Approximation of the step–function, $d_n(W_1^1, L^q)$ and approximate rank.

K.S. Ryutin (joint work with B.S. Kashin, Yu.V. Malykhin)

12 august 2019

Motivation

The main focus of our research is the problem of the order for $d_n(W_1^1, L^q[0,1]), 2 < q < \infty$ (the Kolmogorov width for the convex hull of the step-functions). We show that it is closely related to the order of $\mathrm{rank}_\varepsilon$ for a specific upper-triangular matrix (with 1 on and above the diagonal) and we try to make sharp estimates for this quantity. This problem can be approached with different techniques from harmonic analysis, approximation theory, probability. In this talk we plan to describe these approaches and the results that can be obtained.

The authors acknowledge support from the grant of the Russian Federation Government (project no. 14.W03.31.0031).

Introduction.

The order of decay for Kolmogorov widths of Sobolev classes (the case of small smoothness), similar problems in computer science and some variations.
The main "hero" — the class of step functions in one variable, functions of bounded variation or W_1^1.

Kolmogorov width (поперечник)

Let X be a linear normed space; $W, L \subset X$.

Kolmogorov width (поперечник)

Let X be a linear normed space; $W, L \subset X$.
Deviation of W from L:

$$E(W, L)_X := \sup_{x \in W} \inf_{y \in L} \|x - y\|_X.$$

It shows how well the "worst" elements of W can be approximated by the set L.

Kolmogorov width (поперечник)

Let X be a linear normed space; $W, L \subset X$.
Deviation of W from L:

$$E(W, L)_X := \sup_{x \in W} \inf_{y \in L} \|x - y\|_X.$$

It shows how well the "worst" elements of W can be approximated by the set L.

Kolmogorov n-width of a set W in X:

$$d_n(W, X) := \inf_{\substack{L_n \subset X \\ \dim L_n \leqslant n}} E(W, L_n)_X,$$

where the inf is taken over all linear subspaces of X with dim $\leqslant n$.

The notion was introduced by A.N. Kolmogorov (1936). Motivation: complexity for functional classes, especially classes of functions of several variables (superpositions), applications such as tabulation of functions etc. Very intensive studies since 50s. One of the main problems: orders of decay for Sobolev classes. Different techniques from harmonic analysis, functional analysis, probabilities. Active research on similar problems in computer science and applications (compressed sensing).

Starting point.

Starting point.

The order of decay of the following width is not known ($2 < q < \infty$):

$$c(q,\varepsilon)n^{-1/2}\log^{1/2-\varepsilon} n \leqslant d_n(W_1^1, L_q[0,1]) \leqslant C(q)n^{-1/2}\log n.$$

We (KMR) improved this result (lower estimate).
For $q > 2$ we have

$$d_n(W_1^1, L_q) > c_q(\ln n)^{1/2}n^{-1/2}$$

.

Lemma

Let $M = (M_{i,j})_{i,j=1}^{N}$ *with* $\operatorname{rank} M \leqslant n \leqslant N/4$. *Then at least one of the statements holds:*

(i) *half of diagonal elements are far from* 1*:*

$$\#\{i\colon |M_{i,i} - 1| > 1/2\} \geqslant N/2;$$

(ii) M *is far from the* Id *in the mean:*

$$\sum_{i,j=1}^{N} |M_{i,j} - \delta_{i,j}|^2 \geqslant cN^2/n.$$

Motivation for $\widetilde{B_1^N}$

One may think of the Sobolev class W_1^1 as the set of the "steps"

$$\chi_t(x) = \begin{cases} 1, & x \in [0,t], \\ 0, & x \in (t,1]. \end{cases}.$$

Motivation for $\widetilde{B_1^N}$

One may think of the Sobolev class W_1^1 as the set of the "steps"

$$\chi_t(x) = \begin{cases} 1, & x \in [0,t], \\ 0, & x \in (t,1]. \end{cases}.$$

V.N. Konovalov in 2003 pointed out the *discretization* for this case:

$$d_n(W_1^1, L_q) \asymp N^{-1/q} d_n(\widetilde{B_1^N}, \ell_q^N), \quad \text{for } N > n^{q/2}.$$

