Dimension-free discretizations of the uniform norm by small product sets

Lars Becker[1] · Ohad Klein[2] · Joseph Slote[3] · Alexander Volberg[4,5] · Haonan Zhang[6]

Abstract

Let f be an analytic polynomial of degree at most $K - 1$. A classical inequality of Bernstein compares the supremum norm of f over the unit circle to its supremum norm over the sampling set of the K-th roots of unity. Many extensions of this inequality exist, often understood under the umbrella of Marcinkiewicz–Zygmund-type inequalities for L^p, $1 \leq p \leq \infty$ norms. We study dimension-free extensions of these discretization inequalities in the high-dimension regime, where existing results construct sampling sets with cardinality growing with the total degree of the polynomial. In this work we show that dimension-free discretizations are possible with sampling sets whose cardinality is independent of $\deg(f)$ and is instead governed by the maximum *individual* degree of f; *i.e.*, the largest degree of f when viewed as a univariate polynomial in any coordinate. For example, we find that for n-variate analytic polynomials f of degree at most d and individual degree at most $K - 1$, $\|f\|_{L^\infty(\mathbf{D}^n)} \leq C(X)^d \|f\|_{L^\infty(X^n)}$ for any fixed X in the unit disc $\mathbf{D}$ with $|X| = K$. The dependence on d in the constant is tight for such small sampling sets, which arise naturally for example when studying polynomials of bounded degree coming from functions on products of cyclic groups. As an application we obtain a proof of the cyclic group Bohnenblust–Hille inequality with an explicit constant $\mathcal{O}(\log K)^{2d}$.

1 Introduction and motivation

Discretization inequalities control the L^p, $1 \leq p \leq \infty$ norm of functions f by their L^p norm over some finite sampling set. In this work we present discretization inequalities in a parameter regime motivated by harmonic analysis on high-dimensional discrete spaces. We sketch these motivations now; self-contained statements of our main results are in Sect. 1.1.

Exemplary among applications of harmonic analysis to discrete spaces is harmonic analysis on the hypercube $\{-1, 1\}^n$, which has led both to new proofs (and

sharpenings) of classical results [6, 12], as well as to many discoveries in combinatorics and theoretical computer science—seminal examples of which are [25, 26, 32]; see [42] for more. A guiding application for us is the recent introduction of the Bohnenblust–Hille (BH) inequality to statistical learning theory.

The classical BH inequality [13] states that for an n-variate degree-d analytic polynomial f on the polytorus $\mathbf{T}^n = \{z \in \mathbf{C} : |z| = 1\}^n$,

$$\|\widehat{f}\|_{\frac{2d}{d+1}} \le C(d)\|f\|_{\mathbf{T}^n}, \tag{1.1}$$

for a constant $C(d)$ that, crucially, is independent of dimension n. Here and throughout the norm $\|\cdot\|_X$ for a space X denotes the supremum norm, and $\widehat{f}$ denotes the sequence of coefficients of f. Equation (1.1) has a long history: it generalizes Littlewood's celebrated 4/3 inequality [38] and derives from Bohnenblust and Hille's resolution [13] of the *Bohr's strip problem* [14] on the convergence of Dirichlet series $\sum_n a_n n^{-s}$. Since then, the BH inequality has become an important tool in harmonic analysis and functional analysis, for example to study Sidon constants and the asymptotic behavior of Bohr radii [5, 21]. We remark that in these applications it is important to have good control of the BH constants. See the recent monograph [22] for more discussion and applications to infinite dimensional holomorphy, probability theory, and analytic number theory.

In view of these developments, it is natural to ask for generalizations of the BH inequality to other groups, such as the discrete hypercube. The hypercube BH inequality, which was first proved in [11] and sharpened in [23], gives the analogous estimate for any $f : \{\pm 1\}^n \to \mathbf{R}$ of degree at most d:

$$\|\widehat{f}\|_{\frac{2d}{d+1}} \lesssim_d \|f\|_{\{\pm 1\}^n}. \tag{1.2}$$

Here and in what follows, $A \lesssim_d B$ means $A \le C(d)B$ for a constant $C(d) > 0$ depending only on d. The sharp dependence on degree d of the constants in (1.1) and (1.2) are longstanding open problems; see [20, 22, 23] for more.

Surprisingly, Eskenazis and Ivanisvili [24] recently applied the hypercube BH inequality to obtain exponential improvements in a central task in statistical learning theory called *low-degree learning*: find (with high probability) an L^2 approximation to an unknown degree-d polynomial $f : \{\pm 1\}^n \to [-1, 1]$ using only the data $\{(x^{(j)}, f(x^{(j)}))\}_j$ for $x^{(j)}$'s drawn independently and uniformly from $\{-1, 1\}^n$.

Low-degree learning was introduced in the seminal work of Linial, Mansour, and Nisan [37] who gave an algorithm using $\mathcal{O}(n^d \log n)$ samples $(x^{(j)}, f(x^{(j)}))$. The best dependence on dimension n for this task remained $n^{\Omega(d)}$ until Eskenazis and Ivanisvili [24] obtained the dramatic reduction to $\mathcal{O}(\log n \cdot C^{\mathrm{poly}(d)})$ (one should think $d \ll n$). They applied the hypercube BH inequality to obtain strong Fourier concentration results for bounded low-degree functions, leading to an efficient thresholding approach for estimating f via its Fourier coefficients $\widehat{f}$. Extensions of this idea to quantum learning theory, with corresponding noncommutative (or *qubit*) variants of the BH inequality (1.2), swiftly followed [29, 48].

To expand the reach of these applications, it is necessary to extend harmonic analysis results from the hypercube to other discrete spaces. A primary generalization is

the discrete n-torus,

$$\Omega_K^n := \{e^{2\pi \mathbf{i}k/K} : k = 0, 1, \ldots, K - 1\}^n,$$

which arises naturally as the "hypergrid" in property testing [10, 19], is a key setting for studying the hardness of approximation (*e.g.*, the Plurality is Stablest Conjecture of [35], see also [41]), and is crucial for understanding operators on K-level qudits in the quantum (or noncommutative) setting [46]. Already there is a wealth of applications of harmonic analysis to Ω_K^n in combinatorics [1, 39], number theory [3], and graph theory [2], for example.

All functions $f : \Omega_K^n \to \mathbf{C}$ can be extended to analytic polynomials with individual degree (the maximum degree in any variable) at most $K - 1$ using Fourier expansion. In the spirit of generalizing the hypercube BH inequality (1.2) and learning theory applications to Ω_K^n, we shall pursue Fourier-type estimates for functions $f : \Omega_K^n \to \mathbf{C}$ whose corresponding polynomials also have *total* degree bounded by d. Note that our applications stipulate the parameter regime $K < d \ll n$.

Attempts to prove inequalities that are standard on the hypercube ($K = 2$) meet significant challenges on the discrete n-torus for larger K. For example, if one tries to prove the BH inequality on Ω_K^n (even for $K = 3$) by repeating the argument that worked for both the polytorus and the hypercube, one quickly encounters trouble, as detailed in [45, Appendix A].

We now sketch a difficulty appearing already for a simpler estimate: bounding spectral projections of low-degree polynomials. 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:

$$\|f_\ell\|_{\Omega_K^n} \le C(d, K)\|f\|_{\Omega_K^n}. \tag{1.3}$$

This sort of estimate is a typical dimension-free inequality in harmonic analysis. When the domain is the polytorus ($K = \infty$) this comparison is a trivial Cauchy estimate and bears constant 1. For the hypercube ($K = 2$), this estimate, usually attributed to Figiel [40, §14.6], comes fairly easily as well: 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 $Q(t) := f(tx_1^*, \ldots, tx_n^*)$ for $t \in [-1, 1]$ (f is extended to $[-1, 1]^n$ as a multiaffine 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 $C(d, \ell) \le (1 + \sqrt{2})^d$ [23, 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 a Cauchy estimate for holomorphic functions

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

Unfortunately, there is no simple way to relate $\|Q\|_{\mathbf{T}}$ to $\|f\|_{\Omega_3^n}$. Of course $\|Q\|_{\mathbf{T}} \leq \|f\|_{\mathbf{T}^n}$, but then it seems we would need a dimension-free comparison of the form

$$\|f\|_{\mathbf{T}^n} \lesssim_{d,K} \|f\|_{\Omega_K^n}. \tag{1.4}$$

Naive attempts to prove (1.4) lead to constants with exponential dependence on dimension n. And that is for good reason, as the inequality (1.4) would be quite strong. Not only would it give the spectral projection bound (1.3), but it would also immediately imply the Bohnenblust–Hille inequality for Ω_K^n via:

$$\|\widehat{f}\|_{\frac{2d}{d+1}} \overset{(1.1)}{\lesssim_d} \|f\|_{\mathbf{T}^n} \overset{(1.4)}{\lesssim_{d,K}} \|f\|_{\Omega_K^n}. \tag{1.5}$$

A proof of the inequality (1.4) and its generalizations are the subject of this work.

1.1 Main results: discretizations of L^∞ norms by small sets and a new interpolation formula

Our main result is a generalization of (1.4) where Ω_K is relaxed to any set of points of cardinality K with uniform spacing in the complex unit disk.

Theorem 1 *Let $n \geq 1$ and $K \geq 2$. Consider $Y_n = \prod_{j=1}^n Z_j$ for sets $Z_1, Z_2, \ldots, Z_n \subset \mathbf{D}$ such that for all $1 \leq j \leq n$ we have $|Z_j| = K$, and denote by η the minimum pairwise distance,*

$$\eta = \min_{1 \leq j \leq n} \min_{z \neq z' \in Z_j} |z - z'| > 0. \tag{1.6}$$

Then for any analytic polynomial $f : \mathbf{D}^n \to \mathbf{C}$ of degree d and individual degree $K - 1$,

$$\|f\|_{\mathbf{D}^n} \leq C(d, K) \|f\|_{Y_n}. \tag{1.7}$$

Here the constant $C(d, K) := C(d, K, \eta) = C(K, \eta)^d$, and $C(K, \eta) > 0$ depends only on K and η.

Moreover if all $Z_j = \Omega_K = \{e^{2\pi i k/K} : k = 0, 1, \ldots, K - 1\}$ then $C(d, K) \leq \left(\mathcal{O}(\log K)\right)^{2d}$, which proves (1.4).

The term *individual degree* here means the maximum degree of f viewed as a univariate polynomial in variable z_j, $1 \leq j \leq n$. That is, we say f is of degree at most d and individual degree at most $K - 1$ if it can be expressed as

$$f(z) = \sum_{\boldsymbol{\alpha} \in \{0, 1, \ldots, K-1\}^n : |\boldsymbol{\alpha}| \leq d} \widehat{f}(\boldsymbol{\alpha}) z^{\boldsymbol{\alpha}}, \qquad \widehat{f}(\boldsymbol{\alpha}) \in \mathbf{C}. \tag{1.8}$$

The monomial notation in (1.8) will be used throughout: for any multi-index $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n) \in \mathbf{Z}_{\geq 0}^n$ and point $z = (z_1, \ldots, z_n) \in \mathbf{C}^n$ we define

$$|\boldsymbol{\alpha}| := \sum_{1 \leq j \leq n} \alpha_j, \qquad \text{and} \qquad z^{\boldsymbol{\alpha}} := z_1^{\alpha_1} \cdots z_n^{\alpha_n}.$$

Note also that η is not completely independent of K; there is an upper bound on what η is possible because all points are constrained to be in **D**.

Remark 1 Theorem 1 holds in the real category as well: specialize the inequality to polynomials with real coefficients and choose $Y_n \subset [-1, 1]^n$ to be a real sampling set. See also Corollary 8.

In fact, Theorem 1 comes as an immediate consequence of a novel polynomial interpolation formula.

Theorem 2 *Let Y_n and η be as in Theorem 1. Then for all $z \in \mathbf{D}^n$, there exist coefficients $\{c_{\boldsymbol{\xi}}^{(z)}\}_{\boldsymbol{\xi} \in Y_n} \subset \mathbf{C}$ such that for any degree-d polynomial f,*

$$f(z) = \sum_{\boldsymbol{\xi} \in Y_n} c_{\boldsymbol{\xi}}^{(z)} f(\boldsymbol{\xi}) \tag{1.9}$$

with $\sum_{\boldsymbol{\xi}} |c_{\boldsymbol{\xi}}^{(z)}| \leq C(d, K)$ for a constant $C(d, K) := C(d, K, \eta) = C(K, \eta)^d$ independent of dimension n. When $Y_n = \Omega_K^n$, the explicit bound $C(d, K) \leq \big(\mathcal{O}(\log K)\big)^{2d}$ holds.

