$$\begin{aligned}
\mathcal{T}_\tau(\ldots, \underbrace{\tau \circ \mathcal{N}}_{\text{index } \iota}, \ldots) &= \mathcal{T}_\tau(\ldots, \mathbb{E}_{(L,R)\sim\mu}[\mathcal{T}_{\tau^*}(\tau_{\omega_L} \circ \mathcal{N}_{\omega_L}, \tau_{\omega_R} \circ \mathcal{N}_{\omega_R})], \ldots) && \text{(Claim 6.}i\text{)} \\
&= \mathop{\mathbb{E}}_{(\omega_L,\omega_R)\sim\mu} \Big[\mathcal{T}_\tau(\ldots, \mathcal{T}_{\tau^*}(\tau_{\omega_L} \circ \mathcal{N}_{\omega_L}, \tau_{\omega_R} \circ \mathcal{N}_{\omega_R}), \ldots)\Big] && \text{(Claim 4)} \\
&= \mathop{\mathbb{E}}_{(\omega_L,\omega_R)\sim\mu} \Big[\mathcal{T}_{\tau \circ_\iota \tau^*}(\ldots, \tau_{\omega_L} \circ \mathcal{N}_{\omega_L}, \tau_{\omega_R} \circ \mathcal{N}_{\omega_R}, \ldots)\Big], && \text{(Claim 5)}
\end{aligned}$$

where τ^* has depth 1 and $\text{depth}(\tau_{\omega_L}), \text{depth}(\tau_{\omega_R}) \leq \text{depth}(\tau') - 1$, or we have

$$\begin{aligned}
\mathcal{T}_\tau(\ldots, \underbrace{\tau \circ \mathcal{N}}_{\text{index } \iota}, \ldots) &= \mathcal{T}_\tau(\ldots, \mathbb{E}_{\omega\sim\mu}[\tau_\omega \circ \mathcal{N}_\omega], \ldots) && \text{(Claim 6.}ii\text{)} \\
&= \mathop{\mathbb{E}}_{\omega\sim\mu} \Big[\mathcal{T}_\tau(\ldots, \tau_\omega \circ \mathcal{N}_\omega, \ldots)\Big], && \text{(Claim 4)}
\end{aligned}$$

where $\text{depth}(\tau_\omega) \leq \text{depth}(\tau') - 1$.

If we repeatedly make these transformations on the elements of Γ, we will eventually be left with a distribution over hybrid decision trees $(\tau, \mathcal{G}, \ldots)$ where each channel $\mathcal{G} = \tau' \circ \mathcal{N}$ is in the base case of Claim 6. Such a hybrid tree is equal to a deterministic channel. Hence we are left with a distribution over deterministic channels that is trivially equivalent to a distribution of standard, deterministic decision trees.

Further, it's easy to see that once done, the longest path in any tree of $\text{Supp}(\Gamma)$ is bounded by the longest path in the original tree τ. □

8.4 Discussion

We have seen several pieces of evidence for Conjecture 3, as well as highlighted new connections between quantum complexity theory, nonlocal games, and approximate degree.

If Conjecture 1 is ultimately proved true, we may wish to reach for a stronger no-advantage theorem closer to that of Beals et al. [Bea+01] from query complexity. A natural expression of $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ non-advantage might use the language of Fourier decay.

Question 3. *Does* $\mathsf{AC}^0 \circ \mathsf{QNC}^0$ *exhibit LMN-like Fourier decay? To make this precise for the randomized function* $f \circ \mathcal{C}$*, consider the expectation over the randomness in* $\mathcal{C}$ *to get a function* $F : \{0,1\}^n \to [-1,1]$*. Then we ask, is* $\mathbf{W}^{\geq t}[F] \in \mathcal{O}(\exp(-t))$*?*

As mentioned in the introduction, a similar result is known depth-d QAC^0 circuits with at most $\mathcal{O}(n^{1/d})$ ancillas [NPVY23].

Finally, one may consider any number of variations on the theme of pre-composing a Boolean function with QNC^0. It is natural to ask:

Question 4. *View a* QNC^0 *circuit* $\mathcal{C}$ *as a map from (randomized) Boolean functions to randomized Boolean functions:*

$$f \overset{\mathcal{C}}{\longmapsto} f \circ \mathcal{C}\,.$$

By how much can this map increase influence, sensitivity, or other complexity measures of f*?*

Theorem 53 gives the answer "not at all" to a variant Question 4 where QNC^0 is replaced by nonlocal channels, and the complexity measure is randomized decision tree complexity.

8.5 Acknowledgements

We are grateful to Chris Umans and Thomas Vidick for numerous valuable discussions and for the opportunity to share the contents of this chapter with Henry Yuen and the quantum group at Columbia University in the fall of 2022. We are also grateful to Atul Singh Arora, discussions with whom inspired this project. Finally we thank the anonymous ITCS 2024 reviewers for their generous and meticulous feedback on an earlier draft.

Chapter 9

Testing classical properties from quantum data

In property testing we consider a subset $\mathcal{P}$ of the set of all Boolean functions $f : \{0,1\}^n \to \{0,1\}$ and aim to find fast algorithms for deciding (with high probability) whether an unknown function f has property $\mathcal{P}$ or is ε-far from having property $\mathcal{P}$; that is, we wish to decide between

$$\text{Case } (i) \quad f \in \mathcal{P} \qquad \text{or} \qquad \text{Case } (ii) \quad \min_{g \in \mathcal{P}} \|f - g\|_1 \geq \varepsilon \,,$$

promised one of these is the case. Here $\|f - g\|_1 = \Pr_{x \sim \{0,1\}^n}[f(x) \neq g(x)]$ is the L^1 distance. Property testing began in the context of program checking [BLR90; RS96], where it was shown that only $\mathcal{O}(1)$ queries to f are needed to determine (with high probability) whether f is linear or is $\Theta(1)$-far from linear—which compares very favorably to the $\Omega(n)$ query lower bound for *learning* linear functions. The extreme query efficiency of property testing algorithms soon after played a critical role in interactive proofs and PCP theorems [AS98; Aro+98; Din07]. Since then property testing has developed into a rich landscape of access models, complexity regimes, and separations [Fis04; Rub07; Ron09; Sud10; Gol17].

One of the promises of this broad view of property testing, identified very early on [GGR98], is its potential in data analysis and machine learning: one could run inexpensive property testing algorithms to guide the choice of which long-running learning algorithm to use. But there is an unfortunate catch: the dramatic complexity advantage of testing over learning typically disappears in the natural access model for data analysis and machine learning, where fresh queries to f cannot be made and only a limited dataset $\{(x_j, f(x_j)\}_j$ of random samples from f is available. This setting is known as *passive* or *sample-based testing* [GGR98].

Indeed, many results in passive testing are lower bounds that grow with n, unlike the algorithms available in query-based testing: compare among the "Classical" columns in Table 9.1. In fact, Blais and Yoshida [BY19] showed that if a Boolean property can be tested from $\mathcal{O}(1)$ random samples, then the property is of a rather restricted kind.[1]

[1] In particular, such a property is only a function of the conditional expected values

	Quantum		Classical		
Property	Queries	Examples	Samples	Queries	Learning (from queries)
k-Juntas	$\widetilde{\mathcal{O}}(\sqrt{k})$ [ABRW15]	$\mathcal{O}(k)$ [AS07]	$\Omega(2^{k/2}+k\log n)$ [AHW16]	$\widetilde{\Theta}(k)$ [Bla09; CG04]	$\Omega(2^k+k\log n)$ [AHW16]
Linearity	$\mathcal{O}(1)$	$\Theta(1)$ [BV97]	$n+\Theta(1)$ [AHW16]	$\Theta(1)$ [BLR90]	$n+\Theta(1)$ [AHW16]
$\mathbb{F}_2$ degree-d	$\mathcal{O}(2^d)$	$\mathcal{O}(n^{d-1})$ [ABDY23]	$\Theta(n^d)$ [AHW16]	$\Theta(2^d)$ [Alo+03; Bha+10]	$\Theta(n^d)$ [AHW16]
Monotonicity	$\widetilde{\mathcal{O}}(n^{1/4})$ [BB15]	$\tilde{\mathcal{O}}(n^2)$ [Theorem 63]	$2^{\Omega(\sqrt{n})}$ [Bla24]	$\widetilde{\mathcal{O}}(\sqrt{n})$ [KMS18]	$2^{\Omega(\sqrt{n})}$ [BBL98]
Symmetry	$\mathcal{O}(1)$	$\mathcal{O}(1)$ [Theorem 60]	$\Theta(n^{1/4})$ [AHW16]	$\mathcal{O}(1)$ [BWY15]	$\Theta(n^{1/2})$ [AHW16]
Triangle-freeness	$\mathcal{O}(1)$	$\mathcal{O}(1)$ [Theorem 66]	$\Omega(n)$ via [AHW16]	$\mathcal{O}(1)$ [BWY15]	–

Table 9.1: **Upper and lower bounds for testing and learning in various access models.** All bounds are given for (a sufficiently small) constant $\varepsilon > 0$. Bounds that are given without a reference follow trivially from other bounds in the table.

Remark 1. This is not to say that the classical passive testing model is *uninteresting*; there are many exciting positive results for the model, falling under the umbrella of *sublinear* algorithms. For example, the line of work [FLV15; GR16; DGL23] showed that the existence of certain constant-query testers implies sample-based algorithms with sublinear dependence on n. But passive testers still cannot compete with query-based testing for many important problems, as the lower bounds in Table 9.1 attest.

How could we recover large testing speedups in the context of passive testing from data? In the present chapter we advocate for quantum computing (and "quantum datasets") as an answer. Viewed from the right perspective, early results in quantum complexity theory actually demonstrate that quantum data—in the form *quantum examples*, or copies of the *function state* $|f\rangle := 2^{-n/2}\sum_x |x, f(x)\rangle$—can sometimes suffice for highly efficient property testing. For example, the Bernstein-Vazirani algorithm, usually understood as an $\mathcal{O}(1)$

$\mathbb{E}_x[f(x)|x \in S_j]$ of f for sets S_j forming a constant-cardinality partition of the hypercube, $S_1 \sqcup \cdots \sqcup S_{\mathcal{O}(1)} = \{0,1\}^n$.

quantum *query* algorithm, really only needs $\mathcal{O}(1)$ *function states* to test for linearity [BV97] (vs. $\Omega(n)$ classical samples), and the quantum k-junta tester of Atıcı and Servedio [AS07] also requires only $\mathcal{O}(k)$ quantum examples (*c.f.* the lower bound of $\Omega(2^{k/2} + k \log n)$ classical samples). The present chapter seeks to establish *passive quantum testing* as a fundamental model of property testing by making progress on the question:

What is the extent of quantum advantage in testing classical properties from data?

Before the contributions of this chapter it was not fully clear whether quantum data in the form of quantum examples can lead to testing speedups beyond linearity and k-junta-like properties (such as low Fourier degree): both the Bernstein–Vazirani algorithm and the Atıcı–Servedio junta tester rely only on *quantum Fourier sampling* [BV97], a quantum subroutine which, given copies of $|f\rangle$, returns the label $S \subseteq [n]$ of a Fourier character with probability $\widehat{f}(S)^2$. Despite the success of quantum Fourier sampling, its utility is restricted to properties that are "plainly legible" from the Fourier spectrum.[2]

In this chapter we expand the list of properties with efficient passive quantum testers, including one which provably requires a non-Fourier sampling approach. We also compare the power of quantum data to that of classical queries, finding that they are (essentially) maximally incomparable as resources for testing. Finally, we begin a study of lower bounds for testing monotonicity from quantum data by showing that the ensembles leading to exponential lower bounds for classical sample-based testing yield no nontrivial lower bounds for quantum data-based testing. In the remainder of the introduction we explore each of these points in greater detail.

Remark 2 (Where might quantum data appear?). While from the perspective of complexity theory quantum data leads to a natural counterpart to classical passive testing, and demonstrates a "data-based" quantum advantage, the reader may still feel it is not entirely natural from a practical or "physical" standpoint. To the contrary, we contend that quantum data may be a useful

[2] As an example of a property *not* detectable from the Fourier spectrum, consider the task of testing if f is a quadratic $\mathbb{F}_2$ polynomial. It is well-known (see, *e.g.*, [HHL+19, Claim 2.4]) that degree-2 $\mathbb{F}_2$ polynomials can have Fourier coefficients with uniformly exponentially-small magnitudes, so Fourier sampling is not directly useful for this task. Our Theorem 55 below serves as another example.

component of emerging quantum technologies. We briefly list some scenarios where quantum data may be a natural object.

- Suppose a researcher has time-limited query access to a data-generating process, but does not yet know what questions about the process she will eventually ask. She may prefer to store data in quantum memory rather than classical, to broaden the range of questions that can be answered *post hoc*.

- In high-latency and bandwidth-limited scenarios, back-and-forth (adaptive, query-based) interaction is not feasible, for example in space exploration. If a space probe departing Earth shared some entanglement with a ground station, it could later in its journey encode observations into quantum data and teleport the resulting states back to Earth. In such a scenario, the advantages of quantum data could lead to significant speedups in research and analysis.

- Rather than sharing the source code for a program f, a company may prefer to share a quantum data encoding of it as a form of copy protection—provided the function state is sufficient for the intended application.

Quantum advantage in testing from data: an emerging theme

Our first contribution is to expand the list of properties exhibiting quantum advantage in testing from data. Our algorithms work by finding new quantum ways to exploit insights from prior work in classical testing. See Section 9.1 for proofs.

Symmetry testing. A Boolean function is *symmetric* if $f(x) = f(y)$ when x is a permutation of y. We confirm that projecting $|f\rangle$ onto the symmetric subspace suffices for an $\mathcal{O}(1)$-copy quantum test. For comparison, classical passive symmetry testing requires $\Omega(n^{1/4})$ samples [AHW16].

Monotonicity testing. A Boolean function f is *monotone* if $f(x) \leq f(y)$ when $x \prec y$ in the standard partial order $\prec$ on the hypercube. Monotonicity has been of central importance in the classical property testing literature [Gol+00; BB15; KMS18]. We give a quantum algorithm that tests monotonicity with $\tilde{\mathcal{O}}(n^2)$ copies of the function state for f, in comparison to the lower bound of $2^{\Omega(\sqrt{n})}$ samples for classical passive testing [Bla24].

The algorithm appeals to a characterization of monotonicity in terms of the Fourier spectrum of f. In particular, let ε be the L^1 distance between a Boolean function f and the set of all monotone functions. Then we may relate ε to the Fourier spectrum of f via

$$2\varepsilon \;\leq\; \mathbf{I}[f] - \textstyle\sum_i \widehat{f}(\{i\}) \;\leq\; 4\varepsilon n\,.$$

Here $\mathbf{I}[f]$ is the total influence of f and is equal to the expected size of a subset $S \subseteq [n]$ sampled according to the Fourier distribution of f. $\mathbf{I}[f]$ can thus be easily estimated with Fourier sampling, and the Fourier coefficients $\widehat{f}(\{i\})$ estimated with classical samples. The bounds above follow from a reinterpretation of the "pair tester" characterization of monotonicity [Gol+00], which was not originally Fourier-based.

Triangle-freeness. A Boolean function f is *triangle-free* if there are no x, y such that $(x, y, x+y)$ form a *triangle*: $f(x) = f(y) = f(x+y) = 1$. We give a passive quantum triangle-freeness tester that uses only $\mathcal{O}(1)$ copies of $|f\rangle$, in contrast with the $\Omega(n)$ samples required classically.[3]

It is known that to test triangle-freeness, it suffices to estimate the probability that $(x, y, x+y)$ forms a triangle for uniformly random x, y [Fox11; HST16]. Our test estimates this probability by repeating the following subroutine. First, measuring copies of $|f\rangle$ in the computational basis allows us to find a uniformly random $y \in f^{-1}(1)$. Then by measuring the output register of copies of $|f\rangle$, we obtain copies of the entire 1-preimage state

$$\left|f^{-1}(1)\right\rangle \propto \sum_{x \in \{0,1\}^n,\, f(x)=1} |x\rangle\;.$$

Applying the unitary transformation $U_y\,|x\rangle = |x+y\rangle$ then allows us to transform copies of $|f^{-1}(1)\rangle$ into copies of

$$|f^{-1}(1)+y\rangle \propto \sum_{x \in \{0,1\}^n,\, f(x+y)=1} |x\rangle\;.$$

The overlap $|\,\langle f^{-1}(1)|f^{-1}(1)+y\rangle\,| = \Pr_{x\sim\{0,1\}^n}[f(x) = f(x+y) = 1]$ can now be estimated with a SWAP test [BCWD01].

[3]This classical lower bound can be seen by an argument via linear independence similar to that used in the lower bound proof of [AHW16, Theorem 10].

Fourier sampling does not suffice

Given that Fourier sampling is sufficient to test linearity [BV97], k-juntas [AS07; ABRW15], and (as shown above) monotonicity, one might wonder whether Fourier sampling is "all that quantum data is good for" in the context of property testing Boolean functions. To the contrary, we exhibit a property for which a Fourier sampling-based approach requires super-polynomially more data than the optimal passive quantum tester.

Theorem 55. *There is a property $\mathcal{P}$ of Boolean functions on $2n$ bits such that:*

(i) There is no algorithm for testing $\mathcal{P}$ that uses $2^{o(\sqrt{n})}$ classical samples and any number of Fourier samples.

(ii) There is an efficient quantum algorithm for testing $\mathcal{P}$ from $\mathcal{O}(1)$ copies of $|f\rangle$.

Theorem 55 is proved in Section 9.2 as Theorems 68 and 69. The property $\mathcal{P}$ is the *Maiorana-McFarland (MM)* class of bent functions, which take the form $f(x,y) = \langle x,y\rangle + h(x)$ for h any n-bit Boolean function (see *e.g.,* [CM16] for more).

To prove (i), we show a special subset F_{yes} of MM functions with far-from-constant h are indistinguishable from the set F_{no} of their "duals," defined by replacing $h(x)$ with $h(y)$. Every function in both these sets is *bent*—*i.e.*, all Fourier coefficients have equal magnitude—so Fourier samples cannot help. It thus suffices to lower bound the number of classical samples needed to solve the distinguishing problem. The set F_{yes} is chosen so that for a uniformly-random $\langle x,y\rangle + h(x)$ from F_{yes}, the distribution of truth tables of h is $2^{c\sqrt{n}}$-wise independent. This means that for any number of samples less than $2^{c\sqrt{n}}$, except in the very unlikely event that there is a collision among the sampled points $\{(x^{(i)},y^{(i)})\}_i$, the distribution of values $f(x^{(i)},y^{(i)})$ will look uniformly random, regardless of whether f is sampled uniformly from F_{yes} or F_{no}—and so distinguishing is impossible. The truth tables for h are constructed from certain affine shifts of Reed–Muller codewords.

As for item (ii), the passive quantum tester for this property first applies the unitary U defined by $|x,y,b\rangle \mapsto |x,y,b \oplus \langle x,y\rangle\rangle$ to $|f\rangle$. If f is a MM function the result should be h, a function depending only on the first n variables, while if $|f\rangle$ is far from MM functions, it will have noticeable dependence on coordinates

$n+1, \ldots, 2n$. This dependence can be measured by Fourier-sampling the transformed state.

Comparing access models

Quantum data is always at least as good as classical samples, and from Table 9.1 we see that for a growing list of properties, testing from quantum data is competitive with testing from classical *queries*. In fact, quantum data can be vastly more powerful than classical queries for testing. An extremal example of this is the FORRELATION problem, which can be tested from $\mathcal{O}(1)$ function state copies but requires $\Omega(2^{n/2})$ classical queries [AA15].

Conversely, one may wonder to what extent classical queries may outperform quantum data for property testing. An answer is not so obvious. Although classical queries enable direct access to $f(x)$ at any point x of the algorithm's choosing—a powerful advantage over quantum data—it is not so clear whether this can lead to a separation for property testing. Recall that for a property testing problem, *yes* and *no* instances must be $\Omega(1)$-far in L^1 distance. So to create a hard property for quantum data-based testers, one must find two sets of functions which pairwise differ on a *constant fraction* of the locations in their truth tables, yet still remain hard to distinguish by a quantum algorithm operating on copies of their function states.

We succeed in "hiding" these large differences and identify a testing problem for which classical queries have a dramatic advantage over quantum data.

Theorem 56. *There exists a testing task (3-fold intersection detection) that can be accomplished with $\mathcal{O}(1)$ classical queries but requires $\Omega(2^{n/2})$ copies for quantum testing from data.*

Combined with the FORRELATION separation of [AA15], Theorem 56 entails that quantum data and classical queries are (essentially) maximally incomparable. See Chapter 9 for a full picture of resource inequalities for testing.

Theorem 56 is proved in Section 9.3 as Theorem 70. Given a function $f : \{0,1,2\} \times \{0,1\}^n \to \{0,1\}$ that indicates three subsets of the hypercube $A, B, C \subseteq \{0,1\}^n$, the 3-fold intersection detection task is to determine if the fractional 3-fold intersection $|A \cap B \cap C|/2^n$ is 0 or $\Omega(1)$-far from 0.

This property is readily tested from queries by computing the probability $x \in A \wedge x \in B \wedge x \in C$ for uniformly-random x. To prove the quantum passive testing lower bound, we show the indistinguishability of two ensembles

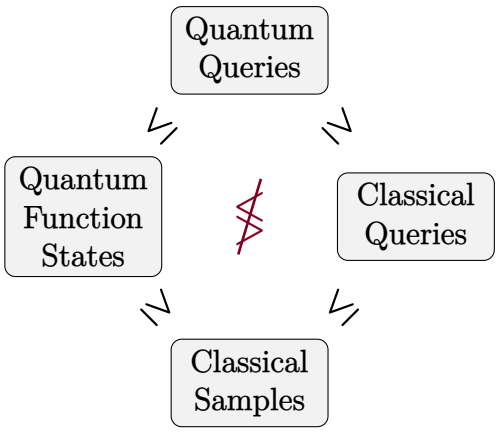

Figure 9.1: **Property testing resource inequalities.** The figure illustrates the connections between four different data access models in property testing, namely classical/quantum example/query access. Here, "resource A $\geq$ resource B" means that access to resource B can be simulated from access to resource A without any overhead. (For example, a single classical query can be used to simulate a single classical sample.) As a consequence of Theorem 56 and [BV97; Sim97; AA15], the only two among these access models that are not trivially comparable are in fact very incomparable.

of function states encoding set triples

$$\left\{(A,B,C)\right\}_{A,B,C \overset{\text{iid}}{\sim} \mathcal{P}\{0,1\}^n} \qquad \text{vs.} \qquad \left\{(A,B,A\Delta B)\right\}_{A,B \overset{\text{iid}}{\sim} \mathcal{P}\{0,1\}^n}.$$

Here Δ denotes symmetric difference, $\mathcal{P}$ denotes the power set, and the samples are uniform. Note the first ensemble has mutual intersection of $\Omega(1)$ density with high probability, while the ensemble always has zero intersection. To obtain the lower bound, the main observation is that the t-copy versions of the two associated function state ensembles are *equal* when projected onto the so-called *distinct subspace* (*i.e.,* the subspace spanned by basis states for which the t input registers are distinct). This projection moves the state ensembles at most $\mathcal{O}(t/2^{n/2})$ in trace distance, so we conclude that for any $t = o(2^{n/2})$, the two ensembles cannot be distinguished using t function state copies.

A challenge: lower bounds for quantum monotonicity testing

We also begin the project of finding lower bounds for the passive quantum testing model. Our main contribution is to establish lower bounds for monotonicity as an important first open problem. In particular, we show the ensembles that entail strong lower bounds for classical passive testing are wholly inadequate for quantum passive testing.

Theorem 57. *The ensembles in Goldreich et al. [Gol+00] and Black [Bla24] can be distinguished by a quantum algorithm with $\mathcal{O}(1/\varepsilon)$ copies of the corresponding function states.*

This theorem says that the best lower bound such ensembles could imply for quantum passive testing is $\Omega(1/\varepsilon)$. But that is no better than the lower bound that exists generically for every (non-trivial) property.[4] To see that $\Omega(1/\varepsilon)$ holds generically, it suffices to consider only two functions, f_{yes} and f_{no}, that are exactly ε-far apart. This is equivalent to $\langle f_{\text{yes}}|f_{\text{no}}\rangle = 1-\varepsilon$, so the trace distance between $|f_{\text{yes}}\rangle^{\otimes t}$ and $|f_{\text{no}}\rangle^{\otimes t}$ is $\sqrt{1-(1-\varepsilon)^{2t}} \leq \sqrt{2t\varepsilon}$. Therefore, distinguishing between f_{yes} and f_{no} with success probability $\geq 2/3$ requires $t \geq \Omega(1/\varepsilon)$ copies of the respective function state.

We prove Theorem 57 via a combinatorial analysis of the spectrum of the matrix

$$A := \mathop{\mathbb{E}}_{\psi\sim E_0} \psi^{\otimes t} - \mathop{\mathbb{E}}_{\phi\sim E_1} \phi^{\otimes t},$$

where E_0 and E_1 are the "yes" and "no" ensembles from [Gol+00] (or, later, from [Bla24]). As neither of our ensembles is close to Haar-random, we cannot directly draw on the rich recent literature on quantum pseudorandomness [JLS18; BS19; GB23; JMW24; MPSY24; Che+24a; SHH24; MH24]. Instead, we notice that our function state is unitarily equivalent to a phase state for a closely-related Boolean function. An intricate index rearrangement reveals A to be block-diagonal, with each block interpretable as the adjacency matrix for a complete bipartite graph. We then determine the spectrum of each block, with eigenvalues and their multiplicities given as functions of certain combinatorial quantities. Exponential generating function techniques lead to explicit formulas for these quantities, and finally the asymptotics can be understood by taking a probabilistic perspective on the counting formulas. Concentration arguments finish the proof and allow us to conclude that $\|A\|_1 \geq \Omega(1)$ (and thus the two ensembles are distinguishable) as soon as $t = \Omega(\varepsilon^{-1})$. This argument is presented in detail in Section 9.4.

A final remark for this section: for certain regimes of ε, the $\Omega(1/\varepsilon)$ lower bound on the number of function state copies already separates passive quantum testing from quantum query-based testing. For example, (adaptive) quantum query complexity upper bounds of $\tilde{\mathcal{O}}(n^{1/4}/\varepsilon^{1/2})$ for monotonicity testing [BB15]

[4] A classical query complexity lower bound of $\Omega(1/\varepsilon)$ also holds for testing any non-trivial property [Fis24].

and of $\tilde{\mathcal{O}}((k/\varepsilon)^{1/2})$ for k-junta testing [ABRW15] are known. However, to the best of our knowledge, the "correct" ε-scaling for quantum property testing of classical functions is far from understood; prior works such as [AS07; BFNR08; CFMW10; AA15; ABRW15; MW16] seem to establish quantum query complexity lower bounds only for constant ε.