Main example

Where we introduce the *skewed octahedron*

$$\widetilde{B_1^N} := \mathrm{conv}\{\pm(1,1,\dots,1,0,\dots,0)\},$$

Main example

Where we introduce the *skewed octahedron*

$$\widetilde{B_1^N} := \mathrm{conv}\{\pm(1,1,\ldots,1,0,\ldots,0)\},$$

And $\Delta^{(N)}$ — the upper-triangular $0/1$-matrix:

$$\Delta_{i,j}^{(N)} = \begin{cases} 1, & 1 \leqslant i \leqslant j \leqslant N, \\ 0, & i > j. \end{cases}.$$

Main example

Where we introduce the *skewed octahedron*

$$\widetilde{B_1^N} := \mathrm{conv}\{\pm(1,1,\dots,1,0,\dots,0)\},$$

And $\Delta^{(N)}$ — the upper-triangular 0/1-matrix:

$$\Delta^{(N)}_{i,j} = \begin{cases} 1, & 1 \leqslant i \leqslant j \leqslant N, \\ 0, & i > j. \end{cases}.$$

We have

$$\mathrm{rank}_\varepsilon(\Delta^{(N)}) \leqslant n \quad\Longleftrightarrow\quad d_n(\widetilde{B_1^N}, \ell_\infty^N) \leqslant \varepsilon.$$

Approximate rank (ε-rank)

Let $A = (A_{i,j})$ be a matrix, and let $\varepsilon > 0$.

Definition

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

Approximate rank (ε-rank)

Let $A = (A_{i,j})$ be a matrix, and let $\varepsilon > 0$.

Definition

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

Motivation:

- Low-rank approximation: many applications in machine learning/statistics, data compression, NLP, numerical methods.

Approximate rank (ε-rank)

Let $A = (A_{i,j})$ be a matrix, and let $\varepsilon > 0$.

Definition

$$\mathrm{rank}_\varepsilon(A) := \min\{\mathrm{rank}\, B\colon \max_{i,j} |A_{i,j} - B_{i,j}| \leqslant \varepsilon\}.$$

Motivation:

- Low-rank approximation: many applications in machine learning/statistics, data compression, NLP, numerical methods. Usually we approximate in Frobenius/spectral norm, where best approximant is just the truncated SVD:

$$B = \sum_{k=1}^{n} \sigma_k u_k v_k^t.$$

In the supremum norm the problem is much more difficult!

Complexity theory

- Complexity of computational problems $\Rightarrow$ complexity measures of related objects (functions, graphs, matrices, ...).

Complexity theory

- Complexity of computational problems $\Rightarrow$ complexity measures of related objects (functions, graphs, matrices, ...).

  Rigid matrices:

  $$\operatorname{rank}(A,k) := \min\{\operatorname{rank} B\colon A \text{ and } B \text{ differ at } \leqslant k \text{ entries}\}.$$

  Problem: construct an explicit family A_N of rigid $N \times N$ matrices: $\operatorname{rank}(A_N, N^{1+\delta}) \geqslant \delta N$. (Walsh matrices are not rigid – 2016.)

Complexity theory

- Complexity of computational problems $\Rightarrow$ complexity measures of related objects (functions, graphs, matrices, ...).

  Rigid matrices:

  $$\operatorname{rank}(A, k) := \min\{\operatorname{rank} B \colon A \text{ and } B \text{ differ at } \leqslant k \text{ entries}\}.$$

  Problem: construct an explicit family A_N of rigid $N \times N$ matrices: $\operatorname{rank}(A_N, N^{1+\delta}) \geqslant \delta N$. (Walsh matrices are not rigid – 2016.)

  Signum-rank:

  $$\operatorname{rank}_{\pm}(A) := \min\{\operatorname{rank} B \colon \operatorname{sign} B_{i,j} \equiv \operatorname{sign} A_{i,j}\}.$$

Complexity theory

- Complexity of computational problems $\Rightarrow$ complexity measures of related objects (functions, graphs, matrices, ...).