In the statement of Theorem 2 we avoid giving explicit forms of the $c_{\boldsymbol{\xi}}^{(z)}$'s as they are complicated (see (3.22)) and a significant portion of the proof is devoted to their development. For now we remark they are naturally expressed in a probabilistic language, a perspective in approximation theory that has been around since Bernstein's probabilistic proof of the Weierstrass approximation theorem [7] (*n.b.* that Bernstein's proof gives an approximation to continuous, univariate f through a probabilistic argument, whereas we find an exact interpolation formula for multivariate polynomial functions with coefficients having ℓ^1-norm free of dimension).

1.2 Theorems 1 and 2 in the context of approximation theory

Let us situate our main results in the context of discretization inequalities in approximation theory, beginning with dimension $n = 1$. Various corollaries and extensions of Theorems 1 and 2, as well as discussions of the optimality of various parameters, are deferred to Sect. 2.

For $1 < p < \infty$, the so-called *Marcinkiewicz–Zygmund inequality* [51, Chapter X, Theorem (7.5)] states that for all analytic polynomials f of degree at most $K - 1$, one has

$$C_p^{-1} \cdot \frac{1}{K} \sum_{z \in \Omega_K} |f(z)|^p \leq \int_{\mathbf{T}} |f(z)|^p dz \leq C_p \cdot \frac{1}{K} \sum_{z \in \Omega_K} |f(z)|^p. \tag{1.10}$$

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 (1.10) 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 [8, 9] and [51, Chapter X, Theorem (7.28)]). In our notation, the $p = \infty$ endpoint of (1.10) reads

$$\|f\|_{\Omega_K} \leq \|f\|_{\mathbf{T}} \leq C(K)\|f\|_{\Omega_K}. \tag{1.11}$$

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 [43, Theorem 5].

We refer to surveys [18, 33] and references therein for more historical background about norm discretizations. Bernstein-type discretization theorems also have some overlap with *discrete Remez-type inequalities*; see [15, 16, 50] for more discussion. (These are discretizations of the classical *Remez inequality*, which controls the supremum norms of bounded-degree polynomials by their absolute suprema over subsets of positive measure; see for example [44, 47] for more.)

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 [34], 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 [17]. Here Dai and Prymak resolved an important problem of Kroó [36] 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 much more relaxed, than the works above. On the one hand, the sampling set Ω_K^n is a fixed product set of small cardinality. Existing Bernstein-type estimates do not seem to apply in the parameter

---

[1] *N.b.*, in the notation of [17] it will be $C_d n^d$ where they used d for the dimension and n for the degree.

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 Sect. 2, dependence of the constant on degree d is unavoidable under these constraints. We also remark that even smaller sampling sets are achievable in our setting, though they are no longer product sets. This is explained in the next section.

2 Discussion: corollaries and aspects of optimality

We now remark on several aspects of optimality of Theorem 1 and summarize the remaining results of this paper.

Sharp degree-dependence of the constant For simplicity we will argue with $Y_n = \Omega_K^n$. Consider the univariate inequality

$$\|f\|_{\mathbf{T}} \leq C(K)\|f\|_{\Omega_K} \tag{2.1}$$

for polynomials f with degree at most $K - 1$. A Lagrange interpolation argument shows the best constant in (2.1) has $C(K) > 1$ for any $K \geq 3$ [45, 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

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

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 in (1.7).

Question 1 What is the optimal dependence on K in the constant in (1.7) of Theorem 1?

On the cardinality of the sampling set *Minimal cardinality of product sampling sets.* The cardinality of Y_n in Theorem 1 is optimal in the following sense. If the sampling sets are of the *product* form $Y_n = \prod_{1 \leq j \leq n} Z_j$ and one expects (1.7) 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 Sect. 4 we show that we may take a "small" part of $\prod_{j=1}^n Z_j$ and retain a dimension-free constant.

Theorem 3 *Let $K \geq 2$. Consider $\{Z_j \subset \mathbf{D}\}_{j \geq 1}$ a sequence of sets such that for all $j \geq 1$ we have $|Z_j| = K$ and*

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

Then for any $\varepsilon > 0$ one can find a subset Y_n of size at most $C_1(d, \epsilon)(1 + \varepsilon)^n$ 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 $K - 1$,

$$\|f\|_{\mathbf{D}^n} \leq C_2(d, K, \eta, \varepsilon) \|f\|_{Y_n}, \tag{2.2}$$

where

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

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

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

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

Sharp dependence of sampling set cardinality on dimension n. On the other 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 4 *Suppose that the uniform norm discretization* (1.7) *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 Sect. 5 for the proof of Theorem 4.

Uniform separation In Theorem 1, 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 1. 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 (1.7) is available for such $(Y_n)_{n \geq 1}$.

Proof ideas of the main result Our approach is probabilistic and we sketch it here with $Y_n = \Omega_K^n$ for simplicity. In one coordinate, polynomial interpolation admits a probabilistic interpretation of the form

$$f(z) = D \cdot \mathbf{E}[R \cdot f(W)], \tag{2.3}$$

where $D = D(K) > 1$ is a constant and R and W are correlated random variables taking values in Ω_4 and Ω_K respectively. Repeating (2.3) coordinatewise gives the identity

$$f(z) = D^n \, \mathbf{E}\left[\left(\prod_{j=1}^n R_j \right) f(W_1, \ldots, W_n) \right], \tag{2.4}$$

which immediately implies a discretization inequality of the desired form, except with exponential dependence on n. The idea is to notice that (2.4) is an expectation over n-many independent pairs of variables (R_j, W_j), while f is of bounded total degree d and thus is not very "aware" that $\left((R_1, W_1), \ldots, (R_n, W_n) \right)$ is a product distribution.

It turns out that by introducing certain correlations among the W_j's, we can reduce the power on D at the expense of picking up an error term:

$$f(z) = D^d \, \mathbf{E}\left[\left(\prod_{j=1}^d S_j \right) f(\widetilde{W}_1, \ldots, \widetilde{W}_n) \right] + \mathrm{error}_{f,z}. \tag{2.5}$$

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 $(\widetilde{W}_1, \ldots, \widetilde{W}_n)$ 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 take advantage of algebraic features of the error's relationship to the introduced correlations. Specifically, the correlation construction actually defines a family of identities similar to (2.5) of the form

$$f(z) = D^m \, \mathbf{E}\left[\left(\prod_{j=1}^m S_j^{(m)} \right) f\left(\widetilde{W}_1^{(m)}, \ldots, \widetilde{W}_n^{(m)} \right) \right] + \mathrm{error}_{f,z}\left(\frac{1}{m} \right),$$

for any integer $m > 1$, and where 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 $m = d, d + 1, \ldots, 2d - 1$ that eliminates the error term:

$$f(z) = \sum_{m=d}^{2d-1} a_m f(z) = \sum_{m=d}^{2d-1} a_m D^m \mathbf{E}\left[\left(\prod_{j=1}^{m} S_j^{(m)}\right) f\left(\widetilde{W}_1^{(m)}, \ldots, \widetilde{W}_n^{(m)}\right)\right], \quad (2.6)$$

and where the absolute sum of the a_m's is suitably small. This directly gives Theorem 1.

A new polynomial interpolation formula All the expectations in (2.6) are over finite probability spaces, so we actually have proved a new interpolation formula of the form

$$f(z) = \sum_{\boldsymbol{\xi} \in Y_n} c_{\boldsymbol{\xi}}^{(z)} f(\boldsymbol{\xi}), \quad (2.7)$$

where $\sum_{\boldsymbol{\xi}} |c_{\boldsymbol{\xi}}^{(z)}|$ is bounded independent of dimension. This is Theorem 2.

Comparing (2.7) 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 (2.7) 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 (2.7) appears in detail as Equation (3.22). 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, $1 \leq p < \infty$, as we describe next.

Dimension-free discretization for L^p norms 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 $\mathbf{T}^n$ [30] and over Ω_K^n [31, 49]. Hypercontractivity is a workhorse of high dimensional analysis [4, 28] and implies dimension-free L^2-L^p comparisons for bounded-degree polynomials when $1 \leq p < \infty$ (see [42, Chap. 9.5] and [21, 22, Chap. 8.4] for discussion). For example, with $2 \leq p < \infty$, and f a degree-d function on Ω_K^n, the argument is

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

where hypercontractivity on the polytorus is applied in the first inequality. Note, however, that such a hypercontractivity argument does not work for $p = \infty$.

In Sect. 6 we show a proof that avoids hypercontractivity altogether by making use of the interpolation formula (2.7) (or more concretely, Equation (3.22)). The main result of Sect. 6 is the following.

Theorem 5 *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 5 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 $1 < p < \infty$.

Consequences It is worthwhile to mention that our results improve various known work.

As outlined before (1.4), we may combine Theorem 1 with a standard Cauchy estimate to obtain the so-called *Figiel's inequality* for spectral projections of polynomials on Ω_K^n.

Corollary 6 *Let $f : \Omega_K^n \to \mathbf{C}$ be a degree-d polynomial with individual degree at most $K - 1$. For $0 \leq \ell \leq d$ let f_ℓ be the degree-ℓ homogeneous part of f. Then*

$$\|f_\ell\|_{\Omega_K^n} \leq \big(\mathcal{O}(\log K)\big)^{2d}\|f\|_{\Omega_K^n}.$$

The chain of inequalities (1.5) shows that Theorem 1 relates the best constants $\mathrm{BH}_{\Omega_K}^{\leq d}$, $\mathrm{BH}_{\mathbf{T}}^{\leq d}$ in the BH inequality for cyclic groups and the polytorus by

$$\mathrm{BH}_{\Omega_K}^{\leq d} \leq \big(\mathcal{O}(\log K)\big)^{2d} \cdot \mathrm{BH}_{\mathbf{T}}^{\leq d}.$$

In [5] the authors obtain the best-known bound for the BH inequality on the polytorus, showing that $\mathrm{BH}_{\mathbf{T}}^{\leq d} \leq C^{\sqrt{d \log d}}$. As a consequence the following BH inequality holds for cyclic groups.

Corollary 7 *Let $f : \Omega_K^n \to \mathbf{C}$ be a degree-d polynomial with individual degree at most $K - 1$. Then*

$$\left(\sum_{|\boldsymbol{\alpha}| \leq d} \big| \widehat{f}(\boldsymbol{\alpha}) \big|^{\frac{2d}{d+1}} \right)^{\frac{d+1}{2d}} \leq \big(\mathcal{O}(\log K)\big)^{2d}\|f\|_{\Omega_K^n}, \tag{2.8}$$

where C is a universal constant.

This improves upon the constant $C(K)^{d^2}$ previously obtained in [45] and, as an application, improves the sample complexity of various classical and quantum learning tasks.

Note that the best constant $\mathrm{BH}_{\{\pm 1\}}^{\leq d}$ in the hypercube BH is also known to be at most $C^{\sqrt{d \log d}}$ [23]. This raises a natural question about the best constant for intermediate K, $3 \leq K < \infty$.

Question 2 What is the best constant for the cyclic-group Bohnenblust–Hille, $\text{BH}_{\Omega_K}^{\leq d}$?
As a starting point, is $\text{BH}_{\Omega_K}^{\leq d}$ subexponential in d?

When specialized to real polynomials, our inequality is related to the discrete variants of Remez inequalities of [15, 50]. Theorem 1 shows that for certain grids, the constant can be dimension-free (*c.f.,* [50, Equation (2.5)]). We formulate here a real variable version for convenience, in the case of the regular grid.

Corollary 8 *Fix $d, n \geq 1$. Let $f : [-1, 1]^n \to \mathbf{R}$ be a real polynomial of degree at most d and individual degree at most $K - 1$, that is,*

$$f(x) = \sum_{\alpha \in \{0, 1, \ldots, K-1\}^n : |\alpha| \leq d} a_\alpha x^\alpha, \qquad a_\alpha \in \mathbf{R}.$$

Then we have

$$\|f\|_{[-1,1]^n} \leq C(d, K) \|f\|_{G_K^n}, \tag{2.9}$$

where G_K is the grid of K points equi-distributed on $[-1, 1]$. Here $C(d, K) > 0$ is a constant depending on d and K only.

Proof Let us use Theorem 1 with f a real polynomial of degree at most d and individual degree at most $K - 1$, and sampling set $G_K^n \subset [-1, 1]^n$. □

3 Discretizations from product sets

Here we prove Theorem 2 and obtain Theorem 1 as an immediate consequence.

3.1 Some preparations

We start with a lemma that records a standard estimate for inverses of Vandermonde matrices.

