I am always doing that which I can not do, in order that I may learn how to do it.

Pablo Picasso

University of Alberta

Smallest singular value of sparse random matrices

by

Omar Daniel Rivasplata

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Mathematics

Department of Mathematical and Statistical Sciences

Spring 2012
Edmonton, Alberta

To my parents

Abstract

In this thesis probability estimates on the smallest singular value of random matrices with independent entries are extended to a class of sparse random matrices. We show that one can relax a previously used condition of uniform boundedness of the variances from below. This allows us to consider matrices with null entries or, more generally, with entries having small variances. Our results do not assume identical distribution of the entries of a random matrix, and help to clarify the role of the variances in the corresponding estimates. We also show that it is enough to require boundedness from above of the r-th moment of the entries, for some $r > 2$.

Acknowledgements

First of all, I sincerely thank my supervisors, Dr. Nicole Tomczak-Jaegermann and Dr. Alexander Litvak, from whom I have received much guidance and support. Working with them I've learned a lot of mathematics and of the mathematician's "tricks of the trade", and thanks to them my experience of the doctoral program has been very enriching. They also made available to me plenty of opportunities for going to conferences and visiting other universities, for which I am very grateful.

I thank the members of my supervisory committee, Dr. Vladimir Troitsky and Dr. Vlad Yaskin. I have been lucky to have them around: I learned a lot in the graduate courses I took with them and I benefited a great deal from discussions with them and from their valuable feedback throughout my program.

For their teachings, supportive friendship, and most valuable advice, I thank the following professors of our Department of Mathematical and Statistical Sciences: Dr. Abel Cadenillas, Dr. Eric Woolgar, Dr. Dragos Hrimiuc, Dr. Byron Schmuland, and Dr. Alexander Melnikov.

This is a neat opportunity to thank a few special professors of my former school, the Pontificia Universidad Católica del Perú: Dr. Cesar Carranza, Dr. Maynard Kong, Dr. Christian Figueroa and Dr. Alfredo Poirier. I started learning mathematics with them as an undergraduate student and one may say with all justice that their teachings have a bearing on the outcome of my graduate program.

I also thank the support staff of our Department of Mathematical and Statistical Sciences for the countless times they have assisted me with various matters such as paperwork, scheduling, supplies, information, among others.

During the time I've been a graduate student here I have had the pleasure of meeting –and befriending– a number of students and postdoctoral fellows. A special note of appreciation, in order of appearance, goes to Jan Rychtar, Peter Pivovarov, Ben Willson, Steven Taschuk, Radek Adamczak, Marton Naszodi, Susanna Spektor, Nikos Dafnis, and David Alonso Gutierrez. It has been great having them around, and so has been learning mathematics alongside them, discussing mathematics with them, or plainly chatting about random stuff with them.

Last, though far from least, I thank my family: my children Gabriel and Sofia, and their mom Mariel. I need no special reason to thank them, which is to say, I am grateful all the time for having them! But they deserve a special place in my acknowledgements for their patience during the many hours that I was away from them while working on my graduate program.

Table of Contents

Introduction

The contents of this thesis pertain the results presented in [29]. Therefore, naturally, there is much overlap between the text written here and that of [29].

Let $N \geq n$ be positive integers. In this thesis we study the smallest singular value of matrices $\Gamma = (\xi_{ji})$, of size $N \times n$, whose entries are real-valued random variables obeying certain probability laws. In particular we are interested in allowing these matrices to contain some null entries (or, more generally, to contain some entries with arbitrarily small variances). In this sense we deal with *sparse* (or *dilute*) random matrices. Sparse random matrices and more generally sparse structures play an important role in many branches of pure and applied mathematics. We refer to Chapter 7 of [5] for definitions, relevant discussions, and references (see also the recent works [18], [47]).

Understanding the behaviour of random matrices is of importance in several fields, including Asymptotic Geometric Analysis, Approximation Theory, Probability and Statistics. The results of classical random matrix theory focused on the limiting properties of random matrices as the dimension grows to infinity. In that context, the limiting behaviour of the extreme singular values (see the definitions in Chapter 1) of random matrices was studied. Such limiting behaviour is now well understood for the case of matrices whose entries are independent identically distributed (*i.i.d.*) random variables. We refer to the following books, surveys, and recent papers for history, results, and open problems in this direction [4], [5], [15], [16], [32], [47], [49].

In the *asymptotic non-limiting* case very little was known till very recently. In this case one studies the rate of convergence, deviation inequalities, and the general asymptotic behaviour of singular values as functions of the dimensions of a matrix. Thus, by 'non-limiting' we mean that the dimensions n, N are fixed, while the name 'asymptotic' is meant to imply that n, N are large (growing to infinity). The case when the entries of a random matrix are independent $\mathcal{N}(0, 1)$ Gaussian was treated in [12] and in [45] (see also [17] for related results, and the survey [10]). In the last decade the attention shifted to other models, like matrices with independent sub-Gaussian entries (the case of symmetric Bernoulli ± 1 appears to be particularly important), independent entries with some moment conditions, as well as matrices with independent columns or rows satisfying some natural restrictions. Major achievements were obtained in [1], [2], [26], [38], [39], [40], [46], [48].

In all previous non-limiting asymptotic results for random matrices with independent entries, an important assumption was that the variances of all the entries are bounded below by one, i.e. in a sense, that all entries are buffered away from zero and thus cannot be too small. Such a condition is not natural for some applications, for instance when dealing with models in the theory of wireless communications, where signals may be lost (or some small noise may appear), or with models in neural network theory, where the neurons are not of full connectivity with each other.

The main goal of our research is to show that one can significantly relax the requirement of boundedness from below of all entries, replacing it by averaging type conditions. Thus our results clarify the role of the variances in the corresponding previous results. In the limiting case of symmetric square matrices (universal Wigner ensemble), many results in this direction were obtained, see survey [13]. Another advantage of our results is that we require only boundedness (from above) of the r-th moments for an arbitrary (fixed) $r > 2$. We would like to emphasize that we don't require identical distributions of all entries of a random matrix nor boundedness of the sub-Gaussian moment of entries (both conditions were crucial for the deep results of [40]). Moreover, the condition on entries "to be identically distributed" is clearly inconsistent with our model, as, under such a condition, if one entry is zero then automatically all entries are zeros.

