ISSN 0081-5438, Proceedings of the Steklov Institute of Mathematics, 2018, Vol. 303, pp. 140–153. © Pleiades Publishing, Ltd., 2018.
Original Russian Text © B.S. Kashin, Yu.V. Malykhin, K.S. Ryutin, 2018, published in Trudy Matematicheskogo Instituta imeni V.A. Steklova, 2018, Vol. 303, pp. 155–168.

Kolmogorov Width and Approximate Rank

B. S. Kashin[a,b], **Yu. V. Malykhin**[a,b], **and K. S. Ryutin**[b]

Received May 30, 2018

Abstract—Closely related notions of the Kolmogorov width and the approximate rank of a matrix are considered. New estimates are established in approximation problems related to the width of the set of characteristic functions of intervals; the multidimensional case (characteristic functions of parallelepipeds) is also considered.

DOI: 10.1134/S0081543818080126

1. INTRODUCTION

Recall the classical definition from approximation theory introduced by A. N. Kolmogorov in 1936. The *Kolmogorov width* of order n of a set W in a normed space X is the quantity

$$d_n(W,X) := \inf_{L_n\subset X}\sup_{x\in W}\inf_{y\in L_n}\|x-y\|_X,$$

where the infimum is taken over linear subspaces L_n in X of dimension at most n. It is clear from the definition that the width characterizes the best possible approximation of a set by linear subspaces of a fixed dimension. For more details on the Kolmogorov width, see, for example, the book [26].

Since the late 1970s, matrix rigidity has been studied in complexity theory; for a matrix A, it can be defined as the minimum possible rank of matrices B obtained from A by varying at most k elements. There are still no constructive examples of a family of $N\times N$ matrices A_N whose rank is stable under the variation of a small part of elements of the matrix, i.e., a family for which a variation of at most $N^{1+\delta}$ elements of a matrix A_N always leads to a matrix B_N with $\operatorname{rank} B_N\ge\varepsilon N$ (for fixed $\varepsilon>0$ and $\delta<1$). If we allow a variation of any number of elements of the matrix, but with preservation of sign, then we arrive at the important notion of a sign-rank. For more details on these quantities, see the survey [25].

In connection with problems of communication complexity theory, the following quantity was studied in [3, 22]. Let $A=(A_{i,j})$ be a matrix and $\varepsilon>0$ a number. The *approximate ε-rank* (or simply the *ε-rank*) is the quantity

$$\operatorname{rank}_\varepsilon(A)=\min\Big\{\operatorname{rank} B\colon \max_{i,j}|A_{i,j}-B_{i,j}|\le\varepsilon\Big\}.$$

The approximate rank is, in a sense, simpler than the rigidity or the sign-rank; moreover, it reduces to the Kolmogorov width. For an $N\times M$ matrix A, define $W_A\subset\mathbb{R}^N$ to be the set of vectors A^j given by the columns of the matrix A ($j=1,\dots,M$). Then the functions $\operatorname{rank}_\varepsilon$ and d_n are mutually inverse: for any $n\in\mathbb{N}$ and $\varepsilon>0$, we have

$$\operatorname{rank}_\varepsilon(A)\le n \qquad\Leftrightarrow\qquad d_n(W_A,\ell_\infty^N)\le\varepsilon.$$

[a] Steklov Mathematical Institute of Russian Academy of Sciences, ul. Gubkina 8, Moscow, 119991 Russia.
[b] Laboratory "High-Dimensional Approximation and Applications," Lomonosov Moscow State University, Moscow, 119991 Russia.

E-mail addresses: kashin@mi-ras.ru (B.S. Kashin), malykhin@mi-ras.ru (Yu.V. Malykhin), kriutin@yahoo.com (K.S. Ryutin).

In particular, the problem of the ε-rank of the identity matrix is equivalent to the well-known problem of the width of an octahedron in ℓ_∞^N. For example, the lower bounds obtained in [1] in terms of matrices follow from the estimates of widths established by Gluskin [9].

We also mention the paper [20], where the authors actually consider the width of special classes of Boolean functions.

Notation. Below, we denote by $c, c_1, \dots, C, C_1, \dots$ positive absolute constants (which are different in different formulas). Dependence on some parameters is pointed out explicitly, for example, $c(d)$. The relation $F \ll G$ means that $F \le CG$ for all values of the arguments on which F and G depend. The relation $F \asymp G$ means that $F \ll G$ and $G \ll F$.

Denote by $|\cdot|$ the Euclidean length of a vector, and by $\|\cdot\|$ the operator norm of a matrix: $\|A\| = \max_{|x|=1}|Ax|$.

We use standard notation of probability theory: P is probability, E is expectation, and Var is variance. For example, if $T = T(\omega; x)$ is a random function of an argument $x \in [0, 2\pi]$, then $\|T\|_p$ denotes the standard L_p norm $\bigl(\int_0^{2\pi}|T(\omega;x)|^p\,dx\bigr)^{1/p}$, and its mean value is thus equal to $\mathsf{E}\|T\|_p$.

The width of the skew octahedron. At present, the orders of the widths of the Sobolev classes of functions of one variable, $d_n(W_p^r, L_q)$, $r \in \mathbb{N}$, $1 \le p, q \le \infty$, are known (see [14]) in all cases except that of $r = 1$, $p = 1$, and $2 < q < \infty$. In this case, the following estimate is known [21]:

$$c(q,\varepsilon)n^{-1/2}\log^{1/2-\varepsilon} n \le d_n(W_1^1, L_q) \le C(q)n^{-1/2}\log n \qquad \forall \varepsilon > 0. \tag{1.1}$$

In the last section of the paper, we slightly sharpen the lower bound in (1.1) and establish

Proposition 1. *For $q \in (2, \infty)$ and $n \in \mathbb{N}$, we have*

$$d_n(W_1^1, L_q) \ge c(q)n^{-1/2}\log^{1/2} n.$$

It is easy to see that the width of the class $W_1^1 = \{f \colon \|f'\|_1 \le 1\}$ is of the same order as the width of the family of step functions $\{\chi_t\}_{0<t<1}$, where $\chi_t(x) = 1$ for $0 \le x \le t$ and $\chi_t(x) = 0$ for $t < x \le 1$.

As early as 2003, V. N. Konovalov pointed out that the width $d_n(W_1^1, L_q)$ admits a natural discretization (the restriction of functions to a uniform N-point grid), namely,

$$d_n(W_1^1, L_q) \asymp N^{-1/q} d_n\bigl(\widetilde{B}_1^N, \ell_q^N\bigr) \qquad \text{for} \quad N > n^{q/2}.$$

Here $\widetilde{B}_1^N$ is the *skew octahedron*, which is the convex hull of the vectors $\pm(1, 1, \dots, 1, 0, 0, \dots, 0)$, i.e., discrete analogs of the step functions χ_t.

Thus, we arrive at the problem of finding the width of a skew octahedron. In our opinion, this problem is of significant independent interest. In the present study, we consider the ℓ_∞ metric; instead of the parameter q, we have a parameter N, which is not related to n.

