DVORETZKY'S THEOREM BY GAUSSIAN METHOD

IVAN MATSAK AND ANATOLIJ PLICHKO

Abstract. A complete proof of the Dvoretzky theorem, accessible to graduate students, is given.

In this note is given a complete proof of the well known Dvoretzky theorem on the almost spherical (or rather ellipsoidal) sections of convex bodies. Our proof follows Pisier [18],[19]. It is accessible to graduate students. In the references we list papers containing other proofs of Dvoretzky's theorem.

1. Gaussian random variables.

Let ξ be a real valued random variable (r.v.) on a probability space $(\Omega, \mathcal{B}, \mathbf{P})$. Put $\mathbf{E}\xi := \int_\Omega \xi(\omega) d\mathbf{P}$. If $\xi \geq 0$, then $\mathbf{E}\xi = \int_0^\infty \mathbf{P}\{\xi \geq t\}dt = \int_0^\infty (1 - \mathbf{P}\{\xi < t\})dt$ [6], 15.6.

Definition 1. An r.v. g with a distribution function

$$\mathbf{P}\{g < t\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \exp(-\frac{u^2}{2}) du \ , \quad t \in \mathbb{R} \ ,$$

is called to be *a standard Gaussian r.v.* Its *characteristic function* is (see f.e. [3], 17.2)

$$\phi(t) := \mathbf{E} \exp(itg) = \exp(-\frac{t^2}{2}) \ .$$

Proposition 1. *Let $(g_j)_1^n$ be independent standard Gaussian r.v. and $\sum_1^n a_j^2 = 1$. Then $\sum_1^n a_j g_j$ is a standard Gaussian r.v. as well.*

Proof. Let us find the characteristic function of r.v. $\sum_1^n a_j g_j$. As is known ([3], 15.12), the characteristic function of a sum of independent r.v. equals to the product of the characteristic functions of summands. So

$$\mathbf{E} \exp(it \sum_1^n a_j g_j) = \prod_{j=1}^n \mathbf{E} \exp(ita_j g_j) =$$

$$\prod_1^n \exp(-\frac{a_j^2 t^2}{2}) = \exp(-\frac{t^2}{2} \sum_1^n a_j^2) = \exp(-\frac{t^2}{2})$$

i.e we obtain the characteristic function of standard Gaussian r.v. However, the distribution of an r.v. is uniquely determined by its characteristic function [3], 10.3. Hence, $\sum_1^n a_j g_j$ is a standard Gaussian r.v. $\square$

Proposition 2. *There exists a constant $c > 0$ such that for any independent standard Gaussian r.v. $(g_j)_1^n$*

$$\mathbf{E} \max_{j \leq n} |g_j| \geq c(\ln n)^{1/2}.$$

Proof. By definition,

$$\mathbf{E}\max_{j\le n}|g_j| = \int_0^\infty [1-\mathbf{P}\{\max_{j\le n}|g_j|<t\}]dt = \int_0^\infty [1-\mathbf{P}\{|g_1|<t,\dots,|g_n|<t\}]dt =$$

(since g_j are independent and identically distributed) $= \displaystyle\int_0^\infty [1-(\mathbf{P}\{|g|<t\})^n]dt =$

$$\int_0^\infty [1-(1-\mathbf{P}\{|g|>t\})^n]dt \ge (\text{for any } a>0) \ge$$

$$\int_0^a [1-(1-\mathbf{P}\{|g|>t\})^n]dt \ge a[1-(1-\mathbf{P}\{|g|>a\})^n].$$

Now

$$\mathbf{P}\{|g|>a\} = \sqrt{\frac{2}{\pi}}\int_a^\infty \exp(-\frac{t^2}{2})dt \ge \sqrt{\frac{2}{\pi}}\int_a^{a+1}\exp(-\frac{t^2}{2})dt \ge \sqrt{\frac{2}{\pi}}\exp(-\frac{(a+1)^2}{2}) .$$

Take $a=(\ln n)^{1/2}$. Then for sufficiently large n

$$\mathbf{P}\{|g|>a\} > \sqrt{\frac{2}{\pi}}\exp(-\frac{((\ln n)^{1/2}+1)^2}{\sqrt{2}\cdot\sqrt{2}}) \ge \sqrt{\frac{2}{\pi}}\exp(-\frac{\ln n}{\sqrt{2}}) \ge \frac{1}{n} .$$

Hence, for some $c>0$

$$\mathbf{E}\max_{j\le n}|g_j| > (\ln n)^{1/2}[1-(1-\frac{1}{n})^n] \ge (\ln n)^{1/2}\cdot(1-\frac{1}{e}) \ge c\cdot(\ln n)^{1/2}$$

for sufficiently large n . $\square$

2. Standard Gaussian random vectors and Gaussian measures.

A random vector $\overrightarrow{g}=(g_j)_1^n$ with independent standard Gaussian components g_j we call *a standard Gaussian random vector*.