  Rigid matrices:

  $$\mathrm{rank}(A, k) := \min\{\mathrm{rank}\, B \colon A \text{ and } B \text{ differ at } \leqslant k \text{ entries}\}.$$

  Problem: construct an explicit family A_N of rigid $N \times N$ matrices: $\mathrm{rank}(A_N, N^{1+\delta}) \geqslant \delta N$. (Walsh matrices are not rigid – 2016.)

  Signum-rank:

  $$\mathrm{rank}_{\pm}(A) := \min\{\mathrm{rank}\, B \colon\ \mathrm{sign}\, B_{i,j} \equiv \mathrm{sign}\, A_{i,j}\}.$$

  Approximate rank: T. Lee, A. Shraibman, 2009. Related to communication complexity.

Equivalence

Let A be an $N \times M$ matrix. Define $W_A \subset R^N$ as the set of the vectors A^j — the columns of A ($j = 1, \dots, M$).

Equivalence

Let A be an $N \times M$ matrix. Define $W_A \subset R^N$ as the set of the vectors A^j — the columns of A ($j = 1, \ldots, M$). Then the functions $\mathrm{rank}_\varepsilon$ and d_n are inverse of each other:

$$\mathrm{rank}_\varepsilon(A) \leqslant n \quad \Longleftrightarrow \quad d_n(W_A, \ell_\infty^N) \leqslant \varepsilon.$$

Equivalence

Let A be an $N \times M$ matrix. Define $W_A \subset R^N$ as the set of the vectors A^j — the columns of A ($j = 1, \ldots, M$). Then the functions $\mathrm{rank}_\varepsilon$ and d_n are inverse of each other:

$$\mathrm{rank}_\varepsilon(A) \leqslant n \quad \Longleftrightarrow \quad d_n(W_A, \ell_\infty^N) \leqslant \varepsilon.$$

(One can replace W_A by $\mathrm{conv}(W_A \cup -W_A)$.)

Equivalence

Let A be an $N \times M$ matrix. Define $W_A \subset R^N$ as the set of the vectors A^j — the columns of A ($j = 1, \ldots, M$). Then the functions $\mathrm{rank}_\varepsilon$ and d_n are inverse of each other:

$$\mathrm{rank}_\varepsilon(A) \leqslant n \quad \Longleftrightarrow \quad d_n(W_A, \ell_\infty^N) \leqslant \varepsilon.$$

(One can replace W_A by $\mathrm{conv}(W_A \cup -W_A)$.)

Example

Approximate rank of the identity matrix $\Leftrightarrow$ Kolmogorov width of the octahedra:

$$\mathrm{rank}_\varepsilon(\mathrm{Id}) \leqslant n \quad \Longleftrightarrow \quad d_n(B_1^N, \ell_\infty^N) \leqslant \varepsilon.$$

Example

For the ε–rank of the identity matrix we have to estimate $d_n(B_1^N, l_\infty^N)$. The order of decay $d_n(B_1^N, l_\infty^N)$ is unknown. The problem is with extremely low–dimensional approximations. Estimates by B.S. Kashin, K. Höllig and others. E.D. Gluskin (1988) —the best lower bound:

$$d_n(B_1^N, l_\infty^N) \geqslant c\sqrt{\frac{\log N}{n\log(1+n/\log N)}}$$

Conjecture. True order!

Random subspace gives $\sqrt{\frac{1+\log(N/n)}{n}}$. Random subspaces are not good enough!

Known bounds

Fix ε; say, $\varepsilon = 1/3$.

Known bounds

Fix ε; say, $\varepsilon = 1/3$. It is known that

$$c \log^2 N \leqslant \operatorname{rank}_{1/3}(\Delta^{(N)}) \leqslant C \log^3 N. \qquad (*)$$

Known bounds

Fix ε; say, $\varepsilon = 1/3$. It is known that

$$c \log^2 N \leqslant \mathrm{rank}_{1/3}(\Delta^{(N)}) \leqslant C \log^3 N. \qquad (*)$$

N. Alon, Lee, Shraibman, S. Vempala (2013) derive this from known results. We don't know how to improve this :(

Known bounds

Fix ε; say, $\varepsilon = 1/3$. It is known that

$$c \log^2 N \leqslant \mathrm{rank}_{1/3}(\Delta^{(N)}) \leqslant C \log^3 N. \qquad (*)$$

