Quantum and Classical Low-Degree Learning via a Dimension-Free Remez Inequality

Ohad Klein
School of Engineering and Computer Science, Hebrew University, Jerusalem, Israel

Joseph Slote
Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA

Alexander Volberg
Department of Mathematics, Michigan State University, Ann Arbor, MI, USA
Hausdorff Center for Mathematics, University of Bonn, Germany

Haonan Zhang
Department of Mathematics, University of South Carolina, Columbia, SC, USA

Abstract

Recent efforts in Analysis of Boolean Functions aim to extend core results to new spaces, including to the slice $\binom{[n]}{k}$, the hypergrid $[K]^n$, and noncommutative spaces (matrix algebras). We present here a new way to relate functions on the hypergrid (or products of cyclic groups) to their harmonic extensions over the polytorus. We show the supremum of a function f over products of the cyclic group $\{\exp(2\pi ik/K)\}_{k=1}^K$ controls the supremum of f over the entire polytorus ($\{z \in \mathbb{C} : |z| = 1\}^n$), with multiplicative constant C depending on K and $\deg(f)$ only. This Remez-type inequality appears to be the first such estimate that is dimension-free (*i.e.,* C does not depend on n).

This dimension-free Remez-type inequality removes the main technical barrier to giving $\mathcal{O}(\log n)$ sample complexity, polytime algorithms for learning low-degree polynomials on the hypergrid and low-degree observables on level-K qudit systems. In particular, our dimension-free Remez inequality implies new Bohnenblust–Hille-type estimates which are central to the learning algorithms and appear unobtainable via standard techniques. Thus we extend to new spaces a recent line of work [10, 13, 23] that gave similarly efficient methods for learning low-degree polynomials on the hypercube and observables on qubits.

An additional product of these efforts is a new class of distributions over which arbitrary quantum observables are well-approximated by their low-degree truncations – a phenomenon that greatly extends the reach of low-degree learning in quantum science [13].

2012 ACM Subject Classification Mathematics of computing → Mathematical analysis; Theory of computation → Boolean function learning; Theory of computation → Quantum computation theory

Keywords and phrases Analysis of Boolean Functions, Remez Inequality, Bohnenblust–Hille Inequality, Statistical Learning Theory, Qudits

Digital Object Identifier 10.4230/LIPIcs.ITCS.2024.69

Related Version *Full Version*: `https://arxiv.org/abs/2301.01438` [16]

Funding Part of this work was started while J.S., A.V., and H.Z. were in residence at the Institute for Computational and Experimental Research in Mathematics in Providence, RI, during the Harmonic Analysis and Convexity program. It is partially supported by NSF DMS-1929284.
Ohad Klein: supported in part by a grant from the Israel Science Foundation (ISF Grant No. 1774/20), and by a grant from the US-Israel Binational Science Foundation and the US National Science Foundation (BSF-NSF Grant No. 2020643).
Joseph Slote: supported by Chris Umans' Simons Foundation Investigator Grant.
Alexander Volberg: supported by NSF grant DMS 2154402 and by the Hausdorff Center of Mathematics, University of Bonn.

1 Introduction

1.1 Motivation: quantum and classical low-degree learning

Recall that any function $f : \{-1,1\}^n \to \mathbb{R}$ admits a unique Fourier expansion

$$f(x) = \sum_{S\subseteq[n]} \widehat{f}(S) \prod_{i\in S} x_i \quad \text{where} \quad \widehat{f}(S) = \mathbb{E}_{x\sim\{-1,1\}^n}[f(x)\cdot\prod_{i\in S} x_i]\,.$$

We say f is *of degree* d ($\deg(f) \le d$) if for all S with $|S| > d$, $\widehat{f}(S) = 0$.

Low-degree learning – which we shall take to mean learning an L_2 approximation to f with $\deg(f) \le d$ and $|f| \le 1$ from uniformly random samples – is a fundamental task in computer science [19]. For a long time, the best polytime algorithm for low-degree learning had a sample complexity exponentially separated in n from what was information theoretically-required ($\mathrm{poly}(n)$ vs. $\mathcal{O}(\log n)$) [18, 15]. But in 2022 Eskenazis and Ivanisvili closed the gap [10], achieving a sample complexity of $\mathcal{O}(\log n)$ for the first time.

Key to their argument is an observation about approximating sparse vectors. Suppose one holds a vector $w \in \mathbb{R}^n$ that is an ℓ_∞ approximation to some unknown vector $v \in \mathbb{R}^n$ (say, $\|v - w\|_\infty \le \varepsilon$). Then it could be that $\|v - w\|_2$ grows unboundedly with n. However, if a constant ℓ_p bound on v for $p < 2$ is promised, it is possible to modify w in a simple way to obtain a new approximation $\widetilde{w}$ to v with $\|\widetilde{w} - v\|_2$ independent of n (and controlled by ε). This is possible because the ℓ_p bound on v implies v is approximately sparse – and in particular that many of its coordinates are small in comparison to ε. Using this observation [10] shows that zeroing-out coordinates in w below a fixed threshold gives a suitable $\widetilde{w}$.

In the context of low-degree learning, v is the vector of true low-degree Fourier coefficients of f, while w is a vector of empirical Fourier coefficients collected through the familiar technique of Fourier sampling [18, 19]. Modifying $w \mapsto \widetilde{w}$ as above we may form the approximate function $\widetilde{f}$ and with Plancherel's theorem conclude $\|f - \widetilde{f}\|_2 = \|v - \widetilde{w}\|_2$ is small.

Thus the Eskenazis–Ivanisvili argument reduces learning low-degree polynomials to finding an $\ell_p, p < 2$ bound on the Fourier coefficients of f. Inequalities of this kind are known as Bohnenblust–Hille-type inequalities, and the state of the art for polynomials over the hypercube was proved recently in [8]:

▶ **Theorem 1** (Hypercube Bohnenblust–Hille [8]). *For any $d \ge 1$, there exists a constant $C_d > 0$ such that for all $n \ge 1$ and all $f : \{-1,1\}^n \to \mathbb{R}$ with $\deg(f) \le d$, we have*

$$\|\widehat{f}\|_{\frac{2d}{d+1}} := \Big(\sum_{S\subseteq[n]} |\widehat{f}(S)|^{\frac{2d}{d+1}}\Big)^{\frac{d+1}{2d}} \le C_d \|f\|_\infty\,.$$

Moreover, there exists a universal constant $c > 0$ such that $C_d \le c^{\sqrt{d\log d}}$.

For our purposes the critical feature of Theorem 1 is the dimension-free-ness of the estimate. Bohnenblust–Hille (BH) inequalities were originally proved for analytic polynomials over the polytorus $\mathbb{T}^n = \{z \in \mathbb{C} : |z| = 1\}^n$ and have a long history in harmonic analysis [9].

One might ask if a similar approach to low-degree learning could work in the quantum world. Quantum observables (Hermitian operators) on a system of n qubits admit a "Fourier-like" decomposition

$$\mathcal{A} = \sum_{\alpha\in\{0,1,2,3\}^n} \widehat{\mathcal{A}}(\alpha)\sigma_\alpha \quad \text{where} \quad \sigma_\alpha = \bigotimes_{i=1}^n \sigma_{\alpha_i} \quad \text{and} \quad \widehat{\mathcal{A}}(\alpha) = 2^{-n}\,\mathrm{tr}[\mathcal{A}\cdot\sigma_\alpha]\,.$$

Here σ_0 is the 2-by-2 identity matrix and $\sigma_i, 1 \leq i \leq 3$ are the Pauli matrices

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

Defining $|\alpha|$ to be the number of nonzero entries in α, we say $\mathcal{A}$ is of degree d if for all α with $|\alpha| > d$ we have $\widehat{\mathcal{A}}(\alpha) = 0$. It was recently identified [20, 13, 23] that this analogy is close-enough to the Boolean case that the Eskenazis–Ivanisvili approach goes through for quantum observables as well, provided a BH-type inequality for Pauli decompositions exists, which was also proved:

▶ **Theorem 2** (Qubit Bohnenblust–Hille [13, 23]). *Suppose that $d \geq 1$. Then there exists a constant $C_d > 0$ such that for all $n \geq 1$ and all $\mathcal{A}$ on n qubits of degree d, we have*

$$\|\widehat{\mathcal{A}}\|_{\frac{2d}{d+1}} := \Big(\sum_{\alpha \in \{0,1,2,3\}^n} |\widehat{\mathcal{A}}(\alpha)|^{\frac{2d}{d+1}} \Big)^{\frac{d+1}{2d}} \leq C_d \|\mathcal{A}\|_{\mathrm{op}},$$

where $\|\mathcal{A}\|_{\mathrm{op}}$ denotes the operator norm.

