Harbin, November 23, 2022

On some applications and connections of Functional Analysis

B. Kashin

Steklov Mathematical Institute

Moscow, RUSSIA

kashin@mi-ras.ru

The main subjects of this talk are two directions of Functional Analysis:

a) asymptotic theory of the finite dimensional normed spaces;

b) approximation theory.

Two current trends:

I) Fast implementation of the theoretical results in practice due to the development of Computer Science and AI.

II) Convergence of the indicated directions of the Functional Analysis to some directions of Theoretical Computer Science including mutual penetration of methods of investigations.

Nowadays the following notions from FA are widely used in practice:

Kolmogorov *n*-width

m-term approximation

Greedy algorithm

Definition 1 (Kolmogorov width)

The *Kolmogorov width of order* n of a set K in a linear metric space X with metric ρ is a quantity

$$d_n(K, X) = \inf_{L_n \subset X} \rho(K, L_n), \quad \rho(K, L_n) = \sup_{x \in K} \rho(x, L_n),$$

where infinum is taken over linear subspaces of a fixed dimension n.

An example of practical application

B_2^N - unit ball in l_2^N.

THEOREM A (B.K., 1977)

Let $\rho > 0$ is fixed and $N = 1, 2, \ldots, n \geqslant \rho N$ then

$$d_n(B_2^N, l_\infty^N) \leqslant \frac{C_\rho}{\sqrt{N}}.$$

More generally

THEOREM B
(B.K., 1977, Garnaev and Gluskin, 1984)

For any (n, N) $1 \leqslant n \leqslant N$

$$d_n(B_2^N, l_\infty^N) \leqslant C \left(\frac{1 + \ln \frac{N}{n}}{n} \right)^{1/2}.$$

Definition 2

A sequence $\{u_i\}_{i=1}^N \subset \mathbb{R}^n$ is called *tight frame* if it satisfied Parseval's identity

$$\|x\|_2^2 = \sum_{i=1}^N |\langle x, u_i \rangle|^2 \qquad \text{for all} \quad x \in \mathbb{R}^n.$$

A frame $\{u_i\}_{i=1}^N$ can be identified with $n \times N$ matrix U with columns u_i.

Let v_j, $j = 1, \ldots, n$ rows of U.

U is a tight frame $\longleftrightarrow \{v_j\}_{j=1}^n$ – orthonormal set in $\mathbb{R}^N$

If $U = \{u_i\}_{i=1}^N$ – tight frame, $x \in \mathbb{R}^n$ then

$$x = \sum_{i=1}^N a_j u_i, \qquad a_i = \langle x, u_i \rangle, \quad i = 1, \ldots, N. \qquad (1)$$

If $N > n$ tight frame is a redundant system and representation (1) is not unique.

Frames are widely used in signal processing. But if, when transmitting coefficients $\{a_i\}$, we loose one very big coefficient then we loose all information about x.

Suppose now that $L \subset \mathbb{R}^N$ – such a subspace that $\dim L = N - n \geqslant \rho N$ (so $n \leqslant (1-\rho)N$),

$$\rho_{l_\infty^N}(B_2^N, L) \leqslant K d_{N-n}(B_2^N, l_\infty^N) \qquad (2)$$

and $\mathbb{R}^N$ can be represented as a orthogonal direct sum: $\mathbb{R}^N = L \oplus U$.

In 2010 Lyubarskii and Vershynin showed that in this case a tight frame U generated by the set of columns of U had the following property: for each $x \in \mathbb{R}^N$

$$x = \sum_{i=1}^{N} c_i u_i, \quad \max_i |c_i| \leqslant \frac{C(\rho, K)}{\sqrt{N}} \|x\|_2 \qquad (3)$$

where coefficients c_i could be found by fast and stable algorithm starting with canonical representation (1).

Why is the representation (3) useful?

Suppose that in the process of transmission
of the vector $\{c_i\}_{i=1}^N$ (see (3))
we loose exact values of $\leqslant \delta N$
coefficients and get distorted vector x^*.

Then

$$\|x - x^*\|_2^2 \leqslant \sum_{i:\ c_i \neq c_i^*} (c_i - c_i^*)^2 \leqslant \frac{C(\rho, K)}{N} \delta N \|x\|_2^2 \leqslant \varepsilon \|x\|_2^2,$$

where ε is small if δ is small enough.

