High-Dimensional Phenomena in Convex Geometry, and Random Matrix Theory

by

Konstantin Tikhomirov

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Mathematics

Department of Mathematical and Statistical Sciences

University of Alberta

Abstract

This thesis is based on six papers. The first three fall into the field of Asymptotic Geometric Analysis, the next two — Random Matrix Theory, and the sixth — high-dimensional Random Walks.

In the first paper, we show that for any $\varepsilon \in (0, 1/2]$ and natural n there is a linear subspace E of $\mathbb{R}^n$ of dimension at least $c \ln n / \ln \frac{1}{\varepsilon}$ such that E is $(1+\varepsilon)$-Euclidean with respect to any 1-symmetric norm in $\mathbb{R}^n$. Here, $c > 0$ is a universal constant.

In the second paper, we show that, given $\varepsilon \in (0, 1/2]$, a natural n, the space ℓ_∞^n, and its random subspace E of dimension $m \geq 2$ uniformly distributed on the corresponding Grassmannian $\mathrm{G}_{n,m}$, E is $(1+\varepsilon)$-spherical with probability at least $1/2$ only if m satisfies $m \leq C\varepsilon \ln n / \ln \frac{1}{\varepsilon}$ for some universal constant $C > 0$.

In the third paper, we show that, given an n-dimensional convex polytope with $n + k$ vertices ($k \leq n$), its Banach–Mazur distance to the Euclidean ball is at least $cn/\sqrt{k}$ for some universal constant $c > 0$.

In the fourth paper, we prove that there are constants $c_1, c_2 > 0$ such that for any natural n and a $2n \times n$ random matrix A with i.i.d. entries a_{ij} satisfying $\mathbb{P}\{|a_{ij} - \lambda| \leq 1\} \leq 1/2$ for all $\lambda \in \mathbb{R}$, we have that the smallest singular value $s_{\min}(A)$ is greater than $c_1\sqrt{n}$ with probability at least $1 - \exp(-c_2 n)$.

In the fifth paper, we generalize a classical theorem of Bai and Yin regarding almost sure convergence of the smallest singular values of a sequence of random matrices with i.i.d. entries. Namely, we remove the assumption that the fourth moment of the matrix entries is bounded.

In the sixth paper (joint work with Pierre Youssef) we show that, given the standard n-dimensional Brownian motion $\mathrm{BM}_n(t)$ in $\mathbb{R}^n$ starting at the origin, and a natural N, the convex hull of $\mathrm{BM}_n(1), \mathrm{BM}_n(2), \ldots, \mathrm{BM}_n(N)$ contains the origin with a high probability whenever $N \geq \exp(Cn)$, and contains the origin with probability close to zero whenever $N \leq \exp(cn)$. Here, $C, c > 0$ are universal constants.

Preface

This thesis is based on six papers, five of which are sole works of the author and the sixth is a joint work with Pierre Youssef (presently working at Université Paris Diderot, France). The publication data for the first five papers, which constitute Chapters 2 and 3 of the thesis, are

1. K. E. Tikhomirov, Almost Euclidean sections in symmetric spaces and concentration of order statistics, J. Funct. Anal. **265** (2013), no. 9, 2074–2088.

2. K. E. Tikhomirov, The Randomized Dvoretzky's theorem in ℓ_∞^n and the χ-distribution, Geometric Aspects of Functional Analysis, Lecture Notes in Mathematics, **2116** (2014), 455–463.

3. K. E. Tikhomirov, On the distance of polytopes with few vertices to the Euclidean ball, Discrete Comput. Geom. **53** (2015), no. 1, 173–181.

4. K. E. Tikhomirov. The smallest singular value of random rectangular matrices with no moment assumptions on entries. Israel Journal of Mathematics, 2016. DOI: 10.1007/s11856-016-1287-8.

5. K. Tikhomirov, The limit of the smallest singular value of random matrices with i.i.d. entries, Adv. Math. **284** (2015), 1–20.

The joint work with Pierre Youssef, which constitutes Chapter 4 of the thesis — *K. Tikhomirov and P. Youssef, When does a discrete-time random walk in* $\mathbb{R}^n$ *absorb the origin into its convex hull? 2015* — is accepted for publication in the Annals of Probability. All parts of this work were written in collaboration with P. Youssef, which makes it impossible to separate our contributions to these results. A permission to include this work in the thesis was obtained from P. Youssef.

Acknowledgments

Undoubtedly, none of the results of this thesis would be possible without constant support of my supervisor Professor Nicole Tomczak-Jaegermann. From the very beginning of our work together, Nicole helped with finding interesting and important problems in the direction of Asymptotic Geometric Analysis — a field which at that time was completely unfamiliar to me. Later, also due to her support, I started to work with random matrices. What I am most grateful to Nicole for is her constant attention and interest in my research. Her suggestions helped to considerably improve the structure, proofs and style within all my works. I am very glad that I have had the opportunity to work with the world-class mathematician during these five years in Alberta!

I would like to thank my co-supervisor Professor Vlad Yaskin for many good discussions and for his support. I would also like to thank Professor Alexander Litvak for introducing me to some very interesting problems in convex geometry (one of which is presented in this thesis); Professor Vladimir Troitsky for his help when I just arrived to Canada and for teaching a very interesting course on Banach lattices. I am very grateful to Pierre Youssef who was a postdoc in Alberta for two years and who is a strong mathematician and a very good friend. I would like to thank Lisa Rebrova for her collaborative work on invertibility of square matrices.

I would like to thank Professor Peter Pivovarov for his invitation to the Analysis Seminar held at the University of Missouri, Columbia in Oct. 2013. I am very grateful to Professor Alain Pajor for the invitation to visit Université Paris-Est Marne-la-Vallée (France) in Nov.–Dec. 2014, and to Professor Djalil Chafaï for invitations to visit Université Paris–Dauphine in Jul. and Oct. 2015. I would like to thank Professor Mark Rudelson for his invitation to visit University of Michigan in Sept. 2015 and Professor Galyna Livshyts for inviting me to the Georgia Institute of Technology in Nov. 2015.

I would like to thank Professors Djalil Chafaï, Ronen Eldan, Alexander Litvak, Anya Lytova, Vitali Milman, Alain Pajor, Mark Rudelson, Roman Vershynin for helpful and valuable discussions.

I am grateful to the Department of Mathematical and Statistical Sciences, University of Alberta, for creating supporting and friendly environment.

My work at the University of Alberta was supported by the following grants and scholarships: J.M. Mitchell Graduare Scholarship and Research Prize (Dept. of Math. and Stats.), Dean's Excellence Award (Faculty of Science), Andrew Stewart Memorial Graduate Prize (Faculty of Graduate Studies and Research), PIMS Graduate Scholarship (Pacific Institute for the Mathematical Sciences).

Contents

1 Introduction **1**

1.1 The Banach–Mazur distance to the Euclidean space . . . 3

1.2 The extreme singular values of random matrices . . . 8

1.3 Convex hulls of random walks in high dimensions . . . 15

2 Asymptotic Geometric Analysis **18**

2.1 Notation . . . 18

2.2 Almost Euclidean Sections in Symmetric Spaces and Concentration of Order Statistics . . . 19

2.2.1 Introduction . . . 19

2.2.2 Preliminaries . . . 20

2.2.3 Dilation operators and concentration of order statistics . . . 22

2.2.4 Dependence on ε in Dvoretzky's theorem . . . 28

2.2.5 Negative results on concentration . . . 30

2.3 The Randomized Dvoretzky's Theorem in ℓ_∞^n and the χ-distribution . . . 35

2.3.1 Introduction . . . 35

2.3.2 Preliminaries . . . 36

2.3.3 Random subspaces of ℓ_∞^n . . . 37

2.4 On the Distance of Polytopes with Few Vertices to the Euclidean Ball . . . 44

2.4.1 Introduction . . . 44

2.4.2 Lower bounds for $\mathrm{dist}(P, B_2^n)$ . . . 45

3 The Smallest Singular Value of Random Matrices **52**

3.1 The Smallest Singular Value of Random Rectangular Matrices with no Moment Assumptions on Entries . . . 52

3.1.1 Introduction . . . 52

3.1.2 Preliminaries . . . 55

3.1.3 The distribution of $\mathrm{d}\big(A_H y, V_{A,B}(H,E)\big)$ . . . 61

3.1.4 Decomposition of S^{n-1} and proof of Theorem 3.1 67
3.2 The Limit of the Smallest Singular Value of Random Matrices with i.i.d. Entries 75
3.2.1 Introduction 75
3.2.2 Preliminaries 78
3.2.3 Norms of coordinate projections of random vectors 80
3.2.4 Matrix truncation and proof of Theorem 3.18 89

4 When does a discrete-time random walk in $\mathbb{R}^n$ absorb the origin into its convex hull? 94
4.1 Introduction 94
4.2 Preliminaries 98
4.3 Escape theorems for random matrices 100
4.4 Applications to random walks in $\mathbb{R}^n$ 105
4.5 Random walks on the sphere 111
4.6 Minimax of the n-dimensional Brownian motion 115
4.6.1 Auxiliary facts 116
4.6.2 Proof of Theorem 4.2 118
4.6.3 Proof of Lemma 4.24 123

5 Conclusion 135

Notation

$\lfloor r \rfloor$	The largest integer not exceeding r
$\lceil r \rceil$	The smallest integer greater or equal to r
Π_n	The set of all permutations on n elements
$\mathbb{N}$	The set of natural numbers (not including zero)
$\mathbb{Z}$	The set of integers
$\mathbb{R}$	The set of real numbers
$\mathbb{R}_+$	The semi-interval $[0, \infty)$
$\mathbb{R}_-$	The semi-interval $(-\infty, 0]$
$\mathbb{R}^n$	n-fold Cartesian product of $\mathbb{R}$ equipped with the standard linear space structure
$e_1^n, e_2^n, \ldots, e_n^n$	The standard basis vectors in $\mathbb{R}^n$
$\langle \cdot, \cdot \rangle_n$	The standard inner product in $\mathbb{R}^n$
$\mathrm{conv} S$	The convex hull of a set S
S^{n-1}	$(n-1)$-dimensional Euclidean unit sphere
σ_{n-1}	The (unique) normalized rotation-invariant Borel measure on S^{n-1}
$\mathrm{G}_{n,m}$	The Grassmannian of m-dimensional linear subspaces of $\mathbb{R}^n$
$\mu_{n,m}$	The normalized rotation-invariant Borel measure on $\mathrm{G}_{n,m}$
$\mathrm{Vol}_n(S)$	Lebesgue n-dimensional volume of a set S
ℓ_p^n	Space $\mathbb{R}^n$ equipped with the norm $\|(x_1, x_2, \ldots, x_n)\|_p = \left(\sum\limits_{i=1}^n \|x_i\|^p\right)^{1/p}$ (for $1 \le p < \infty$) or $\|(x_1, x_2, \ldots, x_n)\|_p = \max\limits_i \|x_i\|$ (for $p = \infty$)
B_p^n	The closed unit ball in ℓ_p^n
$\\| \cdot \\|_{p \to q}$	The operator norm from ℓ_p to ℓ_q
$\mathrm{Im} T$	Image of a linear operator T
$\ker T$	Kernel of a linear operator T
$\mathrm{supp}\, x$	Support of a vector $x = (x_1, x_2, \ldots)$, i.e. the set $\{i : x_i \neq 0\}$
Proj_E	Orthogonal projection onto a linear subspace E
$\mathbb{E}\, \xi$	Expectation of a random variable ξ
$\mathrm{Med}\, \xi$	Median of ξ
i.i.d.	independent identically distributed

Chapter 1

Introduction

In this thesis, we explore phenomena arising within high-dimensional objects of three types: first, convex sets and normed spaces; second, random matrices, and, third, convex hulls of random walks. Accordingly, our work touches three directions within mathematics: Asymptotic Geometric Analysis (AGA), Spectral Theory of Random Matrices and Random Walks.

Asymptotic Geometric Analysis, which started its development in the 1970-es, can be described as the study of high-dimensional convex bodies [80, 86, 114, 5] (see also [78], as well as [115] where a connection to infinite-dimensional Banach space theory is presented). Recall that a *convex body* in $\mathbb{R}^n$ is any compact convex set with non-empty interior. High dimensionality determines central features of the subject. First, it is a variety of "isomorphic" problems, which are uncommon for low-dimensional geometry [5, p. vii], although sharp inequalities (isoperimetric inequality on the sphere, the Brunn–Minkowski, etc. [36]) also play a fundamental role in AGA. Next, it is the concept of randomness. Let us quote [5]: "...in this theory, *randomness* and *pattern* appear together... Objects created by independent identically distributed random processes, while being different from one another, are many times indistinguishable and similar in the statistical sense... The concentration of measure and similar effects caused by the convexity assumption imply in fact a reduction of the diversity with increasing dimension, and the collapse of many different possibilities into one, or, in some cases, a few possibilities only."

In a sense, the situation is similar to the classical law of large numbers or the central limit theorem in Probability Theory. Let us illustrate the above principle by considering the volume distribution in the standard cube. Let $X = (X_1, X_2, \dots, X_n)$ be a random vector uniformly distributed in the cube $[-1,1]^n$ i.e. such that for any Borel subset $A \subset [-1,1]^n$ we have $\mathbb{P}\{X \in A\} = \frac{\mathrm{Vol}_n(A)}{2^n}$. Then the coordinates of X are i.i.d. and each X_i is uniformly distributed in the interval $[-1,1]$. It is easy to see that $\mathbb{E}\,{X_i}^2 = \frac{1}{3}$, so that

$\mathbb{E}\,\|X\|_2^2 = \frac{n}{3}$. It can be easily checked that for small values of n the random quantity $\frac{3}{n}\|X\|_2^2$ is not "concentrated" near 1. However, when n grows to infinity, the Weak Law of Large Numbers tells us that $\frac{3}{n}\|X\|_2^2$ converges to 1 in probability. Geometrically, this means that for large n most of the volume of the cube $[-1,1]^n$ is located in a thin *spherical shell* of radius $\sqrt{n/3}$. In a series of deep results of many researchers (see, in particular, [84, 57, 58, 34, 49]), it was shown that a similar phenomenon holds for any convex body K of high dimension, provided that its center of mass is at the origin, and the covariance matrix of a random vector uniformly distributed in K is the identity (such random vectors are called *isotropic*).

Among major research directions within Asymptotic Geometric Analysis are volume distribution in convex bodies and geometry of their sections and projections. As far as geometry of sections is concerned, a classical result, which to a large extent stimulated the development of the field is a theorem of A. Dvoretzky [26]. For L-Euclidean sections (for a fixed $L > 1$), an optimal form of the theorem is due to V. Milman [76]; it can be stated as follows: *every n-dimensional origin-symmetric convex body contains a section of dimension at least $f(L)\ln n$ which is L-Euclidean.* Here, we say that a convex body K is L-Euclidean if there is an ellipsoid $\mathcal{E}$ such that $\mathcal{E} \subset K \subset L\mathcal{E}$. At the same time, the structure of *almost* Euclidean sections (i.e. L-Euclidean with L arbitrarily close to 1) is far from being understood. In most general form, the question is, given natural m and $\varepsilon > 0$, determine the smallest n such that any n-dimensional convex body contains a $(1+\varepsilon)$-Euclidean m-dimensional section. This problem, as well as the question of estimating the distance of a polytope with very few vertices to the Euclidean ball are discussed in Chapter 2 of the thesis.

Certain geometric objects in AGA can be naturally modelled with help of random matrices (operators). For example, an m-dimensional random subspace of $\mathbb{R}^n$, which is uniformly distributed on the Grassmannian $\mathrm{G}_{n,m}$, can be defined as $G(\mathbb{R}^m)$, with G being the standard $n \times m$ Gaussian matrix. Another example is a class of "extremal" random polytopes which are defined as images of standard cross-polytopes of higher dimension under the action of random linear operators [70]. Let us note here that in Banach space theory, random polytopes of such type emerged in E. Gluskin's groundbreaking paper [39], where the optimal estimate of the diameter of the Minkowski compactum was given. Those questions belong to a large body of problems, including ones coming from statistics, numerical analysis and compressed sensing, which lead to development of a branch of the Random Matrix Theory which is now often called *non-asymptotic* or *non-limiting* [91, 117].

As of now, the central objects of study in the non-asymptotic theory are the *extremal* (largest and smallest) singular values of a random matrix — objects, considered within the classical random matrix theory as well. However, unlike the classical limiting results of the spectral theory, in the non-asymptotic theory one is interested in obtaining quantitative estimates, usually holding with probability very close to one, perhaps at expense of sharpness. A crucial feature of this theory is its methodology which is to a large extent inherited from AGA. For example, covering arguments and arguments involving random projections, which originated within AGA, play a role in the non-asymptotic theory and systematically appear in the literature on the subject [91, 117]. At the same time, it should be emphasized that, on the one hand, extensions of classical limiting theorems which estimate the speed of convergence (see, for example, [6, Chapter 8]), on the other hand, the use of techniques from AGA in some questions regarding the limiting behaviour, make the boundary of non-asymptotic and limiting theories somewhat fuzzy. Our results on the smallest singular value of random matrices, both in limiting and non-asymptotic settings, are presented in Chapter 3 of the thesis.

We conclude the foreword with a brief description of Chapter 4, which deals with geometric properties of convex hulls of random walks in high dimensions. Whereas planar (two-dimensional) convex hulls have been studied a considerable time period, the question on high dimensions attracted attention of researchers relatively recently (let us mention works [29, 30, 54, 53]). The specific problem that we address is the following: *given a random walk in $\mathbb{R}^n$ starting at the origin, how many steps it has to make to absorb the origin into the interior of its convex hull with probability* $1/2$*?* The main reason why we include this work into the thesis is the spectrum of techniques used in the proof: to a large extent, they come from non-asymptotic random matrix theory and, consequently, AGA. In particular, we reduce the original question about random walks to a problem whether the image of certain random operator and certain convex cone have a non-trivial intersection. We believe that the approach employed in our work will find other uses in future.

In the following three subsections of the introduction we will describe the contents of respective chapters in more detail and place the results in a broader context.

1.1 The Banach–Mazur distance to the Euclidean space

Let us recall that *the Banach–Mazur* distance between two n-dimensional convex bodies K_1 and K_2 is defined as the infimum of $\lambda \geq 1$ such that there is an invertible linear operator $T_\lambda : \mathbb{R}^n \to \mathbb{R}^n$ and two vectors x_λ and y_λ satisfying $K_1 \subset T_\lambda(K_2) + x_\lambda \subset \lambda K_1 + y_\lambda$. Clearly, $\operatorname{dist}(K_1, K_2)$ is invariant with respect to affine transformations of K_1 and K_2. It can be shown that when both K_1 and K_2 are origin-symmetric, one can take $x_\lambda = y_\lambda = 0$ in the above definition. The Banach–Mazur distance $\mathrm{dist}(X, Y)$ between two n-dimensional normed spaces X and Y is defined as the distance between their unit balls.

The problem of estimating the Banach–Mazur distance between convex sets has been an important research direction within the local theory of Banach spaces [114]. The classical result in this area is a theorem of F. John [52] which can be stated as follows: *if K is an n-dimensional convex body then* $\mathrm{dist}(K, B_2^n) \le n$. *Moreover, if K is centrally-symmetric then* $\mathrm{dist}(K, B_2^n) \le \sqrt{n}$. At a more technical level, every convex body K is associated with a unique ellipsoid contained in K whose volume is maximal among all ellipsoids in K — *John's ellipsoid*. The theorem of [52] provides necessary and sufficient conditions for B_2^n to be John's ellipsoid for K in terms of the structure of contact points of B_2^n with the boundary of K (see also [9]). In the case when John's ellipsoid coincides with B_2^n, we will say that K is in *John's position*. In particular, the result of [52] states that for K in John's position we have $K \subset nB_2^n$ ($K \subset \sqrt{n}B_2^n$ for origin-symmetric K).

The above estimates for the Banach–Mazur distance to B_2^n cannot in general be improved as is easily seen by considering n-simplex and n-cube. At the same time, for each k *every* convex body of a sufficiently high dimension contains a k-dimensional affine section which is close to an ellipsoid. This fundamental discovery largely influenced the development of Asymptotic Geometric Analysis. Below, when discussing Dvoretzky's theorem and its extensions, we will only consider the case of origin-symmetric convex sets and central sections, although results for non-symmetric sets also exist (see, for example, [61]). We will formulate the statements in terms of norms.

Theorem (A. Dvoretzky, [26])**.** *For any $\varepsilon > 0$ and any natural k there is n depending only on k and ε such that any n-dimensional normed space contains a k-dimensional $(1+\varepsilon)$-Euclidean subspace.*

In the original paper [26], dependence of n on both k and ε was not optimal. Optimal relation of n and k was first obtained by V. Milman, whose proof of the above theorem appeared approximately 10 years later [76]. Milman's approach is one of the cornerstone results of the local theory of Banach spaces. To formulate the result, we need the following definition.

Definition. Let $\|\cdot\|$ be a norm in $\mathbb{R}^n$. Then we define a quantity $k(\|\cdot\|)$ as

$$k(\|\cdot\|) := n\left(\frac{\int_{\mathbb{S}^{n-1}} \|x\|\, d\sigma_{n-1}(x)}{\sup_{x\in\mathbb{S}^{n-1}} \|x\|}\right)^2$$

(see [37, Section 4.3], as well as [86, p. 42] where an equivalent formulation is given in terms of Gaussian variables).

Now, the result of V. Milman [76] can be stated as follows:

Theorem (V. Milman, [76, 80, 98]). *Given $\varepsilon \in (0, 1/2]$, a norm $\|\cdot\|$ in $\mathbb{R}^n$, and a random m-dimensional subspace E uniformly distributed on the Grassmannian $G_{n,m}$, we have that E is $(1+\varepsilon)$-Euclidean with a probability close to one whenever $m \leq \frac{c\varepsilon^2}{\ln\frac{1}{\varepsilon}} k(\|\cdot\|)$. Here, $c > 0$ is a universal constant.*

When the unit ball of the norm $\|\cdot\|$ is in John's position, it follows from the Dvoretzky–Rogers lemma (see, for example, [80, Theorem 3.4] or [86, Lemma 4.13]) that $k(\|\cdot\|) \geq \widetilde{c}\ln n$ for a universal constant $\widetilde{c} > 0$. Taking $\|\cdot\| = \|\cdot\|_\infty$, it is not difficult to see that the lower bound on $k(\|\cdot\|)$ is in general optimal up to the constant multiple $\widetilde{c}$.

The problem of dependence on ε in Dvoretzky's theorem has received considerable attention from researchers. Note that there can be various points of view on the question. First, one may be interested in finding the optimal function $f_n(\varepsilon)$ such that for any norm $\|\cdot\|$ in $\mathbb{R}^n$ there exists a $(1+\varepsilon)$-Euclidean subspace of dimension at least $f_n(\varepsilon)k(\|\cdot\|)$. In this direction, the multiple $\frac{c\varepsilon^2}{\ln\frac{1}{\varepsilon}}$ in the last theorem was improved by Y. Gordon to $c\varepsilon^2$ [43]; shortly an alternative proof of that result was found by G. Schechtman [97]. On the other hand, as is shown by T. Figiel [98], for $n > \varepsilon^{-4}$ we have $f_n(\varepsilon) \leq C\varepsilon^2$ for a universal constant $C > 0$.

Another possible point of view is, fixing natural m, find a function $N_m(\varepsilon)$ such that for any $\varepsilon \in (0, 1/2]$, any normed space of dimension at least $N_m(\varepsilon)$ contains an m-dimensional $(1+\varepsilon)$-Euclidean subspace. As of now, it is an open question whether one can take $N_m(\varepsilon) = h(m)\varepsilon^{-Cm}$ for some universal constant $C > 0$ and some function $h(m)$. There are partial results supporting that. For $m = 2$, it was observed by M. Gromov (see [75]) that one can take $N_2(\varepsilon) = C\varepsilon^{-1/2}$ for a universal constant $C > 0$. For $m \geq 4$, J. Bourgain and J. Lindenstrauss [14] showed that, for some function $h(m)$, any 1-symmetric normed space of dimension at least $h(m)\varepsilon^{-(m-1)/2}|\ln\varepsilon|$, admits a $(1+\varepsilon)$-Euclidean subspace of dimension m. Let us remark that the function $h(m)$ in [14] grows superexponentially with m.

Further, one can be interested in finding a function $\widetilde{h}(\varepsilon)$ such that any n-dimensional normed space contains a $(1+\varepsilon)$-Euclidean subspace of dimension at least $\widetilde{h}(\varepsilon)\ln n$. In this model, it was shown by G. Schechtman that it is possible to take $\widetilde{h}(\varepsilon) = c\varepsilon/\ln^2\frac{1}{\varepsilon}$ [100]. The idea of the proof in [100] is to make use of the following alternative: either the dimension $k(\|\cdot\|)$ of a norm $\|\cdot\|$ (with the unit ball in John's position) is significantly larger than $\ln n$, or there is a large subspace with a small distance to ℓ_∞^m (see also [98]). This assertion, in turn, is based on a result of N. Alon and V. Milman regarding embedding of ℓ_∞^m into finite-dimensional spaces [3].

Apart from the question of existence of large almost Euclidean subspaces, a natural problem is to study how the subspaces are arranged. Of course, the arrangement depends on the position of the unit ball of the norm. For example, it is not difficult to show that if the unit ball B of a norm $\|\cdot\|$ contains an m-dimensional 2-Euclidean section then there is a linear transformation T such that for the random subspace E uniformly distributed on the Grassmannian $\mathrm{G}_{n,m}$, the section $E \cap T(B)$ is C-Euclidean with probability close to one. Accordingly, it is natural to fix a canonical position of the unit ball of a norm, say, John's position. In this setting, G. Schechtman [99] considered random subspaces of ℓ_∞^n. In [99], it was shown that a random m-dimensional subspace of ℓ_∞^n uniformly distributed on $\mathrm{G}_{n,m}$, is $(1+\varepsilon)$-Euclidean with probability close to one as long as $m \leq c\varepsilon \ln n / \ln \frac{1}{\varepsilon}$, and it is *not* $(1+\varepsilon)$-Euclidean with a significant probability provided that $m \geq C\varepsilon \ln n$. Thus, in this setting the dependence of the dimension on ε is much worse than in the question of existence (we recall that ℓ_∞^n contains $(1+\varepsilon)$-Euclidean subspaces of dimension $c \ln n / \ln \frac{1}{\varepsilon}$).

An alternative model of randomness was studied in paper [44], where a problem of embedding arbitrary normed spaces into ℓ_∞^n was considered. Namely, given an m-dimensional normed space X, the authors of [44] introduced a random operator $\Gamma_{n,m} : X \to \ell_\infty^n$ defined by

$$\Gamma_{n,m}(y) = \big(f_1(y), f_2(y), \ldots, f_n(y)\big),$$

where $f_1, f_2, \ldots, f_n$ are independent random points (functionals) uniformly distributed in the unit ball of the dual space X^*. They showed that if m, n and $\varepsilon \in (0, 1/2]$ satisfy the relation $n \geq (8/\varepsilon)^{2m}$ then with high probability the functionals $f_1, f_2, \ldots, f_n$ form an ε-net in B_{X^*} (i.e. the union of $f_i + \varepsilon B_{X^*}$ covers B_{X^*}). In turn, this implies that the Banach–Mazur distance from X to the image $\mathrm{Im}\Gamma_{n,m}$ in ℓ_∞^n is at most $1 + 2\varepsilon$ with high probability.

In connection with the aforementioned problems and results, one can ask the following questions:

Question 1. Whether the results of [44] can be generalized to provide random embeddings into spaces other than ℓ_∞^n, with optimal dependence on ε?

Question 2. In the setting when a random subspace of ℓ_∞^n is uniformly distributed on the corresponding Grassmannian, what is the right dependence of the dimension on ε? (note that the result of G. Schechtman [99] leaves a gap between $c\varepsilon \ln n / \ln \frac{1}{\varepsilon}$ and $C\varepsilon \ln n$)

These questions are addressed in Sections 2.2 and 2.3 of the thesis. The main result of Section 2.2 provides embeddings of ℓ_2^m into arbitrary 1-symmetric spaces. Namely, let the random operator $\Gamma_{n,m} : \mathbb{R}^m \to \mathbb{R}^n$ be defined as

$$\Gamma_{n,m}(y) = \big(\langle y, z_1\rangle_m, \langle y, z_2\rangle_m, \dots, \langle y, z_n\rangle_m\big),$$

where $z_1, z_2, \dots, z_n$ are independent random vectors uniformly distributed on S^{m-1}.

Theorem (Theorem 2.6)**.** *Let* $0 < \varepsilon \le 1/2$ *and* $3 \le m \le c \ln n / \ln(1/\varepsilon)$*, where* $c > 0$ *is a small universal constant. Then with probability close to one the random subspace* $\mathrm{Im}\Gamma_{n,m}$ *has the property that for any* 1*-symmetric norm* $\|\cdot\|$ *in* $\mathbb{R}^n$ *it is* $(1+\varepsilon)$*-Euclidean with respect to* $\|\cdot\|$*.*

The proof of the above theorem is based on the use of standard Chernoff–type inequalities, together with a classical result of the interpolation theory — the theorem of Calderon–Mityagin, which allows us to verify certain geometric properties shared by all symmetric spaces. Let us mention that a relative of Theorem 2.6 which provides *explicit* (non-random) almost isometric embeddings of ℓ_2^m into 1-symmetric normed spaces, was recently obtained by D. Fresen in [35].

We will call a subspace $E \subset \ell_\infty^n$ $(1+\varepsilon)$-*spherical* if

$$\sup_{\substack{x \in E \\ \|x\|_2 = 1}} \|x\|_\infty \Big/ \inf_{\substack{x \in E \\ \|x\|_2 = 1}} \|x\|_\infty \le 1 + \varepsilon.$$

In Section 2.3, we prove the following statement:

Theorem (Theorem 2.13)**.** *Let* $\varepsilon \in (0, 1/2)$ *and* $n > 1$*. Then*

1) *There is a universal constant* $\widetilde{c} > 0$ *such that whenever* $k \le \widetilde{c}\varepsilon \ln n / \ln \frac{1}{\varepsilon}$*, then*

$$\mu_{n,m}\big\{E \in \mathrm{G}_{n,m} : \ E \text{ is } (1+\varepsilon)\text{-spherical subspace of } \ell_\infty^n\big\} \ge 1 - 2n^{-\widetilde{c}\varepsilon};$$

2) *Conversely, if for some* $m > 1$

$$\mu_{n,m}\big\{E \in \mathrm{G}_{n,m} : \ E \text{ is } (1+\varepsilon)\text{-spherical subspace of } \ell_\infty^n\big\} \ge \frac{3}{4}$$

then necessarily $m \le C\varepsilon \ln n / \ln \frac{1}{\varepsilon}$*, where* $C > 0$ *is a universal constant.*

The first part of the above theorem is proved in [99] and is given here only for completeness. Our contribution consists in proving the second assertion, which follows from certain properties of the χ-distribution.

Finally, let us discuss the contents of Section 2.4 of the thesis. Fix any $N \geq n+1$ and let P_N be an n-dimensional convex polytope with N vertices. For $N \geq 2n$, the Banach–Mazur distance of P_N to the Euclidean ball can be estimated from below by $c\sqrt{\frac{n}{\ln\frac{N}{n}}}$ for a universal constant $c > 0$ (we refer to [11], [15, Corollary 9.5], [18] or [38]). This bound is optimal up to the constant multiple: for any $n \in \mathbb{N}$ and $n+1 \leq N \leq 2^n$ there is a polytope P_N in $\mathbb{R}^n$ with N vertices which is $C\sqrt{n/\ln\frac{N}{n}}$–Euclidean for a universal constant $C > 0$ (see [33, p. 96]).

On the other hand, for $N < 2n$ only partial results in this direction were available. Note that in this case the polytope is necessarily non-symmetric. E. Gluskin and A. Litvak showed in [40] that the distance of P_N to the set of all symmetric convex bodies is at least $\frac{cn}{N-n}$. The authors of [40] conjectured that $\mathrm{dist}(P_N, B_2^n) \geq \frac{cn}{\sqrt{N-n}}$. The next result confirms this:

Theorem (Theorem 2.14)**.** *Let $n \in \mathbb{N}$ and let P_N be a convex n-dimensional polytope with N vertices ($n+1 \leq N \leq 2n$). Then $\mathrm{dist}(P_N, B_2^n) \geq cn/\sqrt{N-n}$ where $c > 0$ is a universal constant.*

It can be shown that the above estimate is optimal up to the constant factor c. Let us remark that the question of the distance to the Euclidean ball is naturally connected to coverings of the sphere S^{n-1} by spherical caps of equal radius, which has been considered by several authors in special cases. This will be discussed in more detail in Section 2.4.

1.2 The extreme singular values of random matrices

Before we start our discussion, let us emphasize that, with multiple connections to Mathematical Physics, Computer Science and Statistics, the Random Matrix Theory has developed into a very broad area of research. For an extensive information on the subject, we refer to monographs and surveys [6, 85, 117]. Here, we focus our attention only on those aspects of the theory which are directly related to the results proved in our thesis.

Let M be an $N \times n$ matrix ($N \geq n$) with real-valued entries. Then the *singular* values $s_i(M)$ of M are defined as square roots of the eigenvalues of the $n \times n$ matrix $M^T M$. We will arrange them in non-increasing order, counting multiplicities, i.e.

$$s_i(M) = \sqrt{\lambda_i}, \ \text{ where } \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0 \ \text{ are the eigenvalues of } M^T M.$$

In particular, the largest and the smallest singular values of M can be defined as

$$s_{\max}(M) = s_1(M) = \sup_{y \in \mathrm{S}^{n-1}} \|My\|_2$$

and

$$s_{\min}(M) = s_n(M) = \inf_{y \in S^{n-1}} \|My\|_2.$$

The study of the extremal singular values of random matrices is connected to certain important questions in statistics and analysis of algorithms (see, in particular, [1, 28, 91, 96, 105]).

Given a square non-singular matrix M, the *condition number* $\kappa(M)$ is defined as the ratio $s_{\max}(M)/s_{\min}(M)$. This quantity plays a fundamental role in numerical analysis and, in particular, is used to analyse parameters (precision versus speed) of the Gaussian elimination (see [28, 96] and references therein). When M is random, the distribution of $\kappa(M)$ characterizes the average-case performance and (depending on the model of randomness) is employed in smooth analysis of algorithms [96, 105]. For the $n \times n$ standard Gaussian matrix G, an explicit formula for the joint distribution of the eigenvalues of $G^T G$ is known [51]. This allowed A. Edelman [27] to compute the probability density function of the smallest singular value $s_{\min}(G)$ and establish convergence of the probability density of rescaled condition number $\kappa(G)/n$ (when $n \to \infty$) to the function $f(t) = \frac{2t+4}{t^3} e^{-2/t-2/t^2}$ [27, Theorem 6.1] (see also [107] for large deviation estimates of the singular values). We note here that analogous results for a more general class of random square matrices were obtained by T. Tao and V. Vu in [111]. From the results of A. Edelman it follows, in particular, that the "typical" value of $s_{\min}(G)$ is of order $n^{-1/2}$ and that $\mathbb{P}\{s_{\min}(G) \leq \varepsilon n^{-1/2}\} \leq \varepsilon$ for any $\varepsilon > 0$. In context of smooth analysis of algorithms, A. Sankar, D. Spielman and S.-H. Teng [96] showed that an analogue of the latter estimate holds for a *shifted* Gaussian matrix; namely, if A is $n \times n$ with independent Gaussian entries of unit variance (not necessarily centered) then $\mathbb{P}\{s_{\min}(A) \leq \varepsilon n^{-1/2}\} \leq 2.35\varepsilon$ for all $\varepsilon > 0$. For various generalization of the aforementioned results, we refer to [95, 112, 113, 87] and references therein. Let us also mention an important problem of singularity of a random Bernoulli (± 1) matrix which influenced development of the subject; see [60, 55, 109, 16] for results in that direction.

For rectangular matrices, the study of the extremal singular values has been stimulated by the problem of approximating the covariance matrix of a distribution by a sample covariance matrix, as well as by certain questions in high-dimensional convex geometry. Recall that given an n-dimensional centered random vector X with $\mathbb{E}\,\|X\|_2^2 < \infty$, its *covariance matrix* is defined as the expectation of the outer product $\Sigma = \mathbb{E}\, XX^T$. Further, we say that a random centered vector X (and the underlying distribution) is *isotropic* if its covariance matrix is the identity.

In Section 1.1 we have mentioned a theorem of Y. Gordon [43] which improved the original dependence on ε in Dvoretzky's theorem. As a crucial step, Y. Gordon showed that for the standard $N \times n$ ($N \geq n$) Gaussian matrix G,

$$E_N - E_n \leq \mathbb{E}\, s_{\min}(G) \leq \mathbb{E}\, s_{\max}(G) \leq E_N + E_n,$$

where E_m stands for the expectation of the Euclidean norm of the standard m-dimensional Gaussian vector. We note that together with a well known concentration inequality in the Gauss space (see, for example, [117, Corollary 5.35]), the above relation implies

$$\mathbb{P}\big\{\sqrt{N} - \sqrt{n} - t \leq s_{\min}(G) \leq s_{\max}(G) \leq \sqrt{N} + \sqrt{n} + t\big\} \geq 1 - 2\exp(-ct^2), \quad t > 0, \tag{1.1}$$

for a universal constant $c > 0$. Similar (but somewhat weaker) estimates are known for more general *subgaussian* distributions. Recall that the subgaussian norm of a centered random variable ξ is defined as $\|\xi\|_{\psi_2} = \sup\limits_{p\geq 1} p^{-1/2}\big(\mathbb{E}\,|\xi|^p\big)^{1/p}$. Further, the subgaussian norm of a centered n-dimensional random vector X is defined as

$$\|X\|_{\psi_2} = \sup_{y\in S^{n-1}} \|\langle X, y\rangle_n\|_{\psi_2}.$$

Now, if A is an $N \times n$ random matrix with independent centered isotropic rows $(X^i)^T$ ($1 \leq i \leq N$) with $\|X^i\|_{\psi_2} \leq K$ then

$$\sqrt{N} - L\sqrt{n} - t \leq s_{\min}(A) \leq s_{\max}(A) \leq \sqrt{N} + L\sqrt{n} + t$$

with probability at least $1 - 2\exp(-ct^2)$ for any $t > 0$ and for L depending only on K [117, Theorem 5.39]. Note that the left hand side estimate is non-trivial only for tall matrices with N significantly larger than n. For $N \approx n$, strong estimates for $s_{\min}(A)$ are known in case when the entries of A are i.i.d. subgaussian [66, 93].

The above estimates can be interpreted in the context of sample covariance matrices. Let X be an n-dimensional centered random vector with covariance matrix Σ and let $X^1, X^2, \dots, X^N$ be a collection of independent vectors equidistributed with X. We define the *sample covariance matrix*

$$\widetilde{\Sigma} = \frac{1}{N}\sum_{i=1}^{N} X^i (X^i)^T.$$

If we denote by A an $N \times n$ matrix with rows $(X^i)^T$ ($1 \leq i \leq N$) then it is easy to see that $s_{\max}(A)^2 = N\lambda_{\max}(\widetilde{\Sigma})$ and $s_{\min}(A)^2 = N\lambda_{\min}(\widetilde{\Sigma})$, where $\lambda_{\max}(\widetilde{\Sigma}), \lambda_{\min}(\widetilde{\Sigma})$ are the largest and the smallest eigenvalues of $\widetilde{\Sigma}$. The principal question is how close the sample covariance matrix is to Σ; more specifically, **given $\varepsilon > 0$, how big the number N**

should be to guarantee that the operator norm of the difference $\widetilde{\Sigma} - \Sigma$ is less than $\varepsilon\|\Sigma\|_{2\to 2}$ with probability close to one. Note that the Law of Large Numbers, together with Lipschitzness of the operator norm, imply that such a number N always exists. As an example, let us take X to be the standard n-dimensional Gaussian vector. We have

$$\begin{aligned}\|\widetilde{\Sigma} - \mathbf{Id}_n\|_{2\to 2} &= \max\big(|\lambda_{\max}(\widetilde{\Sigma}) - 1|, |\lambda_{\min}(\widetilde{\Sigma}) - 1|\big) \\ &= \max\Big(\frac{1}{N}s_{\max}(G)^2 - 1, 1 - \frac{1}{N}s_{\min}(G)^2\Big),\end{aligned}$$

where G is the $N \times n$ Gaussian matrix with rows $(X^i)^T$, $i = 1, 2, \ldots, N$. Together with (1.1), the above identity implies

$$\|\widetilde{\Sigma} - \mathbf{Id}_n\|_{2\to 2} \le 3\sqrt{\frac{n}{N}} + \frac{5t}{\sqrt{N}}$$

with probability at least $1-2\exp(-ct^2)$ for any $t \in (0, \sqrt{n})$. Thus, in order to approximate the covariance matrix with precision $\varepsilon \in (0, 1]$ in the Gaussian case, $C\varepsilon^{-2}n$ samples (for a sufficiently large universal constant C) is enough.

It is not difficult to see that given any centered random vector X with covariance matrix Σ, the vector $\Sigma^{-1/2}X$ is isotropic. Now, let $\Sigma^{-1/2}X^1, \Sigma^{-1/2}X^2, \ldots, \Sigma^{-1/2}X^N$ be a sample of size N with respect to $\Sigma^{-1/2}X$. Note that

$$\Big\|\frac{1}{N}\sum_{i=1}^{N}\Sigma^{-1/2}X^i\big(\Sigma^{-1/2}X^i\big)^T - \mathbf{Id}_n\Big\|_{2\to 2} < \varepsilon$$

for some $\varepsilon \in (0, 1]$ only if

$$-\varepsilon\Sigma \prec \frac{1}{N}\sum_{i=1}^{N}X^i(X^i)^T - \Sigma \prec \varepsilon\Sigma,$$

where $M \prec M'$ means that $M' - M$ is positive definite. The last relation, in turn, implies that

$$\Big\|\frac{1}{N}\sum_{i=1}^{N}X^i(X^i)^T - \Sigma\Big\|_{2\to 2} < \varepsilon\|\Sigma\|_{2\to 2}.$$

This observation tells us that whenever we want to verify approximation properties of sample covariance matrices for a class of centered distributions which is closed under linear bijections, it is enough to consider only isotropic distributions from the class. An important example of this kind is the class of *log-concave* distributions.

Recall that a distribution in $\mathbb{R}^n$ is log-concave if its probability density function can be represented in form $\exp(-h(x))$ ($x \in \mathbb{R}^n$), where $h(x)$ is some convex function. Motivated by the problem of computing the volume of a high-dimensional convex set given by a separation oracle, R. Kannan, L. Lovász and M. Simonovits [56] considered the question of approximating the covariance matrix of a log-concave random vector by the sample covariance matrix. The problem was resolved by R. Adamczak, A. Litvak, A. Pajor and N. Tomczak-Jaegermann [1]. Taking into account observations made by the authors in [2], their result can be formulated as follows:

Theorem ([1, 2])**.** *Let X be an isotropic log-concave vector in $\mathbb{R}^n$ and $X^1, X^2, \dots, X^N$ be its independent copies. Then with probability $1 - 2\exp(-c\sqrt{n})$ we have*

$$\Big\| \frac{1}{N} \sum_{i=1}^{N} X^i (X^i)^T - \mathbf{Id}_n \Big\|_{2\to 2} \le C \sqrt{\frac{n}{N}},$$

where $C, c > 0$ are universal constants.

Let us note that the above theorem can be reformulated in terms of singular values as follows: If A is an $N \times n$ random matrix whose rows are independent copies of an isotropic log-concave vector then

$$\sqrt{N} - \widetilde{C}\sqrt{n} \le s_{\min}(A) \le s_{\max}(A) \le \sqrt{N} + \widetilde{C}\sqrt{n}$$

with probability at least $1 - 2\exp(-c\sqrt{n})$. Various generalizations of the above theorem requiring fewer assumptions on the distribution of X have appeared in past years. A principal observation made already in [2] was that the above approximation properties of the sample covariance matrix hold as long as the Euclidean norm of X concentrates sufficiently well around its expectation and one-dimensional marginals $\langle X, y\rangle_n$ ($y \in \mathrm{S}^{n-1}$) have bounded moments of sufficiently high order. For developments in this direction, we refer to [106, 74, 47]. The problem of bounding the smallest singular value of a random matrix with i.i.d. isotropic rows was treated in [59, 118].

Let us turn attention to corresponding *limiting* results. First, we recall the classical *Marčenko–Pastur law.* Given an $n \times n$ matrix M with real eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$, we define the *spectral distribution* F^M as

$$F^M(t) = \frac{1}{n} \big| \{ i \le n : \lambda_i \le t \} \big|, \quad t \in \mathbb{R}.$$

Theorem (the Marčenko–Pastur law; see [71], [119], [6, Theorem 3.6])**.** *Let (a_{ij}) ($1 \le i, j < \infty$) be a set of i.i.d. random variables with zero mean and unit variance and let*

$(m_n)_{n=1}^{\infty}$ be an integer sequence with $\lim\limits_{n\to\infty} \frac{n}{m_n} = z$ for some $z \in (0,1)$. For every $n \in \mathbb{N}$ denote by A_n the random $m_n \times n$ matrix in the top left corner of the array (a_{ij}). Then with probability one the sequence of spectral distributions $\{F^{{A_n}^T A_n}\}$ converges pointwise to a non-random distribution given by

$$F(t) = \begin{cases} 0, & \text{if } t \le r, \\ \frac{1}{2\pi z} \int\limits_r^t \frac{\sqrt{(R-\tau)(\tau-r)}}{\tau}\, d\tau, & \text{if } r \le t \le R, \\ 1, & \text{if } t \ge R. \end{cases}$$

where $r = (1-\sqrt{z})^2$ and $R = (1+\sqrt{z})^2$.

In particular, the Marčenko–Pastur law implies that, with the above notation, for a sufficiently large n the matrix A_n satisfies $s_{\max}(A_n) \ge \sqrt{m_n} + (1-o(1))\sqrt{n}$ and $s_{\min}(A_n) \le \sqrt{m_n} - (1-o(1))\sqrt{n}$ with probability close to one, where $o(1)$ stands for a quantity that goes to zero when $n \to \infty$. A natural question is whether the estimates are sharp, equivalently, whether the extremal eigenvalues of matrix $\frac{1}{m_n} {A_n}^T A_n$ concentrate around the edges of the Marčenko–Pastur distribution. For $s_{\max}(A_n) = \sqrt{\lambda_{\max}({A_n}^T A_n)}$, this question was resolved by Y.Q. Yin, Z.D. Bai and P.R. Krishnaiah in [120]. A corresponding result for $s_{\min}$ was obtained by Z.D. Bai and Y.Q. Yin [8], who also provided a uniform treatment of both singular values of A_n. Those are often called the Bai–Yin theorem in the literature.

Theorem (Bai–Yin, [8])**.** *Let $(a_{ij})_{i,j=1}^{\infty}$ be a two-dimensional infinite array of i.i.d. random variables with zero mean, unit variance and a bounded 4th moment. Further, let $(m_n)_{n=1}^{\infty}$ be an integer sequence such that $\lim\limits_{n\to\infty} \frac{n}{m_n} = z \in (0,1)$. For every $n \in \mathbb{N}$, let A_n be the $m_n \times n$ matrix in the top left corner of the array. Then the sequence $\frac{1}{\sqrt{m_n}} s_{\max}(A_n)$ converges to $1+\sqrt{z}$ almost surely. Similarly, the sequence $\frac{1}{\sqrt{m_n}} s_{\min}(A_n)$ converges to $1-\sqrt{z}$ almost surely.*

Although from the standpoint of approximation of the covariance matrix there is no need to treat the largest and the smallest eigenvalues (singular values) separately, there is definitely a theoretical motivation for this. Indeed, on an intuitive level, the largest singular value is sensitive to spikes — entries of a matrix having a very large value, whereas the smallest singular value should be "stable" as long as the number of spikes is not too big. Further, whereas controlling the largest singular value from above requires certain assumptions on tails of row distributions, the lower bound for $s_{\min}$ (again, on an intuitive level) is likely to depend only on *anticoncentration* properties of respective distributions. Confirming or disproving these intuitive notions can improve one's understanding of spectral properties of random matrices and bring about new techniques for dealing with heavy-tailed entries. The two results we present in Chapter 3 are devoted to this problem.

In Section 3.1, we prove a lower bound on the smallest singular value of a random matrix with independent entries which requires only assumptions on the Lévy concentration function of the entries and no assumptions on moments. Given a real random variable ξ, *the Lévy concentration function* of ξ is defined as

$$\mathcal{Q}(\xi, \alpha) = \sup_{\lambda \in \mathbb{R}} \mathbb{P}\big\{|\xi - \lambda| \leq \alpha\big\}, \quad \alpha \geq 0.$$

We prove the following theorem:

Theorem (Theorem 3.1)**.** *For any real $\beta > 0$ and $\delta > 1$ there are $u, v > 0$ and $N_0 \in \mathbb{N}$ depending only on β and δ with the following property: Let $N, n \in \mathbb{N}$ satisfy $N \geq \max(N_0, \delta n)$; $A = (a_{ij})$ be an $N \times n$ random matrix with i.i.d. entries, such that for some $\alpha > 0$ the concentration function of the entries satisfies*

$$\mathcal{Q}(a_{11}, \alpha) \leq 1 - \beta.$$

Then for any non-random $N \times n$ matrix B we have

$$\mathbb{P}\big\{s_{\min}(A + B) \leq \alpha u \sqrt{N}\big\} \leq \exp(-vN).$$

The next result is a generalization of the aforementioned theorem of Bai and Yin. Let us note that for convergence of the (appropriately rescaled) largest singular value to the right edge of the Marčenko–Pastur distribution the assumption of bounded fourth moment is necessary [103]. However, as we prove in §3.2, for the smallest singular value the situation is different:

Theorem (Theorem 3.18)**.** *Let (a_{ij}) $(1 \leq i, j < \infty)$ be a set of i.i.d. real valued random variables with zero mean and unit variance. Further, let $(m_n)_{n=1}^{\infty}$ be an integer sequence with $\lim\limits_{n\to\infty} \frac{n}{m_n} = z$ for some $z \in (0, 1)$. For every $n \in \mathbb{N}$ we denote by A_n the random $m_n \times n$ matrix with entries a_{ij} $(1 \leq i \leq m_n, 1 \leq j \leq n)$. Then with probability one the sequence*

$$\big({m_n}^{-1/2} s_{\min}(A_n)\big)_{n=1}^{\infty}$$

converges to $1 - \sqrt{z}$.

Further generalization of the Bai–Yin theorem has been recently obtained in [20]. In particular, the result of [20] implies convergence in probability of (appropriately rescaled) singular values of random matrices with i.i.d. isotropic log-concave rows to edges of the Marčenko–Pastur distribution, thereby sharpening the theorem of [1].

Let us note that the proofs of both results in Chapter 3 heavily rely on covering arguments as well as on studying random projection operators. Thus, our methodology in treating these problems is largely inherited from Asymptotic Geometric Analysis.

1.3 Convex hulls of random walks in high dimensions

The standard *Brownian motion* on $[0,\infty)$ is a centered Gaussian process $\mathrm{BM}_1(t)$ such that for any $t,s\in[0,\infty)$ we have $\mathrm{cov}(\mathrm{BM}_1(t),\mathrm{BM}_1(s))=\min(t,s)$. By n-dimensional ($n\geq 1$) Brownian motion $\mathrm{BM}_n(t)$ we understand a vector of n independent standard Brownian motions.

Convex hulls of the *planar* Brownian motion as well as of certain discrete-time 2-dimensional random walks, have been extensively studied in literature; in particular, sharp estimates for the perimeter and the area are known. We refer to [69] for a survey of related results prior to 2010.

In higher dimensions, much less information has been available until recently. In particular, the expected volume of $\mathrm{conv}\{\mathrm{BM}_n(t)\,:\,0\leq t\leq 1\}$ was not known until R. Eldan [30] and Z. Kabluchko and D. Zaporozhets [54] independently showed that

$$\mathbb{E}\,\mathrm{Vol}_n\big(\mathrm{conv}\{\mathrm{BM}_n(t)\,:\,0\leq t\leq 1\}\big)=\Big(\frac{\pi}{2}\Big)^{n/2}\frac{1}{\Gamma(n/2+1)^2}.$$

Moreover, these papers provide explicit formulas for the expectations of all *intristic volumes* of the convex hull.

In [29], R. Eldan studied the following question posed by I. Benjamini: *Let* $t_1,t_2,\ldots,t_N$ *be points generated by the Poisson point process in* $[0,1]$ *of intensity* α. *What is the probability that the convex hull of* $\mathrm{BM}_n(t_i)$ *(*$1\leq i\leq N$*) (where* $\mathrm{BM}_n(t)$ *is independent from the Poisson process) contains the origin?* R. Eldan showed that when the intensity $\alpha\geq\exp(Cn\ln n)$, the origin is contained in the convex hull with probability close to one, whereas, for $\alpha\leq\exp(cn/\ln n)$, the probability is near zero. Analogous questions for the simple random walk on $\mathbb{Z}^n$ and for the spherical Brownian motion were also treated in [29]. In particular, it was shown that for the walk on $\mathbb{Z}^n$ starting at the origin, after $\exp(Cn\ln n)$ steps the probability that the convex hull contains the origin in the interior is almost one and, for less than $\exp(cn/\ln n)$ steps, the probability is close to zero.

The purpose of Chapter 4 of this thesis is to define a novel approach to the above problem based on techniques from non-asymptotic random matrix theory and AGA. This allowed us to strengthen and generalize the aforementioned results of R. Eldan. The first main result of the Chapter is the following theorem:

Theorem (Theorem 4.1). *There exists a constant $C > 0$ such that for any $n \in \mathbb{N}$ and $N \geq \exp(Cn)$ the following holds.*

- *Setting $t_i := i/N$, $i = 1, 2, \ldots, N$, the set $\mathrm{conv}\{BM_n(t_i), i \leq N\}$ contains the origin in its interior with probability at least $1 - \exp(-n)$.*

- *The convex hull of the first N steps of the standard random walk on $\mathbb{Z}^n$ starting at 0, contains the origin in its interior with probability at least $1 - \exp(-n)$.*

The first part of this theorem also holds when $\{t_i\}$ is a homogeneous Poisson process in $[0, 1]$ of intensity at least $\exp(Cn)$. Therefore, our result is strictly stronger than the bound proved in [29]. The first assertion of the above theorem is close to optimal in a sense that for some universal constants $c > 0$ and $n_0 \in \mathbb{N}$ and for $n \geq n_0$ we have

$$\mathbb{P}\big\{0 \in \mathrm{conv}\{\mathrm{BM}_n(t) :\, t \in [1, 2^{cn}]\}\big\} \leq \frac{1}{n}.$$

We provide a proof of this complementary result in Section 4.6 (let us note that the techniques used in its proof are completely different from the "random matrix" approach used to verify the first theorem of the Chapter).

The second main result of Chapter 4 deals with discrete-time random walks on the sphere. For any $\theta \in (0, \pi/2)$, we consider a walk W_θ with values in S^{n-1} such that the angle between two consecutive steps is θ (i.e. $\langle W_\theta(j), W_\theta(j+1)\rangle_n = \cos\theta$, $j \in \mathbb{N}$) and the direction from $W(j)$ to $W(j+1)$ is chosen uniformly at random in the sense that for any $u \in \mathrm{S}^{n-1}$, the distribution of $W_\theta(j+1)$ conditioned on $W_\theta(j) = u$ is uniform on the $(n-2)$-sphere $\mathrm{S}^{n-1} \cap \{x \in \mathbb{R}^n :\, \langle x, u\rangle_n = \cos\theta\}$.

Theorem (Theorem 4.3). *For any $\theta \in (0, \pi/2)$, there exist $L = L(\theta)$ and $n_0 = n_0(\theta)$ depending only on θ such that the following holds: Let $n \geq n_0$ and W_θ be the process with values in S^{n-1} described above. Then for all $N \geq Ln$ we have*

$$\mathbb{P}\big\{0 \textit{ belongs to } \mathrm{conv}\{W_\theta(i) :\, i \leq N\}\big\} \geq 1 - \exp(-n).$$

Let us note that recently, a different treatement of the "absorbtion problem" was proposed by Z. Kabluchko, V. Vysotsky and D. Zaporozhets in [53]. In particular, the authors were able to obtain formulas giving the probability of the convex hull containing the origin for a very large class of high-dimensional random walks, including the discretized

Brownian motion. Nevertheless, we would like to point out that the approach developed in Chapter 4, although providing less precise estimates compared to [53], is a very natural (and technically simple) application of standard AGA methods, and in this respect may be of value for future works on high-dimensional random processes.