An additional feature of low-degree learning specific to the quantum setting shown by [13] is that when averaging over certain distributions of input states, low-degree truncations are good approximations of arbitrary quantum observables – even ones corresponding to exponential-time quantum computations. This means the low-degree learning algorithm can perform well in predicting arbitrarily-complex quantum processes with respect to these distributions, a phenomenon that stands in stark contrast to the classical case.

1.2 From $\mathbb{Z}_2^n$ to $\mathbb{Z}_K^n$ and from qubits to qudits

It is natural to ask whether these algorithms rely on special properties of the hypercube or qubit systems, or whether they extend to larger spaces. In this paper we provide an affirmative answer, extending these classical and quantum learning results to (tensor-)product spaces of arbitrary local size $K \geq 2$.

Classically, this extension works as follows. We shall consider complex-valued functions $f : \mathbb{Z}_K^n \to \mathbb{C}$, where $\mathbb{Z}_K = \{0, 1 \ldots, K-1\}$ is the cyclic group of order K. Then each f has a unique Fourier expansion:

$$f(x) = \sum_{\alpha \in \{0,1,\ldots,K-1\}^n} \widehat{f}(\alpha) \prod_{j=1}^{n} \omega^{\alpha_j x_j} \quad \text{with} \quad \omega := e^{\frac{2\pi i}{K}},$$

where $|\alpha| := \sum_j \alpha_j$. We say f is *of degree d* if $\widehat{f}(\alpha) = 0$ for all α with $|\alpha| > d$. Ultimately, we obtain the following algorithm.

▶ **Theorem 3** (Cyclic Low-degree Learning). *Let $f : \mathbb{Z}_K^n \to \mathbb{D}$ be of degree d. Then with $(\log K)^{\mathcal{O}(d^2)} \log(n/\delta) \varepsilon^{-d-1}$ independent random samples $(x, f(x))$, $x \sim \mathcal{U}(\mathbb{Z}_K^n)$, we may with confidence $1 - \delta$ construct in polynomial time a function $\widetilde{f} : \mathbb{Z}_K^n \to \mathbb{C}$ with $\|f - \widetilde{f}\|_2^2 \leq \varepsilon$.*

(Compare with the naive Fourier sampling algorithm which would require poly(n) samples) Here the L_2 norm $\|\cdot\|_2$ is defined with respect to the uniform probability measure on Ω_K^n. Our efforts in this direction are in the theme of generalizing Analysis of Boolean Functions results to more-general product spaces, *e.g.*, [4].

It could be argued that the quantum case of generalized low0-degree learning is even more important, both for the study of fundamental physics via quantum simulation (*e.g.*, [17, 11]) 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 [11] – and this benefit may very well remain as quantum computing advances. Such systems are called *multilevel system*, or *qudit* quantum computers [24]. While the qu*b*it case gives a conceptual sense of the possibilities for learning on qudit systems, it is practically important to derive guarantees and algorithms that work directly in the native dimension of the quantum system. In so doing we also establish new distributions under which arbitrary quantum processes are well-approximated by low-degree ones (including new distributions for qubit systems beyond those identified in [13]).

▶ **Theorem 4** (Qudit Observable Learning, Informal)**.** *Let $\mathcal{A}$ be any (not necessarily low-degree) bounded quantum observable on $\mathcal{H}_K^{\otimes n}$;* i.e., *on n-many K-level qudits. Then we may via random sampling construct in polynomial time an approximate observable $\widetilde{\mathcal{A}}$ such that for a wide class of distributions μ on states ρ,*

$$\mathop{\mathbb{E}}_{\rho\sim\mu} \left|\mathrm{tr}[\mathcal{A}\rho] - \mathrm{tr}[\widetilde{\mathcal{A}}\rho]\right|^2 \leq \varepsilon\,.$$

The samples are of the form $(\rho, \mathrm{tr}[\mathcal{A}\rho])$ for ρ drawn from the uniform distribution over a certain set of product states. Moreover, the number of samples s required to achieve the guarantee with confidence $1-\delta$ is

$$s \leq \mathcal{O}\Big(\log(\tfrac{n}{\delta})\, C^{\log^2(1/\epsilon)} K^{3/2} \|\mathcal{A}^{\leq t}\|_{\mathrm{op}}^{2t}\Big)\,.$$

The quantity $\|\mathcal{A}^{\leq t}\|_{\mathrm{op}}$ is the operator norm of a degree-t truncation of $\mathcal{A}$, for t roughly $\log(1/\epsilon)$.

There is little prior work on learning qudit observables, but earlier polytime algorithms for learning qubits required at least $\mathcal{O}(n\log(n))$ samples to complete a comparable task [14]. Theorem 4 is proved by combining a low-degree learning algorithm for qudits with a qudit extension of the low-degree approximation lemma of Huang, Chen and Preskill [13, Lemma 14]. The distributions μ admitting this construction are studied in Section 4.2.2.

1.3 Proof ideas: a new Remez-type inequality

The extensions described above essentially amount to proving new Bohnenblust–Hille type inequalities in the associated spaces. Here we briefly describe how these are obtained and introduce our Remez inequality.

There are multiple natural generalizations of the qubit BH inequality (Theorem 2), and we pursue results for operator expansions in both the Heisenberg–Weyl and the Gell-Mann bases (these choices are explained and defined in Section 3). We take inspiration from the technique of [23] to reduce these noncommutative BH inequalities to BH inequalities over classical, commutative spaces. In the case of the Gell-Mann basis, we are able to reduce to the Hypercube BH, so we are done. However, in the very natural and useful Heisenberg–Weyl basis (composed of clock & shift operators), the eigenvalues of the corresponding matrices are the K^{th} roots of unity. Therefore it is natural to reduce to a BH inequality over products of cyclic groups – the same inequality needed for classical cyclic low-degree learning.

So in this way an important version of the qudit BH inequality dovetails with the BH inequality for cyclic groups and hypergrid learning. The cyclic group BH inequality appears to be previously unstudied, despite it being the interpolating case between the hypercube case ($K=2$) and the original polytorus case ($K=\infty$). One quickly discovers why, however: a proof by the "standard recipe" for BH inequalities (*à la* [8, 2, 7, 13]) will not work here.

Let us sketch the difficulty. At a very coarse level, BH inequalities for degree-d polynomials on some product space X^n are typically proved in these steps [9]:

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 for multi-linear 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 Khinchine's inequality.
3. Polarization: estimate the supremum 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. In both the polytorus and hypercube cases, one uses Markov–Bernstein-type inequalities to obtain the intermediate inequality

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

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 ($\|f\|_{\mathbb{D}^n} = \|f\|_{\mathbb{T}^n}$) and for the hypercube by multilinearity ($\|f\|_{[-1,1]^n} = \|f\|_{\{-1,1\}^n}$). But there is no such easy fact in the setting of the multiplicative cyclic group $\Omega_K := \{e^{2\pi ik/K} : k = 0, \dots, K-1\}$ 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 $\|f\|_{\mathrm{conv}(\Omega_K)^n} > \|f\|_{\Omega_K^n}$.

Indeed, it was not at all clear that $\|f\|_{\Omega_K^n}$ should provide any reasonable control over $\|f\|_{\mathrm{conv}(\Omega_K)^n}$, let alone a bound with constant independent of dimension. As a resolution, we here provide a new way to relate the supremum norm of a polynomial f over Ω_K^n to its supremum norm over $\mathbb{T}^n$.