We describe now our setting and results. Our main results present estimates for the smallest singular value $s_n(\Gamma)$ of large matrices Γ of the type described. As it turns out the methods used to establish those estimates depend on the *aspect ratio* of the matrices. The aspect ratio of an $N \times n$ matrix A is the ratio n/N of number of columns to number of rows, or, more intuitively, the ratio "width by height". To have a suggestive terminology, we will say that such matrix A is

- "*tall*" if $\frac{n}{N} \leq c_0$ for a small positive constant c_0;
- "*almost square*" if $\frac{n}{N}$ is close to 1.

Clearly, a matrix is square when its aspect ratio is equal to 1.

We will deal with random matrices under various conditions, which we list now. These conditions allow our matrices to contain many null (or small) entries, which means that we don't have any restrictions on the variance of a particular entry. Our model is different from the models used in [18], [47], where zeros appeared randomly, i.e. each entry, which is a random variable itself, was multiplied by another random variable of type 0/1. Our model is more similar to those considered in [13], where a condition similar to (iii) was used for square symmetric matrices. Naturally, in order for our random matrices to have entries of different kinds, we do not require that the entries are identically distributed. For parameters $r > 2$, $\mu \geq 1$, $a_1 > 0$, $a_2 > 0$, $a_3 \in (0, \mu)$, and $a_4 \in (0, 1]$, we will consider $N \times n$ random matrices $\Gamma = (\xi_{ji})_{j \leq N, i \leq n}$ whose entries are *independent* real-valued *centered* random variables satisfying the following conditions:

(i) Moments: $\quad \mathbb{E}\,|\xi_{ji}|^r \leq \mu^r \quad$ for all j and i.

(ii) Norm: $\quad \mathbb{P}\Big(\|\Gamma\| > a_1\sqrt{N}\Big) \leq e^{-a_2 N}$.

(iii) Columns: $\mathbb{E}\|(\xi_{ji})_{j=1}^N\|_2^2 = \sum_{j=1}^N \mathbb{E}\,\xi_{ji}^2 \geq a_3^2 N$ for each i.

For almost square and for square matrices we also will need the following condition on rows.

(iv) Rows: $|\{i \mid \mathbb{E}\,\xi_{ji}^2 \geq 1\}| \geq a_4 n$ for each j.

It is important to highlight that the parameters μ, r, a_1, a_2, a_3, a_4 should be regarded as constants which do not depend on the dimensions n, N. Note also that the ratio μ/a_3 is of particular importance (μ is responsible for the maximal L_r-norm of entries, while a_3 is an average-type substitution for the lower bound on L_2-norm of entries).

Before we state our main results let us comment on the above conditions. The first condition is a standard requirement saying that the random variables are not "too big". For the limiting case it is known that boundedness of the forth moments is needed. It turns out that for our estimates it is enough to ask boundedness of moments of order $r = 2 + \varepsilon$ only, which improves all previous results. In particular, this was one of the questions raised in [50], where the author proved corresponding estimates for entries with bounded $4 + \varepsilon$ moment, and asked about $2 + \varepsilon$ moment.

The second condition is crucial for many results on random matrices. We recall that the norm of a matrix is understood to be the operator norm from ℓ_2^n to ℓ_2^N, also called the spectral norm, which is equal to the largest singular value. In fact, the question *"What are the models of random matrices satisfying condition (ii)?"* (and more generally, *"What is the behaviour of the largest singular value?"*) is one of the central questions in random matrix theory. Such estimates are well known for the Gaussian and sub-Gaussian cases. We refer to [2], [23] and references therein for other models and recent developments on this problem. We would like to emphasize that condition (ii) is needed in order to get probabilities exponentially close to one. Alternatively, one may substitute this condition by

$$p_N := \mathbb{P}\Big(\|\Gamma\| > a_1\sqrt{N}\Big) < 1,$$

in which case one should add p_N to the estimates of probabilities in our theorems below.

The main novelty in our model are conditions (iii) and (iv). These two conditions substitute the standard condition

$$\mathbb{E}\,|\xi_{ji}|^2 \geq 1 \quad \text{for all} \quad j, i, \tag{1}$$

which was used in all previous works related to the smallest singular value of a random matrix (in the non-limiting case). Removing such strong assumption on *all* entries we allow the possibility of zeros to appear among the entries of a random matrix. Our conditions (iii) and (iv) should be compared with (1.1) and (1.16) in [13].

Of course we want to rule out matrices having a column or a row consisting of zeros only, for if there is a zero column then immediately $s_n(\Gamma) = 0$, while if there is a zero row then the matrix Γ is essentially of size $(N-1) \times n$. Hence we need some general assumptions on the columns and the rows of the matrices under consideration. Our condition (iii) alone implies that each column vector of the matrix has relatively big ℓ_2-norm. Moreover, condition (iii) together with condition (i) guarantee that proportionally many rows have ℓ_2-norms bounded away from 0. It turns out that condition (iii) is already enough for "tall" matrices, when $N > Cn$, as the first theorem below shows. The cases of "almost square" and square matrices are more delicate, because N becomes closer to n, and we need to control the behaviour of rows more carefully. Condition (iv) ensures that each row of the matrix has proportionally many entries with variance at least one.

Now we state our results. The first theorem deals with "tall" matrices and extends the corresponding result from [26] (for mean zero random variables with variances bounded below uniformly and uniformly *bounded above* this was shown in [6]). Note that we use only three conditions, (i), (ii), and (iii), while condition (iv) is not required for this result.

Theorem 1. *Let $r > 2$, $\mu \geq 1$, $a_1, a_2, a_3 > 0$ with $a_3 < \mu$. Let $1 \leq n < N$ be integers, and write N in the form $N = (1+\delta)n$. Suppose Γ is an $N \times n$ matrix whose entries are independent centered random variables such that conditions (i), (ii) and (iii) are satisfied. There exist positive constants c_1, c_2 and δ_0 (depending only on the parameters r, μ, a_1, a_2, a_3) such that whenever $\delta \geq \delta_0$, then*

$$\mathbb{P}\Big(s_n(\Gamma) \leq c_1\sqrt{N}\Big) \leq e^{-c_2 N}.$$

Remark. Our proof gives that $c_1 = c_1(r, \mu, a_3)$ depends on r, μ, and a_3 only; whereas $c_2 = c_2(r, \mu, a_2, a_3)$ and $\delta_0 = \delta_0(r, \mu, a_1, a_3)$.