It is clear that the skew octahedron corresponds to the triangular matrix $\Delta^{(N)} = (\Delta_{i,j})_{1\le i,j\le N}$ with $\Delta_{i,j} = 1$ for $i \le j$ and $\Delta_{i,j} = 0$ for $i > j$:

$$\operatorname{rank}_\varepsilon\bigl(\Delta^{(N)}\bigr) \le n \qquad \Leftrightarrow \qquad d_n\bigl(\widetilde{B}_1^N, \ell_\infty^N\bigr) \le \varepsilon.$$

The problem of the ε-rank of a triangular matrix was considered in [3]. It was pointed out there, in particular, that the known results imply the estimates

$$c\log^2 N \le \operatorname{rank}_{1/3}\bigl(\Delta^{(N)}\bigr) \le C\log^3 N. \tag{1.2}$$

In Section 2, we present the proof of (1.2) which was proposed in [3]. Unfortunately, we could not improve these estimates.

In Section 3, we present a simpler lower bound which is applicable in a more general case. Namely, instead of step functions, we consider standard d-dimensional parallelepipeds $\pi_{\mathbf{s}}^{\mathbf{t}}$ (see Section 3). We establish a d-dimensional analog of (1.2): for $n_1 \asymp \log^{2d+1} N$, $n_2 \asymp \log^{2d} N$, and $0 < c(d) < 1/3$, we have

$$d_{n_1}\big(\{\pi_{\mathbf{s}}^{\mathbf{t}}\}, \ell_\infty^{N^d}\big) \le c(d) \le d_{n_2}\big(\{\pi_{\mathbf{s}}^{\mathbf{t}}\}, \ell_\infty^{N^d}\big).$$

In Sections 4 and 5, we consider a special type of approximation, namely, trigonometric approximation.

Often it is more convenient to consider sign-matrices, i.e., matrices with elements taking the values ± 1. Denote by $\widehat{\Delta}^{(N)}$ the matrix obtained from $\Delta^{(N)}$ by replacing the zeros below the diagonal with -1.

2. ESTIMATE FOR THE WIDTH IN TERMS OF THE γ_2-NORM

In [23, 24], a factorization γ_2-norm was proposed as one of the complexity measures of matrices; this norm has turned out to be useful, in particular, in dealing with the ε-rank and the sign-rank [3, 22].

Let us give a definition. For a matrix $A = (A_{i,j})$, denote by $\gamma_2(A)$ the infimum of those $c > 0$ for which there exist vectors x_i and y_j (in some Euclidean space) such that $A_{i,j} = \langle x_i, y_j\rangle$ and $\max_{i,j} |x_i| \cdot |y_j| \le c$.

For the triangular matrix $\Delta^{(N)}$, we have the relation

$$c\log N \le \gamma_2\big(\Delta^{(N)}\big) \le C\log N. \tag{2.1}$$

The simplest way to derive this relation is to use the well-known equality of the γ_2-norm and the norm of the Schur multiplier:

$$\gamma_2(A) = \max_{\|B\|\le 1} \|A \circ B\|,$$

where $A \circ B = (A_{i,j}B_{i,j})_{i,j}$ is the Schur–Hadamard multiplication. For example, we can estimate $\gamma_2(\Delta^{(N)})$ from below by using the Hilbert matrix $H = (H_{ij})$ as B, with

$$H_{i,j} = \begin{cases} \dfrac{1}{i-j}, & i \ne j, \\ 0, & i = j. \end{cases} \tag{2.2}$$

It is well known [11] that $\|H\| \le \pi$; however, it is easy to see that $\|H \circ \Delta^{(N)}\| \ge c\log N$. (Note that for the dyadic analog of Hilbert matrices (see [18]), after the Schur–Hadamard multiplication by $\Delta^{(N)}$, the norm increases only by a factor of $c\log^{1/2} N$.)

Lower bound on the width. The definition of the γ_2-norm is analogous to the definition of rank, except that here, instead of the dimension of vectors, their length is restricted. Below we will need the following well-known (see [23]) relation between these quantities:

$$\gamma_2(A) \le \sqrt{\operatorname{rank} A} \cdot \max_{i,j} |A_{i,j}|. \tag{2.3}$$

Let A be an arbitrary sign-matrix. Denote by $m(A)$ the least upper bound of those $c > 0$ for which there exist unit-length vectors x_i and y_j such that $\operatorname{sign}\langle x_i, y_j\rangle = A_{i,j}$ and $|\langle x_i, y_j\rangle| \ge c$. It is clear from the definitions that

$$m(\operatorname{sign} B_{i,j}) \ge \frac{\min_{i,j}|B_{i,j}|}{\gamma_2(B)}$$

for any matrix B with nonzero elements.

In [7], the quantity m was exactly calculated for the matrix $\widehat{\Delta}^{(N)}$: it turned out, in particular, that $m(\widehat{\Delta}^{(N)}) \asymp \log^{-1} N$. Moreover, it was actually proved that $\gamma_2(\widehat{\Delta}^{(N)}) = 1/m(\widehat{\Delta}^{(N)})$. It follows from these results and inequality (2.3) that $\operatorname{rank}_\varepsilon(\widehat{\Delta}^{(N)}) \ge c(\varepsilon)\log^2 N$ for $\varepsilon < 1$.

Upper bound on the width. In [3, 22], the following arguments were presented. Let A be an $N \times N$ matrix and $A_{i,j} = \langle x_i, y_j\rangle$. Let us map the vectors x_i and y_j to an n-dimensional space by a random $n \times N$ matrix U, as in the Johnson–Lindenstrauss lemma. For a suitable choice of n and U, $\langle x_i, y_j\rangle \approx \langle Ux_i, Uy_j\rangle$ with large probability, and the original matrix will be approximated by a matrix of rank at most n. For the exact formulation, we use the following recent result (see [2, Theorem 1.2]):

For any vectors $x_1,\dots,x_N,y_1,\dots,y_N \in \mathbb{R}^N$ of length at most 1 and for numbers $0<\varepsilon<1$ and $t = \lfloor C\varepsilon^{-2}\log(2+N\varepsilon^2)\rfloor$, there exist vectors $u_1,\dots,u_N,v_1,\dots,v_N \in \mathbb{R}^t$ such that the inequality $|\langle x_i,y_j\rangle - \langle u_i,v_j\rangle| \le \varepsilon$ holds for all i and j.

In terms of the ε-rank, we obtain the following corollary:

$$\operatorname{rank}_\varepsilon(A) \le C\varepsilon^{-2}\gamma_2^2(A)\log\bigl(2+N\varepsilon^2\gamma_2^{-2}(A)\bigr).$$

Applying this corollary and inequalities (2.1) and making rough estimates, we find $\operatorname{rank}_\varepsilon(\Delta^{(N)}) \le C\varepsilon^{-2}\log^3 N$, or, in terms of widths,

$$d_n\bigl(\widetilde{B}_1^N, \ell_\infty^N\bigr) \le Cn^{-1/2}\log^{3/2} N. \tag{2.4}$$

One can show that for $n \asymp N$ the logarithm can be dropped.

Proposition 2. $d_{N/2}(\widetilde{B}_1^N, \ell_\infty^N) \asymp N^{-1/2}$.