▶ **Theorem 5.** *Fix $K \geq 2$. Let f be an n-variate degree-d polynomial of individual degree at most $K-1$. Then*

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

This can be seen as a sort of maximum principle for Ω_K^n because we may conclude

$$\|f\|_{\mathrm{conv}(\Omega_K)^n} \leq \|f\|_{\mathbb{D}^n} = \|f\|_{\mathbb{T}^n} \lesssim \|f\|_{\Omega_K^n}.$$

However, with Theorem 5 in hand there is no need to repeat any of the steps listed above. The original Bohnenblust–Hille inequality for the polytorus [3] states

$$\|\widehat{f}\|_{\frac{2d}{d+1}} \leq C^{\sqrt{d \log d}} \|f\|_{\mathbb{T}^n}$$

for any degree-d analytic polynomial f. So we immediately obtain the cyclic group BH:

▶ **Corollary 6** (Cyclic Group BH). *Let f be an n-variate degree-d polynomial of individual degree at most $K-1$. Then*

$$\|\widehat{f}\|_{\frac{2d}{d+1}} \leq (\mathcal{O}(\log K))^{d+\sqrt{d \log d}} \|f\|_{\Omega_K^n}. \tag{1}$$

Remez-type inequalities bound the supremum of a low-degree polynomial f over some space X by the supremum of f over some subset $Y \subseteq X$. In this sense Theorem 5 is a discrete Remez-type inequality for the polytorus; moreover it appears to be the first discrete multidimensional Remez inequality with a dimension-free constant (*c.f.* [25] and references therein).

We also remark that Theorem 5 may be improved in certain regimes (when $d \ll K$, or when K is a very composite integer), as well as readily extended to $L_p \to L_p$ comparisons for general p and to spaces with less structure than Ω_K^n. These extensions are to appear elsewhere. We believe Theorem 5 could be of significant general interest, as so much is already known about polynomials over $\mathbb{T}^n$. Theorem 5 provides a bridge from discrete spaces back into classical harmonic analysis.

1.4 Organization

Section 2 is a self-contained proof of the the dimension-free Remez Inequality. In section 3 we then obtain our qudit Bohnenblust–Hille inequalities. In Section 4 we use these results and a slightly generalized Eskenazis–Ivanisvili to give the learning algorithms of Theorem 4 and Theorem 3. In Section 4 we also study the probability distributions of qudit states that allow for accurate low-degree approximations of arbitrary quantum operators.

2 A dimension-free Remez inequality

Let $\mathbb{T} := \{z \in \mathbb{C} : |z| = 1\}$ and denote by $\|f\|_X$ the sup norm of any function $f : X \to \mathbb{C}$. In this section we prove the key technical result of this work. We remark that two proofs of Theorem 5 are actually known; a very different argument was given by three of the authors in [22]. While the proof in [22] is interesting for its own reasons, the argument below gives a better constant which is important for learning applications.

A natural approach to proving Theorem 5 is to consider a specific maximizer $z \in \mathbb{T}^n$ of $|f|$ and approximate it by a linear combination of evaluations of f at points in Ω_K^n. We might begin with this lemma for a single coordinate:

▶ **Lemma 7.** *Suppose $z \in \mathbb{T}$. Then there exists $\boldsymbol{c} := (c_0, \ldots, c_{K-1})$ such that for all $k = 0, 1, \ldots, K-1$,*

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

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

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

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

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

$$z^k = \sum_{j=0}^{K-1} c_j \omega^{jk}. \tag{2}$$

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,

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

In a single coordinate, Lemma 7 provides the desired inequality as follows. With $z \in \mathbb{T}$ a maximizer of $|f(z)|$ we have

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

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 8.** *Let f be a degree-d n-variate polynomial and $\boldsymbol{z} \in \mathbb{T}^n$. Then there is a univariate polynomial $p = p_{f,\boldsymbol{z}}$ 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*

$$f(\boldsymbol{z}) = D^m \mathop{\mathbb{E}}_{R,\boldsymbol{W}}[Rf(\boldsymbol{W})] + p(1/m)\,. \tag{3}$$

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

Lemma 8 is the crux of our argument and we are not aware of a similar statement in the literature. Theorem 5 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 8. We will argue Lemma 8 for $f(\boldsymbol{z}) = \boldsymbol{z}^\alpha$, 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 7 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 7. We may write a decomposition

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

with all $c_j^{(s)} \in \mathbb{R}^{\geq 0}$ 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 7,

$$\|\boldsymbol{c}^{(s)}\|_1 \leq C \tag{4}$$

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 7 in probabilistic form.

Put $D = 4C + 1$ and define $r : [0, D] \to \mathbb{C}$ 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 (4). 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 9. Let T be sampled uniformly from $[0, D]$. Then for all $k = 0, 1, \ldots, K-1$,

$$z^k = D \mathop{\mathbb{E}}_T [r(T) w(T)^k] \,. \tag{5}$$

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

$$D \mathop{\mathbb{E}}_T [r(T)] = 1 \,. \tag{6}$$

This can be seen by direct computation:

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

For $k \geq 1$, consider the joint distribution of $(r(T), w(T)^k)$, whose product appears in (5). 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). ◁

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

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

Instead, we consider only m independent copies of T: $T_1, \ldots, T_m \overset{\text{iid}}{\sim} \mathcal{U}[0, D]$. 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 $\boldsymbol{W}$:

$$R := \prod_{\ell=1}^m R_\ell \quad \text{with} \quad R_\ell := r(T_\ell), 1 \leq \ell \leq m$$
$$\boldsymbol{W} := \Big(w_1\big(T_{P(1)}\big), w_2\big(T_{P(2)}\big), \ldots, w_n\big(T_{P(n)}\big) \Big) =: (W_1, \ldots, W_n) \,.$$

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

▷ Claim 10. Consider $m \geq |\operatorname{supp}(\alpha)|$. Then

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

Proof. 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:

$$\mathop{\mathbb{E}}_{R, \boldsymbol{W}} [R \cdot \boldsymbol{W}^\alpha \mid P = \widetilde{P}] = \textstyle\prod_{\ell=1}^m \mathbb{E}\left[R_\ell \prod_{k \in S_\ell} W_k^{\alpha_k} \right] \tag{7}$$

Because $\widetilde{P}$ is injective on $\operatorname{supp}(\alpha)$, every S_ℓ contains one or zero elements of $\operatorname{supp}(\alpha)$. By Claim 9, 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

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

for the specific j for which $\{j\} = S_\ell \cap \operatorname{supp}(\alpha)$. Substituting these observations into (7) 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 11. For each partition $\mathcal{S}$ of $\operatorname{supp}(\alpha)$ 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. 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, $\widetilde{P}(\operatorname{supp}(\alpha))$), and those that only R depends on. Call these sets $L = \widetilde{P}(\operatorname{supp}(\alpha))$ and L^c respectively. Then by independence of the T_ℓ's,

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

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 $(*)$. ◁

To summarize claims 10 and 11 , we have that for all partitions $\mathcal{S}$ of $\operatorname{supp}(\alpha)$ 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 $\big\{\{j\}\big\}_{j\in\operatorname{supp}(\alpha)}$, we additionally have $E(\mathcal{S}^*) = \boldsymbol{z}^\alpha$.