Lemma 9 *Suppose that $d \geq 1$ and $(x_0, \ldots, x_{d-1}) \in \mathbf{C}^d$ is a vector such that*

$$\max_{0 \leq j \leq d-1} |x_j| \leq M, \qquad \min_{0 \leq j < k \leq d-1} |x_j - x_k| \geq \eta \tag{3.1}$$

with $0 < \eta, M < \infty$. Let $V[x_0, \ldots, x_{d-1}] = [a_{jk}]_{j,k=0}^{d-1}$ be the d-by-d Vandermonde matrix associated to $(x_0, \ldots, x_{d-1})$ with $a_{jk} = x_k^j$. Then its inverse, $V[x_0, \ldots, x_{d-1}]^{-1} = [b_{jk}]_{j,k=0}^{d-1}$ satisfies

$$|b_{jk}| \leq \frac{M^{d-1-k} \binom{d-1}{k}}{\eta^{d-1}}, \qquad 0 \leq j, k \leq d - 1. \tag{3.2}$$

Proof Recall that

$$b_{jk} = \frac{(-1)^{d-1-k} e_{d-1-k}(\{x_0, \ldots, x_{d-1}\} \setminus \{x_j\})}{\prod_{0 \le m \le d-1: m \ne j}(x_j - x_m)},\tag{3.3}$$

where $e_k(\{y_1, \ldots, y_{d-1}\}) := \sum_{1 \le i_1 < \cdots < i_k \le d-1} y_{i_1} \cdots y_{i_k}$. Then the desired estimate follows immediately from the assumption (3.1). $\qquad\square$

Fix a family of distinct points $\{y_j\}_{j=0,\ldots,K-1} \subset \mathbf{D}$ such that

$$\min_{j \ne k} |y_j - y_k| \ge \eta.$$

We denote by $V_{\mathbf{y}}$ the K-by-K Vandermonde matrix associated to $\mathbf{y} := (y_0, \ldots, y_{K-1})$ with $V_{\mathbf{y}} = [y_k^j]_{0 \le j,k \le K-1}$. Then $V_{\mathbf{y}}$ is invertible with $V_{\mathbf{y}}^{-1} = [b_{jk}]_{0 \le j,k \le K-1}$ satisfying

$$|b_{jk}| \le \eta^{1-K} \binom{K-1}{k}, \qquad 0 \le j, k \le K - 1 \tag{3.4}$$

according to Lemma 9. So for any $x \in \mathbf{D}$, the system (putting $0^0 = 1$)

$$\sum_{k=0}^{K-1} c_k(\mathbf{y}; x) y_k^j = \sum_{k=0}^{K-1} c_k(x) y_k^j = x^j, \qquad 0 \le j \le K - 1 \tag{3.5}$$

has a unique solution

$$(c_0(x), c_1(x), \ldots, c_{K-1}(x))^T = V_{\mathbf{y}}^{-1}(1, x, \ldots, x^{K-1})^T \in \mathbf{C}^K$$

such that by (3.4)

$$\sum_{j=0}^{K-1} |c_j(x)| \le \sum_{j=0}^{K-1} \sum_{k=0}^{K-1} |b_{jk}||x|^k \le \frac{K}{\eta^{K-1}} \sum_{k=0}^{K-1} \binom{K-1}{k} = K\left(\frac{2}{\eta}\right)^{K-1} \tag{3.6}$$

uniformly in $x \in \mathbf{D}$. We remark that $c_k(x) = c_k(\mathbf{y}; x)$ depends on $\mathbf{y}$, while most of the time we only need the estimate (3.6) that is uniformly bounded for $\mathbf{y} \in \mathbf{D}^n$. So in the sequel, we omit the dependence on $\mathbf{y}$ whenever no confusion can occur.

We write each complex number $c_k(x)$ in the following form

$$c_k(x) = c_k^{(1)}(x) - c_k^{(-1)}(x) + \mathbf{i} c_k^{(\mathbf{i})}(x) - \mathbf{i} c_k^{(-\mathbf{i})}(x),$$

where $c_k^{(s)}(x) = c_k^{(s)}(\mathbf{y}; x)$ are given by

$$c_k^{(1)}(x) = (\Re c_k(x))_+, \qquad c_k^{(-1)}(x) = (\Re c_k(x))_-,$$

and

$$c_k^{(\mathbf{i})}(x) = (\Im c_k(x))_+, \qquad c_k^{(-\mathbf{i})}(x) = (\Im c_k(x))_-.$$

Note that (3.5) with $j = 0$ becomes $\sum_{k=0}^{K-1} c_k(x) = 1$, so that

$$\sum_{k=0}^{K-1} c_k^{(1)}(x) = \sum_{k=0}^{K-1} c_k^{(-1)}(x) + 1 =: c^{(\Re)}(x) = c^{(\Re)}(\mathbf{y}; x), \tag{3.7}$$

and

$$\sum_{k=0}^{K-1} c_k^{(\mathbf{i})}(x) = \sum_{k=0}^{K-1} c_k^{(-\mathbf{i})}(x) =: c^{(\Im)}(x) = c^{(\Im)}(\mathbf{y}; x). \tag{3.8}$$

Put

$$L = \max_{x \in \mathbf{D}, \, \mathbf{y} \in \mathbf{D}^n} \left\{ c^{(\Re)}(\mathbf{y}; x), c^{(\Im)}(\mathbf{y}; x) \right\} \geq 0.$$

Notice that by definition of L and (3.6)

$$L \leq \max_{x \in \mathbf{D}, \, \mathbf{y} \in \mathbf{D}^n} \sum_{k=0}^{K-1} |c_k(\mathbf{y}; x)| \leq K \left(\frac{2}{\eta} \right)^{K-1}. \tag{3.9}$$

So L is a constant depending only on K and η.

By definition of L, we can choose non-negative $t^{(s)}(x) = t^{(s)}(\mathbf{y}; x), s \in \{\pm 1, \pm \mathbf{i}\}$ such that

$$\sum_{k=0}^{K-1} c_k^{(1)}(x) + t^{(1)}(x) = L + 1, \tag{3.10}$$

$$\sum_{k=0}^{K-1} c_k^{(s)}(x) + t^{(s)}(x) = L, \quad s = -1, \mathbf{i}, -\mathbf{i}. \tag{3.11}$$

It follows immediately from (3.7), (3.8), (3.10), and (3.11) that

$$t^{(1)}(x) = t^{(-1)}(x) \qquad \text{and} \qquad t^{(\mathbf{i})}(x) = t^{(-\mathbf{i})}(x). \tag{3.12}$$

Put $D := 4L + 1$. We need certain functions on $[0, D]$. The first one is $r = r(t)$ that depends only on D. The second one is $w^{(x)}(t) = w_{\mathbf{y}}^{(x)}(t)$ that depends on $x \in \mathbf{D}$ and $\mathbf{y} \in \mathbf{D}^n$. As before, we omit the $\mathbf{y}$-dependence in the notation whenever no confusion can occur.

We first divide the interval $[0, D] = [0, 4L + 1]$ into the disjoint union of intervals

$$[0, D] = I^{(1)} \cup I^{(-1)} \cup I^{(\mathbf{i})} \cup I^{(-\mathbf{i})},$$

with lengths $|I^{(1)}| = L + 1$ and $|I^{(-1)}| = |I^{(\mathbf{i})}| = |I^{(-\mathbf{i})}| = L$. Then the function $r : [0, D] \to \{\pm 1, \pm \mathbf{i}\}$ is defined as follows:

$$r(t) = s, \qquad t \in I^{(s)}.$$

To define $w^{(x)}$ for any $x \in \mathbf{D}$, recall that (3.10) and (3.11) allow us to further decompose each $I^{(s)}$ into disjoint unions:

$$I^{(s)} = \bigcup_{k=0}^{K-1} \left(I_k^{(s)}(x) \cup J_k^{(s)}(x) \right), \qquad s = \pm 1, \pm \mathbf{i}$$

with (recall again the omitted $\mathbf{y}$-dependence so actually $I_k^{(s)}(x) = I_k^{(s)}(\mathbf{y}; x)$ and $J_k^{(s)}(x) = J_k^{(s)}(\mathbf{y}; x)$)

$$|I_k^{(s)}(x)| = c_k^{(s)}(x) \qquad \text{and} \qquad |J_k^{(s)}(x)| = \frac{t^{(s)}(x)}{K}$$

for all $0 \le k \le K-1$ and $s \in \{\pm 1, \pm \mathbf{i}\}$. Then the function $w^{(x)} = w_{\mathbf{y}}^{(x)} : [0, D] \to \{y_0, \ldots, y_{K-1}\}$ is defined as

$$w^{(x)}(t) = y_k, \qquad t \in I_k^{(s)}(x) \cup J_k^{(s)}(x)$$

for all $0 \le k \le K-1$ and $s \in \{\pm 1, \pm \mathbf{i}\}$.

Now assume that U is a random variable uniformly distributed on $[0, D]$. Then by definition

$$\mathbf{E}[r(U)] = \frac{1}{D}. \tag{3.13}$$

When U takes values in $I_k^{(s)}(x) \cup J_k^{(s)}(x)$, we have

$$r(U) = s, \qquad w^{(x)}(U) = y_k.$$

This, together with definitions of $c_k^{(s)}(x)$, $I_k^{(s)}(x)$, $J_k^{(s)}(x)$, and equations (3.5) and (3.12), implies

$$
\begin{aligned}
\mathbf{E}\left[r(U)w^{(x)}(U)^\alpha\right] &= \frac{1}{D} \sum_{s=\pm 1, \pm \mathbf{i}} \sum_{k=0}^{K-1} sy_k^\alpha \left(c_k^{(s)}(x) + \frac{t^{(s)}(x)}{K} \right) \\
&= \frac{1}{D} \sum_{k=0}^{K-1} y_k^\alpha \sum_{s=\pm 1, \pm \mathbf{i}} sc_k^{(s)}(x) + \frac{1}{D} \sum_{s=\pm 1, \pm \mathbf{i}} st^{(s)}(x) \\
&= \frac{1}{D} \sum_{k=0}^{K-1} c_k(x) y_k^\alpha + 0 \\
&= \frac{1}{D} x^\alpha,
\end{aligned}
$$

whenever $0 \le \alpha \le K-1$. We record this fact for later use: For U uniformly distributed on $[0, D]$, we just proved that

$$\mathbf{E}\left[r(U)w^{(x)}(U)^\alpha\right] = \frac{1}{D} x^\alpha, \qquad x \in \mathbf{D}, \quad 0 \le \alpha \le K-1. \tag{3.14}$$

3.2 The proof idea and difficulty

We fix $\boldsymbol{y} = (y_0, \ldots, y_{K-1}) \in \mathbf{D}^K$, $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbf{D}^n$ and follow the notations in the previous subsection. We are going to use the functions $r : [0, D] \to \{\pm 1, \pm \mathbf{i}\}$ and $w^{(x_j)} = w_{\boldsymbol{y}}^{(x_j)} : [0, D] \to \{y_0, \ldots, y_{K-1}\}$ defined above to prove that

$$|f(\boldsymbol{x})| \leq C \max_{z \in \{y_0, \ldots, y_{K-1}\}^n} |f(z)|$$

with a dimension-free constant C for certain polynomials f. The idea is based on (3.13) and (3.14). Let us fix $d \geq 1$ and start with a monomial

$$M(z) = z_{i_1}^{\alpha_{i_1}} \ldots z_{i_k}^{\alpha_{i_k}}, \qquad z = (z_1, \ldots, z_n)$$

where $i_1 < \cdots < i_k$ and $1 \leq \alpha_{i_j} \leq K - 1$. Suppose that $U_1, \ldots, U_n$ are i.i.d. random variables uniformly distributed on $[0, D]$. Then for any $\text{supp}(\boldsymbol{\alpha}) \subset J \subset [n] := \{1, \ldots, n\}$, we have by (3.13), (3.14) and the independence that

$$\mathbf{E}\left[\prod_{j \in J} r(U_j) \cdot M\left(w^{(x_1)}(U_1), \ldots, w^{(x_n)}(U_n)\right)\right]$$

$$= \prod_{j=1}^{k} \mathbf{E}\left[r(U_{i_j}) w^{(x_{i_j})}(U_{i_j})^{\alpha_j}\right] \prod_{j \in J \setminus \text{supp}(\boldsymbol{\alpha})} \mathbf{E}\left[r(U_j)\right]$$

$$= \frac{1}{D^{|J|}} x_{i_1}^{\alpha_1} \cdots x_{i_k}^{\alpha_k} = \frac{1}{D^{|J|}} M(\boldsymbol{x}).$$

Recall that $w^{(x)}$ takes values in $\{y_0, \ldots, y_{K-1}\}$ and $|r(t)| \equiv 1$, and we may then deduce

$$|M(\boldsymbol{x})| \leq D^{|J|} \max_{z \in \{y_0, \ldots, y_{K-1}\}^n} |M(z)|,$$

as desired. Note that for a single monomial of degree at most d, we may choose $J = \text{supp}(\boldsymbol{\alpha})$ and thus $|J| \leq d$. If we consider linear combinations of monomials whose supports are contained in some bounded interval J, then the above argument gives the desired estimate with constant $D^{|J|}$ by linearity in M for fixed J. However, this is no longer the case if the size of J is large enough; that is, when $|J|$ depends on n. Consider polynomials of degree at most $d = 2$ for example. The above argument works for $z_1 z_2 + z_2 z_3$ with a choice of $J = \{1, 2, 3\}$ giving a constant of D^3, but for $z_1 z_2 + z_2 z_3 + \cdots + z_{n-1} z_n$ we need to choose $J = [n]$ which results in a dimension-dependent constant D^n.