Chapter 2

Asymptotic Geometric Analysis

2.1 Notation

Here, we have grouped together notation specific to this Chapter. A norm $\|\cdot\|$ in $\mathbb{R}^n$ is *unconditional* (or 1-*unconditional*) if

$$\|(x_1, x_2, \ldots, x_n)\| = \|(|x_1|, |x_2|, \ldots, |x_n|)\|$$

for any $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$. Further, $\|\cdot\|$ is said to be *symmetric* (or 1-*symmetric*) if it is unconditional and invariant under permutations of coordinates, i.e.

$$\|(x_1, x_2, \ldots, x_n)\| = \|(x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)})\|$$

for all $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ and any permutation $\sigma \in \Pi_n$. For two normed n-dimensional spaces X and Y, the *Banach–Mazur distance* between X and Y is defined as $\mathrm{dist}(X, Y) = \inf \|T\|_{X\to Y}\|T^{-1}\|_{Y\to X}$ where the infimum is taken over all linear bijections $T : X \to Y$. By analogy, the Banach–Mazur distance between two n-dimensional convex bodies K_1 and K_2 (not necessarily symmetric) is defined by

$$\begin{aligned}\mathrm{dist}(K_1, K_2) = \inf\big\{\lambda \geq 1 :\ &\text{there is } T : \mathbb{R}^n \to \mathbb{R}^n \text{ and } x, y \in \mathbb{R}^n \\ &\text{such that } K_1 \subset T(K_2) + x \subset \lambda K_1 + y\big\}.\end{aligned}$$

It is not difficult to see that the last definition agrees with one for normed spaces in the sense that the Banach–Mazur distance between the spaces is equal to the distance between their unit balls. If a convex body K is at the distance at most L from the Euclidean ball then we say that K is *L-Euclidean*. Similarly, if a normed space X is at the distance at most L from ℓ_2^n, it will be called L-Euclidean.

2.2 Almost Euclidean Sections in Symmetric Spaces and Concentration of Order Statistics [1]

2.2.1 Introduction

The classical theorem of A. Dvoretzky [26] states that for arbitrary $\varepsilon > 0$ and $m \in \mathbb{N}$ there is a number $n = n(m, \varepsilon)$ such that any n-dimensional normed space contains an m-dimensional $(1+\varepsilon)$-Euclidean subspace (i.e. its Banach–Mazur distance to ℓ_2^m is at most $1+\varepsilon$). In subsequent years several alternative proofs of the theorem were found, the most influential is due to V. Milman [76] (see [98] for more background information).

V. Milman showed that it is possible to take $n(m,\varepsilon) = \exp(m/f(\varepsilon))$ for certain function $f(\varepsilon)$. Whereas the dependence on m in the last relation is optimal, the behaviour of $f(\varepsilon)$ is unclear. In the original work of V. Milman, $f(\varepsilon) = c\varepsilon^2/\ln(1/\varepsilon)$ where c is a universal constant. This was improved by Y. Gordon [43, Theorem 2.8] to $f(\varepsilon) = c\varepsilon^2$ and later by G. Schechtman [100] to $f(\varepsilon) = c\varepsilon/\ln^2(1/\varepsilon)$.

Instead of considering the general case, one may wish to estimate $n(m,\varepsilon)$ for some particular family of normed spaces. It is well known that $(1+\varepsilon)$-Euclidean subspaces of ℓ_∞^n cannot have dimension larger than $C\ln n/\ln(1/\varepsilon)$. Moreover, this bound is optimal in a sense that these subspaces can be chosen to have dimension at least $c\ln n/\ln(1/\varepsilon)$ (for some universal c). The standard construction is to take a linear operator $T : \mathbb{R}^m \to \mathbb{R}^n$ mapping every $y \in \mathbb{R}^m$ to $(\langle y, z_1\rangle_m, \langle y, z_2\rangle_m, \ldots, \langle y, z_n\rangle_m)$, where $\{z_1, z_2, \ldots, z_n\}$ is a fixed $\sqrt{\varepsilon}$-net on the unit sphere S^{m-1}. Then $\mathrm{Im}T \subset \ell_\infty^n$ is $(1+C\varepsilon)$-Euclidean. This approach was transferred to all spaces with symmetric bases by J. Bourgain and J. Lindenstrauss [14]. The paper [14] deals with the problem of finding the optimal dependence of $n(m,\varepsilon)$ on ε when m is fixed. It was proved that in the class of symmetric spaces the smallest possible $n(m,\varepsilon)$ does not exceed $h(m)\varepsilon^{-(m-1)/2}\ln(1/\varepsilon)$, where $h(m)$ is a function of m only. Disregarding the logarithmic factor $\ln(1/\varepsilon)$, the bound obtained in [14, Theorem 2] is optimal with respect to ε. However, in the proof of Theorem 2, $h(m) \geq m^{cm}$, giving an unsatisfactory estimate for $n(m,\varepsilon)$ when $\ln(1/\varepsilon)/\ln m \ll 1$.

A natural question in connection with embedding ℓ_2^m into ℓ_∞^n is whether the embedding can be randomized. G. Schechtman [99] showed that if for "most" m-dimensional subspaces on the Grassmannian (with respect to the Haar measure) the ℓ_∞^n-norm restricted to the subspace is $(1+\varepsilon)$-equivalent to a multiple of ℓ_2^n-norm, then necessarily $m \leq C\varepsilon\ln n$. As was observed in [44], the situation changes completely when subspaces are generated by the random operator Γ mapping every $y \in \mathbb{R}^m$ to $(\langle y, z_1\rangle_m, \langle y, z_2\rangle_m, \ldots, \langle y, z_n\rangle_m)$,

[1] A version of this section has been published. K. E. Tikhomirov, Almost Euclidean sections in symmetric spaces and concentration of order statistics, J. Funct. Anal. **265** (2013), no. 9, 2074–2088.

with $z_1, z_2, \ldots, z_n$ being independent random vectors uniformly distributed on S^{m-1}. Specifically, Y. Gordon, A. Litvak, A. Pajor and N. Tomczak–Jaegermann proved in [44] that $\mathrm{Im}\Gamma$ is a $(1+\varepsilon)$-Euclidean subspace of ℓ_∞^n with a large probability, provided that $m \le c \ln n / \ln(1/\varepsilon)$ [44, Theorem 4.1]. In fact, the main result of [44] — Theorem 3.3 — produces similar random embeddings into ℓ_∞^n for arbitrary normed spaces, not only ℓ_2^m.

The main motivation for us was to generalize Theorem 4.1 from [44] to all symmetric spaces. We prove that with a large probability $\mathrm{Im}\Gamma$ (for Γ defined above) is actually a $(1+\varepsilon)$-Euclidean subspace of *any* symmetric space $(\mathbb{R}^n, \|\cdot\|)$ as long as $m \le c \ln n / \ln(1/\varepsilon)$ (Theorem 2.6). That is, we can estimate $n(m, \varepsilon)$ in the class of symmetric spaces by ε^{-Cm}, with C a (large) universal constant. Note that our bound for $n(m, \varepsilon)$ is optimal up to the value of C (since in ℓ_∞-case $n(m, \varepsilon)$ cannot be less than ε^{-cm}).

The nice behaviour of the operator Γ is a consequence of a much more general concentration result. As it turns out, for *any* n-dimensional random vector with independent identically distributed coordinates, "sufficiently good" concentration in ℓ_∞^n implies concentration in *all* symmetric norms (Proposition 2.4). This statement follows from the Calderon–Mityagin interpolation theorem and the fact that order statistics of vectors with i.i.d. coordinates are "almost constants". The concentration property of order statistics in the form useful for us is an easy consequence of Chernoff's bounds for the binomial distribution, [22, Theorem 1], by a standard and well-known probabilistic argument. Let us remark here that the use of Chernoff's estimates in geometry of high-dimensional convex bodies was initiated in papers [79] and [4].

The organization of this section is the following. In §2.2.2 we provide some definitions and standard results which form a basis for further discussion. In §2.2.3 we prove the concentration property for vectors with i.i.d. coordinates (Proposition 2.4). In §2.2.4 we deal with Euclidean subspaces of normed spaces. We show that with a high probability the subspace $\mathrm{Im}\Gamma$ is almost Euclidean for all symmetric spaces (Theorem 2.6). Finally, in §2.2.5 we illustrate limitations of this approach by constructing two examples of non-symmetric spaces where $\Gamma(S^{m-1})$ is not "spherical".

2.2.2 Preliminaries

Further, we shall assume a standard lattice order for vectors in $\mathbb{R}^n$, namely,

$$(a_1, a_2, \ldots, a_n) \le (b_1, b_2, \ldots, b_n) \text{ if and only if } a_i \le b_i \text{ for all } i = 1, 2, \ldots, n. \tag{2.1}$$

For any subset $I \subset \{1, 2, \ldots, n\}$, we define the *characteristic function* of I in $\mathbb{R}^n$ as the vector

$$\chi_I = \sum_{i \in I} e_i^n.$$

Let $x = (x_1, x_2, \dots, x_n)$ be a vector of independent identically distributed (i.i.d.) real-valued random variables on a probability space $(\Omega, \Sigma, \mathbb{P})$, each x_i having a cumulative distribution function (cdf) F. For any $\omega \in \Omega$ we denote by $x^*(\omega) = (x_1^*(\omega), x_2^*(\omega), \dots, x_n^*(\omega))$ the non-increasing rearrangement of the vector $(x_1(\omega), x_2(\omega), \dots, x_n(\omega))$, i.e. $x_i^*(\omega)$ is the i-th largest coordinate of $x(\omega)$. We call the random variable x_i^* *the i-th order statistic* of x, $i = 1, 2, \dots, n$. It is easy to check [25, Chapter 2] that the distribution function of the k-th order statistic ($k = 1, 2, \dots, n$) is given by

$$F_k(t) = \sum_{i=n-k+1}^{n} \binom{n}{i} F(t)^i [1 - F(t)]^{n-i} = \sum_{i=0}^{k-1} \binom{n}{i} F(t)^{n-i} [1 - F(t)]^i, \quad t \in \mathbb{R}. \tag{2.2}$$

In particular, the cdf of the largest order statistic

$$F_1(t) = F(t)^n, \quad t \in \mathbb{R}. \tag{2.3}$$

For any $s \in (0, 1)$, *the quantile of order s* is defined by

$$\xi_s = \inf\{t : F(t) > s\}. \tag{2.4}$$

When F is continuous and strictly increasing, ξ_s (as a function of s) is just the inverse of F. Denote by $\eta = (\eta_1, \eta_2, \dots, \eta_n)$ the vector of quantiles of F where

$$\eta_k = \xi_{s(k)} \text{ for } s(k) = 1 - (2k-1)/(2n) \;\; (k = 1, 2, \dots, n). \tag{2.5}$$

We have "$(2k-1)/(2n)$" instead of "k/n" is the last formula to get a well-defined vector for any F. Note that η is non-increasing. We will see that, under some additional conditions, x concentrates near η in any symmetric norm. Here are examples of η for some classical distributions. If F is the uniform distribution on $[0, 1]$ then $\eta_k = 1 - (2k-1)/(2n)$, $k = 1, 2, \dots, n$. If F is Bernoulli with probability of success $1/2$ then $\eta_k = 1$ for $k \le (n+1)/2$ and $\eta_k = 0$ otherwise. In case of the standard normal distribution (with cdf Φ) we have $\eta_k = \Phi^{-1}(1 - (2k-1)/(2n))$; in particular, $\eta_1 - \sqrt{2 \ln 2n} \to 0$ for $n \to \infty$ [25, p. 264].

Estimates for the distributions of order statistics x_k^* play a key role in the proof of the main results of the section. Recall that by Chernoff's theorem [22, Theorem 1], for any

$p \in (0,1)$ and $k \le pn$

$$\sum_{i=0}^{k} \binom{n}{i} p^i (1-p)^{n-i} \le \exp\left(k \ln \frac{pn}{k} + (n-k) \ln \frac{(1-p)n}{n-k} \right). \tag{2.6}$$

Now, let ξ_s be the quantile of order s for some $s \in (0,1)$. Assume for a moment that F is continuous at ξ_s, i.e. $\xi_s = \max\{t : F(t) = s\}$. Then (2.6), together with the inequality $\ln(1+\alpha) \le \alpha - \alpha^2/(2+2\alpha)$ $(\alpha \ge 0)$, implies that for any natural $k < (1-s)n + 1$

$$\mathbb{P}\{x_k^* < \xi_s\} = \sum_{i=0}^{k-1} \binom{n}{i} (1-s)^i s^{n-i} \le \exp\left(-\frac{((1-s)n+1-k)^2}{2n(1-s)} \right). \tag{2.7}$$

Similarly, using the estimate $\ln(1+\alpha) \le \alpha - \alpha^2/2$ $(\alpha \le 0)$ for the right-most logarithm in (2.6), we get: for any $k \ge (1-s)n$

$$\mathbb{P}\{x_k^* > \xi_s\} = \sum_{i=k}^{n} \binom{n}{i} (1-s)^i s^{n-i} \le \exp\left(-\frac{(k-(1-s)n)^2}{2k} \right). \tag{2.8}$$

It is not difficult to show that the estimates in (2.7) and (2.8) hold true for discontinuous F as well.

Clearly, (2.7) and (2.8) imply that for $k \ll \ell$ we have $x_k^* \ge \eta_\ell$ and $x_\ell^* \le \eta_k$ with a large probability. This property shall be described in the next section in terms of dilation operators.

2.2.3 Dilation operators and concentration of order statistics

Fix any n and any $\nu \ge 1$ and consider a linear operator $D_\nu : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$D_\nu e_i^n = \sum_{\ell=1}^{n} \mu([\nu(i-1), \nu i] \cap [\ell-1, \ell]) e_\ell^n, \quad i = 1, 2, \ldots, n,$$

where μ is the Lebesgue measure on $\mathbb{R}$. We say that D_ν is a *dilation operator*. This is a modification of the definition of dilation operators for rearrangement invariant function spaces [65, p. 130].

We shall highlight some of the very basic properties of D_ν. Let $a = (a_1, a_2, \ldots, a_n)$ be a non-increasing sequence. It is not difficult to see that $D_\nu a$ is also non-increasing and $D_\nu a \ge a$. In fact, a stronger relation holds. Fix any $1 \le k \le n$ and let $i = \lceil k/\nu \rceil$. For all

$j > i$ we have $\nu(j-1) \geq k$ which implies $\mu([\nu(j-1), \nu j] \cap [k-1, k]) = 0$. It follows that

$$\langle D_\nu a, e_k^n \rangle_n = \sum_{j=1}^{n} a_j \mu([\nu(j-1), \nu j] \cap [k-1, k]) = \sum_{j=1}^{i} a_j \mu([\nu(j-1), \nu j] \cap [k-1, k]). \quad (2.9)$$

Note the identity

$$\sum_{i=1}^{n} \mu([\nu(i-1), \nu i] \cap [l-1, l]) = 1, \quad l = 1, 2, \ldots, n. \quad (2.10)$$

Together with the fact that a is non-increasing, (2.9) and (2.10) yield

$$\langle D_\nu a, e_k^n \rangle_n \geq \sum_{j=1}^{i} a_i \mu([\nu(j-1), \nu j] \cap [k-1, k]) \geq a_i.$$

Thus,

$$\langle D_\nu a, e_k^n \rangle_n \geq a_i \text{ for all } 1 \leq k \leq n \text{ and } i = \lceil k/\nu \rceil. \quad (2.11)$$

Note that (2.10) implies $\|D_\nu\|_{\infty\to\infty} = 1$. At the same time, for any $i \in \{1, 2, \ldots, n\}$

$$\|D_\nu e_i^n\|_1 \leq \mu([\nu(i-1), \nu i]) = \nu,$$

so $\|D_\nu\|_{1\to1} \leq \nu$. Applying the Calderon–Mityagin interpolation theorem [65, Theorem 2.a.10], we conclude that for any n-dimensional symmetric space E,

$$\|D_\nu\|_{E\to E} \leq \max(\|D_\nu\|_{1\to1}, \|D_\nu\|_{\infty\to\infty}) \leq \nu. \quad (2.12)$$

We now return to the probabilistic setting we discussed earlier. Unitil the end of this section, let $x = (x_1, x_2, \ldots, x_n)$ be a random vector of i.i.d. variables with each coordinate having a cdf F, x^* be the corresponding vector of order statistics and η be defined by (2.5) with respect to F.

Proposition 2.1. *For any $k \in \mathbb{N}$ and $\nu \geq 1$*

$$\max\big(\mathbb{P}\{\langle D_\nu x^*, e_k^n \rangle_n < \eta_k\}, \mathbb{P}\{x_k^* > \langle D_\nu \eta, e_k^n \rangle_n\}\big) \leq \exp\left(-\frac{k}{8}\left(1 - \frac{1}{\nu}\right)^2\right).$$

Proof. Let $i := \lceil k/\nu \rceil$. By (2.11), $\langle D_\nu x^*, e_k^n \rangle_n \geq x_i^*$ and $\langle D_\nu \eta, e_k^n \rangle_n \geq \eta_i$. Consequently,

$$\mathbb{P}\{\langle D_\nu x^*, e_k^n \rangle_n < \eta_k\} \leq \mathbb{P}\{x_i^* < \eta_k\} \text{ and } \mathbb{P}\{x_k^* > \langle D_\nu \eta, e_k^n \rangle_n\} \leq \mathbb{P}\{x_k^* > \eta_i\}.$$

By definition, $\eta_k = \xi_s$ for $s = 1-(2k-1)/(2n)$. Note that $2(k-i+1/2) \geq k-i+1 \geq k-k/\nu$. Then, by (2.7),

$$\begin{aligned}\mathbb{P}\{x_i^* < \eta_k\} &\leq \exp\left(-\frac{((1-s)n+1-i)^2}{2n(1-s)}\right)\\ &= \exp\left(-\frac{(k-i+1/2)^2}{2k-1}\right)\\ &\leq \exp\left(-\frac{k}{8}\left(1-\frac{1}{\nu}\right)^2\right).\end{aligned}$$

Similarly, by (2.8), for $\widetilde{s} = 1-(2i-1)/(2n)$,

$$\begin{aligned}\mathbb{P}\{x_k^* > \eta_i\} &\leq \exp\left(-\frac{(k+n\widetilde{s}-n)^2}{2k}\right)\\ &= \exp\left(-\frac{(k-i+1/2)^2}{2k}\right)\\ &\leq \exp\left(-\frac{k}{8}\left(1-\frac{1}{\nu}\right)^2\right).\end{aligned}$$

□

For a real-valued random variable z with a cdf F_z we say that z is (ε,δ)*–concentrated around* $\tau \in \mathbb{R}$ if

$$\max(\mathbb{P}\{z \leq \tau-\varepsilon\}, \mathbb{P}\{z > \tau+\varepsilon\}) \leq \delta,$$

or, equivalently, $\max(F_z(\tau-\varepsilon), 1-F_z(\tau+\varepsilon)) \leq \delta$.

From (2.3) and the above definition we get: x_1^* is (ε,δ)-concentrated around some $\tau \in \mathbb{R}$ iff

$$F(\tau-\varepsilon) \leq \delta^{1/n} \text{ and } F(\tau+\varepsilon) \geq (1-\delta)^{1/n}. \tag{2.13}$$

Take $k \leq (n+1)/2$ such that $\delta \leq \exp(-(2k-1)\ln 2)$. Then the left-most inequality in (2.13) implies

$$F(\tau-\varepsilon) \leq \exp(-(2k-1)\ln 2/n) = \left(\frac{1}{4}\right)^{(2k-1)/(2n)} \leq 1-\frac{2k-1}{2n}.$$

In view of the definition of η, we get: if x_1^* is (ε,δ)-concentrated around τ then

$$\tau-\varepsilon \leq \eta_k \text{ for all } k \text{ satisfying } k \leq (n+1)/2 \text{ and } \delta \leq \exp(-(2k-1)\ln 2). \tag{2.14}$$

In particular, if $\delta \leq 1/2$ then $\tau-\varepsilon \leq \eta_1$. On the other hand, if $\delta \leq 1-\exp(-1/2)$ then the right-most inequality in (2.13) gives

$$F(\tau+\varepsilon) \geq \exp(-1/(2n)) > 1 - 1/(2n),$$

whence $\tau + \varepsilon \geq \eta_1$. Thus, if $\delta \leq 1 - \exp(-1/2)$ then (ε, δ)-concentration of x_1^* around some $\tau \in \mathbb{R}$ implies $(2\varepsilon, \delta)$-concentration around η_1. Essentially, η_1 is the only possible concentration point for x_1^*.

Lemma 2.2. *Let $k \leq (n+1)/2$ be a natural number and x_1^* be (ε, δ)-concentrated around η_1, where $\delta \leq \exp(-(2k-1)\ln 2)$. Then for any $\nu > 1$*

$$\mathbb{P}\{x^* \leq 2\varepsilon\chi_{\{1,2,\ldots,k\}} + D_\nu \eta\} \geq 1 - \frac{9\nu^2}{(\nu-1)^2} \exp\left(-\frac{k}{8}\left(1-\frac{1}{\nu}\right)^2\right),$$

where the order for vectors is defined by (2.1).

Proof. The conditions on δ imply, by (2.14), that $\eta_k \geq \eta_1 - \varepsilon$. Since $D_\nu \eta \geq \eta$, we obtain

$$\begin{aligned}\mathbb{P}\{x_\ell^* > 2\varepsilon + &\langle D_\nu \eta, e_\ell^n \rangle_n \text{ for some } \ell = 1, \ldots, k\} \\ &\leq \mathbb{P}\{x_\ell^* > 2\varepsilon + \eta_\ell \text{ for some } \ell = 1, \ldots, k\} \\ &\leq \mathbb{P}\{x_1^* > 2\varepsilon + \eta_k\} \\ &\leq \delta < \exp(-k/8).\end{aligned}$$

Further, by Proposition 2.1,

$$\begin{aligned}\mathbb{P}\{x_\ell^* > \langle D_\nu \eta, e_\ell^n \rangle_n \text{ for some } \ell = k+1, \ldots, n\} &\leq \sum_{\ell=k+1}^{n} \mathbb{P}\{x_\ell^* > \langle D_\nu \eta, e_\ell^n \rangle_n\} \\ &\leq \sum_{\ell=k+1}^{n} \exp\left(-\frac{\ell}{8}\left(1-\frac{1}{\nu}\right)^2\right) \\ &\leq \frac{\exp(-(k+1)(1-1/\nu)^2/8)}{1-\exp(-(1-1/\nu)^2/8)} \\ &\leq \frac{8\nu^2}{(\nu-1)^2} \exp\left(-\frac{k}{8}\left(1-\frac{1}{\nu}\right)^2\right).\end{aligned}$$

Combining these estimates, we get

$$\begin{aligned}
&\mathbb{P}\{x^* \not\leq 2\varepsilon\chi_{\{1,2,\dots,k\}} + D_\nu\eta\} \\
&\quad = \mathbb{P}\{x^*_\ell > 2\varepsilon + \langle D_\nu\eta, e^n_\ell\rangle_n \text{ for some } \ell = 1,\dots,k \textbf{ or} \\
&\qquad\quad x^*_\ell > \langle D_\nu\eta, e^n_\ell\rangle_n \text{ for some } \ell = k+1,\dots,n\} \\
&\quad < \frac{9\nu^2}{(\nu-1)^2}\exp\left(-\frac{k}{8}\left(1-\frac{1}{\nu}\right)^2\right).
\end{aligned}$$

□

Lemma 2.3. *Let $\eta_1 - \eta_k \leq \varepsilon$ for some k and ε. Then*

$$\mathbb{P}\{\eta \leq \varepsilon\chi_{\{1,2,\dots,k-1\}} + D_\nu x^*\} \geq 1 - \frac{9\nu^2}{(\nu-1)^2}\exp\left(-\frac{k}{8}\left(1-\frac{1}{\nu}\right)^2\right).$$

Proof. Note that whenever $\langle D_\nu x^*, e^n_\ell\rangle_n < \eta_\ell - \varepsilon$ for some $\ell = 1,2,\dots,k-1$ then, in view of the conditions on η and monotonicity of $D_\nu x^*$,

$$\langle D_\nu x^*, e^n_k\rangle_n \leq \langle D_\nu x^*, e^n_\ell\rangle_n < \eta_\ell - \varepsilon \leq \eta_k.$$

Hence, by Proposition 2.1,

$$\begin{aligned}
&\mathbb{P}\big\{\langle D_\nu x^*, e^n_\ell\rangle_n < \eta_\ell - \varepsilon \text{ for some } \ell = 1,2,\dots,k-1 \textbf{ or } \langle D_\nu x^*, e^n_k\rangle_n < \eta_k\big\} \\
&\quad = \mathbb{P}\{\langle D_\nu x^*, e^n_k\rangle_n < \eta_k\} \\
&\quad \leq \exp\left(-\frac{k}{8}\left(1-\frac{1}{\nu}\right)^2\right).
\end{aligned}$$

Repeating the arguments in the proof of the previous lemma, we get

$$\begin{aligned}
&\mathbb{P}\big\{\langle D_\nu x^*, e^n_\ell\rangle_n < \eta_\ell \text{ for some } \ell = k+1,\dots,n\big\} \\
&\quad \leq \sum_{\ell=k+1}^{n}\mathbb{P}\big\{\langle D_\nu x^*, e^n_\ell\rangle_n < \eta_\ell\big\} \\
&\quad \leq \frac{8\nu^2}{(\nu-1)^2}\exp\left(-\frac{k}{8}\left(1-\frac{1}{\nu}\right)^2\right).
\end{aligned}$$

Finally,

$$\begin{aligned}
&\mathbb{P}\{\eta \not\leq \varepsilon\chi_{\{1,2,\ldots,k-1\}} + D_\nu x^*\} \\
&\quad = \mathbb{P}\{\eta_\ell > \varepsilon + \langle D_\nu x^*, e_\ell^n\rangle_n \text{ for some } \ell = 1,\ldots,k-1 \textbf{ or} \\
&\qquad\quad \eta_\ell > \langle D_\nu x^*, e_\ell^n\rangle_n \text{ for some } \ell = k,\ldots,n\} \\
&\quad \leq \frac{9\nu^2}{(\nu-1)^2}\exp\left(-\frac{k}{8}\left(1-\frac{1}{\nu}\right)^2\right),
\end{aligned}$$

and the result follows. □

We would like to highlight the difference between the assumptions in Lemmas 2.2 and 2.3. Concentration of x_1^* in Lemma 2.2 implies that $\eta_1 - \eta_k \leq \varepsilon$. On the other hand, the condition $\eta_1 - \eta_k \leq \varepsilon$ in Lemma 2.3 does not imply concentration of x_1^* around η_1. For example, let b be a Bernoulli $(0-1)$ random variable with probability of success $\frac{1}{2n}$, u be an independent variable uniformly distributed on $[0;n]$, and let x_i's be independent copies of the product ub. It is not difficult to see that η is the null vector in this case, but

$$\mathbb{P}\{x_1^* > n/2\} = 1 - \left(1 - \frac{1}{4n}\right)^n \geq 1 - \exp(-1/4).$$

Next, we prove the main result of this section:

Proposition 2.4. *Let x be an n-dimensional vector of non-negative i.i.d. random variables on a probability space $(\Omega, \Sigma, \mathbb{P})$, such that $\|x\|_\infty$ is (ε,δ)-concentrated around $\|\eta\|_\infty > 0$ for $\varepsilon \leq \|\eta\|_\infty/2$ and $\delta \leq \exp(-(2k-1)\ln 2)$ for some $k \leq (n+1)/2$. Let $\theta := \varepsilon/\|\eta\|_\infty$. Then with probability greater than $1 - 18\exp(-k\theta^2/8)/\theta^2$, for any symmetric norm $\|\cdot\|$ in $\mathbb{R}^n$ we have*

$$(1-2\theta)\|\eta\| \leq \|x\| \leq (1+6\theta)\|\eta\|. \tag{2.15}$$

Proof. Consider the event

$$A = \left\{\omega \in \Omega :\ x^*(\omega) \leq 2\varepsilon\chi_{\{1,2,\ldots,k\}} + D_\nu\eta \textbf{ and } \eta \leq \varepsilon\chi_{\{1,2,\ldots,k-1\}} + D_\nu x^*(\omega)\right\},$$

where $\nu = \|\eta\|_\infty/(\|\eta\|_\infty - \varepsilon) = 1/(1-\theta)$. Since x and η are non-negative, for any unconditional (and, in particular, symmetric) norm $\|\cdot\|$ in $\mathbb{R}^n$ and any $\omega \in A$ we have

$$\|x^*(\omega)\| \leq 2\varepsilon\|\chi_{\{1,2,\ldots,k\}}\| + \|D_\nu\eta\| \text{ and } \|\eta\| \leq \varepsilon\|\chi_{\{1,2,\ldots,k-1\}}\| + \|D_\nu x^*(\omega)\|. \tag{2.16}$$

By (2.14), $\eta_1 = \|\eta\|_\infty \le \eta_k + \varepsilon$. Then, by Lemmas 2.2 and 2.3,

$$\mathbb{P}(A) \ge 1 - 18\exp(-k\theta^2/8)/\theta^2.$$

Let $\|\cdot\|$ be a symmetric norm in $\mathbb{R}^n$. We shall verify (2.15) for any $\omega \in A$. Since η is non-increasing, $\chi_{\{1,2,\dots,k\}} \le \eta/\eta_k \le \eta/(\|\eta\|_\infty - \varepsilon)$. Then, in view of (2.12) and (2.16),

$$\|x(\omega)\| = \|x^*(\omega)\| \le 2\varepsilon\|\eta\|/(\|\eta\|_\infty - \varepsilon) + \|D_\nu \eta\| \le (2\varepsilon/(\|\eta\|_\infty - \varepsilon) + \nu)\|\eta\|.$$

Note that $\|\eta\|_\infty/(\|\eta\|_\infty - \varepsilon) \le 2$, so

$$2\varepsilon/(\|\eta\|_\infty - \varepsilon) + \nu = 1 + 3\varepsilon/(\|\eta\|_\infty - \varepsilon) \le 1 + 6\theta,$$

and we get the right-most inequality in (2.15). Further, by (2.12) and (2.16),

$$\|\eta\| \le \varepsilon\|\chi_{\{1,2,\dots,k-1\}}\| + \|D_\nu x^*(\omega)\| \le \varepsilon\|\eta\|/(\|\eta\|_\infty - \varepsilon) + \nu\|x(\omega)\|.$$

Hence,

$$\frac{\|\eta\|_\infty - 2\varepsilon}{\|\eta\|_\infty - \varepsilon}\|\eta\| \le \frac{\|\eta\|_\infty}{\|\eta\|_\infty - \varepsilon}\|x(\omega)\|,$$

implying $(1 - 2\theta)\|\eta\|_\infty \le \|x(\omega)\|$, so the left-most inequality in (2.15) is established as well. □

2.2.4 Dependence on ε in Dvoretzky's theorem

As an application of Proposition 2.4, we prove that in any n-dimensional symmetric space there are $(1+\varepsilon)$-Euclidean subspaces of dimension $m = c\ln n/\ln(1/\varepsilon)$ (with c a universal constant). Taking into account the ℓ_∞^n case, this is the optimal order of magnitude for m in the class of symmetric spaces. Thoughout the section, we assume that the distortion ε is less than 1/2.

The following lemma is standard (see, for example, [80, Lemma 4.1] for a similar statement).

Lemma 2.5. *Let $m > 0$ and let $\mathcal{N}$ be an ε-net (with respect to the Euclidean metric in $\mathbb{R}^m$) on S^{m-1} for some $\varepsilon \le 1/2$. Further, let X be a normed space and $T : \mathbb{R}^m \to X$ be a linear operator such that for some $M \in \mathbb{R}_+$ and $\delta \in [0, 1)$ we have*

$$(1-\delta)M \le \|Tz\|_X \le (1+\delta)M \text{ for all } z \in \mathcal{N}.$$

Then for any $y \in \mathrm{S}^{m-1}$ we have

$$(1-2\delta-2\varepsilon)M \le \|Ty\|_X \le (1+2\delta+2\varepsilon)M.$$

Next, pick any $m \ge 2$ and $n \ge 1$ and consider independent random m-dimensional vectors $z_1, z_2, \dots, z_n$ defined on a probability space $(\Omega, \Sigma, \mathbb{P})$ and uniformly distributed on S^{m-1}. We define a random linear operator $\Gamma_{n,m} : \mathbb{R}^m \to \mathbb{R}^n$ by

$$\Gamma_{n,m}(y) := (\langle y, z_i\rangle_m)_{i=1}^n. \tag{2.17}$$

When m and n are clear from the context, we will use "Γ" as an alternative notation. Obviously, for any fixed $y \in \mathrm{S}^{m-1}$, $|\Gamma y| = (|\langle y, z_i\rangle_m|)_{i=1}^n$ is a random vector of non-negative i.i.d. variables. If $\eta = (\eta_1, \eta_2, \dots, \eta_k)$ is the corresponding vector of quantiles defined by (2.5) then for every k the (normalized) measure of a spherical cap in S^{m-1} of geodesic radius $\arccos(\eta_k)$ is $\frac{2k-1}{4n}$.

The operator Γ was studied in [44] in connection with the problem of embedding spaces in ℓ_∞^n. Our next theorem can be viewed as an extension of Theorem 4.1 from [44].

Theorem 2.6. *Let $0 < \varepsilon \le 1/2$ and $3 \le m \le \ln n/(2\ln(64/\varepsilon))$. Further, let $z_1, z_2, \dots, z_n$ be independent random vectors uniformly distributed on S^{m-1}. Then with probability greater than $1 - 2(48/\varepsilon)^{m+2}\exp(-(128/\varepsilon)^{m-2})$ the map Γ has the following property: for any symmetric norm $\|\cdot\|$ in $\mathbb{R}^n$ there exists $M = M(m, n, \|\cdot\|)$ such that*

$$(1-\varepsilon)M \le \|\Gamma y\| \le (1+\varepsilon)M \text{ for all } y \in \mathrm{S}^{m-1}.$$

Proof. Let $\widetilde{\varepsilon} := \varepsilon/16$. Clearly, $m \le \ln n/(2\ln(4/\widetilde{\varepsilon}))$, so by [44, Theorem 4.1], with probability greater than $1 - \exp(-(8/\widetilde{\varepsilon})^m/2)$ we have

$$1 - \widetilde{\varepsilon} \le \|\Gamma y\|_\infty \le 1 \;\text{ for any } y \in \mathrm{S}^{m-1}. \tag{2.18}$$

Let $\mathcal{N}$ be any (non-random) $\widetilde{\varepsilon}$-net on S^{m-1} with $|\mathcal{N}| \le (3/\widetilde{\varepsilon})^m$. Fix for a moment any $z \in \mathcal{N}$ and consider the random vector $|\Gamma z| = (|\langle z, z_i\rangle_m|)_{i=1}^n$. Let η be the vector of quantiles defined by (2.5) with respect to the distribution of $|\langle z, z_1\rangle_m|$ (let us point out that η is completely determined by the numbers m and n). Note that $\eta_1 \in [1-\widetilde{\varepsilon}; 1]$, so in view of (2.18), $\|\Gamma z\|_\infty$ is $(\widetilde{\varepsilon}, \exp(-(8/\widetilde{\varepsilon})^m/2))$-concentrated around $\eta_1 = \|\eta\|_\infty$. Define $\delta = \exp(-(8/\widetilde{\varepsilon})^m/2)$, $k = \lfloor (8/\widetilde{\varepsilon})^m/4\ln 2\rfloor$, $\theta = \widetilde{\varepsilon}/\|\eta\|_\infty$. It is easy to see that $k < (n+1)/2$ and

$$\delta = \exp\left(-2\frac{(8/\widetilde{\varepsilon})^m}{4\ln 2}\ln 2\right) \le \exp(-2k\ln 2) \le \exp(-(2k-1)\ln 2).$$

Then, by Proposition 2.4, with probability greater than $P = 1 - 18\exp(-k\theta^2/8)/\theta^2$, for any symmetric norm $\|\cdot\|$ in $\mathbb{R}^n$ we have

$$(1 - 2\theta)\|\eta\| \le \|\Gamma z\| \le (1 + 6\theta)\|\eta\|. \tag{2.19}$$

Clearly, $\widetilde{\varepsilon} \le \theta$ and $k > (8/\widetilde{\varepsilon})^m/4$, so $P > 1 - 18\exp(-(8/\widetilde{\varepsilon})^{m-2})/\widetilde{\varepsilon}^2$. Now, applying (2.19) to every point $z \in \mathcal{N}$ and using the estimate $\theta \le 9\widetilde{\varepsilon}/8$, we get

$$(1 - 3\widetilde{\varepsilon})\|\eta\| \le \|\Gamma z\| \le (1 + 7\widetilde{\varepsilon})\|\eta\| \text{ for all } z \in \mathcal{N}$$

with probability greater than $1 - 2(3/\widetilde{\varepsilon})^{m+2}\exp(-(8/\widetilde{\varepsilon})^{m-2})$. Finally, by Lemma 2.5, the last identity implies

$$(1 - 16\widetilde{\varepsilon})\|\eta\| \le \|\Gamma y\| \le (1 + 16\widetilde{\varepsilon})\|\eta\| \text{ for all } y \in \mathrm{S}^{m-1},$$

and the proof is complete. □

It is easy to see that for m, n and a sufficiently small ε satisfying the conditions of Theorem 2.6, the subspace $\mathrm{Im}\Gamma_{n,m} \subset \mathbb{R}^n$ is $(1+4\varepsilon)$-Euclidean for any symmetric norm $\|\cdot\|$ in $\mathbb{R}^n$ with a high probability. In particular, with a large probability $\Gamma(\mathrm{S}^{m-1})$ is "almost spherical" in a sense that $\sup_{y\in\mathrm{S}^{m-1}} \|\Gamma y\|_2 \le (1+4\varepsilon) \inf_{y\in\mathrm{S}^{m-1}} \|\Gamma y\|_2$ (see [44, Lemma 4.2] for a direct proof of this fact).

2.2.5 Negative results on concentration

Here, we consider a question which naturally appears in connection with the above results: can Theorem 2.6 be generalized to a wider class of norms? Specifically, for any n and m and any $\varepsilon \in (0,1)$ let $\mathcal{C}(n,m,\varepsilon)$ be the collection of all norms in $\mathbb{R}^n$ satisfying

$$\sup_{y\in\mathrm{S}^{m-1}} \|\Gamma y\| / \inf_{y\in\mathrm{S}^{m-1}} \|\Gamma y\| \le 1 + \varepsilon$$

with a large probability (say, greater than 1/2). Obviously, whatever parameters n, m, ε we take, there are unconditional norms not contained in $\mathcal{C}(n,m,\varepsilon)$: a simple example is a weighted ℓ_∞^n-space with the norm $\|(x_1, x_2, \ldots, x_n)\| = \|(Cx_1, x_2, \ldots, x_n)\|_\infty$ for a large C. On the other hand, this space is not "balanced". We recall some definitions. For a norm $\|\cdot\|$ in $\mathbb{R}^n$, let B be the closed unit ball of $\|\cdot\|$. We say that B is *isotropic* if it has volume 1 and for any $\theta \in S^{n-1}$

$$\int_{\|x\|\le 1} |\langle x,\theta\rangle_n|^2\,dx = L^2$$

for some L independent of θ. Next, B is in *John's position* if the ellipsoid of maximal volume inscribed in B is the unit Euclidean ball B_2^n.

We can consider the following problem: given some (large) m and (small) $\varepsilon > 0$, is it possible to find $n_0 = n_0(m,\varepsilon)$ such that for any $n \ge n_0$ the set $\mathcal{C}(n,m,\varepsilon)$ contains all unconditional norms with unit balls in John's (or isotropic) position? It turns out that such n_0 does not exist. We will need the following elementary lemma:

Lemma 2.7. *Let $P \subset \mathbb{R}^m$ be a centrally-symmetric convex polytope with at most $2N$ facets, and $\|\cdot\|_P$ be the corresponding Minkowski functional. Then* $\operatorname{dist}(P, B_2^m) \ge 1 + (2N)^{-2/(m-1)}/3$.

Proof. Without loss of generality, we may assume that S^{m-1} is inscribed in P (i.e. touches every $(n-1)$-facet of P) and

$$\operatorname{dist}(P, B_2^m) = \sup_{\|y\|_P = 1} \|y\|_2 = 1+\varepsilon$$

for some $\varepsilon > 0$. It is not difficult to show that for $\theta = \arccos\frac{1}{1+\varepsilon}$ the contact points of P and S^{m-1} must form a θ-net on S^{m-1} (with respect to the geodesic metric). Using estimates for the measure of caps (see [80, p. 7]), we get

$$(2N)^{-1} \le \frac{\sqrt{m-2}}{2}\int_0^{\theta} \sin^{m-2} t\,dt \le \theta^{m-1} \le \left((1+\varepsilon)^2 - 1\right)^{(m-1)/2},$$

implying $\varepsilon \ge (2N)^{-2/(m-1)}/3$. □

Proposition 2.8. *For any $m \ge 8$ and $n > 64^2 m$ there is an unconditional norm in $\mathbb{R}^n$ with the unit ball in John's position such that*

$$\mathbb{P}\Big\{\sup_{y\in S^{m-1}} \|\Gamma y\| / \inf_{y\in S^{m-1}} \|\Gamma y\| < 1 + 2^{-4^7}/3\Big\} < 1/2.$$

Proof. Let $k = 64^2 m$ and define $B \subset \mathbb{R}^k \oplus \mathbb{R}^{n-k}$ by

$$B = (\sqrt{k} B_1^k) \oplus_\infty B_\infty^{n-k},$$

where B_p^ℓ is the unit ball of $\|\cdot\|_p$-norm in $\mathbb{R}^\ell$. Let $\|\cdot\|$ be the Minkowski functional for B in $\mathbb{R}^n = \mathbb{R}^k \oplus \mathbb{R}^{n-k}$ (i.e., $\|(x,y)\| = \max\{k^{-1/2}\|x\|_1, \|y\|_\infty\}$ for all $(x,y) \in \mathbb{R}^k \oplus \mathbb{R}^{n-k}$) and denote $E = (\mathbb{R}^n, \|\cdot\|)$. Next, we show that B is in John's position. Note that for an operator $T : \ell_2^n \longrightarrow E$, $T(B_2^n)$ is the ellipsoid of maximal volume in B if and only if $\|T\|_{2\to E} = 1$ and *the nuclear norm* $\nu(\cdot)$ of the inverse T^{-1} is equal to n (for a proof see Theorem 14.2 and Proposition 14.3 in [114]). Now, it is easy to check that for the formal identity operator $I : \ell_2^n \longrightarrow E$, $\|I\|_{2\to E} = 1$ and $\nu(I^{-1}) = n$; thus, B is in John's position.

Let $(\Omega, \Sigma, \mathbb{P})$ be the probability space for the vectors $z_1, z_2, \ldots, z_n$ from the definition of Γ and denote

$$A = \big\{\omega \in \Omega : \ \|\Gamma(\omega)y\| \le 1 \text{ for some } y \in \mathrm{S}^{m-1}\big\}.$$

Using estimates for the measure of spherical caps [80, p. 7], it is easy to check that $\mathbb{E}\,|\langle y_0, z_1\rangle_m| \ge \mathbb{P}\big\{|\langle y_0, z_1\rangle_m| \ge 1/\sqrt{m}\big\}/\sqrt{m} > 1/(8\sqrt{m}) = 8/\sqrt{k}$ for any fixed $y_0 \in \mathrm{S}^{m-1}$. Applying Hoeffding's inequality [50, Theorem 1], we obtain

$$\begin{aligned}
&\mathbb{P}\big\{\sup_{y\in \mathrm{S}^{m-1}} \|\Gamma y\| / \inf_{y\in \mathrm{S}^{m-1}} \|\Gamma y\| \ge 2 \,|\, A\big\} \cdot \mathbb{P}(A) + \mathbb{P}(\Omega \setminus A) \\
&\quad \ge \mathbb{P}\{\|\Gamma y_0\| \ge 2\} \\
&\quad = \mathbb{P}\big\{|\langle y_0, z_1\rangle_m| + |\langle y_0, z_2\rangle_m| + \cdots + |\langle y_0, z_k\rangle_m| \ge 2\sqrt{k}\big\} \\
&\quad \ge 1 - \mathbb{P}\left\{\mathbb{E}\,|\langle y_0, z_1\rangle_m| - \frac{1}{k}\sum_{i=1}^{k} |\langle y_0, z_i\rangle_m| \ge \frac{6}{\sqrt{k}}\right\} \\
&\quad \ge 1 - \exp(-72).
\end{aligned}$$

Next, for $\omega \in \Omega \setminus A$ we have

$$1 < \|\Gamma(\omega)y\| = k^{-1/2}\big(|\langle y, z_1(\omega)\rangle_m| + |\langle y, z_2(\omega)\rangle_m| + \cdots + |\langle y, z_k(\omega)\rangle_m|\big) \text{ for all } y \in \mathrm{S}^{m-1}.$$

Note that $P = \big\{u \in \mathbb{R}^m : |\langle u, z_1(\omega)\rangle_m| + |\langle u, z_2(\omega)\rangle_m| + \cdots + |\langle u, z_k(\omega)\rangle_m| = 1\big\}$ is a convex polytope with no more than 2^k facets and that $\sup_{y\in \mathrm{S}^{m-1}} \|\Gamma(\omega)y\| / \inf_{y\in \mathrm{S}^{m-1}} \|\Gamma(\omega)y\| \ge \mathrm{dist}\big(P, B_2^m\big)$, so by Lemma 2.7

$$\sup_{y\in \mathrm{S}^{m-1}} \|\Gamma(\omega)y\| / \inf_{y\in \mathrm{S}^{m-1}} \|\Gamma(\omega)y\| \ge 1 + 2^{-2k/(m-1)}/3 \ge 1 + 2^{-4^7}/3.$$

The last inequality holds for all $\omega \in \Omega \setminus A$, so finally

$$\begin{aligned}
&\mathbb{P}\big\{ \sup_{y\in \mathrm{S}^{m-1}} \|\Gamma y\| / \inf_{y\in \mathrm{S}^{m-1}} \|\Gamma y\| \geq 1 + 2^{-4^7}/3 \big\} \\
&\quad = \mathbb{P}\big\{ \sup_{y\in \mathrm{S}^{m-1}} \|\Gamma y\| / \inf_{y\in \mathrm{S}^{m-1}} \|\Gamma y\| \geq 1 + 2^{-4^7}/3 \,|\, A \big\} \cdot \mathbb{P}(A) + 1 \cdot \mathbb{P}(\Omega \setminus A) \\
&\quad \geq 1 - \exp(-72).
\end{aligned}$$

□

Proposition 2.9. *For any $m > 2 + 32\pi^2$ and $n > 256m$ there is an unconditional norm in $\mathbb{R}^n$ with the isotropic unit ball and such that*

$$\mathbb{P}\big\{ \sup_{y\in \mathrm{S}^{m-1}} \|\Gamma y\| / \inf_{y\in \mathrm{S}^{m-1}} \|\Gamma y\| < 2/\sqrt{3} \big\} < 1/2.$$

Proof. Take $k = n-1$ and let $B = (\sqrt{(k+1)(k+2)/6}\, B_1^k) \oplus_\infty B_\infty^1$, with the corresponding Minkowski functional

$$\|(x_1, x_2, \ldots, x_n)\| = \max \left\{ \sqrt{\frac{6}{(k+1)(k+2)}} \sum_{i=1}^{k} |x_i|, |x_n| \right\}.$$

It is not difficult to verify that, for some $\lambda > 0$, λB is isotropic. As before, $z_1, z_2 \ldots, z_n$ are the independent random vectors on the probability space $(\Omega, \Sigma, \mathbb{P})$ from the definition of Γ. Now, for any fixed vector $y_0 \in \mathrm{S}^{m-1}$ we have from the concentration inequality on the sphere (see, for example, [80, Corollary 2.2]):

$$\begin{aligned}
\mathbb{E}\,(|\langle y_0, z_1\rangle_m|) &= \int_0^1 \mathbb{P}\{|\langle y_0, z_1\rangle_m| \geq t\}\, dt \\
&\leq \int_0^\infty \sqrt{\frac{\pi}{2}} \exp(-t^2(m-2)/2)\, dt \\
&= \frac{\pi}{2\sqrt{m-2}} \\
&\leq \frac{1}{8\sqrt{2}}.
\end{aligned}$$

Then, by Hoeffding's inequality [50, Theorem 1]

$$\mathbb{P}\big\{ |\langle y_0, z_1\rangle_m| + |\langle y_0, z_2\rangle_m| + \cdots + |\langle y_0, z_k\rangle_m| \geq k/(4\sqrt{2}) \big\} \leq \exp(-k/64). \tag{2.20}$$

Further, let $\mathcal{N}$ be a fixed 1/2-net on S^{m-1} (with respect to the Euclidean metric in $\mathbb{R}^m$) with $|\mathcal{N}| \le 5^m$. Denote

$$A = \big\{\omega \in \Omega : \, |\langle y, z_1(\omega)\rangle_m| + |\langle y, z_2(\omega)\rangle_m| + \cdots + |\langle y, z_k(\omega)\rangle_m| < k/(4\sqrt{2}) \text{ for all } y \in \mathcal{N}\big\}.$$

From (2.20) we have $\mathbb{P}(A) \ge 1 - 5^m \exp(-k/64) \ge 1 - \exp(-2m)$. By successive approximation [80, proof of Lemma 4.1], any $y \in \mathrm{S}^{m-1}$ can be written as $y = y_1 + \sum_{i=2}^{\infty} \delta_i y_i$, where $y_i \in \mathcal{N}$ and $|\delta_i| \le 2^{1-i}$ for all i. Then for all $\omega \in A$ and $y \in \mathrm{S}^{m-1}$ we have

$$|\langle y, z_1(\omega)\rangle_m| + |\langle y, z_2(\omega)\rangle_m| + \cdots + |\langle y, z_k(\omega)\rangle_m| \le k/(4\sqrt{2}) + \sum_{i=2}^{\infty} \delta_i k/(4\sqrt{2}) \le k/(2\sqrt{2}).$$

Combining this inequality with the estimate for $\mathbb{P}(A)$ we get

$$\mathbb{P}\left\{\sqrt{\frac{6}{(k+1)(k+2)}} \sum_{i=1}^{k} |\langle y, z_i\rangle_m| \le \frac{\sqrt{3}}{2} \text{ for all } y \in \mathrm{S}^{m-1}\right\} \ge 1 - \exp(-2m).$$

On the other hand, $\sup\limits_{y \in \mathrm{S}^{m-1}} |\langle y, z_n\rangle_m| = 1$, so in view of the definition of the norm

$$\mathbb{P}\big\{\sup_{y \in \mathrm{S}^{m-1}} \|\Gamma y\| / \inf_{y \in \mathrm{S}^{m-1}} \|\Gamma y\| \ge 2/\sqrt{3}\big\} \ge 1 - \exp(-2m).$$

□

2.3 The Randomized Dvoretzky's Theorem in ℓ_∞^n and the χ-distribution[2]

2.3.1 Introduction

The classical theorem of A. Dvoretzky in the version improved and strengthened by V. Milman, states: *there is a function* $f(\varepsilon) > 0$ *such that for all* $\varepsilon \in (0, 1/2)$, $n > 1$ *and* $1 \le k \le f(\varepsilon) \ln n$, *any* n*-dimensional normed space admits a* k*-dimensional subspace which is* $(1+\varepsilon)$*-Euclidean.*

See [26] and [76], respectively, for the original theorems. A broad perspective of the subject and its developments can be found in the books [80] and [86], as well as in a recent survey [98] and references therein.

The bound $k \le f(\varepsilon) \ln n$ is in general optimal with respect to n, but the form of the function $f(\varepsilon)$ is not clear up to this day. The original formula for $f(\varepsilon)$ from [76] was subsequently improved to $f(\varepsilon) = c\varepsilon^2$ in [43] and then to $f(\varepsilon) = c\varepsilon/(\ln \frac{1}{\varepsilon})^2$ in [100]. As we showed in Section 2.2, in the class of n-dimensional spaces with a 1-symmetric basis we have $f(\varepsilon) = c/\ln \frac{1}{\varepsilon}$.

The problem of optimal dependence on ε in Dvoretzky's theorem can be "randomized" as follows: given an n-dimensional normed space X, determine all k such that a random k-dimensional subspace of X is $(1+\varepsilon)$-Euclidean with a high probability. Of course, the solution depends on the definition of "randomness". For example, in [44] the question was considered for $X = \ell_\infty^n$ and a certain probabilistic model which gives $(1+\varepsilon)$-Euclidean subspaces with a large probability for all $k \le c \ln n / \ln \frac{1}{\varepsilon}$. However, the distribution of the random subspaces in [44] is not invariant under rotations. The (unique) rotation invariant distribution of subspaces of ℓ_∞^n was studied in [99].

It was proved in [99] that the standard Gaussian vector $g = (g_1, g_2, \ldots, g_n)$ in $\mathbb{R}^n$ satisfies

$$\mathbb{P}\{\|g\|_\infty < (1-\varepsilon)\mathrm{Med}\,\|g\|_\infty\} \le 2\exp(-n^{c\varepsilon}),$$
$$\mathbb{P}\{\|g\|_\infty > (1+\varepsilon)\mathrm{Med}\,\|g\|_\infty\} \le 2n^{-c\varepsilon},$$

where $\mathrm{Med}\,\|g\|_\infty$ is the median of the norm of g in ℓ_∞^n and $c > 0$ is a universal constant. A usual covering argument then implies that a random k-dimensional subspace $E \subset \ell_\infty^n$, uniformly distributed on the Grassmannian $\mathrm{G}_{n,k}$, is $(1+\varepsilon)$-spherical with probability at

[2] A version of this section has been published. K. E. Tikhomirov, The Randomized Dvoretzky's theorem in ℓ_∞^n and the χ-distribution, Geometric Aspects of Functional Analysis, Lecture Notes in Mathematics, **2116** (2014), 455–463.

least $1-2n^{-\widetilde{c}\varepsilon}$, provided that $k \le \widetilde{c}\,\varepsilon \ln n/\ln\frac{1}{\varepsilon}$. Of course, a natural question is whether $\ln\frac{1}{\varepsilon}$ in the upper bound for k can be removed. In [99] it was claimed that it is indeed possible.

The main purpose of this work is to show that in fact $\ln\frac{1}{\varepsilon}$ is necessary and the bound $k \le \widetilde{c}\varepsilon \ln n/\ln\frac{1}{\varepsilon}$ is optimal (up to the choice of the constant). In other words, if $k \ge \max(2, C\varepsilon \ln n/\ln\frac{1}{\varepsilon})$ then with a substantial probability the random k-dimensional subspace $E \subset \ell_\infty^n$ uniformly distributed on $\mathrm{G}_{n,k}$ is *not* $(1+\varepsilon)$-spherical. To achieve our goal, we shall link the geometry of "typical" subspaces of ℓ_∞^n to certain properties of the χ-distribution.