Our next theorem is about "almost square" matrices. Here both conditions (iii) and (iv) are needed in order to substitute condition (1). This theorem extends the corresponding result [26, Theorem 3.1].

Theorem 2. *Let $r > 2$, $\mu \geq 1$, $a_1, a_2 > 0$, $a_3 \in (0, \mu)$, $a_4 \in (0, 1]$. Let $1 \leq n < N$ be integers, and write N in the form $N = (1+\delta)n$. Suppose Γ is an $N \times n$ matrix whose entries are independent centered random variables such that conditions (i), (ii), (iii) and (iv) are satisfied. There exist positive constants c_1, c_2, $\tilde{c}_1$ and $\tilde{c}_2$, depending only on the parameters r, μ, a_1, a_2, a_3, a_4, and a positive constant $\gamma = \gamma(r, \mu, a_1, a_3) < 1$, such that if*

$$a_4 > 1 - \gamma \qquad \text{and} \qquad \delta \geq \frac{\tilde{c}_1}{\ln(2 + \tilde{c}_2 n)}$$

then

$$\mathbb{P}\Big(s_n(\Gamma) \leq c_1\sqrt{N}\Big) \leq e^{-c_2 N}.$$

Remarks. 1. Our proof gives that $c_1 = c_1(r, \mu, a_1, a_3, \delta)$, $c_2 = c_2(r, \mu, a_2, a_3)$, $\tilde{c}_1 = \tilde{c}_1(r, \mu, a_1, a_3)$ and $\tilde{c}_2 = \tilde{c}_2(r, \mu, a_1, a_3, a_4)$.
2. Note that for small n, say for $n \le 2/\tilde{c}_2$, Theorem 2 is trivial for every $\delta > 0$, either by adjusting the constant c_2 (for small N) or by using Theorem 1 (for large N).

Let us note that in a sense our Theorems 1 and 2 are incomparable with the corresponding result of [40]. First, we don't restrict our results only to the sub-Gaussian case. The requirement of boundedness of the sub-Gaussian moment is much stronger, implying in particular boundedness of moments of all orders, which naturally yields stronger estimates. Second, another condition essentially used in [40] is "entries are identically distributed." As was mentioned above, such a condition is inconsistent with our model, since having one zero we immediately get the zero matrix.

Our third theorem shows that we can also extend to our setting the corresponding results from [39], where the i.i.d. case was treated, and from [1], which dealt with the case of independent log-concave columns. Note again that we work under the assumption of bounded r-th moment (for a fixed $r > 2$). In fact, in [39] two theorems about square matrices were proved. The first one is for random matrices whose entries have bounded fourth moment. Our Theorem 3 extends this result with much better probability. The second main result of [39] requires the boundedness of sub-Gaussian moments as well as identical distributions of entries in each column, and, thus, is incomparable with Theorem 3.

Theorem 3. *Let $r > 2$, $\mu \ge 1$, $a_1, a_2, a_3, a_4 > 0$ with $a_3 < \mu$. Suppose Γ is an $n \times n$ matrix whose entries are independent centered random variables such that conditions (i), (ii), (iii) and (iv) are satisfied. Then for every $\varepsilon \ge 0$*

$$\mathbb{P}\big(s_n(\Gamma) \le \varepsilon n^{-1/2}\big) \le C\big(\varepsilon + n^{1-r/2}\big),$$

where C depends on the parameters r, μ, a_1, a_2, a_3, a_4.

In Chapter 7 we extend Theorem 1 to the case of random matrices with complex-valued entries. Treating random matrices with real-valued entries in the first place was a natural choice, for instance, this is the case in most theorems of Probability Theory, where real random variables are studied at first, arguably with applications in mind, and later the results are extended to complex random variables.

In the chapter closing this thesis, Chapter 8, we prove that a recent result of Srivastava and Vershynin, which involves projections of random vectors with a product distribution, is valid without requiring that the components have unit variances. Our work in this chapter gives a neat example that the idea of removing conditions of boundedness from below of the variances, as we have done in our results of previous chapters, is useful also in a somewhat different context.

Chapter 1

Notation and Preliminaries

In this chapter we lay down the notations that will be used throughout. We also present here some preliminary definitions.

For $1 \le p \le \infty$, we write $\|x\|_p$ to denote the ℓ_p-norm of $x = (x_i)_{i\ge 1}$, that is, the norm defined by

$$\|x\|_p = \Big(\sum_{i\ge 1} |x_i|^p\Big)^{1/p} \quad \text{for } p < \infty \qquad \text{and} \qquad \|x\|_\infty = \sup_{i\ge 1} |x_i|.$$

Then, as usual, $\ell_p^n = (\mathbb{R}^n, \|\cdot\|_p)$. The unit ball of ℓ_p^n is denoted B_p^n. The unit sphere of ℓ_2^n is denoted S^{n-1}, and $e_1, \dots, e_n$ is the canonical basis of ℓ_2^n.

We write $\langle \cdot, \cdot \rangle$ for the standard inner product on $\mathbb{R}^n$. By $|x|$ we denote the standard Euclidean norm (i.e. ℓ_2-norm) of the vector $x = (x_i)_{i\ge 1}$. On the other hand, when A is a set, we write $|A|$ to denote the cardinality of A.

The support of a vector $x = (x_i)_{i\ge 1}$ is denoted by $\mathrm{supp}(x)$, this is the set of indices corresponding to nonzero coordinates of x, that is,

$$\mathrm{supp}(x) := \big\{ i \;\big|\; x_i \ne 0 \big\}.$$

Given a subspace E of $\mathbb{R}^n$ we write P_E for the orthogonal projection onto E. In the particular case when $E = \mathbb{R}^\sigma$ is the coordinate subspace corresponding to a set of coordinates $\sigma \subset \{1, \dots, n\}$, we will write P_σ as a shorthand for $P_{\mathbb{R}^\sigma}$.

Let $\mathcal{N} \subset D \subset \mathbb{R}^n$ and $\varepsilon > 0$. Recall that $\mathcal{N}$ is called an ε-net of D (in the Euclidean metric) if

$$D \subset \bigcup_{v\in\mathcal{N}} (v + \varepsilon B_2^n).$$

In case D is the unit sphere S^{n-1} or the unit ball B_2^n, a well known volumetric argument establishes that for each $\varepsilon > 0$ there is an ε-net $\mathcal{N}$ of D with cardinality $|\mathcal{N}| \le (1 + 2/\varepsilon)^n$, see for instance [33, Lemma 2.6].