The speed and stability of the algorithm converting (1) to (3) depend on the following property of the subspaces U:
there exist $\delta > 0$, $\eta < 1$ such that
for $y = \{y_i\} \in U$, $\|y\|_2 = 1$, $\Lambda \subset \{1, \ldots, N\}$, $|\Lambda| \leqslant \delta N$

$$\left(\sum_{i \in \Lambda} y_i^2 \right)^{1/2} \leqslant \eta.$$

This property is a consequence of (2).

Another very wide field

of practical applications

of the width estimates

(Theorem B)

is compressed sensing.

Let M_n is a set of all $n \times n$ matrices with real elements. Let $A = \{a_{ij}\} \in M_n$, $\varepsilon > 0$.

Definition 3

The approximate ε-rank (or simply ε-rank) of A is the quantity

$$\operatorname{rank}_\varepsilon(A) = \min\{\operatorname{rank}(B),\ B \in M_n,\ \|A - B\|_\infty \leqslant \varepsilon\},$$

where $\|A - B\|_\infty = \max_{(i,j)} |a_{ij} - b_{ij}|$.

“This parameter is connected to other notions of approximate rank and is motivated by problems from various topics including communication complexity, combinatorial optimization, game theory, computational geometry and learning theory.”

(from the paper by N.Alon, T. Lee, A. Shraibman, S. Vempala, Proc. 45th Symp. on Theory of Computing, 2013, pp. 675–684)

Let $\{V_i\}_{i=1}^n$ – the rows of the matrix A and $V = \operatorname{conv}\left(\{\pm V_i\}_{i=1}^n\right)$. Then

$$\operatorname{rank}_\varepsilon(A) = \min\left\{k\colon d_k\left(\bigcup_i V_i, l_\infty^n\right) \leqslant \varepsilon\right\} =$$

$$= \min\{k\colon d_k(V, l_\infty^n) \leqslant \varepsilon\}.$$

The width of skew octahedron

This problem is important for both approximation theory and computer science. Let me start with one open problem concerning Kolmogorov width of Sobolev class W_1^1.

It is known (Kulanin 1983, Kashin, Malykhin, Ryutin 2018), that

$$c_q \frac{\ln^{1/2} n}{\sqrt{n}} \leqslant d_n(W_1^1, L^q) \leqslant C_q \frac{\ln n}{\sqrt{n}}, \quad 2 < q < \infty.$$

V. N. Konovalov remarked (2003), that for $q > 2$

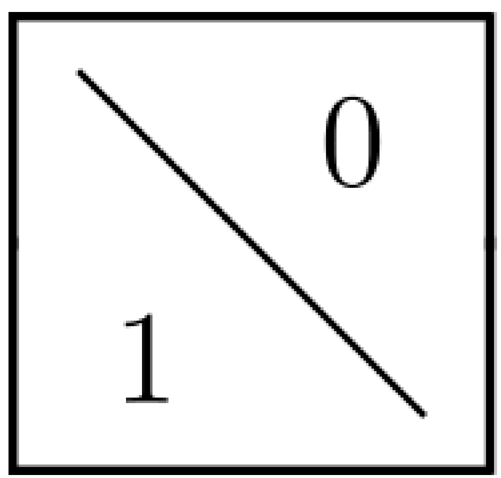
$$d_n(W_1^1, L^q) \asymp N^{-1/q} d_n(Q_N, l_q^N), \quad N \geqslant n^{2/q},$$

where Q_N – the “skew” octahedron.

$$Q_N = \mathrm{conv}\{\pm V_i\}_{i=1}^N, \quad V_i = \{\underbrace{11111}_{i}, 0, 0, 0\}, \ 1 \leqslant i \leqslant N.$$

The case $q = \infty$ – the width $d_n(Q_N, l_\infty^N)$ was not considered in the function theory (maybe because W_1^1 is not a compact in L^∞).

But for computer science this case is important and equivalent to the estimate of ε-rank of $N \times N$ matrix



$= \widetilde{Q}_N$

First of all it make sense to consider the case of fixed $\varepsilon = \text{const} < \frac{1}{2}$, say $\varepsilon = \frac{1}{3}$.