3.3 Proof of partial Theorems 1 and 2: product of many copies of a single subset

As before, we fix $\boldsymbol{y} = (y_0, \ldots, y_{K-1}) \in \mathbf{D}^K$, $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbf{D}^n$. We want to show that for each polynomial $f : \mathbf{D}^n \to \mathbf{C}$ of degree at most d and individual degree at

most $K - 1$, it holds that

$$|f(x)| \le C(d, K) \max_{z \in \{y_0, \ldots, y_{K-1}\}^n} |f(z)|.$$

As explained in the last subsection, the proof relies on the functions r, $w^{(x)}$ and we shall overcome the difficulty of unbounded $|\cup_\alpha \operatorname{supp}(\alpha)|$ mentioned there as follows. For a monomial z^α of degree at most $d \ge 1$, the length of its support $|\operatorname{supp}(\alpha)|$ is at most d as well. Fix $m \ge d$ (one may take $m = d$ for the moment). Consider some map

$$P : [n] \to [m].$$

Instead of working with polynomials in $w^{(x_1)}(U_1), \ldots, w^{(x_n)}(U_n)$, we shall consider

$$w^{(x_1)}(U_{P(1)}), \ldots, w^{(x_n)}(U_{P(n)}),$$

which may potentially resolve the problems of constants being not dimension-free. However, it may break the independence and we need to estimate the error terms. For this, we make P random as well.

Recall that $y = (y_0, \ldots, y_{K-1}) \in \mathbf{D}^K$ is fixed, with respect to which one defines $w^{(x)} = w_y^{(x)}$ function as above for any $x \in \mathbf{D}$. As before, $U_1, \ldots, U_n$ are i.i.d. random variables taking values in $[0, D]$ uniformly.

Proposition 1 *Fix $k \ge 1$. Consider*

$$M(z) = z^\alpha = z_{i_1}^{\alpha_{i_1}} \ldots z_{i_k}^{\alpha_{i_k}}$$

a monomial on $\mathbf{D}^n$ with $1 \le \alpha_{i_j} \le K - 1$ and

$$\operatorname{supp}(\alpha) = \{1 \le i_1 < \cdots < i_k \le n\}.$$

Then for any $x = (x_1, \ldots, x_n) \in \mathbf{D}^n$ there exists a polynomial $p_{\alpha,x,y}$ of degree at most $k - 1$ such that $p_{\alpha,x,y}(0) = 0$ and the following holds. For any $m \ge k$, let $P : [n] \to [m]$ be constructed by choosing for each $j \in [n]$ uniformly at random $P(j) \in [m]$, then we have

$$M(x) = D^m \mathbf{E}_{U,P} \left[\prod_{\ell=1}^m r(U_\ell) \cdot \prod_{j=1}^n w^{(x_j)}(U_{P(j)})^{\alpha_j} \right] + p_{\alpha,x,y}\left(\frac{1}{m}\right). \qquad (3.15)$$

Before proceeding with the proof, note that by the independence of U_ℓ's, for any fixed $P : [n] \to [m]$ we have

$$\mathbf{E}_U \left[\prod_{\ell=1}^m r(U_\ell) \cdot \prod_{j=1}^n w^{(x_j)}(U_{P(j)})^{\alpha_j} \right]$$

$$= \mathbf{E}_U \left[\prod_{\ell=1}^m r(U_\ell) \cdot \prod_{j \in \operatorname{supp}(\alpha)} w^{(x_j)}(U_{P(j)})^{\alpha_j} \right]$$

$$= \mathbf{E}_U\left[\prod_{\ell=1}^m\left(r(U_\ell)\prod_{j\in\text{supp}(\boldsymbol{\alpha}):P(j)=\ell} w^{(x_j)}(U_\ell)^{\alpha_j}\right)\right]$$

$$= \prod_{\ell=1}^m \mathbf{E}_U\left[r(U_\ell)\prod_{j\in\text{supp}(\boldsymbol{\alpha}):P(j)=\ell} w^{(x_j)}(U_\ell)^{\alpha_j}\right].$$

To estimate each term in the product let us consider the partition of $\text{supp}(\boldsymbol{\alpha})$. Fix a partition $\mathcal{S} = \{S_j\}$ of $\text{supp}(\boldsymbol{\alpha})$. For any $P : [n] \to [m]$ we say P *induces* $\mathcal{S}$ if

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

that is

$$\forall j, \quad |P(S_j)| = 1 \quad\text{and}\quad \forall j\neq k, \quad P(S_j)\neq P(S_k).$$

Simple combinatorics give that for any partition $\mathcal{S}$ of $\text{supp}(\boldsymbol{\alpha})$

$$\mathbf{Pr}[P : [n] \to [m] \text{ induces } \mathcal{S}] = \frac{m(m-1)\cdots(m-|\mathcal{S}|+1)}{m^{|\text{supp}(\boldsymbol{\alpha})|}}. \tag{3.16}$$

In particular, it can be represented as

$$\mathbf{Pr}[P \text{ induces } \mathcal{S}] = \begin{cases} 1 + q_{|\mathcal{S}|,|\text{supp}(\boldsymbol{\alpha})|}\left(\frac{1}{m}\right) & \text{if } |\mathcal{S}| = |\text{supp}(\boldsymbol{\alpha})| \\ q_{|\mathcal{S}|,|\text{supp}(\boldsymbol{\alpha})|}\left(\frac{1}{m}\right) & \text{if } |\mathcal{S}| < |\text{supp}(\boldsymbol{\alpha})| \end{cases} \tag{3.17}$$

for some polynomials $q = q_{|\mathcal{S}|,|\text{supp}(\boldsymbol{\alpha})|}$ with $q(0) = 0$ and $\deg(q) < |\text{supp}(\boldsymbol{\alpha})|$.

Proof of Proposition 1 As discussed above, the calculation of expectation in (3.15) depends on the partition $\mathcal{S}$ of $\text{supp}(\boldsymbol{\alpha})$. Clearly, $|\mathcal{S}| \leq |\text{supp}(\boldsymbol{\alpha})| = k$. In the special case when $|\mathcal{S}| = |\text{supp}(\boldsymbol{\alpha})|$, $\mathcal{S}$ is a singleton partition. In this case, we may write $\{j_\ell\} = P^{-1}(\ell), \ell\in[m]$ for P that induces $\mathcal{S}$. For such P we may calculate

$$\mathbf{E}_U\left[r(U_\ell)\prod_{j\in\text{supp}(\boldsymbol{\alpha}):P(j)=\ell} w^{(x_j)}(U_\ell)^{\alpha_j}\right] = \mathbf{E}_U[r(U_\ell)] = \frac{1}{D}$$

if $j_\ell \notin \text{supp}(\boldsymbol{\alpha})$, according to (3.13); and

$$\mathbf{E}_U\left[r(U_\ell)\prod_{j\in\text{supp}(\boldsymbol{\alpha}):P(j)=\ell} w^{(x_j)}(U_\ell)^{\alpha_j}\right] = \mathbf{E}_U\left[r(U_\ell)w^{(x_{j_\ell})}(U_\ell)^{\alpha_{j_\ell}}\right] = \frac{x_{j_\ell}^{\alpha_{j_\ell}}}{D}$$

if $j_\ell \in \text{supp}(\boldsymbol{\alpha})$, according to (3.14). All combined, we find

$$\mathbf{E}_U\left[\prod_{\ell=1}^m r(U_\ell)\prod_{j=1}^n w^{(x_j)}(U_{P(j)})^{\alpha_j}\right] = \prod_{\ell:j_\ell\in\text{supp}(\boldsymbol{\alpha})}\frac{x_{j_\ell}^{\alpha_{j_\ell}}}{D}\prod_{\ell:j_\ell\notin\text{supp}(\boldsymbol{\alpha})}\frac{1}{D}$$

$$= \frac{x^{\alpha}}{D^k}. \tag{3.18}$$

For general P that induces $\mathcal{S}$ with $|\mathcal{S}| < |\text{supp}(\boldsymbol{\alpha})|$, it is not easy to calculate

$$\mathbf{E}_U \left[\prod_{\ell=1}^m r(U_\ell) \prod_{j=1}^n w^{(x_j)}(U_{P(j)})^{\alpha_j} \right].$$

We claim that there exists a constant $E(\mathcal{S}, \boldsymbol{\alpha}, \boldsymbol{x}, \boldsymbol{y})$ independent of m such that

$$\mathbf{E}_U \left[\prod_{\ell=1}^m r(U_\ell) \prod_{j=1}^n w^{(x_j)}(U_{P(j)})^{\alpha_j} \right] = D^{-m} E(\mathcal{S}, \boldsymbol{\alpha}, \boldsymbol{x}, \boldsymbol{y}). \tag{3.19}$$

The exact value of $E(\mathcal{S}, \boldsymbol{\alpha}, \boldsymbol{x}, \boldsymbol{y})$ is not important; we will find a way to eliminate it afterwards. What matters to us is that it is independent from m.

To see the claim, note that for P that induces $\mathcal{S}$ with $|\mathcal{S}| < |\text{supp}(\boldsymbol{\alpha})|$,

$$\mathbf{E}_U \left[\prod_{\ell=1}^m r(U_\ell) \prod_{j=1}^n w^{(x_j)}(U_{P(j)})^{\alpha_j} \right]$$

$$= \prod_{\ell=1}^m \mathbf{E}_U \left[r(U_\ell) \prod_{j \in \text{supp}(\boldsymbol{\alpha}) \cap P^{-1}(\ell)} w^{(x_j)}(U_\ell)^{\alpha_j} \right]$$

$$= \frac{1}{D^{m-|\mathcal{S}|}} \underbrace{\prod_{\ell: \text{supp}(\boldsymbol{\alpha}) \cap P^{-1}(\ell) \neq \emptyset} \mathbf{E}_U \left[r(U_\ell) \prod_{j \in \text{supp}(\boldsymbol{\alpha}) \cap P^{-1}(\ell)} w^{(x_j)}(U_\ell)^{\alpha_j} \right]}_{(*)}$$

where we used (3.13) in the last equality. We observe that the expectation $(*)$ may depend on $(\mathcal{S}, \boldsymbol{\alpha}, \boldsymbol{x}, \boldsymbol{y})$, but does not depend on the specific P inducing $\mathcal{S}$, nor on m. Thus we may define $E(\mathcal{S}, \boldsymbol{\alpha}, \boldsymbol{x}, \boldsymbol{y})$ by setting $D^{-|\mathcal{S}|} E(\mathcal{S}, \boldsymbol{\alpha}, \boldsymbol{x}, \boldsymbol{y})$ equal to $(*)$.