Outlook and future directions

Our results, (most of them) summarized in Table 9.2, highlight passive quantum property testing as a rich testing model deserving of concerted study. We grow the list of properties with efficient passive quantum testers, introduce new techniques for testing, show that the abilities of passive quantum testing extend beyond the reach of Fourier sampling, and highlight subtleties in comparing classical and quantum resources for property testing.

	Quantum function states	Classical samples	Classical queries
Monotonicity testing	$\tilde{\mathcal{O}}(n^2/\varepsilon^2)$ Theorem 63	$\exp\Big(\Omega\big(\min\{\sqrt{n}/\varepsilon, n\}\big)\Big)$ [Gol+00; Bla24]	$\tilde{\mathcal{O}}\Big(\min\Big\{n/\varepsilon, \sqrt{n}/\varepsilon^2\Big\}\Big)$ [Gol+00; KMS18]
Symmetry testing	$\tilde{\mathcal{O}}(1/\varepsilon^2)$ Theorem 60	$\Theta(n^{1/4})$ [AHW16]	$\mathcal{O}(1/\varepsilon)$ [BWY15]
Triangle-freeness testing	$\tilde{\mathcal{O}}\Big(\Big(\mathsf{Tower}\Big(C \cdot \Big\lceil \log\Big(\frac{1}{\varepsilon}\Big)\Big\rceil\Big)\Big)^6\Big)$ Theorem 66	$\Omega(n)$ via [AHW16]	$\mathcal{O}\Big(\mathsf{Tower}\Big(C \cdot \Big\lceil \log\Big(\frac{1}{\varepsilon}\Big)\Big\rceil\Big)\Big)$ via [Fox11; HST16]
3-fold intersection estimation	$\Omega\big(2^{n/2}\big)$ Theorem 56	$\Omega\big(2^{n/2}\big)$ via Theorem 56	$\mathcal{O}(1)$ Theorem 56
FORRELATION	$\mathcal{O}(1)$ [AA15]	$\tilde{\Omega}\big(2^{n/2}\big)$ [AA15]	$\tilde{\Theta}\big(2^{n/2}\big)$ [AA15]

Table 9.2: **Our bounds in context.** The table contrasts our results on property testing from quantum function states with results from the literature (in gray). Where the ε-dependence is not shown explicitly, we have set ε to some suitably small positive constant value. For monotonicity, symmetry, and triangle-freeness, passive quantum testing from function states is (at least) exponentially easier than passive classical testing from samples and at most polynomially harder than classical testing from queries. The testing problem derived from 3-fold intersection estimation is complementary to the FORRELATION problem in that quantum function/phase states and classical queries swap roles in the exponential separation.

In fact, it seems passive quantum testing can make good on the promise of testing from data where classical passive testing cannot. With passive quantum testing, it is possible to generate a dataset about a Boolean function without foreknowledge of the property one would eventually like to test, and still be assured (for a growing list of properties) that testing will be very efficient. In particular, these results suggest that *quantum data*, rather than classical data, could enable the application to machine learning imagined in [GGR98]: as an inexpensive preprocessing procedure that informs the choice of suitable, more data-intensive learning algorithms.

Here we lay out some directions for future work.

More and improved bounds for passive quantum property testing. We have established upper bounds for passive quantum testing of monotonicity, symmetry, and triangle-freeness from function states. These three properties together with linearity testing [BV97] and junta testing [AS07; ABRW15] already demonstrate the power of quantum data for testing a variety of quite different properties, and it seems important to explore quantum datasets in the context of other testing problems. As highlighted in Table 9.1, quantum low-degree testing of Boolean functions is a natural next challenge, with the more general class of locally characterized affine-invariant properties [Bha+13] as a longer-term goal.

One may aim to tighten our bounds to precisely pin down the power of quantum data for these testing tasks. Here, having established that the constructions from classical passive monotonicity testing lower bounds are inadequate for the quantum case, we consider it especially interesting to obtain a n-dependent lower bound for passive quantum monotonicity testing in the constant ε regime. Settling the n-dependence of the quantum sample complexity for passive quantum monotonicity testing is a tantalizing question for future work.

Characterizing properties with constant-complexity passive quantum testers In the classical case, [BY19] gave a complete characterization of those properties that can be tested with a constant number of samples. Achieving an analogous characterization for properties that can be tested from constantly-many function state copies would help demarcate the boundary of quantum advantage for this model.

Intriguingly, the quantum case raises a further question about the constant complexity regime: For properties that admit a constant-copy passive quantum testers, can this always be achieved by algorithms that do not use entangled multi-copy measurements? The role of single- versus multi-copy quantum processing has recently been explored in the literature on learning and testing for quantum objects (see, e.g., [CCHL22; Hua+22; Car24; HH24]) and in quantum computational learning theory [ABDY23], but the picture is far from clear for properties of function states (and of pure states more generally). Concretely, while our testers for monotonicity and symmetry are single-copy algorithms, our triangle-freeness tester uses two-copy SWAP tests and there does not seem to be an immediate way of replacing this by single-copy quantum processing.

One may also ask about the necessity of auxiliary quantum systems in quantum sample-based testers with constant sample complexity. (For example, our symmetry tester relied on auxiliary systems to implement the symmetric subspace projector.) The number of available auxiliary systems is already known to play an important role in, for instance, Pauli channel learning [CZSJ22; Che+24b; CG24], and exploring its relevance for constant-complexity passive quantum testing may shed new light on how these quantum testers achieve their better-than-classical performance.

Other quantum datasets for classical properties We have considered only one kind of quantum representation of classical functions: coherent superpositions of evaluations of f (as function states). Already these are enough to gain major advantages over testing from classical data, but one could ask for more. Are there other, better quantum datasets that lead to even faster testers or extend quantum advantage to more properties? To keep this question interesting, one would require that the dataset be not too tailored to any property.

In fact, this question may be best phrased as a sort of "compression game": we are first given a very long list of questions that we might be asked regarding some black-box function f. We then have $T(n)$ time to interact with an oracle for f, during which period we generate whatever data we would like. What is the best quantum dataset to generate, so that we are best prepared to answer a random (or perhaps worst-case) question from the list?

Passive quantum testing for quantum properties. Recently there has been a growing interest in property testing for quantum objects, such as states [HM13; OW15; HLM17; CHST17; BO20; GNW21; SW22; CCHL22; GIKL23; AD24; CGYZ24; ABD24; BDH24], unitaries [DGRT22; CNY23; SY23], channels [CCHL22; ACQ22; BY23], and Hamiltonians [LW22; BCO24; ADG24]. It is an interesting challenge to design datasets to enable passive versions of these tasks. Just as in the above, we would want quantum datasets that are mostly agnostic to the property to be tested.

In fact, some existing work can be viewed as advocating for quantum datasets. When restricting ourselves to collecting classical data, classical shadows [HKP20; HCP23] serve as a useful representation, but place restrictions on the properties that can be tested after-the-fact. Shadow tomography procedures [Aar18; BO21; Car24] can remove such restrictions but use multi-copy measurements that depend on the properties of interest, and thus in general seem to require quantum data storage to enable passivity.[5] The relevance of data storage in a quantum memory for certain quantum process learning tasks has also been explored in [Bis+10; BDPS11; SBZ19; LKPP22; LK24]. In this context, the contents of this chapter can be viewed as investigating the power of quantum data, stored in quantum memory, for testing properties of diagonal unitary processes arising from classical Boolean functions. We hope that this will inspire future attempts at using quantum data as a resource for passively quantumly testing properties of more general quantum processes.

Related work

Passive classical property testing.[6] Passive (or sample-based) property testing goes back to [GGR98] (see also [KR98]), who introduced it as a testing counterpart to Valiant's model of probably approximately correct (PAC) learning [Val84]. In particular, [GGR98, Proposition 3.1.1] observes that PAC

[5](Non-adaptive) Pauli shadow tomography [HKP21; Car24; KGKB24; CGY24] in some sense interpolates between the (dis-)advantages of classical shadows and shadow tomography for the current discussion: When promised in advance that the properties in question are characterized by expectation values of arbitrary Pauli observables, some of the relevant data can be collected and stored classically in advance, without knowing which specific Pauli observables matter. However, part of the quantum processing still requires knowing the specific Paulis of interest, so to achieve passivity, it seems that some data still has to be stored quantumly.

[6]Due to the vastness of the area of property testing, even when restricting the focus to passive testing, this paragraph is intended to provide context for this chapter rather than an exhaustive bibliography for the field.

learners give rise to passive testers (see also [Ron08, Proposition 2.1]). Later, [BBBY12] proposed active testing as a model interpolating between sample- and query-based testing. Both for passive and active testing, and for a variety of problems, several works have established lower bounds separating them from the more standard query-based testing model. Some notable examples of tasks with such separations include (k-)linearity [BBBY12; AHW16], k-juntaness [AHW16], (partial) symmetry [BWY15; AHW16], low-degreeness [AHW16], and monotonicity [Gol+00; Bla24]. We present these results and how they compare to quantum testing in Table 9.1.

[BY19] gave a full characterization of properties of Boolean-valued functions that admit passive testing with a constant (*i.e.*, independent of domain size) number of uniformly random samples, demonstrating that this is indeed a relatively restricted type of properties. While the works mentioned so far have focused on the case of uniformly random data points (or, in the case of active learning, uniformly random sets of admissible query points), more recently there has been renewed interest in passive distribution-free testing, see for instance [HK07; BFH21]. Finally, the framework of passive testing has also been explored for objects other than Boolean functions, especially for testing geometric properties [MORS10; BBBY12; KNOW14; Nee14; CFSS17; BMR19b; BMR19a].

Quantum property testing. The focus here is on quantumly testing properties of classical functions. This topic, considered for example in [AS07; CFMW10; HA11; AA15; ABRW15], is one of the main directions in quantum property testing, an area that goes back to [BFNR08] and is surveyed in [MW16]. However, quantum property testing also considers quantum algorithms that test properties of other classical objects from quantum data access. Notable examples of other objects to quantumly test include probability distributions [BHH11; CFMW10; GL20], graphs [ACL11; CFMW10], and groups [FSMS09; IL11]. Finally, recently there have also been significant insights in quantum property testing for quantum objects, notably states [HM13; OW15; HLM17; CHST17; BO20; GNW21; SW22; CCHL22; GIKL23; HH24; AD24; CGYZ24; ABD24; BDH24; MT24], unitaries [DGRT22; CNY23; SY23], channels [CCHL22; ACQ22; BY23], and Hamiltonians [LW22; BCO24; ADG24].

9.1 Passive quantum testing upper bounds

Defining passive quantum property testing

As outlined in the introduction, passive property testing considers testing from (non-adaptively chosen) data that does not depend on the property to be tested. We propose a quantum version of this model by considering quantum testing algorithms that have access to copies of a quantum data state. Here, we consider the following form of quantum data encoding for a Boolean function $f : \{0,1\}^n \to \{0,1\}$: We work with *function states*

$$|f\rangle = |\Psi_f\rangle = \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} |x, f(x)\rangle \ . \tag{9.1.1}$$

When the function f is clear from context, we will also use the notation $|\Psi\rangle = |\Psi_f\rangle$. Natural variations of this notation, *e.g.*, $|\Psi'\rangle = |\Psi_{f'}\rangle$, will also be used.

With this, we can now formally define the notion of passive quantum property testing for Boolean functions.

Definition 4 (Passive quantum property testing)**.** *Let $\mathcal{P}_n \subseteq \{0,1\}^{\{0,1\}^n}$ be some property of Boolean functions on n bits, let $\delta, \varepsilon \in (0,1)$. A quantum algorithm is a* passive quantum tester *with accuracy/distance parameter ε and confidence parameter δ for $\mathcal{P}_n$ from $m = m(\varepsilon, \delta)$ function state copies if the following holds: When given m copies of $|\Psi_f\rangle$, the quantum algorithm correctly decides, with success probability $\geq 1 - \delta$, whether*

(i) $f \in \mathcal{P}_n$*, or*

(ii) $\Pr_{x \sim \{0,1\}^n}[f(x) \neq g(x)] \geq \varepsilon$ *holds for all* $g \in \mathcal{P}_n$*,*

promised that f satisfies either (i) or (ii).

This chapter explores Definition 4 for different properties.

Passive quantum symmetry testing

A function $f : \{0,1\}^n \to \{0,1\}$ is called symmetric if $f \circ \pi = f$ holds for all permutations $\pi \in S_n$. Here, $\pi \in S_n$ acts on n-bit strings by permuting coordinates. That is, $\pi(x_1 \dots x_n) = x_{\pi^{-1}(1)} \dots x_{\pi^{-1}(n)}$. This gives rise to the following classical testing problem.

Problem 58 (Classical symmetry testing)**.** *Given query access to an unknown function $f : \{0,1\}^n \to \{0,1\}$ and an accuracy parameter $\varepsilon \in (0,1)$, decide with success probability $\geq 2/3$ whether*

(i) f is symmetric, or

(ii) f is ε-far from all symmetric functions, that is, we have $\Pr_{x\sim\{0,1\}^n}[f(x) \neq g(x)] \geq \varepsilon$ for all symmetric functions $g : \{0,1\}^n \to \{0,1\}$,

promised that f satisfies either (i) or (ii).

Symmetry allows for a (trivial) reformulation in terms of (in general non-local) pairwise comparisons:

Proposition 1. *A function $f : \{0,1\}^n \to \{0,1\}$ is symmetric if and only if for all $x \in \{0,1\}^n$ and for all $\pi \in S_n$, the equality*

$$f(x) = f(\pi(x)) \tag{9.1.2}$$

holds.

This characterization becomes important for testing because of the following result.

Theorem 59 (Soundness of symmetry testing (compare [BWY15, Lemma 3.3]))**.** *If $f : \{0,1\}^n \to \{0,1\}$ is exactly ε-far from all symmetric functions, then*

$$\varepsilon \leq \Pr_{x\sim\{0,1\}^n, \pi\sim S_n}[f(x) \neq f(\pi(x))] \leq 2\varepsilon\,. \tag{9.1.3}$$

Theorem 59 implies that we can classically test symmetry from query access simply by sampling a random permutation π and a random input x and then comparing the function values $f(x)$ and $f(\pi(x))$. Here, $\sim 1/\varepsilon$ many queries suffice to achieve success probability $\geq 2/3$ in symmetry testing.

We now describe how to make use of Theorem 59 to build a passive quantum symmetry tester.

Theorem 60 (Passive quantum symmetry testing)**.** *There is an efficient quantum algorithm that uses $\mathcal{O}\left(\frac{\log(1/\delta)}{\varepsilon^2}\right)$ many copies of the function state $|\Psi\rangle = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} |x, f(x)\rangle$ to decide, with success probability $\geq 1-\delta$, whether f is symmetric or ε-far from all symmetric functions.*

Proof. For a permutation $\pi \in S_n$, write $P(\pi)$ for the representation of that permutation on $(\mathbb{C}^2)^{\otimes n}$ given as $P(\pi) = \sum_{x\in\{0,1\}^n} |\pi(x)\rangle\langle x|$. Then, the orthogonal projector onto the symmetric subspace of $(\mathbb{C}^2)^{\otimes n}$ can be written as

$$P^n_{\mathrm{sym}} = \frac{1}{n!}\sum_{\pi\in S_n} P(\pi)\,.$$

Notice that, if $|\Psi\rangle$ is the function state for $f : \{0,1\}^n \to \{0,1\}$, then

$$\langle\Psi|\,(P^n_{\mathrm{sym}} \otimes \mathbb{1}_2)\,|\Psi\rangle = \Pr_{x\sim\{0,1\}^n,\pi\sim S_n}[f(x) \neq f(\pi(x))]\,.$$

So, by a Chernoff-Hoeffding bound, we can, with success probability $\geq 1-\delta$, obtain a $(\varepsilon/3)$-accurate estimate of the probability in Equation (9.1.3) by independently performing the two-outcome projective measurement $\{P^n_{\mathrm{sym}} \otimes \mathbb{1}_2, \mathbb{1}_2^{\otimes(n+1)} - P^n_{\mathrm{sym}} \otimes \mathbb{1}_2\}$ on $m = \mathcal{O}(\log(1/\delta)/\varepsilon^2)$ many single copies of $|\Psi\rangle$ and then taking the empirical average of the observed outcomes (with outcome 1 associated to $P^n_{\mathrm{sym}}\otimes\mathbb{1}_2$). As the two-outcome measurement $\{P^n_{\mathrm{sym}}\otimes\mathbb{1}_2, \mathbb{1}_2^{\otimes(n+1)} - P^n_{\mathrm{sym}} \otimes \mathbb{1}_2\}$ can be implemented efficiently using $\mathcal{O}(n^2)$ auxiliary qubits and $\mathcal{O}(n^2)$ controlled-SWAP gates [Bar+97; LW22], our quantum symmetry tester is also computationally efficient. □

In contrast to the classical sample complexity of $\Theta(n^{1/4})$ for classical passive symmetry testing [AHW16], our passive quantum symmetry tester in Theorem 60 achieves an n-independent quantum sample complexity. Thus, we have an unbounded separation between classical and quantum for this passive testing task.

Finally, let us comment on two extensions. Firstly, relying on the second inequality in Theorem 59, we can modify the proof of Theorem 60 to obtain an efficient tolerant quantum passive symmetry tester that uses $\mathcal{O}\big(\frac{\log(1/\delta)}{(\varepsilon_2-\varepsilon_1)^2}\big)$ copies of the unknown function state to decide whether f is ε_1-close to or ε_2-far from symmetric, assuming that $\varepsilon_2 > 2C\varepsilon_1$ holds with $C > 1$ some constant. Secondly, as Theorem 59 can be extended to so-called partial symmetric functions (compare again [BWY15, Lemma 3.3]), also our passive quantum symmetry tester can be modified to test for partial symmetry.

Passive quantum monotonicity testing

We define the natural partial order $\preceq$ on the Boolean hypercube $\{0,1\}^n$ via $x \preceq y \;:\Leftrightarrow\; (x_i \leq y_i \text{ holds for all } 1 \leq i \leq n)$. A function $f : \{0,1\}^n \to \{0,1\}$ is called monotone if $f(x) \leq f(y)$ holds for all $x, y \in \{0,1\}^n$ with $x \preceq y$. The associated classical testing problem can be formulated as follows.

Problem 61 (Classical monotonicity testing)**.** *Given query access to an unknown function $f : \{0,1\}^n \to \{0,1\}$ and an accuracy parameter $\varepsilon \in (0,1)$, decide with success probability $\geq 2/3$ whether*

(i) f is monotone, or

(ii) f is ε-far from all monotone functions, that is, we have $\Pr_{x\sim\{0,1\}^n}[f(x) \neq g(x)] \geq \varepsilon$ for all monotone functions $g : \{0,1\}^n \to \{0,1\}$,

promised that f satisfies either (i) or (ii).

Here, as well as in our other property testing tasks below, will think of (i) as the accept case and of (ii) as the reject case. This then allows us to speak of completeness (for getting case (i) right) and soundness (for getting case (ii) right). Here, the chosen success probability of $2/3$ is an arbitrary constant $> 1/2$, it can be boosted arbitrarily close to 1 through repetition and majority voting.

As introduced above, monotonicity is a global property of a function. However, there is a straightforward equivalent local formulation:

Proposition 2 (Local characterization of monotonicity)**.** *A function $f : \{0,1\}^n \to \{0,1\}$ is monotone if and only if for all $x \in \{0,1\}^n$ and for all $i \in [n]$, the following holds:*

$$\neg((x_i = 0 \wedge f(x) = 1 \wedge f(x+e_i) = 0) \vee (x_i = 1 \wedge f(x) = 0 \wedge f(x+e_i) = 1))\,, \tag{9.1.4}$$

where e_i denotes the i^{th} standard basis vector.

It turns out that functions far from the set of all monotone functions necessarily violate Equation (9.1.4) on a non-negligible fraction of all possible x and i. This makes it possible to test for monotonicity by checking Equation (9.1.4) on a small number of randomly chosen x and i.

Theorem 62 (Soundness of monotonicity testing (compare [Gol+00]))**.** *If $f : \{0,1\}^n \to \{0,1\}$ is exactly ε-far from all monotone functions, then*

$$\frac{\varepsilon}{n} \leq \Pr_{x\sim\{0,1\}^n, i\sim[n]}[(x_i = 0 \wedge f(x) = 1 \wedge f(x+e_i) = 0) \vee (x_i = 1 \wedge f(x) = 0 \wedge f(x+e_i) = 1)] \leq 2\varepsilon\,. \tag{9.1.5}$$

Therefore, we can solve Problem 61 from only $\mathcal{O}(n/\varepsilon)$ many queries to the unknown function, which was exactly the celebrated conclusion of [Gol+00].

While this query complexity does depend on n, the dependence is only logarithmic in the size of the function domain, and it in particular is exponentially better than the n-dependence in the query complexity of learning monotone functions [BBL98].

Our passive quantum monotonicity tester also crucially relies on Theorem 62. Here, we first reinterpret the probability appearing in Equation (9.1.5) in terms of Fourier-analytic quantities, which we then estimate based on quantum Fourier sampling. Our procedure is summarized in Algorithm 1, and our next theorem establishes that it is both complete and sound.

Algorithm 1 Monotonicity testing from quantum examples

Input: accuracy parameter $\varepsilon \in (0,1)$; confidence parameter $\delta \in (0,1)$; $\tilde{\mathcal{O}}\Big(\frac{n^2 \log(1/\delta)}{\varepsilon^2}\Big)$ many copies of a function state $|f\rangle = \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} |x, f(x)\rangle$.

Output: "accept" or "reject".

Initialization: $\varepsilon_2 = \frac{\varepsilon}{3}$, $\varepsilon_5 = \frac{\varepsilon}{3n}$, $\delta_1 = \delta_2 = \frac{\delta}{3}$, $\delta_5 = \frac{\delta}{3n}$, $m_1 = \max\{3m_2, \lceil 18 \ln(2/\delta_1) \rceil\}$, $m_2 = \left\lceil \frac{n^2 \ln(2/\delta_2)}{2\varepsilon_2^2} \right\rceil$, $m_4 = m_5 = \left\lceil \frac{4 \ln(2/\delta_5)}{\varepsilon_5^2} \right\rceil$

1: Use m_1 many copies of $|f\rangle$ to produce m_2 many Fourier samples $S_1, \ldots, S_{m_2} \subseteq [n]$ from $g = (-1)^f$.
2: Take $\hat{\mathbf{I}} = \frac{1}{m_2} \sum_{\ell=1}^{m_2} |S_\ell|$.
3: Use m_4 many copies of $|f\rangle$ to generate m_5 many classical samples $(x_1, f(x_1)), \ldots, (x_{m_5}, f(x_{m_5})$ from f.
4: **for** $1 \leq i \leq n$ **do**
5: Take $\tilde{g}_i = \frac{1}{m_5} \sum_{k=1}^{m_5} (-1)^{x_k \cdot e_i + f(x_k)}$.
6: **end for**
7: Set $\hat{p} = \frac{1}{2n} \hat{\mathbf{I}} - \frac{1}{2n} \sum_{i=1}^{n} \tilde{g}_i$.
8: If $\hat{p} \leq \varepsilon/3n$, conclude that f is monotone and accept. If $\hat{p} \geq 2\varepsilon/3n$, conclude that f is ε-far from all monotone functions and reject.

Theorem 63 (Passive quantum monotonicity testing)**.** *Algorithm 1 is an efficient quantum algorithm that uses $\tilde{\mathcal{O}}\Big(\frac{n^2 \log(1/\delta)}{\varepsilon^2}\Big)$ copies of $|f\rangle = \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} |x, f(x)\rangle$ to decide, with success probability $\geq 1 - \delta$, whether f is monotone or ε-far from monotone.*

In particular, Theorem 63 shows that passive quantum testers can exponentially outperform the classical passive monotonicity testing lower bound of $\exp\Big(\Omega\Big(\min\{\sqrt{n}/\varepsilon, n\}\Big)\Big)$ [Gol+00; Bla24].

Proof. We begin with a useful rewriting of the probability from Equation (9.1.5). To this end, as is commonly done in the analysis of Boolean functions, consider the induced function $g : \{-1,1\}^n \to \{-1,1\}$ obtained from f via the relabeling

$0 \leftrightarrow 1$ and $1 \leftrightarrow -1$. Next, we recall the definition of the ith derivative in Boolean analysis (compare, *e.g.*, [ODo21, Definition 2.16]): for $i \in [n]$,

$$D_i g(x) \coloneqq \frac{g(x^{(i \mapsto 1)}) - g(x^{(i \mapsto -1)})}{2},$$

where we used the notation $x^{(i \mapsto b)}$ to denote the n-bit string obtained from x by replacing the ith bit with b. Consequently, we can compute

$$\begin{aligned}\frac{(D_i g(x))^2 - D_i g(x)}{2} &= \begin{cases} 1 & \text{if } g(x^{(i \mapsto 1)}) = -1 \wedge g(x^{(i \mapsto -1)}) = 1 \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} 1 & \text{if } f(x^{(i \mapsto 0)}) = 1 \wedge f(x^{(i \mapsto 1)}) = 0 \\ 0 & \text{otherwise} \end{cases}.\end{aligned}$$

Therefore, we can now rewrite our probability of interest as

$$\begin{aligned}&\Pr_{x \sim \{0,1\}^n, i \sim [n]}[(x_i = 0 \wedge f(x) = 1 \wedge f(x + e_i) = 0) \vee (x_i = 1 \wedge f(x) = 0 \wedge f(x + e_i) = 1)] \\ &= \mathbb{E}_{x \sim \{0,1\}^n, i \sim [n]}\left[\frac{(D_i g(x))^2 - D_i g(x)}{2}\right] \\ &= \frac{1}{2}\mathbb{E}_{i \sim [n]}\mathbb{E}_{x \sim \{0,1\}^n}\left[(D_i g(x))^2\right] - \frac{1}{2}\mathbb{E}_{i \sim [n]}\mathbb{E}_{x \sim \{0,1\}^n}[D_i g(x)] \\ &= \frac{1}{2}\mathbb{E}_{i \sim [n]}[\mathbf{Inf}_i[g]] - \frac{1}{2n}\sum_{i=1}^{n} \widehat{g}(\{i\}) \\ &= \frac{1}{2n}\mathbf{I}[g] - \frac{1}{2n}\sum_{i=1}^{n} \widehat{g}(\{i\}),\end{aligned}$$

where the second-to-last step used the definition of the ith influence (compare [ODo21, Definition 2.17]) as well as [ODo21, Proposition 2.19], and where the last step used the definition of the total influence (compare [ODo21, Definition 2.27]).