2.3.2 Preliminaries

Let us start with some notation. The probability space $(\mathbb{P}, \Sigma, \Omega)$ is fixed. Everywhere in this section, $g_1, g_2, \dots$ are independent standard Gaussian variables and $g = (g_1, \dots, g_n)$ is the standard Gaussian vector in $\mathbb{R}^n$. By $\mathrm{Med}\,\|g\|_2$ ($\mathrm{Med}\,\|g\|_\infty$) we will denote the median of $\|g\|_2$ (respectively, the median of $\|g\|_\infty$). Finally, with some abuse of terminology, we will call a subspace $E \subset \ell_\infty^n$ $(1+\varepsilon)$-*spherical* if

$$\sup_{\substack{x\in E\\ \|x\|_2=1}} \|x\|_\infty / \inf_{\substack{x\in E\\ \|x\|_2=1}} \|x\|_\infty \le 1+\varepsilon.$$

From well known estimates for the Gaussian distribution (see, for example, [25, p. 264] or [32, Lemma VII.1.2]) it follows that for $\alpha \to 0$,

$$\frac{1}{\alpha}\mathbb{P}\Big\{|g_1| \ge \sqrt{2\ln(1/\alpha)} - \frac{\ln\ln(1/\alpha)}{4\sqrt{\ln(1/\alpha)}}\Big\} \longrightarrow 0; \tag{2.21}$$

$$\frac{1}{\alpha}\mathbb{P}\Big\{|g_1| \ge \sqrt{\ln(1/\alpha)}\Big\} \longrightarrow \infty. \tag{2.22}$$

Fix some $k > 1$. The variable $\xi^{(k)} = \sqrt{\sum\limits_{i=1}^{k} g_i^2}$ has the χ-distribution with k degrees of freedom; the distribution density f_k of $\xi^{(k)}$ is given by

$$f_k(t) = \begin{cases} \frac{t^{k-1}e^{-t^2/2}}{2^{k/2-1}\Gamma(k/2)}, & t \ge 0;\\ 0, & \text{otherwise.}\end{cases}$$

where Γ is the Gamma function. It is obvious that $\mathbb{P}\{\xi^{(k)} \ge \tau\} \ge \mathbb{P}\{|g_1| \ge \tau\}$ for all

$\tau > 0$, so in view of (2.22) for all sufficiently small $\alpha > 0$,

$$\mathbb{P}\big\{\xi^{(k)} \geq \sqrt{\ln(1/\alpha)}\big\} \geq \alpha. \tag{2.23}$$

We shall use the formula for f_k to improve the last estimate. Suppose that $\tau = \tau(\alpha)$ satisfies $\mathbb{P}\{\xi^{(k)} \geq \tau\} = \alpha$, i.e.

$$\alpha = \int_{\tau}^{\infty} f_k(t)\, dt.$$

By (2.23), for small α we have $\tau \geq \sqrt{\ln(1/\alpha)}$ and

$$\begin{aligned}\alpha &= \int_{\tau}^{\infty} \frac{t^{k-1}e^{-t^2/2}}{2^{k/2-1}\Gamma(k/2)}\, dt = \int_{\tau^2}^{\infty} \frac{t^{k/2-1}e^{-t/2}}{2^{k/2}\Gamma(k/2)}\, dt \geq \int_{\tau^2}^{\infty} \frac{(\ln(1/\alpha))^{\ell}e^{-t/2}}{2^{k/2}\Gamma(k/2)}\, dt \\ &= \frac{(\ln(1/\alpha))^{\ell}e^{-\tau^2/2}}{2^{\ell}\Gamma(k/2)},\end{aligned}$$

where $\ell = \frac{k}{2} - 1$. Hence,

$$\tau \geq \sqrt{2\ln(1/\alpha) + 2\ln\frac{(\ln(1/\alpha))^{\ell}}{2^{\ell}\Gamma(k/2)}} \geq \sqrt{2\ln(1/\alpha) + 2\ell\ln\frac{\ln(1/\alpha)}{2\ell+1}}. \tag{2.24}$$

Clearly, for $\ell \ll \ln(1/\alpha)$ we have $\ln(1/\alpha) \geq \ell \ln\frac{\ln(1/\alpha)}{2\ell+1}$, and (2.24) implies

$$\tau \geq \sqrt{2\ln(1/\alpha)} + \frac{\ell\ln\frac{\ln(1/\alpha)}{2\ell+1}}{2\sqrt{\ln(1/\alpha)}}.$$

Thus, we have shown the following:

Lemma 2.10. *There are universal constants $\alpha_0 > 0$ and $c_0 > 0$ such that whenever $\alpha \in (0, \alpha_0)$ and $2 \leq k \leq c_0 \ln(1/\alpha)$ then*

$$\mathbb{P}\Big\{\xi^{(k)} \geq \sqrt{2\ln(1/\alpha)} + \frac{(k/2-1)\ln\frac{\ln(1/\alpha)}{k-1}}{2\sqrt{\ln(1/\alpha)}}\Big\} \geq \alpha. \tag{2.25}$$

2.3.3 Random subspaces of ℓ_∞^n

For natural numbers n and k, let $\Gamma_{nk} : \mathbb{R}^k \to \mathbb{R}^n$ be the standard Gaussian operator given by

$$\Gamma_{nk}(z) = \Gamma_{nk}(z_1, z_2, \ldots, z_k) = \sum_{i=1}^{n}\left(\sum_{j=1}^{k} g_{ij}z_j\right)e_i^n$$

for $z = (z_1, z_2, \ldots, z_k) \in \mathbb{R}^k$, where $\{g_{ij}\}$ are independent standard Gaussian variables.

The following proposition, together with the results of [99] on the distribution of $\|g\|_\infty$, is the main tool in proving the central result of the note. It shows that with a substantial probability, the number

$$\sup_{z \in \mathrm{S}^{k-1}} \|\Gamma_{nk} z\|_\infty / \inf_{z \in \mathrm{S}^{k-1}} \|\Gamma_{nk} z\|_\infty$$

is noticeably farther from 1 than $\|g\|_\infty / \mathrm{Med}\,\|g\|_\infty$:

Proposition 2.11. *There are universal constants $c > 0$ and $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ and all $k > 1$,*

$$\sup_{z \in \mathrm{S}^{k-1}} \|\Gamma_{nk} z\|_\infty / \inf_{z \in \mathrm{S}^{k-1}} \|\Gamma_{nk} z\|_\infty > 1 + \frac{ck \ln \frac{c \ln n}{k}}{\ln n} \tag{2.26}$$

with probability greater than $1/2$.

Proof. By (2.21), there exists $\alpha_1 > 0$ such that for all $n \geq \alpha_1^{-1}$ and $k \geq 1$,

$$\begin{aligned} &\mathbb{P}\Big\{ \|\Gamma_{nk}(1, 0, \ldots, 0)\|_\infty \leq \sqrt{2 \ln n} - \frac{\ln \ln n}{4\sqrt{\ln n}} \Big\} \\ &\quad = \mathbb{P}\Big\{ |g_1| \leq \sqrt{2 \ln n} - \frac{\ln \ln n}{4\sqrt{\ln n}} \Big\}^n > \frac{1}{2} + \frac{1}{e}. \end{aligned} \tag{2.27}$$

Define $n_0 = \lceil \max(\alpha_0^{-1}, \alpha_1^{-1}) \rceil$ and $c = \min(c_0, 1/24)$, where α_0 and c_0 are taken from Lemma 2.10. Now, fix any $n \geq n_0$ and $k > 1$. Note that for $k \geq c_0 \ln n$ the statement is trivial so we will assume that $k < c_0 \ln n$. For any point of the probability space $\omega \in \Omega$,

$$\begin{aligned} \sup_{z \in \mathrm{S}^{k-1}} \|\Gamma_{nk}(\omega) z\|_\infty &= \max_i \sup_{(z_1, \ldots, z_k) \in \mathrm{S}^{k-1}} |z_1 g_{i1}(\omega) + \cdots + z_k g_{ik}(\omega)| \\ &= \max_i \sqrt{g_{i1}(\omega)^2 + \cdots + g_{ik}(\omega)^2} \\ &= \max_i \xi_i(\omega), \end{aligned}$$

where $\xi_1, \xi_2, \ldots, \xi_n$ are independent random variables having the χ-distribution with k

degrees of freedom. Letting $\alpha = 1/n$ in (2.25), we get:

$$\begin{aligned}
&\mathbb{P}\Big\{\max_i \xi_i \geq \sqrt{2\ln n} + \frac{(k-2)\ln\frac{\ln n}{k-1}}{4\sqrt{\ln n}}\Big\}\\
&= 1 - \mathbb{P}\Big\{\xi_1 \leq \sqrt{2\ln n} + \frac{(k-2)\ln\frac{\ln n}{k-1}}{4\sqrt{\ln n}}\Big\}^n\\
&\geq 1 - \Big(1 - \frac{1}{n}\Big)^n\\
&\geq 1 - \frac{1}{e}.
\end{aligned}$$

Combining the last estimate with (2.27), we get the result. □

Further, we will need the following version of the covering argument: *Let $k > 1$ and $\mathcal{N}$ be a θ-net (with respect to the Euclidean norm $\|\cdot\|_2$ in $\mathbb{R}^k$) on S^{k-1} for some $\theta < 1/2$. Next, let X be a normed space and $T : \mathbb{R}^k \to X$ be a linear operator such that for some $M > 0$ and $\delta \in [0,1)$*

$$(1-\delta)M \leq \|Ty\|_X \leq (1+\delta)M \ \ \textit{for all } y \in \mathcal{N}.$$

Then for any $z \in \mathrm{S}^{k-1}$

$$(1 - 2\delta - 2\theta)M \leq \|Tz\|_X \leq (1 + 2\delta + 2\theta)M. \tag{2.28}$$

For convenience we give a short proof. For every $z \in \mathrm{S}^{k-1}$, there is $y \in \mathcal{N}$ such that $\|y - z\|_2 \leq \theta$. Then

$$\|Tz\|_X \leq \|Ty\|_X + \|T(y-z)\|_X \leq (1+\delta)M + \theta\|T\|,$$

where $\|T\|$ denotes the operator norm from ℓ_2^k to X. Taking the maximum over z, we get $\|T\| \leq (1+\delta)(1-\theta)^{-1}M$. In particular this implies the right hand side inequality in (2.28). For the left hand side we start with

$$\|Tz\|_X \geq \|Ty\|_X - \|T(y-z)\|_X \geq (1-\delta)M - \theta\|T\|,$$

and then use the estimate for $\|T\|$ obtained above.

The next statement expresses the well known fact that, for $k \ll n$, with a large probability $\|\Gamma_{nk}z\|_2$ is almost a constant on the sphere S^{k-1}. Let $C_2 > 0$ be such that for all $n > 1$ the dimension of any 3/2-spherical subspace of l_∞^n is bounded from above by $C_2 \ln n$.

Lemma 2.12. *There is a universal constant $n_1 \in \mathbb{N}$ such that for all $n \geq n_1$ and $k \leq C_2 \ln n$*

$$\mathbb{P}\Big\{\sup_{z\in \mathrm{S}^{k-1}} \|\Gamma_{nk} z\|_2 / \inf_{z\in \mathrm{S}^{k-1}} \|\Gamma_{nk} z\|_2 \leq 1 + \frac{1}{n^{1/4}}\Big\} \geq \frac{3}{4}. \tag{2.29}$$

Proof. By a concentration inequality for Gaussian vectors (see, for example, [86, Theorem 4.7] or [80, Theorem V.1]) and since $\mathrm{Med}\,\|g\|_2 \approx \sqrt{n}$, we have for some $C_1 > 0$, all $n \geq 1$ and the standard Gaussian vector g in $\mathbb{R}^n$:

$$\mathbb{P}\{|\|g\|_2 - \mathrm{Med}\,\|g\|_2| > \theta \mathrm{Med}\,\|g\|_2\} \leq 2\exp(-C_1\theta^2 n) \ \text{ for any } \theta > 0. \tag{2.30}$$

Then we choose the constant $n_1 \in \mathbb{N}$ so that for all $n \geq n_1$

$$1 - 2(48n^{1/4})^{C_2 \ln n} \exp(-C_1\sqrt{n}/256) \geq 3/4.$$

Fix any $n \geq n_1$ and $1 \leq k \leq C_2 \ln n$; let $\theta = \frac{1}{16n^{1/4}}$ and $\mathcal{N}$ be a θ-net on S^{k-1} of cardinality at most $(3/\theta)^k$. For any point $z \in \mathrm{S}^{k-1}$, $\Gamma_{nk} z$ is the standard Gaussian vector in $\mathbb{R}^n$, so in particular

$$\begin{aligned} &\mathbb{P}\big\{|\|\Gamma_{nk} y\|_2 - \mathrm{Med}\,\|g\|_2| \leq \theta \mathrm{Med}\,\|g\|_2 \text{ for all } y \in \mathcal{N}\big\} \\ &\quad \geq 1 - (3/\theta)^k \mathbb{P}\big\{|\|g\|_2 - \mathrm{Med}\,\|g\|_2| > \theta \mathrm{Med}\,\|g\|_2\big\} \\ &\quad \geq 1 - 2(3/\theta)^k \exp(-C_1\theta^2 n). \end{aligned}$$

The covering argument implies that

$$\begin{aligned} &\mathbb{P}\Big\{\sup_{z\in \mathrm{S}^{k-1}} \|\Gamma_{nk} z\|_2 / \inf_{z\in \mathrm{S}^{k-1}} \|\Gamma_{nk} z\|_2 \leq 1 + 16\theta\Big\} \\ &\quad \geq 1 - 2(3/\theta)^k \exp(-C_1\theta^2 n) \\ &\quad \geq 1 - 2(48n^{1/4})^{C_2 \ln n} \exp(-C_1\sqrt{n}/256) \\ &\quad \geq \frac{3}{4}, \end{aligned}$$

and the result follows. □

To emphasize the geometric character of our main result we shall present it in terms of the Grassmannians. Note that the probabilistic formulations used until now — which are more convenient for calculations — still remain in the proof of the theorem. Note that in view of invariance of the distribution of the Gaussian vector under rotations, we have for any Borel subset $A \subset \mathrm{G}_{n,k}$

$$\mathbb{P}\Big\{\mathrm{Im}\Gamma_{nk} \in A\Big\} = \mu_{n,k}(A). \tag{2.31}$$

Theorem 2.13. *Let* $\varepsilon \in (0, 1/2)$ *and* $n > 1$. *Then*

1) *There is a universal constant* $\widetilde{c} > 0$ *such that whenever* $k \leq \widetilde{c}\varepsilon \ln n / \ln \frac{1}{\varepsilon}$, *then*

$$\mu_{n,k}\{E \in \mathrm{G}_{n,k} : \ E \textit{ is } (1+\varepsilon)\textit{-spherical subspace of } \ell_\infty^n\} \geq 1 - 2n^{-\widetilde{c}\varepsilon}; \tag{2.32}$$

2) *Conversely, if for some* $k > 1$

$$\mu_{n,k}\{E \in \mathrm{G}_{n,k} : \ E \textit{ is } (1+\varepsilon)\textit{-spherical subspace of } \ell_\infty^n\} \geq \frac{3}{4} \tag{2.33}$$

then necessarily $k \leq C\varepsilon \ln n / \ln \frac{1}{\varepsilon}$, *where* $C > 0$ *is a universal constant.*

Proof. The first part of the theorem is essentially proved in [99]. Indeed, by [99, Proposition 1], for some constant $c_1 > 0$

$$\mathbb{P}\Big\{\big|\|g\|_\infty - \mathrm{Med}\,\|g\|_\infty\big| > \varepsilon \mathrm{Med}\,\|g\|_\infty\Big\} \leq 2n^{-c_1\varepsilon}. \tag{2.34}$$

When n is small or $\varepsilon < \frac{1}{\ln n}$, (2.32) is obvious (for a well-chosen constant $\widetilde{c}$), so we can assume that

$$\frac{c_1(64 \ln n)^2}{C_1 n} \leq 1, \ \ \varepsilon \geq \frac{1}{\ln n}, \tag{2.35}$$

where C_1 is taken from (2.30). Pick a natural number $k \leq \frac{c_1 \varepsilon \ln n}{18 \ln(1/\varepsilon)}$. As before, Γ_{nk} is the Gaussian operator. Note that event $\{\mathrm{Im}\Gamma_{nk}$ is $(1+\varepsilon)$-spherical$\}$ contains the event

$$\Big\{ \sup_{z \in \mathrm{S}^{k-1}} \|\Gamma_{nk} z\|_2 / \inf_{z \in \mathrm{S}^{k-1}} \|\Gamma_{nk} z\|_2 \leq 1 + \frac{\varepsilon}{4} \textbf{ and}$$
$$\sup_{z \in \mathrm{S}^{k-1}} \|\Gamma_{nk} z\|_\infty / \inf_{z \in \mathrm{S}^{k-1}} \|\Gamma_{nk} z\|_\infty \leq 1 + \frac{\varepsilon}{4}\Big\}.$$

Then in view of (2.31), (2.30), (2.34) and the covering argument for $\theta = \varepsilon/64$, we have

$$\begin{aligned}
&\mu_{n,k}\{E \in \mathrm{G}_{n,k} :\ E \text{ is } (1+\varepsilon)\text{-spherical}\} \\
&\quad \geq \mathbb{P}\Big\{ \sup_{z\in \mathrm{S}^{k-1}} \|\Gamma_{nk} z\|_2 / \inf_{z\in \mathrm{S}^{k-1}} \|\Gamma_{nk} z\|_2 \leq 1 + \frac{\varepsilon}{4} \textbf{ and} \\
&\qquad \sup_{z\in \mathrm{S}^{k-1}} \|\Gamma_{nk} z\|_\infty / \inf_{z\in \mathrm{S}^{k-1}} \|\Gamma_{nk} z\|_\infty \leq 1 + \frac{\varepsilon}{4} \Big\} \\
&\quad \geq 1 - (3/\theta)^k \mathbb{P}\big\{ \big| \|g\|_2 - \mathrm{Med}\,\|g\|_2 \big| > \theta \mathrm{Med}\,\|g\|_2 \big\} \\
&\qquad - (3/\theta)^k \mathbb{P}\big\{ \big| \|g\|_\infty - \mathrm{Med}\,\|g\|_\infty \big| > \theta \mathrm{Med}\,\|g\|_\infty \big\} \\
&\quad \geq 1 - 2(3/\theta)^k \exp(-C_1 \theta^2 n) - 2(3/\theta)^k n^{-c_1 \varepsilon}.
\end{aligned}$$

By (2.35), $C_1\theta^2 n \geq c_1 \varepsilon \ln n$, hence

$$\mathbb{P}\big\{ \mathrm{Im}\Gamma_{nk} \text{ is } (1+\varepsilon)\text{-spherical} \big\} \geq 1 - 4n^{-c_1\varepsilon} n^{\frac{c_1}{18}\varepsilon \ln\frac{3}{\theta} / \ln\frac{1}{\varepsilon}} \geq 1 - 4n^{-c_1\varepsilon/2}.$$

The statement follows by properly defining $\widetilde{c}$.

Now, we turn to the second part of the theorem. Suppose that $k > 1$ satisfies (2.33). This implies, in particular, that ℓ_∞^n contains $(1+\varepsilon)$-Euclidean subspaces of dimension k, so $k \leq C_3 \ln n / \ln \frac{1}{\varepsilon}$ for a universal constant C_3 (see, for example, [98, Claim 3.3]). Let n_0, n_1, C_2 and c be as they were defined in Proposition 2.11 and Lemma 2.12. The cases when n or $1/\varepsilon$ is small, can be treated in a trivial way, so further we assume

$$n \geq \max(n_0, n_1), \quad \frac{3\ln n}{cn^{1/4}} \leq 1, \quad k \leq \frac{c}{e}\ln n, \quad \varepsilon < \frac{c^2}{3}. \tag{2.36}$$

Obviously, $k \leq C_2 \ln n$. Then (2.29) and (2.33) give

$$\begin{aligned}
&\mathbb{P}\big\{ \sup_{z\in \mathrm{S}^{k-1}} \|\Gamma_{nk} z\|_\infty / \inf_{z\in \mathrm{S}^{k-1}} \|\Gamma_{nk} z\|_\infty \leq (1+\varepsilon)(1 + 1/n^{1/4}) \big\} \\
&\quad \geq \mathbb{P}\Big\{ \sup_{z\in \mathrm{S}^{k-1}} \|\Gamma_{nk} z\|_2 / \inf_{z\in \mathrm{S}^{k-1}} \|\Gamma_{nk} z\|_2 \leq 1 + \frac{1}{n^{1/4}} \textbf{ and} \\
&\qquad \sup_{\substack{x \in \mathrm{Im}\Gamma_{nk} \\ \|x\|_2 = 1}} \|x\|_\infty / \inf_{\substack{x \in \mathrm{Im}\Gamma_{nk} \\ \|x\|_2 = 1}} \|x\|_\infty \leq 1 + \varepsilon \Big\} \\
&\quad \geq \frac{1}{2}.
\end{aligned}$$

Hence, by Proposition 2.11,

$$1 + \frac{ck \ln \frac{c \ln n}{k}}{\ln n} \leq (1+\varepsilon)(1 + 1/n^{1/4}). \tag{2.37}$$

If $\varepsilon \le 1/n^{1/4}$ then, in view of (2.36) and (2.37), $k \le k \ln \frac{c \ln n}{k} \le \frac{3 \ln n}{c n^{1/4}} \le 1$, leading to contradiction. Hence, $\varepsilon > 1/n^{1/4}$, and (2.37) yields

$$k \le \frac{3\varepsilon \ln n}{c \ln \frac{c \ln n}{k}}. \tag{2.38}$$

In particular (2.38) implies $\frac{k}{\ln n} \le \frac{3}{c}\varepsilon$, so $\ln \frac{c \ln n}{k} \ge \ln \frac{c^2}{3\varepsilon}$. Substituting it back to (2.38) we get

$$k \le \frac{3\varepsilon \ln n}{c \ln \frac{c^2}{3\varepsilon}}.$$

□

Remark 2.1. The probability 3/4 in the second part of the Theorem can be replaced with any (fixed) positive number; this only affects the constant.

2.4 On the Distance of Polytopes with Few Vertices to the Euclidean Ball[3]

2.4.1 Introduction

In this section, we consider lower bounds for the Banach–Mazur distance $\mathrm{dist}(P_N, B_2^n)$ when P_N is a convex polytope in $\mathbb{R}^n$ with N vertices, and $N < 2n$.

It is known that for all $N \geq 2n$, the distance $\mathrm{dist}(P_N, B_2^n)$ can be estimated from below as

$$\mathrm{dist}(P_N, B_2^n) \geq c\sqrt{\frac{n}{\ln \frac{N}{n}}}, \tag{2.39}$$

where $c > 0$ is a universal constant (see [11], [15, Corollary 9.5], [18] or [38]). The last relation implies $\mathrm{dist}(P_N, B_2^n) \geq \widetilde{c}\sqrt{n}$ for any convex polytope with at most $2n$ vertices.

The bound (2.39) is optimal up to the constant multiple; in fact, for any $n \in \mathbb{N}$ and $n + 1 \leq N \leq 2^n$ there is a polytope P_N in $\mathbb{R}^n$ with N vertices and $\mathrm{dist}(P_N, B_2^n) \leq C\sqrt{n/\ln \frac{N}{n}}$ for a universal constant $C > 0$ (see [33, p. 96] for $N \geq 2n$ and [13] for $N < 2n$).

Now, let us focus on the polytopes P_N with $N/n \approx 1$. It is known that for an n-simplex Δ, $\mathrm{dist}(\Delta, B_2^n) = n$ (see, for example, [46, §7]). The natural question of the distance of P_N ($n + 1 \leq N < 2n$) to the set of symmetric convex bodies in $\mathbb{R}^n$ was solved in [40]. The main result of [40] implies

$$\mathrm{dist}(P_N, B_2^n) \geq \frac{cn}{N - n}, \quad n + 1 \leq N \leq 2n.$$

However, the above inequality is weaker than (2.39). A.E. Litvak informed us that the authors of [40] conjectured that (2.39) holds for *all* $N \geq n + 1$. The objective of this section is to verify this proposition.

The question of estimating the distance $\mathrm{dist}(P_N, B_2^n)$ is closely related to the problem of optimal covering of S^{n-1} by equal spherical caps (for a detailed discussion of coverings by spherical caps, we refer the reader to [12, Chapter 6]). In fact, our result implies (see [13] or [12, Lemma 6.5.2]) that for all natural n and $N \geq n + 1$, *any* covering of S^{n-1} by N equal spherical caps of geodesic radius ϕ satisfies

$$\cos\phi \leq C\sqrt{\frac{\ln \frac{N}{n}}{n}},$$

[3] A version of this section has been published. K. E. Tikhomirov, On the distance of polytopes with few vertices to the Euclidean ball, Discrete Comput. Geom. **53** (2015), no. 1, 173–181.

where C is a universal constant.

In [13], it was conjectured that for $n \geq 3$ and $n+1 \leq N \leq 2n$, the density of covering S^{n-1} by N equal spherical caps is minimal *only if* the centers of the caps are vertices of $N-n$ pairwise orthogonal regular simplices of circumradius one of dimensions $\lceil \frac{n}{N-n} \rceil$ and $\lfloor \frac{n}{N-n} \rfloor$ (in a less explicit form, the question is asked in [121, A list of open problems, Problem#17]). The result is well known when $N = n+1$ (see, for example, [12, Theorem 6.5.1]). The case $N = 2n = 8$ is treated in [24]; see [31] for $N = 2n = 6$ and [101] for $n = 3, N = 5$. In [13] the conjecture was verified for $N = n+2$. Note that whereas our result does not completely resolve the question, it shows that the configuration of the centers of caps described above is "close" to optimal in the following sense: if ϕ_{min} is the minimal geodesic radius among *all* coverings of S^{n-1} by N equal caps and ϕ_0 is the radius corresponding to the above configuration of the centers of caps then $\cos\phi_0 \geq c\cos\phi_{min}$ for a universal constant $c > 0$.

Throughout the subsection, Δ_r^n is a regular n-simplex in $\mathbb{R}^n$ inscribed into S^{n-1}.

2.4.2 Lower bounds for $\mathrm{dist}(P, B_2^n)$

Note that, for any $n \geq 1$ and $n+1 \leq N \leq 2n$, the quantity $\sqrt{n/\ln\frac{N}{n}}$ is equivalent to $n/\sqrt{N-n}$. Then the main result of the note can be stated as follows:

Theorem 2.14. *Let $n \in \mathbb{N}$ and let P be a convex n-dimensional polytope with $n+k$ vertices. Then $\mathrm{dist}(P, B_2^n) \geq cn/\sqrt{k}$ where $c > 0$ is a universal constant.*

The main step in proving the theorem is to obtain an upper bound for the normalized measure of spherical simplices contained in a cap of a given radius ϕ (Proposition 2.19). Let us first consider several auxiliary statements. The next lemma is taken from [24]:

Lemma 2.15 ([24, Lemma 6])**.** *Let $n > 1$, $\phi \in (0, \pi/2)$ and let $\mathcal{C}$ be a spherical cap on S^{n-1} of radius ϕ. Further, let $\mathcal{S} \subset \mathcal{C}$ be a spherical simplex that has the maximal measure among all simplices lying within $\mathcal{C}$. Then $\mathcal{S}$ is regular.*

In what follows, we will need an upper bound for the normalized surface measure of the set $\{x \in \mathrm{S}^{n-1} : \rho x \in \Delta_r^n\}$. We will show that the quantity is *small* for all $\rho \gg (n\ln(n+1))^{-1/2}$. Note that the case $\rho \gg n^{-1/2}$ can be handled trivially. Indeed, since $\mathrm{Vol}_n(\Delta_r^n) = \frac{(n+1)^{(n+1)/2}}{n!\, n^{n/2}}$ and $\mathrm{Vol}_n(B_2^n) = \frac{\pi^{n/2}}{\Gamma(\frac{n}{2}+1)}$, for some universal constant $\widetilde{c} > 0$ we have

$$\mathrm{Vol}_n(\Delta_r^n) \leq (\widetilde{c}\sqrt{n})^{-n}\mathrm{Vol}_n(B_2^n).$$

Then for $\rho \geq \frac{2}{c} n^{-1/2}$ we get

$$\sigma_{n-1}\big\{x \in \mathrm{S}^{n-1} : \rho x \in \Delta_r^n\big\} \leq \frac{\mathrm{Vol}_n(\Delta_r^n)}{\mathrm{Vol}_n(\rho B_2^n)} \leq 2^{-n}.$$

When $\rho \ll n^{-1/2}$, a more delicate estimate is required.

We say that an m-dimensional random vector X is *Gaussian* if every linear combination of its coordinates is normally distributed. By g we shall denote the standard Gaussian vector in $\mathbb{R}^n$, i.e. a vector with i.i.d. coordinates, each coordinate having normal distribution with zero mean and unit variance. Let us recall Slepian's lemma for Gaussian processes ([104], [63, Corollary 3.12]):

Lemma 2.16 (Slepian)**.** *Let m be a natural number, $X = (x_1, x_2, \dots, x_m)$ and $Y = (y_1, y_2, \dots, y_m)$ be two m-dimensional Gaussian vectors with zero mean satisfying*

$$\begin{aligned} &\mathbb{E}\, x_i^2 = \mathbb{E}\, y_i^2, && i = 1, 2, \dots, m, \\ &\mathbb{E}\,(x_i - x_j)^2 \leq \mathbb{E}\,(y_i - y_j)^2, && 1 \leq i, j \leq m. \end{aligned}$$

Then for all $\lambda \in \mathbb{R}$ we have

$$\mathbb{P}\big\{\max_i y_i \leq \lambda\big\} \leq \mathbb{P}\big\{\max_i x_i \leq \lambda\big\}.$$

As an immediate corollary, we get the following upper bound for the standard Gaussian measure of a regular simplex centered at the origin:

Lemma 2.17. *For any $\rho > 0$ and $n > 1$*

$$\mathbb{P}\big\{g \in \rho \Delta_r^n\big\} \leq \exp\Big(-c_1 n \exp\Big(-\frac{\rho^2}{n^2}\Big)\Big).$$

Proof. From well known estimates of the normal distribution (see, for example, [32, Lemma VII.1.2]) it follows that for some universal constant $c_1 > 0$ and the standard Gaussian vector g

$$\mathbb{P}\big\{\langle g, e_i^n \rangle_n > \tau\big\} \geq c_1 \exp(-\tau^2), \quad \tau \in \mathbb{R}, \quad i = 1, 2, \dots, n. \tag{2.40}$$

Let $v^1, v^2, \dots, v^{n+1}$ be vertices of Δ_r^n, so that

$$\rho \Delta_r^n = \Big\{x \in \mathbb{R}^n : \max_{i \leq n+1} \langle x, -v^i \rangle_n \leq \frac{\rho}{n}\Big\}.$$

Since the angles between any two vertices v^i, v^j $(i \neq j)$ are obtuse, we have

$$\mathbb{E}\left(\langle g, v^i\rangle_n - \langle g, v^j\rangle_n\right)^2 > \mathbb{E}\left(\langle g, e_i^n\rangle_n - \langle g, e_j^n\rangle_n\right)^2, \;\; 1 \le i, j \le n.$$

Then, by Slepian's lemma and (2.40),

$$\begin{aligned}\mathbb{P}\big\{g \in \rho\Delta_r^n\big\} &= \mathbb{P}\Big\{\max_{i\le n} \langle g, -v^i\rangle_n \le \frac{\rho}{n}\Big\}\\ &\le \mathbb{P}\Big\{\max_{i\le n} \langle g, e_i^n\rangle_n \le \frac{\rho}{n}\Big\}\\ &\le \Big(1 - c_1 \exp\Big(-\frac{\rho^2}{n^2}\Big)\Big)^n\\ &\le \exp\Big(-c_1 n \exp\Big(-\frac{\rho^2}{n^2}\Big)\Big).\end{aligned}$$

□

In the next lemma, we give a "geometric" version of the above statement:

Lemma 2.18. *There are universal constants $c_2, C_3 > 0$ such that for any $n > 1$ and $\rho > 0$*

$$\sigma_{n-1}\big\{x \in \mathrm{S}^{n-1} : \rho x \in \Delta_r^n\big\} \le 2\exp\Big(-c_2 n \exp\Big(-\frac{C_3}{\rho^2 n}\Big)\Big). \tag{2.41}$$

Proof. Fix $n > 1$ and $\rho > 0$. We have

$$\begin{aligned}\sigma_{n-1}\big\{x \in \mathrm{S}^{n-1} : \rho x \in \Delta_r^n\big\} &= \mathbb{P}\Big\{\frac{\rho g}{\|g\|_2} \in \Delta_r^n\Big\}\\ &\le \mathbb{P}\{\|g\|_2 > 2\sqrt{n}\} + \mathbb{P}\Big\{g \in \frac{2\sqrt{n}}{\rho}\Delta_r^n\Big\}.\end{aligned}$$

A concentration inequality for Gaussian vectors (see, for example, [86, Theorem 4.7] or [80, Theorem V.1]) implies that $\mathbb{P}\{\|g\|_2 > 2\sqrt{n}\} \le \exp(-\widetilde{c}n)$ for a universal constant $\widetilde{c} > 0$. Then, from the above estimates and Lemma 2.17, we obtain

$$\sigma_{n-1}\big\{x \in \mathrm{S}^{n-1} : \rho x \in \Delta_r^n\big\} \le 2\exp\Big(-c_2 n \exp\Big(-\frac{4}{\rho^2 n}\Big)\Big),$$

where $c_2 = \min(c_1, \widetilde{c})$. □

Remark 2.2. It can be shown that the estimate (2.41) is valid for *any* n-simplex Δ contained in B_2^n (for well chosen universal constants c_2, C_3). However, we will not need the property, as Lemma 2.15 will allow us to work only with regular simplices.

Proposition 2.19. *Let $n \geq 4$ and let $\mathcal{S}$ be a spherical simplex on S^{n-1} lying within a cap of radius $\phi \in \big[\arccos \frac{1}{\sqrt{n}}, \frac{\pi}{2}\big)$. Then*

$$\sigma_{n-1}(\mathcal{S}) \leq \exp\Big(-c_4 n^2 \cos^2\phi \,\ln \frac{e}{n\cos^2\phi}\Big),$$

where $c_4 > 0$ is a universal constant.

Proof. Fix any admissible n and ϕ and let $h = \cos\phi$. In view of Lemma 2.15, we can assume that $\mathcal{S}$ is a regular simplex with the center at $(0,\dots,0,1)$ and with vertices lying in the hyperplane $H = \{x \in \mathbb{R}^n : \langle x, e_n^n\rangle_n = h\}$. We define auxiliary hyperplanes $H_t = \{x \in \mathbb{R}^n : \langle x, e_n^n\rangle_n = t\}$ ($h \leq t \leq 1$). Then, using an estimate $\sigma_{n-1}\big\{x \in \mathrm{S}^{n-1} : \langle x, e_n^n\rangle_n \geq 1/2\big\} \leq \exp(-n/8)$ (see, for example, [80, Corollary 2.2]), we obtain

$$\begin{aligned}\sigma_{n-1}(\mathcal{S}) &= \frac{1}{\mathrm{Vol}_{n-1}(\mathrm{S}^{n-1})}\int_h^1 (1-t^2)^{-1/2}\,\mathrm{Vol}_{n-2}(H_t\cap\mathcal{S})\,dt\\ &< \exp(-n/8) + \frac{2}{\mathrm{Vol}_{n-1}(\mathrm{S}^{n-1})}\int_h^{1/2}\mathrm{Vol}_{n-2}(H_t\cap\mathcal{S})\,dt.\end{aligned}$$

Denote by $\mathrm{Proj} : \big\{x \in \mathrm{S}^{n-1} : \langle x, e_n^n\rangle_n > 0\big\} \longrightarrow H$ a mapping from the upper hemisphere onto H given by

$$\mathrm{Proj}(x) = \frac{h}{\langle x, e_n^n\rangle_n}x, \quad x \in \mathrm{S}^{n-1},\ \langle x, e_n^n\rangle_n > 0.$$

Note that for every $t \in [h, 1/2]$, a point $x \in H$ belongs to $\mathrm{Proj}(H_t \cap \mathcal{S})$ if and only if $\frac{t}{h}x$ belongs to $H_t \cap \mathcal{S}$. Hence,

$$\mathrm{Vol}_{n-2}(H_t\cap\mathcal{S}) = \Big(\frac{t}{h}\Big)^{n-2}\mathrm{Vol}_{n-2}\big(\mathrm{Proj}(H_t\cap\mathcal{S})\big). \tag{2.42}$$

Moreover, the set $\mathrm{Proj}(H_t \cap \mathcal{S})$ is precisely the intersection of the regular $(n-1)$-simplex $\mathrm{Proj}(\mathcal{S})$ with $(n-2)$-sphere $\big\{x \in H : \|x - he_n^n\|_2 = \frac{h\sqrt{1-t^2}}{t}\big\}$, and, setting $u = \frac{h\sqrt{1-t^2}}{t}$, we obtain

$$\begin{aligned}\mathrm{Vol}_{n-2}\big(\mathrm{Proj}(H_t\cap\mathcal{S})\big) &= \mathrm{Vol}_{n-2}\big(\sqrt{1-h^2}\Delta_r^{n-1}\cap u\mathrm{S}^{n-2}\big)\\ &\leq \mathrm{Vol}_{n-2}\big(\Delta_r^{n-1}\cap u\mathrm{S}^{n-2}\big)\\ &= \mathrm{Vol}_{n-2}(u\mathrm{S}^{n-2})\,\sigma_{n-2}\big\{x \in \mathrm{S}^{n-2} : ux \in \Delta_r^{n-1}\big\}.\end{aligned}$$

Now, by Lemma 2.18 we get

$$\mathrm{Vol}_{n-2}\big(\mathrm{Proj}(H_t\cap\mathcal{S})\big)\le 2\mathrm{Vol}_{n-2}(u\mathrm{S}^{n-2})\exp\Big(-c_2(n-1)\exp\Big(-\frac{C_3}{u^2(n-1)}\Big)\Big).$$

Returning to the formula for $\sigma_{n-1}(\mathcal{S})$ and making use of (2.42), the last inequality and the obvious estimate $1-t^2\ge\frac{3}{4}$ for $t\in[0,1/2]$, we obtain

$$\sigma_{n-1}(\mathcal{S})<\exp(-n/8)+\tau_n\int_h^{1/2}(1-t^2)^{n/2}\exp\Big(-c_5 n\exp\Big(-\frac{C_6t^2}{h^2n}\Big)\Big)\,dt,$$

with $\tau_n=6\frac{\mathrm{Vol}_{n-2}(\mathrm{S}^{n-2})}{\mathrm{Vol}_{n-1}(\mathrm{S}^{n-1})}$, $c_5=c_2/2$ and $C_6=2C_3$. A standard formula for the surface area of spheres (see, for example, [23, §7.3]) implies that $\tau_n\le C_7\sqrt{n}$ for some universal constant $C_7>0$. Further, $(1-t^2)^{n/2}\le\exp(-t^2n/2)$. Hence,

$$\sigma_{n-1}(\mathcal{S})<\exp(-n/8)+C_7\sqrt{n}\int_0^{\infty}\exp\Big(-\frac{1}{2}t^2n-c_5n\exp\Big(-\frac{C_6t^2}{h^2n}\Big)\Big)\,dt.$$

Let $t_0=\sqrt{\frac{h^2n}{2C_6}\ln\frac{e}{h^2n}}$. Then

$$c_5\exp\Big(-\frac{C_6t_0^2}{h^2n}\Big)=c_5\sqrt{\frac{h^2n}{e}}\ge\frac{2c_5C_6}{e}t_0^2.$$

Since the function $\exp\big(-\frac{C_6t^2}{h^2n}\big)$ is decreasing on $[0,\infty)$, the sum

$$\frac{1}{2}t^2+c_5\exp\Big(-\frac{C_6t^2}{h^2n}\Big)$$

is greater than $\frac{2c_5C_6}{e}t_0^2$ for all $t\in[0,t_0]$. Let $c_8=\frac{2c_5C_6}{e}$. Then

$$\begin{aligned}\sigma_{n-1}(\mathcal{S})&<\exp(-n/8)+C_7\sqrt{n}t_0\exp\big(-c_8t_0^2n\big)+C_7\sqrt{n}\int_{t_0}^{\infty}\exp\Big(-\frac{1}{2}t^2n\Big)\,dt\\&\le C_9\exp\big(-c_{10}t_0^2n\big),\end{aligned}$$

for some universal constants $C_9\ge1$ and $c_{10}>0$. Finally, we note that $\sigma_{n-1}(\mathcal{S})\le\frac{1}{2}$, which, together with the last estimate, yields

$$\sigma_{n-1}(\mathcal{S})\le\min\big(1/2,C_9\exp\big(-c_{10}t_0^2n\big)\big)\le\exp\big(-c_{11}t_0^2n\big),$$

where $c_{11} = \frac{c_{10} \ln 2}{\ln 2C_9}$. It remains to use the definition of t_0. $\square$

Remark 2.3. Proposition 2.19 can also be proved using the identity $\sigma_{n-1}(\mathcal{S}) = \mathbb{P}\big\{g \in \mathcal{C}\big\}$, where $\mathcal{C}$ is the cone in $\mathbb{R}^n$ generated by $\mathcal{S}$. The probability that the Gaussian vector g belongs to $\mathcal{C}$ can be estimated with help of Lemma 2.17 and calculations very similar to the ones above.

Proof of Theorem 2.14. For small n, the statement follows by choosing a sufficiently small constant $c > 0$. Hence, we can assume that $n \geq 4$. Next, the definition of the Banach–Mazur distance implies that, given a convex body K in $\mathbb{R}^n$, there is an operator $T \in GL_n(\mathbb{R})$ and a point $x \in \mathbb{R}^n$ such that the body $\widetilde{K} = T(K) + x$ contains the origin in its interior, $\widetilde{K} \subset y + B_2^n$ for some $y \in \mathbb{R}^n$ and $\mathrm{dist}(K, B_2^n) = \mathrm{dist}(\widetilde{K}, B_2^n) = \inf\{\lambda \geq 1 :\ B_2^n \subset \lambda\widetilde{K}\}$. Then, clearly, $\frac{1}{2}\widetilde{K} \subset B_2^n$ and $2\mathrm{dist}(K, B_2^n) \geq \inf\{\lambda \geq 1 :\ B_2^n \subset \frac{\lambda}{2}\widetilde{K}\}$. Thus, to prove the theorem, it is enough to check that for any convex polytope $P \subset B_2^n$ with $n + k$ vertices ($1 \leq k \leq n$) and the origin in its interior, we have

$$d := \inf\{\lambda \geq 1 :\ B_2^n \subset \lambda P\} \geq \frac{\widetilde{c}n}{\sqrt{k}}.$$

Without loss of generality, P is simplicial. Let $\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_\ell$ be spherical simplices which are central projections of the facets of P onto S^{n-1} (here, ℓ is the total number of the facets). Let $\phi = \arccos \min(\frac{1}{d}, \frac{1}{\sqrt{n}})$. Then $\phi \in [\arccos \frac{1}{\sqrt{n}}, \frac{\pi}{2})$ and each $\mathcal{S}_i$ is contained in a cap of radius ϕ. Hence, in view of Proposition 2.19, we have

$$\begin{aligned} 1 = \sum_{i=1}^{\ell} \sigma_{n-1}(\mathcal{S}_i) &\leq \ell \exp\Big(-c_4 n^2 \cos^2 \phi \, \ln \frac{e}{n \cos^2 \phi}\Big) \\ &\leq \ell \exp\Big(-n(c_5 n \cos^2 \phi) \, \ln \frac{2e}{c_5 n \cos^2 \phi}\Big) \end{aligned} \tag{2.43}$$

for some universal constant $c_5 > 0$. The number of facets ℓ necessarily satisfies

$$\ell \leq \binom{n+k}{n} \leq \exp\Big(k \ln \frac{e(n+k)}{k}\Big) \leq \exp\Big(n \frac{k}{n} \ln \frac{2en}{k}\Big),$$

which, together with (2.43), implies

$$\frac{k}{n} \ln \frac{2en}{k} \geq c_5 n \cos^2 \phi \, \ln \frac{2e}{c_5 n \cos^2 \phi}.$$

Note that the function $f(t) = t \ln \frac{2e}{t}$ is strictly increasing on $(0, 1]$, so the above inequality yields

$$\frac{k}{n} \geq c_5 n \cos^2 \phi,$$

hence, in view of the definition of ϕ, $\max(\sqrt{n}, d) \geq \sqrt{c_5}n/\sqrt{k}$. By a simple volumetric argument, $d \geq c_6\sqrt{n}$ for some universal constant $c_6 \in (0, 1]$, and finally

$$\frac{\sqrt{c_5}c_6 n}{\sqrt{k}} \leq \max(c_6\sqrt{n}, d) \leq d.$$

□

Chapter 3

The Smallest Singular Value of Random Matrices

3.1 The Smallest Singular Value of Random Rectangular Matrices with no Moment Assumptions on Entries[1]

3.1.1 Introduction

In the last years, spectral properties of random matrices with fixed dimensions (the corresponding theory is often called *non-asymptotic*) have attracted considerable attention of researchers, whose efforts have been mostly concentrated on studying distributions of the largest and the smallest singular values. For detailed information on the development of the subject, we refer the reader to surveys [91, 117].

Let $N \geq n$. Given an $N \times n$ ($N \geq n$) random matrix A, we write $s_{\max}(A) = \sup_{y \in S^{n-1}} \|Ay\|_2$; $s_{\min}(A) = \inf_{y \in S^{n-1}} \|Ay\|_2$. A limiting result of Z.D. Bai and Y.Q. Yin [8] suggests that for an $N \times n$ matrix with i.i.d. mean zero entries with unit variance and a finite fourth moment, its largest and smallest singular values should "concentrate" near $\sqrt{N} + \sqrt{n}$ and $\sqrt{N} - \sqrt{n}$, respectively. In the non-asymptotic setting one is interested, in particular, in finding the weakest possible conditions on random matrices that would imply $s_{\max} \lesssim \sqrt{N} + \sqrt{n}$ and $s_{\min} \gtrsim \sqrt{N} - \sqrt{n}$ with a large probability.

For a random $N \times n$ matrix A with i.i.d. mean zero subgaussian entries, an elemen-

[1] A version of this section has been published. K.E. Tikhomirov. The smallest singular value of random rectangular matrices with no moment assumptions on entries. Israel Journal of Mathematics, 2016. DOI: 10.1007/s11856-016-1287-8.

tary application of the standard covering argument yields $s_{\max}(A) \le C(\sqrt{N}+\sqrt{n})$ with an overwhelming probability. Distribution of the smallest singular value when $N \approx n$ requires a more delicate analysis. A. Litvak, A. Pajor, M. Rudelson and N. Tomczak-Jaegermann showed in [66] that if N and n satisfy $N/n \ge 1+h_1(\ln N)^{-1}$ then $\mathbb{P}\{s_{\min}(A) \le h_2\sqrt{N}\} \le \exp(-h_3 N)$, where h_1, h_3 depend only on the variance and the subgaussian moment and h_2 — on the moments and the aspect ratio N/n. The approach initiated in [66] was further developed by M. Rudelson and R. Vershynin who combined it with certain Littlewood–Offord-type theorems. In [95], Rudelson and Vershynin treated square matrices and later in [93] — rectangular matrices with an arbitrary aspect ratio and i.i.d. mean zero subgaussian entries, thereby sharpening and generalizing the result of [66]. We note that the Littlewood–Offord theory has gained an important role in the study of random matrices primarily due to T. Tao and V. Vu (see, in particular, [108]).

Various estimates for the extremal singular values were obtained when studying the problem of approximating the covariance matrix of a random vector by the empirical covariance matrix. Answering a question of R. Kannan, L. Lovász and M. Simonovits, the authors of [1] treated log-concave random vectors. Later, the log-concavity was replaced by weaker assumptions (see, in particular, [2, 106, 74, 47]).

Recently, it has become apparent that different conditions are required to bound the largest and the smallest singular value, and these two questions should be handled separately. One of results proved by N. Srivastava and R. Vershynin in [106] provides a lower estimate for the second moment of $s_{\min}(A)$, where A is an $N \times n$ matrix with independent isotropic rows satisfying a $(2+\varepsilon)$-moment condition and certain assumptions on the aspect ratio N/n. It is important to note that the conditions imposed on A are too weak to imply the "usual" upper bound $s_{\max}(A) \lesssim \sqrt{N}$ with a large probability [68]. This result of [106] was strengthened by V. Koltchinskii and S. Mendelson in [59] under similar assumptions on the matrix. Another theorem of [59] states the following: given an n-dimensional isotropic random vector X satisfying $\inf\limits_{y \in S^{n-1}} \mathbb{P}\{|\langle X, y\rangle_n| \ge \alpha\} \ge \beta$ for some $\alpha, \beta > 0$, there are $H_1, h_2, h_3 > 0$ depending only on α, β such that for $N \ge H_1 n$ and the $N \times n$ random matrix A with i.i.d. rows distributed like X, one has $\mathbb{P}\{s_{\min}(A) \ge h_2\sqrt{N}\} \ge 1-\exp(-h_3 N)$. We note that a closely related question of bounding random quadratic forms from below was considered by R.I. Oliveira in [82] and that the results of V. Koltchinskii and S. Mendelson were further strengthened in [118].

The isotropy of a random vector or, more generally, boundedness of variances of its coordinates is quite a natural assumption which appears as part of requirements on rows of a matrix in all of the aforementioned papers. However, for a deeper understanding of non-asymptotic characteristics of random matrices, an important question is *whether* *any moment assumptions on entries are really necessary* in order to get satisfactory lower estimates for the smallest singular value.

Unlike in [106] and [59] where the matrix entries within a given row are not necessarily independent, in this section we consider the classical setting when a rectangular matrix has i.i.d. entries. However, in contrast with all the mentioned results, the lower estimate for the smallest singular value that we prove does not use any moment assumptions; the only requirement is that the distribution of the entries satisfies a “spreading” condition given in terms of the Lévy concentration function. Moreover, compared to [106] and [59], we significantly relax the assumptions on the aspect ratio of the matrix.

Given a real random variable ξ, *the concentration function* of ξ is defined as

$$\mathcal{Q}(\xi,t)=\sup_{\lambda\in\mathbb{R}}\mathbb{P}\big\{|\xi-\lambda|\le t\big\},\;\;t\ge 0.$$

The notion of the concentration function was introduced by P. Lévy [64] in context of studying distributions of sums of random variables. Note that for a random variable ξ with zero median satisfying $\mathbb{E}\,|\xi|^p\ge m$ and $\mathbb{E}\,|\xi|^q\le M$ for some $0<p<q$ and $m,M>0$, we have $\mathcal{Q}(\xi,\alpha)\le 1-\beta$ for some $\alpha,\beta>0$ depending only on p,q,m,M. At the same time, the condition $\mathcal{Q}(\xi,\alpha)\le 1-\beta$ for some $\alpha,\beta>0$ does not imply any upper bounds on positive moments of ξ.

The main result of the section is the following theorem:

Theorem 3.1. *For any real $\beta>0$ and $\delta>1$ there are $u,v>0$ and $N_0\in\mathbb{N}$ depending only on β and δ with the following property: Let $N,n\in\mathbb{N}$ satisfy $N\ge\max(N_0,\delta n)$; $A=(a_{ij})$ be an $N\times n$ random matrix with i.i.d. entries, such that for some $\alpha>0$ the concentration function of the entries satisfies*

$$\mathcal{Q}(a_{11},\alpha)\le 1-\beta. \tag{3.1}$$

Then for any non-random $N\times n$ matrix B we have

$$\mathbb{P}\big\{s_{\min}(A+B)\le\alpha u\sqrt{N}\big\}\le\exp(-vN). \tag{3.2}$$

Adding the non-random component B in the theorem does not increase complexity of the proof; on the other hand, it demonstrates “shift-invariance” of the lower estimate. Note that the problem of estimating the smallest singular value of non-random shifts of *square* matrices is important in the analysis of algorithms [96, 105, 112, 113].

It is easy to see that a restriction of type (3.1) is *necessary* for (3.2) to hold. Indeed, suppose that for some $N\times n$ matrix A with i.i.d. entries and some numbers $u,v,\alpha>0$, (3.2) is true whenever $B = \lambda \mathbf{1_{N\times n}}$, $\lambda \in \mathbb{R}$. Then, obviously,

$$\mathbb{P}\Big\{\sum_{i=1}^{N}(a_{i1}-\lambda)^2 \le \alpha^2 u^2 N\Big\} \le \exp(-vN), \quad \lambda \in \mathbb{R},$$

implying $\mathcal{Q}(a_{11}, \alpha u) = \sup_{\lambda\in\mathbb{R}} \mathbb{P}\big\{|a_{11}-\lambda| \le \alpha u\big\} \le \exp(-v)$.

Our proof of Theorem 3.1 is based on two key elements: on a modification of a standard covering ("ε-net") argument for matrices (Proposition 3.3) and on estimates of the distance between a random vector and a fixed linear subspace that follow from a result of [92] (Theorem 3.4 and Corollary 3.6 of the section). Our method is similar in many aspects to the approach developed in [66] and later in [93], [95]. In particular, as in the mentioned papers, we decompose the unit sphere S^{n-1} into several subsets which are studied separately from one another. On the other hand, our modification of the covering argument and its technical realization in regard to splitting a random matrix into "regular" and "non-regular" parts are apparently new.

We will discuss the main idea of the proof more concretely and in more detail at the end of §3.1.2, after we state the modified covering argument.

3.1.2 Preliminaries

Throughout the section, $(\Omega, \Sigma, \mathbb{P})$ denotes a probability space. For an $N\times n$ matrix D and a set $K \subset \mathbb{R}^n$, $D(K)$ is the image of K in $\mathbb{R}^N$ under the action of D. Further, $\mathrm{col}_j(D)$ is the j-th column of D and $\mathrm{span}D$ is the linear span of columns of D in $\mathbb{R}^N$. The $N\times n$ matrix of ones is denoted by $\mathbf{1_{N\times n}}$. For a linear subspace $E\subset\mathbb{R}^n$, $E^\perp$ is the orthogonal complement of E in $\mathbb{R}^n$. In the special case when E is the linear span of a subset $\{e_j^n\}_{j\in J}$ ($J\subset\{1,2,\dots,n\}$) of the standard unit basis in $\mathbb{R}^n$, we will often write $x\chi_J$ in place of $\mathrm{Proj}_E(x)$. By $\mathrm{d}(\cdot,\cdot)$ we denote the standard Euclidean metric in $\mathbb{R}^N$, and set

$$\mathrm{d}(K_1,K_2) := \inf_{y_1\in K_1, y_2\in K_2} \mathrm{d}(y_1,y_2)$$

for any two subsets $K_1, K_2 \subset \mathbb{R}^N$.

In the section, we define many universal constants and functions that are frequently referred to later in the text. For convenience, we add to the name of every such constant or function a subscript indicating the statement where it was defined. For example, $C_{3.12}$ is the universal constant from Lemma 3.12, etc.

Let K be a subset of $\mathbb{R}^n$ and let $\varepsilon \in (0,1]$. A subset $\mathcal{N}\subset K$ is an *ε-net* for K (with respect to the standard Euclidean metric) if for any $y\in K$ there is $y'\in\mathcal{N}$ with

$\|y - y'\|_2 \leq \varepsilon$. We will use a well-known fact that any subset $K \subset B_2^n$ admits an ε-net $\mathcal{N}$ for K with cardinality $|\mathcal{N}| \leq (3/\varepsilon)^n$.

Given an ε-net $\mathcal{N}$ for S^{n-1}, the matrix $A + B$ from Theorem 3.1 trivially satisfies $s_{\min}(A + B) \geq \min\limits_{y' \in \mathcal{N}} \|Ay' + By'\|_2 - \varepsilon\|A + B\|_2$. This standard ε-net argument is not applicable in our setting as $A + B$ may have a very large norm with a large probability. A modification of the method in such a way that $\|A + B\|_2$ does not participate in the estimate for $s_{\min}(A + B)$ is an important element of our proof. Here we provide a "non-probabilistic" form of the argument. Given a non-random $N \times n$ matrix D, we shall represent it as a sum of two matrices D_1 and D_2; then we are able to estimate $s_{\min}(D)$ from below in terms of the norm $\|D_1\|_2$ of the "regular part" of the matrix D and distances between certain vectors and subspaces in $\mathbb{R}^N$ (determined by matrices D_1 and D_2). We start with a simpler version of the argument:

Lemma 3.2. *Let $N, n \in \mathbb{N}$, $h, \varepsilon > 0$ and let D_1, D_2, D be $N \times n$ (non-random) matrices with $D = D_1 + D_2$. Further, let $\mathcal{N}$ be an ε-net on S^{n-1} such that for any $y' \in \mathcal{N}$ we have*

$$\mathrm{d}\big(D_1 y', \mathrm{span} D_2\big) \geq h.$$

Then

$$s_{\min}(D) \geq \inf_{y \in \mathrm{S}^{n-1}} \mathrm{d}\big(D_1 y, \mathrm{span} D_2\big) \geq h - \varepsilon\|D_1\|_2.$$

Proof. Choose any $y \in \mathrm{S}^{n-1}$ and $y' \in \mathcal{N}$ such that $\|y - y'\|_2 \leq \varepsilon$. Then

$$\|Dy\|_2 = \big\|D_1 y + D_2 y\big\|_2 \geq \mathrm{d}\big(D_1 y, \mathrm{span} D_2\big) \geq \mathrm{d}\big(D_1 y', \mathrm{span} D_2\big) - \varepsilon\|D_1\|_2 \geq h - \varepsilon\|D_1\|_2.$$

By taking the infimum over all $y \in \mathrm{S}^{n-1}$, we obtain the result. □

Note that Lemma 3.2 cannot be used to handle matrices with the aspect ratio less than 2. Indeed, the lower estimate $s_{\min}(D) \geq \inf\limits_{y \in \mathrm{S}^{n-1}} \mathrm{d}\big(D_1 y, \mathrm{span} D_2\big)$ is non-trivial only if $\mathrm{span} D_1 \cap \mathrm{span} D_2 = 0$, which is not true when $N < 2n$ and both D_1 and D_2 have full rank. The following strengthening of Lemma 3.2 resolves the problem:

Proposition 3.3. *Let $N, n \in \mathbb{N}$, $S \subset \mathrm{S}^{n-1}$ and let D_1, D_2, D be $N \times n$ (non-random) matrices with $D = D_1 + D_2$. Further, suppose that numbers $h, \varepsilon > 0$, a subset $\mathcal{N} \subset \mathbb{R}^n$ and a collection of linear subspaces $\{E_{y'} \subset \mathbb{R}^n : y' \in \mathcal{N}\}$ satisfy the following three conditions:*

1) $y' \in E_{y'}$ for all $y' \in \mathcal{N}$;

2) for any $y' \in \mathcal{N}$ we have

$$\mathrm{d}\big(D_1 y', D(E_{y'}^{\perp}) + D_2(E_{y'})\big) \geq h; \tag{3.3}$$

3) for any $y \in S$ there is $y' \in \mathcal{N}$ such that $\|\mathrm{Proj}_{E_{y'}}(y) - y'\|_2 \le \varepsilon$.