N. Alon, Lee, Shraibman, S. Vempala (2013) derive this from known results. We don't know how to improve this :(

New results given in our paper:

- generalization of (*) to multivariate case, $\{\chi_D\}$ for parallelepipeds D;
- the same upper bound for trigonometric approximation;
- $\log^3 N$ lower bound for some reasonable random method;

Methods

One should try to use methods related to both of the two equivalent notions:

Methods

One should try to use methods related to both of the two equivalent notions:

- approx rank of a matrix: SVD; SDP; operator norms, especially factorization norm:

$$\gamma_2(A) := \inf_{A=BC} \|C\|_{1\to 2}\|B\|_{2\to\infty},$$

  e.g., the relation $\gamma_2(A) \leqslant \sqrt{\operatorname{rank} A} \cdot \max_{i,j} |A_{i,j}|$;

Methods

One should try to use methods related to both of the two equivalent notions:

- approx rank of a matrix: SVD; SDP; operator norms, especially factorization norm:

$$\gamma_2(A) := \inf_{A=BC} \|C\|_{1\to 2}\|B\|_{2\to\infty},$$

  e.g., the relation $\gamma_2(A) \leqslant \sqrt{\operatorname{rank} A} \cdot \max_{i,j} |A_{i,j}|$;
- width of a convex body: entropy; volumes; orthomassivity (see later!); trigonometric approximation;
  ... and many other.

Upper bounds: JL-lemma

[ALSV] use a random method: let A be an $N \times N$ matrix and $A_{i,j} = \langle x_i, y_j \rangle$. Consider a projection of x_i, y_j on n-dimensional subspace by a random $n \times N$ matrix U (as in Johnson-Lindenstrauss lemma). For appropriate n and U, we have $\langle x_i, y_j \rangle \approx \langle Ux_i, Uy_j \rangle$ with high probability and A is approximated by a matrix of rank $\leqslant n$.

Upper bounds: JL-lemma

[ALSV] use a random method: let A be an $N \times N$ matrix and $A_{i,j} = \langle x_i, y_j \rangle$. Consider a projection of x_i, y_j on n-dimensional subspace by a random $n \times N$ matrix U (as in Johnson-Lindenstrauss lemma). For appropriate n and U, we have $\langle x_i, y_j \rangle \approx \langle Ux_i, Uy_j \rangle$ with high probability and A is approximated by a matrix of rank $\leqslant n$.

Theorem (Alon, Klartag, 2016)

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

This gives an upper bound for $\mathrm{rank}_\varepsilon(A)$ in terms of $\gamma_2(A)$.

Trigonometric width

Let $X = (\mathbb{C}^N, \|\cdot\|)$ be N-dimensional normed space over $\mathbb{C}$, $A \subset X$.
Trigonometrical n–width (R.S. Ismagilov (1974)):

$$d_n^T(A,X) := \inf_{0\leqslant k_1,\dots,k_n<N} \sup_{x\in A} \inf_{c_1,\dots,c_n\in\mathbb{C}} \|x - \sum_1^n c_j e_{k_j}\|_X,$$

where $e_k = (\exp(2\pi i k j/N))_{j=0}^{N-1}$ is the discrete version of *exp* function.
For $A \subset \mathbb{R}^N$ we have $d_{2n}(A, X^{\mathbb{R}}) \leqslant d_n^T(A, X^{\mathbb{C}})$.
Best n-term approximation

$$\sigma_n(x)_X := \inf_{\substack{0\leqslant k_1,\dots,k_n<N \\ c_1,\dots,c_n\in\mathbb{C}}} \|x - \sum_1^n c_j e_{k_j}\|_X.$$

E.S. Belinsky (1987)

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

For the skew octahedron the problem with trigonometric width boils down to n-term approximation of a fixed function:

$$d_n^T(\widetilde{B_1^N}, \ell_\infty^N) \approx \sigma_n(\sum_{k=1}^{aN} \frac{\sin kx}{k})_{\ell_\infty\{\frac{2\pi j}{N}\}_{j=0}^{N-1}},$$

a — large fixed constant.