Theorem 1. *Let $\overrightarrow{g}$ be a standard Gaussian random vector and let U be an orthogonal matrix in $\mathbb{R}^n$. Then $U\overrightarrow{g}$ is a standard Gaussian random vector as well.*

Proof. Let $\phi(\overrightarrow{t}) := \mathbf{E}\exp(i\langle\overrightarrow{t},\overrightarrow{g}\rangle) = \exp(-\frac{1}{2}\sum_{j=1}^n t_j^2)$ be the *characteristic function* of $\overrightarrow{g}$. Let us find the characteristic function $\phi_U(\overrightarrow{t})$ of $U\overrightarrow{g}$

$$\phi_U(\overrightarrow{t}) = \mathbf{E}\exp(i\langle\overrightarrow{t},U\overrightarrow{g}\rangle) = \mathbf{E}\exp(i\langle U^*\overrightarrow{t},\overrightarrow{g}\rangle) =$$
$$=\phi(U^*\overrightarrow{t}) = \exp(-\frac{1}{2}\langle U^*\overrightarrow{t},U^*\overrightarrow{t}\rangle) = \exp(-\frac{1}{2}\langle UU^*\overrightarrow{t},\overrightarrow{t}\rangle) .$$

However, U is an orthogonal matrix, so $UU^*=I$. Hence,

$$\phi_U(\overrightarrow{t}) = \phi(\overrightarrow{t}) = \exp(-\frac{1}{2}\sum_{j=1}^n t_j^2) .$$

But the distribution of a random vector is uniquely determined by its characteristic function ([3], 10.6). Therefore, $U(\overrightarrow{g})$ is a standard Gaussian random vector. $\square$

Definition 2. A measure $\gamma(A) := \mathbf{P}\{g\in A\}$ ($A\subset\mathbb{R}$ is Borel) is called a *Gaussian measure* (generated by a standard Gaussian r.v. g).

Of course, different standard Gaussian r.v. generate the same Gaussian measure γ.

A Gaussian random vector $\overrightarrow{g}$ generates a *Gaussian measure* $\gamma_n(A) = \mathbf{P}\{\overrightarrow{g} \in A\}$ in $\mathbb{R}^n$. The fundamental property of Gaussian measure is its invariance under rotations. More exactly, Theorem 1 implies

Corollary. *Let a Gaussian random vector $\overrightarrow{g}$ generate a Gaussian measure γ_n and let U be an orthogonal matrix in $\mathbb{R}^n$. Then $U\overrightarrow{g}$ generates the same measure γ_n.*

Lemma 1. $\int_{\mathbb{R}} \exp(at)d\gamma(t) = \exp \frac{a^2}{2}$.

Proof.

$$\int_{\mathbb{R}} \exp(at)d\gamma(t) = \int_{\mathbb{R}} \exp(at)\frac{1}{\sqrt{2\pi}}\exp(-\frac{t^2}{2})dt = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \exp(\frac{a^2-(t-a)^2}{2})dt =$$

$$\exp(\frac{a^2}{2})\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} \exp(-\frac{t^2}{2})dt = \exp\frac{a^2}{2} \ . \square$$

Proposition 3. *Let $\overrightarrow{f} = (f_1, \dots, f_n) \in \mathbb{R}^n$. Then*

$$\int_{\mathbb{R}^n} \exp\langle\overrightarrow{f}, \overrightarrow{t}\rangle d\gamma_n(\overrightarrow{t}) = \exp(\frac{1}{2}\|\overrightarrow{f}\|_2^2) \ .$$

Proof. Let $\overrightarrow{t} = (t_1, \dots, t_n)$. Then

$$\int_{\mathbb{R}^n} \exp\langle\overrightarrow{f}, \overrightarrow{t}\rangle d\gamma_n(\overrightarrow{t}) = \int_{\mathbb{R}^n} \exp(\sum_1^n f_i t_i)d\gamma_n(\overrightarrow{t}) = \text{(by the Fubini theorem)} =$$

$$= \prod_1^n \int_{\mathbb{R}} \exp(f_i t_i)d\gamma(t_i) = \text{(by Lemma 1)} = \prod_1^n \exp(\frac{1}{2}f_i^2) = \exp(\frac{1}{2}\|\overrightarrow{f}\|_2^2) \ . \square$$

3. Some mathematical analysis.

Definition 3. A vector function $x(\theta) = (x_1(\theta), \dots, x_n(\theta))$ from $\mathbb{R}$ to $\mathbb{R}^n$ is said to be *differentiable in a point* θ if there are usual derivatives $x_1'(\theta), \dots, x_n'(\theta)$. A vector function $x'(\theta) = (x_1'(\theta), \dots, x_n'(\theta))$ is called a *derivative* of $x(\theta)$. A mapping $F(x)$ from $\mathbb{R}^n$ to $\mathbb{R}$ is called *differentiable in a point* x if there exist a linear functional (denote it by $F'(x)$ and its using to $y \in \mathbb{R}^n$ by $\langle F'(x), y\rangle$) such that

$$F(y) - F(x) = \langle F'(x), y - x\rangle + o(\|y - x\|)$$

if $\|y - x\| \to 0$.