Indeed, using the space of piecewise constant vectors of dimension $N/4$, we "truncate" the step functions to vectors of the form $(0,\dots,0,1,\dots,1,0,\dots,0)$ with at most four ones, whose ℓ_1 norm is therefore not greater than 4. Next, we apply the estimate for the width of an octahedron [13],

$$d_n\bigl(B_1^N, \ell_\infty^N\bigr) \le C(\alpha)n^{-1/2}, \qquad n > N^\alpha, \quad \alpha \in (0,1),$$

where B_p^N, $1 \le p \le \infty$, is the unit ball of the space ℓ_p^N.

3. ESTIMATE FOR THE WIDTHS AND ORTHOMASSIVITY

In [16], the so-called characteristic of orthomassivity of a set K lying in the unit ball of a Hilbert space $\mathcal{H}$ was introduced:

$$\mathrm{OM}_n(K) = n^{-1/2} \sup_{\{\varphi_j\}_1^n} \sup_{\{f_j\}_1^n \subset K} \sum_{j=1}^n \langle \varphi_j, f_j\rangle, \tag{3.1}$$

where the first supremum is taken over all orthonormal systems $\{\varphi_j\}_{j=1}^n$. It is straightforward that instead of the orthonormal system $\{\varphi_j\}_{j=1}^n$ in the definition (3.1), one can take arbitrary sets of elements $\{\varphi_j\}_{j=1}^n \subset \mathcal{H}$ with

$$\|\{\varphi_j\}\| := \sup_{\sum a_j^2 \le 1} \Bigl\| \sum_{j=1}^n a_j\varphi_j \Bigr\|_{\mathcal{H}} \le 1.$$

The orders (as $n \to \infty$) of the quantities (3.1) for the family of step functions $\{\chi_t\}$ and the family of characteristic functions of the d-dimensional parallelepipeds $[0,t_1]\times[0,t_2]\times\ldots\times[0,t_d]$ were found in [17].

The arguments related to orthomassivity allow one to directly estimate the width of interest from below, without using (2.3), and, moreover, to generalize the arguments to the multidimensional case.

Let us fix the dimension $d \in \mathbb{N}$. For $N \in \mathbb{N}$, we pass to a discrete problem by replacing the function $f\colon [0,1]^d \to \mathbb{R}$ with its values on the grid $\{(i_1/N, \dots, i_d/N)\colon 0 \le i_k < N\}$. The obtained vectors (of the values of f) will lie in the space $\mathbb{R}^{N^d}$. We will denote the coordinates in this space as follows:

$$x[i_1, \dots, i_d], \qquad 0 \le i_k < N, \quad k = 1, \dots, d.$$

Under discretization of the characteristic functions of standard parallelepipeds, we obtain vectors $\pi_{\mathbf{s}}^{\mathbf{t}}$ with $\mathbf{s} = (s_1, \dots, s_d)$ and $\mathbf{t} = (t_1, \dots, t_d)$, $s_j, t_j \in \{0, \dots, N-1\}$, where

$$\pi_{\mathbf{s}}^{\mathbf{t}}[i_1, \dots, i_d] = \begin{cases} 1 & \text{if } s_k \le i_k < t_k,\ k = 1, \dots, d, \\ 0 & \text{otherwise.} \end{cases}$$

These vectors are the tensor products of one-dimensional step functions:

$$\pi_{\mathbf{s}}^{\mathbf{t}} = \pi_{s_1}^{t_1} \otimes \dots \otimes \pi_{s_d}^{t_d}. \tag{3.2}$$

For short, we set $\pi^{\mathbf{t}} = \pi_{(0,\dots,0)}^{\mathbf{t}}$.

Proposition 3. *If* $d_n(\{\pi^{\mathbf{t}}\}, \ell_\infty^{N^d}) < 2^{-d-1}$, *then* $n \ge C(d) \log^{2d} N$.

Proof. Consider the dth tensor power $H^{\otimes d}$ of the Hilbert matrix:

$$H^{\otimes d}_{(i_1,\dots,i_d),(j_1,\dots,j_d)} = H_{i_1,j_1} \dots H_{i_d,j_d}.$$

Denote the ith row of the Hilbert matrix by H_i (see (2.2)). In the one-dimensional case, it is easy to see that

$$\langle H_t, \pi^t \rangle \ge \frac{1}{2} \|H_t\|_1 \asymp \log N$$

if $t > N/2$. In the multidimensional case, in view of (3.2) we have

$$\langle H_{\mathbf{t}}, \pi^{\mathbf{t}} \rangle = \langle H_{t_1}, \pi^{t_1} \rangle \dots \langle H_{t_d}, \pi^{t_d} \rangle \ge 2^{-d} \|H_{t_1}\|_1 \dots \|H_{t_d}\|_1 = 2^{-d} \|H_{\mathbf{t}}\|_1 \asymp \log^d N \tag{3.3}$$

if $\mathbf{t} \in \mathcal{T} := \{\mathbf{t}\colon t_1, \dots, t_d > N/2\}$.

Suppose that we have approximated the vectors $\pi^{\mathbf{t}}$ by vectors $g^{\mathbf{t}}$ from an n-dimensional space L_n with an error less than 2^{-d-1}. Then it follows from (3.3) that

$$\langle H_{\mathbf{t}}, g^{\mathbf{t}} \rangle \ge \langle H_{\mathbf{t}}, \pi^{\mathbf{t}} \rangle - 2^{-d-1} \|H_{\mathbf{t}}\|_1 \ge 2^{-d-1} \|H_{\mathbf{t}}\|_1 \asymp \log^d N.$$

Summing over $\mathbf{t}$, we obtain

$$\sum_{\mathbf{t} \in \mathcal{T}} \langle H_{\mathbf{t}}, g^{\mathbf{t}} \rangle \ge C(d) N^d \log^d N.$$

On the other hand, $|g^{\mathbf{t}}|^2 \le 2N^d$ and

$$\sum_{\mathbf{t} \in \mathcal{T}} \langle H_{\mathbf{t}}, g^{\mathbf{t}} \rangle \ll N^d \biggl(\sum_{\mathbf{t} \in \mathcal{T}} |P_{L_n} H_{\mathbf{t}}|^2 \biggr)^{1/2},$$

where P_L is the orthogonal projection onto the space L.

Let $\{\varphi_j\}_{j=1}^n$ be an orthonormal basis of L_n. Then

$$\sum_{\mathbf{t} \in \mathcal{T}} |P_{L_n} H_{\mathbf{t}}|^2 = \sum_{\mathbf{t} \in \mathcal{T}} \sum_{j=1}^n \langle H_{\mathbf{t}}, \varphi_j \rangle^2 = \sum_{j=1}^n |H^{\otimes d} \varphi_j|^2 \ll n,$$

where we used the fact that $\|H^{\otimes d}\| = \|H\|^d \le C(d)$.

Hence,

$$c N^d \log^d N \le \sum_{\mathbf{t} \in \mathcal{T}} \langle H_{\mathbf{t}}, y^{\mathbf{t}} \rangle \ll N^d \sqrt{n},$$

which implies the required estimate for n. $\square$

4. TRIGONOMETRIC WIDTH

Let $X = (\mathbb{C}^N, \|\cdot\|)$ be an N-dimensional normed space over the field $\mathbb{C}$ and $A \subset X$. The trigonometric width differs from the Kolmogorov width in that the approximating spaces are generated by elements of the trigonometric system:

$$d_n^{\mathrm{T}}(A, X) := \inf_{0 \le k_1, \dots, k_n < N} \sup_{x \in A} \inf_{c_1, \dots, c_n \in \mathbb{C}} \left\| x - \sum_{j=1}^n c_j e_{k_j} \right\|_X,$$

where e_k is the discrete exponential $(\exp(2\pi ikj/N))_{j=0}^{N-1}$. Obviously, the trigonometric width provides an upper bound for the Kolmogorov width: if $A \subset \mathbb{R}^N$, then $d_{2n}(A, X^{\mathbb{R}}) \le d_n^{\mathrm{T}}(A, X^{\mathbb{C}})$.