Upper bounds: trigonometric approximation

Upper bounds: trigonometric approximation

We make use of one general theorem on sparse approximation:

Theorem (R. DeVore, V.N. Temlyakov, 1995)

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

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

where $\|\cdot\|_0$ *is the number of non-zero coordinates.*

Upper bounds: trigonometric approximation

We make use of one general theorem on sparse approximation:

Theorem (R. DeVore, V.N. Temlyakov, 1995)

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

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

where $\|\cdot\|_0$ *is the number of non-zero coordinates.*

Apply for trigonometric setting:

$$\sigma_n(x)_{\ell_\infty^N} \leqslant Cn^{-1/2} \log^{1/2}(2 + N/n)\|x\|_A, \quad \|x\|_A := \sum_k |\langle x, e_k\rangle|.$$

So, $O(\log^3 N)$ harmonics suffice to approximate "steps".

Upper bounds: constructive trigonometric approximation

Theorem (Temlyakov, 2005)

There exists a constructive method $A_{N,n}$ that provides an n-term trigonometric polynomial $A_{N,n}(t)$ for any real trigonometric polynomial t with the following approximation property:

$$\|t - A_{N,n}(t)\|_\infty \leqslant Cn^{-1/2} \log^{1/2}(2 + N/n)\|t\|_A.$$

Lower bound for random method

Suppose we want to get a sparse approximation for a vector $x = \sum_{k=1}^{N} x_k e_k$ (here x_k are coefficients, and $\{e_k\}$ is some system).

Lower bound for random method

Suppose we want to get a sparse approximation for a vector $x = \sum_{k=1}^{N} x_k e_k$ (here x_k are coefficients, and $\{e_k\}$ is some system). Consider *selectors*, i.e. random variables $\{\xi_k(\omega)\}_{k=1}^{N}$, such that

- $\mathsf{P}(\xi_k = 0) = 1 - \delta_k$,
- $\mathsf{P}(\xi_k = \delta_k^{-1}) = \delta_k$, hence $\mathsf{E}\xi_k = 1$

Lower bound for random method

Suppose we want to get a sparse approximation for a vector $x = \sum_{k=1}^N x_k e_k$ (here x_k are coefficients, and $\{e_k\}$ is some system). Consider *selectors*, i.e. random variables $\{\xi_k(\omega)\}_{k=1}^N$, such that

- $\mathsf{P}(\xi_k = 0) = 1 - \delta_k$,
- $\mathsf{P}(\xi_k = \delta_k^{-1}) = \delta_k$, hence $\mathsf{E}\xi_k = 1$

If one can prove that the norm $\|\sum_{k=1}^N x_k(1-\xi_k)e_k\|$ is small with high probability, then x will be approximated by the (sparse) sum $\sum_{k:\ \xi_k\neq 0} x_k \xi_k e_k$.

Lower bound for random method

Suppose we want to get a sparse approximation for a vector $x = \sum_{k=1}^{N} x_k e_k$ (here x_k are coefficients, and $\{e_k\}$ is some system). Consider *selectors*, i.e. random variables $\{\xi_k(\omega)\}_{k=1}^{N}$, such that

- $\mathsf{P}(\xi_k = 0) = 1 - \delta_k$,
- $\mathsf{P}(\xi_k = \delta_k^{-1}) = \delta_k$, hence $\mathsf{E}\xi_k = 1$

If one can prove that the norm $\|\sum_{k=1}^{N} x_k(1-\xi_k)e_k\|$ is small with high probability, then x will be approximated by the (sparse) sum $\sum_{k:\ \xi_k \neq 0} x_k \xi_k e_k$.

This method also gives the desired approximation with $O(\log^3 N)$ harmonics. But not better than that!

Lower bound for random method

As trigonometric polynomials are shift-invariant, approximation may be reduced to one “model” function $\sum k^{-1} \sin kx$.

Lower bound for random method

As trigonometric polynomials are shift-invariant, approximation may be reduced to one "model" function $\sum k^{-1}\sin kx$.