With this rewriting established, let us first analyze the probabilities that the different steps of Algorithm 1 succeed and discuss what this implies for the estimator $\hat{p}$. Then, we will see how this gives rise to completeness and soundness. We have the following:

- Using the procedure of [BV97], one copy of $|f\rangle$ suffices to produce one Fourier sample from $g = (-1)^f$—that is, an n-bit string sampled from the probability distribution $\{|\widehat{g}(S)|^2\}_{S \subseteq [n]}$—with success probability $1/2$. Additionally, one knows whether the sampling attempt was successful.[7]

[7] To see this, note that the procedure works as follows: Apply $H^{\otimes(n+1)}$; measure the last qubit; abort if that produces a 0, continue if produces a 1; measure the first n qubits to produce an n-bit string.

So, by simply repeating the above m_1 many times, we see that Step 1 succeeds in producing m_2 Fourier samples with success probability $\geq 1 - \delta_1$.

- By a standard Chernoff-Hoeffding bound, we have $|\hat{\mathbf{I}} - \mathbb{E}_{S \sim \mathcal{S}_g}[|S|]| \leq \varepsilon_2$ with success probability $\geq 1 - \delta_2$. Here, $\mathcal{S}_g$ denotes the Fourier distribution of g, defined via $\mathcal{S}_g(S) = |\widehat{g}(S)|^2$.
- For any $1 \leq i \leq n$, by a standard Chernoff-Hoeffding bound, Step 5 produces a ε_5-accurate estimate $\tilde{g}_i$ of $\widehat{g}(\{i\})$ with probability $\geq 1 - \delta_5$.

Therefore, by a union bound, we have that, with overall success probability $\geq 1 - (\delta_1 + \delta_2 + n\delta_5) = 1 - \delta$, the estimates $\hat{\mathbf{I}}$ and $\tilde{g}_i$ simultaneously satisfy $|\hat{\mathbf{I}} - \mathbb{E}_{S \sim \mathcal{S}_g}[|S|]| \leq \varepsilon_2$ and $|\tilde{g}_i - \widehat{g}(\{i\})| \leq \varepsilon_5$ for all $1 \leq i \leq n$. We condition on this high probability success event for the rest of the proof. In this event, our rewriting of the probability of interest from the beginning of the proof implies:

$$
\begin{aligned}
&\left| \Pr_{x \sim \{0,1\}^n, i \sim [n]}[(x_i = 0 \wedge f(x) = 1 \wedge f(x + e_i) = 0) \vee (x_i = 1 \wedge f(x) = 0 \wedge f(x + e_i) = 1)] - \hat{p} \right| \\
&\leq \frac{1}{2n} \left| \mathbf{I}[g] - \hat{\mathbf{I}} \right| + \frac{1}{2n} \sum_{i=1}^{n} |\widehat{g}(\{i\}) - \tilde{g}_i| \\
&= \frac{1}{2n} \left| \mathbb{E}_{S \sim \mathcal{S}_g}[|S|] - \hat{\mathbf{I}} \right| + \frac{1}{2n} \sum_{i=1}^{n} |\widehat{g}(\{i\}) - \tilde{g}_i| \\
&\leq \frac{\varepsilon_2}{2n} + \frac{\varepsilon_5}{2} \\
&\leq \frac{\varepsilon}{3n},
\end{aligned}
$$

where we used the identity $\mathbf{I}[g] = \mathbb{E}_{S \sim \mathcal{S}_g}[|S|]$ (compare [ODo21, Theorem 2.38]). So, we see that $\hat{p}$ is a $(\varepsilon/3n)$-accurate estimate for our probability of interest.

To prove completeness of Algorithm 1, assume f to be monotone. Then, Proposition 2 and Theorem 62 together with the above inequality imply that $\hat{p} \leq \varepsilon/3n \leq \varepsilon/2n$, and thus the final step in Algorithm 1 correctly concludes that f is monotone and accepts.

To prove soundness, assume f to be ε-far from monotone. Then, the lower bound in Theorem 62 together with the above inequality implies that $\hat{p} \geq 2\varepsilon/3n$, and thus the final step in Algorithm 1 correctly concludes that f is ε-far from monotone and rejects.

The overall number of copies of $|f\rangle$ used by the algorithm is $m_1 + m_4$. Plugging in the choices of the different m_i, we see that

$$\begin{aligned} m_1 + m_4 &\leq \max\{3m_2, \lceil 18\ln(2/\delta_1)\rceil\} + \left\lceil \frac{4\ln(2/\delta_5)}{\varepsilon_5^2} \right\rceil \\ &\leq \max\left\{3\left\lceil \frac{n^2\ln(2/\delta_2)}{2\varepsilon_2^2}\right\rceil, \lceil 18\ln(2/\delta_1)\rceil\right\} + \left\lceil \frac{36n^2\ln(6n/\delta)}{\varepsilon^2}\right\rceil \\ &\leq \max\left\{3\left\lceil \frac{9n^2\ln(6/\delta)}{2\varepsilon^2}\right\rceil, \lceil 18\ln(6/\delta)\rceil\right\} + \left\lceil \frac{36n^2\ln(6n/\delta)}{\varepsilon^2}\right\rceil \\ &\leq \tilde{\mathcal{O}}\left(\frac{n^2\log(1/\delta)}{\varepsilon^2}\right), \end{aligned}$$

where the $\tilde{\mathcal{O}}$ hides a logarithmic dependence on n.

The quantum computational efficiency of Algorithm 1 follows immediately from the efficiency of quantum Fourier sampling. The classical computational efficiency is immediately apparent from our sample complexity bounds and the fact that the classical computation is dominated by the complexity of computing the empirical averages in Steps 2 and 4. □

We further note that because of the second inequality in Theorem 62, the above procedure and proofs can be modified to quantumly efficiently solve the *tolerant* version (as defined in [PRR06]) of the monotonicity testing problem—*i.e.*, distinguishing between f being ε_1-close or ε_2-far from monotone—using $\tilde{\mathcal{O}}\left(\frac{n^2\log(1/\delta)}{(\varepsilon_2-\varepsilon_1)^2}\right)$ copies of a quantum function state, assuming that $\varepsilon_2 > Cn\varepsilon_1$ holds with $C > 1$ some constant.[8] Because of this restrictive assumption on how ε_1 and ε_2 relate, this still falls short of a general tolerant passive quantum monotonicity tester.

Let us also note room for a qualitative improvement in our passive quantum monotonicity tester. Classical query-based testers typically enjoy perfect completeness, *i.e.*, they accept monotone functions with unit probability. In contrast, our tester can be made to accept monotone functions with probability arbitrarily close but not equal to 1. We leave as an open question whether our passive quantum monotonicity tester can be modified to achieve perfect completeness, while enjoying similar guarantees on the quantum sample and time complexity of the procedure.

[8] In more generality, one can see: If an inequality like Equation (9.1.5) holds with lower bound $\ell_n(\varepsilon)$ and upper bound $u_n(\varepsilon)$, satisfying $\ell_n(0) = 0 = u_n(0)$, then estimating the relevant probability to accuracy $\sim \ell_n(\varepsilon_2 - \varepsilon_1)$ suffices for tolerant property testing in the parameter range where there is a constant $c \in (0, 1/2)$ such that $\ell_n(\varepsilon_2) - c\cdot\ell_n(\varepsilon_2 - \varepsilon_1) > u_n(\varepsilon_1) + c\cdot\ell_n(\varepsilon_2 - \varepsilon_1)$.

Passive quantum triangle-freeness testing

For $x, y \in \{0,1\}^n$ and for a Boolean function $f : \{0,1\}^n \to \{0,1\}$, we say that $(x, y, x+y)$ is a triangle in f if $f(x) = f(y) = f(x+y) = 1$. Accordingly, we call the function f triangle-free if no triple $(x, y, x+y)$ is a triangle in f. Testing for triangle-freeness thus becomes the following problem.

Problem 64 (Classical triangle-freeness testing)**.** *Given query access to an unknown function $f : \{0,1\}^n \to \{0,1\}$ and an accuracy parameter $\varepsilon \in (0,1)$, decide with success probability $\geq 2/3$ whether*

(i) f is triangle-free, or

(ii) f is ε-far from all triangle-free functions, that is, we have $\Pr_{x \sim \{0,1\}^n}[f(x) \neq g(x)] \geq \varepsilon$ for all triangle-free functions $g : \{0,1\}^n \to \{0,1\}$,

promised that f satisfies either (i) or (ii).

The natural approach towards testing for triangle-freeness from query access is to choose $x, y \in \{0,1\}^n$ at random and check whether $(x, y, x+y)$ is a triangle in f, and to repeat this sufficiently often. Bounding the number of repetitions needed to succeed with this approach is non-trivial, connecting to Szemerédi's regularity lemma [Sze76] and the triangle removal lemma [RS78]. In our next result, we recall the to our knowledge best known corresponding bounds.

Theorem 65 (Soundness of triangle-freeness testing [Fox11; HST16])**.** *If $f : \{0,1\}^n \to \{0,1\}$ is ε-far from all triangle-free functions, then*

$$\Pr_{x,y \sim \{0,1\}^n}[f(x) = f(y) = f(x+y) = 1] \geq \frac{1}{\mathsf{Tower}\Big(C \cdot \Big\lceil \log\Big(\frac{1}{\varepsilon}\Big) \Big\rceil\Big)}\,, \tag{9.1.6}$$

where $C > 0$ is a universal integer constant.

Here, $\mathsf{Tower}(i)$ denotes a tower of 2's of height i. That is, we define the tower function $\mathsf{Tower} : \mathbb{N} \to \mathbb{N}$ inductively via $\mathsf{Tower}(0) = 1$ and $\mathsf{Tower}(i+1) = 2^{\mathsf{Tower}(i)}$. In a way familiar by now from the two previous subsections, Theorem 65 can be used to show that $\sim \mathsf{Tower}\Big(C \cdot \Big\lceil \log\Big(\frac{1}{\varepsilon}\Big) \Big\rceil\Big)$ many queries suffice for the the simple query-based triangle-freeness tester mentioned above to achieve success probability $2/3$.

We now use Theorem 65 to develop a passive quantum triangle-freeness tester.

Theorem 66 (Passive quantum triangle-freeness testing)**.** *There is an efficient quantum algorithm that uses* $\tilde{\mathcal{O}}\left(\ln(1/\delta)\left(\mathsf{Tower}\left(C\cdot\left\lceil\log\left(\frac{1}{\varepsilon}\right)\right\rceil\right)\right)^6\right)$ *many copies of the function state* $|\Psi\rangle = \frac{1}{\sqrt{2^n}}\sum_{x\in\{0,1\}^n}|x, f(x)\rangle$ *to decide, with success probability* $\geq 1-\delta$*, whether* f *is triangle-free or* ε*-far from all triangle-free functions.*

Proof. Our passive quantum triangle-freeness tester first sets confidence and accuracy parameters $\tilde{\delta} = \delta/(5m)$ and $\tilde{\varepsilon} = \left(\mathsf{Tower}\left(C\cdot\left\lceil\log\left(\frac{1}{\varepsilon}\right)\right\rceil\right)\right)^{-1}$, respectively. Then, it repeats the following for $1 \leq i \leq m = \left\lceil\frac{18\ln(10/\delta)}{\tilde{\varepsilon}^2}\right\rceil$:

1. Take $\frac{\ln(m/\tilde{\delta})}{\varepsilon}$ many copies of $|\Psi\rangle$ and, for each of them, measure the last qubit in the computational basis. If none of these measurements produces outcome 1, abort this iteration, set $\hat{\mu}_i = 0$, and go to the next iteration. Otherwise, take any one of the post-measurement states for which 1 was observed, measure the first n qubits, let the outcome be y_i.

2. Run the procedure from Lemma 83 on $2\left\lceil 162\ln(6/\tilde{\delta})(6/\tilde{\varepsilon})^4\right\rceil\cdot\left\lceil\frac{\ln(2\left\lceil 162\ln(6/\tilde{\delta})(6/\tilde{\varepsilon})^4\right\rceil/\tilde{\delta})}{\varepsilon}\right\rceil$ many copies of $|\Psi\rangle$ to obtain $2\left\lceil 162\ln(6/\tilde{\delta})(6/\tilde{\varepsilon})^4\right\rceil$ many copies of the post-measurement state

$$|\Psi_1\rangle = (|\{x\in\{0,1\}^n : f(x)=1\}|)^{-1/2}\sum_{x\in\{0,1\}^n:f(x)=1}|x\rangle\,,$$

   where we threw away the last qubit. If the procedure from Lemma 83 outputs FAIL, abort this iteration, set $\hat{\mu}_i = 0$, and go to the next iteration.

3. Consider the n-qubit unitary U_{y_i} acting as $U_{y_i}|x\rangle = |x+y_i\rangle$. Run the procedure from Lemma 83 on $2\left\lceil 162\ln(6/\tilde{\delta})(6/\tilde{\varepsilon})^4\right\rceil\cdot\left\lceil\frac{\ln(2\left\lceil 162\ln(6/\tilde{\delta})(6/\tilde{\varepsilon})^4\right\rceil/\tilde{\delta})}{\varepsilon}\right\rceil$ many copies of $(U_{y_i}\otimes\mathbb{1}_2)|\Psi\rangle$ to obtain $2\left\lceil 162\ln(6/\tilde{\delta})(6/\tilde{\varepsilon})^4\right\rceil$ many copies of the post-measurement state

$$|\Psi_{y_i,1}\rangle = (|\{x\in\{0,1\}^n : f(x+y_i)=1\}|)^{-1/2}\sum_{x\in\{0,1\}^n:f(x+y_i)=1}|x\rangle\,,$$

   where we threw away the last qubit. If the procedure from Lemma 83 outputs FAIL, abort this iteration, set $\hat{\mu}_i = 0$, and go to the next iteration.

4. Run the procedure from Corollary 85 on $\left\lceil 162\ln(6/\tilde{\delta})(6/\tilde{\varepsilon})^4\right\rceil$ copies of each of $|\Psi\rangle$, $(U_{y_i}\otimes\mathbb{1}_2)|\Psi\rangle$ and on $2\left\lceil 162\ln(6/\tilde{\delta})(6/\tilde{\varepsilon})^4\right\rceil$ copies of each of $|\Psi_1\rangle$, $|\Psi_{y_i,1}\rangle$ to produce an $(\tilde{\varepsilon}/6)$-accurate estimate $\hat{\mu}_i$ of the probability $\Pr_{x\sim\{0,1\}^n}[f(x)=1=f(x+y_i)]$.

Finally, the tester makes a decision as follows: if $\frac{1}{m}\sum_{i=1}^{m}\hat{\mu}_i \leq \tilde{\varepsilon}/3$, output "triangle-free". Otherwise, output "ε-far from triangle-free".

Let us analyze the completeness and soundness of this tester. First, we consider completeness. So, assume that f is triangle-free. Note the first step above failing in any iteration only ever decreases the empirical average evaluated by the tester in the end, and thus cannot increase the probability of falsely rejecting a triangle-free function. Thus, we can condition on the first step succeeding in all m iterations. In particular, we can assume that $\Pr_{y\sim\{0,1\}^n}[f(y)=1]>0$. By a similar argument, we can also condition on the second and third step succeeding in all iterations. And given these successes, the fourth step will, by Corollary 85, produce, with probability $\geq 1-\tilde{\delta}$, a $\hat{\mu}_i$ satisfying $|\hat{\mu}_i - \Pr_{x\sim\{0,1\}^n}[f(x)=1=f(x+y_i)]| \leq \tilde{\varepsilon}/6$. By a union bound, this means that, with probability $\geq 1-\delta/5$, the empirical average $\frac{1}{m}\sum_{i=1}^{m}\hat{\mu}_i$ is a $(\tilde{\varepsilon}/6)$-accurate estimate of $\frac{1}{m}\sum_{i=1}^{m}\Pr_{x\sim\{0,1\}^n}[f(x)=1=f(x+y_i)]$. We can consider the $\Pr_{x\sim\{0,1\}^n}[f(x)=1=f(x+y_i)]$ for $1\leq i\leq m$ as i.i.d. random variables taking values in $[0,1]$. Hence, by a Chernoff-Hoeffding concentration bound and our choice of m, with probability $\geq 1-\delta/5$, we have

$$\left|\frac{1}{m}\sum_{i=1}^{m}\Pr_{x\sim\{0,1\}^n}[f(x)=1=f(x+y_i)] - \mathbb{E}_{y\sim\{0,1\}^n:f(y)=1}\Big[\Pr_{x\sim\{0,1\}^n}[f(x)=1=f(x+y)]\Big]\right| \leq \tilde{\varepsilon}/6\,.$$

Noticing that

$$\mathbb{E}_{y\sim\{0,1\}^n:f(y)=1}\Big[\Pr_{x\sim\{0,1\}^n}[f(x)=1=f(x+y)]\Big] = \frac{\Pr_{x,y\sim\{0,1\}^n}[f(x)=f(y)=f(x+y)=1]}{\Pr_{y\sim\{0,1\}^n}[f(y)=1]} = 0\,, \tag{9.1.7}$$

since f was assumed to be triangle-free, we conclude (after one more union bound) that $\frac{1}{m}\sum_{i=1}^{m}\hat{\mu}_i \leq \tilde{\varepsilon}/3$ holds with probability $\geq 1-2\delta/5 \geq 1-\delta$, and in this case the tester outputs "triangle-free", thus proving completeness.

Next, we consider soundness. So, assume that f is ε-far from triangle-free. This in particular implies that $\Pr_{x\sim\{0,1\}^n}[f(x)=1]\geq\varepsilon$, since the zero-function is triangle-free. Hence, when measuring the last qubit of $|\Psi\rangle$ in the computational basis, outcome 1 is observed with probability $\geq\varepsilon$. Therefore, in any iteration i, the first step in our sketched procedure above will succeed and produce some y_i with $f(y_i)=1$ with probability $\geq 1-\tilde{\delta}$. We condition on this high-probability event E_1 for the rest of the soundness analysis. By an analogous reasoning, the assumption of Lemma 83 is satisfied in the scenario of steps 2 and 3—with $S=\{0,1\}^n$, $\eta=\varepsilon$ and the function either given directly by f or by $f(\cdot+y_i)$—so in any iteration i, the second and third step each will

succeed with probability $\geq 1 - \tilde{\delta}$. We now further condition on these success events E_2 and E_3. At this point, by the same reasoning as in the completeness case, we know that the fourth step, with success probability $\geq 1 - 2\delta/5$ overall, produces estimates $\hat{\mu}_i$ such that

$$\left| \frac{1}{m} \sum_{i=1}^{m} \hat{\mu}_i - \Pr_{x,y \sim \{0,1\}^n} [f(x) = f(y) = f(x+y) = 1] \right| \leq \tilde{\varepsilon}/3 \,.$$

By Theorem 65, since we assumed f to be ε-far from triangle-free, we have $\Pr_{x,y\sim\{0,1\}^n}[f(x) = f(y) = f(x+y) = 1] \geq \tilde{\varepsilon}$. Thus, by the first equality from Equation (9.1.7), we have

$$\mathbb{E}_{y \sim \{0,1\}^n : f(y)=1} \left[\Pr_{x \sim \{0,1\}^n} [f(x) = 1 = f(x+y)] \right] \geq \tilde{\varepsilon} \,.$$

So, the above implies the inequality $\frac{1}{m}\sum_{i=1}^{m} \geq 2\tilde{\varepsilon}/3$. Hence, the tester will in this case correctly output "ε-far from triangle-free". A final union bound shows that this occurs with probability $\geq 1 - \delta$, which proves soundness.

The quantum sample complexity of a single iteration is given by $2 \cdot 2\left\lceil 162 \ln(6/\tilde{\delta})(6/\tilde{\varepsilon})^4 \right\rceil \cdot \left\lceil \frac{\ln(2\left\lceil 162 \ln(6/\tilde{\delta})(6/\tilde{\varepsilon})^4 \right\rceil / \tilde{\delta})}{\varepsilon} \right\rceil + 2\left\lceil 162 \ln(6/\tilde{\delta})(6/\tilde{\varepsilon})^4 \right\rceil \leq \tilde{\mathcal{O}}\left(\frac{\ln(1/\tilde{\delta})}{\tilde{\varepsilon}^4}\right)$. Thus, the overall quantum sample complexity is $\leq m \cdot \tilde{\mathcal{O}}\left(\frac{\ln(1/\tilde{\delta})}{\tilde{\varepsilon}^4}\right) \leq \tilde{\mathcal{O}}\left(\frac{\ln^2(1/\tilde{\delta})}{\tilde{\varepsilon}^6}\right)$. Plugging in the chosen values for $\tilde{\delta}$ and $\tilde{\varepsilon}$ yields an upper bound of $\tilde{\mathcal{O}}\left(\ln^2(1/\delta)\left(\mathsf{Tower}\left(C \cdot \left\lceil \log\left(\frac{1}{\varepsilon}\right) \right\rceil\right)\right)^6\right)$ on the number of quantum copies used by the tester. To achieve the claimed linear dependence on $\ln(1/\delta)$, one can simply run the protocol described above for a constant confidence parameter (say, $\delta = 1/3$), and then amplify the success probability through majority votes.

Finally, we have to argue that the tester is quantumly computationally efficient. This, however, follows immediately from the efficiency of the procedures from Lemma 83 and Corollary 85, and from the fact that every U_{y_i} can be implemented by at most n Pauli-X gates. □

Classically, passive triangle-freeness testing requires at least $\Omega(n)$ samples.[9] Therefore, just like in symmetry testing, there is an unbounded separation between a classical n-dependent and a quantum n-independent passive testing sample complexity.

[9]This can be seen by an argument via linear independence similar to that used in the lower bound proof of [AHW16, Theorem 10].

9.2 Fourier sampling does not suffice

The passive quantum testers for symmetry and triangle-freeness given in Section 9.1 notably do not use quantum Fourier sampling. One might ask if this is really necessary, given that Fourier sampling (sometimes augmented by classical samples) suffices for so many other learning and testing tasks. This section presents a class of functions, Maiorana–McFarland (bent) functions, which can be tested with $\mathcal{O}(1)$ function state copies, but any algorithm relying solely on Fourier samples and classical samples requires super-polynomial classical samples to succeed.

The *Maiorana–McFarland* functions [CM16, Section 6.1] on $2n$ bits, denoted MM_n, are given by

$$\begin{aligned} f_h : \mathbb{F}_2^n \times \mathbb{F}_2^n &\to \mathbb{F}_2 \\ (x, y) &\mapsto \langle x, y \rangle + h(x), \end{aligned}$$

where h ranges over all functions $\mathbb{F}_2^n \to \mathbb{F}_2$. Maiorana–McFarland functions are a subset of the class of *bent* Boolean functions $g : \{0,1\}^m \to \{0,1\}$, which are those with $\widehat{g}(S)^2 = 1/2^m$ for all $S \subset [m]$ (so they are maximally-far from any $\mathbb{F}_2^n$-linear function).

We begin by proving hardness of testing MM_n using only classical samples $\big(x, f(x)\big)$ and Fourier samples. The proof follows swiftly once we establish the existence of k-wise independent distributions supported only on moderately-biased strings.

Lemma 67. *For sufficiently large n, there is a probability distribution on $\{0,1\}^{2^n}$ that is (i) $2^{0.1n}$-wise independent and (ii) supported on strings x with fractional hamming weight $|x|/2^n$ bounded as*

$$\left| \frac{|x|}{2^n} - \frac{1}{2} \right| \leq 2^{-n/3}.$$

Proof. Let $c > 0$ be a small universal constant to be chosen later. For simplicity we assume cn is an integer, otherwise one may round to a nearest integer without issue. For a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ let $\mathrm{eval}(f)$ denote the truth table (*viz.* a 2^n-bit string) of f. Let $\mathcal{P}_{cn-1}$ denote the set of n-variate $\mathbb{F}_2$ polynomials of degree at most $cn - 1$. We claim there exists a Boolean function f such that the ensemble

$$\Big\{ \mathrm{eval}(p \oplus f) \Big\}_{p \sim \mathcal{P}_{cn-1}}$$

satisfies properties (i) and (ii) above.

Consider the Reed–Muller code $\mathcal{C} = \mathrm{RM}(n - cn, n)$. (See [ASSY23] for the definition and properties of Reed–Muller codes.) It is well-known (*e.g.*, Section 24.2.2 in [GRS23]) that uniformly random codewords from the dual code $\mathcal{C}^\perp$ form a k-wise independent distribution on $\{0,1\}^{2^n}$ for $k := \mathrm{dist}(\mathcal{C}) - 1 = 2^{cn} - 1$, and the codewords of $\mathcal{C}^\perp = \mathrm{RM}(cn - 1, n)$ are exactly $\{\mathrm{eval}(p) : p \in \mathcal{P}_{cn-1}\}$. Invariance of the uniform distribution under addition in $\mathbb{F}_2^k$ implies that for any Boolean function f the uniform distribution over strings $\mathrm{eval}(p \oplus f) = \mathrm{eval}(p) + \mathrm{eval}(f)$ is also k-wise independent. Property (i) follows by decreasing c slightly.

To see (ii), we will argue that a random f suffices with high probability.[10] For any $\mathbb{F}_2$ polynomial p and f a uniformly random Boolean function, we have from Chernoff that

$$\Pr_f\left[\ \Big|\underbrace{\mathop{\mathbb{E}}_x\left[(-1)^{f(x)}(-1)^{p(x)}\right]}_{*}\Big| \geq 2^{-n/3}\ \right] \leq \exp(-\mathsf{const}\cdot 2^{n/3})\,. \tag{9.2.1}$$

Union bounding over the $\leq \exp(2^{3cn})$-many $p \in \mathcal{P}_{cn-1}$ we find

$$\Pr_f\Big[\ \exists p \in \mathcal{P}_{cn-1}, \Big|\mathop{\mathbb{E}}_x\left[(-1)^{f(x)}, (-1)^{p(x)}\right]\Big| \geq 2^{-n/3}\ \Big] \leq \exp\Big(-\mathsf{const}\cdot 2^{n/3} + 2^{3cn}\Big),$$

which goes to 0 for $c = 0.11$, for example. So for large-enough n there exists an f with absolute correlation at most $2^{-n/3}$ with all $p \in \mathcal{P}_{0.11n-1}$.