1.1 Singular values.

Suppose Γ is an $N \times n$ matrix with real entries. The singular values of Γ, denoted $s_k(\Gamma)$, are the eigenvalues of the $n \times n$ matrix $\sqrt{\Gamma^t \Gamma}$, arranged in the decreasing order. It is immediate that the singular values are all non-negative, and furthermore the number of nonzero singular values of Γ equals the rank of Γ.

The largest singular value $s_1(\Gamma)$ and the smallest singular value $s_n(\Gamma)$ are called the extreme singular values of Γ. These are given by the expressions

$$s_1(\Gamma) = \sup\big\{|\Gamma x| \;\big|\; |x| = 1\big\}, \qquad s_n(\Gamma) = \inf\big\{|\Gamma x| \;\big|\; |x| = 1\big\}.$$

In particular $s_1(\Gamma) = \|\Gamma : \; \ell_2^n \to \ell_2^N\|$ is the operator norm of Γ. Observe that for every vector $x \in \mathbb{R}^n$ one has

$$s_n(\Gamma)|x| \le |\Gamma x| \le s_1(\Gamma)|x|. \tag{1.1}$$

The estimate on the left-hand side is trivial (but futile) if $s_n(\Gamma) = 0$. On the other hand, when $s_n(\Gamma) > 0$ the matrix Γ is a bijection on its image, and can be regarded as an embedding from ℓ_2^n into ℓ_2^N, with (1.1) providing an estimate for the distortion of the norms under Γ. In this case, $\|\Gamma^{-1} : \; \Gamma(\ell_2^n) \to \ell_2^n\| = 1/s_n(\Gamma)$.

To estimate the smallest singular number, we will use the following equivalence, which clearly holds for every matrix Γ and every $\lambda \ge 0$:

$$s_n(\Gamma) \le \lambda \quad \Longleftrightarrow \quad \exists x \in S^{n-1} \;\text{ s.t. }\; |\Gamma x| \le \lambda. \tag{1.2}$$

1.2 Compressible and incompressible vectors.

As equivalence (1.2) suggests, to estimate the smallest singular value of a matrix Γ we estimate the norm $|\Gamma x|$ for vectors $x \in S^{n-1}$. More precisely, we will estimate $|\Gamma x|$ individually for vectors in an appropriately chosen ε-net and, for general points of the sphere, we use approximation (to points of the ε-net) and the union bound. In the case of "tall" matrices just one single ε-net is enough for this approximation method to work; but in the case of "almost square" matrices, as well as for square matrices, we need to split the sphere into two parts according to whether the vector x is compressible or incompressible, in the sense that we now define.

Let $m \le n$ and $\rho \in (0, 1)$. A vector $x \in \mathbb{R}^n$ is called

- *m-sparse* if $|\mathrm{supp}(x)| \le m$, that is, if x has at most m nonzero entries.
- *(m, ρ)-compressible* if it is within Euclidean distance ρ from the set of all m-sparse vectors.
- *(m, ρ)-incompressible* if it is not (m, ρ)-compressible.

For convenience we will use the shorthands $Sparse(m)$, $Comp(m,\rho)$, $Incomp(m,\rho)$ to denote the sets of sparse, compressible, and incompressible vectors, respectively. If the parameters m, ρ are clear form the context, sometimes we may write simply $Sparse$, $Comp$ or $Incomp$.

Remark. The idea to split the Euclidean sphere into two parts goes back to Kashin's work [21] on orthogonal decomposition of ℓ_1^{2n}, where the splitting was defined using the ratio of ℓ_2 and ℓ_1 norms. This idea was recently used by Schechtman ([41]) in the same context. The splitting of the sphere essentially as described above appeared in [26], [27], and was later used in many works (e.g. in [39], [40]).

It is clear from these definitions that, for a vector x, the following holds:

$$x \in Comp(m,\rho) \iff \exists \sigma \subset \{1,\dots,n\} \text{ with } |\sigma^c| \le m \text{ such that } |P_\sigma x| \le \rho.$$

Therefore

$$x \in Incomp(m,\rho) \iff \forall \sigma \subset \{1,\dots,n\} \text{ with } |\sigma^c| \le m \text{ one has } |P_\sigma x| > \rho. \tag{1.3}$$

1.3 Two more results.

Here we formulate two results that will be used in the next chapters. The first one is a general form of the Paley-Zygmund inequality (see e.g. [26, Lemma 3.5]). The second one is a quantitative version of the Central Limit Theorem (CLT) which in its original form, under the assumption of finite moments of 3-rd order, is called Berry-Esséen inequality.

Lemma 1.1 (Paley-Zygmund inequality). *Let $p \in (1,\infty)$, $q = p/(p-1)$. Let $f \ge 0$ be a random variable with $\mathbb{E}\, f^{2p} < \infty$. Then for every $0 \le \lambda \le \sqrt{\mathbb{E}\, f^2}$ we have*

$$\mathbb{P}\,(f > \lambda) \ge \frac{(\mathbb{E}\, f^2 - \lambda^2)^q}{(\mathbb{E}\, f^{2p})^{q/p}}.$$

Theorem 1.2 (Berry-Esséen CLT). *Let $2 < r \le 3$. Let $\zeta_1,\dots,\zeta_n$ be independent centered random variables with finite r-th moments and set $\sigma^2 := \sum_{k=1}^n \mathbb{E}|\zeta_k|^2$. Then for all $t \in \mathbb{R}$*

$$\left| \mathbb{P}\Big(\frac{1}{\sigma} \sum_{k=1}^n \zeta_k \le t \Big) - \mathbb{P}\big(g \le t\big) \right| \le \frac{C}{\sigma^r} \sum_{k=1}^n \mathbb{E}|\zeta_k|^r,$$

where $g \sim \mathcal{N}(0,1)$ and C is an absolute constant.

Remarks. 1. The original form of Berry-Esséen inequality requires finite moments of the third order (i.e. is stated for $r = 3$), see e.g. [14, p. 544] or [30, p. 300]. The form presented here, taken from [34] (see Theorem 5.7 there), is a generalization requiring finite moments of order $2+\eta$, a condition in the spirit of Lyapunov's CLT.
2. If $r \ge 3$, then clearly we have boundedness of 3-rd moment for free, and in this case we use the standard form of Berry-Esséen inequality (i.e., with $r = 3$).