Statement

Let $(\eta_k)_{k=1}^N$ *be random variables with* $\mathrm{E}\eta_k = 1$*, such that*

$$\mathrm{E}\eta_k^4 \leqslant C\delta_k^{-3}, \quad \textit{where } \delta_k := \mathrm{P}(\eta_k \neq 0).$$

Then either

$$\sum_{k=1}^N \delta_k \geqslant c\log^3 N,$$

or

$$\mathrm{E}\|\sum_{k=1}^N \frac{\sin kx}{k} - \sum_{k=1}^N \eta_k \frac{\sin kx}{k}\|_\infty \geqslant c.$$

Main ingredient of the proof: QC-norm

$$\Delta_0 = \hat{f}(0), \quad \Delta_s f := \sum_{2^{s-1} \leqslant |k| < 2^s} \hat{f}(k) e^{ikx}, \ s \geqslant 1.$$

Define

$$\|f\|_{QC} := \mathsf{E} \| \sum_{s=0}^{\infty} \pm \Delta_s f \|_\infty.$$

Main ingredient of the proof: QC-norm

$$\Delta_0 = \hat{f}(0), \quad \Delta_s f := \sum_{2^{s-1}\leqslant|k|<2^s} \hat{f}(k)e^{ikx},\ s \geqslant 1.$$

Define

$$\|f\|_{QC} := \mathsf{E}\|\sum_{s=0}^{\infty} \pm\Delta_s f\|_\infty.$$

Theorem (Kashin, Temlyakov, 2007)

For any $f \in L^1(\mathbb{T})$ the inequality holds:

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

A useful version for polynomials:

Lemma

For any polynomial $f(x) = \sum_{k=l+1}^{2l} p_k(x)\cos(4^k x)$, *with* $p_k \in \mathcal{T}_{2^l}$ *we have*

$$\|f\|_\infty \geqslant c\sum_k \|p_k\|_1$$

Multivariate case: orthomassivity

Recall the notion of orthomassivity (Kashin 2002): let K be a subset of the unit ball in Hilbert space H,

$$\mathrm{OM}_n(K) := n^{-1/2} \sup_{\substack{\{\varphi_j\}_1^n \text{ orthonormal} \\ \{f_j\}_1^n \subset K}} \sum_1^n \langle \varphi_j, f_j \rangle.$$

Multivariate case: orthomassivity

Recall the notion of orthomassivity (Kashin 2002): let K be a subset of the unit ball in Hilbert space H,

$$\mathrm{OM}_n(K) := n^{-1/2} \sup_{\substack{\{\varphi_j\}_1^n \text{ orthonormal} \\ \{f_j\}_1^n \subset K}} \sum_1^n \langle \varphi_j, f_j \rangle.$$

It is an equivalent definition if we allow any $\{\varphi_j\}_{j=1}^n \subset H$ with

$$\sup_{\sum a_j^2 \leqslant 1} \|\sum_1^n a_j \varphi_j\|_H \leqslant 1.$$

Multivariate case: orthomassivity

Recall the notion of orthomassivity (Kashin 2002): let K be a subset of the unit ball in Hilbert space H,

$$\mathrm{OM}_n(K) := n^{-1/2} \sup_{\substack{\{\varphi_j\}_1^n \text{ orthonormal} \\ \{f_j\}_1^n \subset K}} \sum_1^n \langle \varphi_j, f_j \rangle.$$

It is an equivalent definition if we allow any $\{\varphi_j\}_{j=1}^n \subset H$ with

$$\sup_{\sum a_j^2 \leqslant 1} \| \sum_1^n a_j \varphi_j \|_H \leqslant 1.$$

For the set of the discrete "steps" $f_i = (1, 1, \ldots, 1, 0, \ldots, 0)$, one has $\mathrm{OM}_N(\{f_i\}_1^N) \asymp N \log N$. We can take $\varphi_i := H_i$, the i-th row of the Hilbert matrix $H_{i,j} = 1/(i-j)$:

$$\langle f_i, H_i \rangle \geqslant c \log N \quad \text{for } i \geqslant N/2, \text{ say.}$$