Proposition 4. *Let a vector function $x(\theta)$ be differentiable in a point θ and let a map $F : \mathbb{R}^n \to \mathbb{R}$ be differentiable in $x(\theta)$. Then the usual function $F(x(\theta))$ is differentiable in θ, moreover*

$$[F(x(\theta))]' = \langle F'(x(\theta)), x'(\theta)\rangle \ .$$

Proof. Indeed,

$$F(x(\vartheta)) - F(x(\theta)) = \langle F'(x(\theta)), x(\vartheta) - x(\theta)\rangle + o(\|x(\vartheta) - x(\theta)\|)$$

Now, divide by $\vartheta - \theta$ and pass to limit. $\square$

Proposition 5. *Let $\Phi(t)$ be a convex measurable scalar function and let $f(x)$ be an integrable function on a measurable space (A, Σ, μ). Then*

$$\Phi[\int_A f(x)d\mu] \leq \frac{1}{\mu(A)} \int_A \Phi[\mu(A) \cdot f(x)]d\mu \ .$$

Proof. Let, first, f be a simple function taking values t_i on sets A_i. Then

$$\Phi[\int_A f(x)d\mu] = \Phi[\sum \frac{\mu(A_i)}{\mu(A)}\mu(A)t_i] \leq \sum \frac{\mu(A_i)}{\mu(A)}\Phi[\mu(A)t_i] = \frac{1}{\mu(A)} \int_A \Phi[\mu(A)\cdot f(x)]d\mu \ .$$

To finish the proof one can use a passage to limit. □

Recall also the *Markov* (*or Chebyshev*) *inequality* $\int_A f(x)d\mu \geq t \cdot \mu\{x : f(x) > t\}$ which (in the case of strictly increasing function Phi) we can write in the form

$$\mu\{x : f(x) > t\} = \mu\{x : \Phi(f(x)) > \Phi(t)\} \leq \frac{\int_A \Phi(f(x))d\mu}{\Phi(t)} \ . \tag{1}$$

4. Sobolev inequality.

Theorem 2. ("Sobolev inequality"). *Let $F : \mathbb{R}^n \to \mathbb{R}$ be a differentiable map and let Φ be a convex measurable function in $\mathbb{R}$. Then*

$$\int_{\mathbb{R}^n} \Phi[F(x) - \int_{\mathbb{R}^n} F(y)d\gamma_n(y)]d\gamma_n(x) \leq \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \Phi[\frac{\pi}{2}\langle F'(x), y\rangle]d\gamma_n(y)d\gamma_n(x) \ .$$

Proof. For fixed elements $x, y \in \mathbb{R}^n$ and $\theta \in [0, \frac{\pi}{2}]$ put $x(\theta) = x \cdot \sin\theta + y \cdot \cos\theta$. Then, by Proposition 4,

$$F(x) - F(y) = F(x(\frac{\pi}{2})) - F(x(0)) = \int_0^{\frac{\pi}{2}} [F(x(\theta))]'d\theta = \int_0^{\frac{\pi}{2}} \langle F'(x(\theta)), x'(\theta)\rangle d\theta \ .$$

This equality and Proposition 5 give

$$\Phi[F(x) - F(y)] = \Phi[\int_0^{\frac{\pi}{2}} \langle F'(x(\theta)), x'(\theta)\rangle d\theta] \leq \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \Phi[\frac{\pi}{2}\langle F'(x(\theta)), x'(\theta)\rangle]d\theta \ .$$

Integrating the last inequality over x and y we obtain

$$\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \Phi[F(x) - F(y)]d\gamma_n(y)d\gamma_n(x) \leq \tag{2}$$

$$\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \frac{2}{\pi} \int_0^{\frac{\pi}{2}} \Phi[\frac{\pi}{2}\langle F'(x(\theta)), x'(\theta)\rangle]d\theta d\gamma_n(y)d\gamma_n(x) \ .$$

Now observe that $x'(\theta) = x\cos\theta - y\sin\theta$, that for every fixed θ the map $(x, y) \to (x(\theta), x'(\theta))$ is a rotation in $\mathbb{R}^n \times \mathbb{R}^n$ an that (*this is the crucial point*) $\gamma_n \times \gamma_n$ is invariant with respect to this map. Thus the right-hand side of (2) is equal to

$$\int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \Phi[\frac{\pi}{2}\langle F'(x), y\rangle]d\gamma_n(y)d\gamma_n(x) \ . \tag{3}$$

By Proposition 5,

$$\int_{\mathbb{R}^n} \Phi[F(x) - \int_{\mathbb{R}^n} F(y)d\gamma_n(y)]d\gamma_n(x) \leq \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \Phi[F(x) - F(y)]d\gamma_n(y)d\gamma_n(x) \ . \tag{4}$$

¿From (4),(2) and (3) we get the theorem. □

5. Gaussian random elements in normed spaces. Second weak moment.

Definition 4. We say that X is a *standard Gaussian random element* (*r.e.*) *in a normed space* E if there are linearly independent elements $(e_i)_1^n$ in E and independent scalar standard Gaussian r.v. $(g_i)_1^n$ such that $X = \sum_1^n g_i e_i$.

Definition 4 is connected with the definition of standard Gaussian random vector (Section 2) in the following way. Consider a map $T : \mathbb{R}^n \to E$:

$$T(a_1, \ldots, a_n) = \sum_1^n a_i e_i \ . \tag{5}$$

This map generates a standard random vector $T^{-1}X(\omega) = (g_1(\omega), \ldots, g_n(\omega))$ in $\mathbb{R}^n$.