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

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

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,

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

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

$$\begin{aligned}\mathop{\mathbb{E}}_{R,\boldsymbol{W}}[R \cdot \boldsymbol{W}^\alpha] &= \sum_{\mathcal{S}} \mathbb{E}[R \cdot \boldsymbol{W}^\alpha \mid P \text{ induces } \mathcal{S}]\Pr[P \text{ induces } \mathcal{S}] \\ &= D^{-m} E(\mathcal{S}^*)\Big(1 + q_{|\operatorname{supp}(\alpha)|}\big(\tfrac{1}{m}\big)\Big) + \sum_{\mathcal{S}, |\mathcal{S}| < |\operatorname{supp}(\alpha)|} D^{-m} E(\mathcal{S}) \cdot q_{|\mathcal{S}|}\big(\tfrac{1}{m}\big) \\ &= D^{-m}\Big[\boldsymbol{z}^\alpha + \sum_{\mathcal{S}} E(\mathcal{S}) \cdot q_{|\mathcal{S}|}\big(\tfrac{1}{m}\big)\Big]\,. \end{aligned} \tag{8}$$

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

$$\begin{aligned}
\mathop{\mathbb{E}}_{R,\boldsymbol{W}}[R \cdot \boldsymbol{W}^\alpha] &= 0 + \sum_{\mathcal{S},|\mathcal{S}|\le m} \mathbb{E}[R \cdot \boldsymbol{W}^\alpha \mid P \text{ induces } \mathcal{S}] \Pr[P \text{ induces } \mathcal{S}] \\
&= \sum_{\mathcal{S},|\mathcal{S}|>m} D^{-m} E(\mathcal{S}) \Pr[P \text{ induces } \mathcal{S}] + \sum_{\mathcal{S},|\mathcal{S}|\le m} D^{-m} E(\mathcal{S}) \Pr[P \text{ induces } \mathcal{S}] \\
&= D^{-m} E(\mathcal{S}^*) \Big(1 + q_{|\operatorname{supp}(\alpha)|}\big(\tfrac{1}{m}\big)\Big) \\
&\qquad + \sum_{|\operatorname{supp}(\alpha)|>|\mathcal{S}|>m} D^{-m} E(\mathcal{S}) \cdot q_{|\mathcal{S}|}\big(\tfrac{1}{m}\big) + \sum_{m\ge|\mathcal{S}|} D^{-m} E(\mathcal{S}) \cdot q_{|\mathcal{S}|}\big(\tfrac{1}{m}\big) \\
&= D^{-m}\Big[\boldsymbol{z}^\alpha + \sum_{\mathcal{S}} E(\mathcal{S}) \cdot q_{|\mathcal{S}|}\big(\tfrac{1}{m}\big)\Big]. \qquad (9)
\end{aligned}$$

Noting that (9) and (8) are identical, we rearrange to find

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

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

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

Proof of Theorem 5. Suppose there were some coefficients $a_m \in \mathbb{C}$ 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 (3) for $m = 1, \dots, d$, weighted by a_m, and get

$$f(z) = \sum_{m=1}^d a_m f(z) = \sum_{m=1}^d a_m D^m \mathbb{E}[R_m f(W_m)] + \sum_{m=1}^d a_m p\big(\tfrac{1}{m}\big) = \sum_{m=1}^d a_m D^m \mathbb{E}[R_m f(W_m)],$$

where R_m, W_m are those $R, \boldsymbol{W}$ from (3) 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 $D \ge 11$ thus $1/(1-1/D) \le 1.1$, so we conclude

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

3 Qudit Bohnenblust–Hille inequalities

Let

$$f(z) = \sum_{\alpha} c_\alpha z^\alpha = \sum_{\alpha} c_\alpha z_1^{\alpha_1} \cdots z_n^{\alpha_n},$$

where $\alpha = (\alpha_1, \ldots, \alpha_n)$ are vectors of non-negative integers, all c_α are nonzero, and the total degree of polynomial f is $d = \max_\alpha(\alpha_1 + \cdots + \alpha_n)$. Here z can be all complex vectors in $\mathbb{T}^n = \{\zeta \in \mathbb{C} : |\zeta| = 1\}^n$ or all sequences of ± 1 in Boolean cube $\{-1, 1\}^n$. Bohnenblust–Hille type of inequalities are the following

$$\Big(\sum_{\alpha} |c_\alpha|^{\frac{2d}{d+1}}\Big)^{\frac{d+1}{2d}} \leq C(d) \sup_z |f(z)|\,. \tag{10}$$

The supremum is taken either over the torus $\mathbb{T}^n$ or, more recently, the Boolean cube $\{-1, 1\}^n$. In both cases this inequality is proven with constant $C(d)$ that is independent of the dimension n and sub-exponential in the degree d. More precisely, denote by $\mathrm{BH}_{\mathbb{T}}^{\leq d}$ and $\mathrm{BH}_{\{\pm 1\}}^{\leq d}$ the best constants in the Bohnenblust–Hille inequalities (10) for degree-d polynomials on $\mathbb{T}^n$ and $\{-1, 1\}^n$, respectively. Then both $\mathrm{BH}_{\mathbb{T}}^{\leq d}$ and $\mathrm{BH}_{\{\pm 1\}}^{\leq d}$ are bounded from above by $e^{c\sqrt{d \log d}}$ for some universal $c > 0$ [2, 8].

One of the key features of this inequality (10) is the dimension-freeness of $C(d)$. This, together with its sub-exponential growth phenomenon in d, plays an important role in resolving some open problems in functional analysis and harmonic analysis [7, 2, 6]. The optimal dependence of $\mathrm{BH}_{\mathbb{T}}^{\leq d}$ and $\mathrm{BH}_{\{\pm 1\}}^{\leq d}$ on the degree d remains open.

The qubit BH inequality, Theorem 2, has received two very different proofs. In [13] Huang, Chen and Preskill pursue a direct proof and notably develop a physically-motivated "algorithmic" procedure to prove the key step in BH-type arguments known as *polarization*. They achieve the dimension-free constant $C_d \leq \mathcal{O}(d^d)$. Another proof approach appears in [23], which works by reducing the qubit BH inequality to the hypercube BH inequality. Let $\mathrm{BH}_{M_2}^{\leq d}$ denote the optimal constant in Theorem 2 (where M_2 designates the 2-by-2 complex matrix algebra). Then [23] showed $\mathrm{BH}_{M_2}^{\leq d} \leq 3^d \mathrm{BH}_{\{\pm 1\}}^{\leq d} \leq C^{\mathcal{O}(d)}$.

Pauli matrices are very special objects, being Hermitian, unitary, and anticommuting, and it was unclear whether the reduction approach in [23] could be extended to the qudit setting, where higher-dimensional generalizations of Pauli matrices are not so well-behaved. In fact we succeed in extending the reduction argument to two bases for the complex matrix algebra $M_K(\mathbb{C})$ (tensors of which form the appropriate space for qudit systems) known as the (generalized) *Gell-Mann basis* and the *Heisenberg–Weyl basis* with the view to reduce to scalar BH inequalities. They are orthonormal with respect to the normalized trace inner product $\frac{1}{K}\mathrm{tr}[A^\dagger B]$, and are respectively Hermitian and unitary generalizations of the 2-dimensional Pauli basis. Our proofs of these extensions reveal some pleasing features of the geometry of the eigenvalues of GM and HW matrices.

▶ **Definition 12** (Gell-Mann Basis). *Let $K \geq 2$ and put $E_{jk} = |e_j\rangle\langle e_k|\,, 1 \leq j, k \leq K$. The generalized Gell-Mann Matrices are a basis of $M_K(\mathbb{C})$ and are comprised of the identity matrix $\mathbf{I}$ along with the following generalizations of the Pauli matrices:*

$$\begin{aligned}
&\textit{symmetric:} && \mathbf{A}_{jk} = \sqrt{\tfrac{K}{2}}\big(E_{jk} + E_{kj}\big) && \textit{for } 1 \leq j < k \leq K \\
&\textit{antisymmetric:} && \mathbf{B}_{jk} = \sqrt{\tfrac{K}{2}}\big(-iE_{jk} + iE_{kj}\big) && \textit{for } 1 \leq j < k \leq K \\
&\textit{diagonal:} && \mathbf{C}_m = \Gamma_m\big(\textstyle\sum_{k=1}^{m} E_{kk} - mE_{m+1,m+1}\big) && \textit{for } 1 \leq m \leq K-1,
\end{aligned}$$

where $\Gamma_m := \sqrt{\frac{K}{m^2+m}}$. *We denote*

$$\mathrm{GM(K)} := \{\mathbf{I}, \mathbf{A}_{jk}, \mathbf{B}_{jk}, \mathbf{C}_m\}_{1\le j<k\le K, 1\le m\le K-1}\,.$$

An observable $\mathcal{A}$ 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 d if $\widehat{\mathcal{A}}(\alpha) = 0$ for all α with $|\alpha| > d$.

We find the Gell-Mann BH inequality enjoys a reduction to the hypercube BH inequality on $\{-1,1\}^{n(K^2-1)}$ and obtain the following.