Multivariate case: idea of the proof

If the "steps" f_i are well approximated by g_i, then $\langle g_i, H_i \rangle$ is large:

$$\sum_{i=1}^{N} \langle g_i, H_i \rangle \geqslant cN \log N.$$

Multivariate case: idea of the proof

If the "steps" f_i are well approximated by g_i, then $\langle g_i, H_i\rangle$ is large:

$$\sum_{i=1}^{N}\langle g_i, H_i\rangle \geqslant cN\log N.$$

On the other hand, g_i come from a low-dimensional space L_n, hence

$$\sum_{i=1}^{N}\langle g_i, h_i\rangle = \sum_{i=1}^{N}\langle g_i, P_{L_n}h_i\rangle \ll N\Big(\sum_{i=1}^{N}|P_{L_n}h_i|^2\Big)^{1/2}.$$

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

$$\sum_{i=1}^{N}|P_{L_n}h_i|^2 = \sum_{i=1}^{N}\sum_{j=1}^{n}\langle h_i, \psi_j\rangle^2 = \sum_{j=1}^{n}|H\psi_j|^2 \leqslant Cn,$$

because $\|H\| \leqslant C$.

Multivariate case: idea of the proof

If the "steps" f_i are well approximated by g_i, then $\langle g_i, H_i\rangle$ is large:

$$\sum_{i=1}^{N}\langle g_i, H_i\rangle \geqslant cN\log N.$$

On the other hand, g_i come from a low-dimensional space L_n, hence

$$\sum_{i=1}^{N}\langle g_i, h_i\rangle = \sum_{i=1}^{N}\langle g_i, P_{L_n}h_i\rangle \ll N(\sum_{i=1}^{N}|P_{L_n}h_i|^2)^{1/2}.$$

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

$$\sum_{i=1}^{N}|P_{L_n}h_i|^2 = \sum_{i=1}^{N}\sum_{j=1}^{n}\langle h_i, \psi_j\rangle^2 = \sum_{j=1}^{n}|H\psi_j|^2 \leqslant Cn,$$

because $\|H\| \leqslant C$. Conclude than $n \gg \log^2 N$ by comparing these inequalities.

Multivariate case: idea of the proof

If the "steps" f_i are well approximated by g_i, then $\langle g_i, H_i\rangle$ is large:

$$\sum_{i=1}^{N}\langle g_i, H_i\rangle \geqslant cN\log N.$$

On the other hand, g_i come from a low-dimensional space L_n, hence

$$\sum_{i=1}^{N}\langle g_i, h_i\rangle = \sum_{i=1}^{N}\langle g_i, P_{L_n}h_i\rangle \ll N(\sum_{i=1}^{N}|P_{L_n}h_i|^2)^{1/2}.$$

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

$$\sum_{i=1}^{N}|P_{L_n}h_i|^2 = \sum_{i=1}^{N}\sum_{j=1}^{n}\langle h_i, \psi_j\rangle^2 = \sum_{j=1}^{n}|H\psi_j|^2 \leqslant Cn,$$

because $\|H\| \leqslant C$. Conclude than $n \gg \log^2 N$ by comparing these inequalities.

Multivariate case: idea of the proof

This generalizes to the multivariate case!
We consider the vectors in $\mathbb{R}^{N^d}$, which are indicators of parallelepipeds: $\pi^{\mathbf{t}}$, where $\mathbf{t} = (t_1, \ldots, t_d)$, $t_j \in \{0, \ldots, N-1\}$, and

$$\pi^{\mathbf{t}}[i_1, \ldots, i_d] = \begin{cases} 1, & i_k < t_k,\ k = 1, \ldots, d, \\ 0, & \text{otherwise.} \end{cases}$$

Multivariate case: idea of the proof

This generalizes to the multivariate case!
We consider the vectors in $\mathbb{R}^{N^d}$, which are indicators of parallelepipeds: $\pi^{\mathbf{t}}$, where $\mathbf{t} = (t_1, \dots, t_d)$, $t_j \in \{0, \dots, N-1\}$, and

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

Statement

For $n_1 \asymp \log^{2d+1} N$, $n_2 \asymp \log^{2d} N$, *and some* $0 < c(d) < 1/3$, *we have*

$$d_{n_1}(\{\pi^{\mathbf{t}}\}, \ell_\infty^{N^d}) \leqslant c(d) \leqslant d_{n_2}(\{\pi^{\mathbf{t}}\}, \ell_\infty^{N^d}).$$