According to the above discussion and (3.17), we have (we omit the constraint that $\mathcal{S}$ is a partition of $\text{supp}(\boldsymbol{\alpha})$ in the summation for notational convenience)

$$\mathbf{E}_{U,P} \left[\prod_{\ell=1}^m r(U_\ell) \prod_{j=1}^n w^{(x_j)}(U_{P(j)})^{\alpha_j} \right]$$

$$= \sum_{\mathcal{S}} \mathbf{E}_{U,P} \left[\prod_{\ell=1}^m r(U_\ell) \prod_{j=1}^n w^{(x_j)}(U_{P(j)})^{\alpha_j} \,\middle|\, P \text{ induces } \mathcal{S} \right] \mathbf{Pr}[P \text{ induces } \mathcal{S}]$$

$$= \sum_{|\mathcal{S}|=|\text{supp}(\boldsymbol{\alpha})|} \frac{x^{\alpha}}{D^m} \mathbf{Pr}[P \text{ induces } \mathcal{S}]$$

$$+ \sum_{|\mathcal{S}|<|\mathrm{supp}(\boldsymbol{\alpha})|} \frac{E(\mathcal{S}, \boldsymbol{\alpha}, \boldsymbol{x}, \boldsymbol{y})}{D^m} \mathbf{Pr}[\mathrm{P} \text{ induces } \mathcal{S}]$$

$$= \sum_{|\mathcal{S}|=|\mathrm{supp}(\boldsymbol{\alpha})|} \frac{\boldsymbol{x}^{\boldsymbol{\alpha}}}{D^m} \left[1 + q_{|\mathcal{S}|,|\mathrm{supp}(\boldsymbol{\alpha})|}\left(\tfrac{1}{m}\right) \right]$$

$$+ \sum_{|\mathcal{S}|<|\mathrm{supp}(\boldsymbol{\alpha})|} \frac{E(\mathcal{S}, \boldsymbol{\alpha}, \boldsymbol{x}, \boldsymbol{y})}{D^m} q_{|\mathcal{S}|,|\mathrm{supp}(\boldsymbol{\alpha})|}\left(\tfrac{1}{m}\right)$$

$$= \frac{\boldsymbol{x}^{\boldsymbol{\alpha}}}{D^m} + \frac{1}{D^m} p_{\boldsymbol{\alpha},\boldsymbol{x},\boldsymbol{y}}\left(\tfrac{1}{m}\right)$$

with

$$p_{\boldsymbol{\alpha},\boldsymbol{x},\boldsymbol{y}}(z) := \sum_{|\mathcal{S}|=|\mathrm{supp}(\boldsymbol{\alpha})|} \boldsymbol{x}^{\boldsymbol{\alpha}} q_{|\mathcal{S}|,|\mathrm{supp}(\boldsymbol{\alpha})|}(z)$$

$$+ \sum_{|\mathcal{S}|<|\mathrm{supp}(\boldsymbol{\alpha})|} E(\mathcal{S}, \boldsymbol{\alpha}, \boldsymbol{x}, \boldsymbol{y}) q_{|\mathcal{S}|,|\mathrm{supp}(\boldsymbol{\alpha})|}(z).$$

Recalling (3.17), $p_{\boldsymbol{\alpha},\boldsymbol{x},\boldsymbol{y}}$ satisfies $p_{\boldsymbol{\alpha},\boldsymbol{x},\boldsymbol{y}}(0) = 0$ and $\deg(p_{\boldsymbol{\alpha},\boldsymbol{x},\boldsymbol{y}}) < |\mathrm{supp}(\boldsymbol{\alpha})|$. $\square$

Proposition 1 generalizes immediately to low-degree polynomials by linearity.

Proposition 2 *Let $d \geq 1$ and suppose that f is an analytic polynomial on $\mathbf{D}^n$ of degree at most d with individual degree at most $K - 1$. Then for any $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbf{D}^n$, there exists a polynomial $p = p_{f,\boldsymbol{x},\boldsymbol{y}}$ of degree at most $d - 1$ such that $p_{f,\boldsymbol{x},\boldsymbol{y}}(0) = 0$ and the following holds. For any $m \geq d$, let $P : [n] \to [m]$ be constructed by choosing for each $j \in [n]$ uniformly at random $P(j) \in [m]$. We have*

$$f(\boldsymbol{x}) = D^m \mathbf{E}_{U,P} \left[\prod_{\ell=1}^m r(U_\ell) \cdot f\left(W^{(\boldsymbol{x})}(U_P) \right) \right] + p_{f,\boldsymbol{x},\boldsymbol{y}}\left(\frac{1}{m} \right) \qquad (3.20)$$

where

$$W^{(\boldsymbol{x})}(U_P) := \left(w^{(x_1)}(U_{P(1)}), \ldots, w^{(x_n)}(U_{P(n)}) \right) \in \{y_0, \ldots, y_{K-1}\}^n = \boldsymbol{y}^n.$$

Proof This follows immediately from (3.15) by linearity and choosing

$$p_{f,\boldsymbol{x},\boldsymbol{y}}(z) = \sum_{|\boldsymbol{\alpha}|\leq d} \sum_{|\mathcal{S}|=|\mathrm{supp}(\boldsymbol{\alpha})|} a_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}} q_{|\mathcal{S}|,|\mathrm{supp}(\boldsymbol{\alpha})|}(z)$$

$$+ \sum_{|\boldsymbol{\alpha}|\leq d} \sum_{|\mathcal{S}|<|\mathrm{supp}(\boldsymbol{\alpha})|} a_{\boldsymbol{\alpha}} E(\mathcal{S}, \boldsymbol{\alpha}, \boldsymbol{x}, \boldsymbol{y}) q_{|\mathcal{S}|,|\mathrm{supp}(\boldsymbol{\alpha})|}(z)$$

for $f(z) = \sum_{|\boldsymbol{\alpha}|\leq d} a_{\boldsymbol{\alpha}} z^{\boldsymbol{\alpha}}$. $\square$

Now we are ready to finish the proof of Theorem 2 when all Z_j's are identical.

Proof of Theorem 2 when Z_j's are identical Assume that

$$Z_j = \{y_0, \ldots, y_{K-1}\} \in \mathbf{D}^K$$

for all j, and we shall follow the above notations, *e.g.* $\mathbf{y} = (y_0, \ldots, y_{K-1})$ and $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbf{D}^n$ is fixed. Choose positive integers

$$d = m_0 < m_1 < \cdots < m_{d-1}$$

and real numbers $a_0, \ldots, a_{d-1}$ in such a way that

$$\sum_{j=0}^{d-1} a_j = 1 \quad \text{and} \quad \sum_{j=0}^{d-1} \frac{a_j}{m_j^k} = 0, \quad k = 1, \ldots, d-1. \tag{3.21}$$

According to Proposition 2, we have for all $m \geq d$ that

$$f(\mathbf{x}) = D^m \mathbf{E}_{U,P} \left[\prod_{\ell=1}^{m} r(U_\ell) \cdot f\left(W^{(\mathbf{x})}(U_P)\right) \right] + p_{f,\mathbf{x},\mathbf{y}}\left(\frac{1}{m}\right).$$

Applying this identity to $m = m_0, \ldots, m_{d-1}$ as above, we get

$$f(\mathbf{x}) = \sum_{j=0}^{d-1} a_j f(\mathbf{x})$$

$$= \sum_{j=0}^{d-1} a_j D^{m_j} \mathbf{E}_{U,P} \left[\prod_{\ell=1}^{m_j} r(U_\ell) \cdot f\left(W^{(\mathbf{x})}(U_P)\right) \right] + \sum_{j=0}^{d-1} a_j p_{f,\mathbf{x},\mathbf{y}}\left(\frac{1}{m_j}\right)$$

$$= \sum_{j=0}^{d-1} a_j D^{m_j} \mathbf{E}_{U,P} \left[\prod_{\ell=1}^{m_j} r(U_\ell) \cdot f\left(W^{(\mathbf{x})}(U_P)\right) \right] \tag{3.22}$$

where in the first and last equalities we used (3.21). Equation (3.22) is the explicit form of (1.9). It remains to instantiate the m_j's and control $\sum_{j=0}^{d-1} |a_j| D^{m_j}$.

Let us choose simple m_j's and estimate the $|a_j|$'s. For this, recall that (3.21) is equivalent to

$$V(a_0, \ldots, a_{d-1})^T = (1, 0, \ldots, 0)^T$$

where $V = [a_{jk}]_{j,k=0}^{d-1}$ is the Vandermonde matrix with $a_{jk} = \left(\frac{1}{m_k}\right)^j$. Denoting $[b_{jk}]_{j,k=0}^{d-1}$ the inverse of V, then $a_j = b_{j,0}$. Now let us estimate the absolute sum of $b_{j,0}$'s.

Now we choose $m_j = d + j, 0 \leq j \leq d - 1$. Then

$$\max_{0 \leq j \leq d-1} |m_j^{-1}| \leq \frac{1}{d}, \qquad \min_{0 \leq j < k \leq d-1} |m_j^{-1} - m_k^{-1}| \geq \frac{1}{4d^2}.$$

By (3.3), we have

$$|a_j| = |b_{j,0}| = \frac{\prod_{0 \le k \le d-1 : k \ne j} m_k^{-1}}{\prod_{0 \le k \le d-1 : k \ne j} |m_j^{-1} - m_k^{-1}|} = \prod_{0 \le k \le d-1 : k \ne j} \frac{m_j}{|m_k - m_j|}.$$

Recalling $m_j = d + j$, one obtains the estimate

$$|a_j| = \frac{(d+j)!}{j!(d-1-j)!} = \frac{d-j}{d+j} \frac{(d+j)^d}{d!} \binom{d}{j} \le C \cdot (2e)^d \left(\frac{d+j}{d}\right)^d \le C \cdot (4e)^d,$$

for all $0 \le j \le d-1$ with some universal $C > 0$, and thus

$$\sum_{j=0}^{d-1} |a_j| D^{m_j} \le Cd(4eD^2)^d \le C(K, \eta)^d, \tag{3.23}$$

for some $C(K, \eta) > 0$ depending on K and η only. Therefore, we have shown that

$$|f(\boldsymbol{x})| \le C(K, \eta)^d \max_{\boldsymbol{z} \in \{y_0, \dots, y_{K-1}\}^n} |f(\boldsymbol{z})|$$

for all $\boldsymbol{x} \in \mathbf{D}^n$. This finishes the proof of Theorem 2 when all the Z_j's are the same. $\quad\square$

Proof of Theorem 1 when Z_j's are identical Let $\boldsymbol{x}$ maximize $|f|$ on $\mathbf{D}^n$. Then directly from (3.22) we conclude

$$\|f\|_{\mathbf{D}^n} = |f(\boldsymbol{x})| \le \sum_{j=0}^{d-1} |a_j| D^{m_j} \cdot \max_{\boldsymbol{z} \in \{y_0, \dots, y_{K-1}\}^n} |f(\boldsymbol{z})|$$

$$\le C(K, \eta)^d \|f\|_{Y_n}. \qquad\qquad \square$$

3.4 Improved estimate for groups Ω_K^n

Now let us prove the second claims of Theorems 1 and 2, that is, we may choose the constant C in (1.7) to be $C(d, K) \le \left(\mathcal{O}(\log(K))\right)^{2d}$ when all $Z_j = \Omega_K$, $j = 1, \dots, n$. This is based on the following lemma.

Lemma 10 *Suppose $z \in \mathbf{D}$, $\omega = e^{\frac{2\pi \mathbf{i}}{K}}$. 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 A direct computation shows that

$$c_j = \frac{1}{K} \sum_{k=0}^{K-1} \omega^{-jk} z^k = \frac{1 - z^K}{K(1 - \omega^{-j}z)}, \qquad 0 \le j \le K - 1. \qquad (3.24)$$

Since $z \in \mathbf{D}$, we have by triangle inequality that

$$|c_j| \le \min\left\{1, \frac{2}{K|\omega^j - z|}\right\}, \qquad 0 \le j \le K - 1. \qquad (3.25)$$

By symmetry, we may assume that $\omega^0 = 1$ is a closest point in Ω_K to z and ω is the next-closest point in Ω_K to z. It can be that $z = 1$, so we use the upper bound $|c_0| \le 1$. We also estimate $|c_1| \le 1$. For $2 \le j \le K - 1$, we use the estimate

$$\sum_{j=2}^{K-1} \frac{2}{K|\omega^j - z|} \le 2 \sum_{j=2}^{\lceil \frac{K-1}{2} \rceil} \frac{2}{K|\omega^j - z|}.$$

For $2 \le j \le \lceil \frac{K-1}{2} \rceil$, we use the estimate

$$|\omega^j - z| \ge 2\sin\left(\frac{(j-1)\pi}{K}\right) \ge \frac{4}{\pi} \cdot \frac{(j-1)\pi}{K} = \frac{4(j-1)}{K}.$$

All combined, we get

$$\sum_{j=0}^{K-1} |c_j| \le 1 + 1 + 2 \sum_{j=2}^{\lceil \frac{K-1}{2} \rceil} \frac{2}{K|\omega^j - z|} \le 2 + 2 \sum_{j=2}^{\lceil \frac{K-1}{2} \rceil} \frac{2}{4j} \le 2 + \sum_{j=2}^{\lceil \frac{K-1}{2} \rceil} \frac{1}{j}. \qquad (3.26)$$

Then the proof is finished using the estimate $\sum_{k=1}^{L} 1/k \le \log(L) + 1$. □

This gives a better upper bound of L, namely $L \le C \log(K)$ improving (3.9). So

$$D = 4L + 1 \le 4C \log(K) + 1 \qquad (3.27)$$

and we obtain the desired bound following the proof of the previous subsection (see (3.23)).