▶ **Theorem 13** (Qudit Bohnenblust–Hille, Gell-Mann Basis). *Fix any $K \ge 2$ and $d \ge 1$. There exists $C(d,K) > 0$ such that for all $n \ge 1$ and GM observable $\mathcal{A} \in M_K(\mathbb{C})^{\otimes n}$ of degree d, we have*

$$\|\widehat{\mathcal{A}}\|_{\frac{2d}{d+1}} \le C(d,K)\|\mathcal{A}\|_{\mathrm{op}}. \tag{11}$$

Moreover, we have $C(d,K) \le \left(\frac{3}{2}(K^2-K)\right)^d \mathrm{BH}_{\{\pm1\}}^{\le d}$.

In particular, for $K = 2$ we recover the main result of [23] exactly.

▶ **Definition 14** (Heisenberg–Weyl Basis). *Fix $K \ge 2$ and let $\omega = \exp(2\pi i/K)$. Define the K-dimensional clock and shift matrices respectively via*

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

Note that $X^K = Z^K = \mathbf{I}$. See more in [1]. Then the Heisenberg–Weyl basis for $M_K(\mathbb{C})$ is

$$\mathrm{HW}(K) := \{X^\ell Z^m\}_{\ell,m\in\{0,1,\ldots,K-1\}}\,.$$

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

$$A = \sum_{\vec{\ell},\vec{m}\in\mathbb{Z}_K^n} \widehat{A}(\vec{\ell},\vec{m}) X^{\ell_1} Z^{m_1} \otimes \cdots \otimes X^{\ell_n} Z^{m_n}, \tag{12}$$

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

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

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

Unlike in the GM expansion, HW Fourier coefficients may be complex-valued. In fact, because the spectra of Heisenberg–Weyl matrices are the roots of unity, it is natural to pursue a reduction to a scalar BH inequality over $\mathbb{Z}_K$ – precisely the inequality needed for classical learning on functions on $\mathbb{Z}_K^n$. This reduction works when K is prime.

Table 1 Best known constants in Bohnenblust–Hille inequalities for (tensor-)product spaces at the time of this writing. Each BH inequality in the right half is proved via a reduction to the scalar BH inequality directly to its left. The results are accurate for all $K \geq 3$, and each appearance of C and C' is a different constant > 1. For the bounds proved in this work, no prior bounds were known.

BH Const.	Best known bound	Source	BH Const.	Best known bound	Source
$\mathrm{BH}^{\leq d}_{\{\pm 1\}}$	$C^{\sqrt{d \log d}}$	[8]	$\mathrm{BH}^{\leq d}_{M_2}$	$3^d \mathrm{BH}^{\leq d}_{\{\pm 1\}} \leq C^d$	[23]
			$\mathrm{BH}^{\leq d}_{\mathrm{GM}(K)}$	$\left(\frac{3}{2}(K^2-K)\right)^d \mathrm{BH}^{\leq d}_{\{\pm 1\}} \leq K^{Cd}$	Thm. 13
$\mathrm{BH}^{\leq d}_{\mathbb{Z}_K}$	$(C \log K)^{d+\sqrt{d \log d}}$	Cor. 6	$\mathrm{BH}^{\leq d}_{\mathrm{HW}(K)}$	$(K+1)^d \cdot \mathrm{BH}^{\leq d}_{\mathbb{Z}_K} \leq (CK \log K)^{C' d}$ *	Thm. 15
$\mathrm{BH}^{\leq d}_{\mathbb{T}}$	$C^{\sqrt{d \log d}}$	[2]	* : For K prime.		

▶ **Theorem 15** (Qudit Bohnenblust–Hille, Heisenberg–Weyl Basis). *Fix a prime number $K \geq 2$ and suppose $d \geq 1$. Consider an observable $A \in M_K(\mathbb{C})^{\otimes n}$ of degree d. Then we have*

$$\|\widehat{A}\|_{\frac{2d}{d+1}} \leq C(d,K)\|A\|_{\mathrm{op}}, \tag{13}$$

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

A summary of the Bohnenblust–Hille inequalities proved in this paper is provided in Table 1, where we denote the best constants in Eqs. (1), (11), and (13) respectively by $\mathrm{BH}^{\leq d}_{\mathbb{Z}_K}$, $\mathrm{BH}^{\leq d}_{\mathrm{GM}(K)}$, and $\mathrm{BH}^{\leq d}_{\mathrm{HW}(K)}$.

3.1 Qudit Bohnenblust–Hille in the Gell-Mann basis

In this section we prove Theorem 13 by reducing (11) to the hypercube Bohnenblust–Hille inequality on $\{-1,1\}^{n(K^2-1)}$. (The $K = 2$ case was done in [23]).

The central part of the reduction is a coordinate-wise construction of density matrices $\rho(\boldsymbol{x}) \in M_K(\mathbb{C})$ parametrized by $\boldsymbol{x} \in \{-1,1\}^{K^2-1} =: H_K$. It will be convenient to partition the coordinates of $\boldsymbol{x}$ as $\boldsymbol{x} = (x,y,z) \in \{-1,1\}^{\binom{K}{2}} \times \{-1,1\}^{\binom{K}{2}} \times \{-1,1\}^{K-1}$ with indices

$$x = (x_{jk})_{1 \leq j < k \leq K}, \qquad y = (y_{jk})_{1 \leq j < k \leq K}, \qquad \text{and} \qquad z = (z_m)_{1 \leq m \leq K-1}\,.$$

▶ **Lemma 16.** *For any $(x,y,z) \in H_K$, there exists a positive semi-definite Hermitian matrix $\rho = \rho(x,y,z)$ with $\mathrm{tr}[\rho] = 3\binom{K}{2}$ such that for all $1 \leq j < k \leq K$ and $1 \leq m \leq K-1$,*

$$\mathrm{tr}[\mathbf{A}_{jk}\rho(x,y,z)] = \sqrt{\tfrac{K}{2}}x_{jk}, \tag{14}$$

$$\mathrm{tr}[\mathbf{B}_{jk}\rho(x,y,z)] = \sqrt{\tfrac{K}{2}}y_{jk}, \tag{15}$$

$$\mathrm{tr}[\mathbf{C}_{m}\rho(x,y,z)] = \sqrt{\tfrac{K}{2}}z_{m}. \tag{16}$$

In comparison to the construction for Pauli matrices in [23], an added difficulty here that anticommutativity does not always hold among the $\mathbf{A}$'s, $\mathbf{B}$'s, and $\mathbf{C}$'s, which was the central property to achieve the equivalent of Lemma 16 for the qubit case. However, with the right construction, we still obtain the right cancellations; please see the full version [16] for a proof of Lemma 16 and Theorem 13.

3.2 Qudit Bohnenblust–Hille in the Heisenberg–Weyl basis

Here we give a proof of Theorem 15 by reduction to the Cyclic BH inequality. We collect first a few facts about X and Z.

▶ **Lemma 17.** *We have the following:*

1. $\{X^\ell Z^m : \ell, m \in \mathbb{Z}_K\}$ *form a basis of* $M_K(\mathbb{C})$.
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 $\ell_1, \ell_2, m_1, m_2 \in \mathbb{Z}_K$*:*

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

3. *If* K *is prime, then for any* $(0,0) \neq (\ell, m) \in \mathbb{Z}_K \times \mathbb{Z}_K$*, the eigenvalues of* $X^\ell Z^m$ *are* $\{1, \omega, \ldots, \omega^{K-1}\}$*. This is not the case if* K *is not prime.*

Proof. Considering each statement one-by-one:

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. Assume K to be prime and $(\ell, m) \neq (0,0)$. If $\ell = 0$ and $m \neq 0$, then the eigenvalues of Z^m are

$$\{\omega^{jm} : j \in \mathbb{Z}_K\} = \{\omega^j : j \in \mathbb{Z}_K\},$$

   since K is prime. If $\ell \neq 0$, then we may relabel the standard basis $\{e_j : j \in \mathbb{Z}_K\}$ as $\{e_{j\ell} : j \in \mathbb{Z}_K\}$. Consider the non-zero vectors

$$\zeta_k := \sum_{j \in \mathbb{Z}_K} \omega^{\frac{1}{2}j(j-1)\ell m - jk} e_{j\ell}, \qquad k \in \mathbb{Z}_K.$$

   A direct computation shows: for all $k \in \mathbb{Z}_K$

$$\begin{aligned} X^\ell Z^m \zeta_k &= \sum_{j \in \mathbb{Z}_K} \omega^{\frac{1}{2}j(j-1)\ell m - jk} \cdot \omega^{j\ell m} X^\ell e_{j\ell} \\ &= \sum_{j \in \mathbb{Z}_K} \omega^{\frac{1}{2}j(j+1)\ell m - jk} e_{(j+1)\ell} \\ &= \sum_{j \in \mathbb{Z}_K} \omega^{\frac{1}{2}j(j-1)\ell m - jk + k} e_{j\ell} \\ &= \omega^k \zeta_k. \end{aligned}$$

   If K is not prime, say $K = K_1 K_2$ with $K_1, K_2 > 1$, then X^{K_1} has 1 as eigenvalue with multiplicity $K_1 > 1$. So we do need K to be prime. ◀

Let us record the following observation as a lemma.