Chapter 2

Sub-Gaussians

This chapter deals with sub-Gaussian random variables and their basic properties. We also present equivalent conditions for a random variable to be sub-Gaussian, and we discuss briefly the structure of the class of sub-Gaussian random variables. All the results contained in this chapter are well-known, a pertinent reference is provided wherever possible, though some of the knowledge presented here seems to be 'folklore' and I have abandoned dishearteadly my efforts to track down original sources.

Intuitively, a random variable is called sub-Gaussian when it is subordinate to a Gaussian random variable, in a certain sense. The precise definition will be presented momentarily. As it turns out, sub-Gaussians are a natural kind of random variables for which the properties of Gaussians can be extended ([8]); probably one of the reasons why sub-Gaussians attracted interest in the first place.

To the best of my knowledge, sub-Gaussian random variables were introduced by Kahane in [19], where they played a role to establish a sufficient condition for the almost-sure uniform convergence of certain random series of functions. The name "sub-Gaussian" is the English counterpart of the French "sous-gaussienne" coined by Kahane in [19]. Subsequent works have studied sub-Gaussian random variables and processes either *per se* or in connection with various other subjects. For instance, sub-Gaussian random variables have been studied in connection with random series in [9]; in connection with the geometry of Banach Spaces in [35]; with the spectral properties of random matrices in [26], [40].

2.1 Sub-Gaussian random variables

A real-valued random variable X is said to be *sub-Gaussian* if it has the property that there is some $b > 0$ such that

$$\forall t \in \mathbb{R}, \quad \mathbb{E}\, e^{tX} \leq e^{b^2t^2/2}.$$

In words, this condition says that there is a positive real number b such that the Laplace transform of X is dominated by the Laplace transform of a $\mathcal{N}(0, b^2)$ random variable. When the condition above is satisfied with a particular value of $b > 0$, we say that X is b-sub-Gaussian, or sub-Gaussian with parameter b.

It is an immediate consequence of this definition that sub-Gaussian random variables are centered, and further their variance has a natural upper bound in terms of the sub-Gaussian parameter. We state this "for the records" in the next proposition, whose proof has been borrowed from [43]; although it should be pointed out that this was known much earlier (see e.g. [8]).

Proposition 2.1. *If X is b-sub-Gaussian, then $\mathbb{E}(X) = 0$ and $\mathrm{Var}(X) \leq b^2$.*

Proof. By Lebesgue's Dominated Convergence Theorem, for any $t \in \mathbb{R}$,

$$\sum_{n=0}^{\infty} \frac{t^n}{n!}\,\mathbb{E}(X^n) = \mathbb{E}\,e^{tX} \leq e^{b^2t^2/2} = \sum_{n=0}^{\infty} \frac{b^{2n}t^{2n}}{2^n n!}.$$

Thus

$$\mathbb{E}(X)t + \mathbb{E}(X^2)\frac{t^2}{2!} \leq \frac{b^2t^2}{2} + o(t^2) \quad \text{as } t \to 0.$$

Dividing through by $t > 0$ and letting $t \to 0$ we get $\mathbb{E}(X) \leq 0$. Dividing through by $t < 0$ and letting $t \to 0$ we get $\mathbb{E}(X) \geq 0$. Thus $\mathbb{E}(X) = 0$. Now that this is established, we divide through by t^2 and let $t \to 0$, thus getting $\mathrm{Var}(X) \leq b^2$. ■

Next we look at three natural examples of sub-Gaussian random variables.

Example 2.2. The most natural example of a sub-Gaussian random variable is that of a centered Gaussian. If X has the distribution $\mathcal{N}(0, \sigma^2)$, then an easy computation shows that for any $t \in \mathbb{R}$,

$$\mathbb{E}\,e^{tX} = e^{\sigma^2t^2/2}.$$

Thus X is sub-Gaussian with parameter σ.

Example 2.3. Let X be a random variable with the Rademacher distribution, meaning that the law of X is $\mathbb{P}_X = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$ [here δ_x is the point mass at x]. Then for any $t \in \mathbb{R}$,

$$\mathbb{E}\,e^{tX} = \frac{1}{2}e^{-t} + \frac{1}{2}e^{t} = \cosh t \leq e^{t^2/2},$$

so X is 1-sub-Gaussian. Random variables with this distribution are also called symmetric ± 1 random variables, or symmetric Bernoulli random variables.

Example 2.4. Suppose X is uniformly distributed over the interval $[-a, a]$ for some fixed $a > 0$, meaning the law of X is $\mathbb{P}_X = \frac{1}{2a}\mathbf{1}_{[-a,a]}\lambda$, where λ is Lebesgue measure. Then for any $t \neq 0$ in $\mathbb{R}$,

$$\mathbb{E}\,e^{tX} = \frac{1}{2a}\int_{-a}^{a} e^{tx}\,dx = \frac{1}{2at}[e^{at} - e^{-at}] = \sum_{n=0}^{\infty} \frac{(at)^{2n}}{(2n+1)!}.$$

Using the inequality $(2n+1)! \geq n!2^n$, we see that X is a-sub-Gaussian.

More generally, any centered and bounded random variable is sub-Gaussian, as we demonstrate now (see e.g. [43, Theorem 9.9]).

Theorem 2.5. *If X is a random variable with $\mathbb{E}(X) = 0$ and $|X| \leq 1$ a.s., then*

$$\mathbb{E}\, e^{tX} \leq \cosh t \qquad \forall t \in \mathbb{R} \tag{2.1}$$

and so X is 1-sub-Gaussian. Moreover, if equality holds in (2.1) for some $t \neq 0$, then X is a Rademacher variable and hence equality holds for all $t \in \mathbb{R}$.