Then

$$\inf_{y \in S} \|Dy\|_2 \ge h - \varepsilon \|D_1\|_2.$$

Proof. Take any $y \in S$ and let $y' \in \mathcal{N}$ be such that $\|\mathrm{Proj}_{E_{y'}}(y) - y'\|_2 \le \varepsilon$. Then

$$\begin{aligned}
\|Dy\|_2 &= \big\| D_1(\mathrm{Proj}_{E_{y'}}(y)) + \big(D(\mathrm{Proj}_{E_{y'}^\perp}(y)) + D_2(\mathrm{Proj}_{E_{y'}}(y))\big)\big\|_2 \\
&\ge \mathrm{d}\big(D_1(\mathrm{Proj}_{E_{y'}}(y)), D(E_{y'}^\perp) + D_2(E_{y'})\big) \\
&\ge \mathrm{d}\big(D_1 y', D(E_{y'}^\perp) + D_2(E_{y'})\big) - \varepsilon \|D_1\|_2 \\
&\ge h - \varepsilon \|D_1\|_2.
\end{aligned}$$

Taking the infimum over S, we get the result. □

To apply Proposition 3.3 we need an estimate for the distance between a random vector in $\mathbb{R}^N$ with independent coordinates and a fixed linear subspace. For any random vector X in $\mathbb{R}^N$ define the concentration function of X by

$$\mathcal{Q}(X, h) = \sup_{\lambda \in \mathbb{R}^N} \mathbb{P}\big\{ \|X - \lambda\|_2 \le h \big\}, \quad h \ge 0.$$

Note that for $N = 1$ the above definition is consistent with that given in the introduction. The following result is proved by M. Rudelson and R. Vershynin in [92]:

Theorem 3.4 ([92])**.** *Let $X = (X_1, X_2, \dots, X_m)$ be a random vector in $\mathbb{R}^m$ with independent coordinates such that*

$$\mathcal{Q}(X_i, h) \le \eta, \quad i = 1, 2, \dots, m$$

for some $h > 0, \eta \in (0,1)$. Then for any $d \in \{1, 2, \dots, m\}$ and any d-dimensional fixed subspace $E \subset \mathbb{R}^m$ we have

$$\mathcal{Q}(\mathrm{Proj}_E X, h\sqrt{d}) \le (C_{3.4}\eta)^d,$$

where $C_{3.4} > 0$ is a (sufficiently large) universal constant.

This theorem gives a nontrivial estimate for the concentration only for η sufficiently close to zero. Below, we provide an elementary extension of this result covering the case of "more concentrated" coordinates. First, let us recall a theorem of B. Rogozin:

Theorem 3.5 ([90]). *Let $k \in \mathbb{N}$, $\xi_1, \xi_2, \dots, \xi_k$ be independent random variables and let h_1, $h_2, \dots,$ $h_k > 0$ be some real numbers. Then for any $h \geq \max\limits_{j=1,2,\dots,k} h_j$,*

$$\mathcal{Q}\Big(\sum_{j=1}^{k} \xi_j, h\Big) \leq C_{3.5} h \Big(\sum_{j=1}^{k} \big(1 - \mathcal{Q}(\xi_j, h_j)\big) h_j^2\Big)^{-1/2},$$

where $C_{3.5} > 0$ is a universal constant.

Now, an easy application of Theorems 3.4 and 3.5 gives

Corollary 3.6. *Let $X = (X_1, X_2, \dots, X_m)$ be a random vector with independent coordinates such that*

$$\mathcal{Q}(X_i, h) \leq 1 - \tau, \quad i = 1, 2, \dots, m$$

for some $h > 0, \tau \in (0, 1)$. Then for any $d \in \{1, 2, \dots, m\}$, $\ell \in \mathbb{N}$ and any d-dimensional non-random subspace $E \subset \mathbb{R}^m$ the concentration function of $\mathrm{Proj}_E X$ satisfies

$$\mathcal{Q}(\mathrm{Proj}_E X, h\sqrt{d}/\ell) \leq \big(C_{3.4} C_{3.5}/\sqrt{\ell\tau}\big)^{d/\ell}.$$

Proof. Let $X^1, X^2, \dots, X^\ell$ be independent copies of X and $S = (S_1, S_2, \dots, S_m) = \sum\limits_{j=1}^{\ell} X^j$. In view of the condition on the coordinates of X and Theorem 3.5, we obtain

$$\mathcal{Q}(S_i, h) \leq C_{3.5}\Big(\ell\big(1 - \mathcal{Q}(X_i, h)\big)\Big)^{-1/2} \leq \frac{C_{3.5}}{\sqrt{\ell\tau}}, \quad i = 1, 2, \dots, m.$$

Then Theorem 3.4 gives

$$\mathcal{Q}(\mathrm{Proj}_E S, h\sqrt{d}) \leq \big(C_{3.4} C_{3.5}/\sqrt{\ell\tau}\big)^d.$$

On the other hand, the definition of S together with the triangle inequality implies that

$$\mathcal{Q}(\mathrm{Proj}_E X, h\sqrt{d}/\ell)^\ell \leq \mathcal{Q}(\mathrm{Proj}_E S, h\sqrt{d}),$$

and the proof is complete. □

Remark 3.1. Note that for any non-zero τ we can choose $\ell \in \mathbb{N}$ such that the upper estimate for the concentration function provided by Corollary 3.6 is non-trivial (strictly less than 1). In fact, a slightly weaker version of Corollary 3.6 still sufficient for our purposes could be proved using the original result of P. Lévy from [64] instead of Theorem 3.5.

As an immediate application of Corollary 3.6, we prove a statement about *peaky* vectors. We call a vector $y \in \mathrm{S}^{n-1}$ *θ-peaky* for some $\theta > 0$ if $\|y\|_\infty \geq \theta$. The set of all θ-peaky unit vectors in $\mathbb{R}^n$ shall be denoted by $\mathrm{S}^{n-1}_p(\theta)$.

Proposition 3.7 (Peaky vectors)**.** *Let $\delta > 1$ and let $n, N \in \mathbb{N}$ satisfy $N \geq \delta n$. Further, assume we are given $\theta, \gamma > 0$ and let $U = (u_{ij})$ be an $N \times n$ random matrix with independent entries (not necessarily identically distributed), each entry u_{ij} satisfying*

$$\mathcal{Q}(u_{ij}, 1) \leq 1 - \gamma.$$

Then

$$\mathbb{P}\Big\{\inf_{y \in \mathrm{S}^{n-1}_p(\theta)} \|Uy\|_2 \leq h_{3.7}\theta\sqrt{N}\Big\} \leq n \exp(-w_{3.7}N),$$

where the $h_{3.7}, w_{3.7} > 0$ depend only on γ and δ.

Proof. By Corollary 3.6, for $d = N - n + 1$, any $\ell \in \mathbb{N}$ and any fixed $(n-1)$-dimensional subspace $F \subset \mathbb{R}^N$ we have

$$\begin{aligned}\mathbb{P}\big\{\mathrm{d}(\mathrm{col}_j(U), F) \leq \sqrt{d}/\ell\big\} &\leq \mathcal{Q}\big(\mathrm{Proj}_{F^\perp}(\mathrm{col}_j(U)), \sqrt{d}/\ell\big) \\ &\leq \big(C_{3.4}C_{3.5}/\sqrt{\ell\gamma}\big)^{d/\ell}, \quad j = 1, 2, \dots, n.\end{aligned}$$

Take $\ell := \lceil 4C_{3.4}^2 C_{3.5}^2/\gamma \rceil$. Since for each $j = 1, 2, \dots, n$, $\mathrm{col}_j(U)$ is independent from the span of the other columns of U, from the above estimate we obtain

$$\mathbb{P}\big\{\mathrm{d}\big(\mathrm{col}_j(U), \mathrm{span}\{\mathrm{col}_k(U)\}_{k \neq j}\big) \leq h\sqrt{d}\big\} \leq \exp\big(-wd\big), \quad j = 1, 2, \dots, n$$

for some $h, w > 0$ depending only on γ. Let

$$\mathcal{E} = \big\{\omega \in \Omega : \mathrm{d}\big(\mathrm{col}_j(U(\omega)), \mathrm{span}\{\mathrm{col}_k(U(\omega))\}_{k \neq j}\big) > h\sqrt{d} \text{ for all } j = 1, 2, \dots, n\big\}.$$

Then $\mathbb{P}(\mathcal{E}) \geq 1 - n\exp(-wd)$. Take arbitrary $\omega \in \mathcal{E}$. For any $y = (y_1, y_2, \dots, y_n)$ in $\mathrm{S}^{n-1}_p(\theta)$ there is $j = j(y)$ such that $|y_j| \geq \theta$, hence

$$\begin{aligned}\|U(\omega)y\|_2 &= \|U(\omega)(y_j e_j^n) + U(\omega)(y - y_j e_j^n)\|_2 \\ &\geq \theta \mathrm{d}\big(\mathrm{col}_j(U(\omega)), \mathrm{span}\{\mathrm{col}_k(U(\omega))\}_{k \neq j}\big) \\ &> h\theta\sqrt{d}.\end{aligned}$$

Thus,

$$\mathbb{P}\Big\{\inf_{y \in \mathrm{S}^{n-1}_p(\theta)} \|Uy\|_2 \leq h\theta\sqrt{d}\Big\} \leq n\exp(-wd),$$

and the statement follows. □

Next, we introduce two notions important for us that will be used throughout the rest of the text. For any number $s \in \mathbb{R}$ and any Borel subset $H \subset \mathbb{R}$, define *the H-part of s* as

$$s_H = \begin{cases} s, & \text{if } s \in H, \\ 0, & \text{otherwise.} \end{cases}$$

The "complementary" $\mathbb{R}\backslash H$-part of s will be denoted by $s_{\overline{H}}$. Obviously, $s = s_H + s_{\overline{H}}$. The name and the notation resemble the *positive* and *negative* part of a real number; in fact $s_+ = s_H$ for $H = [0, \infty)$. For a real-valued random variable ξ we define the H-part of ξ pointwise: $\xi_H(\omega) = \xi(\omega)_H$ for all $\omega \in \Omega$. When a variable has a subscript, we will use parentheses to separate the subscript from the H-part notation, for example $(\xi_1)_H$ is the H-part of a random variable ξ_1. Given a matrix $A = (a_{ij})$, its H-part A_H is defined entry-wise, i.e. $(A_H)_{ij} = (a_{ij})_H$ for all admissible i, j.

For any $N \times n$ matrices M, M' (whether random or not), a Borel set $H \subset \mathbb{R}$ and a linear subspace $E \subset \mathbb{R}^n$ let

$$V_{M,M'}(H, E) := (M + M')(E^\perp) + (M_{\overline{H}} + M')(E).$$

Note that $V_{M,M'}(H, E)$ is a linear subspace of $\mathbb{R}^N$ of dimension at most n. When the matrices M, M' are clear from the context, we shall write $V(H, E)$ in place of $V_{M,M'}(H, E)$. When one or both matrices M, M' are random, $V_{M,M'}(H, E)$ is a *random subspace in* $\mathbb{R}^N$ of dimension at most n.

Let us conclude the section by describing the main idea of the proof of Theorem 3.1. Let S be a subset of S^{n-1}. As we already noted before, the main obstacle in using the standard ε-net argument to get a lower estimate for $\inf\limits_{y \in S} \|Ay + By\|_2$ is the need to control the norm of the matrix $A + B$ which is not possible unless we impose strong restrictions on its entries. Proposition 3.3 provides a workaround: we represent $A + B$ as a sum of two random matrices, "regular" and "irregular", satisfying certain conditions, so that the lower bound for $\inf\limits_{y \in S} \|Ay + By\|_2$ involves the norm of only the "regular" matrix. The splitting shall be defined with help of the above concept of H-part. Namely, for some specially chosen $\lambda \in \mathbb{R}$ and $H \subset \mathbb{R}$ we define the "regular" part as $(A - \lambda \mathbf{1}_{\mathbf{N}\times\mathbf{n}})_H$ and the "irregular" as $A + B - (A - \lambda \mathbf{1}_{\mathbf{N}\times\mathbf{n}})_H$ (which is identical to $(A - \lambda \mathbf{1}_{\mathbf{N}\times\mathbf{n}})_{\overline{H}} + B + \lambda \mathbf{1}_{\mathbf{N}\times\mathbf{n}}$). The set H shall be bounded which implies boundedness of the entries of $(A - \lambda \mathbf{1}_{\mathbf{N}\times\mathbf{n}})_H$. This, together with the appropriately chosen "shift" λ, allows us to easily control $\|(A - \lambda \mathbf{1}_{\mathbf{N}\times\mathbf{n}})_H\|_2$ from above. We will define H as the union of two specially constructed closed intervals on $\mathbb{R}$. The choice of H depends on the set S and may depend on the characteristics of the distribution of the entries of A (we leave this problem for the last section).

The crucial property that our set H shall satisfy is: letting $\widetilde{A} = A - \lambda\mathbf{1_{N\times n}}$ and $\widetilde{B} = B + \lambda\mathbf{1_{N\times n}}$, for certain finite subset of vectors $\mathcal{N} \subset \mathbb{R}^n$ and a collection of linear subspaces $\{E_{y'} \subset \mathbb{R}^n\}_{y'\in\mathcal{N}}$ (see Proposition 3.3) we have

$$\min_{y'\in\mathcal{N}} \mathrm{d}\big(\widetilde{A}_H y', V_{\widetilde{A},\widetilde{B}}(H, E_{y'})\big) \gtrsim \sqrt{N}$$

with a large probability. This restriction on H naturally corresponds to the condition (3.3) in Proposition 3.3. In practice we shall verify this property of H by proving that for every vector $y \in B_2^n$ satisfying certain upper bounds on $\|y\|_\infty$ and lower bounds on $\|y\|_2$ and for $E = \mathrm{span}\{e_j^n\}_{j\in\mathrm{supp}\, y}$, the distance $\mathrm{d}\big(\widetilde{A}_H y, V_{\widetilde{A},\widetilde{B}}(H, E)\big)$ is large with an overwhelming probability. This condition demands a "rich" structure from $\widetilde{A}_H$; consequently, the set H cannot be very small in diameter. On the other hand, the "upper" restrictions on H are dictated by the necessity to control the norm of $\widetilde{A}_H$. Thus, we have to find a balance between the two requirements.

In order to estimate the distance between the random vector $\widetilde{A}_H y$ and the random subspace $V_{\widetilde{A},\widetilde{B}}(H, E)$, we will use Corollary 3.6. However, since in general $V_{\widetilde{A},\widetilde{B}}(H, E)$ is *dependent* (in probabilistic sense) on $\widetilde{A}_H y$, an immediate application of the corollary is not possible; instead, we will combine it with a conditioning argument, which is presented in the next section.

3.1.3 The distribution of $\mathrm{d}\big(A_H y, V_{A,B}(H, E)\big)$

Assume that we are given $\delta > 1$, $N, n \in \mathbb{N}$ with $N \geq \delta n$, a random $N \times n$ matrix A with i.i.d. entries, a non-random $N \times n$ matrix B and a Borel subset $H \subset \mathbb{R}$ with $\mathbb{P}\{a_{11} \in H\} > 0$. The purpose of this section is to study the distribution of the distance between a random vector $A_H y$ and the random subspace $V_{A,B}(H, E) = (A + B)(E^\perp) + (A_{\overline{H}} + B)(E)$, where $E = \mathrm{span}\{e_j^n\}_{j\in\mathrm{supp}\, y}$. We give *sufficient* conditions on A, H and y which guarantee that $\mathrm{d}\big(A_H y, V_{A,B}(H, E)\big)$ is large with a large probability (Proposition 3.11). Note that generally $A_H y$ and $V_{A,B}(H, E)$ are *dependent*. In order to overcome this problem, we apply a decoupling argument.

We adopt the following notation: For any subset $W \subset \{1, 2, \ldots, N\} \times \{1, 2, \ldots, n\}$ let

$$\Omega_W = \big\{\omega \in \Omega :\ a_{ij}(\omega) \in H \text{ for all } (i,j) \in W \textbf{ and } a_{ij}(\omega) \in \overline{H} \text{ for all } (i,j) \notin W\big\}.$$

Given an event $\mathcal{E} \subset \Omega$ with $\mathbb{P}(\mathcal{E}) > 0$, we denote by $(\mathcal{E}, \Sigma_{\mathcal{E}}, \mathbb{P}_{\mathcal{E}})$ the probability space where the σ-algebra $\Sigma_{\mathcal{E}}$ of subsets of $\mathcal{E}$ is naturally induced by the σ-algebra Σ on Ω, and $\mathbb{P}_{\mathcal{E}}$ is defined by $\mathbb{P}_{\mathcal{E}}(K) = \mathbb{P}(\mathcal{E})^{-1}\mathbb{P}(K)$ $(K \in \Sigma_{\mathcal{E}})$.

Lemma 3.8 (Conditional independence)**.** *Let A, B and H be as above, $y \in \mathbb{R}^n$, $E = \operatorname{span}\{e_j^n\}_{j \in \operatorname{supp} y}$ and let $W \subset \{1, 2, \ldots, N\} \times \{1, 2, \ldots, n\}$ be such that $\mathbb{P}(\Omega_W) > 0$. Then the random vector $A_H y$ in $\mathbb{R}^N$ and the random subspace $V_{A,B}(H, E) \subset \mathbb{R}^N$ are conditionally independent given event Ω_W. Moreover, the coordinates of $A_H y$ are conditionally independent given Ω_W.*

The proof of the lemma is quite straightforward, so we omit it. Lemma 3.8 shows that Corollary 3.6 can be applied to $A_H y$ and the subspace $V_{A,B}(H, E)$ "inside" each Ω_W. Hence, to give a satisfactory lower estimate for $\mathrm{d}\big(A_H y, V_{A,B}(H, E)\big)$ on the entire Ω, it is enough to verify that there is a subset $M \subset 2^{\{1,2,\ldots,N\}\times\{1,2,\ldots,n\}}$ such that the $\mathbb{P}$-measure of the union of Ω_W's $(W \in M)$ is close to 1 and for each $W \in M$, the restriction of the vector $A_H y$ to Ω_W has sufficiently "spread" coordinates. Of course, such a set M may exist only under certain assumptions on A, H and y. In Lemma 3.9, we formulate those assumptions using random variables that agree on a part of the probability space and are independent when restricted to the other part of Ω. Let us remark that, whereas the use of such variables has some advantages (in our opinion), it should not be regarded as a necessary ingredient of the proof.

Let ξ, ξ' be two random variables such that $\mathbb{P}\{\xi \in H\} > 0$. We say that ξ, ξ' *are conditionally i.i.d. given event $\{\omega \in \Omega : \xi(\omega) \in H\}$ and identical on $\{\omega \in \Omega : \xi(\omega) \in \overline{H}\}$* if the following is true: setting $\mathcal{E} = \{\omega \in \Omega : \xi(\omega) \in H\}$, the restrictions of ξ, ξ' to the probability space $(\mathcal{E}, \Sigma_{\mathcal{E}}, \mathbb{P}_{\mathcal{E}})$ are i.i.d. and $\xi(\omega) = \xi'(\omega)$ for $\omega \in \Omega \setminus \mathcal{E}$. The definition implies that ξ' has the same individual distribution (on Ω) as ξ and for any Borel subsets $K, K' \subset \mathbb{R}$

$$\mathbb{P}\big\{(\xi, \xi') \in K \times K'\big\} = \frac{\mathbb{P}\{\xi \in H \cap K\}\mathbb{P}\{\xi \in H \cap K'\}}{\mathbb{P}\{\xi \in H\}} + \mathbb{P}\{\xi \in \overline{H} \cap K \cap K'\}; \tag{3.4}$$

in particular, $\mathbb{P}\{(\xi, \xi') \in H \times \overline{H}\} = \mathbb{P}\{(\xi, \xi') \in \overline{H} \times H\} = 0$. Note that $\xi_{\overline{H}}$ and ${\xi'}_{\overline{H}}$ are equal a.s. on Ω. It is a trivial observation that $\xi_H - {\xi'}_H$ is symmetrically distributed.

For any event $\mathcal{E} \subset \Omega$ with $\mathbb{P}(\mathcal{E}) > 0$ and any random variable ξ on Ω, let $\mathcal{Q}_{\mathcal{E}}(\xi, \cdot)$ be the concentration function of the restriction of ξ to the probability space $(\mathcal{E}, \Sigma_{\mathcal{E}}, \mathbb{P}_{\mathcal{E}})$.

Lemma 3.9. *Let H be a Borel subset of $\mathbb{R}$; $N \geq \delta n$ for some $\delta > 1$ and let $A = (a_{ij})$ be an $N \times n$ random matrix with i.i.d. entries and $\mathbb{P}\{a_{11} \in H\} > 0$. Further, let $A' = (a'_{ij})$ be an $N \times n$ random matrix having the same distribution as A such that 2-dimensional vectors*

(a_{ij}, a'_{ij}) *($1 \le i \le N$, $1 \le j \le n$) are i.i.d. and for any admissible i and j the variables* a_{ij} *and* a'_{ij} *are conditionally i.i.d. given event* $\{\omega \in \Omega : a_{ij}(\omega) \in H\}$ *and identical on* $\{\omega \in \Omega : a_{ij}(\omega) \in \overline{H}\}$. *Let* $y = (y_1, y_2, \ldots, y_n) \in \mathbb{R}^n$ *and* $s > 0$ *be such that*

$$\mathbb{P}\Big\{\Big|\sum_{j=1}^{n}\big((a_{ij})_H - (a'_{ij})_H\big)y_j\Big| > s\Big\} \ge \delta^{-1/4}, \quad i = 1, 2, \ldots, N. \tag{3.5}$$

Define M *as the collection of all subsets* $W \subset \{1, 2, \ldots, N\} \times \{1, 2, \ldots, n\}$ *satisfying*

$$\mathbb{P}(\Omega_W) > 0 \ \text{ and } \ \Big|\Big\{i \in \{1, 2, \ldots, N\} : \mathcal{Q}_{\Omega_W}\Big(\sum_{j=1}^{n} (a_{ij})_H y_j, \frac{s}{2}\Big) \le 1 - \tau\Big\}\Big| \ge N\delta^{-1/2}$$

with $\tau = \frac{1}{2}\big(\delta^{-1/4} - \delta^{-1/3}\big)$. *Then*

$$\mathbb{P}\Big(\bigcup_{W \in M} \Omega_W\Big) \ge 1 - \exp(-w_{3.9}N),$$

where $w_{3.9} > 0$ *depends only on* δ.

Proof. For each $i = 1, 2, \ldots, N$ and $J \subset \{1, 2, \ldots, n\}$ let

$$\Omega^i_J = \big\{\omega \in \Omega : a_{ij}(\omega) \in H \text{ for all } j \in J \textbf{ and } a_{ij}(\omega) \in \overline{H} \text{ for all } j \notin J\big\},$$

and for $i = 1, 2, \ldots, N$ define

$$L_i = \Big\{J \subset \{1, 2, \ldots, n\} : \mathbb{P}(\Omega^i_J) > 0 \textbf{ and } \mathcal{Q}_{\Omega^i_J}\Big(\sum_{j=1}^{n} (a_{ij})_H y_j, \frac{s}{2}\Big) \le 1 - \tau\Big\}; \ \mathcal{E}_i = \bigcup_{J \in L_i} \Omega^i_J.$$

It is not difficult to see that the events $\mathcal{E}_i \subset \Omega$ ($i = 1, 2, \ldots, N$) are independent in view of independence of the entries of A.

Fix for a moment any $i \in \{1, 2, \ldots, N\}$. One can verify that for any $j \in \{1, 2, \ldots, n\}$ and $J \subset \{1, 2, \ldots, n\}$ the variables $(a_{ij})_H$ and $(a'_{ij})_H$ are i.i.d. given event Ω^i_J. It follows that

$$\sum_{j=1}^{n} (a_{ij})_H y_j \ \text{ and } \ \sum_{j=1}^{n} (a'_{ij})_H y_j \ \text{ are i.i.d. given } \Omega^i_J, \text{ for all } J \subset \{1, 2, \ldots, n\}. \tag{3.6}$$

Take any subset $J \subset \{1, 2, \ldots, n\}$ satisfying

$$\mathbb{P}(\Omega^i_J) > 0 \text{ and } \mathbb{P}_{\Omega^i_J}\Big\{\Big|\sum_{j=1}^{n}\big((a_{ij})_H - (a'_{ij})_H\big)y_j\Big| > s\Big\} \ge 2\tau. \tag{3.7}$$

For all $\lambda \in \mathbb{R}$ we have, in view of (3.6),

$$\begin{aligned}
&\mathbb{P}_{\Omega^i_J}\Big\{\lambda - \frac{s}{2} \le \sum_{j=1}^n (a_{ij})_H y_j \le \lambda + \frac{s}{2}\Big\}^2 \\
&\quad = \mathbb{P}_{\Omega^i_J}\Big\{\lambda - \frac{s}{2} \le \sum_{j=1}^n (a_{ij})_H y_j \le \lambda + \frac{s}{2} \textbf{ and } \lambda - \frac{s}{2} \le \sum_{j=1}^n (a'_{ij})_H y_j \le \lambda + \frac{s}{2}\Big\} \\
&\quad \le \mathbb{P}_{\Omega^i_J}\Big\{\Big|\sum_{j=1}^n \big((a_{ij})_H - (a'_{ij})_H\big) y_j\Big| \le s\Big\} \\
&\quad \le 1 - 2\tau,
\end{aligned}$$

implying

$$\mathcal{Q}_{\Omega^i_J}\Big(\sum_{j=1}^n (a_{ij})_H y_j, \frac{s}{2}\Big) \le \sqrt{1-2\tau} \le 1 - \tau.$$

Thus, any J satisfying (3.7) belongs to L_i. Clearly,

$$\mathbb{P}\Big\{\Big|\sum_{j=1}^n \big((a_{ij})_H - (a'_{ij})_H\big) y_j\Big| > s\Big\} = \sum_J \mathbb{P}_{\Omega^i_J}\Big\{\Big|\sum_{j=1}^n \big((a_{ij})_H - (a'_{ij})_H\big) y_j\Big| > s\Big\}\mathbb{P}(\Omega^i_J),$$

where the summation is taken over $J \subset \{1, 2, \ldots, n\}$ satisfying $\mathbb{P}(\Omega^i_J) > 0$. Hence, in view of (3.5) and the above observations we get

$$\begin{aligned}
\delta^{-1/4} &\le \sum_J \mathbb{P}_{\Omega^i_J}\Big\{\Big|\sum_{j=1}^n \big((a_{ij})_H - (a'_{ij})_H\big) y_j\Big| > s\Big\}\mathbb{P}(\Omega^i_J) \\
&\le \sum_{J \in L_i} \mathbb{P}(\Omega^i_J) + 2\tau \sum_{J \notin L_i} \mathbb{P}(\Omega^i_J) \\
&\le 2\tau + \mathbb{P}(\mathcal{E}_i),
\end{aligned}$$

implying $\mathbb{P}(\mathcal{E}_i) \ge \delta^{-1/3}$.

We have noted that the events $\mathcal{E}_i$ $(i = 1, 2, \ldots, N)$ are independent and $\mathbb{P}(\mathcal{E}_i) \ge \delta^{-1/3}$ for each i. Now, setting

$$\mathcal{E} = \Big\{\omega \in \Omega : \ \Big|\Big\{i \in \{1, 2, \ldots, N\} : \omega \in \mathcal{E}_i\Big\}\Big| \ge N\delta^{-1/2}\Big\},$$

we obtain by Bernstein's (or Hoeffding's) inequality $\mathbb{P}(\mathcal{E}) \ge 1 - \exp(-w_{3.9}N)$, where $w_{3.9} > 0$ depends only on δ. Finally, we will show that $\mathcal{E} \subset \bigcup_{W \in M} \Omega_W \cup \Omega^0$ for a set Ω^0 of zero probability measure. Define $\Omega^0 = \bigcup_W \Omega_W$, where the union is taken over all W such that $\mathbb{P}(\Omega_W) = 0$. Fix any $\omega \in \mathcal{E} \setminus \Omega^0$ and let $\widetilde{W} \subset \{1, 2, \ldots, N\} \times \{1, 2, \ldots, n\}$

be such that $\omega \in \Omega_{\widetilde{W}}$. In view of the definition of $\mathcal{E}$ and the events $\mathcal{E}_i$, there are indices $i_1 < i_2 < \cdots < i_k$ ($k \geq N\delta^{-1/2}$) such that $w \in \Omega^{i_q}_{J_q}$ for all $q = 1, 2, \ldots, k$, where $J_q = \{j : (i_q, j) \in \widetilde{W}\}$ and

$$\mathcal{Q}_{\Omega^{i_q}_{J_q}}\Big(\sum_{j=1}^{n} (a_{i_q j})_H y_j, \frac{s}{2}\Big) \leq 1 - \tau, \quad q = 1, 2, \ldots, k.$$

Note that, in view of independence of the entries of A and the relation between $\Omega_{\widetilde{W}}$ and $\Omega^{i_q}_{J_q}$, the conditional distribution of the sum $\sum_{j=1}^{n} (a_{i_q j})_H y_j$ given event $\Omega_{\widetilde{W}}$ is the same as its conditional distribution given $\Omega^{i_q}_{J_q}$. Hence,

$$\mathcal{Q}_{\Omega_{\widetilde{W}}}\Big(\sum_{j=1}^{n} (a_{i_q j})_H y_j, \frac{s}{2}\Big) = \mathcal{Q}_{\Omega^{i_q}_{J_q}}\Big(\sum_{j=1}^{n} (a_{i_q j})_H y_j, \frac{s}{2}\Big) \leq 1 - \tau, \quad q = 1, 2, \ldots, k.$$

The last formula implies that $\widetilde{W} \subset M$, so $\omega \in \bigcup_{W \in M} \Omega_W$. The proof is complete. □

Next, we combine the result of Lemma 3.9 with Corollary 3.6:

Lemma 3.10. *Let N, n, δ, H, A, A', y and s be exactly as in Lemma 3.9 and B be a non-random $N \times n$ matrix. Then*

$$\mathbb{P}\big\{\mathrm{d}\big(A_H y, V_{A,B}(H,E)\big) \leq s h_{3.10}\sqrt{N}\big\} \leq 2\exp\big(-w_{3.10} N\big),$$

where $E = \mathrm{span}\{e^n_j\}_{j \in \mathrm{supp}\, y}$ and $h_{3.10} > 0$, $w_{3.10} > 0$ depend only on δ.

Proof. Let M and τ be defined as in Lemma 3.9 and take any $W \in M$. Let

$$m = \Big|\Big\{i \in \{1, 2, \ldots, N\} : \mathcal{Q}_{\Omega_W}\Big(\sum_{j=1}^{n} (a_{ij})_H y_j, \frac{s}{2}\Big) \leq 1 - \tau\Big\}\Big|.$$

By the definition of M, we have $m \geq N\delta^{-1/2} \geq \sqrt{\delta} n$, hence, taking $d = m - n$ and $\ell = 4(C_{3.4}C_{3.5})^2/\tau$, by Corollary 3.6, for $\kappa = \delta^{-1/2} - \delta^{-1}$ and any fixed n-dimensional subspace $F \subset \mathbb{R}^N$ we obtain

$$\mathbb{P}_{\Omega_W}\Big\{\mathrm{d}\big(A_H y, F\big) \leq \frac{s}{2\ell}\sqrt{\kappa N}\Big\} \leq 2^{-\kappa N/\ell}.$$

Now, consider the random subspace $V_{A,B}(H,E) = (A+B)(E^\perp) + (A_{\overline{H}} + B)(E)$. Let us remark that $(A+B)(E^\perp)$ is just the linear span of columns of $A+B$ whose indices do not belong to the support of y, and, similarly, $(A_{\overline{H}} + B)(E)$ is the span of those columns of $A_{\overline{H}} + B$ whose indices belong to the support of y. By Lemma 3.8, $V_{A,B}(H,E)$

and the vector $A_H y$ are conditionally independent given Ω_W, hence the above estimate immediately implies

$$\mathbb{P}_{\Omega_W}\Big\{\mathrm{d}\big(A_H y, V_{A,B}(H,E)\big) \le \frac{s}{2\ell}\sqrt{\kappa N}\Big\} \le 2^{-\kappa N/\ell}.$$

Since the relation holds for all $W \in M$, in view of Lemma 3.9 we obtain

$$\begin{aligned}\mathbb{P}\Big\{\mathrm{d}\big(A_H y, V_{A,B}(H,E)\big) \le \frac{s}{2\ell}\sqrt{\kappa N}\Big\} &\le 2^{-\kappa N/\ell}\,\mathbb{P}\Big(\bigcup_{W\in M}\Omega_W\Big) + 1 - \mathbb{P}\Big(\bigcup_{W\in M}\Omega_W\Big)\\ &\le 2^{-\kappa N/\ell} + \exp(-w_{3.9}N),\end{aligned}$$

and the result follows. □

Finally, we can prove the main result of the section:

Proposition 3.11. *Let $\delta > 1$, $n, N \in \mathbb{N}$, $N \ge \delta n$ and let $A = (a_{ij})$ be an $N \times n$ random matrix with i.i.d. entries and B be any non-random $N \times n$ matrix. Further, for some $d, r > 0$ let H be a Borel subset of $\mathbb{R}$ such that $H = H_1 \cup H_2$ for disjoint Borel sets H_1, H_2 with $\mathrm{d}(H_1, H_2) \ge d$ and $\min\big(\mathbb{P}\{a_{11} \in H_1\}, \mathbb{P}\{a_{11} \in H_2\}\big) \ge r$. For arbitrary $t > 0$ define*

$$h_{3.11} = \frac{1-\delta^{-1/4}}{C_{3.5}}\sqrt{\frac{r}{8}}td$$

and let $y \in \mathbb{R}^n$ be a vector satisfying $\|y\|_2 \ge t$, $\|y\|_\infty \le \frac{2h_{3.11}}{d}$ and $E = \mathrm{span}\{e_j^n\}_{j\in \mathrm{supp}\, y}$. Then

$$\mathbb{P}\big\{\mathrm{d}\big(A_H y, V_{A,B}(H,E)\big) \le h_{3.10}h_{3.11}\sqrt{N}\big\} \le 2\exp(-w_{3.10}N).$$

Proof. Let $A' = (a'_{ij})$ be an $N \times n$ random matrix having the same distribution as A such that 2-dimensional vectors (a_{ij}, a'_{ij}) $(1 \le i \le N,\ 1 \le j \le n)$ are i.i.d. and for any admissible i and j the variables a_{ij} and a'_{ij} are conditionally i.i.d. given event $\{\omega \in \Omega : a_{ij}(\omega) \in H\}$ and identical on $\{\omega \in \Omega : a_{ij}(\omega) \in \overline{H}\}$. For every $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, n$, in view of formula (3.4) for the joint distribution we get

$$\mathbb{P}\big\{\big|(a_{ij})_H - (a'_{ij})_H\big| \ge d\big\} \ge \mathbb{P}\{a_{ij} \in H_1 \textbf{ and } a'_{ij} \in H_2\} + \mathbb{P}\{a_{ij} \in H_2 \textbf{ and } a'_{ij} \in H_1\} \ge r.$$

Since $(a_{ij})_H - (a'_{ij})_H$ is symmetrically distributed, the above relation implies $\mathcal{Q}\big((a_{ij})_H - (a'_{ij})_H, \frac{d}{2}\big) \le 1 - \frac{r}{2}$. Clearly, $h_{3.11} \ge \frac{d|y_j|}{2}$ for every coordinate y_j of the vector y, hence by

Theorem 3.5 for all $i = 1, 2, \ldots, N$ we have

$$\begin{aligned}
&\mathbb{P}\Big\{\Big|\sum_{j=1}^{n}\big((a_{ij})_H - (a'_{ij})_H\big)y_j\Big| \le h_{3.11}\Big\}\\
&\le \mathcal{Q}\Big(\sum_{j=1}^{n}\big((a_{ij})_H - (a'_{ij})_H\big)y_j, h_{3.11}\Big)\\
&\le C_{3.5}h_{3.11}\Big(\frac{1}{4}\sum_{j=1}^{n}\Big(1-\mathcal{Q}\Big(\big((a_{ij})_H - (a'_{ij})_H\big)y_j, \frac{|y_j|d}{2}\Big)\Big)(y_jd)^2\Big)^{-1/2}\\
&\le C_{3.5}h_{3.11}\Big(\frac{r}{8}\sum_{j=1}^{n}(y_jd)^2\Big)^{-1/2}\\
&\le \frac{C_{3.5}h_{3.11}}{td}\sqrt{\frac{8}{r}} = 1-\delta^{-1/4}.
\end{aligned}$$

Thus, vector y satisfies condition (3.5) with $s := h_{3.11}$. Then, by Lemma 3.10,

$$\mathbb{P}\big\{\mathrm{d}\big(A_Hy, V_{A,B}(H,E)\big) \le h_{3.10}h_{3.11}\sqrt{N}\big\} \le 2\exp(-w_{3.10}N).$$

□

3.1.4 Decomposition of S^{n-1} and proof of Theorem 3.1

Recall that in Section 3.1.2 we defined $\mathrm{S}_p^{n-1}(\theta)$ as the set of θ-*peaky* vectors, that is, unit vectors in $\mathbb{R}^n$ whose ℓ_∞^n-norm is at least θ. We say that a vector $y \in \mathrm{S}^{n-1}$ is m-*sparse* if $|\mathrm{supp}\, y| \le m$. Next, $y \in \mathrm{S}^{n-1}$ is *almost* m-*sparse*, if there is a subset $J \subset \{1, 2, \ldots, n\}$ of cardinality at most m, such that $\|y\chi_J\|_2 \ge 1/2$. The set of all almost m-sparse vectors shall be denoted by $\mathrm{S}_a^{n-1}(m)$.

In our proof of Theorem 3.1, we represent S^{n-1} as the union of three subsets:

$$\mathrm{S}^{n-1} = \mathrm{S}_p^{n-1}(\theta) \cup \big(\mathrm{S}_a^{n-1}(\sqrt{N}) \setminus \mathrm{S}_p^{n-1}(\theta)\big) \cup \big(\mathrm{S}^{n-1} \setminus \mathrm{S}_a^{n-1}(\sqrt{N})\big),$$

where θ is a function of the parameters β and δ of the theorem. Then the smallest singular value of $A + B$ can be estimated by bounding separately $\inf_y \|Ay + By\|_2$ over each of the three subsets.

The reasons for such a representation of S^{n-1} are purely technical: Proposition 3.11 proved in the previous section handles vectors with a sufficiently small ℓ_∞^n-norm, so instead we use Proposition 3.7 to deal with the set $\mathrm{S}_p^{n-1}(\theta)$. Further, the separate treatment of almost $\sqrt{N}$-sparse vectors is convenient because, on the one hand, the construction of the set H corresponding to $\mathrm{S}_a^{n-1}(\sqrt{N})\setminus\mathrm{S}_p^{n-1}(\theta)$ is trivial compared to $\mathrm{S}^{n-1}\setminus\mathrm{S}_a^{n-1}(\sqrt{N})$; on the other hand, vectors from $\mathrm{S}^{n-1}\setminus\mathrm{S}_a^{n-1}(\sqrt{N})$ have a useful geometric property (Lemma 3.16) which the almost sparse vectors generally do not possess. We note that the set $\mathrm{S}_a^{n-1}(\sqrt{N})$ in the covering of S^{n-1} can be replaced with $\mathrm{S}_a^{n-1}(N^\kappa)\big)$ for any constant power $\kappa \in (0,1)$; this would only affect the constants in the final estimate.

In our representation of S^{n-1}, we follow an idea from [66], where the unit sphere was split into sets of "close to sparse" and "far from sparse" vectors. A similar splitting was also employed in [93], [95], where the terms "compressible" and "incompressible" were used instead. On the other hand, our "borderline" $\sqrt{N}$ is much smaller than in the mentioned papers.

The next elementary lemma shall be used in conjunction with Proposition 3.3.

Lemma 3.12. *There is a universal constant $C_{3.12} > 0$ with the following property: Let $n, m \in \mathbb{N}$ with $m \le n$, $\varepsilon \in (0,1]$, $S \subset \mathrm{S}^{n-1}$ and let $T \subset B_2^n$ consist of m-sparse vectors and satisfy*

$$\text{for any } y \in S \text{ there is } x = x(y) \in T \text{ with } y\chi_{\operatorname{supp} x} = x. \tag{3.8}$$

Then there is a finite set $\mathcal{N} \subset T$ of cardinality at most $\big(\frac{C_{3.12}n}{\varepsilon m}\big)^m$ such that for any $y \in S$ there is $y' = y'(y) \in \mathcal{N}$ with $\|y\chi_{\operatorname{supp} y'} - y'\|_2 \le \varepsilon$.

Proposition 3.13 (Vectors from $\mathrm{S}_a^{n-1}(\sqrt{N})$ with a small ℓ_∞^n-norm)**.** *For any $\gamma > 0$ and $\delta > 1$ there are $N_{3.13} \in \mathbb{N}$ and $h_{3.13} > 0$ depending only on γ and δ with the following property: Let*

$$\theta_{3.13} = \frac{1-\delta^{-1/4}}{C_{3.5}}\sqrt{\frac{\gamma}{8}},$$

$N \ge \max(N_{3.13}, \delta n)$, $z \in \mathbb{R}$ and let A be an $N \times n$ random matrix with i.i.d. entries such that

$$\min\big(\mathbb{P}\{z-\sqrt{N} \le a_{11} \le z-1\}, \mathbb{P}\{z+1 \le a_{11} \le z+\sqrt{N}\}\big) \ge \gamma.$$

Then for the set $S = \mathrm{S}_a^{n-1}(\sqrt{N}) \setminus S_p^{n-1}(\theta_{3.13})$ and any non-random $N \times n$ matrix B we have

$$\mathbb{P}\big\{\inf_{y\in S}\|Ay + By\|_2 \le h_{3.13}\sqrt{N}\big\} \le \exp(-w_{3.10}N/2).$$

Proof. Fix any $\gamma > 0$ and $\delta > 1$ and define $d := 2$, $r := \gamma$, $t := \frac{1}{2}$; let $h_{3.11}$ be as in Proposition 3.11 and $N_{3.13} = N_{3.13}(\gamma,\delta)$ be the smallest integer greater than $\frac{2}{h_{3.10}h_{3.11}}$ such that for all $N \ge N_{3.13}$

$$2\big(C_{3.12}N\big)^{3\sqrt{N}} \le \exp(w_{3.10}N/2).$$

Now, take any $n \in N$ and $N \ge \max(N_{3.13}, \delta n)$; let z and A safisfy conditions of the lemma and B be any non-random $N \times n$ matrix. We will assume that S is non-empty.

Without loss of generality, $z = 0$ (otherwise, we replace A, B with $A - z\mathbf{1_{N\times n}}$, $B + z\mathbf{1_{N\times n}}$). Define $H_1 = [-\sqrt{N}, -1]$, $H_2 = [1, \sqrt{N}]$, $H = H_1 \cup H_2$. Obviously, $\mathrm{d}(H_1, H_2) = d$ and $\min\bigl(\mathbb{P}\{a_{11} \in H_1\}, \mathbb{P}\{a_{11} \in H_2\}\bigr) \geq r$. Let $T \subset B_2^n$ be the set of $\sqrt{N}$-sparse vectors with the Euclidean norm at least $\frac{1}{2}$ and the maximal norm at most $\theta_{3.13}$. Clearly, T and S satisfy (3.8), hence, by Lemma 3.12, there is a finite subset $\mathcal{N} \subset T$ of cardinality at most $\bigl(C_{3.12}N\bigr)^{3\sqrt{N}}$ such that for any $y \in S$ there is $y' = y'(y) \in \mathcal{N}$ with $\|y\chi_{\operatorname{supp} y'} - y'\|_2 \leq N^{-2}$.

Let $E_{y'} = \operatorname{span}\{e_j^n\}_{j \in \operatorname{supp} y'}$ $(y' \in \mathcal{N})$ and define an event

$$\mathcal{E} = \bigl\{\omega \in \Omega : \mathrm{d}\bigl(A_H(\omega)y', V_{A,B}(H, E_{y'})(\omega)\bigr) > h_{3.10}h_{3.11}\sqrt{N} \text{ for all } y' \in \mathcal{N}\bigr\}.$$

In view of Proposition 3.11, the upper estimate for $|\mathcal{N}|$ and the definition of $N_{3.13}$, we get

$$\mathbb{P}(\mathcal{E}) \geq 1 - 2|\mathcal{N}| \exp(-w_{3.10}N) \geq 1 - \exp(-w_{3.10}N/2).$$

Take any $\omega \in \mathcal{E}$ and define $D_1 = A_H(\omega)$, $D_2 = A_{\overline{H}}(\omega) + B$, $D = D_1 + D_2$. Since all entries of D_1 are bounded by $\sqrt{N}$ by absolute value, we get $\|D_1\|_2 \leq N^{3/2}$; next, for every $y' \in \mathcal{N}$

$$\mathrm{d}\bigl(D_1 y', D(E_{y'}^{\perp}) + D_2(E_{y'})\bigr) > h_{3.10}h_{3.11}\sqrt{N}$$

(note that $D(E_{y'}^{\perp}) + D_2(E_{y'}) = V_{A,B}(H, E_{y'})(\omega)$). Hence, by Proposition 3.3, we obtain

$$\inf_{y \in S} \|Dy\|_2 > h_{3.10}h_{3.11}\sqrt{N} - N^{-1/2} \geq \frac{1}{2}h_{3.10}h_{3.11}\sqrt{N}.$$

Finally, applying the above argument to all $\omega \in \mathcal{E}$, we get the result. □

As we noted before, construction of the set H corresponding to $\mathrm{S}^{n-1} \setminus \mathrm{S}_a^{n-1}(\sqrt{N})$ is not so trivial as in the case of almost $\sqrt{N}$-sparse vectors. The reason is that in general the set $\mathrm{S}^{n-1} \setminus \mathrm{S}_a^{n-1}(\sqrt{N})$ is much larger than $\mathrm{S}_a^{n-1}(\sqrt{N})$, and we have to apply more delicate arguments to get a satisfactory probabilistic estimate. The construction of H for the set of "far from $\sqrt{N}$-sparse" vectors is contained in the following lemma:

Lemma 3.14. *Let ξ be a random variable such that for some $z \in \mathbb{R}$, $\gamma > 0$, $N \in \mathbb{N}$ we have*

$$\min\bigl(\mathbb{P}\bigl\{z - \sqrt{N} \leq \xi \leq z - 1\bigr\}, \mathbb{P}\bigl\{z + 1 \leq \xi \leq z + \sqrt{N}\bigr\}\bigr) \geq \gamma.$$

Then there exists an integer $\ell \in [0, \lfloor \log_2 \sqrt{N} \rfloor]$, $\lambda \in \mathbb{R}$ and disjoint Borel sets $H_1, H_2 \subset [-2^{\ell+2}; 2^{\ell+2}]$ such that $\mathrm{d}(H_1, H_2) \geq 2^{\ell}$, $\min\bigl(\mathbb{P}\{\xi - \lambda \in H_1\}, \mathbb{P}\{\xi - \lambda \in H_2\}\bigr) \geq c_{3.14}\gamma 2^{-\ell/8}$ and $\mathbb{E}\,(\xi - \lambda)_H = 0$ for $H = H_1 \cup H_2$ and a universal constant $c_{3.14} > 0$.

Proof. Without loss of generality we can assume that $z = 0$. Let $c_{3.14} = \Bigl(\sum\limits_{m=0}^{\infty} 2^{-m/8}\Bigr)^{-1}$.

Then, by the conditions on ξ, there are $\ell_1, \ell_2 \in \{0, 1, \ldots, \lfloor \log_2 \sqrt{N} \rfloor\}$ such that

$$\mathbb{P}\{\xi \in [-2^{\ell_1+1}, -2^{\ell_1}]\} \geq c_{3.14}\gamma 2^{-\ell_1/8}; \;\; \mathbb{P}\{\xi \in [2^{\ell_2}, 2^{\ell_2+1}]\} \geq c_{3.14}\gamma 2^{-\ell_2/8}.$$

Now, define λ as the conditional expectation of ξ given the event $\mathcal{M} = \{\omega \in \Omega : \xi(\omega) \in [-2^{\ell_1+1}, -2^{\ell_1}] \cup [2^{\ell_2}, 2^{\ell_2+1}]\}$, i.e.

$$\lambda = \mathbb{P}(\mathcal{M})^{-1} \int\limits_{\mathcal{M}} \xi(\omega) d\,\omega.$$

Let $H_1 = -\lambda + [-2^{\ell_1+1}, -2^{\ell_1}]$ and $H_2 = -\lambda + [2^{\ell_2}, 2^{\ell_2+1}]$. Note that necessarily $\lambda \in [-2^{\ell_1+1}, 2^{\ell_2+1}]$, hence $H_1, H_2 \subset [-2^{\ell+2}, 2^{\ell+2}]$ for $\ell = \max(\ell_1, \ell_2)$. Obviously, $\mathrm{d}(H_1, H_2) \geq 2^{\ell}$ and for $H = H_1 \cup H_2$

$$\mathbb{E}\,(\xi - \lambda)_H = \int\limits_{\{\xi-\lambda \in H\}} (\xi(\omega) - \lambda) d\,\omega = \int\limits_{\mathcal{M}} (\xi(\omega) - \lambda) d\,\omega = 0.$$

Finally,

$$\begin{aligned}\min\big(\mathbb{P}\{\xi - \lambda \in H_1\}, \mathbb{P}\{\xi - \lambda \in H_2\}\big) &= \min\big(\mathbb{P}\{\xi \in [-2^{\ell_1+1}, -2^{\ell_1}]\}, \mathbb{P}\{\xi \in [2^{\ell_2}, 2^{\ell_2+1}]\}\big) \\ &\geq c_{3.14}\gamma 2^{-\ell/8}.\end{aligned}$$

□

Let us recall a folklore estimate of the norm of a random matrix with bounded mean zero entries (see, for example, [91, Proposition 2.4]):

Lemma 3.15. *Let $W = (w_{ij})$ be an $N \times n$ ($N \geq n$) random matrix with i.i.d. mean zero entries; $R > 0$ and assume that $|w_{ij}| \leq R$ a.s. Then for a universal constant $C_{3.15} > 0$*

$$\mathbb{P}\big\{\|W\|_2 \geq C_{3.15} R\sqrt{N}\big\} \leq \exp(-N).$$

The next lemma highlights a useful property of the vectors from $\mathrm{S}^{n-1} \setminus \mathrm{S}_a^{n-1}(\sqrt{N})$:

Lemma 3.16. *For any integer $N \geq n \geq m \geq 1$ and any $y \in \mathrm{S}^{n-1} \setminus \mathrm{S}_a^{n-1}(\sqrt{N})$ there is a set $J = J(y) \subset \{1, 2, \ldots, n\}$ such that $|J| \leq m$, $\|y\chi_J\|_2 \geq \frac{1}{2}\sqrt{\frac{m}{n}}$ and $\|y\chi_J\|_\infty \leq \frac{1}{\lfloor N^{1/4} \rfloor}$.*

Proof. Take any $N \geq n \geq m \geq 1$ and $y = (y_1, y_2, \ldots, y_n) \in \mathrm{S}^{n-1} \setminus \mathrm{S}_a^{n-1}(\sqrt{N})$ and let

$$J'(y) = \Big\{ j \in \{1, 2, \ldots, n\} : \, |y_j| \leq \frac{1}{\lfloor N^{1/4} \rfloor} \Big\}.$$

Obviously, $|J'| \ge n - \sqrt{N} > 0$ and, since y is not almost $\sqrt{N}$-sparse, $\|y\chi_{J'}\|_2 \ge \sqrt{3/4}$. Let $\{J'_1, J'_2, \dots, J'_p\}$ be any partition of J' into pairwise disjoint subsets of cardinality at most m with $p \le \lceil n/m \rceil$. Then, clearly, for some $q \in \{1, 2, \dots, p\}$ we have $\|y\chi_{J_q}\|_2 \ge \|y\chi_{J'}\|_2/\sqrt{p} > \frac{1}{2}\sqrt{\frac{m}{n}}$. Setting, $J(y) = J_q$, we get the result. □

Proposition 3.17 (The set $\mathrm{S}^{n-1} \setminus \mathrm{S}^{n-1}_a(\sqrt{N})$)**.** *For any $\gamma > 0, \delta > 1$ there are $N_{3.17} \in \mathbb{N}$ and $h_{3.17} > 0$ depending only on γ and δ with the following property: Let $N \ge \max(N_{3.17}, \delta n)$ and let A be an $N \times n$ random matrix with i.i.d. entries such that*

$$\min\big(\mathbb{P}\big\{z - \sqrt{N} \le a_{11} \le z - 1\big\}, \mathbb{P}\big\{z + 1 \le a_{11} \le z + \sqrt{N}\big\}\big) \ge \gamma$$

for some $z \in \mathbb{R}$. Then for any non-random $N \times n$ matrix B and the set $S = \mathrm{S}^{n-1} \setminus \mathrm{S}^{n-1}_a(\sqrt{N})$ we have

$$\mathbb{P}\big\{\inf_{y \in S} \|Ay + By\|_2 \le h_{3.17}\sqrt{N}\big\} \le \exp(-w_{3.10}N/2).$$

Proof. Fix any $\gamma > 0$ and $\delta > 1$. To make the notation more compact, denote $f_0 := \frac{(1-\delta^{-1/4})\sqrt{c_{3.14}\gamma}}{C_{3.5}}$ and let $\tau_0 = \tau_0(\gamma, \delta)$ be the largest number in $(0, 1]$ such that for all $s \ge 0$

$$\Big(\frac{16\sqrt{8}C_{3.12}C_{3.15}2^{s/2}}{h_{3.10}f_0\tau_0^{3/2}}\Big)^{2^{-s/4}\tau_0} \le \exp(w_{3.10}/4)$$

(it is not difficult to see that τ_0 is well defined). Then, take $N_{3.17} = N_{3.17}(\gamma, \delta)$ to be the smallest positive integer such that for all $N \ge N_{3.17}$

$$\frac{1}{\lfloor N^{1/4} \rfloor} \le \frac{f_0\sqrt{\tau_0}}{4\sqrt{8}}N^{-3/16} \quad\text{and}\quad \frac{48\sqrt{8N}C_{3.12}C_{3.15}}{h_{3.10}f_0\tau_0^{3/2}} \le \exp(w_{3.10}N/4). \tag{3.9}$$

Let $N \ge N_{3.17}$, $N \ge \delta n$ and let A be an $N \times n$ random matrix with entries satisfying conditions of the lemma and B be any non-random $N \times n$ matrix.

By Lemma 3.14, there is an integer $\ell \in [0, \lfloor \log_2 \sqrt{N} \rfloor]$, $\lambda \in \mathbb{R}$ and disjoint Borel sets $H_1, H_2 \subset [-2^{\ell+2}, 2^{\ell+2}]$ such that $\mathrm{d}(H_1, H_2) \ge 2^\ell$, $\min\big(\mathbb{P}\{a_{11} - \lambda \in H_1\}, \mathbb{P}\{a_{11} - \lambda \in H_2\}\big) \ge c_{3.14}\gamma 2^{-\ell/8}$ and $\mathbb{E}\,(a_{11} - \lambda)_H = 0$ for $H = H_1 \cup H_2$. Denote $\widetilde{A} = A - \lambda \mathbf{1_{N\times n}}$, $\widetilde{B} = B + \lambda \mathbf{1_{N\times n}}$ and let

$$R := 2^{\ell+2},\ d := 2^\ell,\ r := c_{3.14}\gamma 2^{-\ell/8},\ m := \Big\lceil \frac{\tau_0 n}{2^{\ell/4}} \Big\rceil,\ t := \frac{1}{2}\sqrt{\frac{m}{n}},\ \varepsilon := \frac{h_{3.10}h_{3.11}}{2C_{3.15}R},$$

where $h_{3.11}$ is defined as in Proposition 3.11. Assume that S is non-empty and let $T \subset B_2^n$ consist of all m-sparse vectors $y \in B_2^n$ with $\|y\|_2 \ge t$ and $\|y\|_\infty \le \frac{2h_{3.11}}{d}$. The first inequality in (3.9) and a simple calculation show that $\frac{1}{\lfloor N^{1/4}\rfloor} \leq \frac{2h_{3.11}}{d}$. Hence, in view of Lemma 3.16, T is non-empty and satisfies (3.8). By Lemma 3.12, there is a finite subset $\mathcal{N} \subset T$ of cardinality at most $\left(\frac{nC_{3.12}}{m\varepsilon}\right)^m$ such that for any $y \in S$ there is $y' = y'(y) \in \mathcal{N}$ with $\|y\chi_{\operatorname{supp} y'} - y'\|_2 \leq \varepsilon$.