3.5 Proof of Theorems 1 and 2: the general case

In this subsection we prove Theorems 1 and 2 for general $Y_n = \prod_{\ell=1}^{n} Z_\ell \subset \mathbf{D}^n$ with

$$Z_\ell = \{y_{\ell,0}, \dots, y_{\ell,K-1}\}, \qquad 1 \le \ell \le n$$

uniformly separated by $\eta > 0$ as stated in the theorem; that is,

$$\min_{1 \le \ell \le n} \min_{0 \le j < k \le K-1} |y_{\ell,j} - y_{\ell,k}| \ge \eta.$$

The proof is essentially the same as the case when Z_ℓ's are identical. In fact, fix $1 \le \ell \le n$ and put $\mathbf{y}_\ell = (y_{\ell,0}, \ldots, y_{\ell,K-1}) \in \mathbf{D}^K$. As before, for any $x \in \mathbf{D}$ there exist $c_k(x) = c_k(\mathbf{y}_\ell, x)$ such that

$$\sum_{k=0}^{K-1} c_k(\mathbf{y}_\ell, x) y_{\ell,j}^j = x^j, \quad 0 \le j \le K - 1, \tag{3.28}$$

and we have the same universal bound

$$\sum_{k=0}^{K-1} |c_k(\mathbf{y}_\ell, x)| \le K \left(\frac{2}{\eta}\right)^{K-1}$$

as in (3.6). This allows us to choose the same L and $D = 4L + 1$ as in Sect. 3.1 to define the functions r and $w^{(x)}$. Recall that $r : [0, D] \to \{\pm 1, \pm \mathbf{i}\}$ is the same as before, since it depends only on D. The definition of $w_{\mathbf{y}_\ell}^{(x)} : [0, D] \to Z_\ell$ depends on $x \in \mathbf{D}$ and the reference set Z_ℓ as well as some order $\mathbf{y}_\ell$. Then we still have the following identity as in (3.14)

$$\mathbf{E}\left[r(U) w_{\mathbf{y}_\ell}^{(x)}(U)^\alpha \right] = \frac{1}{D} x^\alpha, \quad x \in \mathbf{D}, \quad 0 \le \alpha \le K - 1 \tag{3.29}$$

for all $1 \le \ell \le n$. Repeating the previous proof word by word, as in (3.20) we arrive at

$$f(\mathbf{x}) = D^m \mathbf{E}_{U,P}\left[\prod_{\ell=1}^{m} r(U_\ell) \cdot f\left(W_{Y_n}^{(\mathbf{x})}(U_P) \right) \right] + p_{f,\mathbf{x},Y_n}\left(\frac{1}{m}\right) \tag{3.30}$$

for any $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbf{D}^n$ and $f : \mathbf{D}^n \to \mathbf{C}$ in Theorem 2, where

$$W_{Y_n}^{(\mathbf{x})}(U_P) := \left(w_{\mathbf{y}_1}^{(x_1)}(U_{P(1)}), \ldots, w_{\mathbf{y}_n}^{(x_n)}(U_{P(n)}) \right) \in Y_n,$$

$p_{f,\mathbf{x},Y_n}(0) = 0$ and $p_{f,\mathbf{x},Y_n}$ is of degree at most $d - 1$ as before. Here the notation $W_{Y_n}^{(\mathbf{x})}$ might not be a perfect choice as it depends on the some orders of each Z_ℓ, as indicated in its definition. But we abuse the notation here since we only need the fact that it takes values in Y_n. The rest is the same as the end of Sect. 3.3 and thus we finish the proof of Theorem 2. The general case of Theorem 1 follows immediately.

4 Small sampling sets

In this section we show sampling sets of cardinality $C(1 + \varepsilon)^n$ exist for arbitrary $\varepsilon > 0$. Denote by $C(d, K, \eta)$ the constant obtained in Theorem 1.

Proof of Theorem 3 For any $n \ge 1$, let $k \ge 1$ be an integer such that

$$((d + 1)k^d)^{1/k} \le 1 + \varepsilon. \tag{4.1}$$

We temporarily assume first that n is a multiple of k, $n = n'k$. The reader should think that n' is very large, and k is of the order $\frac{d}{\varepsilon} \log \frac{d}{\varepsilon}$ which is not too large. We are going to choose the desired sampling set in the following way. Write

$$\prod_{j=1}^{n} Z_j = \prod_{j=0}^{n'-1} Z_{jk+1} \times \cdots \times Z_{jk+k},$$

and for each $0 \leq j \leq n' - 1$, we are going to choose M (to be defined later) points in $Z_{jk+1} \times \cdots \times Z_{jk+k}$.

More precisely, define (denoting $\|\boldsymbol{\alpha}\|_p$ the ℓ^p-norm of a vector $\boldsymbol{\alpha}$)

$$\Lambda = \{\boldsymbol{\alpha} \in \mathbf{N}^k : \|\boldsymbol{\alpha}\|_1 \leq d, \|\boldsymbol{\alpha}\|_\infty \leq K - 1\}, \tag{4.2}$$

so that $\{z^{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \Lambda\}$ is a $\mathbb{C}$-vector space basis of the polynomials of degree at most d and individual degree at most $K - 1$ in k variables $z_1, \ldots, z_k \in \mathbf{D}$. The cardinality of Λ is

$$M := |\Lambda| \leq (d + 1)k^d. \tag{4.3}$$

Let $\mathbf{Y} \in (\mathbf{D}^k)^M$. We will write

$$\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_M), \quad \mathbf{y}_j = (y_{j1}, y_{j2}, \ldots, y_{jk}) \in \mathbf{D}^k.$$

For any $z = (z_1, \ldots, z_k) \in \mathbf{D}^k$, we want to solve the system of M equations

$$z^{\boldsymbol{\alpha}} = z_1^{\alpha_1} \ldots z_k^{\alpha_k} = \sum_{j=1}^{M} c_j(z) \mathbf{y}_j^{\boldsymbol{\alpha}} = \sum_{j=1}^{M} c_j(z) \prod_{r=1}^{k} y_{jr}^{\alpha_r}, \quad \boldsymbol{\alpha} \in \Lambda \tag{4.4}$$

where $c_j(z) = c_j(\mathbf{Y}; z)$ depends on $\mathbf{Y}$ and z if it exists. The system is solvable for all $z \in \mathbf{D}^k$ whenever the determinant

$$P(\mathbf{Y}) := \det[\mathbf{y}_j^{\boldsymbol{\alpha}}]_{1 \leq j \leq M, \boldsymbol{\alpha} \in \Lambda} = \det[\mathbf{y}_j^{\boldsymbol{\beta}_i}]_{1 \leq i, j \leq M} = \sum_{\sigma \in S_M} \text{sgn}(\sigma) \prod_{j=1}^{M} \mathbf{y}_j^{\boldsymbol{\beta}_{\sigma(j)}} \tag{4.5}$$

does not vanish, where $\{\boldsymbol{\beta}_i : 1 \leq i \leq M\}$ is some reordering of $\{\boldsymbol{\alpha} : \boldsymbol{\alpha} \in \Lambda\}$, and S_M denotes the permutation group on M letters. If this is the case, we can further explicitly express the solutions of (4.4) using Cramer's rule as

$$c_j(z) = c_j(\mathbf{Y}; z) = \frac{P_j(\mathbf{Y}, z)}{P(\mathbf{Y})},$$

where $P_j(\mathbf{Y}, z)$ is the determinant of the $M \times M$ matrix obtained by replacing the j-th column of $[\mathbf{y}_j^{\boldsymbol{\beta}_i}]_{1 \leq i, j \leq M}$ by $(z^{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \Lambda} = (z^{\boldsymbol{\beta}_i})_{1 \leq i \leq M}$. All the coefficients of this matrix are contained in the unit disc, so by a theorem of Hadamard [27] we have

$$|P_j(\mathbf{Y}, z)| \leq M^{M/2}.$$

On the other hand, the monomials in (4.5) for different σ are all different, hence using orthogonality of monomials in $L^2(\mathbf{T}^{kM})$ we obtain

$$\int_{\mathbf{Y} \in \mathbf{T}^{kM}} |P(\mathbf{Y})|^2 \, d\mathbf{Y} = M!.$$

Here we equip $\mathbf{T}^{kM}$ with the uniform probability measure $d\mathbf{Y}$. Therefore, for this analytic polynomial P over $\mathbf{D}^{kM}$ of degree at most dM and individual degree at most $K-1$, we have

$$\|P\|_{\mathbf{D}^{kM}} = \|P\|_{\mathbf{T}^{kM}} \geq \|P\|_{L^2(\mathbf{T}^{kM}, d\mathbf{Y})} = \sqrt{M!}.$$

By Theorem 1, for any $1 \leq i_1, \ldots, i_k \leq n$ and the sets $Z_{i_1}, \ldots, Z_{i_k}$, there exists $\underline{\mathbf{Y}} \in (Z_{i_1} \times \cdots \times Z_{i_k})^M$ with

$$|P(\underline{\mathbf{Y}})| \geq C(dM, K, \eta)^{-1} \sqrt{M!} > 0. \tag{4.6}$$

For $\underline{\mathbf{Y}}$ satisfying (4.6), the above system (4.4) is solvable and we can estimate the size of the solutions $c_j(\underline{\mathbf{Y}}; z)$ of (4.4):

$$|c_j(\underline{\mathbf{Y}}; z)| \leq \frac{|P_j(\underline{\mathbf{Y}}, z)|}{|P(\underline{\mathbf{Y}})|} \leq C(dM, K, \eta) \frac{M^{M/2}}{\sqrt{M!}}$$

$$\leq C(K, \eta)^{dM} \frac{M^{M/2}}{(M/e)^{M/2}} \leq C(K, \eta)^{dM} e^{M/2}. \tag{4.7}$$

Note that $c_j(\underline{\mathbf{Y}}; z)$ depends on $\underline{\mathbf{Y}}$ and z but this upper bound of $|c_j(\underline{\mathbf{Y}}; z)|$ does not.

We now construct our sampling set. Recall that $n = n'k$ is a multiple of k. For each $s = 0, \ldots, n' - 1$, consider a tuple of points

$$Y^{(s)} = (y_1^{(s)}, \ldots, y_M^{(s)}) \in (Z_{sk+1} \times \cdots \times Z_{sk+k})^M$$

such that (4.6) holds for $Z_{i_j} = Z_{sk+j}$, $j = 1, \ldots, k$. These are the M points that we choose in $Z_{sk+1} \times \cdots \times Z_{sk+k}$. Our sampling set will be

$$\mathbf{Y} := \prod_{s=0}^{n'-1} Y^{(s)},$$

which satisfies, by (4.1) and (4.3),

$$|\mathbf{Y}| \leq M^{n'} \leq ((d+1)k^d)^{n/k} \leq (1+\varepsilon)^n.$$

It remains to show that $\mathbf{Y}$ is a sampling set with dimension-free constant. The argument is essentially the same as in the proof of Theorem 1; the only difference is that the functions w take values in $Y^{(s)} \subset \mathbf{D}^{kM}$. Let $f(z) = \sum_\alpha a_\alpha z^\alpha$ be a polynomial in n variables of degree at most d and individual degree at most $K-1$. Fix

$$z = (z_0, \ldots, z_{n'-1}) \in \mathbf{D}^n = (\mathbf{D}^k)^{n'} \qquad \text{with} \qquad z_s \in \mathbf{D}^k$$

and α such that $a_\alpha \neq 0$. Then $z^\alpha = \prod_{s=0}^{n'-1} z_s^{\alpha_s}$.

Since each $Y^{(s)}$ satisfies (4.6), the determinant of the system (4.4) does not vanish. So we have

$$z_s^{\alpha_s} = \sum_{j=1}^{M} c_j(z_s)(y_j^{(s)})^{\alpha_s}, \qquad \alpha_s \in \Lambda$$

and by (4.7)

$$\sum_{j=1}^{M} |c_j(z_s)| \leq MC(dM, K, \eta)e^{M/2} =: L. \tag{4.8}$$

As before, L is independent of $Y^{(s)}$ and z.