But the correlation $(*)$ in Equation (9.2.1) is nothing but the bias of the function $p \oplus f$. Thus for all $p \in \mathcal{P}_{0.1n} \subseteq \mathcal{P}_{0.11n-1}$, the fractional Hamming weight of $\mathrm{eval}(p \oplus f)$ is at most $2^{-n/3}$ away from $1/2$. □

Theorem 68. *Suppose a tester for* MM_n *using exclusively classical samples and Fourier samples succeeds with probability at least* $1/2 + 2^{-0.7n}$ *for the accuracy parameter* $\varepsilon = 1/2 - 2^{-0.6n}$*. Then the tester uses at least* $2^{0.1n}$ *classical samples.*

Proof. Let $\mathcal{H}$ be the distribution on n-bit functions with truth tables distributed as in Lemma 67. Consider the two ensembles of Boolean functions $\mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2$

$$\mathcal{F}_1 = \Big\{(x,y) \mapsto \langle x, y\rangle + h(x)\Big\}_{h\sim\mathcal{H}} \qquad \text{and} \qquad \mathcal{F}_2 = \Big\{(x,y) \mapsto \langle x, y\rangle + h(y)\Big\}_{h\sim\mathcal{H}},$$

Note that for any $f = \langle x, y\rangle + h(x) \in \mathrm{MM}_n$ and any $g = \langle x, y\rangle + m(y) \in \mathrm{supp}\mathcal{F}_2$,

$$\left|\mathop{\mathbb{E}}_{x,y}(-1)^{f(x,y)}(-1)^{g(x,y)}\right| = \left|\mathop{\mathbb{E}}_{x,y}(-1)^{h(x)}(-1)^{m(y)}\right| = |\mathrm{bias}(h)\mathrm{bias}(m)| \leq 2^{-2n/3}$$

[10] If one desires explicit hard instances one can use the correlation bounds of Smolensky [Smo87b; Smo93] for degree-d $\mathbb{F}_2$ polynomials against the Majority function, but they are quadratically worse (correlation bounded by $\leq \mathcal{O}(d/\sqrt{n})$).

by property (ii) in Lemma 67. Thus all $g \in \mathrm{supp}\mathcal{F}_2$ are at least $(1/2 - 2^{2n/3})$-far from $\mathrm{MM}_n \supseteq \mathrm{supp}\mathcal{F}_1$.

Suppose a testing algorithm using $R \leq 2^{0.1n}$ classical samples succeeds with probability δ. That implies that, given access to a function from $\mathcal{F}_1$ or from $\mathcal{F}_2$ with equal probability the algorithm can guess which of the two ensembles the function was drawn from with success probability δ.

Let $b \sim \{1,2\}$ and $f \sim \mathcal{F}_b$. With probability at least $1 - R^2/2^n$, all $x^{(r)}$'s and all $y^{(r)}$'s in the R-many samples are distinct (collision bound and union bound). Call this event D. Conditioned on D, for all $f \in \mathrm{supp}\mathcal{F}_1$ and all $f \in \mathrm{supp}\mathcal{F}_2$, the distribution of observed values $\left(f(x^{(1)}, y^{(1)}, \ldots, f(x^{(R)}, y^{(R)})\right)$ is uniformly random because $h \sim H$ is $2^{0.1n}$-wise independent (property (i) in Lemma 67) and we assumed $R \leq 2^{0.1n}$. Thus conditioned on D, the data observed is independent of b. Moreover, all functions in $\mathrm{supp}\mathcal{F}_1$ and $\mathrm{supp}\mathcal{F}_2$ are bent, so Fourier sampling provides no information whatsoever. The distinguishing probability is thus bounded by

$$\delta \leq \Pr[D] \cdot \frac{1}{2} + \Pr[D^c] \cdot 1 \leq \frac{1}{2} + \frac{R^2}{2^n} \leq \frac{1}{2} + 2^{-0.8n}. \qquad \square$$

Having established that the Maiorana–McFarland class is hard to test from classical samples and Fourier samples alone, we now give a very efficient passive quantum tester for MM_n. While this tester still uses quantum Fourier sampling at the end of the algorithm, it crucially preprocesses the function state in superposition before applying performing Fourier sampling.

Theorem 69. *There is an efficient quantum algorithm that uses $\mathcal{O}(1)$ copies of the function state $|f\rangle = \frac{1}{2^n}\sum_{x,y\in\{0,1\}^n} |x, y, f(x,y)\rangle$ to decide, with success probability $\geq 2/3$, whether f is in MM_n or $(1/3)$-far from MM_n.*

Proof. Let U denote the $(2n+1)$-qubit unitary acting as $|x, y, b\rangle \mapsto |x, y, b \oplus \langle x, y\rangle\rangle$. Note that U can be implemented by a quantum circuit with $\mathcal{O}(n)$ many 2-qubit gates and depth $\mathcal{O}(\log n)$. Moreover, notice that $U|f\rangle = |\tilde{f}\rangle$ for $\tilde{f}(x,y) := f(x,y) \oplus \langle x, y\rangle$.

The quantum algorithm works as follows. Recall that one function state copy suffices to obtain one Fourier sample with success probability $1/2$, using only $2n$ many single-qubit gates. Applying this Fourier sampling subroutine to $\mathcal{O}(1)$ many copies of $U|f\rangle$ thus suffices to obtain, with success probability $\geq 5/6$, $m \geq C$ many Fourier samples $S_1, \ldots, S_m \subseteq [2n]$ of the function $\tilde{f}$, where

$C > 0$ is a universal constant to be chosen later. Let $J = \{n+1, n+2, \ldots, n\}$ and compute

$$\hat{p} = \frac{1}{m}\left|\left\{1 \le k \le m : J \cap S_k \neq \emptyset\right\}\right|.$$

If $\hat{p} \le 1/9$, output "$f \in \mathrm{MM}_n$". Otherwise, output "f is $(1/3)$-far from MM_n."

First, let us show completeness of the protocol. So, suppose $f \in \mathrm{MM}_n$. Then there is a function $h : \mathbb{F}_2^n \to \mathbb{F}_2$ such that

$$U|f\rangle = \frac{1}{2^n}\sum_{x,y}|x, y, h(x)\rangle = |h\rangle\ ,$$

where we abused notation by using h to denote the function $h : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2$ defined as $h(x,y) = h(x)$. As $h(x,y)$ depends only on the first n variables, we have

$$p \coloneqq \sum_{S \subseteq [2n],\, J \cap S \neq \emptyset} |\hat{h}(S)|^2 = 0\,.$$

The constant C can be chosen such that, conditioned on the high probability event that we obtained at least C many Fourier samples, we have $|p - \hat{p}| \le 1/9$ with probability $\ge 5/6$ (by Chernoff-Hoeffding). So $\hat{p} \le 1/9$, and our tester correctly outputs "$f \in \mathrm{MM}_n$" with probability $\ge 2/3$.

Next, we analyze soundness. So, suppose f is $(1/3)$-far from MM_n. Equivalently, $\tilde{f}(x,y) = f(x,y) \oplus \langle x, y\rangle$ is $(1/3)$-far from any Boolean function h that depends only on the first n variables, $h(x,y) = h(x)$, where we again abused notation. Consider the function g defined as

$$g(x,y) = \sum_{S \subseteq [2n] : J \cap S = \emptyset} \hat{\tilde{f}}(S)\chi_S(x,y)\,.$$

Notice that $g(x,y)$ depends only on x, but g is in general not Boolean. Define $\tilde{g}(x,y) = \mathbf{1}_{g(x,y) \ge 1/2}$. Notice that $\tilde{g}$ is a Boolean function and that $\tilde{g}(x,y)$ depends only on x. Then (compare [AS07, Fact II.2]) we have

$$\tfrac{1}{3} \le \mathbb{P}_{x_1,x_2}[\tilde{f}(x_1,x_2) \neq \tilde{g}_2(x_2)] \le \mathbb{E}_{x_1,x_2}[(\tilde{f}(x_1,x_2) - g(x_1,x_2))^2] = \textstyle\sum_{S \subset [2n] : J \cap S \neq \emptyset}\left|\hat{\tilde{f}}(S)\right|^2 = p\,.$$

Again, conditioned on having produced at least C many Fourier samples, with probability $\ge 5/6$, we have $|p - \hat{p}| \le 1/9$ and thus $\hat{p} \ge 2/9$. So, our tester correctly outputs "f is $(1/3)$-far from MM_n" with probability $\ge 2/3$. □

9.3 Separating passive quantum from query-based classical property testing

In this section we give a property for which classical queries have exponential advantage over quantum testing from function states. This property is closely

related to the inability of quantum computers to measure the intersection of three subset states, where for a subset $S \subseteq \mathbb{F}_2^n$, the corresponding subset state is defined as $|S\rangle = \frac{1}{\sqrt{|S|}} \sum_{x \in S} |x\rangle$. We explain this connection at the end of the section.

The main result of this section is the following theorem.

Theorem 70. *There exist two sets of Boolean functions F_0, F_1 such that*

$$\min_{f_0 \in F_0, f_1 \in F_1} \|f_0 - f_1\|_1 \geq \frac{1}{64}$$

and such that:

- *Any passive quantum tester requires $\Omega(2^{n/2})$ copies of a function state to distinguish F_0 and F_1 with constant probability $2/3$.*
- *F_0 and F_1 may be distinguished with probability $2/3$ from $\mathcal{O}(1)$ classical queries.*

The families F_0 and F_1 arise from certain encodings of triples of subsets $A, B, C \subseteq \{0,1\}^n$. Consider the class of Boolean functions $f_{(A,B,C)} : \{0,1\}^{n+2} \to \{0,1\}$ on $n+2$ bits parameterized by subsets $A, B, C \subseteq \{0,1\}^n$ and defined as follows:

$$f_{(A,B,C)}(x,a) = \begin{cases} \mathbf{1}_{x \in A} & a = 00 \\ \mathbf{1}_{x \in B} & a = 01 \\ \mathbf{1}_{x \in C} & a = 10 \\ 0 & a = 11 \end{cases}.$$

With A, B, C drawn uniformly from subsets of $\{0,1\}^n$, we define two function state ensembles $\{|f_{(A,B,C)}\rangle\}_{A,B,C}$ and $\{|f_{(A,B,A\Delta B)}\rangle\}_{A,B}$, with their mixed state average over t-copy states given by

$$\mathcal{E}_0 = \mathop{\mathbb{E}}_{A,B,C}\left[\left|f_{(A,B,C)}\right\rangle\!\left\langle f_{(A,B,C)}\right|^{\otimes t}\right], \qquad \mathcal{E}_1 = \mathop{\mathbb{E}}_{A,B}\left[\left|f_{(A,B,A\Delta B)}\right\rangle\!\left\langle f_{(A,B,A\Delta B)}\right|^{\otimes t}\right].$$

We now show that these two mixed state averages over t-copy states are close in trace distance unless t scales exponentially in n. This means that exponentially-in-n many copies are needed to distinguish between the two function state ensembles.

Theorem 71. $\|\mathcal{E}_0 - \mathcal{E}_1\|_1 \leq \mathcal{O}(t/2^{n/2})$.

Proof. It will help to reinterpret $\left|f_{(A,B,C)}\right\rangle$ as a subset state via the rewriting

$$|x,a\rangle\,|f_{(A,B,C)}(x,a)\rangle = \sum_{b\in\{0,1\}} \mathbf{1}_{x\in S_{a,b}}\,|x,(a,b)\rangle\,, \tag{9.3.1}$$

where $S_{a,b}$ denotes A, B, C, or $\emptyset$ according to a when $b = 1$, or the respective complements if $b = 0$. Using r to represent the concatenation of a and b we may then write

$$|f_{(A,B,C)}\rangle = \frac{1}{\sqrt{4N}} \sum_{r\in\{0,1\}^3} \sum_{x\in\{0,1\}^n} \mathbf{1}_{x\in S_r}\,|x,r\rangle\,,$$

where $N := 2^n$ and, like before, S_r denotes one of A, B, C, or $\emptyset$ or the complements thereof.

With this notation let us consider the basis for the space of t copies of function states given by

$$\left\{\,|x_1,r_1,\ldots,x_t,r_t\rangle : x_j \in \{0,1\}^n, r_j \in \{0,1\}^3, j = 1,\ldots,t\right\}.$$

Let Π denote the projector onto the subspace spanned by those $|x_1,r_1,\ldots,x_t,r_t\rangle$ for which all x_j are distinct.

First, we claim

$$\|\mathcal{E}_0 - \Pi\mathcal{E}_0\Pi\|_1,\ \|\mathcal{E}_1 - \Pi\mathcal{E}_1\Pi\|_1 \ \le\ \mathcal{O}\left(\frac{t}{\sqrt{N}}\right). \tag{9.3.2}$$

These bounds follow from applying the triangle inequality to the following estimate: for any fixed A, B, C, we have

$$\left\|\left|f_{(A,B,C)}\right\rangle\!\left\langle f_{(A,B,C)}\right|^{\otimes t} - \Pi\left|f_{(A,B,C)}\right\rangle\!\left\langle f_{(A,B,C)}\right|^{\otimes t}\Pi\right\|_1 = \sqrt{\left(1+3\frac{4^tN^{\underline{t}}}{(4N)^t}\right)\left(1-\frac{4^tN^{\underline{t}}}{(4N)^t}\right)} \tag{9.3.3}$$

$$\le \frac{2t}{\sqrt{N}}\,, \tag{9.3.4}$$

where $(x)^{\underline{t}} = x(x-1)\ldots(x-t+1)$ denotes falling factorial, and where in the second step we applied the bound

$$\frac{4^tN^{\underline{t}}}{(4N)^t} \ge \left(1-\frac{t}{N}\right)^t \ge 1-\frac{t^2}{N}.$$

To see Equation (9.3.3), note that

$$M := \left|f_{(A,B,C)}\right\rangle\!\left\langle f_{(A,B,C)}\right|^{\otimes t} - \Pi\left|f_{(A,B,C)}\right\rangle\!\left\langle f_{(A,B,C)}\right|^{\otimes t}\Pi$$

has the following block form after reordering columns:

$$M = \frac{1}{(4N)^t} \underbrace{\left(\begin{array}{c|c|c} 0 & 1 & 0 \\ \hline 1 & 1 & 0 \\ \hline 0 & 0 & 0 \end{array}\right)}_{\substack{4^t N^{\underline{t}} \quad 4^t(N^t - N^{\underline{t}}) \quad (8N)^t - (4N)^t}} \begin{matrix} \} 4^t N^{\underline{t}} \\ \} 4^t(N^t - N^{\underline{t}}) \\ \} (8N)^t - (4N)^t \end{matrix} .$$

This is because $\left|f_{(A,B,C)}\right\rangle\!\left\langle f_{(A,B,C)}\right|^{\otimes t}$ is an all-zeros matrix except for the principal submatrix associated to indices $\big((x_1, r_1), \ldots, (x_t, r_t)\big)$ where $x_j \in S_{r_j}$ for all j, and here it is equal to $(4N)^{-t}$. There are $(4N)^t$ such entries. Moreover, $\Pi\left|f_{(A,B,C)}\right\rangle\!\left\langle f_{(A,B,C)}\right|^{\otimes t}\Pi$ is an all-zeros matrix except for the principal submatrix associated to indices $\big((x_1, r_1), \ldots, (x_t, r_t)\big)$ where $x_j \in S_{r_j}$ for all j and $x_j \neq x_k$ for $j \neq k$ and here it is also equal to $(4N)^{-t}$—and there are $4^t N^{\underline{t}}$ of these entries. M thus has rank 2 and its spectrum is easily determined, leading to the estimate in Equation (9.3.3).

Now we claim that in fact

$$\Pi \mathcal{E}_0 \Pi = \Pi \mathcal{E}_1 \Pi \,. \tag{9.3.5}$$

Let us consider a specific entry in $\Pi\mathcal{E}_0\Pi$ with row and column indices

$$\mathbf{r} = (\ldots, (x_j, r_j), \ldots), \qquad \mathbf{s} = (\ldots, (y_j, s_j), \ldots) \,.$$

It will be useful to write $\mathcal{S}_z = \mathcal{S}_z(\mathbf{r}, \mathbf{s})$ for the set types that appear with a particular string $z \in \{0,1\}$ in $\mathbf{r}$ and $\mathbf{s}$. That is, for any $z \in \{0,1\}^n$ define

$$\mathcal{S}_z = \mathcal{S}_z(\mathbf{r}, \mathbf{s}) = \Big\{ q \in \{0,1\}^3 : (z, q) \in \mathbf{r} \text{ or } (z, q) \in \mathbf{s} \Big\}.$$

Then

$$\begin{aligned} \langle \mathbf{r} | \, \Pi \mathcal{E}_0 \Pi \, | \mathbf{s} \rangle &= (4N)^{-t} \mathop{\mathbb{E}}_{A,B,C} \Big(\textstyle\prod_j \mathbf{1}_{x_j \in S_{r_j}} \Big) \Big(\textstyle\prod_j \mathbf{1}_{y_j \in S_{s_j}} \Big) \\ &= (4N)^{-t} \prod_{z \in \{x_j\}_j \cup \{y_j\}_j} \mathop{\mathbb{E}}_{A,B,C} \textstyle\prod_{q \in \mathcal{S}_z} \mathbf{1}_{z \in S_q} . \end{aligned}$$

It follows from the definition of Π that $|\mathcal{S}_z| \leq 2$ for any $z \in \{x_j\}_j \cup \{y_j\}_j$: there is at most one contribution to $\mathcal{S}_z$ from each of $\mathbf{r}$ and $\mathbf{s}$. As a result we have for any z that

$$\mathop{\mathbb{E}}_{A,B,C \overset{\text{iid}}{\sim} \mathcal{P}\{0,1\}^n} \textstyle\prod_{q \in \mathcal{S}_z} \mathbf{1}_{z \in S_q} = \mathop{\mathbb{E}}_{\substack{A,B \overset{\text{iid}}{\sim} \mathcal{P}\{0,1\}^n \\ C = A \Delta B}} \textstyle\prod_{q \in \mathcal{S}_z} \mathbf{1}_{z \in S_q} .$$

This follows from mild case analysis, the most important part of which is to note that for any $S \neq T \in \{A, B, C, A\Delta B\}$,

$$(\mathbf{1}_{x\in S}, \mathbf{1}_{x\in T}) \sim (b_1, b_2),$$

where b_1 and b_2 are i.i.d. Bernoulli $1/2$ random variables. So we see $\langle \mathbf{r}| \Pi\mathcal{E}_0\Pi |\mathbf{s}\rangle = \langle \mathbf{r}| \Pi\mathcal{E}_1\Pi |\mathbf{s}\rangle$ and Equation (9.3.5) is satisfied.

Combining the triangle inequality with Equation (9.3.2) and Equation (9.3.5) gives the result. □

Proof of Theorem 70. Consider

$$F_0 = \left\{ f_{(A,B,C)} : A, B, C \subseteq \{0,1\}^n, 2^{-n}|A \cap B \cap C| \geq 1/16 \right\}$$
$$\text{and} \quad F_1 = \left\{ f_{(A,B,A\Delta B)} : A, B \subseteq \{0,1\}^n \right\},$$

First we prove the minimum distance between F_0 and F_1. For any $f_0 \in F_0$, there are $2^n/16$ strings $x \in \{0,1\}^n$ such that $f_0(x00) = f_0(x01) = f_0(x10) = 1$. On the other hand, for all $f_1 \in F_1$, by definition there are no strings x with this property. Thus the minimum L_1 distance between F_0 and F_1 is at least

$$\frac{2^n/16}{4 \cdot 2^n} = \frac{1}{64}.$$

Now define the state ensembles

$$\mathcal{E}_0' = \mathop{\mathbb{E}}_{f\sim F_0} |f\rangle\langle f|^{\otimes t} \quad \text{and} \quad \mathcal{E}_1 = \mathop{\mathbb{E}}_{f\sim F_1} |f\rangle\langle f|^{\otimes t}.$$

$\mathcal{E}_1$ here is exactly $\mathcal{E}_1$ from Theorem 71. To compare $\mathcal{E}_0'$ and $\mathcal{E}_0$ from Theorem 71, note that for $A, B, C \overset{\text{iid}}{\sim} \mathcal{P}\{0,1\}^n$, any string x is in $A \cap B \cap C$ with probability $1/8$ and so from Chernoff we have

$$\Pr\left[|A \cap B \cap C| < \frac{1}{2} \cdot \frac{2^n}{8}\right] \leq \exp\left(-\frac{2^n}{64}\right).$$

This dramatic concentration, together with Theorem 71 implies

$$\|\mathcal{E}_0' - \mathcal{E}_1\|_1 \leq \|\mathcal{E}_0' - \mathcal{E}_0\|_1 + \|\mathcal{E}_1 - \mathcal{E}_0\|_1 \leq \mathcal{O}(t/2^{n/2}).$$

To test this property with classical queries, given an unknown $f = f_{(A,B,C)}$ one may simply choose a random $x \in \{0,1\}^n$ and check if $f(x00) = f(x01) = f(x10) = 1$. This test accepts with probability $1/8$ when $f \in F_0$ and accepts with probability 0 when $f \in F_1$. □

k-fold intersection is "unfeelable" for $k \geq 3$

In this subsection, we reinterpret Theorem 70 in the context of subset states. Given access to copies of k different subset states $|S_1\rangle, \ldots, |S_k\rangle$, it is natural to ask how many copies of each are required to estimate the fractional size of the mutual intersection,

$$\frac{|S_1 \cap \cdots \cap S_k|}{2^n}.$$

When $k = 2$, this can be readily accomplished using ideas similar to our algorithms presented above. In the case of intersection estimation with $k = 2$, we have the identity

$$\frac{|S_1 \cap S_2|}{2^n} = \langle S_{\text{all}}|S_1\rangle \langle S_1|S_2\rangle \langle S_2|S_{\text{all}}\rangle,$$

where $S_{\text{all}} := \{0,1\}^n$ denotes the full hypercube. The quantities on the right-hand side are easily estimated via swap tests, so it takes $\mathcal{O}(1)$ copies of $|S_1\rangle, |S_2\rangle$ to estimate the quantity of interest to any constant additive error.

In contrast, it is a consequence of Theorem 70 that the same question for $k = 3$ has a very different answer: it requires $\Omega(2^{n/2})$ copies to achieve constant additive error. To see this, note that from any $|f_{(A,B,C)}\rangle$ one may obtain each of $|A\rangle$, $|B\rangle$, and $|C\rangle$ with constant probability by measuring the a and $f(x,a)$ registers, provided that the minimum among $|A|$, $|B|$, and $|C|$ is at least a constant fraction of 2^n—and this condition is satisfied by the overwhelming majority of functions in the families F_0 and F_1 of Theorem 70. From F_0 and F_1 we obtain:

Corollary 72. *There are two families $\mathcal{S}_0$ and $\mathcal{S}_1$ of triples of subsets of $\{0,1\}^n$ such that*

$$\forall (A_0, B_0, C_0) \in \mathcal{S}_0, \quad |A_0 \cap B_0 \cap C_0|/2^n \geq 1/16$$

and $$\forall (A_1, B_1, C_1) \in \mathcal{S}_1, \quad |A_1 \cap B_1 \cap C_1|/2^n = 0,$$

and yet any quantum algorithm distinguishing the two families via their subset states requires $\Omega(2^{n/2})$ copies of $|A\rangle, |B\rangle,$ or $|C\rangle$.

9.4 A challenge: lower bounds for monotonicity testing

Here we show that the ensembles used in [Gol+00] to establish strong lower bounds on monotonicity testing from samples do not improve upon the basic $\Omega(1/\varepsilon)$ sample complexity lower bound in the quantum case. To prove this, we

consider the pair of distributions over functions from [Gol+00], constructed such that one is supported entirely on monotone functions, and the other with high probability on functions that are ε-far from monotone; we show that the associated t-copy quantum function state ensembles become distinguishable with constant success probability as soon as $t = \Omega(1/\varepsilon)$. At the end of the section, we discuss how to extend our reasoning to the ensembles used in [Bla24].

Distinguishability of twin ensembles

For the proof, it will be useful to also consider *phase states*, which are given by

$$|\Psi_f^{\mathrm{ph}}\rangle = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} (-1)^{f(x)} |x\rangle \ .$$

The proof reduces to the distinguishability of phase state ensembles encoding the following classical sets of functions taken from [Gol+00].

Definition 5 (Twin ensembles)**.** *Let $M = \{(u_i, v_i)\}_{i=1}^m$ be a set of pairs of elements in $\{0,1\}^n$ such that all $u_1, v_1, \ldots, u_m, v_m$ are distinct. Let $\cup M := \cup_{(u,v)\in M}\{u, v\}$ be the complete set of elements in the matching. Fix a function $g : \{0,1\}^n \backslash (\cup M) \to \{0,1\}$. We now define the* twin ensembles *associated to M and g, which are two sets F_0, F_1 of functions on $\{0,1\}^n$.*

For any bipartition of M, $M = A \sqcup B$, define the following two functions.

1. *$f_{A,B}^{(0)}$ is defined as follows:*
   - *For $(u, v) \in A$, we set $f_{A,B}^{(0)}(u) = f_{A,B}^{(0)}(v) = 1$.*
   - *For $(u, v) \in B$, we set $f_{A,B}^{(0)}(u) = f_{A,B}^{(0)}(v) = 0$.*
   - *If $x \notin \cup M$, then define $f_{A,B}^{(0)}(x) = g(x)$.*
2. *$f_{A,B}^{(1)}$ is defined as follows:*
   - *For $(u, v) \in A$, we set $f_{A,B}^{(1)}(u) = 1$ and $f_{A,B}^{(1)}(v) = 0$.*
   - *For $(u, v) \in B$, we set $f_{A,B}^{(1)}(u) = 0$ and $f_{A,B}^{(1)}(v) = 1$.*
   - *If $x \notin \cup M$, then define $f_{A,B}^{(1)}(x) = g(x)$.*

Then the twin ensembles associated to M and g are $F_0 = \{f_{A,B}^{(0)}\}_{A\sqcup B=M}$ and $F_1 = \{f_{A,B}^{(1)}\}_{A\sqcup B=M}$.

Let us recall the reasoning from [Gol+00] that connects these ensembles to monotonicity testing. Take $k = \lceil n/2 \rceil$ and consider the k^{th} and $(k-1)^{\text{th}}$ layer of the Boolean hypercube with respect to the standard partial ordering on strings $\preceq$. These layers we denote by L_k and L_{k-1} respectively; *i.e.*, $L_i = \{x \in \{0,1\}^{n-1} : |x| = i\}$, where $|x|$ denotes the Hamming weight of x. Stirling's formula gives that $|L_k|, |L_{k-1}| = \Omega(2^n/\sqrt{n})$. As argued in [Gol+00], we can find a matching $M = \{(u_i \prec v_i)\}_{i=1}^m \subset L_{k-1} \times L_k$ such that (i) there is no $i \neq j$ such that u_i and v_j are comparable, (ii) $|M|$ is even, and (iii) $m \coloneqq |M| = \varepsilon \cdot 2^n$, for any $\varepsilon = \varepsilon(n)$ with $0 < \varepsilon \leq \mathcal{O}(n^{-3/2})$. Now define

$$\begin{aligned} g : \{0,1\}^n \backslash \cup M &\to \{0,1\} \\ x &\mapsto \mathbf{1}_{|x| \geq n/2} \,. \end{aligned}$$

The choices of g and M define twin ensembles F_0 and F_1.