It is known (see the paper by Alon mentioned above) that

$$c \log^2 N \leqslant \text{rank}_{1/3}(\widetilde{Q}_N) \leqslant C \log^3 N. \tag{4}$$

Using methods and results from approximation theory it is possible to give different proofs for both estimates in (4). But we could not improve it. One approach to the lower estimate (4) use the following.

Definition of orthomassivity (2002)

For the set $K \subset B_H$

$$\mathrm{OM}_n(K) = n^{-1/2} \sup_{\{\varphi_j\}_{j=1}^n} \sup_{\{f_j\} \in K} \sum_{j=1}^{n} (\varphi_j, f_j),$$

where the first supremum is taken over all orthonormal system $\{\varphi_j\}_{j=1}^n$.

Let for $d = 1, 2, \ldots$

$$L^2(I^d) \supset \Pi_d = \{\chi_P \colon P = [0, t_1] \times [0, t_2] \times \cdots \times [0, t_d]\},$$

where $P \subset [0, 1]^d = I^d$.

Proposition 1

$$\mathrm{OM}_n(\Pi_d) \asymp (\log n)^d, \qquad n \to \infty.$$

Let K_2 – the set of indicator functions of convex subsets of $[0,1] \times [0,1] = I^2$.

Proposition 2

$$\mathrm{OM}_n(K_2) \geqslant cn^{1/6}, \qquad c > 0, \quad n = 1, 2, \ldots .$$

P. Grigoriev obtained the following upper estimate

$$\mathrm{OM}_n(K_2) \leqslant Cn^{1/4}(\log n)^{1/2}, \qquad n = 2, 3, \dots,$$

and remarked that the estimate

$$\mathrm{OM}_n(K_2) \leqslant n^{1/6}(\log n)^{4/3}$$

could be obtained as a consequence of the positive solution of the following problem.

Problem (P. Grigoriev)

Let $\{f_{j,k}\}_{j,k=1}^{\infty}$ – O.N.S. For $N = 1, 2, \ldots$ let us define maximal operator with respect to triangle partial sums

$$F_N(x) := \sup_{A,B>0:\ AB\leqslant N} \Big| \sum_{Aj+Bk\leqslant AB} f_{j,k}(x) \Big|.$$

Is it true that

$$\|F_N\|_2 \leqslant C\sqrt{N}(\log N)^2\,?$$

Lovasz Θ-function

Let $G = (V, E)$ – graph. Suppose that $V = \{1, 2, \ldots, n\}$.
Orthonormal representation of G – arbitrary system of unit vectors of Hilbert space $H = \{v_1, \ldots, v_n\}$ such that $(v_i, v_j) = 0$ if $(i, j) \notin E$.

$$\Theta(G) = \sup_{\substack{z,\ \|z\|_H = 1 \\ \{v_i\}_{j=1}^n}} \sum_{i=1}^{n} (v_i, z)^2 \qquad (5)$$

($\{v_i\}_{j=1}^n$ in (5) runs over all orthonormal representation of G).

Another important example of the problem about width of the skew octahedron

For $S \subset \{1, 2, \ldots, n\}$ let $f_S(x_1, \ldots, x_n)$ – Boolean function $\mathbb{OR}$ (disjunction). Here $x_j \in \{-1, 1\}$ and $f_S(x_1, \ldots, x_n) = -1$ if and only if there exists $j \in S$ such that $x_j = -1$ ($-1 =$ "true").

The problem is to estimate

$$d(k, n) \equiv d_k \left(\bigcup_S f_S, l_\infty^{2^n} \right)$$

and first of all to estimate

$$k_0(n) = \min \left\{ k \colon d(k, n) \leqslant \frac{1}{3} \right\}.$$

It was proved in computer science that

$$C_1^{\sqrt{n}} \leqslant k_0(n) \leqslant C_2^{\sqrt{n}}$$

(see A. Klivans, A. Sherstov, 2010
for a lower estimate
and references in the mentioned above
paper by N. Alon for upper estimate)

Let for $S \subset \{1, 2, \ldots, n\}$, $t \in [0, 1]$

$$g_S(t) = -1 + 2^{-|S|+1} \prod_{i \in S} (r_i(t) + 1)),$$

where $r_i(t)$ – Rademacher functions.