Define also the n-term approximation

$$\sigma_n(x)_X := \inf_{\substack{0 \le k_1, \dots, k_n < N \\ c_1, \dots, c_n \in \mathbb{C}}} \left\| x - \sum_{j=1}^n c_j e_{k_j} \right\|_X.$$

The definition of the trigonometric width was proposed by Ismagilov for the continuous case in [12]. Note that for the class W_1^1 in L_q, the rate of decay of the trigonometric width, in contrast to the Kolmogorov width (see (1.1)), is known: Belinskii [5] proved that

$$d_n^{\mathrm{T}}(W_1^1, L_q) \asymp n^{-1/2} \log n, \qquad 2 < q < \infty.$$

It is easy to show that the trigonometric width of the skew octahedron reduces to an n-term approximation of one function, for example, the sawtooth function $f_0(x) := \sum_{k=1}^\infty k^{-1} \sin kx$. It is well known that $f_0(x) = (\pi - x)/2$ for $x \in (0, 2\pi)$, which implies that

$$d_n^{\mathrm{T}}(\widetilde{B}_1^N, \ell_\infty^N) \approx \sigma_n\Big(f_0\big|_{\{2\pi j/N\}_{j=0}^{N-1}}\Big)_{\ell_\infty^N}.$$

(Here the relation $F(n, N) \approx G(n, N)$ means that $F(n, N) \le c_1 G(c_2 n, c_3 N) \le c_4 F(c_5 n, c_6 N)$.)

Consider the problem of approximation of a function f_0 on a grid in greater detail. Recall the inequality from [6]:

Let $Y \subset B_\infty^N$, $|Y| = M$, $x \in \mathbb{C}^N$, and $\|x\|_Y := \max_{y \in Y} |\langle x, y\rangle|$. Then, for any $x \in \mathbb{C}^N$,

$$\min_{\|x^*\|_0 \le n} \|x - x^*\|_Y \le Cn^{-1/2} \log^{1/2}\left(2 + \frac{M}{n}\right) \cdot \|x\|_1,$$

where $\|\cdot\|_0$ is the number of nonzero coordinates.

In trigonometric coordinates, we obtain

$$\sigma_n(x)_{\ell_\infty^N} \le Cn^{-1/2} \log^{1/2}\left(2 + \frac{N}{n}\right) \|x\|_A, \qquad \|x\|_A := \sum_{k=0}^{N-1} |\langle x, e_k\rangle|.$$

(Note that the problems of m-term trigonometric approximations are considered in more detail in the book [28] and in the recent paper [29].) The discrete A-norm of the function f_0 is of order $\log N$; therefore, the above inequality implies the following estimate.

Proposition 4. *$d_n^{\mathrm{T}}(\widetilde{B}_1^N, \ell_\infty^N) \le Cn^{-1/2} \log^{3/2} N$, and similarly in the multidimensional case $d_n(\{\pi_{\mathbf{s}}^{\mathbf{t}}\}, \ell_\infty^{N^d}) \le C(d) n^{-1/2} \log^{d+1/2} N$.*

5. SELECTORS

It is well known that the order of approximation of ℓ_p-balls in the ℓ_q metric is attained in most cases on random subspaces (see [8, 14]). It is straightforward that such subspaces give poor approximation for the skew octahedron. We consider approximation by other random objects and find the limits of such an approximation method.

In this section, we will focus on approximations with a fixed (small) error. It is known that for $a>0$

$$\left\|f_0(x)-\sum_{k=1}^{aN}\frac{\sin kx}{k}\right\|_{\ell_\infty(\{2\pi j/N\}_{j=0}^{N-1})}\le\max_{2\pi/N\le|x|<\pi}\left|\sum_{k>aN}\frac{\sin kx}{k}\right|<\frac{1}{2a}.$$

Hence, for a sufficiently large a, we can replace the function $f_0(x)$ by a polynomial $\sum_{k=1}^{aN}k^{-1}\sin kx$; everywhere below, we will consider this polynomial.

We can approximate $x=\sum_k x_k e_k$ ($\{e_k\}$ is a system of exponentials or another system) by a sparse sum as follows: take a set of "selectors," i.e., random variables $\{\xi_k(\omega)\}$ such that $\mathsf{P}(\xi_k=0)=1-\delta_k$ and $\mathsf{P}(\xi_k=\delta_k^{-1})=\delta_k$ (thus, $\mathsf{E}\,\xi_k=1$). If the norm $\|\sum_k x_k(1-\xi_k)e_k\|$ is small with large probability, then x is approximated by the sum $\sum_{k\colon \xi_k\ne0}x_k\xi_k e_k$ (which is sparse if δ_k are small).

In our case, it suffices to estimate

$$\left\|\sum_{k=1}^{aN}(1-\xi_k)\frac{\sin kx}{k}\right\|_{\ell_\infty(\{2\pi j/N\}_{j=0}^{N-1})} \tag{5.1}$$

by taking appropriate δ_k. Set $\delta_k=1$ (hence, $\xi_k\equiv1$; i.e., we always preserve this harmonic) for $k<k_0:=C\log^2N$ and $\delta_k=k_0/k$ for other k. Let us fix a grid point $x_j=2\pi j/N$. Applying the standard Bernstein inequality for large deviations (in the necessary form, see [27, Lemma 4.3.4]), we obtain

$$\mathsf{P}\left(\left|\sum_{k=k_0}^{aN}(1-\xi_k)\frac{\sin kx_j}{k}\right|\ge t\right)\le 2\exp\left(-\min\left\{\frac{t}{2M},\frac{t^2}{4\sigma^2}\right\}\right),$$

where $M:=\max_k\|(1-\xi_k)\sin kx_j/k\|_\infty\le\max_{k\ge k_0}(k\delta_k)^{-1}=k_0^{-1}$ and

$$\sigma^2:=\sum_k\mathsf{E}\left|(1-\xi_k)\frac{\sin kx_j}{k}\right|^2\le\sum_k k^{-2}\,\mathsf{E}(1-\xi_k)^2\le\sum_{k\ge k_0}(k^2\delta_k)^{-1}\le\frac{\log(aN)}{k_0}.$$

Set $t:=1/3$; then $t/(2M)\gg\log^2N$ and $t^2/(4\sigma^2)\ge ck_0/\log(aN)>2\log N$ (due to the choice of C in the definition of k_0); therefore, the estimated probability is less than $2N^{-2}$. Hence, with probability at least $1-cN^{-1}$, the norm (5.1) is less than $1/3$. It remains to notice that also

$$\#\{k\colon \xi_k\ne0\}\le 2\,\mathsf{E}\big(\#\{k\colon \xi_k\ne0\}\big)=\sum_k\delta_k\asymp k_0\log N\asymp\log^3N$$

with large probability. Thus, we have nontrivially approximated the polynomial $\sum_{k=1}^{aN}k^{-1}\sin kx$ using $O(\log^3N)$ harmonics and proved the upper bound in (1.2) by another method.