Clearly, every $f^{(0)}_{A,B}$ is a monotone function. Let $F_1^{\text{far}} \subset F_1$ be the set of $f^{(1)}_{A,B}$ functions for which $|B| \geq m/4$. Then all functions in F_1^{far} are at least $\Omega(\varepsilon)$-far from monotone [Gol+00]. We thus wish to bound the distinguishability of F_0 from F_1^{far}, which in the quantum case is determined by the 1-norm

$$\left\| \mathop{\mathbb{E}}_{f^{(0)}_{A,B} \sim F_0} \left(|f^{(0)}_{A,B}\rangle\langle f^{(0)}_{A,B}| \right)^{\otimes t} - \mathop{\mathbb{E}}_{f^{(1)}_{A,B} \sim F_1^{\text{far}}} \left(|f^{(1)}_{A,B}\rangle\langle f^{(1)}_{A,B}| \right)^{\otimes t} \right\|_1 .$$

From a standard concentration argument, it suffices to instead bound the distinguishability between F_0 and F_1, which in the quantum case is determined by

$$\left\| \mathop{\mathbb{E}}_{f^{(0)}_{A,B} \sim F_0} \left(|f^{(0)}_{A,B}\rangle\langle f^{(0)}_{A,B}| \right)^{\otimes t} - \mathop{\mathbb{E}}_{f^{(1)}_{A,B} \sim F_1} \left(|f^{(1)}_{A,B}\rangle\langle f^{(1)}_{A,B}| \right)^{\otimes t} \right\|_1 .$$

We will show that, in contrast to the classical case analyzed in [Gol+00], the twin ensembles actually become distinguishable already for $t \sim 1/\varepsilon$. Namely, much of Section 9.4 will be dedicated to proving the following theorem:

Theorem 73. *Define $\varepsilon > 0$ so that $\varepsilon 2^n = m = |M|$. Then*

$$\left\| \mathbb{E}_B \left[\left(|\Psi^{\text{ph}}_{f^{(0)}_{A,B}}\rangle\langle\Psi^{\text{ph}}_{f^{(0)}_{A,B}}| \right)^{\otimes t} \right] - \mathbb{E}_B \left[\left(|\Psi^{\text{ph}}_{f^{(1)}_{A,B}}\rangle\langle\Psi^{\text{ph}}_{f^{(1)}_{A,B}}| \right)^{\otimes t} \right] \right\|_1 \geq \Omega(1) \qquad (9.4.1)$$

for $t = \Omega(1/\varepsilon)$.

Let us note that Theorem 73 implies the same 1-norm lower bound and thus the same distinguishability of the two ensembles also from function state

copies. To see this, we notice that the function state for any Boolean function f defined on n bits is unitarily equivalent to the phase state for a related Boolean function on $n+1$ bits:

$$(I^{\otimes n} \otimes H) |f\rangle = \frac{1}{\sqrt{2^{n+1}}} \sum_{x \in \{0,1\}^n} |x\rangle |0\rangle + \frac{1}{\sqrt{2^{n+1}}} \sum_{x \in \{0,1\}^n} (-1)^{f(x)} |x\rangle |1\rangle =: |\Psi^{\text{ph}}_{\widetilde{f}}\rangle ,$$

where $\widetilde{f}(x_1, \ldots, x_n, x_{n+1}) = (\mathbf{1}_{x_{n+1}=1}) f(x_1, \ldots, x_n)$.

So, in fact the function states for functions in F_0 and F_1 are unitarily equivalent to phase states for another set of twin ensembles $\widetilde{F}^0$ and $\widetilde{F}^1$ with corresponding $\widetilde{M}$ obtained by appending 1 to every string in M and with $\widetilde{g}$ given by $\widetilde{g}(x) = (\mathbf{1}_{x_{n+1}=1}) \cdot g(x_1, \ldots, x_n)$. Theorem 73 implies these phase state ensembles are distinguishable with constant success probability for $t \geq \Omega(1)$, thus the same holds for the function state ensembles for F_0 and F_1.

Difference matrix: the entries

Here, to prove Theorem 73, we pursue a bound on the trace norm distance in Equation (9.4.1).

Define the density matrices

$$A^{(0)} = \mathbb{E}_B \left[\left(|\Psi^{\text{ph}}_{f^{(0)}_{A,B}}\rangle\langle\Psi^{\text{ph}}_{f^{(0)}_{A,B}}| \right)^{\otimes t} \right] \qquad \text{and} \qquad A^{(1)} = \mathbb{E}_B \left[\left(|\Psi^{\text{ph}}_{f^{(1)}_{A,B}}\rangle\langle\Psi^{\text{ph}}_{f^{(1)}_{A,B}}| \right)^{\otimes t} \right],$$

Call $A := A^{(0)} - A^{(1)}$ the *difference matrix*. We now characterize the entries of the difference matrix A by evaluating $A^{(0)}$ and $A^{(1)}$.

Rows (resp. columns) of $A^{(0)}$ and $A^{(1)}$ are indexed by t-tuples $\mathbf{x} = (x_1, \ldots, x_t)$ (resp. $\mathbf{y} = (y_1, \ldots, y_t)$) of strings $x_j \in \{0,1\}^n$ (resp. $y_j \in \{0,1\}^n$), $1 \leq j \leq t$. It turns out that the entries of $A^{(0)}$ and $A^{(1)}$ depend only on the multiset $\{x_1, \ldots, x_t, y_1, \ldots, y_t\}$, and for this we use the notation $\mathbf{x} \cup \mathbf{y}$. In the following it will sometimes be convenient to use $\uparrow$ for exponentiation: $a \uparrow b := a^b$.

Entries of the $A^{(0)}$, the $f^{(0)}$ matrix. The $(\mathbf{x}, \mathbf{y})$ entry of $A^{(0)}$ takes the form

$$A^{(0)}_{\mathbf{x},\mathbf{y}} = \mathop{\mathbb{E}}_{A \sqcup B = M} \left[\frac{1}{2^{nt}} \prod_{z \in \mathbf{x} \cup \mathbf{y}} (-1)^{f^{(0)}_{A,B}(z)} \right] = \frac{1}{2^{nt}} \mathop{\mathbb{E}}_{A \sqcup B = M} \left[(-1) \uparrow \sum_{z \in \mathbf{x} \cup \mathbf{y}} f^{(0)}_{A,B}(z) \right].$$

Evaluating this expectation gives

$$\begin{aligned} A^{(0)}_{\mathbf{x},\mathbf{y}} &= \frac{1}{2^{nt}} \left((-1)^{\uparrow \sum_{z \in \mathbf{x} \cup \mathbf{y}, z \notin \cup M} g(z)} \right) \prod_{p \in M} \mathop{\mathbb{E}}_{A,B} (-1)^{\uparrow \sum_{z \in \mathbf{x} \cup \mathbf{y} : z \in p} f^{(0)}_{A,B}(z)} \\ &= \frac{1}{2^{nt}} \left((-1)^{\uparrow \sum_{z \in \mathbf{x} \cup \mathbf{y}, z \notin \cup M} g(z)} \right) \\ &\qquad \cdot \prod_{p \in M} \mathop{\mathbb{E}}_{A,B} (-1)^{\uparrow \begin{cases} |\{z \in \mathbf{x} \cup \mathbf{y} : z \in p\}| \mod 2 & \text{if } p \in A \\ 0 & \text{otherwise} \end{cases}} . \end{aligned}$$

So, if $|\{z \in \mathbf{x} \cup \mathbf{y} : z \in p\}| = 0 \mod 2$, the expectation is always $(-1)^0$. On the other hand, if $|\{z \in \mathbf{x} \cup \mathbf{y} : z \in p\}| = 1 \mod 2$, the expression inside the expectation is $(-1)^1$ w.p. $1/2$ and $(-1)^0$ w.p. $1/2$. Define

$$s(\mathbf{x} \cup \mathbf{y}) = (-1)^{\uparrow \sum_{z \in \mathbf{x} \cup \mathbf{y}, z \notin \cup M} g(z)} . \tag{9.4.2}$$

Then

$$A^{(0)}_{\mathbf{x},\mathbf{y}} = \begin{cases} \frac{1}{2^{nt}} s(\mathbf{x} \cup \mathbf{y}) & \text{if } \forall p \in M, |\{z \in \mathbf{x} \cup \mathbf{y} : z \in p\}| = 0 \mod 2 \\ 0 & \text{otherwise} . \end{cases}$$

Entries of $A^{(1)}$, the $f^{(1)}$ matrix. Similarly to above we have

$$A^{(1)}_{\mathbf{x},\mathbf{y}} = \frac{1}{2^{nt}} s(\mathbf{x} \cup \mathbf{y}) \prod_{p \in M} \mathop{\mathbb{E}}_{A,B} (-1)^{\uparrow \sum_{z \in \mathbf{x} \cup \mathbf{y} : x \in p} f^{(1)}_{A,B}(z)} .$$

To evaluate the expectation, note there are four cases for each $p = (u, v) \in M$, depending on how many times u occurs in $\mathbf{x} \cup \mathbf{y}$, and how many times v occurs in $\mathbf{x} \cup \mathbf{y}$. Denote these quantities mod 2 as $L_p = L_p(\mathbf{x} \cup \mathbf{y})$ and $U_p = U_p(\mathbf{x} \cup \mathbf{y})$ respectively.

- If $L_p = 0$ and $U_p = 0$, the sum in the exponent is always 0, yielding $(-1)^0 = 1$ w.p. 1.
- If $L_p = 0$ and $U_p = 1$, the sum is either 0 or 1, each w.p. $1/2$, yielding in expectation $(-1)^1/2 + (-1)^0/2 = 0$.
- If $L_p = 1$ and $U_p = 0$, we similarly get 0 for the expectation.
- If $L_p = 1, U_p = 1$, then the sum is always 1, yielding $(-1)^1 = -1$ with probability 1.

In summary,

$$A^{(1)}_{\mathbf{x},\mathbf{y}} = \begin{cases} \frac{1}{2^{nt}}(-1)^{|\{p\in M: L_p = U_p = 1\}|} s(\mathbf{x}\cup\mathbf{y}) & \text{if } \forall p \in M, |\{z \in \mathbf{x}\cup\mathbf{y} : z \in p\}| = 0 \mod 2 \\ 0 & \text{otherwise} \end{cases},$$

which is equivalent to

$$A^{(1)}_{\mathbf{x},\mathbf{y}} = \begin{cases} \frac{1}{2^{nt}}(-1)^{|\{p\in M: L_p = U_p = 1\}|} s(\mathbf{x}\cup\mathbf{y}) & \text{if } \forall p \in M, L_p = U_p \\ 0 & \text{otherwise} \end{cases}.$$

Putting these together, we find

$$A_{\mathbf{x},\mathbf{y}} = A^{(0)}_{\mathbf{x},\mathbf{y}} - A^{(1)}_{\mathbf{x},\mathbf{y}} = \begin{cases} \frac{2}{2^{nt}} s(\mathbf{x}\cup\mathbf{y}) & \text{if } \forall p \in M, L_p = U_p \\ & \quad \text{and } |\{p \in M : L_p = U_p = 1\}| = 1 \mod 2 \\ 0 & \text{otherwise} \end{cases}. \tag{9.4.3}$$

Difference matrix: the spectrum

Here we conduct a fine-grained analysis of the spectrum of the difference matrix A in order to obtain a combinatorial bound on $\|A\|_1$. In Section 9.4 we then understand the asymptotics on this bound in terms of ε, t, and n.

Let D be a diagonal matrix with entries $s(\mathbf{x})$ for all t-tuples $\mathbf{x}$. Then $\tilde{A} := 2^{nt-1} D^{-1} A D$ is similar to $2^{nt-1} A$ and moreover

$$\begin{aligned} \tilde{A}_{\mathbf{x},\mathbf{y}} &= \begin{cases} s(\mathbf{x}) s(\mathbf{x}\cup\mathbf{y}) s(\mathbf{y}) & \text{if } \forall p \in M, L_p = U_p \text{ and } |\{p \in M : L_p = U_p = 1\}| = 1 \mod 2 \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} 1 & \text{if } \forall p \in M, L_p = U_p \text{ and } |\{p \in M : L_p = U_p = 1\}| = 1 \mod 2 \\ 0 & \text{otherwise} \end{cases}. \end{aligned} \tag{9.4.4}$$

Here we used that $s(\mathbf{x}\cup\mathbf{y}) = s(\mathbf{x})s(\mathbf{y})$; the other notation (for L_p and U_p) is as above.

Because of the matrix similarity, we know that A and $\tilde{A}$ have the same spectrum up to a scaling factor of $1/2^{nt-1}$. It will turn out that after a certain permutation of indices, $\tilde{A}$ is block-diagonal, with each block corresponding

to the adjacency matrix of a complete bipartite graph. Towards defining this block structure, we write $\mathbf{x} C \mathbf{y}$ (read "$\mathbf{x}$ is compatible with $\mathbf{y}$") if for all $p \in M, L_p = U_p$ and $|\{p \in M : L_p = U_p = 1\}| = 1 \mod 2$.

Given an index tuple $\mathbf{x}$, we define some technical quantities of $\mathbf{x}$ that are important for combinatorics to follow. These do not depend on the order of elements in $\mathbf{x}$ so we will treat $\mathbf{x}$ as a multiset for this discussion. The multiset $\mathbf{x}$ may be partitioned as:

$$\mathbf{x} = \{ \text{ pairs } \} \sqcup \{ \text{ singletons } \} \sqcup \{ \text{ elements outside of } \cup M \} \tag{9.4.5}$$

To form the "pairs" multiset, we greedily take as many copies of each pair $p \in M$ as we can from $\mathbf{x}$. The "singletons" multiset are the remaining elements in $\mathbf{x}$ from $\cup M$ that cannot be paired up, and the final part corresponds to those elements in $\mathbf{x}$ outside of $\cup M$. This partitioning is unique.

For example, suppose $M = \{(1,2),(3,4),(5,6)\}$, where we identify natural numbers with their n-bit binary expansions. Then, the following multiset has the partition

$$\{0,1,1,1,1,2,2,3,3,3,3,4\} \longmapsto \{(1,2),(1,2),(3,4)\} \sqcup \{1,1,3,3,3\} \sqcup \{0\}\,.$$

Now define $E(\mathbf{x})$ to be the number of pairs (with multiplicity) in $\mathbf{x}$ mod 2, *i.e.,* the cardinality mod 2 of the first part of the partition in Equation (9.4.5). Also, define the *singleton set* of $\mathbf{x}$ as

$$\text{sing}(\mathbf{x}) := \{e \in \text{ "singletons" } : e \text{ occurs an odd number of times in "singletons" } \}\,.$$

So, continuing our example, $E(\mathbf{x}) = 1$ and $\text{sing}(\mathbf{x}) = \{3\}$. We also define the *type* of $\mathbf{x}$, $\text{type}(\mathbf{x})$, to be the pairs with nonzero overlap with its singleton set:

$$\text{type}(\mathbf{x}) = \{p \in M : p \cap \text{sing}(\mathbf{x}) \neq \emptyset\}.$$

Continuing our example, we have that $\text{type}\Big((3,5,5,5)\Big) = \{(3,4),(5,6)\}$. We will also need a quantity on pairs of tuples $\mathbf{x}, \mathbf{y}$ counting the number of elements in their singleton sets that are paired up, mod 2:

$$P(\mathbf{x},\mathbf{y}) = |\{p \in M : p \subseteq \text{sing}(\mathbf{x}) \cup \text{sing}(\mathbf{y})\}| \mod 2\,.$$

So for example, keeping M as before, $P\Big((1,3),(1,4)\Big) = 1$.

Lemma 74. $\mathbf{x} C \mathbf{y}$ *if and only if*

$$\text{type}(\mathbf{x}) = \text{type}(\mathbf{y}) \qquad \textit{and} \qquad E(\mathbf{x}) + E(\mathbf{y}) + P(\mathbf{x},\mathbf{y}) = 1 \mod 2\,.$$

Proof. ($\implies$) Suppose by way of contradiction that $\mathbf{x} C \mathbf{y}$ but $\text{type}(\mathbf{x}) \neq \text{type}(\mathbf{y})$. Then there exists a pair $p = (a, b)$ such that (WLOG) a occurs in the "singletons" partition of $\mathbf{x}$ an even number of times, and a or b occurs in the "singletons" partition of $\mathbf{y}$ an odd number of times. This implies $L_p + U_p = 1 \mod 2$.

So now suppose $\text{type}(\mathbf{x}) = \text{type}(\mathbf{y})$. Because $\mathbf{x} C \mathbf{y}$, we have $|\{p \in M : L_p = U_p = 1\}| = 1 \mod 2$. But $|\{p \in M : L_p = U_p = 1\}| \mod 2$ is precisely $E(\mathbf{x}) + E(\mathbf{y}) + P(\mathbf{x}, \mathbf{y}) \mod 2$, because both quantities count the total number of pairs (with multiplicity) occurring in $\mathbf{x} \cup \mathbf{y}$.

($\impliedby$) If $\text{type}(\mathbf{x}) = \text{type}(\mathbf{y})$, we must have that $L_p + U_p = 0 \mod 2$ for all p. The elements in the "pairs" partition do not affect this condition. For all pairs $p = (a, b)$, either both a, b occur an even number of times in the "singletons" partition of $\mathbf{x}$ and $\mathbf{y}$. Or, if (WLOG) a occurs an odd number of times in $\mathbf{x}$, then a or b occurs an odd number of times in $\mathbf{y}$, preserving the condition.

Furthermore, as before, $|\{p \in M : L_p = U_p = 1\}| \mod 2$ counts the same quantity as $E(\mathbf{x}) + E(\mathbf{y}) + P(\mathbf{x}, \mathbf{y}) \mod 2$. Therefore, if $\text{type}(\mathbf{x}) = \text{type}(\mathbf{y})$ and $E(\mathbf{x}) + E(\mathbf{y}) + P(\mathbf{x}, \mathbf{y}) = 1 \mod 2$, we satisfy the criteria for compatibility. □

To understand the spectrum of the difference matrix we will make repeated use of the following structural fact about compatibility.

Lemma 75. *Let $\mathbf{x}, \mathbf{x}', \mathbf{x}''$ be such that $\mathbf{x} C \mathbf{x}''$ and $\mathbf{x}' C \mathbf{x}''$. Then for all $\mathbf{y}$, $\mathbf{x} C \mathbf{y}$ if and only if $\mathbf{x}' C \mathbf{y}$.*

Proof. By symmetry we need only argue the forward direction. Clearly $\text{type}(\mathbf{x}) = \text{type}(\mathbf{x}') = \text{type}(\mathbf{y})$. Note that for any $\mathbf{z}, \mathbf{z}', \mathbf{z}''$, we have $P(\mathbf{z}, \mathbf{z}') + P(\mathbf{z}', \mathbf{z}'') = P(\mathbf{z}, \mathbf{z}'') \pmod 2$, and because $\mathbf{x} C \mathbf{x}''$ and $\mathbf{x}' C \mathbf{x}''$, Lemma 74 implies

$$E(\mathbf{x}) + P(\mathbf{x}, \mathbf{x}'') = E(\mathbf{x}') + P(\mathbf{x}', \mathbf{x}'') \pmod 2.$$

Then we have the following equivalences modulo 2:

$$\begin{aligned} 1 &\equiv E(\mathbf{x}) + P(\mathbf{x}, \mathbf{y}) + E(\mathbf{y}) && \text{(Lemma 74)} \\ &\equiv E(\mathbf{x}) + P(\mathbf{x}, \mathbf{x}') + P(\mathbf{x}'', \mathbf{y}) + E(\mathbf{y}) \\ &\equiv E(\mathbf{x}') + P(\mathbf{x}', \mathbf{x}'') + P(\mathbf{x}'', \mathbf{y}) + E(\mathbf{y}) \\ &\equiv E(\mathbf{x}') + P(\mathbf{x}', \mathbf{y}) + E(\mathbf{y}) \,. \end{aligned}$$

Appealing to Lemma 74, we conclude $\mathbf{x}' C \mathbf{y}$. □

This implies, for example, that the $\mathbf{x}^{\text{th}}$ and $\mathbf{x}'^{\text{th}}$ rows in $\tilde{A}$ are equal.

We can view $\tilde{A}$ as describing a graph $G_{\tilde{A}}$ with vertices $(\{0,1\}^n)^t$ and edge set

$$\{(\mathbf{x}, \mathbf{x}') \in V \times V \mid \mathbf{x} C \mathbf{x}'\}\,.$$

With Lemmas 74 and 75 in hand, we are prepared to describe the structure of $G_{\tilde{A}}$. It will be useful to define certain combinatorial quantities first.

Definition 6 (Combinatorial quantities)**.** *Let $\Sigma = \Sigma_{odd} \sqcup \Sigma_{even} \sqcup \Sigma_{rest}$ be an alphabet of cardinality $|\Sigma| = 2^n$ partitioned such that Σ_{odd} consists of p pairs, so $|\Sigma_{odd}| = 2p$, and Σ_{even} consists of $m - p$ pairs, so $|\Sigma_{even}| = 2(m-p)$. Define $T(t,p)$ as the number of strings of length t over Σ such that symbols from Σ_{odd} each occur an odd number of times, symbols from Σ_{even} each occur an even number of times, and symbols from Σ_{rest} occur any number of times. Then define*

$$x_1(t) = \sum_{\substack{p=0 \\ p \text{ odd}}}^{\min\{m,t\}} \binom{m}{p} T(t,p) \qquad \text{and} \qquad x_2(t) = \sum_{\substack{p=0 \\ p \text{ even}}}^{\min\{m,t\}} \binom{m}{p} T(t,p)\,. \tag{9.4.6}$$

Further, define $N(t,p)$ as the number of strings of length t over Σ such that for each pair in Σ_{odd}, one of the two symbols occurs an odd number of times while the other occurs an even number of times, and for each pair in Σ_{even}, either both symbols occur an odd number of times or both symbols occur an even number of times.

Lemma 76 (Structure of $G_{\tilde{A}}$)**.** *The graph $G_{\tilde{A}}$ has exactly $\sum_{k=0}^{\min(t,m)} \binom{m}{k}$ connected components, each associated with a specific* type$(\cdot)$ *of vertex.*

The connected component of $G_{\tilde{A}}$ corresponding to the unique type *of cardinality 0 (the* empty type*) is a complete bipartite graph (U, V, E) with parts U, V such that*

$$|U| = x_1(t) \quad \text{and} \quad |V| = x_2(t)\,.$$

For $k \geq 1$, there are $\binom{m}{k}$ connected components, each corresponding to a type *of cardinality k. All such components are complete bipartite graphs (U, V, E) with parts U, V such that*

$$|U|, |V| = N(t,k)/2\,.$$

Proof. First, by Lemma 74, only vertices of the same type can be compatible, and there are a total of $\sum_{k=0}^{\min(t,m)} \binom{m}{k}$ types. For the remainder of the proof, we only need to consider vertices of the same type.

We will analyze the case with $k \geq 1$ first. Here k refers to the size of the type of the vertex. Consider a tuple $\mathbf{x}$ such that $\text{sing}(\mathbf{x}) = \{a_1, a_2, \ldots, a_{k-1}, a_k\}$ and $E(\mathbf{x}) = 0$. Consider a second tuple $\mathbf{x}'$ such that $\text{sing}(\mathbf{x}') = \{a_1, a_2, \ldots, a_{k-1}, b_k\}$, $E(\mathbf{x}') = 0$, and (a_k, b_k) are a pair in M. Two such tuples must exist since $k > 0$. Note that $\mathbf{x}$ is compatible with $\mathbf{x}'$. Furthermore, all elements of the same type as $\mathbf{x}$ and $\mathbf{x}'$ must be compatible with exactly one of $\mathbf{x}$ or $\mathbf{x}'$. To see that we form a complete bipartite graph, consider two elements $\mathbf{y}, \mathbf{y}'$ such that $\mathbf{x} C \mathbf{y}$ and $\mathbf{x}' C \mathbf{y}'$. By Lemma 75, we can conclude that $\mathbf{y} C \mathbf{y}'$. Since $\mathbf{y}, \mathbf{y}'$ were arbitrary, we have a complete bipartite graph.

To get the size of each of the two sets in the bipartition, fix a type of size $k \geq 1$ and take $(a_1, b_1), \ldots, (a_k, b_k)$ to be the pairs representing the type. By our definition of type, we know: For each i, the elements a_i and b_i appear with differing parities; for each of the remaining pairs in M, the two elements of the pair occur with the same parity; and each remaining element, which does not belong to any pair in M, can occur with an arbitrary parity. Thus, the number of vertices with type k is exactly $N(t, k)$. It remains to observe that the two components in the bipartition for type k are of equal size. To see this, take any tuple $\mathbf{x}$ of type k and w.l.o.g. suppose that a_1 occurs with odd parity in $\text{sing}(\mathbf{x})$. (Otherwise, b_1 occurs with odd parity in $\text{sing}(\mathbf{x})$ and the remaining argument is easily modified.) If we construct a tuple $\mathbf{x}'$ from $\mathbf{x}$ by replacing as many occurrences of a_1 as are in $\text{sing}(\mathbf{x})$ with b_1 (while keeping the remaining elements the same), then $\mathbf{x} C \mathbf{x}'$ by Lemma 74. This provides us with a one-to-one mapping between the two components in the bipartition, thus they are of equal size.

For $k = 0$, we instead consider two distinct vertices $\mathbf{x}$ and $\mathbf{x}'$ with $E(\mathbf{x}) = 0$ and $E(\mathbf{x}') = 1$ and $|\text{sing}(\mathbf{x})| = |\text{sing}(\mathbf{x}')| = 0$. Note that $\mathbf{x}$ and $\mathbf{x}'$ are compatible. By an analogous argument to the case of $k > 0$, we must have all vertices connected to either $\mathbf{x}$ and $\mathbf{x}'$, and we get a complete bipartite graph.

For the sizes of the sets in the bipartition, note that the number of vertices connected to $\mathbf{x}$ with $E(\mathbf{x}) = 0$ will be precisely $x_1(t)$ and the number of vertices connected to $\mathbf{x}'$ with $E(\mathbf{x}') = 1$ will be precisely $x_2(t)$, completing our proof. $\square$

As a consequence of this lemma, when suitably ordering the indexing tuples, the matrix $\tilde{A}$ has a block-diagonal structure with blocks corresponding to the connected components of $G_{\tilde{A}}$. Thus, to determine the spectrum of $\tilde{A}$, it suffices to determine the spectrum of each block.

The adjacency matrix of a complete bipartite graph between a- and b-many vertices has two nonzero eigenvalues, each of magnitude $\sqrt{ab}$ [Bol98, Chapter VIII.2]. Instantiating this fact in the context of Lemma 76 and renormalizing by $1/2^{nt-1}$ (recall $A = (D\tilde{A}D^{-1})/2^{nt-1}$), we obtain the following bound on $\|A\|_1$.

Corollary 77. *The 1-norm of the matrix A satisfies:*

$$\|A\|_1 = \frac{2}{2^{nt-1}}\sqrt{x_1(t)\cdot x_2(t)} + \frac{2}{2^{nt-1}}\sum_{k=1}^{\min\{t,m\}}\binom{m}{k}\frac{N(t,k)}{2}.$$

Difference matrix: the trace norm

Here we analyze the growth of $\|A\|_1$ in terms of ε, t, and n. We begin by using exponential generating functions (or EGFs—see [Wil94] for background) to derive explicit expressions for $T(t,p)$, $N(t,p)$, and $x_1(t)+x_2(t)$.