Proof. Define f on $\mathbb{R}$ by $f(t) := e^t[\cosh t - \mathbb{E}(e^{tX})]$. Thus

$$f(t) = \tfrac{1}{2}e^{2t} + \tfrac{1}{2} - \mathbb{E}(e^{t(1+X)}).$$

For convenience let us set $Y := 1 + X$, so $f(t) = \frac{1}{2}e^{2t} + \frac{1}{2} - \mathbb{E}(e^{tY})$. Apply the Mean Value Theorem and Lebesgue's Dominated Convergence Theorem to conclude $f'(t) = e^{2t} - \mathbb{E}(Ye^{tY})$. Using $\mathbb{E}(Y) = 1$,

$$f'(t) = \mathbb{E}\big(Y(e^{2t} - e^{tY})\big).$$

Since $0 \leq Y \leq 2$ a.s., we have

$$t \geq 0 \quad \Longrightarrow \quad Y(e^{2t} - e^{tY}) \geq 0 \quad a.s.$$

It follows that $f' \geq 0$ and f is increasing on $[0, \infty)$. In particular, for $t \geq 0$ we have $f(t) \geq f(0) = 0$, and so (2.1) holds for $t \geq 0$. Since $-X$ satisfies the same hypothesis as X, we have just proved that (2.1) holds for all $t \in \mathbb{R}$.

Now suppose that equality holds in (2.1) for some $t_0 > 0$. Then $f(t_0) = f(0) = 0$, which implies $f(t) = 0$ for all $t \in [0, t_0]$. Thus $f'(t_0) = 0$, and hence $Y(e^{2t_0} - e^{t_0 Y}) = 0$ a.s. Therefore

$$\mathbb{P}(X = -1) + \mathbb{P}(X = 1) = \mathbb{P}(Y = 0) + \mathbb{P}(Y = 2) = 1.$$

Since $\mathbb{E}(X) = 0$, it follows that $\mathbb{P}(X = -1) = \mathbb{P}(X = 1) = 1/2$, and so X is a Rademacher variable. If, on the other hand, equality holds in (2.1) for some $t_0 < 0$, then applying the same argument to $-t_0 > 0$ and $-X$ we see that $-X$ is a Rademacher variable, and hence so is X. ■

Corollary 2.6. *If X is a random variable with $\mathbb{E}(X) = 0$ and $|X| \leq b$ a.s. for some $b > 0$, then X is b-sub-Gaussian.*

We establish now that the set of all sub-Gaussian random variables has a linear structure. The proof that this set is stable under scalar multiples is trivial. For stability under sums the proof we present comes from [8].

Fact 2.7. If the random variable X is b-sub-Gaussian, then for any $\alpha \in \mathbb{R}$, the random variable αX is $|\alpha| b$-sub-Gaussian. If X_1, X_2 are random variables such that X_i is b_i-sub-Gaussian, then $X_1 + X_2$ is $(b_1 + b_2)$-sub-Gaussian.

Proof. Suppose X is b-sub-Gaussian. For $\alpha \neq 0$, we have

$$\mathbb{E}\, e^{t(\alpha X)} \leq e^{b^2\alpha^2 t^2/2} = e^{(|\alpha| b)^2 t^2/2}.$$

Now suppose that X_i is b_i-sub-Gaussian, for $i = 1, 2$. For any $p, q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$, using Hölder inequality,

$$\begin{aligned}\mathbb{E}\, e^{t(X_1+X_2)} &\leq \Big[\mathbb{E}\big(e^{tX_1}\big)^p\Big]^{1/p} \Big[\mathbb{E}\big(e^{tX_2}\big)^q\Big]^{1/q} \\ &\leq \exp\Big\{\frac{t^2}{2}\big(pb_1^2 + qb_2^2\big)\Big\} = \exp\Big\{\frac{t^2}{2}\big(pb_1^2 + \frac{p}{p-1} b_2^2\big)\Big\}.\end{aligned}$$

Minimizing over $p > 1$ we get

$$\mathbb{E}\, e^{t(X_1+X_2)} \leq \exp\Big\{\frac{t^2}{2}\big(b_1 + b_2\big)^2\Big\},$$

and the claim follows. ■

Remark. In the context of Fact 2.7, if X_1, X_2 are required to be independent, then the parameter $b_1 + b_2$ can be improved to $\sqrt{b_1^2 + b_2^2}$ (see e.g. [19]).

As it turns out, the set of sub-Gaussian random variables has a much richer structure. For a centered random variable X, the *sub-Gaussian moment* of X, denoted $\sigma(X)$, is defined as follows

$$\sigma(X) := \inf\Big\{b \geq 0 \;\Big|\; \mathbb{E}\, e^{tX} \leq e^{b^2t^2/2}, \quad \forall t \in \mathbb{R}\Big\}. \tag{2.2}$$

Clearly X is sub-Gaussian if and only if $\sigma(X) < \infty$. Moreover, the functional $\sigma(\cdot)$ is a norm on the space of sub-Gaussian random variables (upon identification of random variables which are equal almost surely), and this normed space is complete (see e.g. [8]).

Remark. We observe that in case $X \sim \mathcal{N}(0, \sigma^2)$ is a centered Gaussian, then $\sigma(X) = \sigma$. Thus for Gaussian variables the sub-Gaussian moment coincides with the standard deviation.

2.2 Characterization of sub-Gaussians

According to the definition, a real-valued random variable is sub-Gaussian when its Laplace transform is dominated by the Laplace transform of a centered Gaussian. The following theorem presents equivalent conditions for a random variable to be sub-Gaussian. The calculations used to prove it are well known, the absence of a reference should not be taken as a claim of originality but rather as reflecting the fact that this is folklore knowledge.

Theorem 2.8. *For a centered random variable X, the following statements are equivalent:*

(1) Laplace transform condition: $\quad \exists b > 0, \quad \forall t \in \mathbb{R}, \quad \mathbb{E}\, e^{tX} \leq e^{b^2t^2/2}$;

(2) sub-Gaussian tail estimate: $\quad \exists c > 0, \quad \forall \lambda > 0, \quad \mathbb{P}(|X| \geq \lambda) \leq 2e^{-c\lambda^2}$;

(3) ψ_2-condition: $\quad \exists a > 0, \quad \mathbb{E}\, e^{aX^2} \leq 2$.