Note that we have actually approximated our polynomial in the L_∞ metric on the entire circle. Indeed, instead of the uniform N-point grid, we could take a grid of $2aN$ points, and for polynomials of degree aN such a grid norm is equivalent to the standard L_∞ norm (see [30, Vol. 2, Ch. X, § 7]). Below, for simplicity, we consider approximation in L_∞; we can also set $a=1$.

Below we prove that using the arbitrary selectors described above (and even using a slightly more general random method), one cannot obtain a better estimate, i.e., one cannot approximate $\sum_{k=1}^{N}k^{-1}\sin kx$ in L_∞ by $o(\log^3N)$ harmonics with large probability. We conjecture that in this case a good approximation is impossible in principle, i.e., the width $d_n^{\mathrm T}(\widetilde B_1^N,\ell_\infty^N)$ does not tend to zero if $n=o(\log^3N)$.

Proposition 5. *Let $(\eta_k)_{k=1}^N$ be independent random variables with $\mathsf{E}\,\eta_k=1$ such that*

$$\mathsf{E}\,\eta_k^4\le C\delta_k^{-3},\qquad \text{where}\quad \delta_k:=\mathsf{P}(\eta_k\ne0).$$

Then

$$\text{either} \quad \sum_{k=1}^{N} \delta_k \ge c\log^3 N \qquad \text{or} \quad \mathsf{E}\left\|\sum_{k=1}^{N} \frac{\sin kx}{k} - \sum_{k=1}^{N} \eta_k \frac{\sin kx}{k}\right\|_\infty \ge c.$$

It is clear that the hypotheses of the proposition are satisfied if η_k takes exactly two values and $\mathsf{E}\,\eta_k = 1$.

Proof. Suppose that $\sum \delta_k \le c\log^3 N$ and

$$\mathsf{E}\left\|\sum_{k=1}^{N} \frac{\sin kx}{k} - \sum_{k=1}^{N} \eta_k \frac{\sin kx}{k}\right\|_\infty \le c, \tag{5.2}$$

where c is sufficiently small. Applying the estimate for the uniform norm of the polynomials $\sum_{k=1}^{N} \sin kx/k$ (see [30, Vol. 1, Ch. II, § 9]), we then have

$$\mathsf{E}\left\|\sum_{k=1}^{N} \eta_k \frac{\sin kx}{k}\right\|_\infty \le 4. \tag{5.3}$$

Let us prove that $\sum_{k=\lfloor\sqrt{N}\rfloor}^{N} \delta_k \ge C_1 \log^3 N$. (Note that the restrictions on η_k are needed only for $k \ge \sqrt{N}$.)

Notice that $\mathsf{E}\,\eta_k^2 \asymp \delta_k^{-1}$ and $\mathsf{E}\,\eta_k^4 \asymp \delta_k^{-3}$. Indeed, we can consider a random variable ρ_k with conditional distribution $\mathsf{Law}\,\rho_k = \mathsf{Law}(\eta_k \,|\, \eta_k \ne 0)$. It follows from the conditions $\mathsf{E}\,\rho_k = \delta_k^{-1}$ and $\mathsf{E}\,\rho_k^4 \ll \delta_k^{-4}$ that the L_p norms of ρ_k are equivalent for $1 \le p \le 4$. Hence, $\mathsf{E}\,\eta_k^2 \asymp \delta_k^{-1}$ and $\mathsf{E}\,\eta_k^4 \asymp \delta_k^{-3}$.

Regularization. Let us construct a sufficiently large subset of indices I such that

$$c\frac{\log N}{k} \le \delta_k \le \frac{\log^2 N}{k}, \qquad k \in I. \tag{5.4}$$

Consider the dyadic blocks $\{2^{s-1}, \dots, 2^s - 1\}$ for $s > (1/2)\log_2 N$. First, we show that in each block there are at most $1/8$ of indices k such that $\delta_k < c/k$. Indeed, otherwise, for the corresponding indices $k_1, \dots, k_l$, $l \ge 2^{s-4}$, the event $A = \{\eta_{k_1} = 0, \dots, \eta_{k_l} = 0\}$ has probability $P(A) = \prod(1 - \delta_{k_j}) \ge (1 - c/2^s)^{2^{s-4}} > c_1$. Then

$$\mathsf{E}\left\|\sum_{k=1}^{N} \frac{\sin kx}{k} - \sum_{k=1}^{N} \eta_k \frac{\sin kx}{k}\right\|_\infty \gg \mathsf{E}\left(\left\|\sum_{k=1}^{N} \frac{\sin kx}{k} - \sum_{k=1}^{N} \eta_k \frac{\sin kx}{k}\right\|_\infty \,\middle|\, A\right)$$
$$= \mathsf{E}\left\|\sum_{k=1}^{N} \frac{\sin kx}{k} - \sum_{k\ne k_j} \eta_k \frac{\sin kx}{k}\right\|_\infty \ge \left\|\mathsf{E}\left(\sum_{k=1}^{N} \frac{\sin kx}{k} - \sum_{k\ne k_j} \eta_k \frac{\sin kx}{k}\right)\right\|_\infty = \left\|\sum_{j=1}^{l} \frac{\sin k_j x}{k_j}\right\|_\infty.$$

The value of the last function is not less than c_2 at the point $x = 2^{-s-1}\pi$, which contradicts (5.2).

In the sth dyadic block, we distinguish a subset I_s of cardinality 2^{s-3} formed by the indices of those δ_k that have middle values (i.e., we arrange the numbers $\{\delta_k\}_{k=2^{s-1}}^{2^s-1}$ in increasing order: $\delta_1^* \le \dots \le \delta_{2^{s-1}}^*$, take the numbers δ_j^* with $3\cdot 2^{s-4} \le j \le 5\cdot 2^{s-4}$, and take the indices of the corresponding δ_k). Then $\delta_k < \log^2 N/2^s$ for all $k \in I_s$ in at least $2/3$ of the blocks: otherwise the sum of all δ_k is greater than $c\log^3 N$, and we are done. Moreover, in at least $2/3$ of the blocks, we have $\delta_k > c\log N/k$, $k \in I_s$. Indeed, if this is not the case in the sth block, then, for at least $1/8$ of the indices in this block, we have $1/\delta_k > k/(c\log N)$ and $\delta_k > c/k$; denote this set of indices by J_s. It follows that

$$\sum_{k\in J_s} \frac{1}{k^2\delta_k} \gg 2^s \frac{1}{2^{2s}} \frac{2^s}{c\log N} \asymp (c\log N)^{-1}.$$

Summing over the blocks (with the number of blocks being of order $\log N$), we find that the sum $\sum(k^2\delta_k)^{-1}$ over $k \in J := \bigcup J_s$ is large enough. However, then there cannot be a good approximation even in L_2, because

$$\mathsf{E}\left\|\sum_{k\in J}\eta_k\frac{\sin kx}{k}\right\|_2^2 \asymp \sum_{k\in J}\frac{1}{k^2\delta_k} \ge c_0.$$