Proposition 6. *Let $(X_k)_1^m$ be independent identically distributed standard Gaussian r.e. in a normed space E and let $\sum_1^m a_k^2 = 1$. Then $\sum_1^m a_k X_k$ is a standard Gaussian r.e. with the same distribution as X_k.*

Proof. Let $X_k = \sum_{i=1}^n g_{ki} e_i$. Then

$$\sum_1^m a_k X_k = \sum_{k=1}^m a_k \sum_{i=1}^n g_{ki} e_i = \sum_{i=1}^n (\sum_{k=1}^m a_k g_{ki}) e_i \ .$$

Since $(g_{ki})_{k,i}$ all are independent and (by Proposition 1) for every i , $\sum_{k=1}^m a_k g_{ki}$ is a standard Gaussian r.v., hence $\sum_{k=1}^m a_k X_k$ is a standard Gaussian r.e. with the same distribution as X_k. $\square$

Definition 5. *The second weak moment* of r.e. X in a normed space E is, by definition,

$$\sigma = \sigma(X) := \sup_{\|f\|=1, f\in E^*} (\mathbf{E} f^2(X))^{1/2} \ .$$

Proposition 7. *Let $X = \sum_1^n g_i e_i$ be a standard Gaussian r.e. in a normed space E and let T be defined by* (5)*. Then $\sigma = \|T\|$.*

Proof. Obviously, without restriction of generality, one can suppose $\text{lin}(e_i)_1^n = E$. Let (f_i) be the biorthogonal to (e_i) functionals. Then for $f \in E^*$ we have $f = \sum_1^n c_i f_i$ and

$$\sigma = \sup_{\|f\|=1} (\mathbf{E}\langle \sum_1^n g_i e_i, \sum_1^n c_i f_i \rangle^2)^{1/2} = \sup_{\|f\|=1} (\mathbf{E}(\sum_1^n g_i c_i)^2)^{1/2} = \sup_{\|f\|=1} (\sum_1^n c_i^2 \mathbf{E} g_i^2)^{1/2} =$$

$$\sup_{\|f\|=1} (\sum_1^n c_i^2)^{1/2} = \sup_{\|f\|=1} \|T^* f\| = \|T^*\| = \|T\| \ . \square$$

6. Tail behavior (the concentration of measure phenomenon).

Theorem 3. (Maurey-Pisier; [19], Theorem 4.7). *Let $X = \sum_1^n g_i e_i$ be a standard Gaussian r.e. in a normed space E. Then for any $t > 0$*

$$\mathbf{P}\{|\|X\| - \mathbf{E}\|X\|| > t\} \le 2\exp(-\frac{2t^2}{\pi^2\sigma^2}) \ . \tag{6}$$

Proof. Let T be the operator of Section 5. Then (6) is equivalent to

$$\gamma_n\{x \in \mathbb{R}^n : |\|Tx\| - \int_{\mathbb{R}^n} \|Tx\| d\gamma_n| > t\} \le 2\exp(-\frac{2t^2}{\pi^2\sigma^2}) \ . \tag{7}$$

Put $F(x) = \|Tx\|$ and suppose temporary that $\|\cdot\|$ is differentiable. Since for any $x, y \in \mathbb{R}^n$

$$|F(x) - F(y)| \le \|T(x-y)\| \le \|T\|\|x-y\|_2 \ ,$$

$\|F'(x)\|_2 \le \sigma$ for every $x \in \mathbb{R}^n$ (recall, $\sigma = \|T\|$!). Thus, by Proposition 3, for any $\lambda \in \mathbb{R}$

$$\int_{\mathbb{R}^n} \exp[\frac{\lambda\pi}{2}\langle F'(x), y\rangle] d\gamma_n(y) = \exp[\frac{1}{2}(\frac{\lambda\pi}{2})^2 \|F'(x)\|_2^2] \le \exp[\frac{1}{2}(\frac{\lambda\pi}{2})^2\sigma^2] \ . \tag{8}$$

Taking in Theorem 2 $\Phi(t) = \exp(\lambda t)$, we get

$$\begin{aligned} &\int_{\mathbb{R}^n} \exp[\lambda\{F(x) - \int_{\mathbb{R}^n} F(y) d\gamma_n(y)\}] d\gamma_n(x) \le \\ &\int_{\mathbb{R}^n}\int_{\mathbb{R}^n} \exp[\lambda\frac{\pi}{2}\langle F'(x), y\rangle] d\gamma_n(y) d\gamma_n(x) \le \\ &(\text{by (8)}) \le \int_{\mathbb{R}^n} \exp[\frac{1}{2}(\frac{\lambda\pi\sigma}{2})^2] d\gamma_n(x) = \\ &\exp[\frac{1}{2}(\frac{\lambda\pi\sigma}{2})^2] \ . \end{aligned} \tag{9}$$

Using the Markov inequality (1), taking there $\Phi(t) = \exp(\lambda t)$, $f(x) = F(x) - \int_{\mathbb{R}^n} F(y) d\gamma_n(y)$ and $\mu = \gamma_n(x)$, we get from (9)

$$\gamma_n\{x : F(x) - \int F(y) d\gamma_n(y) > t\} \le \exp[\frac{1}{2}(\frac{\lambda\pi\sigma}{2})^2 - \lambda t] \ .$$

Putting $\lambda = \frac{4t}{(\pi\sigma)^2}$ we obtain

$$\gamma_n\{x : F(x) - \int_{\mathbb{R}^n} F(y) d\gamma_n(y) > t\} \le \exp[-\frac{2t^2}{(\pi\sigma)^2}] \ .$$

Clearly, the same inequality holds for $-F$, so we finally obtain (7), and hence (6), for differentiable norms.