Proof. (1) $\Rightarrow$ (2) Using Markov's inequality, for any $t > 0$ we have

$$\mathbb{P}(X \geq \lambda) = \mathbb{P}(tX \geq t\lambda) \leq \frac{\mathbb{E}\, e^{tX}}{e^{t\lambda}} \leq e^{-t\lambda + b^2t^2/2},$$

hence, minimizing over $t > 0$ we get

$$\mathbb{P}(X \geq \lambda) \leq \inf_{t>0} e^{-t\lambda + b^2t^2/2} = e^{-\lambda^2/2b^2}.$$

Similarly one sees that $\mathbb{P}(X \leq -\lambda) \leq e^{-\lambda^2/2b^2}$. Then, using the union bound, we get $\mathbb{P}(|X| \geq \lambda) \leq 2e^{-\lambda^2/2b^2}$, and the assertion is proved with $c = 1/(2b^2)$.

(2) $\Rightarrow$ (3) Assuming the sub-Gaussian tail estimate is satisfied with a constant c, for any a with $0 < a < c$ we have

$$\begin{aligned}
\mathbb{E}\, e^{aX^2} &\leq 1 + \int_0^\infty 2ate^{at^2} \cdot \mathbb{P}(|X| > t)dt \\
&\leq 1 + \int_0^\infty 2at \cdot 2e^{-(c-a)t^2} dt = 1 + \frac{2a}{c-a}.
\end{aligned}$$

Then by taking a small enough (e.g. $a = c/3$) we get $\mathbb{E}\, e^{aX^2} \leq 2$.

(3) $\Rightarrow$ (1) Assume that $\mathbb{E}\, e^{aX^2} \leq 2$ for some $a > 0$. Recalling that X is centered, we have

$$\begin{aligned}
\mathbb{E}\, e^{tX} = 1 + \int_0^1 (1-y)\, \mathbb{E}\big[(tX)^2 e^{ytX}\big] dy &\leq 1 + \frac{t^2}{2}\, \mathbb{E}\big[X^2 e^{|tX|}\big] \\
&\leq 1 + \frac{t^2}{2} e^{t^2/2a}\, \mathbb{E}\big[X^2 e^{aX^2/2}\big] \\
&\leq 1 + \frac{t^2}{2a} e^{t^2/2a}\, \mathbb{E}\, e^{aX^2} \\
&\leq \Big(1 + \frac{t^2}{a}\Big) e^{t^2/2a},
\end{aligned}$$

From this it is clear that X is sub-Gaussian with parameter $\beta = \sqrt{\frac{3}{a}}$. ■

Remark. If the random variable X has the Gaussian distribution $\mathcal{N}(0, \sigma^2)$, then for each $p > 0$ one has

$$\mathbb{E}|X|^p = \sqrt{\frac{2^p}{\pi}}\, \sigma^p\, \Gamma\Big(\frac{p+1}{2}\Big).$$

For instance, this can be deduced from the following identity, which holds for all $\beta > 0$ and $r > -1$,

$$\int_0^\infty t^r e^{-\frac{t^2}{2\beta^2}} dt = 2^{\frac{r-1}{2}} \beta^{r+1} \Gamma\Big(\frac{r+1}{2}\Big).$$

In fact, if the random variable X is sub-Gaussian, then its (absolute) moments are bounded above by an expression involving the sub-Gaussian parameter and the gamma function, somewhat similar to the right hand side of the above expression for the moments of a Gaussian.

Fact 2.9. If X is b-sub-Gaussian, then for any $p > 0$ one has

$$\mathbb{E}|X|^p \leq p\, 2^{\frac{p}{2}}\, b^p\, \Gamma\big(\frac{p}{2}\big).$$

Consequently, for $p \geq 1$,

$$\|X\|_{L_p} = \big(\mathbb{E}|X|^p\big)^{1/p} \leq Cb\,\sqrt{p}.$$

Conversely, if a centered random variable X satisfies $\big(\mathbb{E}|X|^p\big)^{1/p} \leq Cb\,\sqrt{p}$ for all $p \geq 1$, then X is sub-Gaussian.

2.3 The Orlicz space L_{ψ_2}

By ψ_2 we denote the Orlicz function $\psi_2(x) = e^{x^2} - 1$. The purpose of this section is to construct a special normed space associated to this function, and to give some insight into the ψ_2-condition. This material is borrowed from [36]. We define

$$\mathcal{L}_{\psi_2} = \left\{ f : \Omega \to \mathbb{R} \ \text{ measurable } \ \middle| \ \mathbb{E}\,\psi_2\Big(\frac{|f|}{t}\Big) < \infty \quad \text{for some } t > 0 \right\}.$$

We claim that this is a linear space. For it is clear that the zero function is in $\mathcal{L}_{\psi_2}$. Given any $f \in \mathcal{L}_{\psi_2}$ and real number $\lambda \neq 0$, let $t > 0$ be such that $\mathbb{E}\,\psi_2\big(|f|/t\big) < \infty$, and set $t' = |\lambda| t$. We have

$$\mathbb{E}\,\psi_2\Big(\frac{|\lambda f|}{t'}\Big) = \mathbb{E}\,\psi_2\Big(\frac{|f|}{t}\Big) < \infty,$$

which proves that $\lambda f \in \mathcal{L}_{\psi_2}$. Finally, if $f, g \in \mathcal{L}_{\psi_2}$, choose $t, s > 0$ such that $\mathbb{E}\,\psi_2\big(|f|/t\big) < \infty$ and $\mathbb{E}\,\psi_2\big(|g|/s\big) < \infty$. Since the function ψ_2 is increasing and convex, we have

$$\begin{aligned}\psi_2\Big(\frac{|f+g|}{t+s}\Big) &\leq \psi_2\Big(\frac{|f|+|g|}{t+s}\Big) \\ &\leq \frac{t}{t+s}\psi_2\Big(\frac{|f|}{t}\Big) + \frac{s}{t+s}\psi_2\Big(\frac{|g|}{s}\Big).\end{aligned}$$

Then, taking expectations,

$$\mathbb{E}\,\psi_2\Big(\frac{|f+g|}{t+s}\Big) \leq \frac{t}{t+s}\,\mathbb{E}\,\psi_2\Big(\frac{|f|}{t}\Big) + \frac{s}{t+s}\,\mathbb{E}\,\psi_2\Big(\frac{|g|}{s}\Big). \tag{2.3}$$

Since the right hand side is finite, we see that $f + g \in \mathcal{L}_{\psi_2}$.