Having estimated the mean of the squared norm, we can now estimate the required mean value of the norm. Set $X := \left\|\sum_{k\in J}\eta_k \sin kx/k\right\|_2^2$. It is known that $\mathsf{E}\,X \ge c_0$. The required assertion follows from the elementary inequality

$$\mathsf{P}\left(\xi \ge \frac{1}{2}\,\mathsf{E}\,\xi\right) \ge \frac{(\mathsf{E}\,\xi)^2}{4\,\mathsf{E}\,\xi^2} \qquad \text{if} \quad \xi \ge 0 \ \text{ a.s.} \tag{5.5}$$

and an estimate for $\mathsf{E}\,X^2$. We have $\mathsf{E}\,X^2 = \mathsf{Var}\,X + (\mathsf{E}\,X)^2$; since $\delta_k > c/k$, we find

$$\mathsf{Var}\,X \le \sum_{k\in J}\frac{\mathsf{E}\,\eta_k^4}{k^4} \asymp \sum_{k\in J}\delta_k^{-3}k^{-4} \ll \sum_{k\in J}\delta_k^{-1}k^{-2} = \mathsf{E}\,X \ll (\mathsf{E}\,X)^2.$$

Thus, there exists a set S of blocks with numbers $s > (1/2)\log N$, $|S| \asymp \log N$, such that (5.4) is satisfied for $k \in I_s$, $s \in S$. Finally, we set $I := \bigcup_{s\in S} I_s$.

Symmetrization. Consider the sum (5.3). For $k \in I$, we replace η_k in the kth term by an independent copy η'_k:

$$\mathsf{E}\left\|\sum_{k\notin I}\eta_k\frac{\sin kx}{k} + \sum_{k\in I}\eta'_k\frac{\sin kx}{k}\right\|_\infty \le C_1. \tag{5.6}$$

Subtracting inequality (5.3) from (5.6), we obtain

$$\mathsf{E}\left\|\sum_{k\in I}(\eta_k - \eta'_k)\frac{\sin kx}{k}\right\|_\infty \le C_2. \tag{5.7}$$

Everywhere below, we will consider $k \in I$.

QC-*norm.* We will need a result from [19]. Denote by $\Delta_s f$ the sth block of the Fourier series, i.e.,

$$\Delta_0 f = \widehat{f}(0), \qquad \Delta_s f := \sum_{2^{s-1}\le|k|<2^s}\widehat{f}(k)e^{ikx}, \quad s \ge 1.$$

In [19], the norm $\|f\|_{\mathrm{QC}} := \mathsf{E}\left\|\sum_{s=0}^{\infty}(\pm\Delta_s f)\right\|_\infty$ is defined, where $\pm$ are random signs.

Proposition [19]**.** *The following inequality holds for any real function $f \in L^1[0,2\pi]$:*

$$\|f\|_{\mathrm{QC}} \ge \frac{1}{16}\sum_{s=0}^{\infty}\|\Delta_s f\|_1.$$

Let us apply this proposition to the random function $f = \sum_{k\in I}(\eta_k - \eta'_k)\sin kx/k$ from (5.7). Since the multiplication of blocks (and even the multiplication of individual terms!) by random signs does not change the distribution of f, we obtain

$$\mathsf{E}\|f\|_\infty = \mathsf{E}\|f\|_{\mathrm{QC}} \gg \sum_{s\in S}\mathsf{E}\left\|\sum_{k\in I_s}(\eta_k - \eta'_k)\frac{\sin kx}{k}\right\|_1. \tag{5.8}$$

Estimate of the L_1 and L_2 norms. Fix an $s\in S$ and consider a term in the sth block:

$$T_s:=\sum_{k\in I_s}(\eta_k-\eta_k')\frac{\sin kx}{k}.$$

Let us prove the following relations:

$$\mathsf{P}\big(\|T_s\|_1\gg\|T_s\|_2\big)=1-o(1), \tag{5.9}$$

$$\mathsf{P}\big(\|T_s\|_2\gg(\mathsf{E}\|T_s\|_2^2)^{1/2}\big)\ge c>0. \tag{5.10}$$

For the inequality $\|T_s\|_1\gg\|T_s\|_2$ to hold, it suffices to have the relation $\|T_s\|_2\asymp\|T_s\|_4$; the latter will hold (see [4, Ch. IX, § 10]) if the set Λ_s of numbers of nonzero Fourier coefficients of T_s is a Sidon set, i.e., if there are no nontrivial equalities of the form

$$k_1+k_2=k_3+k_4,\qquad k_i\in\Lambda_s,\quad i=1,\dots,4.$$

However, the probability of such a relation is small (for simplicity, we assume that k_j are different):

$$\mathsf{P}(k\in\Lambda_s)\asymp\delta_k=\frac{O\big(\log^{O(1)}N\big)}{k},$$

$$\sum_{k_1+k_2=k_3+k_4}\mathsf{P}(k_i\in\Lambda_s)\ll(2^s)^3\left(\frac{\log^{O(1)}N}{2^s}\right)^4\ll\frac{\log^{O(1)}N}{N^{1/2}}=o(1).$$

Consider (5.10). Set $X:=\|T_s\|_2^2=\sum_{k\in I_s}(\eta_k-\eta_k')^2/k^2$. We proceed using (5.5), as before. Then

$$\mathsf{E}\,X=\sum_{k\in I_s}\frac{\mathsf{E}(\eta_k-\eta_k')^2}{k^2}\asymp 2^{-2s}\sum_{k\in I_s}\delta_k^{-1}.$$

Moreover, since $|I_s|\asymp 2^s$, we have $\mathsf{E}\,X\asymp|I_s|^{-1}\sum_{k\in I_s}(k\delta_k)^{-1}$. Next, $\mathsf{E}\,X^2=(\mathsf{E}\,X)^2+\mathsf{Var}\,X$, and

$$\mathsf{Var}\,X\le\sum_{k\in I_s}\mathsf{E}(\eta_k-\eta_k')^4k^{-4}\asymp\frac{1}{|I_s|}\sum_{k\in I_s}\frac{1}{(k\delta_k)^3}.$$

Let us apply inequalities (5.4): we have $(k\delta_k)^{-2}\ll\log^{-2}N$, and the mean value of $(k\delta_k)^{-1}$ is not less than $\log^{-2}N$, so

$$\mathsf{Var}\,X\ll\log^{-2}N\frac{1}{|I_s|}\sum_{k\in I_s}\frac{1}{k\delta_k}\ll\left(\frac{1}{|I_s|}\sum_{k\in I_s}\frac{1}{k\delta_k}\right)^2\asymp(\mathsf{E}\,X)^2.$$

It follows from the relations proved that

$$\mathsf{E}\|T_s\|_1\gg(\mathsf{E}\|T_s\|_2^2)^{1/2}\asymp 2^{-s}\left\{\sum_{k\in I_s}\delta_k^{-1}\right\}^{1/2}.$$

Summing over s in (5.8), we obtain the estimate

$$\sum_{s\in S}2^{-s}\left\{\sum_{k\in I_s}\delta_k^{-1}\right\}^{1/2}\le C.$$

Hence, using the fact that $|S|\asymp\log N$ and $|I_s|\asymp 2^s$, we arrive at the estimate $\sum_{k=1}^N\delta_k\gg\log^3N$. Indeed, by the inequality for the power means, we have $\sum_{s\in S}\big\{\sum_{k\in I_s}\delta_k\big\}^{-1/2}\le C$. Applying it once again (for the exponents $-1/2$ and 1), we obtain the required estimate. $\square$

6. PROOF OF PROPOSITION 1

Let $q > 2$. Let us prove that $d_n(W_1^1, L_q) > c(q)n^{-1/2}\log^{1/2} n$. We will need the following lemma.