Now for the convenience of the reader we sketch the construction in Sect. 3 with minor modifications. Again, we write each complex number $c_j(z_s)$ in the following form

$$c_j(z_s) = c_j^{(1)}(z_s) - c_j^{(-1)}(z_s) + \mathbf{i}c_j^{(\mathbf{i})}(z_s) - \mathbf{i}c_j^{(-\mathbf{i})}(z_s),$$

where all $c_j^{(r)}(z_s)$'s are non-negative such that

$$c_j^{(1)}(z_s) = (\Re c_j(z_s))_+, \qquad c_j^{(-1)}(z_s) = (\Re c_j(z_s))_-,$$

and

$$c_j^{(\mathbf{i})}(z_s) = (\Im c_j(z_s))_+, \qquad c_j^{(-\mathbf{i})}(z_s) = (\Im c_j(z_s))_-.$$

Same as before, we may choose non-negative $t^{(1)}(z_s) = t^{(-1)}(z_s), t^{(\mathbf{i})}(z_s) = t^{(-\mathbf{i})}(z_s)$ in such a way that

$$\sum_{j=1}^{M} c_j^{(1)}(z_s) + t^{(1)}(z_s) = L + 1, \tag{4.9}$$

$$\sum_{j=1}^{M} c_j^{(r)}(z_s) + t^{(r)}(z_s) = L, \quad r = -1, \mathbf{i}, -\mathbf{i}. \tag{4.10}$$

Moreover,

$$t^{(1)}(z_s) = t^{(-1)}(z_s) \qquad \text{and} \qquad t^{(\mathbf{i})}(z_s) = t^{(-\mathbf{i})}(z_s). \tag{4.11}$$

Again, we put $D = 4L + 1$ and divide the interval $[0, D]$ into the disjoint union

$$[0, D] = I^{(1)} \cup I^{(-1)} \cup I^{(\mathbf{i})} \cup I^{(-\mathbf{i})},$$

with $|I^{(1)}| = L + 1$ and $|I^{(-1)}| = |I^{(\mathbf{i})}| = |I^{(-\mathbf{i})}| = L$. Define the function $r : [0, D] \to \{\pm 1, \pm \mathbf{i}\}$ as before

$$r(t) = s, \qquad t \in I^{(s)}.$$

For each z_s we further decompose each $I^{(r)}$ into disjoint unions:

$$I^{(r)} = \bigcup_{k=1}^{M} \left(I_k^{(r)}(z_s) \cup J_k^{(r)}(z_s) \right), \qquad r = \pm 1, \pm \mathbf{i}$$

with

$$|I_k^{(r)}(z_s)| = c_k^{(r)}(z_s) \qquad \text{and} \qquad |J_k^{(r)}(z_s)| = \frac{t^{(r)}(z_s)}{K}, \qquad 1 \le k \le M.$$

Then we define the function $w_{Y^{(s)}}^{(z_s)} : [0, D] \to Y^{(s)}$ as

$$w_{Y^{(s)}}^{(z_s)}(t) = y_k^{(s)}, \qquad t \in I_k^{(r)}(z_s) \cup J_k^{(r)}(z_s)$$

for all $1 \le k \le M$ and $r \in \{\pm 1, \pm \mathbf{i}\}$.

Suppose U is a random variable uniformly distributed on $[0, D]$, then

$$\mathbf{E}[r(U)] = \frac{1}{D}, \tag{4.12}$$

and

$$\mathbf{E}\left[r(U) w_{Y^{(s)}}^{(z_s)}(U)^{\alpha_s} \right] = \frac{1}{D} z_s^{\alpha_s}, \qquad z_s \in \mathbf{D}^k, \alpha_s \in \Lambda.$$

Arguing as before, we may deduce that there exists a polynomial $p = p_{f,z,\mathbf{Y}}$ of degree at most $d-1$ such that $p_{f,z,\mathbf{Y}}(0) = 0$ and the following holds. For any $m \ge d$, let $P : [n'] \to [m]$ be constructed by choosing for each $i \in [n']$ uniformly at random $P(i) \in [m]$. We have

$$f(z) = D^m \mathbf{E}_{U,P}\left[\prod_{j=1}^{m} r(U_j) \cdot f\left(W_{\mathbf{Y}}^{(z)}(U_P) \right) \right] + p_{f,z,\mathbf{Y}}\left(\frac{1}{m} \right) \tag{4.13}$$

where

$$W_{\mathbf{Y}}^{(z)}(U_P) := \left(w_{Y_0}^{(z_0)}(U_{P(1)}), \ldots, w_{Y_{n'-1}}^{(z_{n'-1})}(U_{P(n')}) \right) \in \mathbf{Y},$$

$p_{f,z,\mathbf{Y}}(0) = 0$ and $p_{f,z,\mathbf{Y}}$ is of degree at most $d-1$ as before.

The rest of proof is exactly as before. Thus $\mathbf{Y}$ discretizes the uniform norm with constant at most

$$Cd(4eD^2)^d. \tag{4.14}$$

This completes the proof when n is a multiple of k.

In general, if $n'k < n < (n'+1)k \le n+k-1$, then we may project the sampling set from $\mathbf{D}^{(n'+1)k}$ onto $\mathbf{D}^n$ so that there exists a sampling set (Y_n) of size at most

$$|Y_n| \le (1+\varepsilon)^{n+k-1} = C_1(d, \varepsilon)(1+\varepsilon)^n. \tag{4.15}$$

Finally, let us compute the precise dependence of $C_1(d, \epsilon)$ and $C_2(d, K, \eta, \epsilon)$ on ε. The condition (4.1) on k holds for all $d \geq 1$ and $0 < \varepsilon \leq 1/2$ if k is the largest integer smaller than

$$100 \frac{d}{\varepsilon} \log\left(\frac{d}{\varepsilon}\right).$$

This yields for $C_1(d, \varepsilon)$, using (4.15)

$$C_1(d, \varepsilon) \leq (1 + \varepsilon)^k \leq \exp(100d \log\left(\frac{d}{\varepsilon}\right)) = \left(\frac{d}{\varepsilon}\right)^{100d}.$$

For $C_2(d, K, \eta, \varepsilon)$, we obtain with (4.3), (4.8) and (4.14) the bound

$$C_2(d, K, \eta, \varepsilon) \leq \exp(C(d, K, \eta)(\varepsilon^{-1} \log(\varepsilon^{-1}))^d).$$

The number 100 in the exponent can be brought arbitrarily close to 1, if ε is assumed to be sufficiently small. $\qquad\square$

5 Necessity of exponential-cardinality sampling sets

Now we prove Theorem 4 which says sampling sets with cardinality sub-exponential in dimension cannot exist.

Proof of Theorem 4 For any $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n) \in \{-1, 1\}^n$, consider the polynomials $f_{\boldsymbol{\varepsilon}}(\boldsymbol{x}) = \sum_{j=1}^n \varepsilon_j x_j$ on $\{-1, 1\}^n$ of degree at most 1. Then by definition,

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

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

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

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

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

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

$$|\Lambda_{\boldsymbol{v}}| = 2^n \mathbf{Pr}\left[\left| \sum_{j=1}^n v_j \varepsilon_j \right| \geq 2\delta n \right]$$

$$\leq 2^n \mathbf{Pr}\left[\left|\sum_{j=1}^n \Re(v_j)\varepsilon_j\right| \geq \delta n\right] + 2^n \mathbf{Pr}\left[\left|\sum_{j=1}^n \Im(v_j)\varepsilon_j\right| \geq \delta n\right]$$

$$\leq 2^{n+1}\exp\left(-\frac{\delta^2 n^2}{2\|a\|_2^2}\right) + 2^{n+1}\exp\left(-\frac{\delta^2 n^2}{2\|b\|_2^2}\right),$$

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

6 Discretizations of L^p norms

In this section, we prove Theorem 5 about an L^p version of dimension-free discretizations for products of cyclic groups.

Proof of Theorem 5 Write $Y_n = \prod_{j=1}^n Z_j$ with $Z_j = \Omega_K$ and $\mathbf{y} = (1, \omega, \ldots, \omega^{K-1})$ with $\omega = e^{\frac{2\pi i}{K}}$. According to (3.22), we have for all $\mathbf{x} = (x_1, \ldots, x_n) \in \mathbf{T}^n$ that

$$f(\mathbf{x}) = \sum_{j=0}^{d-1} a_j D^{m_j}\mathbf{E}_{U,P}\left[\prod_{\ell=1}^{m_j} r(U_\ell) \cdot f\left(W_{Y_n}^{(\mathbf{x})}(U_P)\right)\right]$$

where $a_j, m_j, 0 \leq j \leq d-1$ are as defined in Sect. 3.3 and

$$W_{Y_n}^{(\mathbf{x})}(U_P) := \left(w_{\mathbf{y}}^{(x_1)}(U_{P(1)}), \ldots, w_{\mathbf{y}}^{(x_n)}(U_{P(n)})\right) \in Y_n = \Omega_K^n.$$

Again, we abuse the notation $W_{Y_n}^{(\mathbf{x})}$ which actually depends on some order of each Z_j as specified in its definition. See Sect. 3.5. For any $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_n) \in \Omega_K^n$, denote $\mathbf{x}\boldsymbol{\xi} = (x_1\xi_1, \ldots, x_n\xi_n)$ the multiplication of $\mathbf{x}$ and $\boldsymbol{\xi}$ as elements in the group $\mathbf{T}^n$. Then $\mathbf{x}\boldsymbol{\xi} \in \mathbf{T}^n$ and

$$f(\mathbf{x}\boldsymbol{\xi}) = \sum_{j=0}^{d-1} a_j D^{m_j}\mathbf{E}_{U,P}\left[\prod_{\ell=1}^{m_j} r(U_\ell) \cdot f\left(W_{Y_n}^{(\mathbf{x}\boldsymbol{\xi})}(U_P)\right)\right].$$

For any $0 \leq j \leq K - 1$, recall that $\omega_{\mathbf{y}}^{(x_j \xi_j)}$ takes values in Ω_K with

$$\mathbf{Pr}\left[\omega_{\mathbf{y}}^{(x_j \xi_j)}(U) = \omega^k\right] = \frac{1}{D} \sum_{s=\pm 1, \pm \mathbf{i}} \left(c_k^{(s)}(\mathbf{y}; x_j \xi_j) + \frac{t^{(s)}(\mathbf{y}; x_j \xi_j)}{K}\right)$$

$$= \frac{1}{K} + \frac{1}{D}\left\{ \sum_{s=\pm 1, \pm \mathbf{i}} c_k^{(s)}(\mathbf{y}; x_j \xi_j) - \frac{1}{K} \sum_{\ell=0}^{K-1} c_\ell^{(s)}(\mathbf{y}; x_j \xi_j) \right\},$$

where we used (3.10) and (3.11). Also, $(c_k(\mathbf{y}; x_j \xi_j))_k$ is uniquely determined by

$$\sum_{k=0}^{K-1} c_k(\mathbf{y}; x_j \xi_j) \omega^{k\ell} = (x_j \xi_j)^\ell, \qquad 0 \leq \ell \leq K - 1,$$

or equivalently

$$\sum_{k=0}^{K-1} c_k(\mathbf{y}; x_j \xi_j)(\omega^k \overline{\xi_j})^\ell = x_j^\ell, \qquad 0 \leq \ell \leq K - 1.$$

So $c_k(\mathbf{y}; x_j \xi_j) = c_k(\mathbf{y}_j; x_j)$ with $\mathbf{y}_j = (\overline{\xi_j}, \overline{\xi_j}\omega, \ldots, \overline{\xi_j}\omega^{K-1})$ that is the same as Ω_K as a set. Here we used the group structure of Ω_K. Therefore,

$$\mathbf{Pr}\left[\omega_{\mathbf{y}}^{(x_j \xi_j)}(U) = \omega^k\right] = \frac{1}{K} + \frac{1}{D}\left\{ \sum_{s=\pm 1, \pm \mathbf{i}} c_k^{(s)}(\mathbf{y}; x_j \xi_j) - \frac{1}{K} \sum_{\ell=0}^{K-1} c_\ell^{(s)}(\mathbf{y}; x_j \xi_j) \right\}$$

$$= \frac{1}{K} + \frac{1}{D}\left\{ \sum_{s=\pm 1, \pm \mathbf{i}} c_k^{(s)}(\mathbf{y}_j; x_j) - \frac{1}{K} \sum_{\ell=0}^{K-1} c_\ell^{(s)}(\mathbf{y}_j; x_j) \right\}$$

$$= \mathbf{Pr}\left[\omega_{\mathbf{y}_j}^{(x_j)}(U) = \overline{\xi_j}\omega^k\right].$$

So we just argued that $\omega_{\mathbf{y}}^{(x_j \xi_j)}(U) = \xi_j \omega_{\mathbf{y}_j}^{(x_j)}(U)$. Thus we find

$$f(\mathbf{x}\boldsymbol{\xi}) = \sum_{j=0}^{d-1} a_j D^{m_j} \mathbf{E}_{U,P} \left[\prod_{\ell=1}^{m_j} r(U_\ell) \cdot f\left(W_{Y_n}^{(\mathbf{x}\boldsymbol{\xi})}(U_P)\right) \right]$$

$$= \sum_{j=0}^{d-1} a_j D^{m_j} \mathbf{E}_{U,P} \left[\prod_{\ell=1}^{m_j} r(U_\ell) \cdot f\left(\boldsymbol{\xi} W_{Y_n'}^{(\mathbf{x})}(U_P)\right) \right]$$

where $Y_n' = Y_n$ with a different ordering of each Z_j:

$$W_{Y_n'}^{(\mathbf{x})}(U_P) = \left(w_{\mathbf{y}_1}^{(x_1)}(U_{P(1)}), \ldots, w_{\mathbf{y}_n}^{(x_n)}(U_{P(n)})\right) \in \Omega_K^n.$$