For each $y' \in \mathcal{N}$ denote $E_{y'} = \operatorname{span}\{e_j^n\}_{j \in \operatorname{supp} y'}$. By Proposition 3.11,

$$\mathbb{P}\big\{\mathrm{d}\big(\widetilde{A}_H y', V_{\widetilde{A},\widetilde{B}}(H, E_{y'})\big) \leq h_{3.10} h_{3.11}\sqrt{N}\big\} \leq 2\exp(-w_{3.10}N).$$

Define an event

$$\begin{aligned}\mathcal{E} = \big\{\omega \in \Omega :\ &\mathrm{d}\big(\widetilde{A}_H(\omega) y', V_{\widetilde{A},\widetilde{B}}(H, E_{y'})(\omega)\big) > h_{3.10} h_{3.11}\sqrt{N} \\ &\text{for all } y' \in \mathcal{N} \text{ and } \|\widetilde{A}_H(\omega)\|_2 \leq C_{3.15} R\sqrt{N}\big\}.\end{aligned}$$

By the above probability estimates and Lemma 3.15,

$$\begin{aligned}\mathbb{P}(\mathcal{E}) &\geq 1 - \exp(-N) - 2|\mathcal{N}|\exp\big(-w_{3.10}N\big) \\ &\geq 1 - \exp(-N) - 2\Big(\frac{C_{3.12}n}{m\varepsilon}\Big)^m \exp\big(-w_{3.10}N\big).\end{aligned}$$

Using the definition of ε, m, τ_0 and the second inequality in (3.9), we can estimate the probability as

$$\begin{aligned}\mathbb{P}(\mathcal{E}) &\geq 1 - 3\Big(\frac{8C_{3.12}C_{3.15}2^{\ell+\ell/4}}{\tau_0 h_{3.10} h_{3.11}}\Big)^{2^{-\ell/4}\tau_0 n + 1} \exp(-w_{3.10}N) \\ &\geq 1 - 3\Big(\frac{16\sqrt{8}C_{3.12}C_{3.15}2^{\ell/2}}{h_{3.10} f_0 \tau_0^{3/2}}\Big)^{2^{-\ell/4}\tau_0 n + 1} \exp(-w_{3.10}N) \\ &\geq 1 - \exp(-w_{3.10}N/2).\end{aligned}$$

Take any $\omega \in \mathcal{E}$ and define $D_1 = \widetilde{A}_H(\omega)$, $D_2 = \widetilde{A}_{\overline{H}}(\omega) + \widetilde{B}$, $D = A(\omega) + B(\omega) = D_1 + D_2$. Then $\|D_1\|_2 \leq C_{3.15}R\sqrt{N}$ and for every $y' \in \mathcal{N}$ we have

$$\mathrm{d}\big(D_1 y', D(E_{y'}^{\perp}) + D_2(E_{y'})\big) > h_{3.10} h_{3.11}\sqrt{N}.$$

Hence, by Proposition 3.3 and the definition of ε, we get

$$\inf_{y \in S}\|Dy\|_2 > h_{3.10} h_{3.11}\sqrt{N} - \varepsilon C_{3.15} R\sqrt{N} = \frac{1}{2} h_{3.10} h_{3.11}\sqrt{N} \geq \frac{h_{3.10} f_0 \sqrt{\tau_0}}{4\sqrt{8}}\sqrt{N}.$$

Finally, applying the above argument to the entire set $\mathcal{E}$, we obtain the result. $\square$

Proof of Theorem 3.1. In view of the trivial identity $\mathcal{Q}(a_{ij},\alpha)=\mathcal{Q}(a_{ij}/\alpha,1)$, it is enough to prove the theorem for $\alpha=1$. Fix any $\delta>0$ and $\beta>0$, let $\gamma=\beta/4$ and let $N_0=N_0(\beta,\delta)$ be the smallest integer such that $N_0\geq\max(N_{3.13},N_{3.17})$ and for all $N\geq N_0$

$$N\leq\exp(w_{3.7}N/2)\ \text{ and }\ 3\leq\exp\bigl(\min(w_{3.7},w_{3.10})N/4\bigr).$$

Take any $N,n\in\mathbb{N}$ with $N\geq\max(N_0,\delta n)$, let $A=(a_{ij})$ be a $N\times n$ random matrix with i.i.d. entries satisfying $\mathcal{Q}(a_{11},1)\leq 1-\beta$ and let B be any non-random $N\times n$ matrix. By the right-continuity of the cdf of a_{11}, there is $z\in\mathbb{R}$ such that

$$\mathbb{P}\{a_{11}\leq z-1\}\geq\frac{\beta}{2}\ \text{ and }\ \mathbb{P}\{a_{11}<z-1\}\leq\frac{\beta}{2}.$$

Then

$$\mathbb{P}\{a_{11}\geq z+1\}\geq 1-\mathbb{P}\{a_{11}<z-1\}-\mathcal{Q}(a_{11},1)\geq\frac{\beta}{2}.$$

Let us consider three cases.

1) $\mathbb{P}\{z+1\leq a_{11}\leq z+\sqrt{N}\}\leq\gamma$. Then $\mathcal{Q}(a_{11},\sqrt{N}/8)\leq\mathcal{Q}(a_{11},(\sqrt{N}-1)/2)\leq 1-\gamma$. Obviously, any vector on S^{n-1} is $N^{-1/2}$-peaky. Then, applying Proposition 3.7 with the "scaling factor" $\sqrt{N}/8$, we get

$$\begin{aligned}\mathbb{P}\bigl\{s_{\min}(A+B)\leq h_{3.7}\sqrt{N}/8\bigr\}&=\mathbb{P}\bigl\{\inf_{y\in\mathrm{S}^{n-1}}\|Ay+By\|_2\leq h_{3.7}\sqrt{N}/8\bigr\}\\&\leq n\exp(-w_{3.7}N)\\&\leq\exp(-w_{3.7}N/2).\end{aligned}$$

2) $\mathbb{P}\{z-\sqrt{N}\leq a_{11}\leq z-1\}\leq\gamma$. Treated as above.

3) $\min\bigl(\mathbb{P}\{z-\sqrt{N}\leq a_{11}\leq z-1\},\mathbb{P}\{z+1\leq a_{11}\leq z+\sqrt{N}\}\bigr)\geq\gamma$. Define $\theta_{3.13}$ as in Proposition 3.13. By Proposition 3.7 for peaky vectors,

$$\mathbb{P}\bigl\{\inf_{y\in\mathrm{S}_p^{n-1}(\theta_{3.13})}\|Ay+By\|_2\leq h_{3.7}\theta_{3.13}\sqrt{N}\bigr\}\leq n\exp(-w_{3.7}N)\leq\exp(-w_{3.7}N/2).$$

By Propositions 3.13 and 3.17 for $S=\mathrm{S}_a^{n-1}(\sqrt{N})\setminus\mathrm{S}_p^{n-1}(\theta_{3.13})$ and $S'=\mathrm{S}^{n-1}\setminus\mathrm{S}_a^{n-1}(\sqrt{N})$ we have

$$\begin{aligned}&\mathbb{P}\bigl\{\inf_{y\in S}\|Ay+By\|_2\leq h_{3.13}\sqrt{N}\bigr\}\leq\exp(-w_{3.10}N/2);\\&\mathbb{P}\bigl\{\inf_{y\in S'}\|Ay+By\|_2\leq h_{3.17}\sqrt{N}\bigr\}\leq\exp(-w_{3.10}N/2).\end{aligned}$$

Combining the estimates, we get for $h = \min\big(h_{3.7}\theta_{3.13}, h_{3.13}, h_{3.17}\big)$:

$$\begin{aligned}\mathbb{P}\big\{s_{\min}(A+B) \le h\sqrt{N}\big\} &\le \exp(-w_{3.7}N/2) + 2\exp(-w_{3.10}N/2)\\ &\le \exp\big(-\min(w_{3.7}, w_{3.10})N/4\big).\end{aligned}$$

This completes the proof. □

3.2 The Limit of the Smallest Singular Value of Random Matrices with i.i.d. Entries[2]

3.2.1 Introduction

For $N \geq m$ and an $N \times m$ real-valued matrix B, its *singular values* $s_1(B)$, $s_2(B), \dots,$ $s_m(B)$ are the eigenvalues of the matrix $\sqrt{B^T B}$ arranged in non-increasing order, where multiplicities are counted. In particular, *the largest* and *the smallest* singular values are given by

$$s_{\max}(B) = \sup_{y \in \mathrm{S}^{m-1}} \|By\|_2 = \|B\|_2; \quad s_{\min}(B) = \inf_{y \in \mathrm{S}^{m-1}} \|By\|_2.$$

In this section, we establish convergence of the smallest singular values of a sequence random matrices with i.i.d. entries under minimal moment assumptions.

The extreme singular values of random matrices attract considerable attention of researchers both in *limiting* and *non-limiting* settings. We refer the reader to surveys and monographs [6, 85, 91, 117] for extensive information on the spectral theory of random matrices. Here, we shall focus on the following specific question: for matrices with i.i.d. entries, what are the weakest possible assumptions on the entries which are sufficient for the smallest singular value to "concentrate"?

We note that a corresponding problem for the *largest* singular value (i.e. the operator norm) was essentially resolved in the i.i.d. case, where finiteness of the fourth moment of the entries turns out to be crucial both in limiting and non-limiting settings. We refer the reader to [120] and [7] for results on a.s. convergence of the largest singular value, and [62] for the non-limiting case (see also [103], [68] for some negative results on concentration of the operator norm).

For the *smallest* singular value, its concentration properties are relatively well understood in the i.i.d. case provided that the fourth moment of the matrix entries is bounded. A classical theorem of Z.D. Bai and Y.Q. Yin [8] (see also [6, Theorem 5.11]) states the following: given an array $\{a_{ij}\}$ $(1 \leq i, j < \infty)$ of i.i.d. random variables such that $\mathbb{E}\, a_{ij} = 0$, $\mathbb{E}\, {a_{ij}}^2 = 1$ and $\mathbb{E}\, {a_{ij}}^4 < \infty$, and an integer sequence $(N_m)_{m=1}^{\infty}$ with $m/N_m \longrightarrow z$ for some $z \in (0,1)$, the $N_m \times m$ matrices $A_m = (a_{ij})$ $(1 \leq i \leq N_m, 1 \leq j \leq m)$ satisfy

$${N_m}^{-1/2} s_{\min}(A_m) \longrightarrow 1 - \sqrt{z} \;\text{ almost surely.}$$

Further, it is proved in [94, 95] that for square $m \times m$ matrices with i.i.d. centered entries

[2]A version of this section has been published. K. Tikhomirov, The limit of the smallest singular value of random matrices with i.i.d. entries, Adv. Math. **284** (2015), 1–20.

with unit variance and a bounded fourth moment, $s_{\min}(A)$ is of order $m^{-1/2}$ with a large probability.

A natural question in connection with the mentioned results is *whether the assumption on the fourth moment is necessary for the least singular value to "concentrate"*; in particular, whether any assumptions on moments of a_{ij}'s higher than the 2-nd are required for the a.s. convergence in the Bai–Yin theorem. This question is discussed in [6] on p. 6. Solving the problem was a motivation for our work.

A considerable progress has been made recently in the direction of weakening the moment assumptions on matrix entries. For square matrices, given a sufficiently large m and an $m \times m$ matrix with i.i.d. entries with zero mean and unit variance, its smallest singular value is bounded from below by a constant (negative) power of m with probability close to one [110, Theorem 2.1] (see also [41, Theorem 4.1] for sparse matrices).

For *tall* rectangular matrices, N. Srivastava and R. Vershynin proved in [106] that for any $\varepsilon, \eta > 0$ and an $N \times m$ random matrix A with independent isotropic rows X_i such that $\sup_{y \in \mathbb{S}^{m-1}} \mathbb{E}\,|\langle X_i, y\rangle_m|^{2+\eta} \le C$, the singular value $s_{\min}(A)$ satisfies $\mathbb{E}\, s_{\min}(A)^2 \ge (1-\varepsilon)N$ provided that the aspect ratio N/m is bounded from below by a certain function of ε and η. This result of [106] was strengthened by V. Koltchinskii and S. Mendelson [59] who proved that, under similar assumptions on the matrix, $s_{\min}(A) \ge (1-\varepsilon)\sqrt{N}$ with a very large probability. Moreover, another theorem of [59] states that, for a sufficiently tall $N \times m$ random matrix A with i.i.d. isotropic rows satisfying certain "spreading" condition, $s_{\min}(A) \gtrsim \sqrt{N}$ with probability very close to one. Some further strengthening of the results of [59] is obtained in [118].

In Section 3.1, we considered a situation when no upper bounds for moments of the matrix entries are given. Our result can be used to show that in the limiting setup of the Bai–Yin theorem but without the assumptions on moments higher than the 2-nd, the sequence $\big({N_m}^{-1/2} s_{\min}(A_m)\big)_{m=1}^{\infty}$ satisfies

$$\liminf_{m\to\infty} \big({N_m}^{-1/2} s_{\min}(A_m)\big) \ge r > 0 \text{ almost surely,}$$

where r is a certain function of $z = \lim m/N_m$ and the distribution of a_{ij}'s. The same conclusion can be derived from [59, Theorem 1.4], if we additionally assume that the limiting aspect ratio z is bounded from above by a sufficiently small positive quantity (i.e. the matrices are tall). However, both Theorem 3.1 and [59, Theorem 1.4] do not give the precise asymptotics.

This problem is resolved here. The main result is the following

Theorem 3.18. *Let $\{a_{ij}\}$ $(1 \le i, j < \infty)$ be a set of i.i.d. real valued random variables* *with zero mean and unit variance. Further, let* $(N_m)_{m=1}^\infty$ *be an integer sequence satisfying* $m/N_m \longrightarrow z$ *for some* $z \in (0,1)$*. For every* $m \in \mathbb{N}$ *we denote by* A_m *the random* $N_m \times m$ *matrix with entries* a_{ij} $(1 \le i \le N_m, 1 \le j \le m)$*. Then with probability one the sequence*

$$\left({N_m}^{-1/2} s_{\min}(A_m)\right)_{m=1}^\infty$$

converges to $1 - \sqrt{z}$*.*

Theorem 3.18 in a strong form establishes the *asymmetry* of the limiting behaviour of the extreme singular values: whereas the fourth moment is necessary for the operator norm, the second moment is sufficient for the convergence of the smallest singular value.

Let us briefly describe our approach to proving Theorem 3.18. We shall "approximate" the matrices A_m by matrices with truncated and centered entries. Namely, for $M > 0$ and all $m \ge 1$ let $\widetilde{A}_m$ be the $N_m \times m$ matrix with the entries

$$\widetilde{a}_{ij} = a_{ij}\chi_{\{|a_{ij}| \le M\}} - \mathbb{E}\left(a_{ij}\chi_{\{|a_{ij}| \le M\}}\right), \;\; 1 \le i \le N_m, \;\; 1 \le j \le m,$$

where $\chi_{\mathcal{E}}$ is the indicator of an event $\mathcal{E}$. If the truncation level M is large enough then it turns out that for *all sufficiently large* m we have $s_{\min}(\widetilde{A}_m) \approx s_{\min}(A_m)$ with probability close to one. In fact, we need only one-sided estimate for our proof. To be more precise, we will show that with a large probability the quantity

$$\limsup_{m\to\infty} {N_m}^{-1/2}\left(s_{\min}(\widetilde{A}) - s_{\min}(A)\right)$$

is bounded from above by a positive number which depends only on M and can be made arbitrarily small by increasing the truncation level (in a more technical form, this is stated in Theorem 3.32 of the note). Then, applying the Bai–Yin theorem [8] to the truncated matrices $\widetilde{A}_m$, we get

$$\liminf_{m\to\infty} {N_m}^{-1/2} s_{\min}(A) \gtrsim \liminf_{m\to\infty} {N_m}^{-1/2} s_{\min}(\widetilde{A}) \gtrsim 1 - \sqrt{z} \;\; \text{almost surely},$$

which implies the result. Thus, the argument of the paper [8] remains the crucial element of the proof, although we apply it only to the truncated variables, for which all positive moments are bounded. Let us emphasize that, whereas a truncation procedure for matrices also appears as a technical step in [8], in our approach the truncation level M is *not a function of* m.

Note that the equivalence $s_{\min}(A_m) \approx s_{\min}(\widetilde{A}_m)$ would follow immediately if the difference $A_m - \widetilde{A}_m$ had the operator norm very small compared to $\sqrt{N_m}$ with a large probability. However, the moment assumptions that we impose on a_{ij}'s are too weak to expect a good upper bound for $\|A_m - \widetilde{A}_m\|_{2\to 2}$. To overcome this problem, we shall consider a special *non-convex* function of the matrix $A_m - \widetilde{A}_m$ which has much better concentration properties than the norm and which shall act as a "replacement" for the norm in our calculations. This quantity and its concentration properties are discussed in Section 3.2.3 and are the main novel igredient of this part of the thesis.

3.2.2 Preliminaries

Here, we present some classical or elementary facts, which we include for an easier referencing.

We denote by $(\Omega, \Sigma, \mathbb{P})$ a probability space. Universal constants are denoted by C, c_1, etc. A numerical subscript in the name of a constant determines the statement where the constant is defined. Similarly, a function defined within a statement and intended to be used further in the text, has the statement number as a subscript.

Let T be a subset of $\mathbb{R}^n$ and $\|\cdot\|_B$ be a norm on $\mathbb{R}^n$ with the unit ball B. A subset $\mathcal{N} \subset T$ is called *an ε-net in T with respect to* $\|\cdot\|_B$ if for any $y \in T$ there is $y' \in \mathcal{N}$ satisfying $\|y - y'\|_B \le \varepsilon$. We shall omit the reference to $\|\cdot\|_B$ when $B = B_2^n$.

Lemma 3.19. *For any $n \in \mathbb{N}$ and $\varepsilon \in (0,1]$ there exists an ε-net in B_2^n of cardinality at most $\left(\frac{3}{\varepsilon}\right)^n$.*

Lemma 3.20. *For any $n \in \mathbb{N}$ and any $T \subset \mathrm{S}^{n-1}$ there is an $n^{-1/2}$-net in T with respect to $\|\cdot\|_\infty$ of cardinality at most* $\exp(C_{3.20}n)$. *Here, $C_{3.20} > 0$ is a universal constant.*

Remark 3.2. Both lemmas above follow from a well known estimate for covering numbers for pairs of convex sets in $\mathbb{R}^n$ (see, for example, [86, Lemma 4.16]). For Lemma 3.20, the estimate for the pair (B_2^n, B_∞^n) yields an existence of a $(4n)^{-1/2}$-net $\bar{\mathcal{N}}$ in B_2^n with respect to $\|\cdot\|_\infty$ of cardinality at most $\exp(C_{3.20}n)$ for a universal constant $C_{3.20} > 0$. Then $\mathcal{N} \subset T$ can be constructed by picking a point from every non-empty intersection of the form $(y' + (4n)^{-1/2}B_\infty^n) \cap T$, $y' \in \bar{\mathcal{N}}$.

The next statement, which is sometimes called the Bernstein (or Hoeffding's) inequality, can be derived from classical Khintchine's inequality for the sum of weighted independent signs by a symmetrization procedure:

Lemma 3.21 (see, for ex., [117, Proposition 5.10])**.** *Let $n \in \mathbb{N}$, $M > 0$, $y = (y_1, y_2, \dots, y_n)$ with $\|y\| = 1$, and let $a_1, a_2 \dots, a_n$ be independent mean zero random variables with $|a_j| \le$*

M a.s. $(j = 1, 2, \dots, n)$. *Then*

$$\mathbb{P}\Big\{\Big|\sum_{j=1}^{n} a_j y_j\Big| \geq \tau\Big\} \leq 2\exp(-c_{3.21}\tau^2/M^2), \ \ \tau > 0,$$

where $c_{3.21} > 0$ *is a universal constant.*

The lemma below is a law of large numbers, where instead of the arithmetic mean of a collection of random variables we consider more general weighted sums. As in the case of the classical weak LLN, the statement can be proved by applying Lévy's continuity theorem for characteristic functions.

Lemma 3.22. *Let* $a_1, a_2, \dots$ *be i.i.d. random variables with zero mean. Then for any* $\varepsilon > 0$ *there is* $\delta > 0$ *depending only on* ε *and the distribution of* a_j*'s with the following property: whenever* $(t_j)_{j=1}^{\infty}$ *is a sequence of non-negative real numbers such that* $\sum_{j=1}^{\infty} t_j = 1$ *and* $\max t_j \leq \delta$, *we have*

$$\mathbb{P}\Big\{\Big|\sum_{j=1}^{\infty} a_j t_j\Big| > \varepsilon\Big\} < \varepsilon.$$

Given an $m \times m$ random symmetric matrix T with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_m$, *the empirical spectral distribution* of T is the function on $\mathbb{R}$ given by

$$F^T(t) = \frac{1}{m}\big|\{j \leq m :\ \lambda_j \leq t\}\big|, \ \ t \in \mathbb{R}.$$

Theorem 3.23 (Marčenko–Pastur law; see [71], [119], [6, Theorem 3.6])**.** *Let* $\{a_{ij}\}$ $(1 \leq i, j \leq \infty)$ *be a set of i.i.d. random variables with zero mean and unit variance and let* $(N_m)_{m=1}^{\infty}$ *be an integer sequence satisfying* $m/N_m \longrightarrow z$ *for some* $z \in (0, 1)$. *For every* $m \in \mathbb{N}$ *denote by* A_m *the random* $N_m \times m$ *matrix with entries* a_{ij} $(1 \leq i \leq N_m, 1 \leq j \leq m)$ *and by* T_m *the matrix* $\frac{1}{N_m} A_m^T A_m$. *Then with probability one the sequence of empirical spectral distributions* $\{F^{T_m}\}$ *converges pointwise to a non-random distribution given by*

$$F_{MP}(t) = \begin{cases} 0, & \text{if } t \leq r, \\ \frac{1}{2\pi z}\int\limits_{r}^{t} \frac{\sqrt{(R-\tau)(\tau - r)}}{\tau}\, d\tau, & \text{if } r \leq t \leq R, \\ 1, & \text{if } t \geq R. \end{cases}$$

where $r = (1 - \sqrt{z})^2$ *and* $R = (1 + \sqrt{z})^2$.

Remark 3.3. Note that the above theorem does not require any assumptions on moments higher than the 2nd, and so can be applied in our setting. For our proof, we will actually need a much weaker result than Theorem 3.23, namely, that $\limsup\limits_{m\to\infty} \frac{s_{\min}(A_m)}{\sqrt{N_m}} \le 1-\sqrt{z}$ almost surely. The latter can be immediately verified with help of Theorem 3.23: for every fixed $t > (1-\sqrt{z})^2$, we have $\lim\limits_{m\to\infty} F^{T_m}(t) = F_{MP}(t) > 0$ with probability one, hence the smallest non-zero eigenvalues $\lambda_{\min}(T_m)$ of matrices T_m satisfy $\limsup\limits_{m\to\infty} \lambda_{\min}(T_m) \le t$ a.s. This implies $\limsup\limits_{m\to\infty} \frac{s_{\min}(A_m)}{\sqrt{N_m}} \le \sqrt{t}$ a.s., which gives the required estimate by letting $t \to (1-\sqrt{z})^2$.

3.2.3 Norms of coordinate projections of random vectors

For any $N \in \mathbb{N}$ and a subset $I \subset \{1, 2, \ldots, N\}$, let us denote by $\mathrm{Proj}_I : \mathbb{R}^N \to \mathbb{R}^N$ the coordinate projection onto the subspace spanned by $\{e_i^N\}_{i\in I}$. Throughout the rest of the section, we will often use expressions of the form $\min\limits_{|I|\ge r} \|\mathrm{Proj}_I x\|_2$, where x is some vector in $\mathbb{R}^N$ and r is a positive real number. This notation should be interpreted as the minimum of $\|\mathrm{Proj}_I x\|_2$ over *all* subsets $I \subset \{1, 2, \ldots, N\}$ of cardinality at least r.

The goal of this section is to show that, given a sufficiently large random $N \times n$ matrix A with i.i.d. entries with zero mean and unit variance, the quantity

$$\sup_{y\in \mathrm{S}^{n-1}} \min_{|I|\ge N-\varepsilon N} \|\mathrm{Proj}_I A y\|_2 \tag{3.10}$$

is of order $\sqrt{N}$ with a very large probability (the probability shall depend on $\varepsilon > 0$). It shall act as a "replacement" of the matrix norm $\|A\|_{2\to 2}$ which in our setting may be much greater than $\sqrt{N}$ with probability close to one. We remark here that a quantity

$$\max_{|I|=m} \|\mathrm{Proj}_I D\|_{2\to 2} = \sup_{y\in \mathrm{S}^{n-1}} \max_{|I|=m} \|\mathrm{Proj}_I D y\|_2,$$

where $m \le N$ and D is an $N \times n$ random matrix with i.i.d. isotropic log-concave rows, played a crucial role in the paper [1] by R. Adamczak, A. Litvak, A. Pajor and N. Tomczak-Jaegermann, dealing with the problem of approximating covariance matrix of a log-concave random vector by the sample covariance matrix. In our case, however, the latter quantity is inapplicable as it may not concentrate near $\sqrt{N}$ (even for small m).

First, we prove the required estimate for (3.10) under the additional assumption that the entries of A are symmetrically distributed (Lemma 3.29). Then we generalize the result to non-symmetric distributions in Proposition 3.30.

Let us outline the proof of Lemma 3.29. A crucial observation (that will be formally justified later) is that there exist finite sets $\mathcal{N}_1 \subset 2B_2^n$ and $\mathcal{N}_2 \subset B_2^n \cap (n^{-1/4} B_\infty^n)$ with $|\mathcal{N}_1| \lesssim n^{\sqrt{n}}$ and $|\mathcal{N}_2| \le \exp(Cn)$ such that S^{n-1} is a subset of the Minkowski sum $\mathcal{N}_1 +$

$\mathcal{N}_2 + \frac{2}{\sqrt{n}} B_\infty^n$. In this way, estimating the supremum over the unit sphere can be reduced to considering separately

$$\sup_{y \in S} \min_{|I| \geq N - \varepsilon N/3} \|\mathrm{Proj}_I Ay\|_2,$$

where $S = \mathcal{N}_1, \mathcal{N}_2, \frac{2}{\sqrt{n}} B_\infty^n$. For $S = \mathcal{N}_1, \mathcal{N}_2$, we shall use estimates for individual vectors (Lemmas 3.24 and 3.27 below) and then apply the union bound. The cube is treated in Lemma 3.28.

Lemma 3.24. *For each $\varepsilon \in (0,1]$ there is $N_{3.24} = N_{3.24}(\varepsilon) > 0$ depending only on ε with the following property: let $N \geq N_{3.24}$ and let $X = (X_1, X_2, \ldots, X_N)$ be a random vector of independent variables, each X_i having zero mean and unit variance. Then*

$$\min_{|I| \geq N - \varepsilon N} \|\mathrm{Proj}_I X\|_2 \leq C_{3.24} \sqrt{N}$$

with probability at least $1 - \exp(-c_{3.24} \varepsilon N)$, where $C_{3.24}, c_{3.24} > 0$ are universal constants.

Proof. Fix any $\varepsilon \in (0,1]$ and define $N_{3.24}$ as the smallest positive integer such that

$$\left(\frac{e}{4}\right)^{\varepsilon N} + \exp(-\varepsilon e N/4) \leq \exp(-\varepsilon N/3)$$

for all $N \geq N_{3.24}$. Choose any $N \geq N_{3.24}$ and let X be as stated above. Set $M = \frac{4}{\varepsilon}$. In view of Markov's inequality,

$$\mathbb{P}\big\{|\{i \leq N : |X_i| \geq \sqrt{M}\}| \geq 4N/M\big\} \leq \binom{N}{\lceil 4N/M \rceil} \left(\frac{1}{M}\right)^{\lceil 4N/M \rceil} \leq \left(\frac{e}{4}\right)^{\lceil 4N/M \rceil}.$$

Let $\widetilde{X} = (\widetilde{X}_1, \widetilde{X}_2, \ldots, \widetilde{X}_N)$ be a vector of truncations of X_i's, with

$$\widetilde{X}_i(\omega) = \begin{cases} X_i(\omega), & \text{if } |X_i(\omega)| \leq \sqrt{M}; \\ 0, & \text{otherwise.} \end{cases}$$

Then, from the above estimate,

$$\mathbb{P}\big\{\min_{|I| \geq N - \varepsilon N} \|\mathrm{Proj}_I X\|_2 > \|\widetilde{X}\|_2\big\} \leq \mathbb{P}\big\{|\{i \leq N : |X_i| \geq \sqrt{M}\}| \geq \varepsilon N\big\} \leq \left(\frac{e}{4}\right)^{\varepsilon N}.$$

Now, let us estimate the Euclidean norm of $\widetilde{X}$ using the Laplace transform. Set $\lambda = \frac{1}{M}$.

We have

$$
\begin{aligned}
\mathbb{E}\,\exp(\lambda\|\widetilde{X}\|_2^2) &= \prod_{i=1}^{N} \mathbb{E}\,\exp(\lambda \widetilde{X}_i^2) \\
&= \prod_{i=1}^{N}\Big(1+\int_1^{\exp(\lambda M)} \mathbb{P}\big\{\exp(\lambda\widetilde{X}_i^2)\ge\tau\big\}\,d\tau\Big) \\
&\le \prod_{i=1}^{N}\Big(1+\int_1^{e} \mathbb{P}\Big\{\widetilde{X}_i^2\ge\frac{\tau-1}{e\lambda}\Big\}\,d\tau\Big) \\
&\le \prod_{i=1}^{N}\Big(1+e\lambda\mathbb{E}\,\widetilde{X}_i^2\Big) \\
&\le \big(1+e\lambda\big)^N \\
&\le \exp(eN/M).
\end{aligned}
$$

Hence,

$$\mathbb{P}\big\{\|\widetilde{X}\|_2\ge\sqrt{2eN}\big\}\le\exp(-eN/M).$$

Finally, using the definition of $N_{3.24}$, we get

$$
\begin{aligned}
&\mathbb{P}\big\{\min_{|I|\ge N-\varepsilon N}\|\mathrm{Proj}_I X\|_2>\sqrt{2eN}\big\} \\
&\quad\le \mathbb{P}\big\{\min_{|I|\ge N-\varepsilon N}\|\mathrm{Proj}_I X\|_2>\|\widetilde{X}\|_2\big\}+\mathbb{P}\big\{\|\widetilde{X}\|_2\ge\sqrt{2eN}\big\} \\
&\quad\le \Big(\frac{e}{4}\Big)^{\varepsilon N}+\exp(-\varepsilon eN/4) \\
&\quad\le \exp(-\varepsilon N/3).
\end{aligned}
$$

□

Lemma 3.25. *For every $K>0$ there is $L_{3.25}=L_{3.25}(K)>0$ depending only on K with the following property: Let $N,n\in\mathbb{N}$, $N\ge n$, and let $A=(a_{ij})$ be an $N\times n$ random matrix with i.i.d. symmetrically distributed entries with unit variance. For each $y=(y_1,y_2,\dots,y_n)\in\mathrm{S}^{n-1}$ let $I_y:\Omega\to 2^{\{1,2,\dots,N\}}$ be a random subset of $\{1,2,\dots,N\}$ defined as*

$$I_y=\Big\{i\le N:\ \sum_{j=1}^{n}a_{ij}^2y_j^2\le 2\Big\}.$$

Then for every $y\in\mathrm{S}^{n-1}$ we have

$$\mathbb{P}\big\{\|\mathrm{Proj}_{I_y}Ay\|_2\ge L_{3.25}\sqrt{N}\big\}\le\exp(-KN).$$

Proof. Fix any $K > 0$ and let N, n and $A = (a_{ij})$ be as stated above. Let r_{ij} ($1 \le i \le N$, $1 \le j \le n$) be Rademacher variables jointly independent with A, and let $\bar{A}$ denote the random $N \times n$ matrix $(r_{ij} a_{ij})$. Then, since a_{ij}'s are symmetrically distributed, for any fixed vector $y = (y_1, y_2, \ldots, y_n) \in \mathrm{S}^{n-1}$ the distribution of $\|\mathrm{Proj}_{I_y} Ay\|_2$ is the same as that of $\|\mathrm{Proj}_{I_y} \bar{A}y\|_2$. Define a subset of (non-random) $N \times n$ matrices:

$$\mathcal{M}_y = \Big\{ B = (b_{ij}) \in \mathbb{R}^{N\times n} : \sum_{j=1}^{n} b_{ij}^2 y_j^2 \le 2 \text{ for all } i = 1, 2, \ldots, N \Big\}$$

and for every $B = (b_{ij}) \in \mathcal{M}_y$ denote by $\bar{B}$ the random matrix $(r_{ij} b_{ij})$. Note that at every point ω of the probability space the matrix $\mathrm{Proj}_{I_{y(\omega)}} \bar{A}(\omega)$ belongs to $\mathcal{M}_y$. Then, conditioning on a_{ij}'s, we get for every $\tau > 0$:

$$\mathbb{P}\big\{ \|\mathrm{Proj}_{I_y} Ay\|_2 \ge \tau \big\} = \mathbb{P}\big\{ \|\mathrm{Proj}_{I_y} \bar{A}y\|_2 \ge \tau \big\} \le \sup_{B \in \mathcal{M}_y} \mathbb{P}\big\{ \|\bar{B}y\|_2 \ge \tau \big\}. \tag{3.11}$$

Note that for each $B \in \mathcal{M}_y$ and $i \le N$, the i-th coordinate of the vector $\bar{B}y$ satisfies in view of Lemma 3.21:

$$\mathbb{P}\big\{ |\langle \bar{B}y, e_i^N \rangle_N| \ge \tau \big\} \le 2\exp\big(-c_{3.21}\tau^2/2\big), \ \ \tau > 0.$$

A standard application of the Laplace transform then yields

$$\mathbb{P}\big\{ \|\bar{B}y\|_2 \ge L_{3.25}\sqrt{N} \big\} \le \exp(-KN)$$

for some $L_{3.25} > 0$ depending only on K. This, together with (3.11), proves the result. □

Lemma 3.26. *Let ξ be a symmetrically distributed random variable with unit variance. For every $\varepsilon > 0$ and $K > 0$ there is $\delta_{3.26} = \delta_{3.26}(\varepsilon, K) > 0$ depending on ε, K and the distribution of ξ with the following property: whenever $N, n \in \mathbb{N}$, $N \ge n$; $A = (a_{ij})$ is an $N \times n$ random matrix with i.i.d. entries distributed as ξ and $y \in \mathrm{S}^{n-1}$ is a vector satisfying $\|y\|_\infty \le \delta_{3.26}$, we have*

$$\mathbb{P}\{|I_y| \le N - \varepsilon N\} \le \exp(-KN),$$

where I_y is defined as in Lemma 3.25.

Proof. Fix any $K > 0$ and $\varepsilon \in (0, 1]$. In view of Lemma 3.22, there is $\delta > 0$ such that for all $y = (y_1, y_2, \ldots) \in \ell_2$ with $\|y\|_2 = 1$ and $\|y\|_\infty \le \delta$, and for a sequence of independent random variables $a_1, a_2 \dots$ distributed as ξ, we have

$$\mathbb{P}\Big\{\sum_{j=1}^{\infty} a_j^2 y_j^2 > 2\Big\} \le \varepsilon \exp\big(-1 - K/\varepsilon\big).$$

Now, fix $N, n \in \mathbb{N}$ with $N \ge n$ and $y \in \mathrm{S}^{n-1}$ with $\|y\|_\infty \le \delta$, and let A be defined as above. Then, using the last estimate, we obtain

$$\begin{aligned}\mathbb{P}\big\{|I_y| \le N - \varepsilon N\big\} &= \mathbb{P}\Big\{\big|\big\{i \le N : \sum_j a_{ij}^2 y_j^2 > 2\big\}\big| \ge \varepsilon N\Big\} \\ &\le \binom{N}{\lceil \varepsilon N\rceil}\Big(\frac{\varepsilon}{e}\Big)^{\lceil \varepsilon N\rceil} \exp(-KN) \\ &\le \exp(-KN).\end{aligned}$$

□

As an elementary consequence of Lemmas 3.25 and 3.26 we get

Lemma 3.27. *Let ξ be a symmetrically distributed random variable with unit variance. For every $\varepsilon > 0$ and $K > 0$ there are $\delta_{3.27} = \delta_{3.27}(\varepsilon, K) > 0$ depending on ε, K and the distribution of ξ, and $L_{3.27} = L_{3.27}(K) > 0$ depending only on K such that, whenever $N, n \in \mathbb{N}$, $N \ge n$; $A = (a_{ij})$ is an $N \times n$ random matrix with i.i.d. entries distributed as ξ, and $y \in \mathrm{S}^{n-1}$ is a vector satisfying $\|y\|_\infty \le \delta_{3.27}$, we have*

$$\mathbb{P}\big\{\min_{|I| \ge N - \varepsilon N} \|\mathrm{Proj}_I Ay\|_2 \ge L_{3.27}\sqrt{N}\big\} \le \exp(-KN).$$

Lemma 3.28. *Let ξ be a symmetrically distributed random variable with unit variance. For every $\varepsilon > 0$ and $K > 0$ there are $n_{3.28} = n_{3.28}(\varepsilon, K) \in \mathbb{N}$ depending on ε, K and the distribution of ξ, and $L_{3.28} = L_{3.28}(K) > 0$ depending only on K such that, whenever $N \ge n \ge n_{3.28}$ and $A = (a_{ij})$ is an $N \times n$ random matrix with i.i.d. entries distributed as ξ, we have*

$$\mathbb{P}\big\{\min_{|I| \ge N - \varepsilon N} \sup_{y \in B_\infty^n} \|\mathrm{Proj}_I Ay\|_2 \ge L_{3.28}\sqrt{nN}\big\} \le \exp(-KN).$$

Proof. Fix any $K > 0$ and $\varepsilon > 0$ and define $n_{3.28} = \lceil \delta_{3.26}(\varepsilon, K+1)^{-2}\rceil$, where $\delta_{3.26} > 0$ is taken from Lemma 3.26. Now, choose any $N \ge n \ge n_{3.28}$ and let $A = (a_{ij})$ be an $N \times n$ random matrix with i.i.d. entries distributed as ξ. Let V be the set of vertices of the cube $\frac{1}{\sqrt{n}} B_\infty^n = [-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}]^n$. In view of Lemma 3.26, any $v \in V$ satisfies

$$\mathbb{P}\{|I_v| \le N - \varepsilon N\} \le \exp\big(-(K+1)N\big).$$

Next, by Lemma 3.25, for $L = L_{3.25}(K+2) > 0$ we have

$$\mathbb{P}\big\{\|\mathrm{Proj}_{I_v} Av\|_2 \geq L\sqrt{N}\big\} \leq \exp\big(-(K+2)N\big)$$

for all $v \in V$. Note that for any $u, v \in V$ the random sets I_u and I_v coincide everywhere on Ω. Hence, together with the above estimates, we get

$$\begin{aligned}&\mathbb{P}\big\{\min_{|I|\geq N-\varepsilon N}\max_{v\in V}\|\mathrm{Proj}_I Av\|_2 \geq L\sqrt{N}\big\}\\ &\leq \exp\big(-(K+1)N\big) + \mathbb{P}\big\{\max_{v\in V}\|\mathrm{Proj}_{I_v} Av\|_2 \geq L\sqrt{N}\big\}\\ &\leq \exp(-KN).\end{aligned}$$

It remains to note that for any $I \subset \{1, 2, \ldots, N\}$ and $y \in B_\infty^n$ we have

$$\|\mathrm{Proj}_I Ay\|_2 \leq \sqrt{n}\max_{v\in V}\|\mathrm{Proj}_I Av\|_2$$

everywhere on Ω. □

In the following statement, we bound the quantity (3.10) assuming that the matrix entries are symmetrically distributed. As we already mentioned above, to derive an estimate for the supremum over the sphere, we shall embed S^{n-1} into Minkowski sum of a multiple of B_∞^n and two specially chosen finite sets (see (3.12) in the proof below). This way each vector $y \in \mathrm{S}^{n-1}$ can be "decomposed" as a sum of three vectors with particular characteristics. This approach is similar to splitting the unit sphere into sets of "close to sparse" and "far from sparse" vectors introduced in [66] and subsequently used in [93], [95].

Lemma 3.29. *Let ξ be a symmetrically distributed random variable with unit variance, and let $\varepsilon \in (0,1]$. Then there are $N_{3.29} = N_{3.29}(\varepsilon) \in \mathbb{N}$ depending on ε and the distribution of ξ and $w_{3.29} = w_{3.29}(\varepsilon) > 0$ depending only on ε such that, whenever $N \geq N_{3.29}$, $n \leq N$ and $A = (a_{ij})$ is an $N \times n$ random matrix with i.i.d. entries distributed as ξ, we have*

$$\mathbb{P}\big\{\sup_{y\in \mathrm{S}^{n-1}}\min_{|I|\geq N-\varepsilon N}\|\mathrm{Proj}_I Ay\|_2 \leq C_{3.29}\sqrt{N}\big\} \geq 1 - \exp(-w_{3.29}N),$$

where $C_{3.29} > 0$ is a universal constant.

Proof. Fix $\varepsilon \in (0,1]$ and let $N_{3.29}$ be the smallest integer such that

1) $\lfloor N_{3.29}^{1/4}\rfloor \delta_{3.27}(\varepsilon/3, 2C_{3.20}) \geq 1$;

2) $N_{3.29} \geq \max\bigl(N_{3.24}(\varepsilon/3), n_{3.28}(\varepsilon/3,1)\bigr)$;

3) for all $N \geq N_{3.29}$,

$$(12eN)^{\sqrt{N}} \exp(-c_{3.24}\varepsilon N/3) + e^{-C_{3.20}N} + e^{-N} \leq \exp\bigl(-\min(c_{3.24}\varepsilon/6, C_{3.20}/2, 1/2)N\bigr).$$

Choose $N \geq N_{3.29}$. Without loss of generality, we can assume that $n = N$. Let A be as stated above.

We say that a vector $y \in \mathbb{R}^N$ is m-sparse if it has at most m non-zero coordinates. It is not difficult to verify, using Lemma 3.19, that the set of all $\sqrt{N}$-sparse vectors in $2B_2^N$ admits a $N^{-1/2}$-net $\mathcal{N}_1$ of cardinality at most $\binom{N}{\lfloor\sqrt{N}\rfloor}(6\sqrt{N})^{\sqrt{N}} \leq (12eN)^{\sqrt{N}}$. Denote

$$T = \bigl\{y \in \mathrm{S}^{N-1} :\ \|y\|_\infty \leq 1/\lfloor N^{1/4}\rfloor\bigr\}.$$

By Lemma 3.20, there is a finite subset $\mathcal{N}_2 \subset T$ of cardinality at most $\exp(C_{3.20}N)$ such that for any $y \in T$ there is $y' \in \mathcal{N}_2$ with $\|y - y'\|_\infty \leq N^{-1/2}$.

Now, we claim that

$$\mathrm{S}^{N-1} \subset \mathcal{N}_1 + \mathcal{N}_2 + \frac{2}{\sqrt{N}}B_\infty^N, \tag{3.12}$$

i.e. any vector $y = (y_1, y_2, \dots, y_N) \in \mathrm{S}^{N-1}$ can be represented as $y = y^1 + y^2 + y^3$ for some $y^1 \in \mathcal{N}_1$, $y^2 \in \mathcal{N}_2$ and $y^3 \in \frac{2}{\sqrt{N}}B_\infty^N$. Indeed, we can always find a subset $J \subset \{1, 2, \dots, N\}$ of cardinality $\lfloor\sqrt{N}\rfloor$ such that $|y_j| \leq 1/\lfloor N^{1/4}\rfloor$ whenever $j \notin J$. Denote $r = \sqrt{1 - \|y - \mathrm{Proj}_J y\|_2^2}$ and $\widetilde{y} = \mathrm{Proj}_J y - r|J|^{-1/2}\sum_{j\in J} e_j^N$. Note that $\widetilde{y}$ is $\sqrt{N}$-sparse and has the Euclidean norm at most 2, so there is $y^1 \in \mathcal{N}_1$ such that $\|\widetilde{y} - y^1\|_\infty \leq \|\widetilde{y} - y^1\|_2 \leq N^{-1/2}$. Next, the vector $y - \widetilde{y}$ satisfies $\|y - \widetilde{y}\|_2 = 1$ and $\|y - \widetilde{y}\|_\infty \leq 1/\lfloor N^{1/4}\rfloor$, i.e. $y - \widetilde{y} \in T$. Hence there is $y^2 \in \mathcal{N}_2$ such that $\|y - \widetilde{y} - y^2\|_\infty \leq N^{-1/2}$. Finally, for the vector $y^3 = y - y^1 - y^2$ we get

$$\|y - y^1 - y^2\|_\infty \leq \|\widetilde{y} - y^1\|_\infty + \|y - \widetilde{y} - y^2\|_\infty \leq \frac{2}{\sqrt{N}},$$

so $y^3 \in \frac{2}{\sqrt{N}}B_\infty^N$. This proves (3.12).

For each $y^1 \in \mathcal{N}_1$, in view of Lemma 3.24 and the condition $N \geq N_{3.24}(\varepsilon/3)$, we have

$$\mathbb{P}\Bigl\{\min_{|I|\geq N-\varepsilon N/3} \|\mathrm{Proj}_I A y^1\|_2 > 2C_{3.24}\sqrt{N}\Bigr\} \leq \exp(-c_{3.24}\varepsilon N/3).$$

Next, for every $y^2 \in \mathcal{N}_2$, Lemma 3.27, with the inequality $\lfloor N^{1/4}\rfloor\delta_{3.27}(\varepsilon/3, 2C_{3.20}) \geq 1$ and

$\|y^2\|_\infty \le 1/\lfloor N^{1/4} \rfloor$ implies that

$$\mathbb{P}\Big\{\min_{|I|\ge N-\varepsilon N/3} \|\mathrm{Proj}_I A y^2\|_2 \ge L_{3.27}\sqrt{N}\Big\} \le \exp(-2C_{3.20}N)$$

for some constant $L_{3.27} > 0$. Finally, by Lemma 3.28 and in view of the condition $N \ge n_{3.28}(\varepsilon/3, 1)$ we have

$$\mathbb{P}\Big\{\min_{|I|\ge N-\varepsilon N/3} \sup_{y\in \frac{1}{\sqrt{N}}B_\infty^N} \|\mathrm{Proj}_I A y\|_2 \ge L_{3.28}\sqrt{N}\Big\} \le \exp(-N),$$

where $L_{3.28} > 0$ is a universal constant. Let $\mathcal{E}$ denote the event

$$\begin{aligned}\mathcal{E} = \Big\{\omega \in \Omega : &\text{for every } y^1 \in \mathcal{N}_1 \text{ there is a set } I_1 = I_1(y^1) \text{ with } |I_1| \ge N - \varepsilon N/3\\ &\text{such that } \|\mathrm{Proj}_{I_1} A(\omega) y^1\|_2 \le 2C_{3.24}\sqrt{N} \textbf{ and}\\ &\text{for every } y^2 \in \mathcal{N}_2 \text{ there is a set } I_2 = I_2(y^2) \text{ with } |I_2| \ge N - \varepsilon N/3\\ &\text{such that } \|\mathrm{Proj}_{I_2} A(\omega) y^2\|_2 \le L_{3.27}\sqrt{N} \textbf{ and}\\ &\text{there is a set } I_3 \text{ with } |I_3| \ge N - \varepsilon N/3\\ &\text{such that } \sup_{y\in \frac{2}{\sqrt{N}}B_\infty^N} \|\mathrm{Proj}_{I_3} A(\omega) y\|_2 \le 2L_{3.28}\sqrt{N}\Big\}.\end{aligned}$$

Then from the above probability estimates and the definition of $N_{3.29}$ we obtain

$$\mathbb{P}(\mathcal{E}) \ge 1 - (12eN)^{\sqrt{N}} \exp(-c_{3.24}\varepsilon N/3) - \exp(-C_{3.20}N) - \exp(-N) \ge 1 - \exp(-w_{3.29}N),$$

where $w_{3.29} = \min\big(\frac{c_{3.24}\varepsilon}{6}, \frac{C_{3.20}}{2}, \frac{1}{2}\big)$.

Finally, take any $\omega \in \mathcal{E}$ and any $y \in \mathrm{S}^{N-1}$, and let $y^1 \in \mathcal{N}_1$, $y^2 \in \mathcal{N}_2$ and $y^3 \in \frac{2}{\sqrt{N}}B_\infty^N$ satisfy $y = y^1 + y^2 + y^3$. Then, by the definition of $\mathcal{E}$, there are sets $I_1, I_2, I_3 \subset \{1, 2, \ldots, N\}$ with $|I_\ell| \ge N - \varepsilon N/3$ ($\ell = 1, 2, 3$) such that

$$\begin{aligned}&\|\mathrm{Proj}_{I_1} A(\omega) y^1\|_2 \le 2C_{3.24}\sqrt{N};\\ &\|\mathrm{Proj}_{I_2} A(\omega) y^2\|_2 \le L_{3.27}\sqrt{N};\\ &\|\mathrm{Proj}_{I_3} A(\omega) y^3\|_2 \le 2L_{3.28}\sqrt{N}.\end{aligned}$$

Note that the intersection $I = I_1 \cap I_2 \cap I_3$ necessarily satisfies $|I| \ge N - \varepsilon N$, and from the last inequalities we get $\|\mathrm{Proj}_I A(\omega) y\|_2 \le (2C_{3.24} + L_{3.27} + 2L_{3.28})\sqrt{N}$. Since our choice of

$y \in \mathrm{S}^{N-1}$ and $\omega \in \mathcal{E}$ was arbitrary, we get

$$\mathbb{P}\big\{\sup_{y\in \mathrm{S}^{N-1}}\min_{|I|\geq N-\varepsilon N}\|\mathrm{Proj}_I Ay\|_2 \leq (2C_{3.24}+L_{3.27}+2L_{3.28})\sqrt{N}\big\} \geq \mathbb{P}(\mathcal{E}) \geq 1-\exp\big(-w_{3.29}N\big).$$

□

Finally, we can state the main result of the section.

Proposition 3.30. *Let ξ be a random variable with zero mean and unit variance, and let $\varepsilon \in (0,1]$. Then there are $N_{3.30} = N_{3.30}(\varepsilon) \in \mathbb{N}$ depending on ε and the distribution of ξ and $w_{3.30} = w_{3.30}(\varepsilon) > 0$ depending only on ε such that, whenever $N \geq N_{3.30}$, $n \leq N$ and $A = (a_{ij})$ is an $N \times n$ random matrix with i.i.d. entries distributed as ξ, we have*

$$\mathbb{P}\big\{\sup_{y\in \mathrm{S}^{n-1}}\min_{|I|\geq N-\varepsilon N}\|\mathrm{Proj}_I Ay\|_2 \leq C_{3.30}\sqrt{N}\big\} \geq 1-\exp(-w_{3.30}N),$$

where $C_{3.30} > 0$ is a universal constant.

Proof. Fix any $\varepsilon \in (0,1]$ and let ξ' be an independent copy of ξ. Then $\frac{1}{\sqrt{2}}(\xi-\xi')$ is symmetrically distributed and $\mathbb{E}\left(\frac{1}{\sqrt{2}}(\xi-\xi')\right)^2 = 1$. Let $N_{3.29}, w_{3.29}$ from Lemma 3.29 be defined with respect to ε and the distribution of $\frac{1}{\sqrt{2}}(\xi-\xi')$, and let $N_{3.30}$ be the smallest integer greater than $N_{3.29}$ such that $\exp(w_{3.29}N_{3.30}/2) \geq \frac{4}{3}$. Take any $N \geq N_{3.30}$ and $n \leq N$ and let A be an $N \times n$ random matrix with i.i.d. entries distributed as ξ, and A' be an independent copy of A. We can find a Borel function $f : \mathbb{R}^{N\times n} \to \mathrm{S}^{n-1}$ such that for any $B \in \mathbb{R}^{N\times n}$ we have

$$\min_{|I|\geq N-\varepsilon N}\|\mathrm{Proj}_I Bf(B)\|_2 \geq \sup_{y\in \mathrm{S}^{n-1}}\min_{|I|\geq N-\varepsilon N}\|\mathrm{Proj}_I By\|_2 - 1$$

(the term "-1" above allows us to construct a piecewise constant function f, thus avoiding any measurability questions). Then we define a random vector $\widetilde{Y} : \Omega \to \mathrm{S}^{n-1}$ as $\widetilde{Y}(\omega) = f(A(\omega))$. Conditioning on A, we obtain

$$\begin{aligned}
&\mathbb{P}\big\{\min_{|I|\geq N-\varepsilon N}\|\mathrm{Proj}_I A\widetilde{Y}\|_2 > (\sqrt{2}C_{3.29}+2)\sqrt{N} \textbf{ and } \|A'\widetilde{Y}\|_2 \leq 2\sqrt{N}\big\}\\
&\quad\geq \inf_{y\in \mathrm{S}^{n-1}}\mathbb{P}\big\{\|A'y\|_2 \leq 2\sqrt{N}\big\}\,\mathbb{P}\big\{\min_{|I|\geq N-\varepsilon N}\|\mathrm{Proj}_I A\widetilde{Y}\|_2 > (\sqrt{2}C_{3.29}+2)\sqrt{N}\big\}\\
&\quad\geq \frac{3}{4}\mathbb{P}\big\{\min_{|I|\geq N-\varepsilon N}\|\mathrm{Proj}_I A\widetilde{Y}\|_2 > (\sqrt{2}C_{3.29}+2)\sqrt{N}\big\},
\end{aligned}$$

where the estimate $\inf_{y\in \mathrm{S}^{n-1}}\mathbb{P}\big\{\|A'y\|_2 \leq 2\sqrt{N}\big\} \geq 3/4$ follows from Markov's inequality. Hence, taking into consideration that the entries of $A-A'$ are distributed as $\xi-\xi'$ and using Lemma 3.29, we get

$$
\begin{aligned}
&\mathbb{P}\big\{\sup_{y\in S^{n-1}}\min_{|I|\geq N-\varepsilon N}\|\mathrm{Proj}_I Ay\|_2 > (\sqrt{2}C_{3.29}+3)\sqrt{N}\big\}\\
&\quad\leq \mathbb{P}\big\{\min_{|I|\geq N-\varepsilon N}\|\mathrm{Proj}_I A\widetilde{Y}\|_2 > (\sqrt{2}C_{3.29}+2)\sqrt{N}\big\}\\
&\quad\leq \frac{4}{3}\mathbb{P}\big\{\min_{|I|\geq N-\varepsilon N}\|\mathrm{Proj}_I A\widetilde{Y}\|_2 > (\sqrt{2}C_{3.29}+2)\sqrt{N}\ \textbf{and}\ \|A'\widetilde{Y}\|_2\leq 2\sqrt{N}\big\}\\
&\quad\leq \frac{4}{3}\mathbb{P}\big\{\min_{|I|\geq N-\varepsilon N}\|\mathrm{Proj}_I (A-A')\widetilde{Y}\|_2 > \sqrt{2}C_{3.29}\sqrt{N}\big\}\\
&\quad\leq \frac{4}{3}\mathbb{P}\big\{\sup_{y\in S^{n-1}}\min_{|I|\geq N-\varepsilon N}\|\mathrm{Proj}_I (A-A')y\|_2 > \sqrt{2}C_{3.29}\sqrt{N}\big\}\\
&\quad\leq \frac{4}{3}\exp(-w_{3.29}N)\\
&\quad\leq \exp(-w_{3.29}N/2).
\end{aligned}
$$

□

3.2.4 Matrix truncation and proof of Theorem 3.18

In the next statement, we compare the n-th largest singular value of a random $N\times n$ matrix A with bounded entries to $s_{\min}(\mathrm{Proj}_I A)$. Obviously,

$$s_{\min}(\mathrm{Proj}_I A)\leq s_{\min}(A)\ \text{ for any } I\subset\{1,2,\ldots,N\}.$$

We will need an inequality in the opposite direction when $|I|/N\approx 1$. A theorem of A. Litvak, A. Pajor, M. Rudelson and N. Tomczak-Jaegermann [66, Theorem 3.1] implies that for any $\delta>1$ and $M>0$ there are $h>0$ and $\varepsilon>0$ depending only on δ and M with the following property: whenever $N\geq\delta n$ and A is an $N\times n$ random matrix with i.i.d. entries with mean zero, variance one and a.s. bounded by M, we have

$$\mathbb{P}\big\{\min_{|I|\geq N-\varepsilon N}s_{\min}(\mathrm{Proj}_I A)\geq h\sqrt{N}\big\}\geq 1-2\exp(\varepsilon N).$$

This, together with an upper bound for $s_{\min}(A)$, gives an estimate

$$s_{\min}(A)\leq L\min_{|I|\geq N-\varepsilon N}s_{\min}(\mathrm{Proj}_I A)$$

with a large probability, where $L>0$ depends only on δ and M. However, such an estimate would be insufficient for our needs, and we shall apply a more direct argument to get a stronger relation.

Proposition 3.31. *Let ξ be a random variable with zero mean such that $|\xi| \le M$ a.s. for some $M > 0$. For any $\eta > 0$ there are $\varepsilon_{3.31} = \varepsilon_{3.31}(\eta, M) > 0$ and $N_{3.31} = N_{3.31}(\eta, M) \in \mathbb{N}$ (both depending only on η and M) with the following property: whenever $N \ge N_{3.31}$, $n \le N$ and $A = (a_{ij})$ is an $N \times n$ random matrix with i.i.d. entries distributed as ξ, we have*

$$\mathbb{P}\Big\{s_{\min}(A) \le \min_{|I| \ge N - \varepsilon_{3.31} N} s_{\min}(\mathrm{Proj}_I A) + \eta\sqrt{N}\Big\} \ge 1 - \exp(-\varepsilon_{3.31} N).$$

Proof. Fix any $\eta > 0$, let $\varepsilon = \varepsilon_{3.31}(\eta, M)$ be the largest number in $(0, 1]$ satisfying

$$\frac{c_{3.21}}{2M^2}\eta^2 \ge \varepsilon\big(1 + \ln\frac{6e}{\varepsilon}\big),$$

and $N_{3.31} \in \mathbb{N}$ be the smallest number such that $N - \lceil N - \varepsilon N\rceil \ge \varepsilon N/2$ for all $N \ge N_{3.31}$.