Lemma 1. *Let $M = (M_{i,j})_{i,j=1}^N$ be a matrix with* $\operatorname{rank} M \le n \le N/4$. *Then at least one of the following assertions holds*:

(i) *half of elements on the diagonal are far from unity, i.e.,*

$$\#\left\{i\colon\ |M_{i,i} - 1| > \frac{1}{2}\right\} \ge \frac{N}{2};$$

(ii) *the matrix is far from the identity matrix "in the mean," i.e.,*

$$\sum_{i,j=1}^{N} |M_{i,j} - \delta_{i,j}|^2 \ge \frac{cN^2}{n},$$

where $\delta_{i,j} = 1$ for $i = j$ and $\delta_{i,j} = 0$ for $i \ne j$.

Proof. We reduce the matter to the case when all diagonal elements are equal to one. Let $I = \{i\colon |M_{i,i} - 1| \le 1/2\}$. If $|I| \le N/2$, then we are done. In the case of $|I| > N/2$, consider the matrix $(M_{i,j})_{i,j\in I}$. Next, we normalize the rows of this matrix by dividing them by $M_{i,i} \in [1/2, 3/2]$ to obtain one on the diagonal; in this case, the rank of the matrix M is preserved.

So, without loss of generality, we can assume $M_{i,i} = 1$ for all $1 \le i \le N$ and $\operatorname{rank} M = n \le N/2$; let us estimate the sum $\sum_{i,j=1}^N |M_{i,j} - \delta_{i,j}|^2$. Here the arguments used in [15] for an octahedron in ℓ_q apply. Set $d_{i,j} = |M_{i,j} - \delta_{i,j}|^2$. Take a random subset $\Omega \subset \{1,\dots,N\}$ of cardinality $2n$ and consider the projection onto the subspace $\mathbb{R}^\Omega$:

$$\mathsf{E}\sum_{i,j\in\Omega} d_{i,j} = \sum_{i=1}^{N} d_{i,i}\,\mathsf{P}(i\in\Omega) + \sum_{\substack{i,j=1\\ i\ne j}}^{N} d_{i,j}\,\mathsf{P}(i,j\in\Omega).$$

By construction, $d_{i,i} = 0$; it is clear that

$$\mathsf{P}(i,j\in\Omega) = \frac{2n}{N}\,\frac{2n-1}{N-1}, \qquad \text{and hence} \qquad \sum_{i,j} d_{i,j} \gg \left(\frac{N}{n}\right)^2 \mathsf{E}\sum_{i,j\in\Omega} d_{i,j}.$$

For a fixed Ω of cardinality $2n$, we have

$$\sum_{i,j\in\Omega} d_{i,j} = \sum_{i,j\in\Omega} |\delta_{i,j} - M_{i,j}|^2 \ge \sum_{i\in\Omega} d^2(e_i, L_n)_{\ell_2(\Omega)} \ge n,$$

which is a standard estimate for the approximation of the octahedron B_1^{2n}; hence,

$$\sum_{i,j=1,\dots,N} d_{i,j} \gg \left(\frac{N}{n}\right)^2 \mathsf{E}\sum_{i,j\in\Omega} d_{i,j} \gg \frac{N^2}{n}. \quad \square$$

Set

$$\square_{k,i,j} := \left[\frac{i-1}{2^k}, \frac{i}{2^k}\right] \times \left[\frac{j-1}{2^k}, \frac{j}{2^k}\right].$$

Let $p \in (1,\infty)$, $F \in L_p[0,1]^2$, S_kF be a function obtained by averaging F on the squares $\square_{k,i,j}$, $1 \le i,j \le 2^k$, and $Q_kF = S_kF - S_{k-1}F$ (where $S_{-1}F := 0$). Then, as is well known (see [10, Theorem 6.4.7]),

$$\|F\|_q \asymp \|QF\|_q, \qquad \text{where} \quad QF := \left(\sum_{k=0}^{\infty} (Q_kF)^2\right)^{1/2}.$$

Let us reduce the original problem to the "matrix form." Set

$$\Phi(x,t) := f_t(x) := \begin{cases} -1, & x \le t, \\ 1, & x > t. \end{cases}$$

It is clear that $f_t \in 2W_1^1$. Assume that we have almost optimally approximated $f_t \approx g_t$, with $\dim \operatorname{span}\{g_t\}_{t\in[0,1]} \le n$, and set $G(x,t) = g_t(x)$. Then

$$d_n^q(2W_1^1, L_q) \ge 2 \int_0^1 \|f_t - g_t\|_q^q \, dt = 2\|\Phi - G\|_{L_q([0,1]^2)}^q$$

(it is clear that one can always guarantee the measurability of the function G). Let us apply Lemma 1:

$$\|\Phi - G\|_{L_q([0,1]^2)}^q \asymp \left\| \left(\sum_{k=0}^{\infty} (Q_k\Phi - Q_kG)^2 \right)^{1/2} \right\|_q^q = \int_0^1\int_0^1 \left| \sum_{k=0}^{\infty} (Q_k\Phi - Q_kG)^2 \right|^{q/2} dx\, dt$$
$$\ge \sum_{k=0}^{\infty} \int_0^1\int_0^1 |Q_k\Phi - Q_kG|^q \, dx\, dt.$$

Set $N_k := 2^k$. For $F \in L_q([0,1]^2)$, the values of Q_kF on dyadic squares form a matrix $\bigl(F_{i,j}^{(k)}\bigr)_{i,j=1}^{N_k}$:

$$Q_kF = \sum_{i,j=1}^{N_k} F_{i,j}^{(k)} \cdot 1_{\square_{k,i,j}}, \qquad \|Q_kF\|_q^q = N_k^{-2} \sum_{i,j=1}^{N_k} \bigl|F_{i,j}^{(k)}\bigr|^q.$$

Denote by $\Phi^{(k)}$ the matrix of values of $Q_k\Phi$ and by $G^{(k)}$ the matrix of values of Q_kG. It is easy to calculate $\Phi^{(k)}$: the mean value of Φ on the square $\square_{k,i,j}$ is equal to zero on the diagonal (for $i = j$), to one below the diagonal, and to minus one above the diagonal. Hence, $\Phi^{(k)}$ has blocks $\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ on the diagonal, and the other cells are zero. The rank of $G^{(k)}$ is easily seen to be at most n. Denote by π_k a transformation of matrices that maps $\Phi^{(k)}$ into the identity matrix (π_k permutes the columns in pairs and multiplies half of them by -1); since π_k does not change the rank, the rank of $M^{(k)} := \pi_kG^{(k)}$ is also at most n. Finally, it follows from all the aforesaid that

$$d_n^q \gg \sum_{k=0}^{\infty} \|Q_k\Phi - Q_kG\|_q^q = \sum_{k=0}^{\infty} N_k^{-2} \sum_{i,j=1}^{N_k} \bigl|M_{i,j}^{(k)} - \delta_{i,j}\bigr|^q \tag{6.1}$$

for any $q \ge 2$.