It is easy to check that

$$d(k, n) = d_k \left(\bigcup_S g_S(t), L^\infty(0, 1) \right).$$

It is natural to formulate the following

PROBLEM

Let $\{v_j\}_{j=1}^n \subset \mathbb{R}^n$.

Suppose that we know the Gram matrix

$$G = \{\langle v_i, v_j \rangle\}_{i,j=1}^n$$

How to estimate the width of the skew octahedron

$$d_k(V, l_\infty^n), \quad V = \mathrm{conv}(\{\pm v_j\}_{j=1}^n)?$$

The table of widths for nonlinear operator

Let X, Y – Banach spaces and $\Psi\colon X \to Y$ (nonlinear) operator such that $\Psi(B_X)$ is bounded. (B_X – the unit ball in X.)

For $n = 1, 2, \ldots$ and $m \geqslant n$

$$D_\Psi(n, m) = \sup_{L \subset X,\ \dim L \leqslant n} d_m(\Psi(B_L), Y)$$

This numbers were defined in 1988
(B. Kashin, Vestnik Moscow State University).

I consider the case

$$X = Y = L^2(0,1), \quad \Psi_0(f) = |f|$$

Proposition (B. K., 2021)
For each $\varepsilon < 1$ there exists C_ε such that for $n = 1, 2, \dots$

$$D_{\Psi_0}(n, [\exp(C_\varepsilon \sqrt{n} \log n)]) \leqslant \varepsilon.$$

Signum-rank
Definition 4

For a matrix $A = \{a_{ij}\}_{i,j=1}^{N}$ with $a_{ij} = \pm 1$

$$\text{sign-rank}\,(A) = \min\{\operatorname{rank} B,\ \ B = \{b_{ij}\} \colon \operatorname{sign} b_{ij} = a_{ij}\}.$$

THEOREM C (Forster)

$$\forall A \qquad \text{sign - rank}\,(A) \geqslant \frac{N}{\|A\|}.$$

Recently the notion of signum-rank was used
by Yu. Malykhin in order to estimate of m-term approximation.

Let me recall the definitions.

Let (X, ρ) – linear metric space
and $\Phi = \{\varphi_j\}$ – system of elements of X
or more general subset of X.

By Π_m we define the set of polynomials with
respect to Φ with $\leqslant m$ nonzero coefficients.

For $f \in X$ the best m-term approximation is

$$e_m(f, \Phi, X) = \inf_{P \in \Pi_m} \rho(f, P).$$

If F – the subset of X then

$$e_m(F, \Phi, X) = \sup_{f \in F} e_m(f, \Phi, X).$$

Important example of m-term approximation
is the following case.
Let $\mathbb{I}^d = [0,1]^d$ and for $f \in L^p(\mathbb{I}^d)$

$$\Theta_m(f, L^p) = \inf_{u^{i,s}\colon [0,1]\to\mathbb{R}} \|f - \sum_{s=1}^{m} u^{1,s}(x_1)\dots u^{d,s}(x_d)\|_{L^p(\mathbb{I}^d)}$$

Using the notion of signum-rank
Yu. Malykhin got the sharp order
of the quantity $\Theta_m(W^r_p, L^q(\mathbb{I}^d))$ for Sobolev classes
if $d \geqslant 3$ (if $d = 2$ it was known).

THEOREM

For $d \geqslant 3$, $r > 0$, $\mathbf{r} = (r, \dots, r)$

$$\Theta_m(W^r_p, L^q(\mathbb{I}^d)) \asymp m^{-\frac{rd}{d-1}}, \qquad 1 \leqslant q \leqslant 2.$$

Here W^r_p – standard functional class.
(For $r \in \mathbb{N}$ W^r_p – class of function
with all mixed derivatives up to the order r
– are bounded by 1 in L^p norm.)