Let $N \ge N_{3.31}$, $n \le N$ and A be an $N \times n$ random matrix defined as above. We shall prove the statement by contradiction. Let us assume that

$$\mathbb{P}\Big\{s_{\min}(A) > \min_{|I| \ge N - \varepsilon N} s_{\min}(\mathrm{Proj}_I A) + \eta\sqrt{N}\Big\} > \exp(-\varepsilon N).$$

Cardinality of the set $T = \big\{I \subset \{1, 2, \dots, N\} : |I| = \lceil N - \varepsilon N\rceil\big\}$ can be estimated as

$$|T| \le \binom{N}{\lceil N - \varepsilon N\rceil} \le \Big(\frac{eN}{N - \lceil N - \varepsilon N\rceil}\Big)^{N - \lceil N - \varepsilon N\rceil} \le \Big(\frac{2e}{\varepsilon}\Big)^{\varepsilon N}.$$

Hence, our assumption implies that there is a set $I_0 \in T$ such that

$$\mathbb{P}\Big\{s_{\min}(A) > s_{\min}(\mathrm{Proj}_{I_0} A) + \eta\sqrt{N}\Big\} > \exp(-\varepsilon N)\Big(\frac{2e}{\varepsilon}\Big)^{-\varepsilon N}. \tag{3.13}$$

Let $f : \mathbb{R}^{N\times n} \to \mathrm{S}^{n-1}$ be a Borel function such that for every $B \in \mathbb{R}^{N\times n}$, $f(B) \in \mathrm{S}^{n-1}$ is an eigenvector of $B^T B$ corresponding to its smallest eigenvalue. So, we have $\|Bf(B)\|_2 = s_{\min}(B)$. Then we define a random vector $\widetilde{Y} : \Omega \to \mathrm{S}^{n-1}$ as $\widetilde{Y}(\omega) = f(\mathrm{Proj}_{I_0} A(\omega))$. It is not difficult to see that such a definition implies that $\widetilde{Y}$ and a_{ij} ($i \notin I_0$, $1 \le j \le n$) are jointly independent. Hence,

$$\begin{aligned}\mathbb{P}\Big\{s_{\min}(A) > s_{\min}(\mathrm{Proj}_{I_0} A) + \eta\sqrt{N}\Big\} &\le \mathbb{P}\Big\{\|A\widetilde{Y}\|_2 > \|\mathrm{Proj}_{I_0} A\widetilde{Y}\|_2 + \eta\sqrt{N}\Big\}\\ &\le \mathbb{P}\Big\{\|\mathrm{Proj}_{\{1,2,\dots,N\}\setminus I_0} A\widetilde{Y}\|_2 > \eta\sqrt{N}\Big\}\\ &\le \sup_{y \in \mathrm{S}^{n-1}} \mathbb{P}\Big\{\|\mathrm{Proj}_{\{1,2,\dots,N\}\setminus I_0} Ay\|_2 > \eta\sqrt{N}\Big\}.\end{aligned}$$

Now, for every $y = (y_1, y_2, \dots, y_n) \in \mathrm{S}^{n-1}$, Lemma 3.21 and the standard procedure with the Laplace transform give for $\lambda = \frac{c_{3.21}}{2M^2}$:

$$\begin{aligned}
&\mathbb{P}\big\{\|\mathrm{Proj}_{\{1,2,\ldots,N\}\setminus I_0} Ay\|_2 > \eta\sqrt{N}\big\}\\
&\quad= \mathbb{P}\Big\{\sum_{i\notin I_0}\Big(\sum_{j=1}^{n} a_{ij}y_j\Big)^2 > \eta^2 N\Big\}\\
&\quad\le \frac{\Big(\mathbb{E}\,\exp\big(\lambda\big(\sum_{j=1}^{n} a_{1j}y_j\big)^2\big)\Big)^{N-\lceil N-\varepsilon N\rceil}}{\exp(\lambda\eta^2 N)}\\
&\quad= \exp(-\lambda\eta^2 N)\Big(1+\int_1^\infty \mathbb{P}\Big\{\Big|\sum_{j=1}^{n} a_{1j}y_j\Big| \ge \sqrt{\ln\tau/\lambda}\Big\}\,d\tau\Big)^{N-\lceil N-\varepsilon N\rceil}\\
&\quad\le \exp(-\lambda\eta^2 N)\Big(1+2\int_1^\infty \exp\Big(-\frac{c_{3.21}\ln\tau}{\lambda M^2}\Big)\,d\tau\Big)^{N-\lceil N-\varepsilon N\rceil}\\
&\quad= \exp(-\lambda\eta^2 N)\,3^{N-\lceil N-\varepsilon N\rceil}\\
&\quad\le \exp\big(-\lambda\eta^2 N + \varepsilon N\ln 3\big).
\end{aligned}$$

Together with (3.13), the last estimate implies

$$-\lambda\eta^2 + \varepsilon\ln 3 > -\varepsilon - \varepsilon\ln\frac{2e}{\varepsilon}.$$

However, this contradicts our choice of ε. Thus, the initial assumption was wrong, and the statement is proved. □

Let ξ be a random variable with zero mean. Then for any $M > 0$ we call the variable

$$\xi\chi_{\{|\xi|\le M\}} - \mathbb{E}\,(\xi\chi_{\{|\xi|\le M\}})$$

the centered M-truncation of ξ. Here, $\chi_{\{|\xi|\le M\}}$ is the indicator of the event $\big\{\omega\in\Omega : |\xi(\omega)|\le M\big\}$.

Denote $\widetilde{\xi}_M = \xi\chi_{\{|\xi|\le M\}} - \mathbb{E}\,(\xi\chi_{\{|\xi|\le M\}})$ and $\theta_M = \xi - \widetilde{\xi}_M = \xi\chi_{\{|\xi|>M\}} + \mathbb{E}\,(\xi\chi_{\{|\xi|\le M\}})$. Obviously, $\mathbb{E}\,\widetilde{\xi}_M = \mathbb{E}\,\theta_M = 0$ and $|\widetilde{\xi}_M| \le 2M$ everywhere on Ω for any $M > 0$. Further, if the second moment of ξ is bounded then

$$\begin{aligned}
&\mathbb{E}\,\widetilde{\xi}_M^2 = \mathbb{E}\,(\xi\chi_{\{|\xi|\le M\}})^2 - \big(\mathbb{E}\,(\xi\chi_{\{|\xi|\le M\}})\big)^2 \longrightarrow \mathbb{E}\,\xi^2 \ \text{ and}\\
&\mathbb{E}\,{\theta_M}^2 = \mathbb{E}\,\xi^2 - 2\mathbb{E}\,\big(\xi^2\chi_{\{|\xi|\le M\}}\big) + \mathbb{E}\,\widetilde{\xi}_M^2 \longrightarrow 0 \ \text{ when } M\to\infty.
\end{aligned}$$

Theorem 3.32. *Let ξ be a random variable with zero mean and unit variance. For any $M > 0$ and $\eta > 0$ there are $N_{3.32}\in\mathbb{N}$ depending on M,η and the distribution of ξ, and $w_{3.32} > 0$ depending only on M and η with the following property: Let $N \ge N_{3.32}$, $n \le N$*

and let $A = (a_{ij})$ be an $N \times n$ random matrix with i.i.d. entries distributed as ξ. Further, let $\widetilde{A}$ be an $N \times n$ matrix with the entries $\widetilde{a}_{ij} = a_{ij}\chi_{\{|a_{ij}| \le M\}} - \mathbb{E}\,(a_{ij}\chi_{\{|a_{ij}| \le M\}})$ and denote $\theta = \xi\chi_{\{|\xi| > M\}} + \mathbb{E}\,(\xi\chi_{\{|\xi| \le M\}})$. Then

$$\mathbb{P}\big\{s_{\min}(A) \ge s_{\min}(\widetilde{A}) - \eta\sqrt{N} - C_{3.30}\sqrt{N\mathbb{E}\,\theta^2}\big\} \ge 1 - \exp(-w_{3.32}N).$$

Proof. Fix any $M > 0$ and $\eta > 0$ and let θ be as above. We will assume that $\mathbb{P}\{\theta = 0\} < 1$; otherwise the truncation leaves the variable unchanged and there is nothing to prove. Let $N_{3.31} = N_{3.31}(\eta, 2M)$ and $\varepsilon = \varepsilon_{3.31}(\eta, 2M)$ be taken from Proposition 3.31. Let also $N_{3.30}$ and $w_{3.30}$ be defined as in Proposition 3.30 with respect to ε and the distribution of the "normalized tail" $\theta/\sqrt{\mathbb{E}\,\theta^2}$. Now, let $N_{3.32}$ be the smallest integer greater than $\max(N_{3.31}, N_{3.30})$ such that for all $N \ge N_{3.32}$ we have

$$\exp(-\varepsilon N) + \exp(-w_{3.30}N) \le \exp\big(-\min(\varepsilon/2, w_{3.30}/2)N\big).$$

Take any $N \ge N_{3.32}$, $n \le N$, and let $A, \widetilde{A}$ be as stated above. By Proposition 3.31, we have

$$\mathbb{P}\big\{s_{\min}(\widetilde{A}) > \min_{|I| \ge N - \varepsilon N} s_{\min}(\mathrm{Proj}_I\widetilde{A}) + \eta\sqrt{N}\big\} \le \exp(-\varepsilon N),$$

and, by Proposition 3.30,

$$\mathbb{P}\big\{\sup_{y \in \mathrm{S}^{n-1}} \min_{|I| \ge N - \varepsilon N} \|\mathrm{Proj}_I(A - \widetilde{A})y\|_2 > C_{3.30}\sqrt{N\mathbb{E}\,\theta^2}\big\} \le \exp(-w_{3.30}N).$$

Combining the two relations, we get

$$\begin{aligned}
&\mathbb{P}\big\{s_{\min}(A) < s_{\min}(\widetilde{A}) - \eta\sqrt{N} - C_{3.30}\sqrt{N\mathbb{E}\,\theta^2}\big\}\\
&\quad\le \mathbb{P}\big\{s_{\min}(\widetilde{A}) > \min_{|I| \ge N - \varepsilon N} s_{\min}(\mathrm{Proj}_I\widetilde{A}) + \eta\sqrt{N}\big\}\\
&\qquad + \mathbb{P}\big\{s_{\min}(A) < \min_{|I| \ge N - \varepsilon N} s_{\min}(\mathrm{Proj}_I\widetilde{A}) - C_{3.30}\sqrt{N\mathbb{E}\,\theta^2}\big\}\\
&\quad\le \exp(-\varepsilon N)\\
&\qquad + \mathbb{P}\big\{\exists y \in \mathrm{S}^{n-1} :\ \min_{|I| \ge N - \varepsilon N} s_{\min}(\mathrm{Proj}_I\widetilde{A}) - \|Ay\|_2 > C_{3.30}\sqrt{N\mathbb{E}\,\theta^2}\big\}\\
&\quad\le \exp(-\varepsilon N)\\
&\qquad + \mathbb{P}\big\{\exists y \in \mathrm{S}^{n-1} :\ \min_{|I| \ge N - \varepsilon N} (\|\mathrm{Proj}_I\widetilde{A}y\|_2 - \|\mathrm{Proj}_I Ay\|_2) > C_{3.30}\sqrt{N\mathbb{E}\,\theta^2}\big\}\\
&\quad\le \exp(-\varepsilon N) + \exp(-w_{3.30}N)\\
&\quad\le \exp\big(-\min(\varepsilon/2, w_{3.30}/2)N\big).
\end{aligned}$$

□

Proof of Theorem 3.18. Let $\{a_{ij}\}$ $(1 \le i,j < \infty)$ be a two-dimensional array of i.i.d. random variables with zero mean and unit variance and let $(N_m)_{m=1}^{\infty}$ be an integer sequence satisfying $m/N_m \longrightarrow z$ for some $z \in (0,1)$. Recall that for every $m \in \mathbb{N}$, A_m denotes the random $N_m \times m$ matrix with entries a_{ij} $(1 \le i \le N_m, 1 \le j \le m)$. The Marčenko–Pastur law (see Theorem 3.23 and Remark 3.3) implies that

$$\limsup_{m\to\infty} \frac{s_{\min}(A_m)}{\sqrt{N_m}} \le 1 - \sqrt{z} \text{ almost surely.}$$

Thus, it suffices to prove the lower estimate

$$\liminf_{m\to\infty} \frac{s_{\min}(A_m)}{\sqrt{N_m}} \ge 1 - \sqrt{z} \text{ a.s.}$$

Now, choose arbitrary $\eta > 0$ and let $M > 0$ be such that

$$\mathbb{E}\left(a_{11}\chi_{\{|a_{11}|\le M\}} - \mathbb{E}\left(a_{11}\chi_{\{|a_{11}|\le M\}}\right)\right)^2 \ge (1-\eta)^2 \text{ and}$$
$$\mathbb{E}\left(a_{11}\chi_{\{|a_{11}|> M\}} + \mathbb{E}\left(a_{11}\chi_{\{|a_{11}|\le M\}}\right)\right)^2 \le \eta^2.$$

For every $m \in \mathbb{N}$, let $\widetilde{A}_m$ be the $N_m \times m$ matrix of truncated and centered variables $\widetilde{a}_{ij} = a_{ij}\chi_{\{|a_{ij}|\le M\}} - \mathbb{E}\left(a_{ij}\chi_{\{|a_{ij}|\le M\}}\right)$ $(1 \le i \le N_m, 1 \le j \le m)$. Theorem 3.32 and the conditions on the sequence $(N_m)_{m=1}^{\infty}$ imply that there are $m_0 \in \mathbb{N}$ and $w > 0$ such that for all $k \ge m_0$

$$\mathbb{P}\Big\{s_{\min}(A_m) \ge s_{\min}(\widetilde{A}_m) - (1 + C_{3.30})\eta\sqrt{N_m} \text{ for all } m \ge k\Big\} \ge 1 - \sum_{m=k}^{\infty} \exp(-wN_m),$$

where the quantity on the right-hand side goes to 1 as k tends to infinity. Hence, we obtain

$$\mathbb{P}\Big\{\liminf_{m\to\infty} \frac{s_{\min}(A_m)}{\sqrt{N_m}} \ge \liminf_{m\to\infty} \frac{s_{\min}(\widetilde{A}_m)}{\sqrt{N_m}} - (1 + C_{3.30})\eta\Big\} = 1.$$

On the other hand, the theorem of Bai and Yin [8] implies that $\lim\limits_{m\to\infty} \frac{s_{\min}\left(\widetilde{A}_m\right)}{\sqrt{N_m}} \ge (1-\eta)(1-\sqrt{z})$ a.s. Thus, we come to the estimate

$$\liminf_{m\to\infty} \frac{s_{\min}(A_m)}{\sqrt{N_m}} \ge (1-\eta)(1-\sqrt{z}) - (1 + C_{3.30})\eta \text{ a.s.}$$

Since $\eta > 0$ was arbitrary, this proves the result. □

Chapter 4

When does a discrete-time random walk in $\mathbb{R}^n$ absorb the origin into its convex hull?[1]

4.1 Introduction

The goal of this Chapter is to study certain convexity aspects of high-dimensional random walks. Given a discrete-time random walk $W(i)$ with values in $\mathbb{R}^n$, we are interested in estimating the number of steps N when the origin becomes an interior point of the convex hull of $\{W(i)\}_{i\leq N}$. This question was raised by I. Benjamini and considered by R. Eldan in [29]. Three models of random walks are treated here: a walk given by a discretization of the standard Brownian motion in $\mathbb{R}^n$, the standard random walk on $\mathbb{Z}^n$ and a random walk on the unit sphere S^{n-1}. We employ a novel approach that reduces the problem to a question about certain geometric properties of random matrices. Random matrix theory has strong connections with asymptotic geometric analysis (see, for example, [19] and [116]); in particular, random matrices appear in Gordon's escape theorem [42] and in various estimates of diameters of random sections of convex sets [77], [83]. The interconnection between random walks, random matrix theory and high-dimensional convex geometry is at the heart of our work.

The *standard Brownian motion* $\mathrm{BM}_1(t)$ ($t\in[0,\infty)$) with values in $\mathbb{R}$ is defined as a centered Gaussian process, such that the covariance $\mathrm{cov}\,(\mathrm{BM}_1(t),\ \mathrm{BM}_1(s)) = \min(t,s)$ for all $t,s\in[0,\infty)$. The Brownian motion in $\mathbb{R}^n$, denoted by BM_n, is a vector of n

[1] A version of this Chapter is accepted for publication in the Annals of Probability. K. Tikhomirov and P.Youssef, When does a discrete-time random walk in $\mathbb{R}^n$ absorb the origin into its convex hull? 2015, arXiv:1410.0458

independent one-dimensional Brownian motions. We refer the reader to [81] for extensive information on the process BM_n. Various properties of the convex hull of the Brownian motion in high dimensions were studied recently in [29], [30] and [54]; in particular, results on interior and extremal points of the convex hulls were obtained. It is easy to see that the interior of $\mathrm{conv}\{\mathrm{BM}_n(t) : 0 < t < 1\}$ (with "conv" denoting the convex hull) contains the origin almost surely. In the case when the domain $t \in (0, 1)$ is replaced by a finite subset of the unit interval, the origin is outside of the convex hull with a non-zero probability. Our work is motivated by the following problem which in a more specific form was considered by Eldan in [29]:

Let $t_1 < t_2 < \cdots < t_N$ be points in $[0, 1]$. How is the probability that the origin belongs to the interior of $\mathrm{conv}\{\mathrm{BM}_n(t_i) : i \leq N\}$ related to the structure of the set $\{t_i\}_{i \leq N}$?

In [29], the numbers N and $t_1, t_2, \ldots, t_N$ were generated by a homogeneous Poisson point process in $[0, 1]$. It was shown that when the expected number of generated points N is greater than $e^{Cn\log(n)}$, the origin belongs to the interior of $\mathrm{conv}\{\mathrm{BM}_n(t_i) : i \leq N\}$ with high probability [29, Theorem 3.1]. A related result of [29] dealing with the standard walk on $\mathbb{Z}^n$ states that, with probability close to one, $e^{Cn\log(n)}$ steps are sufficient for the convex hull of the walk to absorb the origin. It was not clear, however, whether the bound $e^{Cn\log(n)}$ was sharp. This question is addressed in the first main theorem of this Chapter:

Theorem 4.1. *There exists a constant $C > 0$ such that for any $n \in \mathbb{N}$ and $N \geq \exp(Cn)$ the following holds.*

- *Setting $t_i := i/N$, $i = 1, 2, \ldots, N$, the set $\mathrm{conv}\{BM_n(t_i), i \leq N\}$ contains the origin in its interior with probability at least $1 - \exp(-n)$.*
- *The convex hull of the first N steps of the standard random walk on $\mathbb{Z}^n$ starting at 0, contains the origin in its interior with probability at least $1 - \exp(-n)$.*

The first part of this theorem also holds when $\{t_i\}$ is a homogeneous Poisson process in $[0, 1]$ of intensity at least $\exp(Cn)$. Therefore, our result is strictly stronger than the bound proved in [29].

Let us discuss optimality of the estimates in Theorem 4.1. Regarding the second assertion, it was proved in [29] that if the number of steps N is less than $\exp(cn/\log n)$ then with probability close to one the origin does not belong to the interior of the convex hull of the standard walk on $\mathbb{Z}^n$.

For the first assertion of Theorem 4.1, we prove that it is optimal in the sense that the number of points N must be exponential in n in order to have, say, $\mathbb{P}\{0 \in \mathrm{conv}\{\mathrm{BM}_n(t_i), i \leq N\}\} \geq 1/2$. More precisely, we prove the following:

Theorem 4.2. *There exist universal constants $c > 0$ and $n_0 \in \mathbb{N}$ with the following property: let $n \geq n_0$ and $\mathrm{BM}_n(t)$ $(0 \leq t < \infty)$ be the standard Brownian motion in $\mathbb{R}^n$. Then*

$$\mathbb{P}\big\{0 \in \mathrm{conv}\{\mathrm{BM}_n(t) : t \in [1, 2^{cn}]\}\big\} \leq \frac{1}{n}.$$

Remark 4.1. The bound $\frac{1}{n}$ in the above theorem can be replaced with $\frac{1}{n^L}$ for any constant $L > 0$ at expense of decreasing c and increasing n_0.

The statement of Theorem 4.2 is equivalent to the estimate

$$\mathbb{P}\big\{\inf_{u \in \mathrm{S}^{n-1}} \sup_{t \in [1, 2^{cn}]} \langle u, \mathrm{BM}_n(t)\rangle_n < 0\big\} \geq 1 - \frac{1}{n}, \tag{4.1}$$

where the quantity in the brackets is the "minimax" of 1-dimensional Gaussian process $\langle u, \mathrm{BM}_n(t)\rangle_n$ indexed over $\mathrm{S}^{n-1} \times [1, 2^{cn}]$. We note that a comparison theorem for the minimax of doubly indexed Gaussian processes was obtained in [43] (see also [63, Corollary 3.13 and Theorem 3.16]), and was the central ingredient in proving the escape theorem of [42].

The second main result of this Chapter deals with discrete-time random walks on the sphere. For any $\theta \in (0, \pi/2)$, we consider a Markov chain W_θ with values in S^{n-1} such that the angle between two consecutive steps is θ (i.e. $\langle W_\theta(j), W_\theta(j+1)\rangle_n = \cos\theta$, $j \in \mathbb{N}$) and the direction from $W(j)$ to $W(j+1)$ is chosen *uniformly at random* in the sense that for any $u \in \mathrm{S}^{n-1}$, the distribution of $W_\theta(j+1)$ conditioned on $W_\theta(j) = u$ is uniform on the $(n-2)$-sphere $\mathrm{S}^{n-1} \cap \{x \in \mathbb{R}^n : \langle x, u\rangle_n = \cos\theta\}$.

Theorem 4.3. *For any $\theta \in (0, \pi/2)$, there exist $L = L(\theta)$ and $n_0 = n_0(\theta)$ depending only on θ such that the following holds: Let $n \geq n_0$ and W_θ be the process with values in S^{n-1} described above. Then for all $N \geq Ln$ we have*

$$\mathbb{P}\big\{0 \text{ belongs to } \mathrm{conv}\{W_\theta(i) : i \leq N\}\big\} \geq 1 - \exp(-n).$$

It follows from dimension considerations that the estimate of the number of steps is optimal up to a factor depending only on θ. We note here that a related problem for the standard spherical Brownian motion was studied in [29].

Let us outline the main ideas behind the proofs of Theorems 4.1 and 4.3. The following simple observation relates the question about random walks to a problem dealing with random matrices:

Let $X(t)$ ($t \in [0, \infty)$ or $t \in \mathbb{N} \cup \{0\}$) be a random process with values in $\mathbb{R}^n$, with $X(0) = 0$; let $0 = t_0 < t_1 < \cdots < t_N$ be a collection of non-random points and assume that the increments $X(t_i) - X(t_{i-1})$ are independent. Define A as the $N \times n$ random matrix with independent rows obtained by appropriately rescaling the increments $X(t_i) - X(t_{i-1})$, $i = 1, 2, \ldots, N$. Then there exists a non-random $N \times N$ lower-triangular matrix F such that the rows of FA are precisely $X(t_i)$, $i = 1, 2, \ldots, N$. Thus, we can restate our problem about the convex hull of $X(t_i)$'s in terms of certain properties of the matrix FA. Namely, the convex hull of $X(t_i)$'s contains the origin in its interior if and only if for any unit vector y in $\mathbb{R}^n$, the vector FAy has at least one negative coordinate. Geometrically, this problem is reduced to estimating the probability that the image of A *escapes* (i.e. does not intersect) the set $F^{-1}(\mathbb{R}_+^N) \cap \mathrm{S}^{N-1}$, where $\mathbb{R}_+^N$ denotes the cone of positive vectors. For the standard Brownian motion, A is the $N \times n$ standard Gaussian matrix. In this case, we apply Gordon's escape theorem [42] which estimates the probability that a random subspace uniformly distributed on the Grassmannian does not intersect with a given subset of S^{N-1}. In a more general case, when the image of A is not uniformly distributed, Gordon's theorem cannot be applied. To treat that scenario, we prove a statement which can be seen as an extension of Gordon's theorem to a broad class of random matrices, however, with considerable restrictions on the subsets of S^{N-1}:

Theorem 4.4. *For any $\tau, \delta \in (0, 1]$ and any $K > 1$, there exist L and $\eta > 0$ depending only on τ, δ and K with the following property: Let $N \geq Ln$ and let A be an $N \times n$ random matrix with independent rows $(R_i)_{i \leq N}$ satisfying*

$$\mathbb{P}\{\langle R_i, y\rangle_n < -\tau\} \geq \delta, \text{ for any } y \in \mathrm{S}^{n-1} \text{ and any } i \leq N.$$

Then for any $N \times N$ random matrix F, matrix FA satisfies

$$\mathbb{P}\big\{\exists y \in \mathrm{S}^{n-1},\ FAy \in \mathbb{R}_+^N\big\} \leq \exp(-\delta^2 N/4) + \mathbb{P}\big\{\|A\|_{2\to 2} > K\sqrt{N}\big\} + \mathbb{P}\big\{\|F - \mathbf{Id}_N\|_{2\to 2} > \eta\big\}.$$

We use this result to deal with the random walk on $\mathbb{Z}^n$. For the random walks W_θ on the sphere we follow, with some modifications, the same scheme as for processes in $\mathbb{R}^n$ with independent increments.

The Chapter is organized as follows: Section 4.2 contains preliminaries. Results about random matrices are given in Section 4.3, while corollaries for the Brownian motion and the standard random walk on $\mathbb{Z}^n$ are stated in Section 4.4. In Section 4.5, we consider random walks on the sphere. Finally, we prove Theorem 4.2 in Section 4.6.

4.2 Preliminaries

In this section we state some auxiliary facts. For $N \geq n$ and an $N \times n$ matrix A, let $s_{\max}(A)$ and $s_{\min}(A)$ be its largest and smallest singular values, respectively, i.e. $s_{\max}(A) = \|A\|_{2\to 2}$ (the operator norm of A) and $s_{\min}(A) = \inf\limits_{y \in S^{n-1}} \|Ay\|_2$. When A is an $N \times N$ invertible matrix, the condition number of A is $\|A\|_{2\to 2} \cdot \|A^{-1}\|_{2\to 2}$. Note that the condition number is equal to the ratio of the largest and the smallest singular values of A.

Throughout this Chapter, g denotes a standard Gaussian variable. The following estimate is well known (see, for example, [32, Lemma VII.1.2]):

$$\mathbb{P}\{g \geq t\} = \frac{1}{\sqrt{2\pi}} \int_t^\infty \exp(-r^2/2)\, dr < \frac{1}{\sqrt{2\pi}t} \exp(-t^2/2), \quad t > 0. \tag{4.2}$$

A centered random vector X in $\mathbb{R}^n$ is *isotropic* if $\mathbb{E}\, X = 0$ and the covariance matrix of X is the identity i.e. $\mathbb{E}\, XX^t = \mathbf{Id}_n$. The *standard Gaussian vector* Y in $\mathbb{R}^n$ is a random vector with i.i.d. coordinates having the same law as g. As a corollary of a concentration inequality for Gaussian variables (see [86, Theorem 4.7] or [80, Theorem V.1]), we have for any $\varepsilon > 0$:

$$\mathbb{P}\big\{(1-\varepsilon)\sqrt{n} \leq \|Y\|_2 \leq (1+\varepsilon)\sqrt{n}\big\} \geq 1 - 2\exp(-\tilde{c}\varepsilon^2 n) \tag{4.3}$$

for a universal constant $\tilde{c} > 0$. An $N \times n$ matrix is called the *standard Gaussian matrix* if its entries are i.i.d. having the same law as g. We denote this matrix by G (and recall that $N \geq n$). Then for any $t \geq 0$ we have

$$\begin{aligned} &\mathbb{P}\big\{\sqrt{N} - \sqrt{n} - t \leq s_{\min}(G) \leq s_{\max}(G) \leq \sqrt{N} + \sqrt{n} + t\big\} \\ &\quad \geq 1 - 2\exp(-t^2/2) \end{aligned} \tag{4.4}$$

(see, for example, [117, Corollary 5.35]).

Given a vector $x \in \mathbb{R}^N$, we denote by x_+ and x_- its positive and negative part, respectively, i.e.

$$x_+ = \sum_{i=1}^N \max(0, \langle x, e_i^N \rangle_N)\, e_i^N \quad \text{and} \quad x_- = \sum_{i=1}^N \max(0, -\langle x, e_i^N \rangle_N)\, e_i^N.$$

The following simple observation will be useful in the proof of the main theorems.

Lemma 4.5. *Let $x, y \in \mathbb{R}^N$. Then $\|x_-\|_2 \geq \|y_-\|_2 - \|x - y\|_2$.*

Proof. Writing $x = x_+ - x_-$ and $y = y_+ - y_-$, we obtain

$$\begin{aligned}\|x - y\|_2^2 &= \|(x_+ - y_+) - (x_- - y_-)\|_2^2 \\ &= \|x_- - y_-\|_2^2 + \|x_+ - y_+\|_2^2 - 2\langle x_+ - y_+, x_- - y_-\rangle_N \\ &\geq \|x_- - y_-\|_2^2 \\ &\geq \left(\|y_-\|_2 - \|x_-\|_2\right)^2,\end{aligned}$$

where the first inequality in the above formula holds since $\langle x_+ - y_+, x_- - y_-\rangle_N$ is non-positive. □

Given a compact set $S \subset \mathbb{R}^N$, the *Gaussian width* of S is defined by

$$w(S) := \mathbb{E} \sup_{x \in S} \langle Y, x\rangle_N,$$

where Y is the standard Gaussian vector in $\mathbb{R}^N$ (see [17], [21], [116]). The following is a consequence of Urysohn's inequality (see, for example, Corollary 1.4 in [86]) and the relation between the Gaussian and the mean width:

$$\sqrt{N-1}\left(\frac{\mathrm{Vol}_N(S)}{\mathrm{Vol}_N(B_2^N)}\right)^{1/N} \leq w(S). \tag{4.5}$$

Given a convex cone C in $\mathbb{R}^N$, the *polar cone* C^* of C is defined by

$$C^* := \{x \in \mathbb{R}^N,\ \langle x, y\rangle_N \leq 0 \ \text{ for any } y \in C\}.$$

The next Lemma provides a useful relation between the Gaussian widths of the parts of a convex cone and its polar enclosed in the unit Euclidean ball. The lemma is proved in [21] for intersections of cones with the unit sphere (see [21, Lemma 3.7]); we put it here in a version more convenient for us.

Lemma 4.6. *Let $C \subset \mathbb{R}^N$ be a nonempty closed convex cone. Then*

$$w\left(C \cap B_2^N\right)^2 + w\left(C^* \cap B_2^N\right)^2 \leq N.$$

Proof. For any $x \in \mathbb{R}^N$, let $P_C x := \arg\inf_{y \in C} \|x - y\|_2$ be the projection of x onto C. It can be checked that each vector $x \in \mathbb{R}^N$ can be decomposed as

$$x = P_C x + P_{C^*} x, \tag{4.6}$$

with $\langle P_C x, P_{C^*} x\rangle_N = 0$. As before, let Y be the standard Gaussian vector in $\mathbb{R}^N$. Having decomposition (4.6) in mind, we can write

$$w(C \cap B_2^N) = \mathbb{E} \sup_{x \in C \cap B_2^N} \langle Y, x \rangle_N \leq \mathbb{E} \sup_{x \in C \cap B_2^N} \langle P_C Y, x \rangle_N,$$

where the last inequality holds since $\langle P_{C^*} Y, x \rangle_N \leq 0$ for all $x \in C$. We deduce that

$$w(C \cap B_2^N) \leq \mathbb{E}\, \|P_C Y\|_2. \tag{4.7}$$

Now using the decomposition (4.6) and the above inequality, we obtain

$$w(C \cap B_2^N)^2 \leq \mathbb{E}\, \|P_C Y\|_2^2 = \mathbb{E}\, \|Y\|_2^2 - \mathbb{E}\, \|P_{C^*} Y\|_2^2 = N - \mathbb{E}\, \|P_{C^*} Y\|_2^2. \tag{4.8}$$

Note that (4.7) applied to the cone C^* yields $w(C^* \cap B_2^N)^2 \leq \mathbb{E}\, \|P_{C^*} Y\|_2^2$. Plugging it into (4.8), we complete the proof. □

4.3 Escape theorems for random matrices

In this section, we estimate the probability that the image of a random $N \times n$ matrix A *escapes* the intersection of a given cone with the unit sphere S^{N-1} (we shall restrict ourselves to considering a special family of convex cones in $\mathbb{R}^N$). Similar questions have attracted considerable attention recently in connection with the theory of compressed sensing [17].

Given a closed subset $S \subset \mathrm{S}^{N-1}$, the problem of estimating the probability $\mathbb{P}\{\mathrm{Im}(A) \cap S = \emptyset\}$ can be treated in different ways. One may look at it as the question of bounding the diameter of the random section $\mathrm{conv}(S, -S) \cap \mathrm{Im}(A)$ of the convex set $\mathrm{conv}(S, -S)$: clearly, $\mathrm{Im}(A) \cap S = \emptyset$ if and only if $\mathrm{diam}\big(\mathrm{conv}(S, -S) \cap \mathrm{Im}(A)\big) < 2$. The study of random sections of convex sets is a central theme in the area of asymptotic geometric analysis and its importance has been highlighted in Milman's proof of Dvoretzky's theorem [80], [86]. The question of estimating diameters of random sections of proportional dimension was originally considered in [77] and [83] in the case when the corresponding random subspace is uniformly distributed on the Grassmannian (i.e. the randomness is given by a standard Gaussian matrix). More recently, results for much more general distributions of sections given by kernels and images of random matrices were obtained, among others, in papers [67] and [73]. In our setting, however, these papers do not seem directly applicable as they provide estimates for diameters up to a constant multiple: in particular, if a convex set K, say, satisfies $K \subset B_2^N \subset 2K$, and E is a random subspace given by a kernel or an image of a random matrix, those results only give a trivial bound $\mathrm{diam}\big(K \cap E\big) < C$

for a large constant C. At the same time, if $S = \mathrm{S}^{N-1} \cap \mathbb{R}^N_+$ then it is easy to show that $\mathrm{conv}(S, -S) \subset B_2^N \subset \sqrt{2}\,\mathrm{conv}(S, -S)$.

When the matrix A is Gaussian, a way of estimating the probability $\mathbb{P}\{\mathrm{Im}(A) \cap S = \emptyset\}$ which is more suitable in our setting is to apply the following result of Gordon (see Corollary 3.4 in [42]):

Theorem 4.7 (Gordon's escape theorem)**.** Let S be a subset of the unit Euclidean sphere S^{N-1} in $\mathbb{R}^N$. Let E be a random n-dimensional subspace of $\mathbb{R}^N$, distributed uniformly on the Grassmannian with respect to the associated Haar measure. Assume that $w(S) < \sqrt{N-n}$. Then $E \cap S = \emptyset$ with probability at least

$$1 - 3.5\ \exp\Big(-\frac{1}{18}\Big(\frac{N-n}{\sqrt{N-n+1}} - w(S)\Big)^2\Big).$$

For the standard Gaussian matrix G, its image is uniformly distributed on the Grassmannian, and Gordon's result provides an efficient estimate of probability $\mathbb{P}\{\mathrm{Im}G \cap S = \emptyset\}$, as long as we have control over the Gaussian width of the set S. In our setting, the choice of S is determined by the applications to random walks; in fact, S shall always be a spherical simplex satisfying certain additional assumptions. A standard approach would be to bound $w(S)$ in terms of the covering numbers of S using the classical Dudley's inequality (see, for example, [63, Theorem 11.17]). However, in our case the set S is relatively large, so the upper bound given by Dudley's inequality is trivial (greater than $\sqrt{N}$). Instead, we will estimate the Gaussian width of S using the following proposition which is a direct consequence of Lemma 4.6 and inequality (4.5):

Proposition 4.8. *Let C be a convex cone in $\mathbb{R}^N$ and denote by C^* its polar cone. Then*

$$w(C \cap B_2^N)^2 \leq N - (N-1)\left(\frac{\mathrm{Vol}_N(C^* \cap B_2^N)}{\mathrm{Vol}_N(B_2^N)}\right)^{2/N}.$$

The next theorem will be applied in §4.4 and §4.5 to the discretized Brownian motion and to random walks on the sphere.

Theorem 4.9. *For any $\gamma \in (0,1]$ there exist positive L, κ and η depending on γ such that the following is true: For $N \geq Ln$, let F be an $N \times N$ random matrix and $\tilde{F}$ be a deterministic invertible $N \times N$ matrix with the condition number satisfying $\|\tilde{F}\|_{2\to 2} \cdot \|\tilde{F}^{-1}\|_{2\to 2} \leq \gamma^{-1}$. If G is the $N \times n$ standard Gaussian matrix, then*

$$\mathbb{P}\big\{\exists y \in \mathrm{S}^{n-1},\ FGy \in \mathbb{R}^N_+\big\} \leq 5.5\ \exp(-\kappa N) + \mathbb{P}\big\{\big\|F - \tilde{F}\big\|_{2\to 2} > \eta \|\tilde{F}\|_{2\to 2}\big\}.$$

The statement holds with $L = 64/\gamma^2$, $\kappa = 2L^{-2}/9$ and $\eta = \gamma/4L$.

Proof. Let $\gamma \in (0,1)$ and take L, κ, and η as stated above. In view of Lemma 4.5 we have

$$\begin{aligned}
&\mathbb{P}\left\{\exists y \in \mathrm{S}^{n-1},\ (FGy)_- = 0\right\} \\
&\quad \le \mathbb{P}\left\{\exists y \in \mathrm{S}^{n-1},\ \|(\tilde{F}Gy)_-\|_2 \le \|(F-\tilde{F})Gy\|_2\right\} \\
&\quad \le \mathbb{P}\left\{\exists y \in \mathrm{S}^{n-1},\ \|(\tilde{F}Gy)_-\|_2 \le \eta\|\tilde{F}\|_{2\to 2}\cdot\|G\|_{2\to 2}\right\} + \mathbb{P}\{\|F-\tilde{F}\|_{2\to 2} > \eta\|\tilde{F}\|_{2\to 2}\}.
\end{aligned}$$

Further,

$$\begin{aligned}
&\mathbb{P}\left\{\exists y \in \mathrm{S}^{n-1},\ \|(\tilde{F}Gy)_-\|_2 \le \eta\|\tilde{F}\|_{2\to 2}\cdot\|G\|_{2\to 2}\right\} \\
&\quad \le \mathbb{P}\left\{\exists y \in \mathrm{S}^{n-1},\ \tilde{F}Gy \in \mathbb{R}^N_+ + \eta\|\tilde{F}\|_{2\to 2}\cdot\|G\|_{2\to 2}B_2^N\right\} \\
&\quad \le \mathbb{P}\left\{\exists y \in \mathrm{S}^{n-1},\ \frac{Gy}{\|Gy\|_2} \in \tilde{F}^{-1}(\mathbb{R}^N_+) + \eta\|\tilde{F}\|_{2\to 2}\frac{\|G\|_{2\to 2}}{s_{\min}(G)}\tilde{F}^{-1}\left(B_2^N\right)\right\} \\
&\quad \le \mathbb{P}\left\{\exists y \in \mathrm{S}^{n-1},\ \frac{Gy}{\|Gy\|_2} \in \tilde{F}^{-1}(\mathbb{R}^N_+) + 2\eta\cdot\gamma^{-1}B_2^N\right\} \\
&\qquad\quad + \mathbb{P}\left\{\|G\|_{2\to 2} > 2s_{\min}(G)\right\} \\
&\quad \le \mathbb{P}\left\{\mathrm{Im}(G) \cap \left(\tilde{F}^{-1}(\mathbb{R}^N_+) + 2\eta\cdot\gamma^{-1}B_2^N\right) \cap \mathrm{S}^{N-1} \ne \emptyset\right\} + 2e^{-N/128},
\end{aligned} \tag{4.9}$$

where the last estimate follows from (4.4).

To control the probability of escaping in (4.9) with help of Theorem 4.7, we have to estimate the Gaussian width of the set

$$\Gamma := \left(\tilde{F}^{-1}(\mathbb{R}^N_+) + 2\eta\cdot\gamma^{-1}B_2^N\right) \cap \mathrm{S}^{N-1}.$$

Note that $\Gamma \subset (1+2\eta\cdot\gamma^{-1})\tilde{F}^{-1}(\mathbb{R}^N_+) \cap B_2^N + 2\eta\cdot\gamma^{-1}B_2^N$. Therefore

$$w(\Gamma) \le (1+2\eta\cdot\gamma^{-1})\cdot w\left(\tilde{F}^{-1}(\mathbb{R}^N_+)\cap B_2^N\right) + 2\eta\cdot\gamma^{-1}\sqrt{N}. \tag{4.10}$$

It remains to bound the Gaussian width of $\tilde{F}^{-1}(\mathbb{R}^N_+)\cap B_2^N$. Denote by C the cone $\tilde{F}^{-1}(\mathbb{R}^N_+)$ and note that $C^* = \tilde{F}^t(\mathbb{R}^N_-)$. Then we have

$$\begin{aligned}
\mathrm{Vol}_N\left(\tilde{F}^t(\mathbb{R}^N_-)\cap B_2^N\right) &= |\det(\tilde{F})|\cdot\mathrm{Vol}_N\left(\mathbb{R}^N_-\cap(\tilde{F}^t)^{-1}(B_2^N)\right) \\
&\ge |\det(\tilde{F})|\cdot\|\tilde{F}\|_{2\to 2}^{-N}\cdot\mathrm{Vol}_N\left(\mathbb{R}^N_-\cap B_2^N\right).
\end{aligned}$$

Since $|\det(\tilde{F})| \ge \|\tilde{F}^{-1}\|_{2\to 2}^{-N}$, we get $\mathrm{Vol}_N\left(C^*\cap B_2^N\right) \ge (\gamma/2)^N\cdot\mathrm{Vol}_N\left(B_2^N\right)$. Now, applying Proposition 4.8, we deduce that

$$w(C \cap B_2^N) \le \sqrt{(1-\gamma^2/8)N}. \tag{4.11}$$

Putting (4.10) and (4.11) together, we get that

$$w(\Gamma) \le \big(1 + 4\eta \cdot \gamma^{-1} - \gamma^2/16\big)\sqrt{N}.$$

The proof is finished by a straightforward application of Theorem 4.7. □

As we will see in the next sections, Theorem 4.9 provides a way to deal with the standard Brownian motion in $\mathbb{R}^n$ and random walks W_θ on the sphere. To treat the standard walk on $\mathbb{Z}^n$, we shall derive a statement covering a rather broad class of random matrices. Let us introduce the following

Definition. A random variable ξ is said to have property $\mathcal{P}(\tau,\delta)$ (or safisfy condition $\mathcal{P}(\tau,\delta)$) for some $\tau,\delta \in (0,1]$ if

$$\mathbb{P}\{\xi < -\tau\} \ge \delta.$$

A random vector X in $\mathbb{R}^n$ is said to have property $\mathcal{P}(\tau,\delta)$ for $\tau,\delta \in (0,1]$ if for any $y \in \mathrm{S}^{n-1}$, the random variable $\langle X, y\rangle_n$ satisfies $\mathcal{P}(\tau,\delta)$.

Obviously, the above property holds (for some τ and δ) for any non-zero r.v. ξ with $\mathbb{E}\,\xi = 0$. As the next elementary lemma shows, with some additional assumptions on moments of ξ, the numbers τ and δ can be chosen as certain functions of the moments:

Lemma 4.10. *Any random variable ξ such that $\mathbb{E}\,\xi = 0$, $\mathbb{E}\,\xi^2 = 1$ and $\mathbb{E}\,|\xi|^{2+\varepsilon} \le B < \infty$ for some $\varepsilon > 0$, has the property $\mathcal{P}(\tau,\delta)$, with τ and δ depending only on ε and B.*

Proof. Indeed, an easy calculation shows that such ξ satisfies

$$\int\limits_{L_\xi{}^2}^{\infty} \mathbb{P}\{\xi^2 \ge t\}\,dt \le \frac{1}{2}$$

for some $L_\xi > 0$ depending only on B and ε. Then

$$\mathbb{E}\,|\xi| \ge \int\limits_0^{L_\xi} \mathbb{P}\{|\xi| \ge t\}\,dt \ge \frac{1}{2L_\xi}\int\limits_0^{L_\xi{}^2} \mathbb{P}\{\xi^2 \ge t\}\,dt \ge \frac{1}{4L_\xi},$$

implying, as $\mathbb{E}\max(0,-\xi)=\frac{1}{2}\mathbb{E}\,|\xi|$,

$$\begin{aligned}\frac{1}{8L_\xi} &\le \int_0^\infty \mathbb{P}\{\xi\le -t\}\,dt\\ &\le \int_0^{8L_\xi}\mathbb{P}\{\xi\le -t\}\,dt+\int_{64{L_\xi}^2}^\infty \frac{1}{2\sqrt{t}}\mathbb{P}\{\xi^2\ge t\}\,dt\\ &\le \int_0^{8L_\xi}\mathbb{P}\{\xi\le -t\}\,dt+\frac{1}{16L_\xi}.\end{aligned}$$

Hence, $\mathbb{P}\{\xi<-2^{-5}{L_\xi}^{-1}\}\ge 2^{-8}{L_\xi}^{-2}$. □

The following theorem will be used to treat the standard walk on $\mathbb{Z}^n$:

Theorem 4.11. *For any $\tau,\delta\in(0,1]$ and any $K>1$, there exist L and $\eta>0$ depending only on τ, δ and K with the following property: Let $N\ge Ln$ and let A be an $N\times n$ random matrix with independent rows having property $\mathcal{P}(\tau,\delta)$. Then for any $N\times N$ random matrix F, matrix FA satisfies*

$$\begin{aligned}\mathbb{P}\big\{\exists y\in \mathrm{S}^{n-1},\ FAy\in\mathbb{R}^N_+\big\}&\le \exp(-\delta^2N/4)\\ &\quad+\mathbb{P}\big\{\|A\|_{2\to2}>K\sqrt{N}\big\}+\mathbb{P}\big\{\|F-\mathbf{Id}_N\|_{2\to2}>\eta\big\}.\end{aligned}$$

Proof. Define L as the smallest positive number satisfying

$$\Big(\frac{3}{\eta}\Big)^{1/L}\le\exp(\delta^2/4),$$

where $\eta:=\frac{\sqrt{\delta}\,\tau}{2\sqrt{2}K}$. Now, take any admissible $N\ge Ln$ and let A and F be as stated above.

Let $\mathcal{N}$ be an η-net on S^{n-1} of cardinality at most $\big(\frac{3}{\eta}\big)^n$. In view of Lemma 4.5 we have

$$\begin{aligned}&\mathbb{P}\big\{\exists y\in \mathrm{S}^{n-1},\ FAy\in\mathbb{R}^N_+\big\}\\ &\quad\le \mathbb{P}\big\{\exists y\in \mathrm{S}^{n-1},\ \|(Ay)_-\|_2\le\|(F-\mathbf{Id}_N)Ay\|_2\big\}\\ &\quad\le \mathbb{P}\big\{\exists y\in \mathrm{S}^{n-1},\ \|(Ay)_-\|_2\le\eta\|A\|_{2\to2}\big\}+\mathbb{P}\big\{\|F-\mathbf{Id}_N\|_{2\to2}>\eta\big\}\\ &\quad\le \mathbb{P}\big\{\exists y'\in \mathcal{N},\ \|(Ay')_-\|_2\le2\eta\|A\|_{2\to2}\big\}+\mathbb{P}\big\{\|F-\mathbf{Id}_N\|_{2\to2}>\eta\big\}.\end{aligned}$$

Further,

$$\begin{aligned}&\mathbb{P}\big\{\exists y' \in \mathcal{N},\ \|(Ay')_-\|_2 \le 2\eta\|A\|_{2\to 2}\big\}\\ &\quad\le \mathbb{P}\big\{\exists y' \in \mathcal{N},\ \|(Ay')_-\|_2 \le 2K\eta\sqrt{N}\big\} + \mathbb{P}\big\{\|A\|_{2\to 2} > K\sqrt{N}\big\}.\end{aligned} \tag{4.12}$$

Fix any $y' \in \mathcal{N}$. For all $i = 1, 2, \ldots, N$, the random variable $\langle Ay', e_i^N\rangle_N$ satisfies the property $\mathcal{P}(\tau, \delta)$. For any $i \le N$, denote by χ_i the indicator function of the event $\{\langle Ay', e_i^N\rangle_N < -\tau\}$. Then $(\chi_i)_{i\le N}$ are independent and $\mathbb{E}\chi_i \ge \delta$. Applying Hoeffding's inequality (see [50, Theorem 1]), we get

$$\begin{aligned}\mathbb{P}\Big\{|\{i \le N : \langle Ay', e_i^N\rangle_N < -\tau\}| \le \frac{\delta N}{2}\Big\} &\le \mathbb{P}\Big\{\frac{1}{N}\sum_{i\le N}(\chi_i - \mathbb{E}\chi_i) \le -\frac{\delta}{2}\Big\}\\ &\le \exp(-\delta^2 N/2).\end{aligned}$$

Therefore for any fixed $y' \in \mathcal{N}$, we have

$$\begin{aligned}\mathbb{P}\big\{\|(Ay')_-\|_2 \le 2K\eta\sqrt{N}\big\} &\le \mathbb{P}\big\{|\{i \le N : \langle Ay', e_i^N\rangle_N \le -\tau\}| \le 4K^2\eta^2 N/\tau^2\big\}\\ &\le \exp(-\delta^2 N/2).\end{aligned}$$

Combining the last estimate with (4.12) and the upper estimate for $|\mathcal{N}|$, we get

$$\begin{aligned}&\mathbb{P}\big\{\exists y \in \mathrm{S}^{n-1},\ FAy \in \mathbb{R}_+^N\big\}\\ &\quad\le \Big(\frac{3}{\eta}\Big)^n \exp(-\delta^2 N/2) + \mathbb{P}\big\{\|A\|_{2\to 2} > K\sqrt{N}\big\} + \mathbb{P}\big\{\|F - \mathbf{Id}_N\|_{2\to 2} > \eta\big\}.\end{aligned}$$

The result follows by the choice of L. □

Remark 4.2. Theorem 4.11, applied to the Gaussian matrix G, gives a weaker form of Theorem 4.9 (with more restrictions on the choice of F). Let us emphasize that the theorems do not require F to be independent from G. This will be important in §4.5.

4.4 Applications to random walks in $\mathbb{R}^n$

In this section, we will apply the statements about random matrices to the Brownian motion and the standard walk on $\mathbb{Z}^n$.

Corollary 4.12. *For any $K > 1$, there are constants L and κ depending only on K such that the following holds. Let $N \ge Ln$ and $t_1, \ldots, t_N$ be such that $t_i \ge K \cdot t_{i-1}$ for any*

$i = 2 \dots N$ *and* $t_1 > 0$. *Then*

$$\mathbb{P}\big\{0 \text{ belongs to the interior of } \operatorname{conv}\{\mathrm{BM}_n(t_i) : i \le N\}\big\} \ge 1 - 5.5\exp(-\kappa N).$$

Proof. Let $c_K := 1 + (K-1)^{-1/2} \sum_{j\ge 0} K^{-j/2}$ and $\gamma := c_K^{-1} \cdot (1 + (K-1)^{-1/2})^{-1}$ be two constants depending only on K and take $L = 64/\gamma^2$ and $\kappa := 2L^{-2}/9$.

Denote $\delta_1 := \sqrt{t_1}$ and $\delta_i := \sqrt{t_i - t_{i-1}}$ for any $i = 2 \dots N$. Observe that for any $j < i$, we have $\delta_i \ge K^{\frac{i-j-1}{2}} \sqrt{K-1} \cdot \delta_j$.

Define F as the $N \times N$ lower triangular matrix whose entries are given by $f_{ii} = 1$ for any $i \le N$ and $f_{ij} = \frac{\delta_j}{\delta_i}$ for any $i > j$. One can easily check that $\|F\|_{2\to 2} \le c_K$. Moreover, the inverse of F is a lower bidiagonal matrix with 1 on the main diagonal and $(\delta_i/\delta_{i+1})_{i<N}$ on the diagonal below. Hence $\|F^{-1}\|_{2\to 2} \le 1 + (K-1)^{-1/2}$, and the condition number of F satisfies

$$\|F\|_{2\to 2} \cdot \|F^{-1}\|_{2\to 2} \le \gamma^{-1}.$$

Let $(R_i)_{i\le N}$ be the rows of FG. One can check that $R_i = \mathrm{BM}_n(t_i)/\delta_i$ and therefore

$$0 \in \operatorname{conv}\{\mathrm{BM}_n(t_i) : i \le N\} \Leftrightarrow 0 \in \operatorname{conv}\{R_i : i \le N\}$$

Note that, by a standard separation argument, 0 does not belong to the interior of $\operatorname{conv}\{R_i : i \le N\}$ if and only if $\operatorname{rank}(FG) < n$ or there is a vector $y \in \mathrm{S}^{n-1}$ such that $\langle FGy, e_i^N \rangle_N = \langle y, R_i \rangle_n \ge 0$ for any $i \le N$, where $(e_i^N)_{i\le N}$ denotes the canonical basis of $\mathbb{R}^N$. Since with probability one we have $\operatorname{rank}(FG) = n$, the result follows by applying Theorem 4.9 with $\tilde{F} := F$. □

Suppose (t_i) is a finite increasing sequence of points in $[0,1]$. The above statement tells us that if (t_i) contains a geometrically growing subsequence of length Ln for an appropriate $L > 0$ then with high probability the origin of $\mathbb{R}^n$ is contained in the interior of $\mathrm{BM}_n(t_i)$'s. We shall apply this result to the case when the t_i's are generated by the Poisson point process independent from BM_n.

Recall that *the homogeneous Poisson point process in* $[0,1]$ of intensity $s > 0$ is a random discrete measure N_s on $[0,1]$ such that 1) for each Borel subset $B \subset [0,1]$, the random variable $N_s(B)$ has the Poisson distribution with parameter $s\mu(B)$, where μ is the usual Lebesgue measure on $\mathbb{R}$, and 2) for any $j \in \mathbb{N}$ and pairwise disjoint Borel sets $B_1, B_2 \dots, B_j \subset [0,1]$, the random variables $N_s(B_1)$, $N_s(B_2), \dots, N_s(B_j)$ are jointly independent. The measure N_s admits a representation of the form

$$N_s = \sum_{i=1}^{\tau} \delta_{\xi_i},$$

where $\xi_1, \xi_2, \dots$ are i.i.d. random variables uniformly distributed on $[0,1]$, δ_{ξ_i} is the Dirac measure with the mass at ξ_i and τ is the random non-negative integer with the Poisson distribution with parameter s.