▶ **Lemma 18.** *Suppose that $k \geq 1$, A, B are two unitary matrices such that $B^k = \mathbf{I}$, $AB = \lambda BA$ with $\lambda \in \mathbb{C}$ and $\lambda \neq 1$. Suppose that ξ is a non-zero vector such that $B\xi = \mu\xi$ ($\mu \neq 0$ since $\mu^k = 1$). Then*

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

Here $\langle \cdot, \cdot \rangle$ denotes the inner product on $\mathbb{C}^n$ that is linear in the second argument.

Proof. By assumption

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

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

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

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

Now we are ready to prove Theorem 15:

Proof of Theorem 15. Fix a prime number $K \geq 2$. Recall that $\omega = e^{\frac{2\pi i}{K}}$. Consider the generator set of $\mathbb{Z}_K \times \mathbb{Z}_K$

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

For any $z \in \Omega_K$ and $(\ell, m) \in \Sigma_K$, we denote by $e_z^{\ell,m}$ the unit eigenvector of $X^\ell Z^m$ corresponding to the eigenvalue z. For any vector $\vec{\omega} \in (\Omega_K)^{(K+1)n}$ of the form

$$\vec{\omega} = (\vec{\omega}^{\ell,m})_{(\ell,m)\in\Sigma_K}, \qquad \vec{\omega}^{\ell,m} = (\omega_1^{\ell,m}, \ldots, \omega_n^{\ell,m}) \in (\Omega_K)^{(K+1)n}, \tag{17}$$

we consider the matrix

$$\rho(\vec{\omega}) := \rho_1(\vec{\omega}) \otimes \cdots \otimes \rho_n(\vec{\omega}) \qquad \text{where} \qquad \rho_k(\vec{\omega}) := \frac{1}{K+1} \sum_{(\ell,m)\in\Sigma_K} |e^{\ell,m}_{\omega_k^{\ell,m}}\rangle\langle e^{\ell,m}_{\omega_k^{\ell,m}}|.$$

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

Suppose that $(\ell, m) \in \Sigma_K$ and $(\ell', m') \notin \{(k\ell, km) : (\ell, m) \in \Sigma_K\}$, then by Lemma 17

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

From our choice $\omega^{\ell m' - \ell' m} \neq 1$. By Lemmas 17 and 18

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

Suppose that $(\ell, m) \in \Sigma_K$ and $1 \leq k \leq K-1$. Then by Lemma 17

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

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

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

Recall that any degree-d polynomial in $M_K(\mathbb{C})^{\otimes n}$ is a linear combination of monomials

$$A(\vec{k},\vec{\ell},\vec{m};\vec{i}) := \cdots \otimes X^{k_1\ell_1}Z^{k_1m_1} \otimes \cdots \otimes X^{k_\kappa\ell_\kappa}Z^{k_\kappa m_\kappa} \otimes \cdots$$

where

- $\vec{k} = (k_1,\ldots,k_\kappa) \in \{1,\ldots,K-1\}^\kappa$ with $0 \le \sum_{j=1}^\kappa k_j \le d$;
- $\vec{\ell} = (\ell_1,\ldots,\ell_\kappa), \vec{m} = (m_1,\ldots,m_\kappa)$ with each $(\ell_j,m_j) \in \Sigma_K$;
- $\vec{i} = (i_1,\ldots,i_\kappa)$ with $1 \le i_1 < \cdots < i_\kappa \le n$;
- $X^{k_j\ell_j}Z^{k_jm_j}$ appears in the i_j-th place, $1 \le j \le \kappa$, 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 (17) we have from the above discussion that

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

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

Now for general polynomial $A \in M_K(\mathbb{C})^{\otimes n}$ of degree-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 $c(\vec{k},\vec{\ell},\vec{m};\vec{i}) \in \mathbb{C}$ 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 different $(\vec{k},\vec{\ell},\vec{m},\vec{i})$. Therefore,

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

Using the Bohnenblust–Hille inequalities for cyclic groups (Corollary 6), we have

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

for some $C(d) > 0$. All combined, we obtain for prime K

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

4 Applications to learning

We now apply our new BH inequalities to obtain learning results for functions on $\mathbb{Z}_K^n$ and operators on qudits. In the latter case, as for qubits [13], this result may be extended to quantum observables of arbitrary complexity.

Fourier sampling is approached differently in the cyclic and qudit contexts, so we isolate the Eskenazis–Ivanisvili approximation principle from the Boolean function learning aspects of [10]. We'll also need it for vectors in $\mathbb{C}$ rather than $\mathbb{R}$ as it appeared originally but the proof is essentially identical.

▶ **Theorem 19** (Generic Eskenazis–Ivanisvili). *Let $d \in \mathbb{K}$ and $\eta, B > 0$. Suppose $v, w \in \mathbb{C}^n$ with $\|v - w\|_\infty \leq \eta$ and $\|v\|_{\frac{2d}{d+1}} \leq B$. Then for $\widetilde{w}$ defined as $\widetilde{w}_j = w_j \mathbb{1}_{[|w_j| \geq \eta(1+\sqrt{d+1})]}$ we have the bound*

$$\|\widetilde{w} - v\|_2^2 \leq (e^5 \eta^2 d B^{2d})^{\frac{1}{d+1}}.$$

See the full version [16] of this paper for a proof. 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.

4.1 Cyclic group learning

▶ **Theorem 20.** *Let $f : \mathbb{Z}_K^n \to \mathbb{D}$ be a degree-d function. Then with $(\log K)^{\mathcal{O}(d^2)} \log(n/\delta)\varepsilon^{-d-1}$ independent random samples $(x, f(x))$, $x \sim \mathcal{U}(\mathbb{Z}_K^n)$, we may with confidence $1 - \delta$ learn a function $\widetilde{f} : \mathbb{Z}_K^n \to \mathbb{C}$ with $\|f - \widetilde{f}\|_2^2 \leq \varepsilon$.*

With Theorem 19 established, one may mimic the proof approach of Eskenazis and Ivanisvili in the setting of functions on products of cyclic groups. Please see the full paper [16] for the proof.

4.2 Qudit learning

We first pursue a learning algorithm that finds a (normalized) L_2 approximation to a low-degree operator $\mathcal{A}$. Then we'll see how this extends to an algorithm finding an approximation $\widetilde{\mathcal{A}}$ with good mean-squared error over certain distributions of states for target operators $\mathcal{A}$ of any degree.

We make a couple assumptions for clarity and brevity. First, we assume the unknown observable $\mathcal{A}$ has operator norm $\|\mathcal{A}\|_{\mathrm{op}} \leq 1$. We will also assume that for a mixed state ρ the quantity $\mathrm{tr}[\mathcal{A}\rho]$ can be directly computed. Of course this is not true in practice; in the lab one must take many copies of ρ, collect observations $m_1, \ldots, m_s$ and form the estimate $\frac{1}{s}\sum_j m_j \approx \mathrm{tr}[\mathcal{A}\rho]$. The analysis required to relax these assumptions from the following results are routine so we omit them.