Lemma 78. *Let $T(t,p)$ be as in Definition 6. Then*

$$T(t,p) = \left(\frac{1}{2}\right)^{2m}\sum_{j=0}^{2m-2p}\sum_{k=0}^{2p}(-1)^k\binom{2m-2p}{j}\binom{2p}{k}(2^n-2j-2k)^t.$$

Proof. For readability let us use a for the number of symbols occurring an even number of times, b for the number of symbols occurring an odd number of times and c for the rest. Then $T(t,p)$ has EGF

$$f(x) = \left(\frac{e^x+e^{-x}}{2}\right)^a\left(\frac{e^x-e^{-x}}{2}\right)^b\left(e^x\right)^c.$$

Rearranging, we find

$$f(x) = \left(\frac{1}{2}\right)^{a+b}\sum_{j=0}^{a}\sum_{k=0}^{b}\binom{a}{j}\binom{b}{k}(-1)^k e^{(a-2j+b-2k+c)x}.$$

And we assume $c > a+b$, so we use that the EGF $e^{\eta x}$ corresponds to the sequence $\{\eta^t\}_{t=0}^{\infty}$, read off the relevant coefficient to derive the formula for $T(t,p)$:

$$T(t,p) = \left(\frac{1}{2}\right)^{a+b}\sum_{j=0}^{a}\sum_{k=0}^{b}\binom{a}{j}\binom{b}{k}(-1)^k(a-2j+b-2k+c)^t.$$

Substituting for a, b, c yields the result. □

Lemma 79. *With x_1 and x_2 defined as in Equation* (9.4.6)*,*

$$x_1(t)+x_2(t) = \left(\frac{1}{2}\right)^m\sum_{k=0}^{m}\binom{m}{k}(2^n-4k)^t.$$

Proof. The combinatorial interpretation of $x_1(t)+x_2(t)$ is the number of strings of length t over Σ where symbols from $A \sqcup B, |A \sqcup B| = 2m$ are paired up and within each pair, they must appear with the same parity.

The EGF for strings with 2 elements appearing with same parity is

$$\left(\frac{e^x + e^{-x}}{2}\right)^2 + \left(\frac{e^x - e^{-x}}{2}\right)^2 = \frac{e^{2x} + e^{-2x}}{2}.$$

To construct the desired strings, we combine m copies of this with $2^n - m$ copies of the unrestricted EGF, leading to the overall EGF

$$2^{-m}\left(e^{2x} + e^{-2x}\right)^m e^{(2^n - 2m)x}.$$

The result follows from simplifying this EGF and recognizing the related counting formula. □

Lemma 80. *Let $N(t,p)$ be as in Definition 6. Then*

$$N(t,p) = \left(\frac{1}{2}\right)^m \sum_{j=0}^{p} \sum_{k=0}^{m-p} (-1)^j \binom{p}{j} \binom{m-p}{k} (2^n - 4j - 4k)^t.$$

Proof. The proof uses EGFs analogously to how we derived the expression for $T(t,p)$. The EGF for $N(t,p)$ is

$$f(x) = \left(\frac{e^{2x} - e^{-2x}}{2}\right)^p \left(\frac{e^{2x} + e^{-2x}}{2}\right)^{m-p} \left(e^x\right)^{2^n - 2m}.$$

We can now rearrange this product of sums into a sum of products and, using again that the EGF $e^{\eta x}$ corresponds to the sequence $\{\eta^t\}_{t=0}^{\infty}$, read off the expression for $N(t,p)$. □

Lemma 81. *With x_1 and x_2 defined as in Equation* (9.4.6)*,*

$$\frac{1}{2^{nt}} \sum_{k=1}^{\min\{t,m\}} \binom{m}{k} \frac{N(t,k)}{2} = \frac{1}{2}\left(1 - \frac{x_1(t) + x_2(t)}{2^{nt}}\right).$$

Proof. First, observe that $\sum_{k=0}^{\min\{t,m\}} \binom{m}{k} N(t,k) = 2^{nt}$, since every one of the overall 2^{nt} tuples belongs to some type and we are summing over all sizes of types. Consequently,

$$\begin{aligned}
\frac{1}{2^{nt}} \sum_{k=1}^{\min\{t,m\}} \binom{m}{k} \frac{N(t,k)}{2} &= \frac{1}{2} - \frac{1}{2}\binom{m}{0}\frac{N(t,0)}{2^{nt}} \\
&= \frac{1}{2} - \frac{1}{2^{m+1}} \sum_{k=0}^{m} \binom{m}{k} \left(1 - \frac{4k}{2^n}\right)^t \\
&= \frac{1}{2} - \frac{1}{2}\frac{x_1(t) + x_2(t)}{2^{nt}},
\end{aligned}$$

where the second-to-last step used Lemma 80 and the last step used Lemma 79. □

We now further study the trace norm of A. We will need the following technical fact.

Proposition 3. *Suppose $m = |M| = \varepsilon 2^n$ and $\varepsilon = \varepsilon(n) < 1$. Let $D \in \mathbb{R}$ be a fixed constant. Then, for any $t = t(n) \leq \mathcal{O}(2^{bn})$ with $0 \leq b < \frac{1}{4}$, we have*

$$\mathop{\mathbb{E}}_{K \sim \mathcal{B}(m,1/2)} \left(1 - \frac{DK}{2^n}\right)^t = \left(1 - \frac{D\frac{m}{2}}{2^n}\right)^t + o(1)\,.$$

In Proposition 3 and what follows, $\mathcal{B}(\ell, 1/2)$ denotes the Binomial distribution with ℓ trials and success probability $1/2$.

Proof. Notice that, for any $t \in \mathbb{N}$, the function $f : [0, B) \to \mathbb{R}$ given by $f(x) = (1-x)^t$ is Lipschitz with Lipschitz constant $t \max_{0 \leq x \leq B} |1-x|^{t-1}$. Using that $|1 - Dk/2^n| \leq 1$ holds for all $0 \leq k \leq m$ for sufficiently large n, this implies

$$\mathop{\mathbb{E}}_{K \sim \mathcal{B}(m,1/2)} \left| \left(1 - \frac{DK}{2^n}\right)^t - \left(1 - \frac{D\frac{m}{2}}{2^n}\right)^t \right| \leq \mathop{\mathbb{E}}_{K \sim \mathcal{B}(m,1/2)} \left| t\left(\frac{DK}{2^n} - \frac{D\frac{m}{2}}{2^n}\right) \right|.$$

From Chernoff we have that

$$\Pr\left[\left|\frac{DK}{2^n} - \frac{D\frac{m}{2}}{2^n}\right| \geq \eta\right] \leq 2\exp\left(-\mathsf{const} \cdot 2^n \cdot \frac{\eta^2}{\varepsilon}\right).$$

Call the low-probability event above E. Then

$$\begin{aligned}\mathop{\mathbb{E}}_{K \sim \mathcal{B}(m,1/2)} \left| t\left(\frac{DK}{2^n} - \frac{D\frac{m}{2}}{2^n}\right) \right| &\leq t \Pr[E^c]\eta + t \Pr[E] \frac{D\frac{m}{2}}{2^n} \\ &\leq t\eta + 2tD\varepsilon \exp\left(-\mathsf{const} \cdot 2^n \cdot \frac{\eta^2}{\varepsilon}\right).\end{aligned}$$

We can set $\eta = \min\{1/n, 1/t^2\}$, then, because of our assumption on $t = t(n)$, both summands go to 0 as $n \to \infty$, finishing the proof. □

Theorem 82. *$\|A\|_1 = \Omega(1)$ for $t = \Omega\big(\varepsilon(n)^{-1}\big)$, assuming $\varepsilon = \varepsilon(n) \geq \Omega(2^{-bn})$ with $0 \leq b < \frac{1}{4}$.*

Proof. Let us label the expression for $\|A\|_1$ from Corollary 77 as

$$\|A\|_1 = \frac{2}{2^{nt-1}} \sqrt{x_1(t) \cdot x_2(t)} + \underbrace{\frac{2}{2^{nt-1}} \sum_{k=1}^{\min\{t,m\}} \binom{m}{k} \frac{N(t,k)}{2}}_{(*)}\,. \tag{9.4.7}$$

We will lower-bound $\|A\|_1$ by lower-bounding the second summand here, $(*)$. Lemma 79 implies

$$x_1(t) + x_2(t) = \left(\frac{1}{2}\right)^m \sum_{k=0}^{m} \binom{m}{k} (2^n - 4k)^t = \mathop{\mathbb{E}}_{K \sim \mathcal{B}(m,1/2)} (2^n - 4K)^t, \quad (9.4.8)$$

where $\mathcal{B}(m, 1/2)$ refers to the Binomial distribution. Returning to $(*)$ and using Lemma 81, we have

$$(*) = \frac{4}{2}\left(1 - \frac{x_1(t) + x_2(t)}{2^{nt}}\right) = 2\left(1 - \mathbb{E}\left(1 - \frac{4K}{2^n}\right)^t\right).$$

Using Proposition 3, we obtain that, as long as $t \leq \mathcal{O}(2^{bn})$ with $0 \leq b < \frac{1}{4}$,

$$(*) = 2\left(1 - \left(1 - \frac{2m}{2^n}\right)^t + o(1)\right) = 2\Big(1 - (1 - 2\varepsilon)^t + o(1)\Big).$$

Notice that our assumption on $\varepsilon = \varepsilon(n)$ ensures that $\frac{1}{\varepsilon(n)} \leq \mathcal{O}(2^{bn})$ with $0 \leq b < \frac{1}{4}$. Therefore, we can consider $t \geq \Omega(1/\varepsilon)$, we get $1 - (1 - 2\varepsilon)^t \geq \Omega(1)$, and therefore $\|A\|_1 \geq (*) \geq \Omega(1)$.

□

Remark 3. We can extend the above quantum distinguishability analysis to the ensembles from [Bla24]. The construction in [Bla24], based on *Talagrands random DNFs* [Tal96], establishes a lower bound of $\exp(\Omega(\sqrt{n}/\varepsilon))$ for passive classical monotonicity testing via a birthday paradox argument. The construction randomly selects DNF terms of fixed width to define a partial partition of the Boolean cube into disjoint sets U_j such that any two points in different U_j are incomparable. The difference between the monotone D_{yes} and non-monotone D_{no} case lies in the function value assignments: in D_{yes}, values within each disjoint set U_j are structured monotonically while in D_{no}, values with each U_j are randomly assigned. Classically, distinguishing these distributions requires $\exp(\Omega(\sqrt{2^n}/\varepsilon))$ samples, as a tester must sample at least two points from the same U_j to gain information. This leads to an exponential lower bound when parameters are chosen appropriately.

Following an argument structured similarly to the one above, one may see that the difference matrix between the induced function state ensembles in the quantum setting decomposes into blocks corresponding to complete multipartite graphs. To see this, in analogy to the analysis from Section 9.4, we can define a notion of compatibility between any two index tuples. Given a collection of sets U_j and an index tuple $\mathbf{x}$, we first remove duplicates from the tuple

and analyze its intersection pattern with each U_j. For instance, if $U_1 = \{1, 2\}$, $U_2 = \{3, 4\}$, and $U_3 = \{5, 6\}$, and our index tuple is $(1, 1, 1, 2, 3, 3, 3, 3, 5)$, then after removing duplicate, the corresponding subsets under the U_j sets are $[1, 2], [\,], [5]$. Two tuples are said to be compatible if, for every j, the number of elements from each U_j that appear in the tuple is even but not identical across the tuples after removing duplicates and decomposing. For example, the tuple $(1, 1, 1, 2, 3, 3, 3, 3, 5)$ is compatible with $(1, 1, 1, 2, 3, 3, 3, 3, 6)$ but not with $(1, 2, 3, 3, 3, 3, 3, 3, 5)$, as the latter shares the same decomposition as the original.

Importantly, this compatibility is transitive: if $\mathbf{x}$ is compatible with $\mathbf{x}'$ and if $\mathbf{x}'$ is compatible with $\mathbf{x}''$, then $\mathbf{x}$ is also compatible with $\mathbf{x}''$. This transitivity induces complete multipartite graph blocks with each block corresponding to a compatibility class. Therefore, the trace distance between the two ensembles equals the sum of the trace norms of multipartite graphs of various sizes. Bounds on the eigenvalues of such a graph in terms of the sizes of its parts can be found in, for example, [EH80; Meh23]. Also in analogy to our analysis of the [Gol+00] construction, we find the ensembles remain distinguishable when $t = \Omega(1/\varepsilon)$, and thus they achieve no improvement over the generic lower bound.

Acknowledgments

The authors thank Srinivasan Arunachalam, Fernando Jeronimo, Kyle Gulshen, Jiaqing Jiang, John Bostanci, Yeongwoo Hwang, Shivam Nadimpali, John Preskill, Akshar Ramkumar, Abdulrahman Sahmoud, Mehdi Soleimanifar, and Thomas Vidick for enlightening discussions. Moreover, we thank Nathan Harms for suggesting we look at monotonicity and general discussions on classical bounds for related problems. We thank Fermi Ma for helpful discussions that provided inspiration for Theorem 70. We are grateful to Francisco Escudero Gutiérrez for pointing us to the Fourier-analytic characterization of monotonicity which allowed us to improve a previous passive quantum monotonicity tester. Finally, we thank the anonymous STOC 2025 reviewers for helpful feedback. MCC was partially supported by a DAAD PRIME fellowship and by the BMBF (QPIC-1). JS is funded by Chris Umans' Simons Foundation Investigator Grant. Parts of contributions of this chapter were completed while JS was visiting the Simons Institute for the Theory of Computing, supported by DOE QSA grant #FP00010905.

9.5 Appendix: Useful facts

In this appendix we collect some simple lemmas that are used as subroutines in the main body. First, we make a simple observation about the possibility of post-selecting on a desired function value in a subset function state.

Lemma 83. *Let $m \in \mathbb{N}$, $S \subseteq \{0,1\}^n$, $f : \{0,1\}^n \to \{0,1\}$, $b \in \{0,1\}$, $\eta \in (0,1]$, and $\delta \in (0,1)$. Assume that $\Pr_{x\sim S}[f(x) = b] \geq \eta$. There is an efficient quantum algorithm that given $m\lceil \frac{\ln(m/\delta)}{\eta} \rceil$ many copies of the state $|\Psi_{S,f}\rangle = \frac{1}{\sqrt{|S|}} \sum_{x\in S} |x, f(x)\rangle$, outputs, with success probability $\geq 1-\delta$, at least m many copies of the state $|\Psi_{S,f,b}\rangle \propto \sum_{x\in S : f(x)=b} |x\rangle$. Moreover, if the algorithm fails, then the algorithm explicitly outputs* FAIL.

We note that via standard Chernoff-Hoeffding bounds, one can achieve the same guarantee as in Lemma 83 using $\max\{\lceil 2m/\eta \rceil, \lceil 2\ln(1/\delta)/\eta^2 \rceil\}$ many copies of the state $|\Psi_{S,f}\rangle$. This improves the m-dependence, but in general comes at the cost of a worse η-dependence.

Proof. By a union bound, it suffices to show that $\tilde{m} = \lceil \frac{\ln(1/\tilde{\delta})}{\eta} \rceil$ many copies of $|\Psi_{S,f}\rangle$ suffice to quantumly efficiently obtain one copy of $|\Psi_{S,f,b}\rangle$ with success probability $\geq 1-\tilde{\delta}$. So, let's assume $m = 1$. Also here the procedure is clear: For each copy of $|\Psi_{S,f}\rangle$, measure the last qubit in the computational basis. If the outcome is $b \in \{0,1\}$, then the post-measurement state after discarding the last qubit is $|\Psi_{S,f,b}\rangle$. If, after measuring on all the copies, outcome b has never been observed, output FAIL. Otherwise, output one copy one of the post-measurement states from rounds in which outcome b was observed.

Again, the analysis of the failure probability is simple:

$$\begin{aligned}\Pr[\text{FAIL}] &= \Pr[\text{outcome } b \text{ never occurs}] \\ &\leq (1-\eta)^{\tilde{m}} \\ &\leq \exp(-\eta\tilde{m}) \\ &\leq \tilde{\delta}\,.\end{aligned}$$

Here, we used the assumption $\Pr_{x\sim S}[f(x) = b] \geq \eta$ and our choice of $\tilde{m}$. □

We also require the following standard routine for estimating the overlap of two pure quantum states.

Lemma 84. *Let $\varepsilon, \delta \in (0,1)$. There is an efficient quantum algorithm that, given $\lceil \frac{2\ln(2/\delta)}{\varepsilon^4} \rceil$ many copies of each of two pure quantum states $|\psi\rangle$ and $|\phi\rangle$,*

outputs, with success probability $\geq 1-\delta$, an ε-accurate estimate (in $[0,1]$) of the (non-squared) overlap $|\langle\phi|\psi\rangle|$.

Proof. The procedure is as follows: Let $m = \lceil \frac{2\ln(2/\delta)}{\varepsilon^4} \rceil$ be the number of copies that are available for each of $|\psi\rangle$ and $|\phi\rangle$. For $1 \leq i \leq m$, perform a SWAP test between one copy of $|\psi\rangle$ and one copy of $|\phi\rangle$, let the outcome be $\hat{o}_i$. Define $\hat{o} = \frac{1}{m}\sum_{i=1}^m \hat{o}_i$ and output the estimate $\hat{\mu} = \sqrt{2(\hat{o} - \frac{1}{2})}$. Let us analyze the success probability of this procedure.

First, notice that each SWAP test accepts (*i.e.*, outputs 1) with probability $\frac{1+|\langle\phi|\psi\rangle|^2}{2}$ [BCWD01]. Thus, the $\hat{o}_i$ are i.i.d. Bernoulli($\frac{1+|\langle\phi|\psi\rangle|^2}{2}$) random variables. So, by a standard Chernoff-Hoeffding bound, we have $\left|\hat{o} - \frac{1+|\langle\phi|\psi\rangle|^2}{2}\right| \leq \varepsilon^2/2$ with probability $\geq 1 - 2\exp(-m\varepsilon^4/2) \geq 1-\delta$, by our choice of m. As $|\sqrt{x} - \sqrt{y}| \leq \sqrt{|x-y|}$ holds for all $x, y \geq 0$, this implies that also $|\hat{\mu} - |\langle\phi|\psi\rangle|| = \left|\sqrt{2(\hat{o} - \frac{1}{2})} - \sqrt{2(\frac{1+|\langle\phi|\psi\rangle|^2}{2} - \frac{1}{2})}\right| \leq \varepsilon$, with probability $\geq 1-\delta$, as desired. □

Finally, using the overlap estimation routine, one can start from a function state and two function subset states to estimate the probability of an input lying in both

Corollary 85. *Let $S, S' \subseteq \{0,1\}^n$, $f, f' : \{0,1\}^n \to \{0,1\}$, and $b, b' \in \{0,1\}$. Let $\varepsilon, \delta \in (0,1)$. There is an efficient quantum algorithm that, given $\lceil \frac{162\ln(6/\delta)}{\varepsilon^4} \rceil$ many copies of each of the states $|\Psi\rangle$, $|\Psi'\rangle$, and $2\lceil \frac{162\ln(6/\delta)}{\varepsilon^4} \rceil$ copies of each of the states $|\Psi_{S,f,b}\rangle$ and $|\Psi_{S',f',b'}\rangle$, outputs, with success probability $\geq 1-\delta$, an ε-accurate estimate of $\Pr_{x\sim\{0,1\}^n}[x \in S \cap S', f(x) = b, f'(x) = b']$.*

Proof. The procedure combines the ingredients developed above. To do so, notice the following equalities:

$$|\langle\Psi|\Psi_{S,f,b}\rangle| = \langle\Psi|\Psi_{S,f,b}\rangle = \sqrt{\frac{|S \cap f^{-1}(b)|}{2^n}},$$

$$|\langle\Psi'|\Psi_{S',f',b'}\rangle| = \langle\Psi'|\Psi_{S',f',b'}\rangle = \sqrt{\frac{|S' \cap f'^{-1}(b')|}{2^n}},$$

$$|\langle\Psi_{S,f,b}|\Psi_{S',f',b'}\rangle| = \langle\Psi_{S,f,b}|\Psi_{S',f',b'}\rangle = \sqrt{\frac{|S \cap f^{-1}(b) \cap S' \cap f'^{-1}(b')|^2}{|S \cap f^{-1}(b)| \cdot |S' \cap f'^{-1}(b')|}},$$

$$\frac{|S \cap f^{-1}(b) \cap S' \cap f'^{-1}(b')|}{2^n} = \sqrt{\frac{|S \cap f^{-1}(b) \cap S' \cap f'^{-1}(b')|^2}{|S \cap f^{-1}(b)| \cdot |S' \cap f'^{-1}(b')|} \cdot \frac{|S \cap f^{-1}(b)|}{2^n} \cdot \frac{|S' \cap f'^{-1}(b')|}{2^n}}.$$

So, we can estimate $\Pr_{x\sim\{0,1\}^n}[x \in S \cap S', f(x) = b, f'(x) = b'] = \frac{|S \cap f^{-1}(b) \cap S' \cap f'^{-1}(b')|}{2^n}$ as follows:

1. Via the procedure in Lemma 84, use $\lceil \frac{162 \ln(6/\delta)}{\varepsilon^4} \rceil$ many copies of each of $|\Psi\rangle$ and $|\Psi_{S,f,b}\rangle$ to output, with success probability $\geq 1 - \frac{\delta}{3}$, an $(\varepsilon/3)$-accurate estimate $\hat{\alpha}$ of $|\langle \Psi | \Psi_{S,f,b} \rangle|$.

2. Via the procedure in Lemma 84, use $\lceil \frac{162 \ln(6/\delta)}{\varepsilon^4} \rceil$ many copies of each of $|\Psi'\rangle$ and $|\Psi_{S',f',b'}\rangle$ to output, with success probability $\geq 1 - \frac{\delta}{3}$, an $(\varepsilon/3)$-accurate estimate $\hat{\alpha}'$ of $|\langle \Psi' | \Psi_{S',f',b'} \rangle|$.

3. Via the procedure in Lemma 84, use $\lceil \frac{162 \ln(6/\delta)}{\varepsilon^4} \rceil$ many copies of each of $|\Psi_{S,f,b}\rangle$ and $|\Psi_{S',f',b'}\rangle$ to output, with success probability $\geq 1 - \frac{\delta}{3}$, an $(\varepsilon/3)$-accurate estimate $\hat{\beta}$ of $|\langle \Psi_{S,f,b} | \Psi_{S',f',b'} \rangle|$.

4. Output the estimate $\hat{\gamma} = \hat{\beta}\hat{\alpha}\hat{\alpha}'$.

By a union bound, the probability that Steps 1-3 all succeed is $\geq 1 - \delta$. In this case, we have

$$\begin{aligned}
&\left| \hat{\gamma} - \Pr_{x \sim \{0,1\}^n}[x \in S \cap S', f(x) = b, f'(x) = b'] \right| \\
&= \left| \hat{\beta}\hat{\alpha}\hat{\alpha}' - \sqrt{\frac{|S \cap f^{-1}(b) \cap S' \cap f'^{-1}(b')|^2}{|S \cap f^{-1}(b)| \cdot |S' \cap f'^{-1}(b')|} \cdot \frac{|S \cap f^{-1}(b)|}{2^n} \cdot \frac{|S' \cap f'^{-1}(b')|}{2^n}} \right| \\
&\leq \left| \hat{\beta} - \sqrt{\frac{|S \cap f^{-1}(b) \cap S' \cap f'^{-1}(b')|^2}{|S \cap f^{-1}(b)| \cdot |S' \cap f'^{-1}(b')|}} \right| + \left| \hat{\alpha} - \sqrt{\frac{|S \cap f^{-1}(b)|}{2^n}} \right| + \left| \hat{\alpha}' - \sqrt{\frac{|S' \cap f'^{-1}(b')|}{2^n}} \right| \\
&\leq 3 \cdot \frac{\varepsilon}{3} \\
&= \varepsilon .
\end{aligned}$$

Here, the second step used the triangle inequality together with $\hat{\beta}, \hat{\alpha}, \hat{\alpha}', \sqrt{\frac{|S \cap f^{-1}(b) \cap S' \cap f'^{-1}(b')|^2}{|S \cap f^{-1}(b)| \cdot |S' \cap f'^{-1}(b')|}}, \sqrt{\frac{|S \cap f^{-1}(b)|}{2^n}}, \sqrt{\frac{|S' \cap f'^{-1}(b')|}{2^n}} \in [0,1]$. □

Bibliography

[AA15] Scott Aaronson and Andris Ambainis. "Forrelation: A Problem that Optimally Separates Quantum from Classical Computing". In: *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*. STOC '15. Portland, Oregon, USA: Association for Computing Machinery, 2015, pp. 307–316. DOI: 10.1145/2746539.2746547.

[Aar18] Scott Aaronson. "Shadow Tomography of Quantum States". In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2018. Los Angeles, CA, USA: Association for Computing Machinery, 2018, pp. 325–338. DOI: 10.1145/3188745.3188802.

[ABD24] Srinivasan Arunachalam, Sergey Bravyi, and Arkopal Dutt. *A note on polynomial-time tolerant testing stabilizer states*. 2024. arXiv: 2410.22220 [quant-ph].

[ABDY23] Srinivasan Arunachalam, Sergey Bravyi, Arkopal Dutt, and Theodore J. Yoder. "Optimal Algorithms for Learning Quantum Phase States". In: *18th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2023)*. Ed. by Omar Fawzi and Michael Walter. Vol. 266. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023, 3:1–3:24. DOI: 10.4230/LIPIcs.TQC.2023.3.

[ABE93] Richard Aron, Bernard Beauzamy, and Per Enflo. "Polynomials in many variables: real vs complex norms". In: *J. Approx. Theory* 74.2 (1993), pp. 181–198. DOI: 10.1006/jath.1993.1060.

[ABN23] Anurag Anshu, Nikolas P. Breuckmann, and Chinmay Nirkhe. "NLTS Hamiltonians from good quantum codes". In: *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*. arXiv:2206.13228 [cond-mat, physics:quant-ph]. June 2023, pp. 1090–1096. DOI: 10.1145/3564246.3585114.

[ABRW15] Andris Ambainis, Aleksandrs Belovs, Oded Regev, and Ronald de Wolf. "Efficient Quantum Algorithms for (Gapped) Group Testing and Junta Testing". In: *Proceedings of the 2016 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. Proceedings. Society for Industrial and Applied Mathematics, Dec. 2015, pp. 903–922. DOI: 10.1137/1.9781611974331.ch65.

[ACL11] Andris Ambainis, Andrew M. Childs, and Yi-Kai Liu. "Quantum property testing for bounded-degree graphs". In: *International Workshop on Approximation Algorithms for Combinatorial Op-*

timization. Springer. 2011, pp. 365–376. DOI: 10.1007/978-3-642-22935-0_31.