To finish the proof, observe that in the finite dimensional space $\mathrm{lin}(e_i)_1^n$ one can approximate any norm by differentiable one as exact as need. □

Remark. The estimation (6) is an important result in theory of Banach valued Gaussian r.e. It demonstrates that the distribution of $\|X\|$ is concentrated near $\mathbf{E}\|X\|$ as dense as the distribution of Gaussian r.v. $\sigma \cdot g$ near 0. In a somewhat different form:

$$\mathbf{P}\{|\|X\| - \mathbf{m}\|X\|| > t\} \le 2(1 - \Phi(\frac{t}{\sigma})) \ , \tag{10}$$

(where $\mathbf{m}\|X\|$ is the median of $\|X\|$ and $\Phi(t)$ is the distribution function of g), the concentration phenomenon is presented in books Ledoux-Talagrand ([14], 3.1) and Lifshits ([15], Chapt.12).

In fact, the concentration phenomenon in the form (10) was contained as far as in Borell [2] and Sudakov-Tsirelson [21].

In connection with Theorem 3, note also papers Fernique [7], Landau-Shepp [13] and Skorokhod [20], where for the first time were established exponential integrability of the norm for Gaussian r.e., and Yurinskii [24], where for the first time was proposed the martingale approach to estimations of probabilities of large deviations for the norm of sums of independent r.e. in Banach spaces.

7. Lévy's inequality.

Recall that an r.e. X in a normed space E is called *symmetric* if $\mathbf{P}\{X \in A\} = \mathbf{P}\{X \in -A\}$ for any Borel set $A \subset E$.

Theorem 4. (Lévy's inequality; [11], II.3). *Let $(X_i)_1^n$ be symmetric independent r.e. in a normed space E and $S = \sum_1^n X_i$. Then for every $t > 0$*

$$2\mathbf{P}\{\|S\| > t\} \geq \mathbf{P}\{\max_{i \leq n} \|X_i\| > t\} .$$

Proof. Put $\tau = \tau(\omega) = \min\{i : \|X_i\| > t\}$. Then

$$\mathbf{P}\{\|S\| > t\} \geq \sum_{i=1}^{n} \mathbf{P}\{\|S\| > t\ ,\ \tau = i\} . \tag{11}$$

Now since $(X_1, \ldots, X_n)$ and $(-X_1, \ldots, -X_{i-1}, X_i, -X_{i+1}, \ldots, -X_n)$ are identically distributed and τ depends on $\{\|X_i\|\}_1^n$ only, we have

$$\mathbf{P}\{\|S\| > t\ ,\ \tau = i\} \ = \ \mathbf{P}\{\|X_i - R_i\| > t\ ,\ \tau = i\} ,$$

where $R_i = S - X_i$. This equality and (11) imply

$$2\mathbf{P}\{\|S\| > t\} \geq \sum_{i=1}^{n} \mathbf{P}\{\|S\| > t\ ,\ \tau = i\} + \mathbf{P}\{\|X_i - R_i\| > t\ ,\ \tau = i\} . \tag{12}$$

By triangle inequality

$$2\|X_i\| \leq \|X_i + R_i\| + \|X_i - R_i\| = \|S\| + \|X_i - R_i\| .$$

Hence

$$\begin{aligned}
&\mathbf{P}\{\|S\| > t\ ,\ \tau = i\} + \mathbf{P}\{\|X_i - R_i\| > t\ ,\ \tau = i\} \geq \\
&\mathbf{P}\{\|S\| > t \bigvee \|X_i - R_i\| > t\ ,\ \tau = i\} \geq \mathbf{P}\{\|S\| + \|X_i - R_i\| > 2t\ ,\ \tau = i\} \geq \\
&\mathbf{P}\{2\|X_i\| > 2t\ ,\ \tau = i\} = \mathbf{P}\{\|X_i\| > t\ ,\ \tau = i\} = \mathbf{P}\{\tau = i\} .
\end{aligned}$$

This inequality and (12) imply

$$2\mathbf{P}\{\|S\| > t\} \geq \sum_{i=1}^{n} \mathbf{P}\{\tau = i\} = \mathbf{P}\{\max_{i \leq n} \|X_i\| > t\} . \ \square$$

Corollary. *There exists $c > 0$ such that for any normed space E and any standard Gaussian r.e. $X = \sum_1^n g_i e_i$ in E*

$$\mathbf{E}\|X\| \geq c \cdot \min_{i \leq n} \|e_i\| \cdot (\ln n)^{1/2} .$$

Proof. Evidently, standard Gaussian r.e. are symmetric. Hence one can use Lévy's inequality taking $S = X$ and $X_i = g_i e_i$. We have for any t

$$2\mathbf{P}\{\|X\| > t\} \geq \mathbf{P}\{\max_{i \leq n} \|g_i e_i\| > t\} .$$

Integrating over t we receive

$$\begin{aligned}
\mathbf{E}\|X\| \geq \frac{1}{2}\mathbf{E}\max_{i \leq n} \|g_i e_i\| \geq \frac{1}{2}\min_{i \leq n} \|e_i\| \cdot \mathbf{E}\max_{i \leq n} |g_i| \geq \\
\text{(Proposition 2)} \geq c \cdot \min_{i \leq n} \|e_i\| \cdot (\ln n)^{1/2} . \ \square
\end{aligned}$$

8. Dvoretzky-Rogers theorem.

Theorem 5. (Dvoretzky-Rogers, [5]). *Let $\|\cdot\|$ be a norm on $\mathbb{R}^N$ and let D be the ellipsoid of maximal volume inscribed in the unit $\|\cdot\|$-ball B. Then there exists a basis $(e_i)_1^N$, orthonormal with respect to D, such that* $1 \geq \|e_i\| \geq 2^{-(N-1)/(N-i)}$, $i = 1, \dots, N-1$.