Theorem 3.1 of [29] states that if τ and the points $\xi_1, \xi_2, \dots, \xi_\tau$ are generated by the homogeneous PPP in $[0,1]$ of intensity $s \geq n^{Cn}$ then the convex hull of $\mathrm{BM}_n(\xi_i)$'s contains the origin in its interior with probability at least $1 - n^{-n}$. In our next statement, we weaken the assumptions on s at expense of decreasing the probability to $1 - \exp(-n)$:

Corollary 4.13. *There is a universal constant $\tilde{C} > 0$ with the following property: Let $n \in \mathbb{N}$ and let $\mathrm{BM}_n(t)$, $t \in [0,\infty)$, be the standard Brownian motion in $\mathbb{R}^n$. Further, let τ and the points $\xi_1, \xi_2, \dots, \xi_\tau$ be given by the homogeneous Poisson process on $[0,1]$ of intensity $s \geq \exp(\tilde{C}n)$, which is independent from $\mathrm{BM}_n(t)$. Then*

$$\mathbb{P}\big\{0 \text{ belongs to the interior of } \operatorname{conv}\{\mathrm{BM}_n(\xi_i) : i \leq \tau\}\big\} \geq 1 - \exp(-n).$$

Proof. Let $K := 2$ and κ, L be as in Corollary 4.12. Then we define the constant $\tilde{C} := \max\big(\frac{32}{\kappa}, 8L\big)$. Let $n \in \mathbb{N}$ and let N_s be as stated above. Take $m := \lfloor \tilde{C}n \rfloor$ and

$$I_1 := [0, K^{-m+1}]; \;\; I_j := (K^{j-m-1}, K^{j-m}],\ j = 2, 3, \dots, m.$$

From the definition of N_s, we have

$$\begin{aligned}\mathbb{P}\big\{N_s(I_j) > 0 \text{ for all } j = 1, 2, \dots, m\big\} &\geq 1 - \sum_{j=1}^{m} \exp\big(-s\mu(I_j)\big)\\ &\geq 1 - m\exp\big(-sK^{-m}\big).\end{aligned}$$

In particular, with probability at least $1 - m\exp\big(-sK^{-m}\big)$ the set $\{\xi_i\}_{i=1}^{\tau}$ contains a subset $\{\xi_{i_1}, \xi_{i_2}, \dots, \xi_{i_m}\}$ such that $\xi_{i_j} \in I_j$ for every admissible j, hence $\xi_{i_{j+2}} \geq K\xi_{i_j}$ for any $j \leq m-2$. Conditioning on the realization of N_s, we obtain by Corollary 4.12:

$$\begin{aligned}&\mathbb{P}\big\{0 \text{ belongs to the interior of } \operatorname{conv}\{\mathrm{BM}_n(\xi_i) : i \leq \tau\}\big\}\\ &\quad\geq 1 - m\exp\big(-sK^{-m}\big) - 5.5\exp(-\kappa\lfloor m/2 \rfloor)\\ &\quad\geq 1 - \exp(-n),\end{aligned}$$

and the proof is complete. □

The last result of this section concerns the standard random walk $W(j)$ on $\mathbb{Z}^n$, which is defined as a walk with independent increments such that each increment $W(j+1) - W(j)$ is uniformly distributed on the set $\{\pm e_j^n\}_{j\le n}$. We note that the random variables $\langle \sqrt{n/m}W(m), y\rangle_n$ ($m \in \mathbb{N}$, $y \in \mathrm{S}^{n-1}$) are *not* uniformly subgaussian; to be more precise, their subgaussian moment depends on the dimension n. At the same time, the vectors $W(m)$ still have very strong concentration properties as the next lemma shows:

Lemma 4.14. *Let $W(j)$ ($j \ge 0$) be the standard walk on $\mathbb{Z}^n$ starting at the origin, and $m \ge n^4$ be any fixed integer. Then the vector $X := \sqrt{n/m}W(m)$ is isotropic and satisfies for any $y \in \mathrm{S}^{n-1}$:*

$$\mathbb{P}\big\{|\langle X, y\rangle_n| \ge t\big\} \le \exp\big(-2(mn)^{1/4}\big) + 2\exp(-t^2/4), \quad t > 0.$$

In particular, $\mathbb{E}\,|\langle X, y\rangle_n|^3 \le 100$ for all $y \in \mathrm{S}^{n-1}$, and X has the property $\mathcal{P}(\tau,\delta)$ for some universal constants τ,δ.

Proof. The isotropicity of X can be easily checked. Fix for a moment any vector $y \in \mathrm{S}^{n-1}$. The random variable $\langle X, y\rangle_n$ can be represented as

$$\langle X, y\rangle_n = \sqrt{n/m}\sum_{k=1}^m s_k,$$

where the variables $s_1, s_2, \dots, s_m$ are i.i.d. and each

$$s_k := \langle W(k) - W(k-1), y\rangle_n$$

is symmetrically distributed, has variance $\mathbb{E}\,{s_k}^2 = \frac{1}{n}$ and takes values in the interval $[-1,1]$. Applying Hoeffding's inequality to the sum $\sum_{k=1}^m {s_k}^2$, we get

$$\mathbb{P}\Big\{\sum_{k=1}^m {s_k}^2 \ge \frac{2m}{n}\Big\} \le \exp(-2m/n^2). \tag{4.13}$$

Further, since s_k is symmetric, the distribution of the sum $\sum_{k=1}^m s_k$ is the same as the distribution of $\sum_{k=1}^m r_k s_k$, where $r_1, r_2, \dots, r_m$ are Rademacher variables jointly independent with $s_1, s_2, \dots, s_m$. Conditioning on the values of s_k and using (4.13) and the Khintchine inequality, we obtain for every $t > 0$:

$$\begin{aligned}
&\mathbb{P}\Big\{\Big|\sum_{k=1}^m s_k\Big| \geq mt\Big\} \\
&= \mathbb{P}\Big\{\Big|\sum_{k=1}^m r_k s_k\Big| \geq mt\Big\} \\
&\leq \mathbb{P}\Big\{\sum_{k=1}^m {s_k}^2 \geq \frac{2m}{n}\Big\} + \mathbb{P}\Big\{\sum_{k=1}^m {s_k}^2 \leq \frac{2m}{n} \textbf{ and } \Big|\sum_{k=1}^m r_k s_k\Big| \geq mt\Big\} \\
&\leq \exp(-2m/n^2) + 2\exp(-mnt^2/4).
\end{aligned}$$

Whence, in view of the bound $m \geq n^2(mn)^{1/4}$, we get

$$\mathbb{P}\big\{|\langle X, y\rangle_n| \geq t\big\} \leq \exp\big(-2(mn)^{1/4}\big) + 2\exp(-t^2/4), \quad t > 0. \tag{4.14}$$

The condition (4.14), together with the bound $\|X\|_2 \leq \sqrt{mn}$, gives $\mathbb{E}\,|\langle X, y\rangle_n|^3 \leq 100$. It remains to apply Lemma 4.10. $\square$

The next lemma follows from well known concentration inequalities for subexponential random variables (see, for example, [117, Corollary 5.17]):

Lemma 4.15. *There is a universal constant $\tilde{C} > 0$ such that for any $N \in \mathbb{N}$ and independent centered random variables $\tilde{\xi}_1, \tilde{\xi}_2, \ldots, \tilde{\xi}_N$, each satisfying*

$$\mathbb{P}\big\{\tilde{\xi}_i \geq t\big\} \leq 3\exp(-t/4), \quad t > 0, \tag{4.15}$$

we have

$$\mathbb{P}\Big\{\sum_{i=1}^N \tilde{\xi}_i \geq \tilde{C}N\Big\} \leq 40^{-N}. \tag{4.16}$$

In the next result, compared to Theorem 1.2 of [29], we decrease the bound on the number of steps N of the walk on $\mathbb{Z}^n$ sufficient to absorb the origin with high probability.

Corollary 4.16. *There is a universal constant $C > 0$ with the following property: Let $n, R \in \mathbb{N}$, $R \geq \exp(Cn)$ and let $W(j)$, $j \geq 0$, be the standard random walk on $\mathbb{Z}^n$ starting at the origin. Then*

$$\mathbb{P}\big\{0 \textit{ belongs to the interior of } \operatorname{conv}\{W(j) : j = 1, \ldots, R\}\big\} \geq 1 - 2\exp(-n).$$

Proof. Definition of constants and the matrix A. Let $\tau, \delta > 0$ be taken from Lemma 4.14 and $\tilde{C}$ — from Lemma 4.15. Now, we define $K := 2\sqrt{\tilde{C}}$ and let L and η be taken from

Theorem 4.11. Finally, we define $C > 0$ as the smallest positive number satisfying

$$\exp(Cn) \geq (28N)^4 \Big\lceil \frac{4}{\eta^2} + 1 \Big\rceil^N$$

for any $n \in \mathbb{N}$ and $N = n\lceil \max(L, 4/\delta^2) \rceil$.

Fix any numbers $n > 0$ and $R \geq \exp(Cn)$, and let $N := n\lceil \max(L, 4/\delta^2) \rceil$. Further, let t_i ($i = 0, 1, \ldots, N$) be numbers from $\{0, 1, \ldots, R\}$, with $t_0 = 0$, $t_1 = (28N)^4$ and $t_i = \lceil \frac{4}{\eta^2} + 1 \rceil t_{i-1}$, $i = 2, 3, \ldots, N$. Denote

$$X_i := \sqrt{n}(t_i - t_{i-1})^{-1/2}\big(W(t_i) - W(t_{i-1})\big), \quad i = 1, 2, \ldots, N.$$

Then the vectors are isotropic, jointly independent and, in view of Lemma 4.14, satisfy

$$\mathbb{P}\big\{|\langle X_i, y\rangle_n| \geq t\big\} \leq \exp\big(-2(nt_i - nt_{i-1})^{1/4}\big) + 2\exp(-t^2/4), \quad t > 0 \tag{4.17}$$

for all $y \in \mathrm{S}^{n-1}$. We let A to be the $N \times n$ random matrix with rows X_i.

Estimating the norm of A. Let $\mathcal{N}$ be a 1/2-net on S^{n-1} of cardinality at most 5^n. Fix any $y' \in \mathcal{N}$. For each $i = 1, 2, \ldots, N$, let $\xi_i := {\langle X_i, y'\rangle_n}^2$, and let $\tilde{\xi}_i$ be its truncation at level $(nt_i - nt_{i-1})^{1/4}$, i.e.

$$\tilde{\xi}_i(\omega) = \begin{cases} \xi_i(\omega), & \text{if } \xi_i(\omega) \leq (nt_i - nt_{i-1})^{1/4}, \\ 0, & \text{otherwise.} \end{cases}$$

Note that, in view of (4.17), the variables $\tilde{\xi}_i$ satisfy (4.15), and

$$\mathbb{P}\{\xi_i \neq \tilde{\xi}_i\} \leq 3\exp\big(-(nt_i - nt_{i-1})^{1/4}/4\big).$$

Hence, by (4.16) and the above estimate, we have

$$\begin{aligned}
\mathbb{P}\{\|Ay'\|_2 \geq \sqrt{\tilde{C}N}\} &= \mathbb{P}\Big\{\sum_{i=1}^N \xi_i \geq \tilde{C}N\Big\} \\
&\leq 40^{-n} + \mathbb{P}\big\{\xi_i \neq \tilde{\xi}_i \text{ for some } i \in \{1, 2, \ldots, N\}\big\} \\
&\leq 40^{-n} + 3\sum_{i=1}^N \exp\big(-(nt_i - nt_{i-1})^{1/4}/4\big) \\
&\leq 40^{-n} + 3N\exp\big(-7Nn^{1/4}\big) \\
&\leq 20^{-n}.
\end{aligned}$$

Taking the union bound for all $y' \in \mathcal{N}$ and applying the standard approximation argument, we obtain $\|A\|_{2\to 2} \le 2\sqrt{CN} = K\sqrt{N}$ with probability at least $1 - \exp(-n)$.

Construction of the matrix F and application of Theorem 4.11. Let F be the $N \times N$ non-random lower-triangular matrix, with the entries

$$f_{ij} = \sqrt{\frac{t_j - t_{j-1}}{t_i - t_{i-1}}}, \quad i \ge j.$$

Obviously, FA is the matrix whose i-th row ($i = 1, \dots, N$) is precisely the vector

$$\sqrt{\frac{n}{t_i - t_{i-1}}} W(t_i).$$

Then, in view of the definition of t_i's, we have

$$\|F - \mathbf{Id}_N\|_{2\to 2} \le \frac{\eta/2}{1 - \eta/2} \le \eta.$$

Finally, applying Theorem 4.11, we obtain

$$\begin{aligned}
&\mathbb{P}\big\{0 \text{ belongs to the interior of } \mathrm{conv}\{W(j) : j = 1, 2, \dots, R\}\big\} \\
&\quad \ge \mathbb{P}\big\{0 \text{ belongs to the interior of } \mathrm{conv}\{W(t_i) : i = 1, 2, \dots, N\}\big\} \\
&\quad = \mathbb{P}\big\{\mathrm{rank} A = n \textbf{ and } \mathrm{Im}(FA) \cap \mathbb{R}^n_+ = \{0\}\big\} \\
&\quad \ge 1 - 2\exp(-n).
\end{aligned}$$

□

4.5 Random walks on the sphere

Let $n > 1$ and $\theta \in (0, \pi/2)$. Here, we consider the Markov chain W_θ taking values on S^{n-1} such that the angle between two consecutive steps is θ i.e. for any $i \ge 1$ we have $\langle W_\theta(i), W_\theta(i+1)\rangle_n = \cos\theta$ a.s., and the direction from $W_\theta(i)$ to $W_\theta(i+1)$ is chosen uniformly at random. The latter condition means that for any $u \in \mathrm{S}^{n-1}$, the distribution of $W_\theta(i+1)$ conditioned on $W_\theta(i) = u$, is uniform on the $(n-2)$-sphere $\mathrm{S}^{n-1} \cap \{x \in \mathbb{R}^n : \langle x, u\rangle_n = \cos\theta\}$. See [89] for a study of these walks and some of their generalizations.

The question addressed in this section is how many steps it takes for W_θ to absorb the origin into its convex hull. Note that the answer does not depend on the distribution of the first vector $W_\theta(1)$, and we shall further assume that $W_\theta(1)$ is uniformly distributed on the sphere. The question can be equivalently reformulated as a problem of estimating

$\pi/2$-covering time of W_θ. For $\phi \in (0, \pi/2]$, a *ϕ-covering* of S^{n-1} is any subset S of the sphere such that the geodesic distance from any point of the sphere to S is at most ϕ. Then the ϕ-covering time for W_θ is the random variable

$$T = \min\big\{N : \text{ the set } \{W_\theta(i),\, i \le N\} \text{ is a } \phi\text{-covering of } S^{n-1}\big\}.$$

A related problem of estimating ϕ-covering time of the *spherical Brownian motion* was considered in [72] and [29], for $\phi \to 0$ and $\phi = \pi/2$, respectively. It is not clear whether the argument developed in [29] can be adopted to the walks W_θ. Our approach to the above problem is based on the results of §4.3 and is completely different from the argument in [29].

The walk W_θ can be constructively described as follows: Let $Y_1, Y_2, \dots$ be a sequence of independent standard Gaussian vectors in $\mathbb{R}^n$. Let $\beta_1 := \|Y_1\|_2$ and define

$$W_\theta(1) := \frac{Y_1}{\|Y_1\|_2} = \frac{Y_1}{\beta_1}.$$

Further, for any $i \ge 1$ let

$$W_\theta(i+1) := \frac{\alpha_{i+1} W_\theta(i) + Y_{i+1}}{\beta_{i+1}}, \tag{4.18}$$

where

$$\begin{aligned} \beta_{i+1} &:= \|\alpha_{i+1} W_\theta(i) + Y_{i+1}\|_2 \quad \text{ and} \\ \alpha_{i+1} &:= \cot\theta\, \|P_i Y_{i+1}\|_2 - \langle Y_{i+1}, W_\theta(i)\rangle_n, \;\; i \ge 1, \end{aligned} \tag{4.19}$$

with P_i denoting the (random) orthogonal projection onto the hyperplane orthogonal to $W_\theta(i)$. It can be easily checked that

$$\beta_i = \frac{\|P_{i-1} Y_i\|_2}{\sin\theta}, \;\; i \ge 2,$$

and that W_θ is the Markov process described at the beginning of the section. For any $i = 2, 3, \dots$ the coefficients α_i and β_i are random variables depending on Y_i and $W_\theta(i-1)$. Using (4.2) and (4.3), one can deduce the following concentration inequalities:

Lemma 4.17. *There exist a universal constant $c > 0$ such that for $\delta_\theta := c\min(1, \cot\theta)$ and for any $i = 2, 3, \dots$ and $\varepsilon > 0$ we have*

$$\mathbb{P}\big\{(1-\varepsilon)\sqrt{n}\cot\theta \le \alpha_i \le (1+\varepsilon)\sqrt{n}\cot\theta\big\} \ge 1 - 2\exp(-{\delta_\theta}^2\varepsilon^2 n)$$

and

$$\mathbb{P}\big\{(1-\varepsilon)\sin\theta/\sqrt{n} \le {\beta_i}^{-1} \le (1+\varepsilon)\sin\theta/\sqrt{n}\big\} \ge 1 - 2\exp(-{\delta_\theta}^2\varepsilon^2 n).$$

Moreover, (4.3) immediately implies

$$\mathbb{P}\big\{(1-\varepsilon)/\sqrt{n} \le {\beta_1}^{-1} \le (1+\varepsilon)/\sqrt{n}\big\} \ge 1 - 2\exp(-c\varepsilon^2 n), \quad \varepsilon > 0, \tag{4.20}$$

provided that the constant c is sufficiently small. Before we state the main result of the section, let us consider the following elementary lemma:

Lemma 4.18. *For any $q \in (0,1)$ and $0 < \varepsilon \le \frac{1-q}{8}$ we have*

$$\sum_{k=0}^{\infty}\big((1+\varepsilon)^{2k+1} - 1\big)q^k \le \frac{4\varepsilon}{(1-q)^2}.$$

Proof. First, note that the conditions on ε and q imply

$$q(1+\varepsilon)^2 \le \frac{81q}{64} - \frac{9q^2}{32} + \frac{q^3}{64} \le q + \frac{17q}{64} - \frac{17q^2}{64} \le \frac{1+q}{2},$$

whence

$$1 - q(1+\varepsilon)^2 \ge \frac{1-q}{2}.$$

Using the last inequality, we obtain

$$\begin{aligned}\sum_{k=0}^{\infty}\big((1+\varepsilon)^{2k+1} - 1\big)q^k &= (1+\varepsilon)\sum_{k=0}^{\infty}\big(q(1+\varepsilon)^2\big)^k - \sum_{k=0}^{\infty} q^k \\ &= \frac{(1+\varepsilon)}{1-q(1+\varepsilon)^2} - \frac{1}{1-q} \\ &= \frac{\varepsilon + \varepsilon q + \varepsilon^2 q}{(1-q)(1-q(1+\varepsilon)^2)} \\ &\le \frac{4\varepsilon}{(1-q)^2}.\end{aligned}$$

□

Theorem 4.19. *For any $\theta \in (0, \pi/2)$ there exist $n_0 = n_0(\theta)$ and $K = K(\theta)$ depending only on θ such that the following holds: Let $n \ge n_0$ and let W_θ be the random walk on S^{n-1} defined above. Then for all $N \ge Kn$ we have*

$$\mathbb{P}\big\{0 \textit{ belongs to } \mathrm{conv}\{W_\theta(i) : i \le N\}\big\} \ge 1 - \exp(-n).$$

Proof. Fix an angle $\theta \in (0, \pi/2)$. Let $\gamma := \frac{\sin\theta\,(1-\cos\theta)}{1+\cos\theta}$ and let η, L and κ be as in Theorem 4.9. Define $\varepsilon := \eta \sin\theta\,(1-\cos\theta)^2/4$ and let n_0 be the smallest integer such that for all $n \ge n_0$ we have

$$5.5\exp(-\kappa\lceil Ln\rceil) + 4\lceil Ln\rceil \exp(-{\delta_\theta}^2\varepsilon^2 n) \le \exp(-\mu\, n),$$

where $\mu = \frac{1}{2}\min\bigl(\kappa, {\delta_\theta}^2\varepsilon^2\bigr)$ and δ_θ is taken from Lemma 4.17.

Fix $n \ge n_0$. First, we show that $\tilde{N} := \lceil Ln\rceil$ steps is sufficient to get the origin in the convex hull of $W_\theta(i)$ $(i \le \tilde{N})$ with probability $1-\exp(-\mu\, n)$. This shall be done by using the representation (4.18) for the walk W_θ and by applying Theorem 4.9. Then we will augment the probability estimate to $1 - \exp(-n)$ by increasing the number of steps.

Let G be the standard $\tilde{N}\times n$ Gaussian matrix with rows Y_i $(i \le \tilde{N})$. We shall construct a random lower-triangular $\tilde{N} \times \tilde{N}$ matrix F such that the i-th row of FG is $W_\theta(i)$. Define $F := (f_{ij})$ with

$$f_{ij} := \frac{\prod_{k=j+1}^{i} \alpha_k}{\prod_{k=j}^{i} \beta_k} \text{ for } j < i \le \tilde{N} \quad \text{and} \quad f_{ii} := \frac{1}{\beta_i} \text{ for } i \le \tilde{N},$$

where α_k and β_k are given by (4.19). Since $FG = (W_\theta(1), W_\theta(2), \ldots, W_\theta(\tilde{N}))^t$, the origin does not belong to $\mathrm{conv}\{W_\theta(i) :\ i \le \tilde{N}\}$ only if there exists $y \in \mathrm{S}^{n-1}$ such that $FGy \in \mathbb{R}_+^{\tilde{N}}$. Now define $\tilde{F}$ as the $\tilde{N} \times \tilde{N}$ lower triangular matrix whose entries are given by

$$\tilde{f}_{i1} = \frac{(\cos\theta)^{i-1}}{\sqrt{n}} \text{ for any } i \le \tilde{N} \text{ and } \tilde{f}_{ij} := \sin\theta\,\frac{(\cos\theta)^{i-j}}{\sqrt{n}} \text{ for } 2 \le j \le i.$$

It is not difficult to see that

$$\frac{\sin\theta}{\sqrt{n}} \le \|\tilde{F}\|_{2\to 2} \le \frac{1}{(1-\cos\theta)\sqrt{n}}. \tag{4.21}$$

Further, let Q be the matrix obtained from $\tilde{F}$ by multiplying the first column of $\tilde{F}$ by $\sin\theta$ and leaving the other columns unchanged. Then, clearly, $s_{\min}(Q) \le s_{\min}(\tilde{F})$ implying $\|\tilde{F}^{-1}\|_{2\to 2} \le \|Q^{-1}\|_{2\to 2}$. On the other hand, the inverse of Q is a lower bidiagonal matrix with $\frac{\sqrt{n}}{\sin\theta}$ on the main diagonal and $-\cos\theta\frac{\sqrt{n}}{\sin\theta}$ on the diagonal below. Hence, $\|\tilde{F}^{-1}\|_{2\to 2} \le \|Q^{-1}\|_{2\to 2} \le (1+\cos\theta)\frac{\sqrt{n}}{\sin\theta}$, and the condition number of $\tilde{F}$ satisfies

$$\|\tilde{F}\|_{2\to 2} \cdot \|\tilde{F}^{-1}\|_{2\to 2} \le \frac{1+\cos\theta}{\sin\theta\,(1-\cos\theta)} = \gamma^{-1}.$$

Applying Theorem 4.9, we get

$$\mathbb{P}\big\{\exists y \in \mathrm{S}^{n-1},\ FGy \in \mathbb{R}_+^{\tilde N}\big\} \le 5.5\exp(-\kappa \tilde N) + \mathbb{P}\big\{\big\|F-\tilde F\big\|_{2\to 2} > \eta\|\tilde F\|_{2\to 2}\big\}.$$

It remains to bound the probability $\mathbb{P}\big\{\big\|F-\tilde F\big\|_{2\to 2} > \eta\|\tilde F\|_{2\to 2}\big\}$. In view of Lemma 4.17 and (4.20), with probability at least $1-4\tilde N\exp(-{\delta_\theta}^2\varepsilon^2 n)$ we have

$$\big|f_{ij}-\tilde f_{ij}\big| \le \big((1+\varepsilon)^{2(i-j)+1}-1\big)\,\tilde f_{ij} \ \text{ for any } j\le i.$$

This, together with Lemma 4.18 and (4.21), implies that

$$\|F-\tilde F\|_{2\to 2} \le \frac{1}{\sqrt n}\sum_{k=0}^{\infty}\big((1+\varepsilon)^{2k+1}-1\big)(\cos\theta)^k \le \frac{4\varepsilon}{(1-\cos\theta)^2\sqrt n} \le \eta\|\tilde F\|_{2\to 2}$$

with probability at least $1-4\tilde N\exp(-{\delta_\theta}^2\varepsilon^2 n)$. Hence, by the restriction on n_0,

$$\mathbb{P}\big\{\exists y \in \mathrm{S}^{n-1},\ FGy \in \mathbb{R}_+^{\tilde N}\big\} \le 5.5\exp(-\kappa\tilde N) + 4\tilde N\exp(-{\delta_\theta}^2\varepsilon^2 n) \le \exp(-\mu n),$$

where $\mu = \frac12\min\big(\kappa, {\delta_\theta}^2\varepsilon^2\big)$. Finally, if $N \ge \lceil \mu^{-1}\rceil \tilde N$ then the above estimate implies

$$\begin{aligned}&\mathbb{P}\big\{0 \text{ does not belong to } \mathrm{conv}\{W_\theta(i) : i\le N\}\big\}\\ &\quad\le \mathbb{P}\big\{0 \text{ does not belong to } \mathrm{conv}\{W_\theta(i) : i\le \tilde N\}\big\}^{\lceil \mu^{-1}\rceil}\\ &\quad\le \exp(-n).\end{aligned}$$

□

4.6 Minimax of the n-dimensional Brownian motion

In this section we will prove Theorem 4.2 which, as noted in the introduction to this Chapter, is equivalent to estimate (4.1).

Let us give an informal description of the proof. We construct a random unit vector $\bar v$ in $\mathbb{R}^n$ such that with probability close to one

$$\langle \bar v, \mathrm{BM}_n(t)\rangle_n > 0 \quad \text{for any } t\in[1,2^{cn}]. \tag{4.22}$$

The construction procedure shall be divided into a series of steps. At the initial step, we produce a random vector $\bar{v}_0$ such that

$$\langle \bar{v}_0, \mathrm{BM}_n(2^i)\rangle_n > 0 \quad \text{for any } i = 0, 1, \ldots, cn.$$

(In fact, $\bar{v}_0$ will satisfy a stronger condition). At a step k, $k \geq 1$, we update the vector $\bar{v}_{k-1}$ by adding a small perturbation in such a way that

$$\langle \bar{v}_k, \mathrm{BM}_n(2^{j2^{-k}})\rangle_n > 0 \quad \text{for any } j = 0, 1, \ldots, 2^k cn.$$

(Again $\bar{v}_k$ will in fact satisfy a stronger condition). Finally, using some standard properties of the Brownian bridge, we verify that $\bar{v} := \bar{v}_{\log_2 \ln n}$ satisfies (4.22) with a large probability.

4.6.1 Auxiliary facts

In this paragraph we introduce several auxiliary results that will be used within the proof. The proof of the next lemma is straightforward, so we omit it.

Lemma 4.20. *Let* $\mathrm{BM}_n(t)$ *(*$0 \leq t < \infty$*) be the standard Brownian motion in* $\mathbb{R}^n$ *and let* $0 < a < b$*. Fix any* $s \in (a, b)$ *and set*

$$w(s) := \frac{b-s}{b-a}\mathrm{BM}_n(a) + \frac{s-a}{b-a}\mathrm{BM}_n(b); \;\; u(s) := \mathrm{BM}_n(s) - w(s).$$

Then the process $u(s)$*,* $s \in (a, b)$*, is a Brownian bridge, and*

1. $u(s)$ *is a centered Gaussian vector with the covariance matrix*

$$\frac{(b-s)(s-a)}{b-a}\mathbf{Id}_n.$$

2. *The random vector* $u(s)$ *is independent from the process* $\mathrm{BM}_n(t)$ *indexed over* $t \in (0, a] \cup [b, \infty)$*.*

Lemma 4.21. *Let* $d, m \in \mathbb{N}$ *be such that* $m \leq d/2$*. Let* $X_1, X_2, \ldots, X_m$ *be independent standard Gaussian vectors in* $\mathbb{R}^d$*. Then for any non-random vector* $b \in \mathrm{S}^{m-1}$*, there exists a random unit vector* $\bar{u}_b \in \mathbb{R}^d$ *such that*

$$\mathbb{P}\Big\{\langle \bar{u}_b, X_i\rangle_d \geq c_{4.21}\sqrt{d}|b_i| \text{ for all } i = 1, 2, \ldots, m\Big\} \geq 1 - \exp(-c_{4.21}d),$$

where $c_{4.21}$ *is a universal constant and* b_i*'s are the coordinates of* b*. Moreover,* $\bar{u}_b$ *can be defined as a Borel function of* X_i*'s and* b*.*

Proof. Without loss of generality, we can assume that $b_i \neq 0$ for any $i \leq m$ and that X_i's are linearly independent on the entire probability space. Denote by E the random affine subspace of $\mathbb{R}^d$ spanned by $\{|b_i|^{-1}X_i\}_{i\leq m}$. Define $\bar{u}_b$ as the unique unit vector in $\mathrm{span}\{X_1,\dots,X_m\}$ such that $\bar{u}_b$ is orthogonal to E and for any $i \leq m$ we have

$$\langle \bar{u}_b, |b_i|^{-1}X_i\rangle_d = \mathrm{d}(0,E),$$

where $\mathrm{d}(0,E)$ stands for the distance from the origin to E. Then we have

$$\sum_{i\leq m}\langle \bar{u}_b, X_i\rangle_d^{\,2} = \sum_{i\leq m}\Big\langle \bar{u}_b, \frac{X_i}{|b_i|}\Big\rangle_d^{2}|b_i|^2 = \sum_{i\leq m}\mathrm{d}(0,E)^2\cdot|b_i|^2 = \mathrm{d}(0,E)^2. \tag{4.23}$$

Let G be the $d\times m$ standard Gaussian matrix with columns X_i, $i=1,2,\dots,m$. Using the definition of $\bar{u}_b$ together with (4.23), we obtain for any $\tau>0$:

$$\begin{aligned}\mathbb{P}\Big\{\langle \bar{u}_b, X_i\rangle_d \geq \tau\sqrt{d}|b_i| \text{ for all } i=1,2,\dots,m\Big\} &= \mathbb{P}\Big\{\mathrm{d}(0,E)\geq \tau\sqrt{d}\Big\}\\ &= \mathbb{P}\Big\{\sqrt{\sum_{i\leq m}\langle \bar{u}_b, X_i\rangle_d^{\,2}}\geq \tau\sqrt{d}\Big\}\\ &= \mathbb{P}\Big\{\|G^t\bar{u}_b\|_2\geq \tau\sqrt{d}\Big\}\\ &\geq \mathbb{P}\Big\{s_{\min}(G)\geq \tau\sqrt{d}\Big\},\end{aligned}$$

where the last inequality holds since $\bar{u}_b\in \mathrm{Im}\,G$. The proof is finished by choosing a sufficiently small $c_{4.21} := \tau$ and applying (4.4). □

Lemma 4.22. *Let $q\in\mathbb{N}$ and $r\in\mathbb{R}$ with $e\leq r\leq\sqrt{\ln q}$, and let $g_1,g_2,\dots,g_q$ be independent standard Gaussian variables. Define a random vector $b=(b_1,b_2,\dots,b_q)\in\mathbb{R}^q$ by $b_i := \max(0, g_i - r)$, $i\leq q$. Then*

$$\mathbb{P}\Big\{\|b\|_2\leq 4\sqrt{q}\exp(-r^2/8)\Big\}\geq 1-\exp(-2\sqrt{q}).$$

Proof. Let $\lambda\in(0,1/2)$. We have

$$\mathbb{E}\,e^{\lambda\|b\|_2^2} = \prod_{i=1}^{q}\mathbb{E}\,e^{\lambda {b_i}^2} = \Big(1+\int_1^\infty \mathbb{P}\{e^{\lambda {b_1}^2}\geq\tau\}d\tau\Big)^q.$$

Next, using (4.2), we get

$$\begin{aligned}\int_1^\infty \mathbb{P}\{e^{\lambda b_1^2}\geq \tau\}d\tau &\leq (r-1)\mathbb{P}\{g_1>r\}+\int_r^\infty \mathbb{P}\{e^{\lambda b_1^2}\geq \tau\}d\tau\\ &\leq e^{-r^2/2}+\int_r^\infty \mathbb{P}\Big\{g_1\geq \sqrt{\frac{\ln\tau}{\lambda}}\Big\}d\tau\\ &\leq e^{-r^2/2}+\int_r^\infty \tau^{-\frac{1}{2\lambda}}d\tau\\ &= e^{-r^2/2}+\frac{r^{1-\frac{1}{2\lambda}}}{\frac{1}{2\lambda}-1}.\end{aligned}$$

Now, take $\lambda=\big(2+\frac{r^2}{\ln r}\big)^{-1}$ so that $\frac{1}{2\lambda}-1=\frac{r^2}{2\ln r}$. After replacing λ with its value, we deduce that

$$\mathbb{E}\,e^{\lambda\|b\|_2^2}\leq \big(1+2e^{-r^2/2}\big)^q\leq \exp(2qe^{-r^2/2}). \tag{4.24}$$

Using Markov's inequality together with (4.24), we obtain

$$\mathbb{P}\{\lambda\|b\|_2^2\geq 4qe^{-r^2/2}\}\leq \exp(-2qe^{-r^2/2})\leq \exp(-2\sqrt{q}),$$

where the last inequality holds since $r\leq\sqrt{\ln q}$. To finish the proof, it remains to note that

$$\frac{4qe^{-r^2/2}}{\lambda}\leq 8qr^2e^{-r^2/2}\leq 16qe^{-r^2/4}.$$

□

4.6.2 Proof of Theorem 4.2

Throughout this part, we assume that $c>0$ and $n_0\in\mathbb{N}$ are appropriately chosen constants (with c sufficiently small and n_0 sufficiently large) and $n\geq n_0$ is fixed. The admissible values for c and n_0 can be recovered from the proof, however, we prefer to avoid these technical details. Further, in order not to overload the presentation, from now on we treat certain real-valued parameters are integers. In particular, this concerns the product cn, as well as several other quantities depending on n (we will point them out later). To prove relation (4.1), we will construct a random unit vector $\bar{v}\in\mathbb{R}^n$ such that

$$\langle\bar{v},\mathrm{BM}_n(t)\rangle_n>0\quad \text{for any } t\in[1,2^{cn}] \tag{4.25}$$

with probability close to one.

Let $N := cn$ and define

$$a_0 := 0 \text{ and } a_i := 2^{i-1}, \ \ i = 1, 2, \ldots, N+1.$$

The starting point of the proof is to define a random vector $\bar{v}_0$ such that $\langle \bar{v}_0, \mathrm{BM}_n(a_i)\rangle_n$ is large for all $i \le N+1$. For this, we will use Lemma 4.21 taking all coordinates of the vector b equal. It will be more convenient to state the next lemma (which is a direct consequence of Lemma 4.21) with generic parameters m and d instead of N, n.

Lemma 4.23. *Let $d, m \in \mathbb{N}$ with $m \le d/2$ and $\mathrm{BM}_d(t)$ be the standard Brownian motion in $\mathbb{R}^d$. Then there exists a random unit vector $\bar{v}_0 \in \mathbb{R}^d$ such that*

$$\mathbb{P}\Big\{\langle \bar{v}_0, \mathrm{BM}_d(a_{i+1}) - \mathrm{BM}_d(a_i)\rangle_d \ge \frac{c_{4.21}}{2}\sqrt{\frac{d a_{i+1}}{m}},\ i = 0, \ldots m\Big\}$$
$$\ge 1 - \exp(-c_{4.21} d).$$

We note that, conditioned on a realization of $\mathrm{BM}_d(a_1), \ldots, \mathrm{BM}_d(a_{m+1})$ (hence, $\bar{v}_0$), for each admissible $i \ge 1$ the process

$$\langle \bar{v}_0, \mathrm{BM}_d\big(a_i + t(a_{i+1} - a_i)\big)\rangle_d, \ \ t \in [0, 1],$$

is a (non-centered) Brownian bridge, and standard estimates (see, for example, [102, p. 34]) together with above lemma imply that given i, we have $\langle \bar{v}_0, \mathrm{BM}_d(a_i + t(a_{i+1} - a_i))\rangle_d > 0$ for all $t \in [0, 1]$ with probability at least $1 - 2\exp(-c'' d/m)$ for a universal constant c''. If $m \ll d/\ln d$ then applying the union bound we get $\langle \bar{v}_0, \mathrm{BM}_d(t)\rangle_d > 0$ for all $1 \le t \le a_{m+1}$ with high probability.

The argument described above is given in [29]. Note that for $m \gg d/\ln d$ the probability that the i-th Brownian bridge is not positive becomes too large to apply the union bound over all i. For this reason, we significantly modified the approach of [29]. Let $M := \log_2 \ln n$ (we will further treat the quantity as an integer, omitting a truncation operation). Our construction will be iterative: after defining vector $\bar{v}_0$ as described above, we will produce a sequence of random vectors $\bar{v}_k$, $k = 1, \ldots, M$, where each $\bar{v}_k$ with a high probability satisfies $\langle \bar{v}_k, \mathrm{BM}_n(t)\rangle_n > 0$ for all t in a certain discrete subset of $[1, 2^{cn}]$. The subset for $\bar{v}_k$ is obtained by zooming in and adding mid-points between every two neighbouring points of the subset generated for $\bar{v}_{k-1}$. The size of those discrete subsets grows with k exponentially, so that the vector $\bar{v} := \bar{v}_M$ will possess the required property (4.25) with probability close to one. The definition of the subsets is made more precise below.

We split the interval $[0, a_{N+1}]$ into blocks. For each admissible $i \geq 0$, *the i-th block* is the interval $[a_i, a_{i+1}]$. With the i-th block, we associate a sequence of sets I_k^i, $k = 0, 1, \ldots, M$, in the following way: for $i = 0$ we have $I_k^i = \emptyset$ for all $k \geq 0$; for $i \geq 1$, we set $I_0^i = \emptyset$ and

$$I_k^i := \{2^{1/2^k} a_i, 2^{2/2^k} a_i, 2^{3/2^k} a_i, \ldots, 2^{(2^k-1)/2^k} a_i\}, \quad k = 1, 2, \ldots, M.$$

Given any $0 < k \leq M$, the vector $\bar{v}_k$ will be a small perturbation of the vector $\bar{v}_{k-1}$. The operation of constructing $\bar{v}_k$ will be referred to as *the k-th step* of the construction. We must admit that the construction is rather technical. In fact, each step itself is divided into a sequence of *substeps*. To make the exposition of the proof as clear as possible, we won't provide all the details at once but instead introduce them sequentially.

At each step, to avoid issues connected with probabilistic dependencies, the already constructed vector $\bar{v}_{k-1}$ and the perturbation added to it will be defined on disjoint coordinate subspaces of $\mathbb{R}^n$. Namely, we split $\mathbb{R}^n$ into $M+1$ coordinate subspaces as follows

$$\mathbb{R}^n := \prod_{k=0}^{M} \mathbb{R}^{J^k},$$

where J^k are pairwise disjoint subsets of $\{1, \ldots, n\}$ with $|J^k| = \tilde{c}n2^{-k/8}$ for an appropriate constant $\tilde{c}$ (chosen so that $\sum_{k \leq M} |J^k| = n$) and $\mathbb{R}^{J^k} = \operatorname{span}\{e_i^n\}_{i \in J^k}$. Again, for a lighter exposition we treat the quantities $\tilde{c}n2^{-k/8}$ as integers. For every $k \leq M$, define $\operatorname{Proj}^k : \mathbb{R}^n \to \mathbb{R}^n$ as the orthogonal projection onto $\mathbb{R}^{J^k}$.

Let $F, H : \mathbb{N} \to \mathbb{R}_+$ be a decreasing and an increasing function, respectively, satisfying the relations

$$8c\, F(1)^2 = \tilde{c}\, c_{4.21}{}^2 \quad \text{and} \quad \forall k \leq M, \quad F(k) \geq C_f \geq 2H(k), \tag{4.26}$$

where $C_f > 0$ is a constant which will be determined later.

Now, we can state more precisely what we mean by the k-th step of the construction $(k = 0, 1, \ldots, M)$. **The goal of the k-th step is to produce a random unit vector**

$\bar{v}_k$ **with the following properties:**

$$\mathbf{1}.\ \bar{v}_k \text{ is supported on } \prod_{p=0}^{k} \mathbb{R}^{J^p}; \tag{4.27}$$

$$\begin{aligned}&\mathbf{2}.\ \bar{v}_k \text{ is measurable with respect to the } \sigma\text{-algebra generated by}\\ &\operatorname{Proj}^p(\mathrm{BM}_n(t)) \text{ for all } 0 \le p \le k,\ t \in \textstyle\bigcup_{i=0}^{N}\big(\{a_{i+1}\} \cup I_k^i\big);\end{aligned} \tag{4.28}$$

3. The event

$$\begin{aligned}\mathcal{E}_k := \Big\{&\langle \bar{v}_k, \mathrm{BM}_n(t) - \mathrm{BM}_n(a_i)\rangle_n \ge -H(k+1)\sqrt{a_i} \textbf{ and}\\ &\langle \bar{v}_k, \mathrm{BM}_n(a_{i+1}) - \mathrm{BM}_n(a_i)\rangle_n \ge F(k+1)\sqrt{a_{i+1}}\\ &\text{for all } t \in I_k^i \text{ and } i = 0, 1, \dots, N\Big\}\end{aligned}$$

has probability close to one.

Quantitative estimates of $\mathbb{P}(\mathcal{E}_k)$ are provided by the following lemma which will be proved in the next section.

Lemma 4.24 (k-th Step)**.** *For a small enough constant $c > 0$ and a large enough $C_f > 0$, there exist F and H satisfying (4.26) such that the following holds. Let $1 \le k \le M$ and assume that a random unit vector $\bar{v}_{k-1}$ satisfying properties* (4.27), (4.28) *has been constructed. Then there exists a random unit vector $\bar{v}_k$ satisfying* (4.27)—(4.28) *and such that*

$$\mathbb{P}(\mathcal{E}_k) \ge \mathbb{P}(\mathcal{E}_{k-1}) - \frac{1}{n^2}.$$

Proof of Theorem 4.2. In view of the relation (4.26), we have

$$2F(1) = c_{4.21}\sqrt{\frac{\tilde{c}}{2c}} \le c_{4.21}\sqrt{\frac{|J^0|}{N}}.$$

Hence, in view of Lemma 4.23 (applied with $m = N$ and $d = |J^0|$), there exists a random unit vector $\bar{v}_0 \in \mathbb{R}^{J^0}$ measurable with respect to the σ-algebra generated by vectors $\operatorname{Proj}^0(\mathrm{BM}_n(a_{i+1}) - \mathrm{BM}_n(a_i))$, $i = 0, 1, \dots, N$, and such that

$$\begin{aligned}\mathbb{P}(\mathcal{E}_0) &= \mathbb{P}\big\{\langle \bar{v}_0, \mathrm{BM}_n(a_{i+1}) - \mathrm{BM}_n(a_i)\rangle_n \ge F(1)\sqrt{a_{i+1}} \text{ for } i = 0, 1, \dots, N\big\}\\ &\ge 1 - \exp(-c_{4.21}|J^0|)\\ &\ge 1 - \frac{1}{n^2}.\end{aligned}$$

Applying Lemma 4.24 M times, we obtain a random unit vector $\bar{v}_M$ satisfying (4.27)–

(4.28) such that

$$\mathbb{P}(\mathcal{E}_M) \geq 1 - \frac{M+1}{n^2}.$$

Note that everywhere on $\mathcal{E}_M$, we have

$$\langle \bar{v}_M, \mathrm{BM}_n(a_{i+1})\rangle_n \geq \langle \bar{v}_M, \mathrm{BM}_n(a_{i+1}) - \mathrm{BM}_n(a_i)\rangle_n \geq C_f\sqrt{a_{i+1}}$$

and

$$\langle \bar{v}_M, \mathrm{BM}_n(t)\rangle_n \geq \langle \bar{v}_M, \mathrm{BM}_n(a_i)\rangle_n - \frac{C_f}{2}\sqrt{a_i} \geq \frac{C_f}{2}\sqrt{a_i}, \quad t \in I_k^i$$

for all $i = 0, 1, \ldots, N$. Hence, denoting $Q := \{a_1, a_2, \ldots, a_{N+1}\} \cup \bigcup_{i=1}^N I_M^i$, we get

$$\mathcal{E}_M \subset \Big\{ \big\langle \bar{v}_M, \frac{\mathrm{BM}_n(t)}{\sqrt{t}} \big\rangle_n \geq \frac{C_f}{4}, \;\; t \in Q \Big\}. \tag{4.29}$$

Now, take any two neighbouring points $t_1 < t_2$ from Q. Note that, conditioned on a realization of vectors $\mathrm{BM}_n(t)$, $t \in Q$, the random process

$$X(s) = \big\langle \bar{v}_M, \frac{s\mathrm{BM}_n(t_2) + (1-s)\mathrm{BM}_n(t_1)}{\sqrt{t_2 - t_1}} \big\rangle_n - \big\langle \bar{v}_M, \frac{\mathrm{BM}_n(t_1 + s(t_2 - t_1))}{\sqrt{t_2 - t_1}} \big\rangle_n,$$

defined for $s \in [0,1]$, is a standard Brownian bridge. Hence (see, for example, [102, p. 34]), we have for any $\tau > 0$

$$\mathbb{P}\big\{X(s) \geq \tau \text{ for some } s \in [0,1]\big\} = \exp(-2\tau^2).$$

Taking $\tau := 2\sqrt{\ln n}$, we obtain

$$\begin{aligned} \mathbb{P}\Big\{\big\langle \bar{v}_M, \mathrm{BM}_n(t)\big\rangle_n \leq \min\big(&\langle \bar{v}_M, \mathrm{BM}_n(t_1)\rangle_n, \langle \bar{v}_M, \mathrm{BM}_n(t_2)\rangle_n\big) \\ &-2\sqrt{t_2 - t_1}\sqrt{\ln n} \text{ for some } t \in [t_1, t_2]\Big\} \\ \leq \frac{1}{n^8}&. \end{aligned}$$

Finally, note that, in view of (4.29), everywhere on $\mathcal{E}_M$ we have

$$\begin{aligned} &(t_2 - t_1)^{-1/2} \min\big(\langle \bar{v}_M, \mathrm{BM}_n(t_1)\rangle_n, \langle \bar{v}_M, \mathrm{BM}_n(t_2)\rangle_n\big) - 2\sqrt{\ln n} \\ &\geq \frac{C_f}{4}\sqrt{\frac{t_1}{t_2 - t_1}} - 2\sqrt{\ln n} \\ &\geq 2^{M/2-3} C_f - 2\sqrt{\ln n} \\ &> 0. \end{aligned}$$

Taking the union bound over all adjacent pairs in Q (clearly, $|Q| \le n^2$), we come to the relation

$$\mathbb{P}\big\{\langle \bar v_M, \mathrm{BM}_n(t)\rangle_n > 0 \text{ for all } t \in [1, 2^{cn}]\big\} \ge \mathbb{P}(\mathcal{E}_M) - \frac{|Q|}{n^8} \ge 1 - \frac{1}{n}.$$

□

4.6.3 Proof of Lemma 4.24

Let $M' = \frac{1}{4}\log_2 \ln n$. For every $k \le M$, we split J^k into pairwise disjoint subsets J^k_ℓ, $\ell \le M'$, with $|J^k_\ell| = c' n 2^{-(k+\ell)/8}$ for an appropriate constant c', chosen so that $\sum_{\ell \le M'} |J^k_\ell| = |J^k|$ (to make computations lighter, we will treat the quantities $c'n2^{-(k+\ell)/8}$, $k \le M, \ell \le M'$, as integers). For every $k \le M, \ell \le M'$, define $\mathrm{Proj}^k_\ell : \mathbb{R}^n \to \mathbb{R}^n$ as the orthogonal projection onto $\mathbb{R}^{J^k_\ell}$.

Further, we define two functions $f, h : \mathbb{N} \times \mathbb{N}_0 \to \mathbb{R}_+$ as follows:

1. f is decreasing in both arguments; $f(1,0) = C_f + 2^{-1/2}(1-2^{-1/4})^{-2} C_f$; for each $k > 0$ and $\ell > 0$ we have $f(k, \ell-1) - f(k,\ell) = C_f 2^{-(k+\ell)/4}$; finally, $f(k, 0) = \lim\limits_{\ell\to\infty} f(k-1, \ell)$ for all $k > 1$. The constant $C_f > 0$ is defined via the relation $8cf(1,0)^2 = \tilde c {c_{4.21}}^2$, where $\tilde c$ is taken from the definition of sets J^k and $c_{4.21}$ comes from Lemma 4.21.

2. h is increasing in both arguments; $h(1,0) = 0$; for each $k > 0$ and $\ell > 0$ we have $h(k,\ell) - h(k, \ell-1) = C_h 2^{-(k+\ell)/4}$; moreover, $h(k,0) = \lim\limits_{\ell\to\infty} h(k-1,\ell)$ for all $k > 1$. The constant C_h is defined by $C_h = 2^{-1/2}(1-2^{-1/4})^2 C_f$.

Now define $F : \mathbb{N} \to \mathbb{R}$ and $H : \mathbb{N} \to \mathbb{R}$ by $F(k) := f(k,0)$ and $H(k) := h(k,0)$ for any $k \in \mathbb{N}$. Note that F and H satisfy (4.26).

Fix $k \ge 1$. Assuming that the vector $\bar v_{k-1}$ is already constructed, the aim is to construct $\bar v_k$ such that the event $\mathcal{E}_k$ has large probability. The vector $\bar v_k$ is obtained via an embedded iteration procedure realized as a sequence of substeps. Namely, we set $\bar v_{k,0} := \bar v_{k-1}$ and inductively construct random vectors $\bar v_{k,\ell}$, $1 \le \ell \le M'$ and take $\bar v_k = \bar v_{k,M'}$. Let us give a partial description of the procedure, omitting some details.

For each $\ell = 1, 2, \ldots, M'+1$ and every block $i = 0, 1, 2, \ldots, N$ *the i-th block statistic* is

$$\begin{aligned}\mathcal{B}_i(k,\ell) := \max\Big(0, \max_{t \in I^i_k} \big\langle \bar v_{k,\ell-1}, \frac{\mathrm{BM}_n(a_i) - \mathrm{BM}_n(t)}{\sqrt{a_i}}\big\rangle_n - h(k,\ell), \\ \big\langle \bar v_{k,\ell-1}, \frac{\mathrm{BM}_n(a_i) - \mathrm{BM}_n(a_{i+1})}{\sqrt{a_{i+1}}}\big\rangle_n + f(k,\ell)\Big).\end{aligned} \tag{4.30}$$

Note that the statistic for the zero block is simply

$$\max\Big(0, -\big\langle \bar{v}_{k,\ell-1}, \mathrm{BM}_n(a_1)\big\rangle_n + f(k,\ell)\Big).$$

The $(N+1)$-dimensional vector $\big(\mathcal{B}_0(k,\ell), \dots, \mathcal{B}_N(k,\ell)\big)$ will be denoted by $\mathcal{B}(k,\ell)$. Let us also denote

$$\mathcal{I}(k,\ell) := \big\{i : \mathcal{B}_i(k,\ell) \neq 0\big\}.$$

Note that the event $\{\mathcal{I}(k, M'+1) = \emptyset\}$ is contained inside $\mathcal{E}_k$. At each substep, using information about the statistics $\mathcal{B}(k,\ell)$ and choosing an appropriate perturbation of $\bar{v}_{k,\ell-1}$ to obtain $\bar{v}_{k,\ell}$, we will control the measure of the event $\{\mathcal{I}(k,\ell+1) = \emptyset\}$, and in this way will be able to estimate the probability of $\mathcal{E}_k$ from below.

Given $\bar{v}_{k,\ell-1}$, the goal of the ℓ-th substep is to construct a random unit vector $\bar{v}_{k,\ell}$ such that

1. $\bar{v}_{k,\ell}$ is supported on $\prod_{(p,q)\precsim(k,\ell)} \mathbb{R}^{J_q^p}$, where the notation $(p,q) \precsim (k,\ell)$ means "$p < k$ or $p = k,\ q \leq \ell$"; (4.31)

2. $\bar{v}_{k,\ell}$ is measurable with respect to the σ-algebra generated by $\mathrm{Proj}_q^p(\mathrm{BM}_n(t))$ for all $(p,q) \precsim (k,\ell)$ and $t \in \bigcup_{i=0}^N \big(\{a_{i+1}\} \cup I_k^i\big)$; (4.32)

3. $\|\mathcal{B}(k,\ell+1)\|_2$ is typically smaller than $\|\mathcal{B}(k,\ell)\|_2$.

The third property will be made more precise later. For now, we note that the typical value of $\|\mathcal{B}(k,\ell)\|_2$ will decrease with ℓ in such a way that, after the M'-th substep, the vector $\mathcal{B}(k, M'+1)$ will be zero with probability close to one.

The vector $\bar{v}_{k,\ell}$ will be defined as

$$\bar{v}_{k,\ell} = \frac{\bar{v}_{k,\ell-1} + \alpha_{k,\ell}\bar{\Delta}_{k,\ell}}{\sqrt{1+{\alpha_{k,\ell}}^2}}, \tag{4.33}$$

where $\bar{\Delta}_{k,\ell}$ is a random unit vector (perturbation) and $\alpha_{k,\ell} := 16^{-k-\ell}$.

The vector $\bar{\Delta}_{k,\ell}$ will satisfy the following properties:

1. $\bar{\Delta}_{k,\ell}$ is supported on $\mathbb{R}^{J_\ell^k}$; (4.34)

2. $\bar{\Delta}_{k,\ell}$ is measurable with respect to the σ-algebra generated by $\mathrm{Proj}_q^p(\mathrm{BM}_n(t))$ for all admissible $(p,q) \precsim (k,\ell)$, $t \in \bigcup_{i=0}^N \big(\{a_{i+1}\} \cup I_k^i\big)$; (4.35)

3. For any subset $I \subset \{0,1,\dots,N\}$ such that $\mathbb{P}\{\mathcal{I}(k,\ell) = I\} > 0$, $\bar{\Delta}_{k,\ell}$ is *conditionally* independent from the collection of vectors

$$\big\{\mathrm{Proj}_\ell^k(\mathrm{BM}_n(t) - \mathrm{BM}_n(a_i)),\ \ t \in I_k^i \cup \{a_{i+1}\},\ \ i \notin I\big\} \tag{4.36}$$

given the event $\{\mathcal{I}(k,\ell) = I\}$.

4. The event

$$\mathcal{E}_{k,\ell} := \big\{\mathcal{B}_i(k,\ell+1) = 0 \text{ for all } i \in \mathcal{I}(k,\ell)\big\}$$

has probability close to one.

Again, we will make the last property more precise later.

Let us sum up the construction procedure. We sequentially produce random unit vectors $\bar{v}_0 = \bar{v}_{1,0},\ \bar{v}_{1,1},\ \bar{v}_{1,2},\dots,\ \bar{v}_{1,M'} = \bar{v}_1 = \bar{v}_{2,0},\ \bar{v}_{2,1},\ \bar{v}_{2,2},\dots,\ \bar{v}_{2,M'} = \bar{v}_2 = \bar{v}_{3,0},\dots,$ $\dots,\ \bar{v}_{M,M'} = \bar{v}_M$ (in the given order). Each next vector is a random perturbation of the previous one. In a certain sense (quantified with help of order statistics $\mathcal{B}(k,\ell)$), each newly produced vector is a refinement of the previous one in such a way that $\bar{v}_M = \bar{v}$ will possess the required characteristics.

In the next two lemmas, we establish certain important properties of the block statistics.