[ACQ22] Dorit Aharonov, Jordan Cotler, and Xiao-Liang Qi. "Quantum algorithmic measurement". In: *Nature Communications* 13.1 (2022), pp. 1–9. DOI: 10.1038/s41467-021-27922-0.

[AD24] Srinivasan Arunachalam and Arkopal Dutt. *Towards tolerant testing stabilizer states*. 2024. arXiv: 2408.06289 [quant-ph].

[ADFS04] N. Alon, I. Dinur, E. Friedgut, and B. Sudakov. "Graph products, Fourier analysis and spectral techniques". In: *Geom. Funct. Anal.* 14.5 (2004), pp. 913–940. DOI: 10.1007/s00039-004-0478-3.

[ADG24] Srinivasan Arunachalam, Arkopal Dutt, and Francisco Escudero Gutiérrez. *Testing and learning structured quantum Hamiltonians*. 2024. arXiv: 2411.00082 [quant-ph].

[ADOY25] Anurag Anshu, Yangjing Dong, Fengning Ou, and Penghui Yao. "On the Computational Power of QAC with Barely Superlinear Ancillae". In: *Proceedings of the 28th Conference on Quantum Information Processing (QIP 2025)*. Accepted talk. Extended abstract available at arXiv:2410.06499. 2025.

[AEHK16] Ali Asadian, Paul Erker, Marcus Huber, and Claude Klöckl. "Heisenberg-Weyl Observables: Bloch vectors in phase space". In: *Phys. Rev. A* 94 (1 July 2016), p. 010301. DOI: 10.1103/PhysRevA.94.010301.

[AHW16] Noga Alon, Rani Hod, and Amit Weinstein. "On Active and Passive Testing". In: *Comb. Probab. Comput.* 25.1 (2016), pp. 1–20. DOI: 10.1017/S0963548315000292.

[AIK22] Scott Aaronson, DeVon Ingram, and William Kretschmer. "The acrobatics of BQP". In: *Proceedings of the 37th Computational Complexity Conference*. CCC '22. Dagstuhl, DEU: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Sept. 2022, pp. 1–17. DOI: 10.4230/LIPIcs.CCC.2022.20.

[ALM91] N. Alon, N. Linial, and R. Meshulam. "Additive bases of vector spaces over prime fields". In: *J. Combin. Theory Ser. A* 57.2 (1991), pp. 203–210. DOI: https://doi.org/10.1016/0097-3165(91)90045-I.

[Alo+03] Noga Alon, Tali Kaufman, Michael Krivelevich, Simon Litsyn, and Dana Ron. "Testing low-degree polynomials over GF (2)". In:

International Workshop on Randomization and Approximation Techniques in Computer Science. Springer. 2003, pp. 188–199. DOI: 10.1007/978-3-540-45198-3_17.

[Aro+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. "Proof verification and the hardness of approximation problems". In: *J. ACM* 45.3 (May 1998), pp. 501–555. DOI: 10.1145/278298.278306.

[AS07] Alp Atc and Rocco A Servedio. "Quantum algorithms for learning and testing juntas". In: *Quantum Information Processing* 6.5 (2007), pp. 323–348. DOI: 10.1007/s11128-007-0061-6.

[AS98] Sanjeev Arora and Shmuel Safra. "Probabilistic checking of proofs: A new characterization of NP". In: *Journal of the ACM (JACM)* 45.1 (1998), pp. 70–122. DOI: 10.1145/273865.273901.

[ASSY23] Emmanuel Abbe, Ori Sberlo, Amir Shpilka, and Min Ye. "Reed-Muller Codes". In: *Foundations and Trendső in Communications and Information Theory* 20.12 (2023), pp. 1–156. DOI: 10.1561/0100000123.

[Bak22] Alan Baker. *Transcendental number theory*. Cambridge Mathematical Library. With an introduction by David Masser, Reprint of the 1975 original [0422171]. Cambridge University Press, Cambridge, 2022, pp. xiv+169. DOI: 10.1017/9781009229937.

[Bar+97] Adriano Barenco, André Berthiaume, David Deutsch, Artur Ekert, Richard Jozsa, and Chiara Macchiavello. "Stabilization of quantum computations by symmetrization". In: *SIAM Journal on Computing* 26.5 (1997), pp. 1541–1557. DOI: 10.1137/S0097539796302452.

[BB15] Aleksandrs Belovs and Eric Blais. "Quantum Algorithm for Monotonicity Testing on the Hypercube". In: *Theory of Computing* 11.1 (2015), pp. 403–412. DOI: 10.4086/toc.2015.v011a016.

[BBBY12] Maria-Florina Balcan, Eric Blais, Avrim Blum, and Liu Yang. "Active Property Testing". In: *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*. ISSN: 0272-5428. Oct. 2012, pp. 21–30. DOI: 10.1109/FOCS.2012.64.

[BBL98] Avrim Blum, Carl Burch, and John Langford. "On Learning Monotone Boolean Functions". In: *39th Annual Symposium on Foundations of Computer Science, FOCS '98, November 8-11,*

1998, Palo Alto, California, USA. IEEE Computer Society, 1998, pp. 408–415. DOI: 10.1109/SFCS.1998.743491.

[BCO24] Andreas Bluhm, Matthias C. Caro, and Aadil Oufkir. *Hamiltonian Property Testing*. 2024. arXiv: 2403.02968 [quant-ph].

[BCS23] Hadley Black, Deeparnab Chakrabarty, and C. Seshadhri. "Directed Isoperimetric Theorems for Boolean Functions on the Hypergrid and an $\widetilde{\mathcal{O}}(n\sqrt{d})$ Monotonicity Tester". In: *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*. STOC 2023. Orlando, FL, USA: Association for Computing Machinery, 2023, pp. 233–241. DOI: 10.1145/3564246.3585167.

[BCWD01] Harry Buhrman, Richard Cleve, John Watrous, and Ronald De Wolf. "Quantum fingerprinting". In: *Physical review letters* 87.16 (2001), p. 167902. DOI: 10.1103/PhysRevLett.87.167902.

[BDH24] Zongbo Bao, Philippe van Dordrecht, and Jonas Helsen. *Tolerant testing of stabilizer states with a polynomial gap via a generalized uncertainty relation*. 2024. arXiv: 2410.21811 [quant-ph].

[BDPS11] Alessandro Bisio, Giacomo Mauro D'Ariano, Paolo Perinotti, and Michal Sedlák. "Quantum learning algorithms for quantum measurements". In: *Physics Letters A* 375.39 (2011), pp. 3425–3434. DOI: 10.1016/j.physleta.2011.08.002.

[Bea+01] Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. "Quantum lower bounds by polynomials". In: *Journal of the ACM* 48.4 (July 2001), pp. 778–797. DOI: 10.1145/502090.502097.

[Bec+25] Lars Becker, Ohad Klein, Joseph Slote, Alexander Volberg, and Haonan Zhang. "Dimension-free discretizations of the uniform norm by small product sets". In: *Invent. Math.* 239.2 (2025), pp. 469–503. DOI: 10.1007/s00222-024-01306-9.

[Ben+16] Itai Benjamini, David Ellis, Ehud Friedgut, Nathan Keller, and Arnab Sen. "Juntas in the ℓ^1-grid and Lipschitz maps between discrete tori". In: *Random Struct. Algorithms* 49.2 (2016), pp. 253–279. DOI: 10.1002/RSA.20623.

[Ber31] S. Bernstein. "Sur une classe de formules d'interpolation." French. In: *Bull. Acad. Sci. URSS* 1931.9 (1931), pp. 1151–1161.

[Ber32] Serge Bernstein. *Sur une modification de la formule d'interpolation de Lagrange*. French. Commun. Soc. Math. Kharkow et Inst. Sci. Math. Ukraine, IV. Ser. 5, 49-57 (1932). 1932.

[BFH21] Eric Blais, Renato Ferreira Pinto Jr, and Nathaniel Harms. "VC dimension and distribution-free sample-based testing". In: *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. 2021, pp. 504–517. DOI: 10.1145/3406325.3451104.

[BFNR08] Harry Buhrman, Lance Fortnow, Ilan Newman, and Hein Röhrig. "Quantum property testing". In: *SIAM Journal on Computing* 37.5 (2008), pp. 1387–1400. DOI: 10.1137/S0097539704442416.

[BG73] Ju. A. Brudnyi and M. I. Ganzburg. "A certain extremal problem for polynomials in n variables". In: *Izv. Akad. Nauk SSSR Ser. Mat.* 37 (1973), pp. 344–355.

[BGK18] Sergey Bravyi, David Gosset, and Robert König. "Quantum advantage with shallow circuits". In: *Science* 362.6412 (Oct. 2018). Publisher: American Association for the Advancement of Science, pp. 308–311. DOI: 10.1126/science.aar3106.

[BGKT20] Sergey Bravyi, David Gosset, Robert König, and Marco Tomamichel. "Quantum advantage with noisy shallow circuits". en. In: *Nature Physics* 16.10 (Oct. 2020). Number: 10 Publisher: Nature Publishing Group, pp. 1040–1045. DOI: 10.1038/s41567-020-0948-z.

[BGL14] Dominique Bakry, Ivan Gentil, and Michel Ledoux. *Analysis and geometry of Markov diffusion operators*. Vol. 348. Grundlehren der mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer, Cham, 2014, pp. xx+552. DOI: 10.1007/978-3-319-00227-9.

[BH31] H. F. Bohnenblust and Einar Hille. "On the absolute convergence of Dirichlet series". In: *Ann. of Math. (2)* 32.3 (1931), pp. 600–622. DOI: 10.2307/1968255.

[Bha+10] Arnab Bhattacharyya, Swastik Kopparty, Grant Schoenebeck, Madhu Sudan, and David Zuckerman. "Optimal testing of Reed-Muller codes". In: *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*. IEEE. 2010, pp. 488–497. DOI: 10.1109/FOCS.2010.54.

[Bha+13] Arnab Bhattacharyya, Eldar Fischer, Hamed Hatami, Pooya Hatami, and Shachar Lovett. "Every locally characterized affine-invariant property is testable". In: *Proceedings of the forty-fifth*

annual ACM symposium on Theory of computing. 2013, pp. 429–436. DOI: 10.1145/2488608.2488662.

[BHH11] Sergey Bravyi, Aram W. Harrow, and Avinatan Hassidim. "Quantum algorithms for testing properties of distributions". In: *IEEE Transactions on Information Theory* 57.6 (2011), pp. 3971–3981. DOI: 10.1109/TIT.2011.2134250.

[Bis+10] Alessandro Bisio, Giulio Chiribella, Giacomo Mauro D'Ariano, Stefano Facchini, and Paolo Perinotti. "Optimal quantum learning of a unitary transformation". In: *Physical Review AAtomic, Molecular, and Optical Physics* 81.3 (2010), p. 032324. DOI: 10.1103/PhysRevA.81.032324.

[BIVW16] Andrej Bogdanov, Yuval Ishai, Emanuele Viola, and Christopher Williamson. "Bounded Indistinguishability and the Complexity of Recovering Secrets". In: *Proceedings, Part III, of the 36th Annual International Cryptology Conference on Advances in Cryptology — CRYPTO 2016 - Volume 9816*. Berlin, Heidelberg: Springer-Verlag, Aug. 2016, pp. 593–618. DOI: 10.1007/978-3-662-53015-3_21.

[BKT19] Mark Bun, Robin Kothari, and Justin Thaler. "Quantum algorithms and approximating polynomials for composed functions with shared inputs". In: *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '19. USA: Society for Industrial and Applied Mathematics, Jan. 2019, pp. 662–678.

[Bla09] Eric Blais. "Testing juntas nearly optimally". In: *Proceedings of the forty-first annual ACM symposium on Theory of computing*. STOC '09. New York, NY, USA: Association for Computing Machinery, May 2009, pp. 151–158. DOI: 10.1145/1536414.1536437.

[Bla24] Hadley Black. "Nearly Optimal Bounds for Sample-Based Testing and Learning of k-Monotone Functions". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2024)*. Ed. by Amit Kumar and Noga Ron-Zewi. Vol. 317. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2024, 37:1–37:23. DOI: 10.4230/LIPIcs.APPROX/RANDOM.2024.37.

[Ble01] Ron Blei. *Analysis in Integer and Fractional Dimensions*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2001. DOI: 10.1017/CBO9780511543012.

[BLR90] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. "Self-testing/correcting with applications to numerical problems". In: *Proceedings of the twenty-second annual ACM symposium on Theory of computing*. 1990, pp. 73–83. DOI: 10.1145/100216.100225.

[BMR19a] Piotr Berman, Meiram Murzabulatov, and Sofya Raskhodnikova. "Testing convexity of figures under the uniform distribution". In: *Random Structures & Algorithms* 54.3 (2019), pp. 413–443. DOI: https://doi.org/10.1002/rsa.20797. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/rsa.20797.

[BMR19b] Piotr Berman, Meiram Murzabulatov, and Sofya Raskhodnikova. "The Power and Limitations of Uniform Samples in Testing Properties of Figures". In: *Algorithmica* 81.3 (Mar. 2019), pp. 1247–1266. DOI: 10.1007/s00453-018-0467-9.

[BO20] Costin Bdescu and Ryan ODonnell. "Lower bounds for testing complete positivity and quantum separability". In: *Latin American Symposium on Theoretical Informatics*. Springer. 2020, pp. 375–386. DOI: 10.1007/978-3-030-61792-9_30.

[BO21] Costin Bdescu and Ryan O'Donnell. "Improved quantum data analysis". In: *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. 2021, pp. 1398–1411. DOI: 10.1145/3406325.3451109.

[Bol98] Béla Bollobás. *Modern Graph Theory*. Springer New York, 1998. DOI: 10.1007/978-1-4612-0619-4.

[BPS14] Frédéric Bayart, Daniel Pellegrino, and Juan B. Seoane-Sepúlveda. "The Bohr radius of the n-dimensional polydisk is equivalent to $(\log n)/n$". In: *Advances in Mathematics* 264 (2014), pp. 726–746. DOI: https://doi.org/10.1016/j.aim.2014.07.029.

[BS19] Zvika Brakerski and Omri Shmueli. "(Pseudo) random quantum states with binary phase". In: *Theory of Cryptography Conference*. Springer. 2019, pp. 229–250. DOI: 10.1007/978-3-030-36030-6_10.

[BS80] Roger C. Baker and Wolfgang M. Schmidt. "Diophantine problems in variables restricted to the values 0 and 1". In: *J. Number Theory* 12.4 (1980), pp. 460–486. DOI: 10.1016/0022-314X(80)90039-6.

[BT22] Mark Bun and Justin Thaler. "Approximate Degree in Classical and Quantum Computing". English. In: *Foundations and Trendső in Theoretical Computer Science* 15.3-4 (Dec. 2022). Publisher: Now Publishers, Inc., pp. 229–423. DOI: 10.1561/0400000107.

[BV97] Ethan Bernstein and Umesh Vazirani. "Quantum complexity theory". In: *SIAM J. Comput.* 26.5 (1997), pp. 1411–1473. DOI: 10.1137/S0097539796300921.

[BW02] Harry Buhrman and Ronald de Wolf. "Complexity measures and decision tree complexity: a survey". In: *Theoretical Computer Science*. Complexity and Logic 288.1 (Oct. 2002), pp. 21–43. DOI: 10.1016/S0304-3975(01)00144-X.

[BWY15] Eric Blais, Amit Weinstein, and Yuichi Yoshida. "Partially symmetric functions are efficiently isomorphism testable". In: *SIAM Journal on Computing* 44.2 (2015), pp. 411–432. DOI: 10.1137/140971877.

[BY16] A. Brudnyi and Y. Yomdin. "Norming sets and related Remez-type inequalities". In: *J. Aust. Math. Soc.* 100.2 (2016), pp. 163–181. DOI: 10.1017/S1446788715000488.

[BY19] Eric Blais and Yuichi Yoshida. "A characterization of constant-sample testable properties". In: *Random Struct. Algorithms* 55.1 (2019), pp. 73–88. DOI: 10.1002/RSA.20807.

[BY23] Zongbo Bao and Penghui Yao. "On Testing and Learning Quantum Junta Channels". In: *Proceedings of Thirty Sixth Conference on Learning Theory*. Ed. by Gergely Neu and Lorenzo Rosasco. Vol. 195. Proceedings of Machine Learning Research. PMLR, July 2023, pp. 1064–1094.

[Car24] Matthias C. Caro. "Learning quantum processes and Hamiltonians via the Pauli transfer matrix". In: *ACM Transactions on Quantum Computing* 5.2 (June 2024). DOI: 10.1145/3670418.

[CCHL22] Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. "Exponential separations between learning with and without quantum memory". In: *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE. 2022, pp. 574–585. DOI: 10.1109/FOCS52979.2021.00063.

[CFMW10] Sourav Chakraborty, Eldar Fischer, Arie Matsliah, and Ronald de Wolf. "New Results on Quantum Property Testing". In: *IARCS Annual Conference on Foundations of Software Technology and*

Theoretical Computer Science (FSTTCS 2010). Ed. by Kamal Lodaya and Meena Mahajan. Vol. 8. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2010, pp. 145–156. DOI: 10.4230/LIPIcs.FSTTCS.2010.145.

[CFSS17] Xi Chen, Adam Freilich, Rocco A. Servedio, and Timothy Sun. "Sample-Based High-Dimensional Convexity Testing". In: *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2017)*. Ed. by Klaus Jansen, José D. P. Rolim, David P. Williamson, and Santosh S. Vempala. Vol. 81. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2017, 37:1–37:20. DOI: 10.4230/LIPIcs.APPROX-RANDOM.2017.37.

[CG04] Hana Chockler and Dan Gutfreund. "A lower bound for testing juntas". In: *Information Processing Letters* 90.6 (June 2004), pp. 301–305. DOI: 10.1016/j.ipl.2004.01.023.

[CG24] Sitan Chen and Weiyuan Gong. *Efficient Pauli channel estimation with logarithmic quantum memory*. 2024. arXiv: 2309.14326 [quant-ph].

[CGY24] Sitan Chen, Weiyuan Gong, and Qi Ye. *Optimal tradeoffs for estimating Pauli observables*. 2024. arXiv: 2404.19105 [quant-ph].

[CGYZ24] Sitan Chen, Weiyuan Gong, Qi Ye, and Zhihan Zhang. *Stabilizer bootstrapping: A recipe for efficient agnostic tomography and magic estimation*. 2024. arXiv: 2408.06967 [quant-ph].

[Che+24a] Chi-Fang Chen, Adam Bouland, Fernando G. S. L. Brandão, Jordan Docter, Patrick Hayden, and Michelle Xu. *Efficient unitary designs and pseudorandom unitaries from permutations*. 2024. arXiv: 2404.16751 [quant-ph].

[Che+24b] Senrui Chen, Changhun Oh, Sisi Zhou, Hsin-Yuan Huang, and Liang Jiang. "Tight bounds on Pauli channel learning without entanglement". In: *Physical Review Letters* 132.18 (2024), p. 180805. DOI: 10.1103/PhysRevLett.132.180805.

[CHST17] Claudio Carmeli, Teiko Heinosaari, Jussi Schultz, and Alessandro Toigo. "Probing quantum state space: does one have to learn everything to learn something?" In: *Proceedings of the Royal Soci-*

ety A: Mathematical, Physical and Engineering Sciences 473.2201 (2017), p. 20160866. DOI: 10.1098/rspa.2016.0866.

[CM16] Claude Carlet and Sihem Mesnager. "Four decades of research on bent functions". In: *Designs, codes and cryptography* 78.1 (2016), pp. 5–50. DOI: 10.1007/s10623-015-0145-8.

[CNY23] Thomas Chen, Shivam Nadimpalli, and Henry Yuen. "Testing and learning quantum juntas nearly optimally". In: *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM. 2023, pp. 1163–1185. DOI: 10.1137/1.9781611977554.ch43.

[CS14] Deeparnab Chakrabarty and C. Seshadhri. "An Optimal Lower Bound for Monotonicity Testing over Hypergrids". In: *Theory of Computing* 10.17 (2014), pp. 453–464. DOI: 10.4086/toc.2014.v010a017.

[CZSJ22] Senrui Chen, Sisi Zhou, Alireza Seif, and Liang Jiang. "Quantum advantages for Pauli channel estimation". In: *Physical Review A* 105.3 (2022), p. 032435. DOI: 10.1103/PhysRevA.105.032435.

[Def+11] Andreas Defant, Leonhard Frerick, Joaquim Ortega-Cerdà, Myriam Ounaïes, and Kristian Seip. "The Bohnenblust-Hille inequality for homogeneous polynomials is hypercontractive". In: *Ann. of Math. (2)* 174.1 (2011), pp. 485–497. DOI: 10.4007/annals.2011.174.1.13.

[DFKO07] Irit Dinur, Ehud Friedgut, Guy Kindler, and Ryan O'Donnell. "On the Fourier tails of bounded functions over the discrete cube". In: *Israel Journal of Mathematics* 160.1 (Aug. 2007), pp. 389–412. DOI: 10.1007/s11856-007-0068-9.

[DGL23] Marcel Dall'Agnol, Tom Gur, and Oded Lachish. "A Structural Theorem for Local Algorithms with Applications to Coding, Testing, and Verification". In: *SIAM J. Comput.* 52.6 (2023), pp. 1413–1463. DOI: 10.1137/21M1422781.

[DGMS19] Andreas Defant, Domingo García, Manuel Maestre, and Pablo Sevilla-Peris. *Dirichlet Series and Holomorphic Functions in High Dimensions*. New Mathematical Monographs. Cambridge University Press, 2019. DOI: 10.1017/9781108691611.

[DGRT22] Marcel Dall'Agnol, Tom Gur, Subhayan Roy Moulik, and Justin Thaler. "Quantum Proofs of Proximity". In: *Quantum* 6 (Oct. 2022), p. 834. DOI: 10.22331/q-2022-10-13-834.

[Din07] Irit Dinur. "The PCP theorem by gap amplification". In: *J. ACM* 54.3 (June 2007), 12–es. DOI: 10.1145/1236457.1236459.

[DMP18] Andreas Defant, Mieczysaw Mastyo, and Antonio Pérez. "On the Fourier spectrum of functions on Boolean cubes". In: *Mathematische Annalen* 374.1-2 (Sept. 2018), pp. 653–680. DOI: 10.1007/s00208-018-1756-y.

[DP24] Feng Dai and Andriy Prymak. "Optimal polynomial meshes exist on any multivariate convex domain". In: *Found. Comput. Math.* 24.3 (2024), pp. 989–1018. DOI: 10.1007/s10208-023-09606-x.

[DPTT19] F. Dai, A. Prymak, V. N. Temlyakov, and S. Yu. Tikhonov. "Integral norm discretization and related problems". In: *Uspekhi Mat. Nauk* 74.4(448) (2019), pp. 3–58. DOI: 10.4213/rm9892.

[DS14] Andreas Defant and Pablo Sevilla-Peris. "The Bohnenblust-Hille cycle of ideas from a modern point of view". In: *Functiones et Approximatio Commentarii Mathematici* 50.1 (2014), pp. 55–127. DOI: 10.7169/facm/2014.50.1.2.

[EH80] Friedrich Esser and Frank Harary. "On the spectrum of a complete multipartite graph". In: *European J. Combin.* 1.3 (1980), pp. 211–218. DOI: 10.1016/S0195-6698(80)80004-7.

[EI22] Alexandros Eskenazis and Paata Ivanisvili. "Learning Low-Degree Functions from a Logarithmic Number of Random Queries". In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2022. Rome, Italy: Association for Computing Machinery, 2022, pp. 203–207. DOI: 10.1145/3519935.3519981.

[Fis04] Eldar Fischer. "The art of uninformed decisions: A primer to property testing". In: *Current Trends in Theoretical Computer Science: The Challenge of the New Century Vol 1: Algorithms and Complexity Vol 2: Formal Models and Semantics*. World Scientific, 2004, pp. 229–263. DOI: 10.1142/9789812562494_0014.

[Fis24] Eldar Fischer. *A basic lower bound for property testing*. 2024. arXiv: 2403.04999 [cs.DS].

[FLV15] Eldar Fischer, Oded Lachish, and Yadu Vasudev. "Trading Query Complexity for Sample-Based Testing and Multi-testing Scalability". In: *IEEE 56th Annual Symposium on Foundations of Computer Science, FOCS 2015, Berkeley, CA, USA, 17-20 Oc-*

tober, 2015. Ed. by Venkatesan Guruswami. IEEE Computer Society, 2015, pp. 1163–1182. DOI: 10.1109/FOCS.2015.75.

[Fon06] Natacha Fontes-Merz. "A multidimensional version of Turán's lemma". In: *J. Approx. Theory* 140.1 (2006), pp. 27–30. DOI: 10.1016/j.jat.2005.11.012.

[Fox11] Jacob Fox. "A new proof of the graph removal lemma". In: *Annals of Mathematics* (2011), pp. 561–579. DOI: 10.4007/annals.2011.174.1.17.

[Fra09] Matthieu Fradelizi. "Concentration inequalities for s-concave measures of dilations of Borel sets and applications". In: *Electron. J. Probab.* 14 (2009), no. 71, 2068–2090. DOI: 10.1214/EJP.v14-695.

[FSMS09] Katalin Friedl, Miklos Santha, Frédéric Magniez, and Pranab Sen. "Quantum testers for hidden group properties". In: *Fundamenta Informaticae* 91.2 (2009), pp. 325–340. DOI: 10.3233/FI-2009-0046.

[FY13] Omer Friedland and Yosef Yomdin. "An observation on the Turán-Nazarov inequality". In: *Studia Math.* 218.1 (2013), pp. 27–39. DOI: 10.4064/sm218-1-2.

[Gal20] François Le Gall. "Average-case quantum advantage with shallow circuits". In: *Proceedings of the 34th Computational Complexity Conference*. CCC '19. Dagstuhl, DEU: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Sept. 2020, pp. 1–20. DOI: 10.4230/LIPIcs.CCC.2019.21.

[Gan17] Michael I. Ganzburg. "A multivariate Remez-type inequality with φ-concave weights". In: *Colloq. Math.* 147.2 (2017), pp. 221–240. DOI: 10.4064/cm6884-7-2016.

[GB23] Tudor Giurgica-Tiron and Adam Bouland. *Pseudorandomness from Subset States*. 2023. arXiv: 2312.09206 [quant-ph].

[GGR98] Oded Goldreich, Shafi Goldwasser, and Dana Ron. "Property testing and its connection to learning and approximation". In: *Journal of the ACM (JACM)* 45.4 (1998), pp. 653–750. DOI: 10.1145/285055.285060.

[GIKL23] Sabee Grewal, Vishnu Iyer, William Kretschmer, and Daniel Liang. *Improved Stabilizer Estimation via Bell Difference Sampling*. 2023. arXiv: 2304.13915 [quant-ph].

[GKK18] Alexandru Gheorghiu, Theodoros Kapourniotis, and Elham Kashefi. "Verification of Quantum Computation: An Overview of Existing Approaches". In: *Theory of Computing Systems* 63.4 (July 2018), pp. 715–808. DOI: 10.1007/s00224-018-9872-3.