Proof. We choose the basis (e_i) inductively in the following way.

Let e_1 be a vector in D with maximal norm (clearly $\|e_1\| = 1$).

Given $e_1, \dots, e_i$, choose e_{i+1} in $D \cap (e_1, \dots, e_i)^\perp$ with maximal possible norm ($\perp$ is the orthogonal complement with respect to D).

Then for any $x \in \mathrm{lin}(e_i, \dots, e_N) \cap D$

$$\|x\| \leq \|e_i\| . \tag{13}$$

Now consider the ellipsoid

$$U = \{\sum_{j=1}^N a_j e_j : \frac{\sum_1^{i-1} a_j^2}{a^2} + \frac{\sum_i^N a_j^2}{b^2} \leq 1\} \ .$$

Of course, this ellipsoid depends on i and scalars a and b which we shall choose late on.

If $\sum_{j=1}^N a_j e_j \in U$ then $\sum_1^{i-1} a_j e_j \in aD$ and thus $\| \sum_1^{i-1} a_j e_j\| \leq a$. In the same way, $\| \sum_{j=i}^N a_j e_j\| \leq b$ and thus, by (13), $\| \sum_i^N a_j e_j\| \leq b\|e_i\|$.

Choosing $a = 1/2$, $b = 1/(2\|e_i\|)$, we get that $U \subset B$. On the other hand

$$\mathrm{vol} U = \frac{1}{2^{i-1}} \cdot \frac{1}{(2\|e_i\|)^{N-i}} \mathrm{vol} D$$

so necessarily

$$\frac{1}{2^{i-1}} \cdot \frac{1}{(2\|e_i\|)^{N-i}} \leq 1$$

or $\|e_i\| \geq 2^{-(N-1)/(N-i)}$. $\square$

Corollary 1. *Let E be an N-dimensional normed space and $\overline{N} = [\frac{N}{2}]$. Then there exist $(e_i)_1^{\overline{N}} \subset E$ such that $\|e_i\| \geq \frac{1}{2}$ and for any $(a_i)_1^{\overline{N}} \subset \mathbb{R}$*

$$\| \sum_1^{\overline{N}} a_i e_i \| \leq (\sum_1^{\overline{N}} a_i^2)^{1/2} . \tag{14}$$

Corollary 2. *For every N-dimensional normed space E there exists a subspace $\overline{E} \subset E$ of dimension $\overline{N} = [\frac{N}{2}]$ and $\overline{E}$-valued standard Gaussian r.e. X with $\mathbf{E}\|X\| = 1$ and $\sigma^2(X) \leq c/\ln N$, where $c > 0$ is an absolute constant.*

Proof. Let $\overline{X} = \sum_1^{\overline{N}} g_i e_i$ where (g_i) are independent Gaussian r.v. and (e_i) are from Corollary 1. Then, by (14),

$$\sigma(\overline{X}) := \sup_{\|f\|=1} (\mathbf{E} f^2(\overline{X}))^{1/2} = \sup_{\|f\|=1} (\sum_{i=1}^{\overline{N}} f^2(e_i))^{1/2} = \sup_{\|a\|_2=1} \| \sum_{i=1}^{\overline{N}} a_i e_i \|^{1/2} \leq 1 \ .$$

On the other hand, by Corollary of Theorem 4

$$\mathbf{E}\|\overline{X}\| \geq c_1 (\ln \overline{N})^{1/2} \geq c_2 (\ln N)^{1/2} \ .$$

Now put $X = \overline{X}/\mathbf{E}\|\overline{X}\|$. $\square$

9. Two geometric lemmas.

Lemma 2. *Let $\|\cdot\|$ be any norm on $\mathbb{R}^n$ with unit ball B and unit sphere S. Let $\delta > 0$. There is a δ-net $A \subset S$ with cardinality*

$$\mathrm{card}A \leq (1 + \frac{2}{\delta})^n \ . \tag{15}$$

Proof. Let A be a maximal subset of S such that $\|a - b\| \geq \delta$ for all $a, b \in A$, $a \neq b$. Clearly, by maximality, A is a δ-net of S. To majorize card$A$ we note that balls $a + \frac{\delta}{2}B$, $a \in A$ are disjoint and included into $(1 + \frac{\delta}{2})B$. Therefore

$$\sum_{a \in A} \mathrm{vol}(a + \frac{\delta}{2}B) \leq \mathrm{vol}((1 + \frac{\delta}{2})B) = (1 + \frac{\delta}{2})^n \mathrm{vol}B \ ;$$

hence

$$\mathrm{card}A \cdot (\frac{\delta}{2})^n \mathrm{vol}B \leq (1 + \frac{\delta}{2})^n \mathrm{vol}B \ .$$