So by Jensen's inequality, and recalling (3.23) and (3.27),

$$|f(\boldsymbol{x}\boldsymbol{\xi})|^p \leq d^{p-1} \sum_{j=0}^{d-1} |a_j D^{m_j}|^p \left| \mathbf{E}_{U,P} \left[\prod_{\ell=1}^{m_j} r(U_\ell) \cdot f\left(\boldsymbol{\xi} W_{Y_n'}^{(\boldsymbol{x})}(U_P)\right) \right] \right|^p$$

$$\leq d^{p-1} \sum_{j=0}^{d-1} |a_j D^{m_j}|^p \mathbf{E}_{U,P} \left| f\left(\boldsymbol{\xi} W_{Y_n'}^{(\boldsymbol{x})}(U_P)\right) \right|^p$$

$$\leq d^{p-1} (C_1 \log(K) + C_2)^{dp} \mathbf{E}_{U,P} \left| f\left(\boldsymbol{\xi} W_{Y_n'}^{(\boldsymbol{x})}(U_P)\right) \right|^p.$$

Since $W_{Y_n'}^{(\boldsymbol{x})}(U_P)$ takes values in Ω_K^n and Ω_K^n is a group, we have

$$\mathbf{E}_{\boldsymbol{\xi} \sim \Omega_K^n} \left| f\left(\boldsymbol{\xi} W_{Y_n'}^{(\boldsymbol{x})}(U_P)\right) \right|^p = \|f\|_{L^p(\Omega_K^n)}^p.$$

Similarly, for any $\boldsymbol{\xi} \in \Omega_K^n \subset \mathbf{T}^n$, we have

$$\mathbf{E}_{\boldsymbol{x} \sim \mathbf{T}^n} |f(\boldsymbol{x}\boldsymbol{\xi})|^p = \|f\|_{L^p(\mathbf{T}^n)}^p.$$

All combined, we conclude

$$\|f\|_{L^p(\mathbf{T}^n)}^p = \mathbf{E}_{\boldsymbol{\xi} \sim \Omega_K^n} \mathbf{E}_{\boldsymbol{x} \sim \mathbf{T}^n} |f(\boldsymbol{x}\boldsymbol{\xi})|^p$$

$$\leq d^{p-1} (C_1 \log(K) + C_2)^{dp} \mathbf{E}_{U,P} \mathbf{E}_{\boldsymbol{x} \sim \mathbf{T}^n} \mathbf{E}_{\boldsymbol{\xi} \sim \Omega_K^n} \left| f\left(\boldsymbol{\xi} W_{Y_n'}^{(\boldsymbol{x})}(U_P)\right) \right|^p$$

$$= d^{p-1} (C_1 \log(K) + C_2)^{dp} \|f\|_{L^p(\Omega_K^n)}^p.$$

This finishes the proof of the theorem. $\square$

References

1. Alon, N., Linial, N., Meshulam, R.: Additive bases of vector spaces over prime fields. J. Comb. Theory, Ser. A **57**(2), 203–210 (1991)

2. Alon, N., Dinur, I., Friedgut, E., Sudakov, B.: Graph products, Fourier analysis and spectral techniques. Geom. Funct. Anal. **14**(5), 913–940 (2004)

3. Baker, R.C., Schmidt, W.M.: Diophantine problems in variables restricted to the values 0 and 1. J. Number Theory **12**(4), 460–486 (1980)

4. Bakry, D., Gentil, I., Ledoux, M.: Analysis and Geometry of Markov Diffusion Operators. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 348, p. xx+552. Springer, Cham (2014)

5. Bayart, F., Pellegrino, D., Seoane-Sepúlveda, J.B.: The Bohr radius of the n-dimensional polydisk is equivalent to $\sqrt{(\log n)/n}$. Adv. Math. **264**, 726–746 (2014)

6. Beckner, W.: Inequalities in Fourier analysis on $\mathbf{R}^n$. Proc. Natl. Acad. Sci. **72**(2), 638–641 (1975)

7. Bernstein, S.: Démonstration du théorème de Weierstrass fondée sur le calcul des probabilités. In: Communications of the Kharkov Mathematical Society XIII (1912)

8. Bernstein, S.: Sur une classe de formules d'interpolation. French. Bull. Acad. Sci. URSS **1931**(9), 1151–1161 (1931)

9. Bernstein, S.: 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)

10. Black, H., Chakrabarty, D., Seshadri, C.: Directed isoperimetric theorems for Boolean functions on the hypergrid and an $\tilde{\mathcal{O}}(n\sqrt{d})$ monotonicity tester. In: Proceedings of the 55th Annual ACM Symposium on Theory of Computing. STOC 2023, pp. 233–241. Association for Computing Machinery, Orlando (2023)

11. Blei, R.: Analysis in Integer and Fractional Dimensions. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge (2001)

12. Bobkov, S.G.: An isoperimetric inequality on the discrete cube, and an elementary proof of the isoperimetric inequality in Gauss space. Ann. Probab. **25**(1), 206–214 (1997)

13. Bohnenblust, H.F., Hille, E.: On the absolute convergence of Dirichlet series. Ann. Math. (2) **32**(3), 600–622 (1931)

14. Bohr, H.: Über die Bedeutung der Potenzreihen unendlich vieler Variabeln in der Theorie der Dirichletschen Reihen $\sum \frac{a_n}{n^s}$. German. Nachr. Ges. Wiss. Göttingen, Math.-Phys. Kl. **1913**, 441–488 (1913)

15. Brudnyi, A., Yomdin, Y.: Norming sets and related Remez-type inequalities. J. Aust. Math. Soc. **100**(2), 163–181 (2016)

16. Buhovsky, L., Logunov, A., Malinnikova, E., Sodin, M.: A discrete harmonic function bounded on a large portion of $\mathbb{Z}^2$ is constant. Duke Math. J. **171**(6), 1349–1378 (2022)

17. Dai, F., Prymak, A.: Optimal polynomial meshes exist on any multivariate convex domain. Found. Comput. Math. **24**(3), 989–1018 (2024)

18. Dai, F., Prymak, A., Temlyakov, V.N., Tikhonov, S.Y.: Integral norm discretization and related problems. Usp. Mat. Nauk **74.4**(448), 3–58 (2019)

19. Deeparnab, C., Seshadri, C.: An optimal lower bound for monotonicity testing over hypergrids. Theory Comput. **10**(17), 453–464 (2014)

20. Defant, A., Sevilla-Peris, P.: The Bohnenblust-Hille cycle of ideas from a modern point of view. Funct. Approx. Comment. Math. **50**(1), 55–127 (2014)

21. Defant, A., Frerick, L., Ortega-Cerdà, J., Ounaïes, M., Seip, K.: The Bohnenblust-Hille inequality for homogeneous polynomials is hypercontractive. Ann. Math. (2) **174**(1), 485–497 (2011)

22. Defant, A., García, D., Maestre, M., Sevilla-Peris, P.: Dirichlet Series and Holomorphic Functions in High Dimensions. New Mathematical Monographs. Cambridge University Press, Cambridge (2019)

23. Defant, A., Mastyło, M., Pérez, A.: On the Fourier spectrum of functions on Boolean cubes. Math. Ann. **374**, 653–680 (2019)

24. Eskenazis, A., Ivanisvili, P.: 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, pp. 203–207. Association for Computing Machinery, Rome (2022)

25. Friedgut, E.: Boolean functions with low average sensitivity depend on few coordinates. Combinatorica **18**(1), 27–35 (1998)

26. Friedgut, E., Kalai, G., Naor, A.: Boolean functions whose Fourier transform is concentrated on the first two levels. Adv. Appl. Math. **29**(3), 427–437 (2002)

27. Hadamard, J.: Résolution d'une question relative aux déterminants. French. Bull. Sci. Math., II. Sér. **17**, 240–246 (1893)
28. Hu, Y.: Analysis on Gaussian Spaces, p. xi+470. World Scientific Publishing Co. Pte. Ltd., Hackensack (2017)
29. Huang, H.-Y., Chen, S., Preskill, J.: Learning to predict arbitrary quantum processes. PRX Quantum **4**(4), 040337 (2023)
30. Janson, S.: On complex hypercontractivity. J. Funct. Anal. **151**(1), 270–280 (1997)
31. Junge, M., Palazuelos, C., Parcet, J., Perrin, M.: Hypercontractivity in group von Neumann algebras. Mem. Am. Math. Soc. **249**(1183), xii+83 (2017)
32. Kahn, J., Kalai, G., Linial, N.: The influence of variables on Boolean functions. In: [Proceedings 1988] 29th Annual Symposium on Foundations of Computer Science, pp. 68–80 (1988)
33. Kashin, B., Kosov, E., Limonova, I., Temlyakov, V.: Sampling discretization and related problems. J. Complex. **71**, Article ID 101653 (2022)
34. Kashin, B., Konyagin, S., Temlyakov, V.: Sampling discretization of the uniform norm. Constr. Approx. **57**(2), 663–694 (2023)
35. Khot, S., Kindler, G., Mossel, E., O'Donnell, R.: Optimal inapproximability results for MAX-CUT and other 2-variable CSPs? SIAM J. Comput. **37**(1), 319–357 (2007)
36. Kroó, A.: On optimal polynomial meshes. J. Approx. Theory **163**(9), 1107–1124 (2011)
37. Linial, N., Mansour, Y., Nisan, N.: Constant depth circuits, Fourier transform, and learnability. J. Assoc. Comput. Mach. **40**, 607–620 (1993)
38. Littlewood, J.E.: On bounded bilinear forms in an infinite number of variables. Q. J. Math. **os-1.1**, 164–174 (1930)
39. Meshulam, R.: On subsets of finite Abelian groups with no 3-term arithmetic progressions. J. Comb. Theory, Ser. A **71**(1), 168–172 (1995)
40. Milman, V.D., Schechtman, G.: Asymptotic Theory of Finite Dimensional Normed Spaces. Lecture Notes in Mathematics. Springer, Berlin (1986)
41. Mossel, E., O'Donnell, R., Oleszkiewicz, K.: Noise stability of functions with low influences: invariance and optimality. Ann. Math. **171**(1), 295–341 (2010)
42. O'Donnell, R.: Analysis of Boolean Functions, p. xx+423. Cambridge University Press, New York (2014)
43. Ortega-Cerdà, J., Saludes, J.: Marcinkiewicz-Zygmund inequalities. J. Approx. Theory **145**(2), 237–252 (2007)
44. Remez, E.J.: Sur une propriété des polynômes de Tchebycheff. Comm. Inst. Sci. Kharkov **13**, 93–95 (1936)
45. Slote, J., Volberg, A., Zhang, H.: Bohnenblust-Hille inequality for cyclic groups. Adv. Math. **452**, Article ID 109824 (2024)
46. Slote, J., Volberg, A., Zhang, H.: Noncommutative Bohnenblust–Hille Inequality for qudit systems. arXiv:2406.08509. (2024)
47. Tikhonov, S., Yuditskii, P.: Sharp Remez inequality. Constr. Approx. **52**(2), 233–246 (2020)
48. Volberg, A., Zhang, H.: Noncommutative Bohnenblust–Hille inequalities. Math. Ann. **389**(2), 1657–1676 (2024)
49. Weissler, F.B.: Logarithmic Sobolev inequalities and hypercontractive estimates on the circle. J. Funct. Anal. **37**(2), 218–234 (1980)
50. Yomdin, Y.: Remez-type inequality for discrete sets. Isr. J. Math. **186**, 45–60 (2011)
51. Zygmund, A.: 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

Authors and Affiliations

Lars Becker[1] · **Ohad Klein**[2] · **Joseph Slote**[3] · **Alexander Volberg**[4,5] · **Haonan Zhang**[6]

✉ L. Becker
becker@math.uni-bonn.de

O. Klein
ohadkel@gmail.com

J. Slote
jslote@caltech.edu

A. Volberg
volberg@math.msu.edu

H. Zhang
haonanzhangmath@gmail.com

1   Mathematical Institute, University of Bonn, Endenicher Allee 60, 53115, Bonn, Germany

2   School of Computer Science and Engineering, Hebrew University, Jerusalem, Israel

3   Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA 91125, USA

4   Department of Mathematics, MSU, East Lansing, MI 48823, USA

5   Hausdorff Center of Mathematics, Bonn, Germany

6   Department of Mathematics, University of South Carolina, Columbia SC 29208, USA