[GL20] András Gilyén and Tongyang Li. "Distributional Property Testing in a Quantum World". In: *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)*. Ed. by Thomas Vidick. Vol. 151. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020, 25:1–25:19. DOI: 10.4230/LIPIcs.ITCS.2020.25.

[GNW21] David Gross, Sepehr Nezami, and Michael Walter. "Schur-Weyl duality for the Clifford group with applications: property testing, a robust Hudson theorem, and de Finetti representations". In: *Comm. Math. Phys.* 385.3 (2021), pp. 1325–1393. DOI: 10.1007/s00220-021-04118-7.

[Gol+00] Oded Goldreich, Shafi Goldwasser, Eric Lehman, Dana Ron, and Alex Samorodnitsky. "Testing Monotonicity". In: *Combinatorica* 20.3 (Mar. 2000), pp. 301–337. DOI: 10.1007/s004930070011.

[Gol17] Oded Goldreich. *Introduction to Property Testing*. Cambridge University Press, 2017. DOI: 10.1017/9781108135252.

[Gon+22] Daniel González-Cuadra, Torsten V. Zache, Jose Carrasco, Barbara Kraus, and Peter Zoller. "Hardware Efficient Quantum Simulation of Non-Abelian Gauge Theories with Qudits on Rydberg Platforms". In: *Phys. Rev. Lett.* 129 (16 Oct. 2022), p. 160501. DOI: 10.1103/PhysRevLett.129.160501.

[GR16] Oded Goldreich and Dana Ron. "On Sample-Based Testers". In: *ACM Transactions on Computation Theory* 8.2 (Apr. 2016), 7:1–7:54. DOI: 10.1145/2898355.

[GRS23] Venkatesan Guruswami, Atri Rudra, and Madhu Sudan. "Essential Coding Theory". Book draft. 2023.

[GS20] Daniel Grier and Luke Schaeffer. "Interactive shallow Clifford circuits: Quantum advantage against NC1 and beyond". In: *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2020. New York, NY, USA: Association for Computing Machinery, June 2020, pp. 875–888. DOI: 10.1145/3357713.3384332.

[HA11] Mark Hillery and Erika Andersson. "Quantum tests for the linearity and permutation invariance of Boolean functions". In: *Physical Review A* 84.6 (2011), p. 062329. DOI: 10.1103/PhysRevA.84.062329.

[Hås86] J Håstad. "Almost optimal lower bounds for small depth circuits". In: *Proceedings of the eighteenth annual ACM symposium on Theory of computing*. STOC '86. New York, NY, USA: Association for Computing Machinery, Nov. 1986, pp. 6–20. DOI: 10.1145/12130.12132.

[HCP22] Hsin-Yuan Huang, Sitan Chen, and John Preskill. "Learning to predict arbitrary quantum processes". In: (Oct. 2022). arXiv: 2210.14894 [quant-ph].

[HCP23] Hsin-Yuan Huang, Sitan Chen, and John Preskill. "Learning to predict arbitrary quantum processes". In: *PRX Quantum* 4.4 (2023), p. 040337. DOI: 10.1103/PRXQuantum.4.040337.

[HH24] Marcel Hinsche and Jonas Helsen. *Single-copy stabilizer testing*. 2024. arXiv: 2410.07986 [quant-ph].

[HHL+19] Hamed Hatami, Pooya Hatami, Shachar Lovett, et al. "Higher-order fourier analysis and applications". In: *Foundations and Trendső in Theoretical Computer Science* 13.4 (2019), pp. 247–448. DOI: 10.1561/0400000064.

[HK07] Shirley Halevy and Eyal Kushilevitz. "Distribution-Free Property-Testing". In: *SIAM Journal on Computing* 37.4 (2007), pp. 1107–1138. DOI: 10.1137/050645804.

[HKP20] Hsin-Yuan Huang, Richard Kueng, and John Preskill. "Predicting many properties of a quantum system from very few measurements". In: *Nature Physics* 16 (10 2020), pp. 1050–1057.

[HKP21] Hsin-Yuan Huang, Richard Kueng, and John Preskill. "Information-theoretic bounds on quantum advantage in machine learning". In: *Physical Review Letters* 126.19 (2021), p. 190505. DOI: 10.1103/PhysRevLett.126.190505.

[HLM17] Aram W. Harrow, Cedric Yen-Yu Lin, and Ashley Montanaro. "Sequential measurements, disturbance and property testing". In: *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM. 2017, pp. 1598–1611. DOI: 10.1137/1.9781611974782.105.

[HM13] Aram W Harrow and Ashley Montanaro. "Testing product states, quantum Merlin-Arthur games and tensor optimization". In: *Journal of the ACM (JACM)* 60.1 (2013), pp. 1–43. DOI: 10.1145/2432622.2432625.

[H03] Peter Høyer and Robert palek. "Quantum Circuits with Unbounded Fan-out". en. In: *STACS 2003*. Ed. by Helmut Alt and Michel Habib. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2003, pp. 234–246. DOI: 10.1007/3-540-36494-3_22.

[HST16] Pooya Hatami, Sushant Sachdeva, and Madhur Tulsiani. *An Arithmetic Analogue of Fox's Triangle Removal Argument*. 2016. arXiv: 1304.4921 [math.CO].

[Hu17] Yaozhong Hu. *Analysis on Gaussian spaces*. World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ, 2017, pp. xi+470.

[Hua+22] Hsin-Yuan Huang, Michael Broughton, Jordan Cotler, Sitan Chen, Jerry Li, Masoud Mohseni, Hartmut Neven, Ryan Babbush, Richard Kueng, John Preskill, and Jarrod R. McClean. "Quantum advantage in learning from experiments". In: *Science* 376.6598 (2022), pp. 1182–1186. DOI: 10.1126/science.abn7293.

[IL11] Yoshifumi Inui and François Le Gall. "Quantum property testing of group solvability". In: *Algorithmica* 59 (2011), pp. 35–47. DOI: 10.1007/s00453-009-9338-8.

[Iye+21] Siddharth Iyer, Anup Rao, Victor Reis, Thomas Rothvoss, and Amir Yehudayoff. "Tight bounds on the Fourier growth of bounded functions on the hypercube". In: *Electron. Colloquium Comput. Complex.* TR21-102 (2021). ECCC: TR21-102.

[Jan97] Svante Janson. "On complex hypercontractivity". In: *J. Funct. Anal.* 151.1 (1997), pp. 270–280. DOI: 10.1006/jfan.1997.3144.

[JLS18] Zhengfeng Ji, Yi-Kai Liu, and Fang Song. "Pseudorandom quantum states". In: *Advances in Cryptology–CRYPTO 2018: 38th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 19–23, 2018, Proceedings, Part III 38*. Springer. 2018, pp. 126–152. DOI: 10.1007/978-3-319-96878-0_5.

[JMW24] Fernando Granha Jeronimo, Nir Magrafta, and Pei Wu. *Pseudorandom and Pseudoentangled States from Subset States*. 2024. arXiv: 2312.15285 [quant-ph].

[JPPP17] Marius Junge, Carlos Palazuelos, Javier Parcet, and Mathilde Perrin. "Hypercontractivity in group von Neumann algebras". In: *Mem. Amer. Math. Soc.* 249.1183 (2017), pp. xii+83. DOI: 10.1090/memo/1183.

[KGKB24] Robbie King, David Gosset, Robin Kothari, and Ryan Babbush. *Triply efficient shadow tomography*. 2024. arXiv: 2404.19211 [quant-ph].

[KKLT22] B. Kashin, E. Kosov, I. Limonova, and V. Temlyakov. "Sampling discretization and related problems". In: *J. Complexity* 71 (2022), Paper No. 101653, 55. DOI: 10.1016/j.jco.2022.101653.

[KKMO07] Subhash Khot, Guy Kindler, Elchanan Mossel, and Ryan O'Donnell. "Optimal inapproximability results for MAX-CUT and other 2-variable CSPs?" In: *SIAM Journal on Computing* 37.1 (2007), pp. 319–357.

[KKT23] B. Kashin, S. Konyagin, and V. Temlyakov. "Sampling discretization of the uniform norm". In: *Constr. Approx.* 57.2 (2023), pp. 663–694. DOI: 10.1007/s00365-023-09618-4.

[Kli95] Maciej Klimek. "Metrics associated with extremal plurisubharmonic functions". In: *Proc. Amer. Math. Soc.* 123.9 (1995), pp. 2763–2770. DOI: 10.2307/2160572.

[KMS18] Subhash Khot, Dor Minzer, and Muli Safra. "On Monotonicity Testing and Boolean Isoperimetric-type Theorems". In: *SIAM J. Comput.* 47.6 (2018), pp. 2238–2276. DOI: 10.1137/16M1065872.

[KNOW14] Pravesh Kothari, Amir Nayyeri, Ryan O'Donnell, and Chenggang Wu. "Testing surface area". In: *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*. SIAM. 2014, pp. 1204–1214. DOI: 10.1137/1.9781611973402.89.

[KR98] Michael Kearns and Dana Ron. "Testing problems with sublearning sample complexity". In: *Proceedings of the eleventh annual conference on Computational learning theory*. 1998, pp. 268–279. DOI: 10.1145/279943.279996.

[Kro11] András Kroó. "On optimal polynomial meshes". In: *J. Approx. Theory* 163.9 (2011), pp. 1107–1124. DOI: 10.1016/j.jat.2011.03.007.

[KS97] András Kroó and Darrell Schmidt. “Some extremal problems for multivariate polynomials on convex bodies”. In: *J. Approx. Theory* 90.3 (1997), pp. 415–434. DOI: 10.1006/jath.1996.3083.

[KSVZ24] Ohad Klein, Joseph Slote, Alexander Volberg, and Haonan Zhang. “Quantum and Classical Low-Degree Learning via a Dimension-Free Remez Inequality”. In: *15th Innovations in Theoretical Computer Science Conference, ITCS 2024, January 30 to February 2, 2024, Berkeley, CA, USA*. Ed. by Venkatesan Guruswami. Vol. 287. LIPIcs. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2024, 69:1–69:22. DOI: 10.4230/LIPICS.ITCS.2024.69.

[Kur+21] Doga Murat Kurkcuoglu, M. Sohaib Alam, Joshua Adam Job, Andy C. Y. Li, Alexandru Macridin, Gabriel N. Perdue, and Stephen Providence. “Quantum simulation of ϕ^4 theories in qudit systems”. In: (2021). DOI: 10.48550/ARXIV.2108.13357.

[LK24] Paulina Lewandowska and Ryszard Kukulski. “Storage and retrieval of von Neumann measurements via indefinite causal order structures”. In: *Physical Review A* 110.4 (2024), p. 042422. DOI: 10.1103/PhysRevA.110.042422.

[LKPP22] Paulina Lewandowska, Ryszard Kukulski, ukasz Pawela, and Zbigniew Puchaa. “Storage and retrieval of von Neumann measurements”. In: *Physical Review A* 106.5 (2022), p. 052423. DOI: 10.1103/PhysRevA.106.052423.

[LMN93a] Nathan Linial, Yishay Mansour, and Noam Nisan. “Constant Depth Circuits, Fourier Transform, and Learnability”. In: *J. ACM* 40.3 (July 1993), pp. 607–620. DOI: 10.1145/174130.174138.

[LMN93b] Nathan Linial, Yishay Mansour, and Noam Nisan. “Constant depth circuits, Fourier transform, and learnability”. In: *Journal of the ACM* 40.3 (July 1993), pp. 607–620. DOI: 10.1145/174130.174138.

[Lun85] Magnus Lundin. “The extremal PSH for the complement of convex, symmetric subsets of $\mathbf{R}^N$”. In: *Michigan Math. J.* 32.2 (1985), pp. 197–201. DOI: 10.1307/mmj/1029003186.

[LV11] Shachar Lovett and Emanuele Viola. “Bounded-Depth Circuits Cannot Sample Good Codes”. In: *2011 IEEE 26th Annual Conference on Computational Complexity*. ISSN: 1093-0159. June 2011, pp. 243–251. DOI: 10.1109/CCC.2011.11.

[LW22] Margarite L. LaBorde and Mark M. Wilde. "Quantum algorithms for testing Hamiltonian symmetry". In: *Physical Review Letters* 129.16 (2022), p. 160503. DOI: 10.1103/PhysRevLett.129.160503.

[Mah18] Urmila Mahadev. "Classical Verification of Quantum Computations". In: *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*. 2018, pp. 259–267. DOI: 10.1109/FOCS.2018.00033.

[Meh23] Ranjit Mehatari. "Eigenvalues of multipart matrices and their applications". In: *American Journal of Combinatorics Research Article* 2 (2023), pp. 59–71.

[Mes95] Roy Meshulam. "On subsets of finite abelian groups with no 3-term arithmetic progressions". In: *J. Combin. Theory Ser. A* 71.1 (1995), pp. 168–172. DOI: 10.1016/0097-3165(95)90024-1.

[MH24] Fermi Ma and Hsin-Yuan Huang. *How to Construct Random Unitaries*. 2024. arXiv: 2410.10116 [quant-ph].

[MOO10] Elchanan Mossel, Ryan O'Donnell, and Krzysztof Oleszkiewicz. "Noise stability of functions with low influences: Invariance and optimality". In: *Ann. of Math.* 171.1 (Mar. 2010), pp. 295–341. DOI: 10.4007/annals.2010.171.295.

[Moo99] Cristopher Moore. *Quantum Circuits: Fanout, Parity, and Counting*. en. Tech. rep. TR99-032. ISSN: 1433-8092. Electronic Colloquium on Computational Complexity (ECCC), Sept. 1999.

[MORS10] Kevin Matulef, Ryan O'Donnell, Ronitt Rubinfeld, and Rocco A. Servedio. "Testing Halfspaces". In: *SIAM Journal on Computing* 39.5 (2010), pp. 2004–2047. DOI: 10.1137/070707890.

[MPSY24] Tony Metger, Alexander Poremba, Makrand Sinha, and Henry Yuen. *Simple constructions of linear-depth t-designs and pseudorandom unitaries*. 2024. arXiv: 2404.12647 [quant-ph].

[MS86] Vitali D. Milman and Gideon Schechtman. *Asymptotic theory of finite dimensional normed spaces*. Lecture notes in mathematics. Berlin, Heidelberg: Springer Berlin Heidelberg, 1986.

[MT24] Saeed Mehraban and Mehrdad Tahmasbi. *Improved bounds for testing low stabilizer complexity states*. 2024. arXiv: 2410.24202 [quant-ph].

[MW16] Ashley Montanaro and Ronald de Wolf. *A Survey of Quantum Property Testing*. Graduate Surveys 7. Theory of Computing Library, 2016, pp. 1–81. DOI: 10.4086/toc.gs.2016.007.

[Naz93] F. L. Nazarov. "Local estimates for exponential polynomials and their applications to inequalities of the uncertainty principle type". In: *Algebra i Analiz* 5.4 (1993), pp. 3–66.

[Nee14] Joe Neeman. "Testing surface area with arbitrary accuracy". In: *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*. STOC '14. New York, New York: Association for Computing Machinery, 2014, pp. 393–397. DOI: 10.1145/2591796.2591807.

[NPVY23] Shivam Nadimpalli, Natalie Parham, Francisca Vasconcelos, and Henry Yuen. *On the Pauli Spectrum of QAC0*. 2023. DOI: 10.48550/ARXIV.2311.09631.

[NSV02] F. Nazarov, M. Sodin, and A. Volberg. "The geometric Kannan-Lovász-Simonovits lemma, dimension-free estimates for the distribution of the values of polynomials, and the distribution of the zeros of random analytic functions". In: *Algebra i Analiz* 14.2 (2002), pp. 214–234.

[NSV03] F. Nazarov, M. Sodin, and A. Volberg. "Local dimension-free estimates for volumes of sublevel sets of analytic functions". In: *Israel J. Math.* 133 (2003), pp. 269–283. DOI: 10.1007/BF02773070.

[ODo14] Ryan O'Donnell. *Analysis of Boolean functions*. Cambridge University Press, New York, 2014, pp. xx+423. DOI: 10.1017/CBO9781139814782.

[ODo21] Ryan O'Donnell. *Analysis of Boolean Functions*. 2021. arXiv: 2105.10386 [cs.DM].

[OS07] Joaquim Ortega-Cerdà and Jordi Saludes. "Marcinkiewicz-Zygmund inequalities". In: *J. Approx. Theory* 145.2 (2007), pp. 237–252. DOI: 10.1016/j.jat.2006.09.001.

[OW15] Ryan O'Donnell and John Wright. "Quantum spectrum testing". In: *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*. 2015, pp. 529–538. DOI: 10.1145/2746539.2746582.

[Pat17] Siddhi Pathak. "On a conjecture of Livingston". In: *Canad. Math. Bull.* 60.1 (2017), pp. 184–195. DOI: 10.4153/CMB-2016-065-1.

[PFGT20] Daniel Padé, Stephen Fenner, Daniel Grier, and Thomas Thierauf. *Depth-2 QAC circuits cannot simulate quantum parity.* arXiv:2005.12169 [quant-ph]. May 2020. DOI: 10.48550/arXiv.2005.12169.

[PRR06] Michal Parnas, Dana Ron, and Ronitt Rubinfeld. "Tolerant property testing and distance approximation". In: *Journal of Computer and System Sciences* 72.6 (2006), pp. 1012–1042. DOI: 10.1016/j.jcss.2006.03.002.

[Raz87] A. A. Razborov. "Lower bounds on the size of bounded depth circuits over a complete basis with logical addition". en. In: *Mathematical notes of the Academy of Sciences of the USSR* 41.4 (Apr. 1987), pp. 333–338. DOI: 10.1007/BF01137685.

[Rem36] E. Remes. *Sur une propriété extrémale des polynomes de Tchébychef.* French. Commun. Inst. Sci. math. méc. Univ. Kharkoff Soc. math. Kharkoff (4) 13, 93-95 (1936). 1936.

[Ron08] Dana Ron. "Property testing: A learning theory perspective". In: *Foundations and Trendső in Machine Learning* 1.3 (2008), pp. 307–402. DOI: 10.1561/2200000004.

[Ron09] Dana Ron. "Algorithmic and Analysis Techniques in Property Testing". In: *Found. Trends Theor. Comput. Sci.* 5.2 (2009), pp. 73–205. DOI: 10.1561/0400000029.

[Ros21] Gregory Rosenthal. "Bounds on the QAC^0 Complexity of Approximating Parity". In: *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*. Ed. by James R. Lee. Vol. 185. Leibniz International Proceedings in Informatics (LIPIcs). ISSN: 1868-8969. Dagstuhl, Germany: Schloss DagstuhlLeibniz-Zentrum für Informatik, 2021, 32:1–32:20. DOI: 10.4230/LIPIcs.ITCS.2021.32.

[RS78] Imre Z Ruzsa and Endre Szemerédi. "Triple systems with no six points carrying three triangles". In: *Combinatorics (Keszthely, 1976), Coll. Math. Soc. J. Bolyai* 2.18 (1978), pp. 939–945.

[RS96] Ronitt Rubinfeld and Madhu Sudan. "Robust characterizations of polynomials with applications to program testing". In: *SIAM Journal on Computing* 25.2 (1996), pp. 252–271. DOI: 10.1137/S0097539793255151.

[Rub07] Ronitt Rubinfeld. "Sublinear time algorithms". In: *Proceedings of the International Congress of Mathematicians Madrid, August*

2230, 2006. EMS Press, May 2007, pp. 1095–1110. DOI: 10.4171/022-3/53.

[RWZ24a] Cambyse Rouzé, Melchior Wirth, and Haonan Zhang. "Quantum Talagrand, KKL and Friedgut's theorems and the learnability of quantum Boolean functions". In: *Comm. Math. Phys.* 405.4 (2024), Paper No. 95, 47. DOI: 10.1007/s00220-024-04981-0.

[RWZ24b] Cambyse Rouzé, Melchior Wirth, and Haonan Zhang. "Quantum Talagrand, KKL and Friedgut's theorems and the learnability of quantum Boolean functions". In: *Comm. Math. Phys.* 405.4 (2024), Paper No. 95, 47. DOI: 10.1007/s00220-024-04981-0.

[SBZ19] Michal Sedlák, Alessandro Bisio, and Mário Ziman. "Optimal probabilistic storage and retrieval of unitary channels". In: *Physical review letters* 122.17 (2019), p. 170502. DOI: 10.1103/PhysRevLett.122.170502.

[She22] Alexander A. Sherstov. "The approximate degree of DNF and CNF formulas". In: *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2022. New York, NY, USA: Association for Computing Machinery, June 2022, pp. 1194–1207. DOI: 10.1145/3519935.3520000.

[SHH24] Thomas Schuster, Jonas Haferkamp, and Hsin-Yuan Huang. *Random unitaries in extremely low depth*. 2024. arXiv: 2407.07754 [quant-ph].

[Sim97] Daniel R Simon. "On the power of quantum computation". In: *SIAM journal on computing* 26.5 (1997), pp. 1474–1483. DOI: 10.1137/S0097539796298637.

[Smo87a] R. 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. New York, NY, USA: Association for Computing Machinery, Jan. 1987, pp. 77–82. DOI: 10.1145/28395.28404.

[Smo87b] Roman Smolensky. "Algebraic Methods in the Theory of Lower Bounds for Boolean Circuit Complexity". In: *Proceedings of the 19th Annual ACM Symposium on Theory of Computing, 1987, New York, New York, USA*. Ed. by Alfred V. Aho. ACM, 1987, pp. 77–82. DOI: 10.1145/28395.28404.

[Smo93] Roman Smolensky. "On Representations by Low-Degree Polynomials". In: *34th Annual Symposium on Foundations of Computer*

Science, Palo Alto, California, USA, 3-5 November 1993. IEEE Computer Society, 1993, pp. 130–138. DOI: 10.1109/SFCS.1993.366874.

[Sud10] Madhu Sudan. "Invariance in Property Testing". In: *Property Testing - Current Research and Surveys*. Ed. by Oded Goldreich. Vol. 6390. Lecture Notes in Computer Science. Springer, 2010, pp. 211–227. DOI: 10.1007/978-3-642-16367-8_12.

[SVZ24a] Joseph Slote, Alexander Volberg, and Haonan Zhang. "Bohnenblust-Hille inequality for cyclic groups". In: *Adv. Math.* 452 (2024), Paper No. 109824, 35. DOI: 10.1016/j.aim.2024.109824.

[SVZ24b] Joseph Slote, Alexander Volberg, and Haonan Zhang. *Noncommutative Bohnenblust–Hille Inequality for qudit systems*. 2024. arXiv: 2406.08509 [math.FA].

[SVZ25] Joseph Slote, Alexander Volberg, and Haonan Zhang. "A dimension-free Remez-type inequality on the polytorus". In: *Discrete Analysis (to appear)* (2025). arXiv: 2305.10828 [math.CA].

[SW22] Mehdi Soleimanifar and John Wright. "Testing matrix product states". In: *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. SIAM. 2022, pp. 1679–1701. DOI: 10.1137/1.9781611977073.68.

[SY23] Adrian She and Henry Yuen. "Unitary Property Testing Lower Bounds by Polynomials". In: *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*. Ed. by Yael Tauman Kalai. Vol. 251. Leibniz International Proceedings in Informatics (LIPIcs). Dagstuhl, Germany: Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2023, 96:1–96:17. DOI: 10.4230/LIPIcs.ITCS.2023.96.

[Sze76] Endre Szemerédi. "Regular partitions of graphs". In: *Proc. Colloq. Inter. CNRS, 260 – Problèmes combinatoires et théorie des graphes, Orsay*. Vol. 260. 1976, pp. 399–401.

[Tal17] Avishay Tal. "Tight bounds on the Fourier spectrum of AC0". In: *Proceedings of the 32nd Computational Complexity Conference*. CCC '17. Dagstuhl, DEU: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, July 2017, pp. 1–31.

[Tal96] Michel Talagrand. "How much are increasing sets positively correlated?" In: *Combinatorica* 16.2 (1996), pp. 243–258.

[Tur53] Paul Turán. *Eine neue Methode in der Analysis und deren Anwendungen*. Budapest: Akadémiai Kiadó, 1953.

[Val15] Gregory Valiant. "Finding Correlations in Subquadratic Time, with Applications to Learning Parities and the Closest Pair Problem". In: *J. ACM* 62.2 (2015), 13:1–13:45. DOI: 10.1145/2728167.

[Val84] Leslie G. Valiant. "A theory of the learnable". In: *Communications of the ACM* 27.11 (1984), pp. 1134–1142. DOI: 10.1145/1968.1972.

[VZ23] Alexander Volberg and Haonan Zhang. "Noncommutative Bohnenblust–Hille inequalities". In: *Mathematische Annalen* (2023), pp. 1–20. DOI: 10.1007/s00208-023-02680-0.

[Wei80] Fred B. Weissler. "Logarithmic Sobolev inequalities and hypercontractive estimates on the circle". In: *J. Functional Analysis* 37.2 (1980), pp. 218–234. DOI: 10.1016/0022-1236(80)90042-7.

[WHSK20] Yuchen Wang, Zixuan Hu, Barry C. Sanders, and Sabre Kais. "Qudits and High-Dimensional Quantum Computing". In: *Frontiers in Physics* 8 (2020). DOI: 10.3389/fphy.2020.589504.

[Wil94] Herbert S. Wilf. *generatingfunctionology*. Second Edition. San Diego: Academic Press, 1994. DOI: https://doi.org/10.1016/B978-0-08-057151-5.50006-X.

[WKST19] Adam Bene Watts, Robin Kothari, Luke Schaeffer, and Avishay Tal. "Exponential separation between shallow quantum circuits and unbounded fan-in shallow classical circuits". In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*. STOC 2019. New York, NY, USA: Association for Computing Machinery, June 2019, pp. 515–526. DOI: 10.1145/3313276.3316404.

[WP23] Adam Bene Watts and Natalie Parham. *Unconditional Quantum Advantage for Sampling with Shallow Circuits*. arXiv:2301.00995 [quant-ph]. Jan. 2023. DOI: 10.48550/arXiv.2301.00995.

[Yom11] Y. Yomdin. "Remez-type inequality for discrete sets". In: *Israel Journal of Mathematics* 186.1 (Nov. 2011), pp. 45–60. DOI: 10.1007/s11856-011-0131-4.

[YZ22] Takashi Yamakawa and Mark Zhandry. "Verifiable Quantum Advantage without Structure". English. In: IEEE Computer Society, Oct. 2022, pp. 69–74. DOI: 10.1109/FOCS54457.2022.00014.

[Zyg02] A. Zygmund. *Trigonometric series. Vol. I, II.* Third. Cambridge Mathematical Library. With a foreword by Robert A. Fefferman. Cambridge University Press, Cambridge, 2002, xii, Vol. I: xiv+383 pp., Vol. II: viii+364.

Index