This inequality implies (15). □

Lemma 3. *For each $\varepsilon > 0$ there is a $\delta = \delta(\varepsilon)$, $0 < \delta < 1$, with the following property. Let $\|\cdot\|$ be an arbitrary norm on $\mathbb{R}^n$ with the unit sphere S. Let A be a δ-net in S and let $x_1, \ldots, x_n$ be elements of a normed space E . If for every $a = (a_1, \ldots, a_n) \in A$*

$$1 - \delta \leq \|\sum_1^n a_k x_k\| \leq 1 + \delta \ ,$$

then for every $a \in S$

$$(1 + \varepsilon)^{-1} \leq \|\sum_1^n a_k x_k\| \leq 1 + \varepsilon; .$$

Proof. There is a^0 in A such that $\|a - a^0\| \leq \delta$ hence $a = a^0 + \lambda_1 a'$ with $|\lambda_1| \leq \delta$ and $a' \in S$. Continuing this process we obtain $a = a^0 + \lambda_1 a^1 + \lambda_2 a^2 + \cdots$ with $a^j \in A$ and $|\lambda_j| \leq \delta^j$. It follows that

$$\|\sum_1^n a_k x_k\| \leq \sum_{j \geq 0} \delta^j \|\sum_{k=1}^n a_k^j x_k\| \leq \frac{1 + \delta}{1 - \delta} \ .$$

Similarly

$$\|\sum_1^n a_k x_k\| \geq \|\sum_1^n a_k^0 x_k\| - \frac{\delta(1 + \delta)}{1 - \delta} \geq 1 - \delta - \frac{\delta(1 + \delta)}{1 - \delta} = \frac{1 - 3\delta}{1 - \delta} \ .$$

Hence, if $\delta > 0$ is chosen small enough so that

$$\frac{1 - 3\delta}{1 - \delta} \geq \frac{1}{1 + \varepsilon} \ \text{ and } \ \frac{1 + \delta}{1 - \delta} \leq 1 + \varepsilon \ ,$$

we obtain the announced result. Note that one can find a suitable δ depending only on ε (and *independent* on n). □

10. Dvoretzky's theorem.

Theorem 6. (Dvoretzky). *For each $\varepsilon > 0$, there is a number $\eta = \eta(\varepsilon) > 0$ with the following property. Every normed space E of dimension N contains a subspace of dimension $n = [\eta \ln N]$ which is $(1 + \varepsilon)^2$–isomorphic to l_2^n.*

Let us present first the idea of proof. We take independent copies $X_1, \ldots, X_n$ of r.e. X from Corollary 2 of Dvoretzky-Rogers theorem, $n \approx \ln N$, which is determined on a probability space $(\Omega, \mathcal{B}, \mathbf{P})$. Let $\Omega(n, \varepsilon)$ be the set of all ω in Ω such that for every $a = (a_1, \ldots, a_n)$ in the unit sphere S of Euclidean space $\mathbb{R}^n$

$$(1+\varepsilon)^{-1} \le \| \sum_1^n a_k X_i(\omega) \| \le 1 + \varepsilon \ .$$

We will show that $\mathbf{P}\{\Omega(n,\varepsilon)\} > 0$ provided that n is not too large and precisely provided $n \le \eta/\sigma^2$. This clearly yields Theorem 6.

Now the

Proof. Let $\overline{E}$ be the subspace and X be an r.e. from Corollary 2 of Theorem 5. Let $X_1, \ldots, X_n$ be independent copies of X (n we choose late on). Then for any $a = (a_1, \ldots, a_n)$ in S the r.e. $\sum_1^n a_k X_k$ has the same distribution as X (Proposition 6).

Thus we have $\mathbf{E}\| \sum_1^n a_k X_k \| = 1$. Therefore, by Theorem 3, for any $\delta > 0$

$$\mathbf{P}\{ |\| \sum_1^n a_k X_k \| - 1| > \delta \} \le 2\exp(-\frac{2\delta^2}{\pi^2\sigma^2}) < 2\exp(-\frac{\delta^2}{\sigma^2}) \ .$$

Let A be as in Lemmas 2 and 3 with $\|a\| = \|a\|_2$. Then preceding inequality implies

$$\begin{aligned} \mathbf{P}\{\exists a \in A : |\| \sum_1^n a_k X_k \| - 1| > \delta \} &\le 2(\mathrm{card} A)\exp(-\frac{\delta^2}{\sigma^2}) \le \\ (\text{by Lemma 2}) &\le 2(1+\frac{2}{\delta})^n \exp(-\frac{\delta^2}{\sigma^2}) \le \\ 2\exp(\frac{2n}{\delta})\exp(-\frac{\delta^2}{\sigma^2}) &= 2\exp(\frac{2n}{\delta} - \frac{\delta^2}{\sigma^2}) \ . \end{aligned} \tag{16}$$

Now let $\delta = \delta(\varepsilon)$ be the function of ε, given by Lemma 3. Let we choose n so that

$$\frac{2n}{\delta} \le \frac{\delta^2}{2\sigma^2} \ . \tag{17}$$

Then the probability (16) is not greater than $2\exp(-\frac{\delta^2}{2\sigma^2})$. Clearly, we can always assume that σ is small enough (say $\sigma < \delta/2$), otherwise there is nothing to prove. Hence we can assume that the right side of (16) < 1. We than obtain that with *positive* probability, for every $a \in A$

$$|\| \sum a_k X_k(\omega) \| - 1| < \delta$$

i.e.

$$1 - \delta \le \| \sum_1^n a_k X_k(\omega) \| \le 1 + \delta \ .$$

By Lemma 3 (recall, we choose $\delta = \delta(\varepsilon)$!), we conclude that with positive probability for all $a \in S$

$$(1+\varepsilon)^{-1} \le \| \sum_1^n a_k X_k(\omega) \| \le 1 + \varepsilon \ .$$

Therefore, there exists $\omega_0 \in \Omega$ such that for $x_k = X_k(\omega_0)$ and for every $a \in S$

$$(1+\varepsilon)^{-1} \le \| \sum_1^n a_k x_k \| \le 1 + \varepsilon \ .$$

By homogenity of the norm it means that $\mathrm{lin}(x_k)_1^n$ is $(1+\varepsilon)^2$-isomorphic to l_2^n .