Matrix rigidity

For a matrix $A = \{a_{ij}\}_{ij=1}^{N}$ with $a_{ij} \in \mathbb{R}$ and $r = 1, \ldots, N-1$

$$R_A(r) = \min\{|\Lambda| \subset [1, N] \times [1, N] : \\ \exists\, B = \{b_{ij}\},\ \operatorname{rank} B = r, \\ a_{ij} = b_{ij} \text{ при } (i, j) \notin \Lambda\}.$$

The famous problem in discrete mathematics: to find EFFECTIVE examples of matrices with big value of $R_A(r)$.

<u>This problem is unsolved</u>.

Until recently the candidates for such example were Walsh matrices W_N, $N = 2^s$, $s = 2, 3, \dots$.

It was known (Kashin, Razborov, 1998), that for $r \leqslant \frac{N}{2}$

$$R_{W_N}(r) \geqslant c\,\frac{N^2}{r}.$$

For $r \asymp N$ this estimate becomes trivial.

THEOREM D

(J. Alman, R. Williams, 2017)

For sufficiently small $\varepsilon > 0$

$$R_{W_N}(N^{1-f(\varepsilon)}) \leqslant N^{1+\varepsilon},$$

where $f(\varepsilon) \geqslant c\,\frac{\varepsilon^2}{\log(1/\varepsilon)}$.

In fact in the proof of Theorem D it was shown, that

$$d_{N^{1-\delta}}(\{w_0, \ldots, w_N\},\, R_H^N) \leqslant N^{\delta},$$

$\delta > 0$ – absolute constant,
R_H^N – space $\mathbb{R}^N$ with Hamming metric h,

$$h(x, y) = \#\{i\colon\ x_i \neq y_i\}.$$

In order to get lower estimate for a width in Hamming metric we can replace it by the metric "convergence in measure": for $x, y \in \mathbb{R}^N$

$$\rho(x,y) = \frac{1}{N}\sum_{i=1}^{N} \frac{|x_i - y_i|}{1 + |x_i - y_i|}$$

For upper estimate of this width we can replace it by any L^p, $0 < p < \infty$, metric.

THEOREM E (Yu. Malykhin)

Let $w_1, w_2, \dots$ – Walsh-Paley system.
For any $p < 2$ there exists $\delta = \delta(p) > 0$
such that for any big enough N

$$d_{N^{1-\delta}}(\{w_1, \dots, w_N\}, L^p(0,1)) \leqslant N^{-\delta}.$$

Definition 4

Let $F = \{f_i\}_{i=1}^s$ – the set of elements of linear metric space (X, ρ).
Averaged n-width of the set F is

$$d_{n,p}^{\mathrm{avg}}(F, X) = \inf_{L_n \subset X} \left(\frac{1}{s} \sum_{i=1}^{s} \rho^p(f_i, L_n) \right)^{1/p}, \quad (6)$$

where $1 \leqslant p \leqslant \infty$ and infinum is taken over all linear subspaces of dimension n.

Classical result from linear algebra implies that for any orthonormal system $\Psi = \{\psi_1, \ldots, \psi_N\} \subset L^2(0,1)\}$

$$d_{n,2}^{\mathrm{avg}}(\Psi, L^2) = 1 - \frac{n}{N}, \qquad 1 \leqslant n \leqslant N.$$

Recently some estimates of averaged n-width were obtained by Yu. Malykhin and B. Kashin.

I used this quantity in order to get lower estimates for n-term approximation in L^0_n metric.

For $n = 1, 2, \dots$, let

$$E_n = \{\varepsilon = \{\varepsilon_\nu\}_{\nu=1}^n \colon \varepsilon_\nu = \pm 1,\ \nu = 1, 2, \dots, n\}$$

and let μ_n be a natural measure on E_n: for $A \subset E_n$, we put $\mu_n(A) = |A| \cdot 2^{-n}$, where $|A|$ is the cardinality of A.

Theorem 5. *There are positive absolute constants c_1 and c_2 such that*

$$\max_{\varepsilon \in E_n} e_m\left(\sum_{\nu=1}^{n} \varepsilon_\nu \psi_\nu, \Phi, L_n^0\right) \geqslant c_2$$

for any $n = 1, 2, \dots$, any orthonormal bases $\Phi = \{\varphi_j\}_{j=1}^n$ and $\Psi = \{\psi_\nu\}_{\nu=1}^n$ in $\mathbb{R}^n$, and all $m \leqslant c_1 n$.

Theorem 5 is a consequence of the following result.