4.2.1 Low-degree Qudit learning

▶ **Theorem 21** (Low-degree Qudit Learning). *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*

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

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.

We elect to use the Gell-Mann basis to perform out qudit learning algorithm because, as we will see in the next section, it extends more easily to learning arbitrary qudit observables. With the Gell-Mann basis BH inequality established, very little further work is required to prove 21; see the full version [16] for details.

4.2.2 Learning arbitrary qudit observables

As observed by Huang, Chen, and Preskill in [13], there are certain distributions μ of input states ρ for which a low-degree truncation $\widetilde{\mathcal{A}}$ of any observable $\mathcal{A}$ gives a suitable approximation as measured by $\mathbb{E}_{\rho\sim\mu}\,|\mathrm{tr}[\widetilde{\mathcal{A}}\rho] - \mathrm{tr}[\mathcal{A}\rho]|^2$. This observation extends easily to qudits, which in turn ends up generalizing the phenomenon in the context of qubits as well.

▶ **Definition 22.** *For a 1-qudit unitary U let $U_j := \mathbf{I}^{\otimes j-1}\otimes U\otimes \mathbf{I}^{\otimes n-j}$. Then for a probability distribution μ on n-qudit densities and $j \in [n]$ let $\mathrm{Stab}_j(\mu)$ be the set of unitaries $U \in \mathrm{U}(K)$ such that for all n-qudit densities ρ,*

$$\mu(\rho) = \mu(U_j\rho U_j^\dagger)\,.$$

▶ Remark 23. For a set S of states, define $\mathrm{Stab}_j(S) = \{U \in \mathrm{U}(K) : U_j S U_j^\dagger \subseteq S\}$. Then it can be seen easily that $\mathrm{Stab}_j(\mu)$ is equal to the intersection of the stabilizers of the level sets of μ. That is, $\mathrm{Stab}_j(\mu) = \bigcap_{0\leq r\leq 1}\mathrm{Stab}_j\left(\mu^{-1}(r)\right)$.

Recall the definition of a unitary t-design.

▶ **Definition 24.** *Consider $P_{t,t}(U)$, a polynomial of degree at most t in the matrix elements of unitary $U \in \mathrm{U}(K)$ and of degree at most t in the matrix elements of $U^\dagger$. Then a finite subset S of the unitary group $\mathrm{U}(K)$ is a* unitary t-design *if for all such $P_{t,t}$,*

$$\frac{1}{|S|}\sum_{U\in S} P_{t,t}(U) = \mathop{\mathbb{E}}_{U\sim\mathrm{Haar}(\mathrm{U}(K))}[P_{t,t}(U)]\,.$$

We are ready to name the distributions for which low-degree truncation is possible without losing much accuracy.

▶ **Definition 25.** *Call a distribution μ on n-qudit densities* locally 2-design invariant *(L2DI) if for all $j \in [n]$, $\mathrm{Stab}_j(\mu)$ contains a unitary 2-design.*

Of course, the n-fold tensor product of Haar-random qudits is an L2DI distribution, but there are many other possible distributions and in general they can be highly entangled. When $K = 2$ and the 2-design leaving μ locally invariant is the single-qubit Clifford group, such distributions are termed *locally flat* in [13]. For any prime K the Clifford group on $\mathcal{H}_K$ is a 2-design [12]. Importantly, however, any distribution just on classical inputs $|x\rangle, x \in \{0, 1, \ldots, K-1\}^n$ is not L2DI, as a consequence of the following general observation:

▶ **Proposition 26.** *Suppose μ is an L2DI distribution over pure product states $\rho = \bigotimes_{j=1}^{n} |\psi_j\rangle\langle\psi_j|$. For $j \in [n]$ define $S_j = \{\rho_{\{j\}} : \rho \in \mathrm{supp}(\mu)\}$. Then $|S_j| \geq K^2$.*

Proof. By the L2DI property we have a 2-design $X \subseteq \mathrm{SU}(K)$ under which μ is locally invariant. Let ρ be a pure state in S_j and note that $P := \{U\rho U^\dagger : U \in X\} \subseteq S_j$. On the other hand, P forms a complex-projective 2-design [5]. And any complex-projective 2-design in K dimensions must have cardinality at least K^2 [21, Theorem 4]. ◀

The truncation theorem for qudits goes through much the same way as it does for locally flat qubit distributions in [13]. A proof is included in the full version of this paper [16].

▶ **Theorem 27.** *Let $\mathcal{A}$ be an operator on $\mathcal{H}_K^{\otimes n}$ and μ a probability distribution on densities. Then if μ is L2DI we have*

$$\mathop{\mathbb{E}}_{\rho\sim\mu} \mathrm{tr}[\mathcal{A}\rho]^2 \leq \sum_{\alpha} \left(\tfrac{K}{K^2-1}\right)^{|\alpha|} \widehat{\mathcal{A}}(\alpha)^2 .$$

The reader will notice a marked similarity to the Fourier-basis expression of noise stability $\mathbb{E}_{x\sim_\delta y} f(x)f(y) = \langle f, \mathrm{T}_\delta f\rangle = \sum_{S\subseteq[n]} \delta^{|S|} \widehat{f}(S)^2$ for Boolean functions (*e.g.*, [19]).

▶ **Definition 28.** *Let $\mathcal{A}$ be an operator with Gell-Mann decomposition $\mathcal{A} = \sum_\alpha \widehat{\mathcal{A}}_\alpha M_\alpha$. Then for $d \in [n]$ define its* degree d truncation *to be $\mathcal{A}^{\leq d} = \sum_{\alpha,|\alpha|\leq d} \widehat{\mathcal{A}}_\alpha M_\alpha$.*

▶ Remark 29. The choice of the GM decomposition here is essentially without loss of generality: consider any basis B for $M_K(\mathbb{C})$ containing the identity and define $\mathcal{A}_B^{\leq d}$ analogously to Definition 28, keeping the definition of degree analogous to that for the GM basis too. Then we have $\mathcal{A}_B^{\leq d} = \mathcal{A}^{\leq d}$, as can be seen easily by expanding one basis in the other.

▶ **Corollary 30.** $\mathbb{E}_{\rho\sim\mu} |\mathrm{tr}[\mathcal{A}\rho] - \mathrm{tr}[\mathcal{A}^{\leq d}\rho]|^2 \leq \left(\frac{K}{K^2-1}\right)^d \|\mathcal{A}\|_2^2$.

Proof. We apply Theorem 27 to obtain

$$\begin{aligned}\mathop{\mathbb{E}}_{\rho\sim\mu} |\mathrm{tr}[\mathcal{A}\rho] - \mathrm{tr}[\mathcal{A}^{\leq d}\rho]|^2 &= \mathop{\mathbb{E}}_{\rho\sim\mu} \mathrm{tr}[(\mathcal{A}-\mathcal{A}^{\leq d})\rho]^2 \\ &\leq \sum_{\alpha,|\alpha|>d} \left(\tfrac{K}{K^2-1}\right)^{|\alpha|} \widehat{\mathcal{A}}(\alpha)^2 \leq \left(\tfrac{K}{K^2-1}\right)^d \|\mathcal{A}\|_2^2 .\end{aligned}$$ ◀

▶ **Theorem 31.** *Let $\mathcal{A}$ be any observable on $\mathcal{H}_K^{\otimes n}$, of any degree, with $\|\mathcal{A}\|_{\mathrm{op}} \leq 1$. Fix an error threshold $\epsilon > 0$ and a failure probability $\delta > 0$ and put $t = \log_{K^2-1}(4/\epsilon)$. Then there is a set S of product states such that with a number*

$$s = \mathcal{O}\Big(K^{3/2} \log\left(\tfrac{n}{\delta}\right) e^{c\cdot\log^2(\frac{1}{\varepsilon})} \|\mathcal{A}^{\leq t}\|_{\mathrm{op}}^{2t}\Big)$$

of samples $(\rho, \mathrm{tr}[\mathcal{A}\rho])$, $\rho \sim \mathcal{U}(S)$, an approximate operator $\widetilde{\mathcal{A}}$ may be formed in time $\mathrm{poly}(n)$ *with confidence $1-\delta$ such that*

$$\mathop{\mathbb{E}}_{\rho\sim\mu} |\mathrm{tr}[\widetilde{\mathcal{A}}\rho] - \mathrm{tr}[\mathcal{A}\rho]|^2 \leq \varepsilon$$

for any L2DI distribution μ.