Lemma 4.25 (Initial substep for block statistics)**.** *Fix any $1 \le k \le M$ and assume that a random unit vector $\bar{v}_{k,0} := \bar{v}_{k-1}$ satisfying properties* (4.27) *and* (4.28) *has been constructed. Then*

$$\mathbb{P}\Big\{|\mathcal{I}(k,1)| \le N\exp(-{C_h}^2 2^{k/2}/16) \textbf{ and } \|\mathcal{B}(k,1)\|_2 \le \frac{8\sqrt{N}}{\exp({C_h}^2 2^{k/2}/32)}\Big\}$$
$$\ge \mathbb{P}(\mathcal{E}_{k-1}) - 2\exp(-2\sqrt{N}).$$

Proof. Let $i > 0$ so that $I_k^i \neq \emptyset$. For each $t \in I_k^i \setminus I_{k-1}^i$, let t_L be the maximal number in $\{a_i\} \cup I_{k-1}^i$ strictly less than t (“left neighbour”) and, similarly, t_R be the minimal number in $I^i_{k-1} \cup \{a_{i+1}\}$ strictly greater than t ("right neighbour"). For every such t, let

$$w_t := \frac{t_R - t}{t_R - t_L}\mathrm{BM}_n(t_L) + \frac{t - t_L}{t_R - t_L}\mathrm{BM}_n(t_R); \quad u_t := \mathrm{BM}_n(t) - w_t.$$

It is not difficult to see that

$$\begin{aligned}
&\Big\langle \bar v_{k,0}, \frac{\mathrm{BM}_n(a_i) - w_t}{\sqrt{a_i}}\Big\rangle_n \\
&\le \max\Big(\Big\langle \bar v_{k,0}, \frac{\mathrm{BM}_n(a_i) - \mathrm{BM}_n(t_L)}{\sqrt{a_i}}\Big\rangle_n, \Big\langle \bar v_{k,0}, \frac{\mathrm{BM}_n(a_i) - \mathrm{BM}_n(t_R)}{\sqrt{a_i}}\Big\rangle_n\Big) \\
&\le \max\Big(0, \max_{\tau \in I^i_{k-1}} \Big\langle \bar v_{k,0}, \frac{\mathrm{BM}_n(a_i) - \mathrm{BM}_n(\tau)}{\sqrt{a_i}}\Big\rangle_n, \\
&\qquad \Big\langle 2\bar v_{k,0}, \frac{\mathrm{BM}_n(a_i) - \mathrm{BM}_n(a_{i+1})}{\sqrt{a_{i+1}}}\Big\rangle_n\Big).
\end{aligned}$$

Hence, the i-th block statistic (for $i = 0, 1, \ldots, N$) can be (deterministically) bounded as

$$\begin{aligned}
\mathcal{B}_i(k,1) &\le \max\Big(0, \max_{t \in I^i_{k-1}} \Big\langle \bar v_{k,0}, \frac{\mathrm{BM}_n(a_i) - \mathrm{BM}_n(t)}{\sqrt{a_i}}\Big\rangle_n - h(k,1), \\
&\qquad \max_{t \in I^i_k \setminus I^i_{k-1}} \Big\langle \bar v_{k,0}, \frac{\mathrm{BM}_n(a_i) - w_t}{\sqrt{a_i}}\Big\rangle_n - h(k,1) + \max_{t \in I^i_k \setminus I^i_{k-1}} \Big\langle \bar v_{k,0}, \frac{-u_t}{\sqrt{a_i}}\Big\rangle_n, \\
&\qquad \Big\langle \bar v_{k,0}, \frac{\mathrm{BM}_n(a_i) - \mathrm{BM}_n(a_{i+1})}{\sqrt{a_{i+1}}}\Big\rangle_n + f(k,1)\Big) \\
&\le \max\Big(0, \max_{t \in I^i_{k-1}} \Big\langle \bar v_{k,0}, \frac{\mathrm{BM}_n(a_i) - \mathrm{BM}_n(t)}{\sqrt{a_i}}\Big\rangle_n - h(k,0), \\
&\qquad \Big\langle 2\bar v_{k,0}, \frac{\mathrm{BM}_n(a_i) - \mathrm{BM}_n(a_{i+1})}{\sqrt{a_{i+1}}}\Big\rangle_n + 2f(k,0)\Big) \\
&\qquad + \max\Big(0, \max_{t \in I^i_k \setminus I^i_{k-1}} \Big\langle \bar v_{k,0}, \frac{-u_t}{\sqrt{a_i}}\Big\rangle_n + h(k,0) - h(k,1)\Big).
\end{aligned}$$

Let us denote the first summand in the last estimate by ξ_i, so that

$$\mathcal{B}_i(k,1) \le \xi_i + \max\Big(0, \max_{t \in I^i_k \setminus I^i_{k-1}} \Big\langle \bar v_{k,0}, \frac{-u_t}{\sqrt{a_i}}\Big\rangle_n + h(k,0) - h(k,1)\Big).$$

Note that

$$\mathcal{E}_{k-1} = \big\{\xi_i = 0 \text{ for all } i = 0, 1, \ldots, N\big\}. \tag{4.37}$$

Further, the property (4.28) of the vector $\bar v_{k,0} = \bar v_{k-1}$, together with Lemma 4.20 and the independence of the Brownian motion on disjoint intervals, imply that the Gaussian variables $\big\langle \bar v_{k,0}, \frac{-u_t}{\sqrt{a_i}}\big\rangle_n$ are jointly independent for $t \in I^i_k \setminus I^i_{k-1}$, $i = 1, 2, \ldots, N$, and the variance of each one can be estimated from above by 2^{1-k}. Thus, the vector $\mathcal{B}(k,1)$ can be majorized coordinate-wise by the vector

$$\big(\xi_i + \max_{t\in I_k^i\setminus I_{k-1}^i}(0, 2^{(1-k)/2}g_t + h(k,0) - h(k,1))\big)_{i=0}^N,$$

where g_t ($t \in I_k^i \setminus I_{k-1}^i$, $i = 0,1,\dots,N$) are i.i.d. standard Gaussians (in fact, appropriate scalar multiples of $\langle \bar v_{k,0}, \frac{-u_t}{\sqrt{a_i}}\rangle_n$). Denoting by g the standard Gaussian variable, we get from the definition of h:

$$\begin{aligned}\mathbb{P}\big\{\max_{t\in I_k^i\setminus I_{k-1}^i}(0, 2^{(1-k)/2}g_t + h(k,0) - h(k,1)) > 0\big\} &\le 2^k\mathbb{P}\{g > C_h 2^{k/4}/2\}\\ &\le 2^k\exp(-{C_h}^2 2^{k/2}/8)\\ &\le \frac{1}{2}\exp(-{C_h}^2 2^{k/2}/16).\end{aligned}$$

(In the last two inequalities, we assumed that C_h is sufficiently large). Applying Hoeffding's inequality to corresponding indicators, we infer

$$|\mathcal{I}(k,1)| \le |\{i:\, \xi_i \ne 0\}| + N\exp(-{C_h}^2 2^{k/2}/16)$$

with probability at least $1-\exp(-2\sqrt{N})$ (we note that, in view of the inequality $k \le M$, we have $\frac{1}{2}\exp(-{C_h}^2 2^{k/2}/16) \ge N^{-1/4}$). Next, it is not hard to see that the Euclidean norm of $\mathcal{B}(k,1)$ is majorized (deterministically) by the sum

$$\big\|(\xi_i)_{i=0}^N\big\|_2 + 2^{(1-k)/2}\big\|\big(\max(0, g_t - C_h 2^{k/4}/2)\big)_t\big\|_2,$$

with the second vector having $\sum_{i=0}^N |I_k^i \setminus I_{k-1}^i| \le 2^k N$ coordinates. Applying Lemma 4.22 to the second vector (note that for sufficiently large n we have $C_h 2^{k/4}/2 \le \sqrt{\ln N}$), we get

$$\|\mathcal{B}(k,1)\|_2 \le \big\|(\xi_i)_{i=0}^N\big\|_2 + \frac{8\sqrt{N}}{\exp({C_h}^2 2^{k/2}/32)}$$

with probability at least $1-\exp(-2\sqrt{N})$. Combining the estimates with (4.37), we obtain the result. □

Lemma 4.26 (Subsequent substeps for block statistics)**.** *Fix any* $1 \le k \le M$ *and* $1 < \ell \le M'+1$ *and assume that the random unit vectors* $\bar v_{k,\ell-2}$ *and* $\bar\Delta_{k,\ell-1}$ *satisfying properties* (4.31)—(4.32) *and* (4.34)—(4.35)—(4.36)*, respectively, are constructed, and*

$\bar{v}_{k,\ell-1}$ *is defined according to formula* (4.33)*. Then*

$$\mathbb{P}\Big\{|\mathcal{I}(k,\ell)| \le N\exp(-{C_h}^2 2^{(k+\ell)/2}) \textbf{ and } \|\mathcal{B}(k,\ell)\|_2 \le \frac{\sqrt{N}}{\exp({C_h}^2 2^{(k+\ell)/2})}\Big\}$$
$$\ge \mathbb{P}(\mathcal{E}_{k,\ell-1}) - 2\exp(-2\sqrt{N}).$$

Moreover,

$$\mathbb{P}\big\{\mathcal{I}(k,\ell) \ne \emptyset\big\} \le N\exp(-{C_h}^2/\alpha_{k,\ell-1}) + 1 - \mathbb{P}(\mathcal{E}_{k,\ell-1}).$$

Proof. To shorten the notation, we will use α in place of $\alpha_{k,\ell-1}$ within the proof. Using the definition of $\bar{v}_{k,\ell-1}$ in terms of $\bar{v}_{k,\ell-2}$ and $\bar{\Delta}_{k,\ell-1}$, we get for every $i = 0, 1, \dots, N$

$$\begin{aligned}
\mathcal{B}_i(k,\ell) = \max\Big(0, \max_{t\in I_k^i}\big\langle \frac{\bar{v}_{k,\ell-2}+\alpha\bar{\Delta}_{k,\ell-1}}{\sqrt{1+\alpha^2}}, \frac{\mathrm{BM}_n(a_i)-\mathrm{BM}_n(t)}{\sqrt{a_i}}\big\rangle_n - h(k,\ell),\\
\big\langle \frac{\bar{v}_{k,\ell-2}+\alpha\bar{\Delta}_{k,\ell-1}}{\sqrt{1+\alpha^2}}, \frac{\mathrm{BM}_n(a_i)-\mathrm{BM}_n(a_{i+1})}{\sqrt{a_{i+1}}}\big\rangle_n + f(k,\ell)\Big)\\
\le \frac{\mathcal{B}_i(k,\ell-1)}{\sqrt{1+\alpha^2}}\\
+ \max\Big(0, \max_{t\in I_k^i}\big\langle \alpha\bar{\Delta}_{k,\ell-1}, \frac{\mathrm{BM}_n(a_i)-\mathrm{BM}_n(t)}{\sqrt{a_i}}\big\rangle_n + h(k,\ell-1) - h(k,\ell),\\
\big\langle \alpha\bar{\Delta}_{k,\ell-1}, \frac{\mathrm{BM}_n(a_i)-\mathrm{BM}_n(a_{i+1})}{\sqrt{a_{i+1}}}\big\rangle_n + \sqrt{1+\alpha^2}f(k,\ell) - f(k,\ell-1)\Big).
\end{aligned}$$

Let us denote the second summand by η_i so that

$$\mathcal{B}_i(k,\ell) \le \frac{\mathcal{B}_i(k,\ell-1)}{\sqrt{1+\alpha^2}} + \eta_i.$$

Fix for a moment any subset I of $\{0,1,\dots,N\}$ such that $\mathbb{P}\{\mathcal{I}(k,\ell-1) = I\} > 0$. A crucial observation is that, conditioned on the event $\mathcal{I}(k,\ell-1) = I$, the variables η_i, $i \notin I$, are jointly independent. This follows from properties (4.34), (4.36) of $\bar{\Delta}_{k,\ell-1}$ and from independence of the Brownian motion on disjoint intervals. Next, the same properties tell us that, conditioned on $\mathcal{I}(k,\ell-1) = I$, each variable $\langle \bar{\Delta}_{k,\ell-1}, \frac{\mathrm{BM}_n(a_i)-\mathrm{BM}_n(t)}{\sqrt{a_i}}\rangle_n$, $t \in I_k^i$, and $\langle \bar{\Delta}_{k,\ell-1}, \frac{\mathrm{BM}_n(a_i)-\mathrm{BM}_n(a_{i+1})}{\sqrt{a_{i+1}}}\rangle_n$ have Gaussian distributions with variances at most 1. Further, note that, by the choice of α and the functions f and h, we have

$$\sqrt{1+\alpha^2}f(k,\ell) - f(k,\ell-1) \le h(k,\ell-1) - h(k,\ell) = -C_h 2^{(-k-\ell)/4}.$$

Thus, denoting by g the standard Gaussian variable, we get

$$\begin{aligned}\mathbb{P}\{\eta_i > 0 \,|\, \mathcal{I}(k,\ell-1) = I\} &\le 2^k \mathbb{P}\{g > \alpha^{-1} C_h 2^{(-k-\ell)/4}\} \\ &\le \frac{1}{2}\exp(-{C_h}^2\alpha^{-1}), \quad i \in \{0,1,\ldots,N\} \setminus I. \end{aligned} \tag{4.38}$$

Hence, by Hoeffding's inequality (note that $\exp(-{C_h}^2 2^{(k+\ell)/2}) > 2N^{-1/4}$):

$$\mathbb{P}\big\{|\{i \notin I :\, \eta_i > 0\}| \ge N\exp(-{C_h}^2 2^{(k+\ell)/2}) \,|\, \mathcal{I}(k,\ell-1) = I\big\} \le \exp(-2\sqrt{N}).$$

Next, it is not difficult to see that for any $\tau > 0$ and $i \notin I$

$$\begin{aligned}&\mathbb{P}\{\eta_i^2 \ge \tau \,|\, \mathcal{I}(k,\ell-1) = I\} \\ &\quad\le 2^k \mathbb{P}\{\max(0, \alpha g - C_h 2^{(-k-\ell)/4})^2 \ge \tau\} \\ &\quad\le 1 - \exp\big(-2^{k+1}\mathbb{P}\{\max(0, \alpha g - C_h 2^{(-k-\ell)/4})^2 \ge \tau\}\big) \\ &\quad\le 1 - \mathbb{P}\big\{\max(0, \alpha g - C_h 2^{(-k-\ell)/4})^2 < \tau\big\}^{2^{k+1}} \\ &\quad\le \mathbb{P}\Big\{\sum_{j=1}^{2^{k+1}} \max(0, \alpha g_j - C_h 2^{(-k-\ell)/4})^2 \ge \tau\Big\} \\ &\quad\le \mathbb{P}\Big\{\sum_{j=1}^{2^{k+1}} \max(0, \alpha g_j - 4\alpha C_h 2^{(k+\ell)/4})^2 \ge \tau\Big\},\end{aligned}$$

where g_j ($j = 1, 2, \ldots, 2^{k+1}$) are i.i.d. copies of g. Hence, the conditional cdf of $\|(\eta_i)_{i\notin I}\|_2$ given $\mathcal{I}(k,\ell-1) = I$ majorizes the cdf of

$$\alpha\big\|\big(\max(0, g_j - 4C_h 2^{(k+\ell)/4})\big)_{j=1}^{2^{k+1}N}\big\|_2 =: \alpha Z$$

for i.i.d. standard Gaussians g_j, $j = 1, 2, \ldots, 2^{k+1}N$. Applying Lemma 4.22 (note that $4C_h 2^{(k+\ell)/4} \le \sqrt{\ln N}$), we obtain

$$\begin{aligned}&\mathbb{P}\Big\{\|(\eta_i)_{i\notin I}\|_2 > \frac{\sqrt{N}}{\exp({C_h}^2 2^{(k+\ell)/2})} \,\Big|\, \mathcal{I}(k,\ell-1) = I\Big\} \\ &\quad\le \mathbb{P}\Big\{Z > \frac{\alpha^{-1}\sqrt{N}}{\exp({C_h}^2 2^{(k+\ell)/2})} \,\Big|\, \mathcal{I}(k,\ell-1) = I\Big\} \\ &\quad\le \mathbb{P}\Big\{Z > \frac{4\sqrt{2^{k+1}N}}{\exp(2{C_h}^2 2^{(k+\ell)/2})} \,\Big|\, \mathcal{I}(k,\ell-1) = I\Big\} \\ &\quad\le \exp(-2\sqrt{N}).\end{aligned}$$

Clearly $\mathcal{B}_i(k,\ell-1)=0$ for all $i\notin I$ given $\mathcal{I}(k,\ell-1)=I$. Hence, the above estimates give

$$\begin{aligned}&\mathbb{P}\Big\{|\mathcal{I}(k,\ell)|\geq N\exp(-{C_h}^2 2^{(k+\ell)/2})\\&\qquad\text{or } \|\mathcal{B}(k,\ell)\|_2>\frac{\sqrt{N}}{\exp({C_h}^2 2^{(k+\ell)/2})}\,\Big|\,\mathcal{I}(k,\ell-1)=I\Big\}\\&\qquad\leq\mathbb{P}\big\{\mathcal{B}_i(k,\ell)>0\text{ for some }i\in I\,|\,\mathcal{I}(k,\ell-1)=I\big\}+2\exp(-2\sqrt{N}).\end{aligned}$$

Summing over all admissible subsets I, we get

$$\begin{aligned}&\mathbb{P}\Big\{|\mathcal{I}(k,\ell)|\geq N\exp(-{C_h}^2 2^{(k+\ell)/2})\textbf{ or }\|\mathcal{B}(k,\ell)\|_2>\frac{\sqrt{N}}{\exp({C_h}^2 2^{(k+\ell)/2})}\Big\}\\&\leq 2\exp(-2\sqrt{N})\\&\quad+\sum_I\mathbb{P}\big\{\mathcal{B}_i(k,\ell)>0\text{ for some }i\in I\,|\,\mathcal{I}(k,\ell-1)=I\big\}\mathbb{P}\{\mathcal{I}(k,\ell-1)=I\}\\&=2\exp(-2\sqrt{N})+\mathbb{P}\big\{\mathcal{B}_i(k,\ell)>0\text{ for some }i\in\mathcal{I}(k,\ell-1)\big\}\\&=2\exp(-2\sqrt{N})+1-\mathbb{P}(\mathcal{E}_{k,\ell-1}).\end{aligned}$$

By analogous argument, as a corollary of (4.38),

$$\mathbb{P}\big\{\mathcal{I}(k,\ell)\neq\emptyset\big\}\leq N\exp(-{C_h}^2\alpha^{-1})+1-\mathbb{P}(\mathcal{E}_{k,\ell-1}).$$

□

The next lemma, which is the heart of the proof, provides a construction procedure for the perturbation $\bar{\Delta}_{k,\ell}$. Given vector $\bar{v}_{k,\ell-1}$, we examine its block statistics $\mathcal{B}(k,\ell)$, and define the perturbation in such a way that its inner product with increments of the Brownian motion is large on bad blocks $\mathcal{I}(k,\ell)$ (in fact, it will be proportional to the values of corresponding $\mathcal{B}_i(k,\ell)$), and random on other blocks. This is achieved using Lemma 4.21.

Lemma 4.27 (Construction of $\bar{\Delta}_{k,\ell}$)**.** *Let $1\leq k\leq M$ and $1\leq\ell\leq M'$ and assume that the random unit vector $\bar{v}_{k,\ell-1}$ satisfying properties* (4.31) *and* (4.32) *has been constructed. Then one can construct a random unit vector $\bar{\Delta}_{k,\ell}$ satisfying properties* (4.34)—(4.35)—(4.36) *and such that*

$$\begin{aligned}\mathbb{P}(\mathcal{E}_{k,\ell})&\geq\mathbb{P}(\mathcal{E}_{k,\ell-1})-3\exp(-\sqrt{N})\qquad\textit{if } \ell>1,\textit{ or}\\\mathbb{P}(\mathcal{E}_{k,\ell})&\geq\mathbb{P}(\mathcal{E}_{k-1})-3\exp(-\sqrt{N})\qquad\textit{if } \ell=1.\end{aligned}$$

Proof. Fix for a moment any subset $I \subset \{0, 1, \ldots, N\}$ such that the event

$$\Gamma_I = \{\mathcal{I}(k, \ell) = I\}$$

has a non-zero probability. If $|I| > N \exp(-{C_h}^2 2^{(k+\ell)/2}/32)$ then define a random vector $\bar{\Delta}^I_{k,\ell}$ on Γ_I by setting $\bar{\Delta}^I_{k,\ell} := u$ for an arbitrary fixed unit vector $u \in \mathbb{R}^{J^k_\ell}$. Otherwise, if $|I| \le N \exp(-{C_h}^2 2^{(k+\ell)/2}/32)$, we proceed as follows:

Define a set of double indices

$$T_I := \big\{(i,p) :\, i \in I \setminus \{0\},\, p \in \{1, \ldots, 2^k - 1\}\big\} \cup \bigcup_{i \in I} \{(i, 0)\}.$$

For each $(i,p) \in T_I$, define an increment $X_{i,p}$ on the probability space $(\Gamma_I, \mathbb{P}(\cdot|\Gamma_I))$ by

$$X_{i,p} := \frac{\mathrm{Proj}^k_\ell\big(\mathrm{BM}_n(t_{i,p+1}) - \mathrm{BM}_n(t_{i,p})\big)}{\sqrt{t_{i,p+1} - t_{i,p}}},$$

where $t_{i,p} = 2^{i-1+p2^{-k}}$ for $p = 0, 1, \ldots, 2^k$ and $i \in I \setminus \{0\}$; additionally, if $0 \in I$, then $t_{0,1} = 1$ and $t_{0,0} = 0$.

Note that $\mathcal{B}(k, \ell)$ is measurable with respect to the σ-algebra generated by processes $\mathrm{Proj}^q_s \mathrm{BM}_n(t)$, $(q, s) \precsim (k, \ell - 1)$, where the notation "$\precsim$" is taken from (4.31); see formula (4.30). It implies that $\mathrm{Proj}^k_\ell(\mathrm{BM}_n(t))$ (on Ω) is independent from the event Γ_I; moreover, considered on the space $(\Gamma_I, \mathbb{P}(\cdot|\Gamma_I))$, the set $\{X_{i,p},\ (i,p) \in T_I\}$ is a collection of standard Gaussian vectors, such that all $X_{i,p}$ and the vector $\mathcal{B}(k, \ell)$ are *jointly independent*. Let us define a random vector $\tilde{b}^I \in \mathbb{R}^{T_I}$ on $(\Gamma_I, \mathbb{P}(\cdot|\Gamma_I))$ by

$$\tilde{b}^I_{i,p} = \begin{cases} 2^{-k/2}\mathcal{B}_i(k,\ell)/\|\mathcal{B}(k,\ell)\|_2, & \text{if } \mathcal{B}(k,\ell) \neq \mathbf{0}; \\ 0, & \text{otherwise.} \end{cases}$$

It is easy to see that $\|\tilde{b}^I\|_2 \le 1$ (deterministically) and that

$$|T_I| \le 2^k |I| \le 2^k N \exp(-{C_h}^2 2^{(k+\ell)/2}/32) \le \frac{1}{2}|J^k_\ell|.$$

(In the last estimate, we used the assumption that C_h is a large constant). Hence, in view of Lemma 4.21, there exists a random unit vector $\bar{\Delta}^I_{k,\ell}$ on the space $(\Gamma_I, \mathbb{P}(\cdot|\Gamma_I))$ with values in $\mathbb{R}^{J_\ell^k}$, which is a Borel function of $X_{i,p}$ and $\tilde{b}^I$, and such that

$$\mathbb{P}\Big\{\langle\bar{\Delta}^I_{k,\ell}, X_{i,p}\rangle_{|J_\ell^k|} \geq c_{4.21}\sqrt{|J_\ell^k|}\,\tilde{b}^I_{i,p} \text{ for all } (i,p)\in T_I \,|\, \Gamma_I\Big\} \geq 1-\exp(-c_{4.21}|J_\ell^k|)$$
$$\geq 1-\exp(-\sqrt{N}).$$

It will be convenient for us to denote by $\tilde{\Gamma}_I$ the event

$$\Big\{\langle\bar{\Delta}^I_{k,\ell}, X_{i,p}\rangle_{|J_\ell^k|} \geq c_{4.21}\sqrt{|J_\ell^k|}\,\tilde{b}^I_{i,p} \text{ for all } (i,p)\in T_I\Big\} \subset \Gamma_I.$$

By glueing together $\bar{\Delta}^I_{k,\ell}$ for all I, we obtain a random vector $\bar{\Delta}_{k,\ell}$ defined on the entire probability space Ω.

Clearly, $\bar{\Delta}_{k,\ell}$ satisfies properties (4.34) and (4.35). Next, on each Γ_I with $\mathbb{P}(\Gamma_I)>0$ the vector $\bar{\Delta}_{k,\ell}$ was defined as a Borel function of $\mathcal{B}(k,\ell)$ and $\mathrm{Proj}_\ell^k(\mathrm{BM}_n(t)-\mathrm{BM}_n(\tau))$, $t,\tau\in I_k^i\cup\{a_i,a_{i+1}\}$, $i\in I$, so, in view of independence of the Brownian motion on disjoint intervals, $\bar{\Delta}_{k,\ell}$ satisfies (4.36).

Finally, we shall estimate the probability of $\mathcal{E}_{k,\ell}$. Define

$$\mathcal{E}=\Big\{|\mathcal{I}(k,\ell)|\leq N\exp(-{C_h}^2 2^{(k+\ell)/2}/32) \textbf{ and }$$
$$\|\mathcal{B}(k,\ell)\|_2 \leq \frac{\sqrt{N}}{\exp({C_h}^2 2^{(k+\ell)/2}/64)}\Big\}.$$

Note that, according to Lemmas 4.25 and 4.26, the probability of $\mathcal{E}$ can be estimated from below by $\mathbb{P}(\mathcal{E}_{k,\ell-1})-2\exp(-2\sqrt{N})$ for $\ell>1$ and $\mathbb{P}(\mathcal{E}_{k-1})-2\exp(-2\sqrt{N})$ for $\ell=1$.

Take any subset $I\subset\{0,1,\ldots,N\}$ with $|I|\leq N\exp(-{C_h}^2 2^{(k+\ell)/2}/32)$ and such that $\tilde{\Gamma}_I\cap\mathcal{E}\neq\emptyset$, and let $\omega\in\tilde{\Gamma}_I\cap\mathcal{E}$. If $\mathcal{I}(k,\ell)=\emptyset$ at point ω then, obviously, $\omega\in\mathcal{E}_{k,\ell}$. Otherwise, we have

$$\Big\langle\bar{\Delta}_{k,\ell}(\omega), \frac{\mathrm{BM}_n(t_{i,p+1})(\omega)-\mathrm{BM}_n(t_{i,p})(\omega)}{\sqrt{t_{i,p+1}-t_{i,p}}}\Big\rangle_n$$
$$\geq \frac{c_{4.21}2^{-k/2}\sqrt{|J_\ell^k|}\,\mathcal{B}_i(k,\ell)(\omega)}{\|\mathcal{B}(k,\ell)(\omega)\|_2} \text{ for all } (i,p)\in T_I,$$

whence, using the estimate $t_{i,p+1}-t_{i,p}\geq\frac{2^{i-k}}{4}$ ($(i,p)\in T_I$), we obtain for any $i\in I$ and

$t \in I_k^i \cup \{a_{i+1}\}$:

$$\begin{aligned}
&\big\langle \bar{\Delta}_{k,\ell}(\omega), \mathrm{BM}_n(t)(\omega) - \mathrm{BM}_n(a_i)(\omega)\big\rangle_n \\
&\quad = \sum_{p:\, t_{i,p} < t} \big\langle \bar{\Delta}_{k,\ell}(\omega), \mathrm{BM}_n(t_{i,p+1})(\omega) - \mathrm{BM}_n(t_{i,p})(\omega)\big\rangle_n \\
&\quad \geq \frac{c_{4.21} 2^{-k-1} \sqrt{a_{i+1} |J_\ell^k|}\, \mathcal{B}_i(k,\ell)(\omega)}{\|\mathcal{B}(k,\ell)(\omega)\|_2}.
\end{aligned}$$

Further,

$$\frac{c_{4.21} 2^{-k-1} \sqrt{|J_\ell^k|}}{\|\mathcal{B}(k,\ell)(\omega)\|_2} \geq \frac{c_{4.21} 2^{-k-1} \sqrt{c' n 2^{(-k-\ell)/8}} \exp(C_h{}^2 2^{(k+\ell)/2}/64)}{\sqrt{N}} \geq \frac{1}{\alpha_{k,\ell}}.$$

Using the definition of $\bar{v}_{k,\ell}$ in terms of $\bar{v}_{k,\ell-1}$ and $\bar{\Delta}_{k,\ell}$ and the above estimates, we get

$$\begin{aligned}
&\Big\langle \bar{v}_{k,\ell}(\omega), \frac{\mathrm{BM}_n(t)(\omega) - \mathrm{BM}_n(a_i)(\omega)}{\sqrt{a_i}} \Big\rangle_n \\
&\geq \frac{\alpha_{k,\ell}}{\sqrt{1+\alpha_{k,\ell}{}^2}} \Big\langle \bar{\Delta}_{k,\ell}(\omega), \frac{\mathrm{BM}_n(t)(\omega) - \mathrm{BM}_n(a_i)(\omega)}{\sqrt{a_i}} \Big\rangle_n - \frac{h(k,\ell) + \mathcal{B}_i(k,\ell)(\omega)}{\sqrt{1+\alpha_{k,\ell}{}^2}} \\
&\geq \frac{-h(k,\ell)}{\sqrt{1+\alpha_{k,\ell}{}^2}} \\
&\geq -h(k,\ell+1), \quad t \in I_k^i, \quad i \in I,
\end{aligned}$$

and, similarly,

$$\Big\langle \bar{v}_{k,\ell}(\omega), \frac{\mathrm{BM}_n(a_{i+1})(\omega) - \mathrm{BM}_n(a_i)(\omega)}{\sqrt{a_{i+1}}} \Big\rangle_n \geq \frac{f(k,\ell)}{\sqrt{1+\alpha_{k,\ell}{}^2}} \geq f(k,\ell+1), \quad i \in I.$$

Thus, by the definition of the event $\mathcal{E}_{k,\ell}$, we get $\omega \in \mathcal{E}_{k,\ell}$.

The above argument shows that

$$\mathbb{P}(\mathcal{E}_{k,\ell}) \geq \sum_I \mathbb{P}(\tilde{\Gamma}_I \cap \mathcal{E}),$$

where the sum is taken over all I with $|I| \leq N \exp(-C_h{}^2 2^{(k+\ell)/2}/32)$. Finally,

$$\sum_I \mathbb{P}(\tilde{\Gamma}_I \cap \mathcal{E}) \geq \sum_I \mathbb{P}(\Gamma_I \cap \mathcal{E}) - \sum_I \mathbb{P}(\Gamma_I \setminus \tilde{\Gamma}_I) \geq \mathbb{P}(\mathcal{E}) - \exp(-\sqrt{N}),$$

and we get the result. □

Proof of Lemma 4.24. As before, we set $\bar{v}_{k,0} := \bar{v}_{k-1}$. Consecutively applying Lemma 4.27

and formula (4.33) M' times, we obtain a random unit vector $\bar{v}_{k,M'}$ satisfying (4.31) and (4.32). Moreover, the same lemma provides the estimate $\mathbb{P}(\mathcal{E}_{k,M}) \geq \mathbb{P}(\mathcal{E}_{k-1}) - 3M' \exp(-\sqrt{N})$. Then, in view of Lemma 4.26 and the definition of M', we have

$$\mathbb{P}\big\{\mathcal{I}(k, M'+1) \neq \emptyset\big\} \leq N \exp(-{C_h}^2/\alpha_{k,M'}) + 1 - \mathbb{P}(\mathcal{E}_{k,M'}) \leq \frac{1}{n^2} + 1 - \mathbb{P}(\mathcal{E}_{k-1}).$$

Combining the above estimate with the definition of $\mathcal{E}_k$, we get for $\bar{v}_k := \bar{v}_{k,M'}$ that

$$\mathbb{P}(\mathcal{E}_k) \geq \mathbb{P}(\mathcal{E}_{k-1}) - \frac{1}{n^2}.$$

□

Chapter 5

Conclusion

In this thesis, we have considered three directions. In Chapter 2, we have studied almost Euclidean sections of convex bodies and (in Section 2.4) estimated the distance from a convex polytope with few vertices to the Euclidean ball. In Chapter 3, we have obtained bounds (both non-asymptotic and limiting) for the smallest singular value of a random matrix with i.i.d. heavy-tailed entries. In Chapter 4, we have estimated the number of steps by a high-dimensional random walk needed to absorb the origin into its convex hull.

All the results are connected via the techniques used in the proofs, which originated within Asymptotic Geometric Analysis. Namely, we have made an extensive use of covering arguments and various concentration inequalities for random projections. Thus, the thesis can be seen as a manifestation of the strength and universality of those methods.

Let us conclude the thesis with some open problems.

Dependence on ε in Dvoretzky's theorem

Question 1. *Given $m \geq 2$ and $\varepsilon \in (0,1]$, what is the least number $n = n(m,\varepsilon)$ such that any n-dimensional normed space contains a $(1+\varepsilon)$-Euclidean m-dimensional subspace?*

As we have already mentioned, the answer is known for $m = 2$ and is due to M. Gromov (see [75]). A result of J. Bourgain and J. Lindenstrauss [14], as well as the theorem of Section 2.2 (see also [35]), provide (essentially) optimal dependence of n on m and ε within the class of 1-symmetric normed spaces. In the general setting, the best known result up to now is due to G. Schechtman [100], however, optimality of that estimate is unclear.

Apart from the problem of existence of large almost Euclidean sections (subspaces), one may be interested in studying dependence on ε in various randomized constructions (we proved results in this direction in Sections 2.2 and 2.3). For a n-dimensional origin-symmetric convex body in John's position the random m-dimensional section uniformly distributed on the Grassmannian is $(1+\varepsilon)$-Euclidean with a high probability provided that $m \leq c\varepsilon^2 \ln n$ for a universal constant $c > 0$ [76, 43, 97]. In this connection, one may ask the following general question:

Question 2. *Let K be an origin-symmetric convex body in some natural position (other than John's) and E be an m-dimensional subspace uniformly distributed on the Grassmannian $\mathrm{G}_{n,m}$. Under what conditions on m and ε the section $E \cap K$ is $(1+\varepsilon)$-Euclidean with probability close to one?*

Let us emphasize that we are interested in a *natural* position, for example, when the unit Euclidean ball is an extremal (in some sense) ellipsoid for our convex body (see [114, Chapter 3], [86, Chapter 3] and [37] for an extensive information on extremal ellipsoids).

Optimal coverings of the sphere by equal spherical caps

Let us mention a problem which was briefly discussed in the introduction to Section 2.4.

Question 3. *For $n \geq 2$ and $n+1 \leq N \leq 2n$, let N equal spherical caps of minimal radius cover S^{n-1}. Is it true that in this case the centers of the caps are vertices of pairwise orthogonal $\lceil \frac{n}{N-n} \rceil$ and $\lfloor \frac{n}{N-n} \rfloor$-dimensional regular simplices of circumradius one, with the total number of the simplices equal to $N-n$?*

As we already mentioned in Section 2.4, this question is directly connected with the problem of estimating the minimal Banach–Mazur distance of a convex polytope with N vertices to the Euclidean ball. The above question is stated in [13] (see also [12, Section 6.6]). The answer is known to be positive in the case $N = n+2$ [13] as well as for $n = 2, 3$ and for $n = N/2 = 4$ (see the bibliography in §2.4.1 of this thesis).

The largest singular value of a random matrix

In Chapter 3, we proved two theorems concerning the behaviour of the smallest singular value of a rectangular random matrix with i.i.d. entries. Let us mention here two open problems concerning the *largest* singular value.

Question 4. *Let A be an $n \times n$ random matrix with independent (but not identical) centered Gaussian entries. Is it true that the expectation of the largest singular value of A can be estimated as*

$$\mathbb{E}\, s_{\max}(A) \leq C\mathbb{E} \max_i \|\mathrm{Row}_i\|_2 + C\mathbb{E} \max_j \|\mathrm{Column}_j\|_2,$$

where Row_i and Column_j are i-th row and j-th column of A, respectively, and $C > 0$ is a universal constant?

Partial results in the direction of solving the above question are due to R. Latala [62], S. Riemer and C. Schütt [88], A. S. Bandeira and R. v. Handel [10].

Question 5. *Let $\varepsilon, B, L > 0$; $N \geq n$ and let A be an $N \times n$ random matrix with i.i.d. isotropic rows satisfying*

$$\sup_{y \in S^{n-1}} \mathbb{E} \left| \langle y, \mathrm{Row}_i \rangle_n \right|^{2+\varepsilon} \leq B$$

and $\|\mathrm{Row}_i\|_2 \leq L\sqrt{n}$ a.s. Is it true that with probability close to one we have $s_{\max}(A) \leq (1+\delta)\sqrt{N}$, where δ is a function of ε, B, L and the ratio N/n, such that $\delta \to 0$ with $\frac{N}{n} \to \infty$ for any fixed triple ε, B, L?

The last question is directly connected to the problem of approximating the covariance matrix of a multidimensional distribution by the sample covariance matrix. As of now, the most general result in this direction yields a positive answer for any $\varepsilon > 2$ [47, 48]. At the same time, the case $0 < \varepsilon \leq 2$ is open. We refer to the introductions to Sections 3.1 and 3.2 of the thesis, as well as paper [47], for more information and further references.

Bibliography

[1] R. Adamczak, A.E. Litvak, A. Pajor, N. Tomczak-Jaegermann, Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles, J. Amer. Math. Soc. **23** (2010), no. 2, 535–561. MR2601042

[2] R. Adamczak, A.E. Litvak, A. Pajor, N. Tomczak-Jaegermann, Sharp bounds on the rate of convergence of the empirical covariance matrix, C. R. Math. Acad. Sci. Paris **349** (2011), no. 3-4, 195–200. MR2769907

[3] N. Alon and V. D. Milman, Embedding of l_∞^k in finite-dimensional Banach spaces, Israel J. Math. **45** (1983), no. 4, 265–280. MR0720303

[4] S. Artstein-Avidan, O. Friedland and V. Milman, Geometric applications of Chernoff-type estimates, in *Geometric aspects of functional analysis*, 45–75, Lecture Notes in Math., 1910, Springer, Berlin. MR2347039

[5] S. Artstein-Avidan, A. Giannopoulos and V. D. Milman, *Asymptotic geometric analysis. Part I*, Mathematical Surveys and Monographs, 202, Amer. Math. Soc., Providence, RI, 2015. MR3331351

[6] Z. Bai and J. W. Silverstein, *Spectral analysis of large dimensional random matrices*, second edition, Springer Series in Statistics, Springer, New York, 2010. MR2567175

[7] Z. D. Bai, J. W. Silverstein and Y. Q. Yin, A note on the largest eigenvalue of a large-dimensional sample covariance matrix, J. Multivariate Anal. **26** (1988), no. 2, 166–168. MR0963829

[8] Z. D. Bai and Y. Q. Yin, Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix, Ann. Probab. **21** (1993), no. 3, 1275–1294. MR1235416

[9] K. Ball, Ellipsoids of maximal volume in convex bodies, Geom. Dedicata **41** (1992), no. 2, 241–250. MR1153987

[10] A. S. Bandeira and R. v. Handel, Sharp nonasymptotic bounds on the norm of random matrices with independent entries, Annals of Probability, to appear, 2015.

[11] I. Bárány and Z. Füredi, Approximation of the sphere by polytopes having few vertices, Proc. Amer. Math. Soc. **102** (1988), no. 3, 651–659. MR0928998

[12] K. Böröczky, Jr., *Finite packing and covering*, Cambridge Tracts in Mathematics, 154, Cambridge Univ. Press, Cambridge, 2004. MR2078625

[13] K. Böröczky, Jr. and G. Wintsche, Covering the sphere by equal spherical balls, in *Discrete and computational geometry*, 235–251, Algorithms Combin., 25, Springer, Berlin. MR2038476

[14] J. Bourgain and J. Lindenstrauss, Almost Euclidean sections in spaces with a symmetric basis, in *Geometric aspects of functional analysis (1987–88)*, 278–288, Lecture Notes in Math., 1376, Springer, Berlin. MR1008730

[15] J. Bourgain, J. Lindenstrauss and V. Milman, Approximation of zonoids by zonotopes, Acta Math. **162** (1989), no. 1-2, 73–141. MR0981200

[16] J. Bourgain, V. H. Vu and P. M. Wood, On the singularity probability of discrete random matrices, J. Funct. Anal. **258** (2010), no. 2, 559–603. MR2557947

[17] E. J. Candès, Mathematics of sparsity (and a few other things), *Proceedings of the International Congress of Mathematicians*, Seoul, South Korea, 2014.

[18] B. Carl and A. Pajor, Gel'fand numbers of operators with values in a Hilbert space, Invent. Math. **94** (1988), no. 3, 479–504. MR0969241

[19] D. Chafaï, O. Guédon, G. Lecué and A. Pajor., *Interactions between compressed sensing random matrices and high dimensional geometry*, Panoramas et Synthèses, 37, Soc. Math. France, Paris, 2012. MR3113826

[20] D. Chafaï and K. Tikhomirov, On the convergence of the extremal eigenvalues of empirical covariance matrices with dependence. Preprint. arXiv:1509.02231

[21] V. Chandrasekaran, B. Recht, P. A. Parrilo and A. S. Willsky, The convex geometry of linear inverse problems, Found. Comput. Math. **12** (2012), no. 6, 805–849. MR2989474

[22] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, Ann. Math. Statistics **23** (1952), 493–507. MR0057518

[23] H. S. M. Coxeter, *Regular polytopes*, third edition, Dover, New York, 1973. MR0370327

[24] L. Dalla, D. G. Larman, P. Mani-Levitska and C. Zong, The blocking numbers of convex bodies, Discrete Comput. Geom. **24** (2000), no. 2-3, 267–277. MR1758049

[25] H. A. David, *Order statistics*, second edition, Wiley, New York, 1981. MR0597893

[26] A. Dvoretzky, Some results on convex bodies and Banach spaces, in *Proc. Internat. Sympos. Linear Spaces (Jerusalem, 1960)*, 123–160, Jerusalem Academic Press, Jerusalem. MR0139079

[27] A. Edelman, Eigenvalues and condition numbers of random matrices, SIAM J. Matrix Anal. Appl. **9** (1988), no. 4, 543–560. MR0964668

[28] A. Edelman and N. R. Rao, Random matrix theory, Acta Numer. **14** (2005), 233–297. MR2168344

[29] R. Eldan, Extremal points of high-dimensional random walks and mixing times of a Brownian motion on the sphere, Ann. Inst. Henri Poincaré Probab. Stat. **50** (2014), no. 1, 95–110. MR3161524

[30] R. Eldan, Volumetric properties of the convex hull of an n-dimensional Brownian motion, Electron. J. Probab. **19** (2014), no. 45, 34 pp. MR3210546

[31] L. Fejes Tóth, *Regular figures*, A Pergamon Press Book, Macmillan, New York, 1964. MR0165423

[32] W. Feller, *An introduction to probability theory and its applications. Vol. I*, Third edition, Wiley, New York, 1968. MR0228020

[33] T. Figiel and W. B. Johnson, Large subspaces of l_∞^n and estimates of the Gordon-Lewis constant, Israel J. Math. **37** (1980), no. 1-2, 92–112. MR0599305

[34] B. Fleury, Concentration in a thin Euclidean shell for log-concave measures, J. Funct. Anal. **259** (2010), no. 4, 832–841. MR2652173

[35] D. J. Fresen, Explicit Euclidean embeddings in permutation invariant normed spaces, Adv. Math. **266** (2014), 1–16. MR3262353

[36] R. J. Gardner, The Brunn-Minkowski inequality, Bull. Amer. Math. Soc. (N.S.) **39** (2002), no. 3, 355–405. MR1898210

[37] A. A. Giannopoulos and V. D. Milman, Euclidean structure in finite dimensional normed spaces, in *Handbook of the geometry of Banach spaces, Vol. I*, 707–779, North-Holland, Amsterdam. MR1863705

[38] E. D. Gluskin, Extremal properties of orthogonal parallelepipeds and their applications to the geometry of Banach spaces. (Russian) Mat. Sb. (N.S.) **136(178)** (1988), no. 1, 85–96; translation in Math. USSR-Sb. **64** (1989), no. 1, 85–96. MR0945901

[39] E. D. Gluskin, The diameter of the Minkowski compactum is roughly equal to n, Funktsional. Anal. i Prilozhen. **15** (1981), no. 1, 72–73. MR0609798

[40] E. D. Gluskin and A. E. Litvak, On the asymmetry constant of a body with few vertices, Geom. Dedicata **90** (2002), 45–48. MR1898150

[41] F. Götze and A. Tikhomirov, The circular law for random matrices, Ann. Probab. **38** (2010), no. 4, 1444–1491. MR2663633

[42] Y. Gordon, On Milman's inequality and random subspaces which escape through a mesh in $\mathbf{R}^n$, in *Geometric aspects of functional analysis (1986/87)*, 84–106, Lecture Notes in Math., 1317, Springer, Berlin. MR0950977

[43] Y. Gordon, Some inequalities for Gaussian processes and applications, Israel J. Math. **50** (1985), no. 4, 265–289. MR0800188

[44] Y. Gordon, A.E. Litvak, A. Pajor and N. Tomczak–Jaegermann, Random ϵ-nets and embeddings in l_∞^N, Studia Math. **178** (2007), no. 1, 91–98. MR2282492

[45] P. M. Gruber, The space of convex bodies, in *Handbook of convex geometry, Vol. A, B*, 301–318, North-Holland, Amsterdam. MR1242983

[46] B. Grünbaum, Measures of symmetry for convex sets, in *Proc. Sympos. Pure Math., Vol. VII*, 233–270, Amer. Math. Soc., Providence, RI. MR0156259

[47] O. Guédon, A. E. Litvak, A. Pajor, N. Tomczak-Jaegermann, On the interval of fluctuation of the singular values of random matrices, Journal of the European Mathematical Society, to appear.

[48] O. Guédon, A.E. Litvak, A. Pajor, N. Tomczak-Jaegermann, Restricted isometry property for random matrices with heavy-tailed columns, C. R. Math. Acad. Sci. Paris **352** (2014), no. 5, 431–434. MR3194251

[49] O. Guédon and E. Milman, Interpolating thin-shell and sharp large-deviation estimates for isotropic log-concave measures, Geom. Funct. Anal. **21** (2011), no. 5, 1043–1068. MR2846382

[50] W. Hoeffding, Probability inequalities for sums of bounded random variables, J. Amer. Statist. Assoc. **58** (1963), 13–30. MR0144363

[51] A. T. James, Distributions of matrix variates and latent roots derived from normal samples, Ann. Math. Statist. **35** (1964), 475–501. MR0181057

[52] F. John, Extremum problems with inequalities as subsidiary conditions, in *Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948*, 187–204, Interscience Publishers, Inc., New York, NY. MR0030135

[53] Z. Kabluchko, V. Vysotsky, D. Zaporozhets, Convex hulls of random walks, hyperplane arrangements, and Weyl chambers, Preprint, arXiv:1510.04073

[54] Z. Kabluchko, D. Zaporozhets, Intrinsic volumes of Sobolev balls with applications to Brownian convex hulls, Preprint, arXiv:1404.6113

[55] J. Kahn, J. Komlós and E. Szemerédi, On the probability that a random ± 1-matrix is singular, J. Amer. Math. Soc. **8** (1995), no. 1, 223–240. MR1260107

[56] R. Kannan, L. Lovász and M. Simonovits, Random walks and an $O^*(n^5)$ volume algorithm for convex bodies, Random Structures Algorithms **11** (1997), no. 1, 1–50. MR1608200

[57] B. Klartag, A central limit theorem for convex sets, Invent. Math. **168** (2007), no. 1, 91–131. MR2285748

[58] B. Klartag, Power-law estimates for the central limit theorem for convex sets, J. Funct. Anal. **245** (2007), no. 1, 284–310. MR2311626

[59] V. Koltchinskii and S. Mendelson, Bounding the Smallest Singular Value of a Random Matrix Without Concentration, Int. Math. Res. Not. IMRN **2015**, no. 23, 12991–13008. MR3431642

[60] J. Komlós, On the determinant of (0, 1) matrices, Studia Sci. Math. Hungar **2** (1967), 7–21. MR0221962

[61] D. G. Larman and P. Mani, Almost ellipsoidal sections and projections of convex bodies, Math. Proc. Cambridge Philos. Soc. **77** (1975), 529–546. MR0377697

[62] R. Latała, Some estimates of norms of random matrices, Proc. Amer. Math. Soc. **133** (2005), no. 5, 1273–1282 (electronic). MR2111932

[63] M. Ledoux and M. Talagrand, *Probability in Banach spaces*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3), 23, Springer, Berlin, 1991. MR1102015

[64] Lévy, P. Théorie de l'addition des variables aléatoires, Second edition, 1954.

[65] J. Lindenstrauss and L. Tzafriri, *Classical Banach spaces. II*, Ergebnisse der Mathematik und ihrer Grenzgebiete, 97, Springer, Berlin, 1979. MR0540367

[66] A. E. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, Smallest singular value of random matrices and geometry of random polytopes, Adv. Math. **195** (2005), no. 2, 491–523. MR2146352

[67] A. E. Litvak, A. Pajor and N. Tomczak-Jaegermann, Diameters of sections and coverings of convex bodies, J. Funct. Anal. **231** (2006), no. 2, 438–457. MR2195339

[68] A. E. Litvak and S. Spektor, Quantitative version of a Silverstein's result, in *Geometric aspects of functional analysis*, 335–340, Lecture Notes in Math., 2116, Springer, Cham. MR3364695

[69] S. N. Majumdar, A. Comtet and J. Randon-Furling, Random convex hulls and extreme value statistics, J. Stat. Phys. **138** (2010), no. 6, 955–1009. MR2601420

[70] P. Mankiewicz and N. Tomczak-Jaegermann, Quotients of finite-dimensional Banach spaces; random phenomena, in *Handbook of the geometry of Banach spaces, Vol. 2*, 1201–1246, North-Holland, Amsterdam. MR1999195

[71] V. A. Marčenko and L. A. Pastur, Distribution of eigenvalues in certain sets of random matrices, Mat. Sb. (N.S.) **72 (114)** (1967), 507–536. MR0208649

[72] P. Matthews, Covering problems for Brownian motion on spheres, Ann. Probab. **16** (1988), no. 1, 189–199. MR0920264

[73] S. Mendelson, A remark on the diameter of random sections of convex bodies, in *Geometric aspects of functional analysis*, 395–404, Lecture Notes in Math., 2116, Springer, Cham. MR3364699

[74] S. Mendelson and G. Paouris, On the singular values of random matrices, J. Eur. Math. Soc. (JEMS) **16** (2014), no. 4, 823–834. MR3191978

[75] V. D. Milman, A few observations on the connections between local theory and some other fields, in *Geometric aspects of functional analysis (1986/87)*, 283–289, Lecture Notes in Math., 1317, Springer, Berlin. MR0950988

[76] V. D. Milman, A new proof of A. Dvoretzky's theorem on cross-sections of convex bodies, Funkcional. Anal. i Priložen. **5** (1971), no. 4, 28–37. MR0293374

[77] V. D. Milman, Random subspaces of proportional dimension of finite-dimensional normed spaces: approach through the isoperimetric inequality, in *Banach spaces (Columbia, Mo., 1984)*, 106–115, Lecture Notes in Math., 1166, Springer, Berlin. MR0827766

[78] V. Milman, Randomness and pattern in convex geometric analysis, Doc. Math. **1998**, Extra Vol. II, 665–677. MR1648115

[79] V. D. Milman and A. Pajor, Regularization of star bodies by random hyperplane cut off, Studia Math. **159** (2003), no. 2, 247–261. MR2052221

[80] V. D. Milman and G. Schechtman, *Asymptotic theory of finite-dimensional normed spaces*, Lecture Notes in Mathematics, 1200, Springer, Berlin, 1986. MR0856576

[81] P. Mörters and Y. Peres, *Brownian motion*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge Univ. Press, Cambridge, 2010. MR2604525

[82] Oliveira, R.I. The lower tail of random quadratic forms, with applications to ordinary least squares and restricted eigenvalue properties. arXiv:1312.2903.

[83] A. Pajor and N. Tomczak-Jaegermann, Subspaces of small codimension of finite-dimensional Banach spaces, Proc. Amer. Math. Soc. **97** (1986), no. 4, 637–642. MR0845980

[84] G. Paouris, Concentration of mass on convex bodies, Geom. Funct. Anal. **16** (2006), no. 5, 1021–1049. MR2276533

[85] L. Pastur and M. Shcherbina, *Eigenvalue distribution of large random matrices*, Mathematical Surveys and Monographs, 171, Amer. Math. Soc., Providence, RI, 2011. MR2808038

[86] G. Pisier, *The volume of convex bodies and Banach space geometry*, Cambridge Tracts in Mathematics, 94, Cambridge Univ. Press, Cambridge, 1989. MR1036275

[87] E. Rebrova and K. Tikhomirov, Coverings of random ellipsoids, and invertibility of matrices with i.i.d. heavy-tailed entries. Preprint. arXiv:1508.06690

[88] S. Riemer and C. Schütt, On the expectation of the norm of random matrices with non-identically distributed entries, Electron. J. Probab. **18** (2013), no. 29, 13 pp. MR3035757

[89] P. H. Roberts and H. D. Ursell, Random walk on a sphere and on a Riemannian manifold, Philos. Trans. Roy. Soc. London. Ser. A **252** (1960), 317–356. MR0117795

[90] B. A. Rogozin, On the increase of dispersion of sums of independent random variables, Teor. Verojatnost. i Primenen **6** (1961), 106–108. MR0131894

[91] M. Rudelson and R. Vershynin, Non-asymptotic theory of random matrices: extreme singular values, in *Proceedings of the International Congress of Mathematicians. Volume III*, 1576–1602, Hindustan Book Agency, New Delhi. MR2827856

[92] M. Rudelson and R. Vershynin, Small Ball Probabilities for Linear Images of High-Dimensional Distributions, Int. Math. Res. Not. IMRN **2015**, no. 19, 9594–9617. MR3431603

[93] M. Rudelson and R. Vershynin, Smallest singular value of a random rectangular matrix, Comm. Pure Appl. Math. **62** (2009), no. 12, 1707–1739. MR2569075

[94] M. Rudelson and R. Vershynin, The least singular value of a random square matrix is $O(n^{-1/2})$, C. R. Math. Acad. Sci. Paris **346** (2008), no. 15-16, 893–896. MR2441928

[95] M. Rudelson and R. Vershynin, The Littlewood-Offord problem and invertibility of random matrices, Adv. Math. **218** (2008), no. 2, 600–633. MR2407948

[96] A. Sankar, D. A. Spielman and S.-H. Teng, Smoothed analysis of the condition numbers and growth factors of matrices, SIAM J. Matrix Anal. Appl. **28** (2006), no. 2, 446–476 (electronic). MR2255338

[97] G. Schechtman, A remark concerning the dependence on ε in Dvoretzky's theorem, in *Geometric aspects of functional analysis (1987–88)*, 274–277, Lecture Notes in Math., 1376, Springer, Berlin. MR1008729

[98] G. Schechtman, Euclidean sections of convex bodies, in *Asymptotic geometric analysis*, 271–288, Fields Inst. Commun., 68, Springer, New York. MR3076155

[99] G. Schechtman, The random version of Dvoretzky's theorem in ℓ_∞^n, in *Geometric aspects of functional analysis*, 265–270, Lecture Notes in Math., 1910, Springer, Berlin. MR2349612

[100] G. Schechtman, Two observations regarding embedding subsets of Euclidean spaces in normed spaces, Adv. Math. **200** (2006), no. 1, 125–135. MR2199631

[101] K. Schütte, Überdeckungen der Kugel mit höchstens acht Kreisen, Math. Ann. **129** (1955), 181–186. MR0069528

[102] G. R. Shorack and J. A. Wellner, *Empirical processes with applications to statistics*, reprint of the 1986 original [MR0838963], Classics in Applied Mathematics, 59, SIAM, Philadelphia, PA, 2009. MR3396731

[103] J. W. Silverstein, On the weak limit of the largest eigenvalue of a large-dimensional sample covariance matrix, J. Multivariate Anal. **30** (1989), no. 2, 307–311. MR1015375

[104] D. Slepian, The one-sided barrier problem for Gaussian noise, Bell System Tech. J. **41** (1962), 463–501. MR0133183

[105] D. A. Spielman and S.-H. Teng, Smoothed analysis of algorithms, in *Proceedings of the International Congress of Mathematicians, Vol. I (Beijing, 2002)*, 597–606, Higher Ed. Press, Beijing. MR1989210

[106] N. Srivastava and R. Vershynin, Covariance estimation for distributions with $2+\varepsilon$ moments, Ann. Probab. **41** (2013), no. 5, 3081–3111. MR3127875

[107] S. J. Szarek, Condition numbers of random matrices, J. Complexity **7** (1991), no. 2, 131–149. MR1108773

[108] T. Tao and V. H. Vu, Inverse Littlewood-Offord theorems and the condition number of random discrete matrices, Ann. of Math. (2) **169** (2009), no. 2, 595–632. MR2480613

[109] T. Tao and V. Vu, On the singularity probability of random Bernoulli matrices, J. Amer. Math. Soc. **20** (2007), no. 3, 603–628. MR2291914

[110] T. Tao and V. Vu, Random matrices: the circular law, Commun. Contemp. Math. **10** (2008), no. 2, 261–307. MR2409368

[111] T. Tao and V. Vu, Random matrices: the distribution of the smallest singular values, Geom. Funct. Anal. **20** (2010), no. 1, 260–297. MR2647142

[112] T. Tao and V. Vu, Smooth analysis of the condition number and the least singular value, Math. Comp. **79** (2010), no. 272, 2333–2352. MR2684367

[113] T. Tao and V. Vu, The condition number of a randomly perturbed matrix, in *STOC'07—Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, 248–255, ACM, New York. MR2402448

[114] N. Tomczak-Jaegermann, *Banach-Mazur distances and finite-dimensional operator ideals*, Pitman Monographs and Surveys in Pure and Applied Mathematics, 38, Longman Sci. Tech., Harlow, 1989. MR0993774

[115] N. Tomczak-Jaegermann, From finite- to infinite-dimensional phenomena in geometric functional analysis on local and asymptotic levels, Doc. Math. **1998**, Extra Vol. II, 731–742 (electronic). MR1648121

[116] R. Vershynin, Estimation in high dimensions: a geometric perspective. Available at arXiv:1405.5103

[117] R. Vershynin, Introduction to the non-asymptotic analysis of random matrices, in *Compressed sensing*, 210–268, Cambridge Univ. Press, Cambridge. MR2963170

[118] P. Yaskov, Lower bounds on the smallest eigenvalue of a sample covariance matrix, Electron. Commun. Probab. **19** (2014), no. 83, 10 pp. MR3291620

[119] Y. Q. Yin, Limiting spectral distribution for a class of random matrices, J. Multivariate Anal. **20** (1986), no. 1, 50–68. MR0862241

[120] Y. Q. Yin, Z. D. Bai and P. R. Krishnaiah, On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix, Probab. Theory Related Fields **78** (1988), no. 4, 509–521. MR0950344

[121] Fourier analytic methods in convex geometry. Workshop in Palo Alto, California, August 20–24, 2007; http://aimath.org/pastworkshops/fourierconvex.html.