Let us proceed to the concluding part of the proof. Given $q > 2$, we fix $q_1 \in (2, q)$ and consider numbers k for which $4n < N_k := 2^k < n^{q_1/2}$. There is a dichotomy: either, for some such number k, the matrix $M^{(k)}$ satisfies property (i) in Lemma 1, or all these matrices satisfy property (ii).

In the first case, we have

$$d_n^q \gg N_k^{-2} \sum_{i,j=1}^{N_k} \bigl|M_{i,j}^{(k)} - \delta_{i,j}\bigr|^q \ge N_k^{-1} = 2^{-k} > n^{-q_1/2}$$

for such a k; hence $d_n \gg n^{-q_1/(2q)} \gg n^{-1/2} \log^{1/2} n$.

In the second case, we apply (6.1) for $q = 2$:

$$d_n^2 \gg \sum_{k=0}^{\infty} N_k^{-2} \sum_{i,j=1}^{N_k} \left|M_{i,j}^{(k)} - \delta_{i,j}\right|^2 \geq \sum_{4n<2^k<n^{q_1/2}} \frac{N_k^{-2} c N_k^2}{n} \gg \Big(\frac{q_1}{2} - 1\Big) n^{-1} \log_2 n;$$

hence again $d_n \gg n^{-1/2} \log^{1/2} n$.

ACKNOWLEDGMENTS

This work was supported by a grant of the Government of the Russian Federation (contract no. 14.W03.31.0031).

REFERENCES

1. N. Alon, "Perturbed identity matrices have high rank: proof and applications," Comb. Probab. Comput. **18** (1–2), 3–15 (2009).
2. N. Alon and B. Klartag, "Optimal compression of approximate inner products and dimension reduction," arXiv: 1610.00239 [math.MG].
3. N. Alon, T. Lee, A. Shraibman, and S. Vempala, "The approximate rank of a matrix and its algorithmic applications: approximate rank," in *Proc. 45th Annu. ACM Symp. on Theory of Computing, Palo Alto, CA (USA), 2013* (ACM, New York, 2013), pp. 675–684.
4. N. K. Bary, *A Treatise on Trigonometric Series* (Pergamon, Oxford, 1964), Vols. I, II.
5. E. S. Belinskii, "Approximation of periodic functions by a 'floating' system of exponentials, and trigonometric widths," in *Studies on the Theory of Functions of Many Real Variables* (Yaroslav. Gos. Univ., Yaroslavl, 1984), pp. 10–24 [in Russian].
6. R. A. DeVore and V. N. Temlyakov, "Nonlinear approximation by trigonometric sums," J. Fourier Anal. Appl. **2** (1), 29–48 (1995).
7. J. Forster, N. Schmitt, H. U. Simon, and T. Suttorp, "Estimating the optimal margins of embeddings in Euclidean half spaces," Mach. Learn. **51** (3), 263–281 (2003).
8. E. D. Gluskin, "Norms of random matrices and widths of finite-dimensional sets," Math. USSR, Sb. **48** (1), 173–182 (1984) [transl. from Mat. Sb. **120** (2), 180–189 (1983)].
9. E. D. Gluskin, "The octahedron is badly approximated by random subspaces," Funct. Anal. Appl. **20** (1), 11–16 (1986) [transl. from Funkts. Anal. Prilozh. **20** (1), 14–20 (1986)].
10. L. Grafakos, *Classical Fourier Analysis*, 3rd ed. (Springer, New York, 2014).
11. G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities* (Univ. Press, Cambridge, 1934).
12. R. S. Ismagilov, "Diameters of sets in normed linear spaces and the approximation of functions by trigonometric polynomials," Russ. Math. Surv. **29** (3), 169–186 (1974) [transl. from Usp. Mat. Nauk **29** (3), 161–178 (1974)].
13. B. S. Kashin, "The diameters of octahedra," Usp. Mat. Nauk **30** (4), 251–252 (1975).
14. B. S. Kashin, "Diameters of some finite-dimensional sets and classes of smooth functions," Math. USSR, Izv. **11** (2), 317–333 (1977) [transl. from Izv. Akad. Nauk SSSR, Ser. Mat. **41** (2), 334–351 (1977)].
15. B. S. Kashin, "On certain properties of matrices of bounded operators from ℓ_2^n into ℓ_2^m," Izv. Akad. Nauk Arm. SSR **15** (5), 379–394 (1980).
16. B. S. Kashin, "Lower bounds for n-term approximations of plane convex sets and related topics," Math. Notes **72** (4), 473–478 (2002) [transl. from Mat. Zametki **72** (4), 509–515 (2002)].
17. B. S. Kashin, "Remark on estimates of orthomassivity," Math. Notes **84** (6), 879–882 (2008) [transl. from Mat. Zametki **84** (6), 948–951 (2008)].
18. B. S. Kashin, "Dyadic analogues of Hilbert matrices," Russ. Math. Surv. **71** (6), 1135–1136 (2016) [transl. from Usp. Mat. Nauk **71** (6), 155–156 (2016)].
19. B. S. Kashin and V. N. Temlyakov, "On a norm and approximate characteristics of classes of multivariable functions," J. Math. Sci. **155** (1), 57–80 (2008).
20. A. R. Klivans and A. A. Sherstov, "Lower bounds for agnostic learning via approximate rank," Comput. Complexity **19** (4), 581–604 (2010).
21. E. D. Kulanin, "On the diameters of a class of functions of bounded variation in the space $L^q(0,1)$, $2 < q < \infty$," Russ. Math. Surv. **38** (5), 146–147 (1983) [transl. from Usp. Mat. Nauk **38** (5), 191–192 (1983)].
22. T. Lee and A. Shraibman, "An approximation algorithm for approximation rank," in *Proc. 24th Annu. IEEE Conf. on Computational Complexity, Paris, 2009* (IEEE, Los Alamitos, CA, 2009), pp. 351–357.

23. N. Linial, S. Mendelson, G. Schechtman, and A. Shraibman, "Complexity measures of sign matrices," Combinatorica **27** (4), 439–463 (2007).
24. N. Linial and A. Shraibman, "Lower bounds in communication complexity based on factorization norms," Random Struct. Algorithms **34** (3), 368–394 (2009).
25. S. V. Lokam, "Complexity lower bounds using linear algebra," Found. Trends Theor. Comput. Sci. **4** (1–2), 1–155 (2009).
26. G. G. Lorentz, M. von Golitschek, and Y. Makovoz, *Constructive Approximation: Advanced Problems* (Springer, Berlin, 1996).
27. M. Talagrand, *Upper and Lower Bounds for Stochastic Processes: Modern Methods and Classical Problems* (Springer, Berlin, 2014), Ergeb. Math. Grenzgeb. 3. Flg. **60**.
28. V. Temlyakov, *Sparse Approximation with Bases* (Birkhäuser, Basel, 2015).
29. V. Temlyakov, "Constructive sparse trigonometric approximation for functions with small mixed smoothness," Constr. Approx. **45** (3), 467–495 (2017).
30. A. Zygmund, *Trigonometric Series* (Univ. Press, Cambridge, 1959), Vols. 1, 2.

Translated by I. Nikitin