To satisfy (17) put

$$n = [\frac{\delta^3}{4\sigma^2}] = \eta \ln N$$

(recall, $\sigma^2 \le c/\ln N$, by Corollary 2 of Theorem 5). $\square$

Acknowledgement. The first version of this note was written when the second author visited Politecnico di Milano. He express his thanks to P.Terenzi for hospitality and valuable remarks. The authors also express they thanks to M.Ostrovskii for valuable references.

References

[1] Y.Benyamini and J.Lindenstrauss, *Geometric nonlinear functional analysis.* v.I, Colloq. Publ. Amer. Math. Soc., Rhode Island, 2000.

[2] C.Borell, *The Brunn-Minkowski inequality in Gauss space ,* Invent. Math., 30(1975), 207-216.

[3] H.Cramér, *Mathematical methods of statistics,* Princeton Univ. Press, N.J., 1946.

[4] A.Dvoretzky, *Some results on convex bodies and Banach spaces,* Proc. Int. Symp. on Linear Spaces, Jerusalem, 1961, 123-160.

[5] A.Dvoretzky and C.A.Rogers, *Absolute and unconditional convergence in normed linear spaces,* Proc. Nat. Acad. Sci. USA., 36(1950), 192-197.

[6] W.Feller, *An introduction to probability theory and its applications,* v.II, John Wiley and Sons, New York e.a., 1966.

[7] K.Fernique, *Intégrabilité des vecteurs gaussiens*, C. r. Acad. sci., ser.A, 270(1970), 1698-1699.

[8] T.Figiel, *A short proof of Dvoretzky's theorem on almost spherical sections,* Compositio Math., 33(1976), 297-301.

[9] T.Figiel, J.Lindenstrauss and V.D.Milman, *The dimension of almost spherical sections of convex bodies,* Acta Math., 139(1977), 53-94.

[10] Y.Gordon, *On Dvoretzky's theorem and extensions of Slepian's lemma,* Israel Semin. Geom. Aspects Funct. Anal., Technion preprint series, No. MT-636, 1983(4), p.II.1.1-II.1.25.

[11] J.P.Kahane, *Some random series of functions,* D.C.Heath and Co, Lexington, 1968.

[12] J.L.Krivine, *Sous-espaces de dimension finie des espaces de Banach reticules,* Ann. of Math. (2) 104(1976), no. 1, 1-29.

[13] H.J.Landau and L.A.Shepp, *On the supremum of a Gaussian process*, Sankhya, 32(1970), 369-378.

[14] M.Ledoux and M.Talagrand, *Probability in Banach spaces. Isoperimetry and processes,* Springer, Berlin e.a., 1991.

[15] M.A.Lifshits, *Gaussian random functions*, TViMS, Kyiv, 1995 (Russian).

[16] V.D.Milman, *New proof of the theorem of Dvoretzky on sections of convex bodies,* Funkts. Anal. Prilozh., 5(1971), No 5, 28-37 (Russian).

[17] V.D.Milman and G.Schechtman, *Asymptotic theory of finite dimensional normed spaces,* Lect. Notes in Math., 1200(1986).

[18] G.Pisier, *Probabilistic methods in the geometry of Banach spaces,* Lecture Notes in Math., 1206(1986), 167-241.

[19] G.Pisier, *The wolume of convex bodies and Banach space geometry,* Cambridge Univ. Press, Cambridge e.a., 1989.

[20] A.V.Skorokhod, *A note on Gaussian measures in a Banach space*, Teor. Veroyatn. Prim., 15(1970), 519-520 (Russian).

[21] V.N.Sudakov and B.S.Tsirel'son, *Extremal properties of half-spaces for spherically invariant measures*, Zap. Nauch. Sem. L.O.M.I., 41(1974), 14-24 (Russian).

[22] A.Szankowski, *On Dvoretzky's theorem on almost spherical sections of convex bodies,* Israel J. Math., 17(1974), 325-338.

[23] L.Tzafriri, *On Banach spaces with unconditional bases,* Israel J. Math., 17(1974), 84-93.

[24] V.V.Yurinskii, *Exponential bounds for large deviations*, Teor. Veroyatn. Prim., 19(1974), 152-153 (Russian).

Ivan Matsak
Kyiv State University of Technology and Design
2, Nemirovich-Danchenko str.
Kyiv 04011
UKRAINE
infor1@valtek.kiev.ua

Anatolij Plichko
Department of Mathematics
Kirovograd State Pedagogical University
1, Shevchenko str.
Kirovograd 25006
UKRAINE
aplichko@kspu.kr.ua