Theorem 6. *There are absolute constants* $0<\gamma_0<1$, $c_3>0$, *and* $c_4>0$ *such that*

$$\mu_n\left\{\varepsilon\in E_n : \rho\left(\sum_{\nu=1}^{n}\varepsilon_\nu\psi_\nu, L\right)\leqslant c_4\right\}\leqslant\gamma_0^n$$

for any $n=1,2,\ldots$, *any orthonormal basis* $\Psi=\{\psi_\nu\}_{\nu=1}^{n}$ *in* $\mathbb{R}^n$, *and any linear subspace* L *of* $\mathbb{R}^n$ *of dimension* $\dim L\leqslant c_3 n$.

Gram matrices of the systems of uniformly bounded functions

Classical Grothendieck inequality is equivalent to the following

Proposition

Let $Z = \{z_j\}_{j=1}^N \subset \mathbb{R}^N$, $|z_j| \leqslant 1$, $j = 1, \dots, N$ and $W = \{w_k\}_{k=1}^N \subset \mathbb{R}^N$, $|w_k| \leqslant 1$, $k = 1, \dots, N$.

There exists the set of functions $\{f_j\}_{j=1}^N$, $\{g_k\}_{k=1}^N$ with $\|f_j\|_{L^\infty(0,1)} \leqslant 2$, $\|g_k\|_{L^\infty(0,1)} \leqslant 2$, $j, k = 1, \dots, N$, such that

$$\langle z_j, w_k \rangle = \int_0^1 f_j(t) g_k(t)\, dt, \quad j, k = 1, \dots, N.$$

The point is that if $Z = W$ it is not always possible to find $\{f_j\}_{j=1}^N$, $\|f_j\|_{L^\infty(0,1)} \leqslant K$, $j = 1, \ldots, N$, such that

$$\langle z_j, z_k \rangle = \int_0^1 f_j(t) f_k(t)\, dt. \quad (7)$$

The best possible estimate of $\max_j \|f_j\|_{L^\infty(0,1)}$ under requirement (7) is $(\log N)^{1/2}$.

The problem about conditions on Z which guarantee the existence of uniformly bounded functions $\{f_j\}_{j=1}^N$ such that (7) holds is important for the orthogonal series theory.

It is important also for computer science.

Let $G = (V, E)$ – graph.

Definition

Grothendieck constant of the graph G is a smallest constant K such that for any $A\colon E \to R$

$$\sup_{\{f_k\}\subset B_H} \sum_{(u,v)\in E} A\{(u,v)\}\langle f_k, f_v\rangle \leqslant$$

$$\leqslant K \sup_{\varepsilon_u=\pm 1} \sum_{(u,v)\in E} A\{(u,v)\}\varepsilon_k \cdot \varepsilon_v$$

(here B_H – the unit ball of some Hilbert space).

The problem mentioned above (see (7)) is closely connected with the estimate of Grothendieck constant.

A. Olevskii in 1975 stated the following

PROBLEM

Suppose that $\{z_j\}_{j=1}^N \subset \mathbb{R}^N$, $|z_j| \leqslant 1$, $j = 1, \dots, N$, and

$$\|G_Z\|_{\mathrm{op}} \leqslant R \quad G_Z = \{\langle z_j, z_k\rangle\}_{j,k=1}^N. \qquad (8)$$

It is possible to find the set of functions $\{f_j\}_{j=1}^N$ such that $\|f_j\|_{L^\infty(0,1)} \leqslant C(R)$ and (7) is satisfied?

THEOREM E
(B.K., Russian Math. Surv., 2022, № 1)

If the conditions (8) are satisfied for the set $\{z_j\}_{j=1}^N \subset \mathbb{R}^N$, $|z_j| \leqslant 1$, $j = 1, \dots, N$, then there exists the set of functions $F = \{f_j\}_{j=1}^N \subset L^\infty(0,1)$ such that

1) $|f_j(x)| = (2R)^{1/2}$ for almost all x and $j = 1, \dots, N$;

2) $\langle z_j, z_k \rangle = \int_0^1 f_j f_k \, dt \quad \text{if } j \neq k,\ 1 \leqslant j, k \leqslant N.$

Thank you for your attention!