Proof. Choose the truncation degree $d = \log_{K^2-1}(4/\varepsilon)$. Then the triangle inequality and Corollary 30 give

$$\begin{aligned}\mathop{\mathbb{E}}_{\rho\sim\mu} |\mathrm{tr}[\widetilde{\mathcal{A}}^{\leq d}\rho] - \mathrm{tr}[\mathcal{A}\rho]|^2 &\leq 2 \mathop{\mathbb{E}}_{\rho\sim\mu} |\mathrm{tr}[\mathcal{A}^{\leq d}\rho] - \mathrm{tr}[\mathcal{A}\rho]|^2 + 2 \mathop{\mathbb{E}}_{\rho\sim\mu} |\mathrm{tr}[\widetilde{\mathcal{A}}^{\leq d}\rho] - \mathrm{tr}[\mathcal{A}^{\leq d}\rho]|^2 \\ &\leq 2\left(\tfrac{1}{K^2-1}\right)^d + 2\|\widetilde{\mathcal{A}}^{\leq d} - \mathcal{A}^{\leq d}\|_2^2 \\ &\leq \varepsilon/2 + 2\|\widetilde{\mathcal{A}}^{\leq d} - \mathcal{A}^{\leq d}\|_2^2 .\end{aligned}$$

So we need to choose a number of samples such that with confidence $1-\delta$, the low-degree qudit learning algorithm (Theorem 21) yields a $\mathcal{A}^{\leq d}$ such that $\|\widetilde{\mathcal{A}}^{\leq d} - \mathcal{A}^{\leq d}\|_2^2 \leq \varepsilon/4$. This requires no more than

$$CK^{3/2}\log\left(\frac{2en}{\delta}\right) e^{C'\log^2(4/\varepsilon)} \|\mathcal{A}^{\leq t}\|_{\mathrm{op}}^{2t}$$

samples, where $t = \log_{K^2-1}(4/\epsilon)$ and C, C' are constants > 1. ◀

This learning theorem may be of interest even in the context of qubits. In particular, for a small divisor k of n, a system of n qubits may be interpreted as n/k-many 2^k-level qudits, and there may be interesting distributions over states in this system which are only L2DI when viewed as "virtual qudits" in this way.

5 Conclusions

Our efforts to extend recent low-degree learning results to new spaces have led to discoveries in approximation theory, namely a dimension-free Remez-type inequality on the polytorus (Theorem 5). It would be nice to find more applications of this inequality.

In the quantum setting, obtaining qudit BH inequalities required understanding the relationships among eigenspaces of basis elements in the Gell-Mann and Heisenberg–Weyl bases. This is mostly complete, though it remains open what can be said for the HW basis when K is composite. And regarding the constants in the quantum BH inequalities, it is interesting to consider whether the exponential dependence of $\mathrm{BH}_{M_2}^{\leq d}$, $\mathrm{BH}_{\mathrm{GM}(K)}^{\leq d}$, and $\mathrm{BH}_{\mathrm{HW}(K)}^{\leq d}$ on d is necessary. (Recall that in the BH inequalities for the polytorus and hypercube, the best known constant is subexponential in d).

References

1 Ali Asadian, Paul Erker, Marcus Huber, and Claude Klöckl. Heisenberg-Weyl observables: Bloch vectors in phase space. *Phys. Rev. A*, 94:010301, July 2016. doi:10.1103/PhysRevA.94.010301.

2 Frédéric Bayart, Daniel Pellegrino, and Juan B. Seoane-Sepúlveda. The Bohr radius of the n-dimensional polydisk is equivalent to $(\log n)/n$. *Advances in Mathematics*, 264:726–746, 2014. doi:10.1016/j.aim.2014.07.029.

3 Henri F. Bohnenblust and Einar Hille. On the absolute convergence of dirichlet series. *Annals of Mathematics*, 32(3):600–622, 1931. URL: http://www.jstor.org/stable/1968255.

4 Jean Bourgain, Jeff Kahn, Gil Kalai, Yitzhak Katznelson, and Nathan Linial. The influence of variables in product spaces. *Israel Journal of Mathematics*, 77(1-2):55–64, February 1992. doi:10.1007/bf02808010.

5 Christoph Dankert. Efficient simulation of random quantum states and operators, 2005. doi:10.48550/arXiv.quant-ph/0512217.

6 Andreas Defant, Leonhard Frerick, Joaquim Ortega-Cerdà, Myriam Ounaïes, and Kristian Seip. The Bohnenblust-Hille inequality for homogeneous polynomials is hypercontractive. *Annals of Mathematics*, 174(1):485–497, July 2011. doi:10.4007/annals.2011.174.1.13.

7 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.

8 Andreas Defant, Mieczysław Mastyło, and Antonio Pérez. On the Fourier spectrum of functions on boolean cubes. *Mathematische Annalen*, 374(1-2):653–680, September 2018. doi:10.1007/s00208-018-1756-y.

9 Andreas Defant and Pablo Sevilla-Peris. The Bohnenblust-Hille cycle of ideas from a modern point of view. *Functiones et Approximatio Commentarii Mathematici*, 50(1):55–127, 2014. doi:10.7169/facm/2014.50.1.2.

10 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, pages 203–207, New York, NY, USA, 2022. Association for Computing Machinery. doi:10.1145/3519935.3519981.

11 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. *Phys. Rev. Lett.*, 129:160501, October 2022. doi:10.1103/PhysRevLett.129.160501.

12 David Gross, Koenraad Audenaert, and Jens Eisert. Evenly distributed unitaries: On the structure of unitary designs. *Journal of Mathematical Physics*, 48(5):052104, 2007. doi:10.1063/1.2716992.

13 Hsin-Yuan Huang, Sitan Chen, and John Preskill. Learning to predict arbitrary quantum processes. *CoRR*, October 2022. arXiv:2210.14894.

14 Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, June 2020. doi:10.1038/s41567-020-0932-7.

15 Siddharth Iyer, Anup Rao, Victor Reis, Thomas Rothvoss, and Amir Yehudayoff. Tight bounds on the Fourier growth of bounded functions on the hypercube. *Electron. Colloquium Comput. Complex.*, TR21-102, 2021. arXiv:TR21-102.

16 Ohad Klein, Joseph Slote, Alexander Volberg, and Haonan Zhang. Quantum and classical low-degree learning via a dimension-free remez inequality, 2023. arXiv:2301.01438.

17 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. *CoRR*, 2021. doi:10.48550/arXiv.2108.13357.

18 Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, fourier transform, and learnability. *J. ACM*, 40(3):607–620, July 1993. doi:10.1145/174130.174138.

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

20 Cambyse Rouzé, Melchior Wirth, and Haonan Zhang. Quantum Talagrand, KKL and Friedgut's theorems and the learnability of quantum Boolean functions. *CoRR*, September 2022. arXiv:2209.07279.

21 Andrew J. Scott. Tight informationally complete quantum measurements. *Journal of Physics A: Mathematical and General*, 39(43):13507, October 2006. doi:10.1088/0305-4470/39/43/009.

22 Joseph Slote, Alexander Volberg, and Haonan Zhang. A dimension-free remez-type inequality on the polytorus, 2023. arXiv:2305.10828.

23 Alexander Volberg and Haonan Zhang. Noncommutative Bohnenblust–Hille inequalities. *Mathematische Annalen*, pages 1–20, 2023. doi:10.1007/s00208-023-02680-0.

24 Yuchen Wang, Zixuan Hu, Barry C. Sanders, and Sabre Kais. Qudits and high-dimensional quantum computing. *Frontiers in Physics*, 8, 2020. doi:10.3389/fphy.2020.589504.

25 Y. Yomdin. Remez-type inequality for discrete sets. *Israel Journal of Mathematics*, 186(1):45–60, November 2011. doi:10.1007/s11856-011-0131-4.