We define a functional $\|\cdot\|_{\psi_2} : \mathcal{L}_{\psi_2} \to \mathbb{R}$ by setting

$$\|f\|_{\psi_2} = \inf\left\{ t > 0 \;\middle|\; \mathbb{E}\,\psi_2\Big(\frac{|f|}{t}\Big) \leq 1 \right\}. \tag{2.4}$$

Given $f \in \mathcal{L}_{\psi_2}$, choose $t > 0$ such that $\mathbb{E}\,\psi_2\big(|f|/t\big) < \infty$. Since ψ_2 is increasing for positive values, it follows that $\mathbb{E}\,\psi_2\big(|f|/s\big) < \infty$ for all $s \geq t$. Then, using Lebesgue's Dominated Convergence Theorem,

$$\lim_{s\to\infty} \mathbb{E}\,\psi_2\Big(\frac{|f|}{s}\Big) = 0.$$

Then there is some $t_0 > 0$ such that $\mathbb{E}\,\psi_2\big(|f|/t_0\big) \leq 1$. This proves that $\|f\|_{\psi_2} < \infty$ for $f \in \mathcal{L}_{\psi_2}$, showing that $\|\cdot\|_{\psi_2}$ is well defined. It is clear that $\|\cdot\|_{\psi_2} \geq 0$.

If $f = 0$ a.e., then clearly $\|f\|_{\psi_2} = 0$. Conversely, let $f \in \mathcal{L}_{\psi_2}$ be such that $\|f\|_{\psi_2} = 0$. It follows that $\mathbb{E}\,\psi_2\big(n|f|\big) \leq 1$ for all $n \geq 1$. Assuming $\mathbb{P}(|f| > 0) > 0$, we may find some positive number δ such that the event $A := \{\omega \in \Omega \mid |f(\omega)| \geq \delta\}$ has $\mathbb{P}(A) > 0$. Then we have

$$\psi_2\big(n\delta\big)\mathbb{P}(A) = \int_A \psi_2\big(n\delta\big)d\,\mathbb{P} \leq \int_A \psi_2\big(n|f|\big)d\,\mathbb{P} \leq \mathbb{E}\,\psi_2\big(n|f|\big) \leq 1,$$

and letting $n \to \infty$ we get a contradiction. Hence $f = 0$ a.e.

It is clear that if $f \in \mathcal{L}_{\psi_2}$ and $\lambda \in \mathbb{R}$, then $\|\lambda f\|_{\psi_2} = |\lambda|\cdot\|f\|_{\psi_2}$. This is obvious for $\lambda = 0$, and for $\lambda \neq 0$ it follows form properties of the infimum. Thus the functional $\|\cdot\|_{\psi_2}$ is (positively) homogeneous.

To see that $\|\cdot\|_{\psi_2}$ satisfies the triangle inequality, let $f, g \in \mathcal{L}_{\psi_2}$, and choose $t, s > 0$ such that $\mathbb{E}\,\psi_2\big(|f|/t\big) \leq 1$ and $\mathbb{E}\,\psi_2\big(|g|/s\big) \leq 1$. Using inequality (2.3), we obtain

$$\mathbb{E}\,\psi_2\Big(\frac{|f+g|}{t+s}\Big) \leq \frac{t}{t+s}\,\mathbb{E}\,\psi_2\Big(\frac{|f|}{t}\Big) + \frac{s}{t+s}\,\mathbb{E}\,\psi_2\Big(\frac{|g|}{s}\Big) \leq 1.$$

Thus $\|f+g\|_{\psi_2} \leq t + s$, and taking infimum one at a time over t and s we get $\|f+g\|_{\psi_2} \leq \|f\|_{\psi_2} + \|g\|_{\psi_2}$.

We have thus established that $\|\cdot\|_{\psi_2}$ is a seminorm on $\mathcal{L}_{\psi_2}$. Upon identifying functions in $\mathcal{L}_{\psi_2}$ which are equal almost everywhere, we obtain a normed space $(L_{\psi_2}, \|\cdot\|_{\psi_2})$, called the Orlicz space associated to ψ_2. As is done with the Lebesgue spaces L_p, we regard the elements of L_{ψ_2} as functions, thus avoiding the awkward treatment of 'classes of functions' and 'representatives' and so on. However, we should keep in mind that equality of elements in L_{ψ_2} means equality a.e.

Remark. Orlicz spaces can be defined for more general functions than ψ_2, the reader is referred to [36] to see the requirements on a function $\psi : [0,\infty) \to \mathbb{R}$ that allow the construction of a space L_ψ and a norm $\|\cdot\|_\psi$ with the same properties as L_{ψ_2} and $\|\cdot\|_{\psi_2}$ as we did above for $\psi_2(x) = e^{x^2} - 1$.

We close this chapter by establishing the link between sub-Gaussian random variables and the Orlicz space L_{ψ_2}.

Proposition 2.10.

$$\|f\|_{\psi_2} \leq 1 \quad \textit{if and only if} \quad \mathbb{E}\,\psi_2\big(|f|\big) \leq 1 \quad \textit{if and only if} \quad \mathbb{E}\,e^{f^2} \leq 2.$$

Proof. The first equivalence is clear if $\|f\|_{\psi_2} = 0$; and in case $\|f\|_{\psi_2} > 0$, setting $a := \|f\|_{\psi_2}$ we first note that

$$\mathbb{E}\,\psi_2\Big(\frac{|f|}{a}\Big) \leq 1.$$

If $a \leq 1$, then $\mathbb{E}\,\psi_2\big(|f|\big) \leq 1$ by the monotonicity of the function ψ_2. Conversely, if $\mathbb{E}\,\psi_2\big(|f|\big) \leq 1$, then $1 \in \big\{t > 0 \;\big|\; \mathbb{E}\,\psi_2\big(|f|/t\big)\big\}$, so upon taking the infimum of this set we get $a \leq 1$.

The second equivalence is obvious. ∎

Comparing Proposition 2.10 and the ψ_2-condition from Theorem 2.8, it is now evident that a random variable is sub-Gaussian precisely when it is an element of the space L_{ψ_2}.

Chapter 3

Small ball probabilities

In this chapter we gather auxiliary results related to random sums, their small ball probabilities, etc., which are needed later. In fact, we adjust corresponding results from [26] and [39] to our setting. These results are also of independent interest. We provide proofs for the sake of completeness.

The following lemma provides a lower bound on the small ball probability of a random sum. Its proof follows the steps of [26, Lemma 3.6] with the appropriate modification to deal with centered random variables (rather than symmetric), to remove the assumption that the variances are bounded from below uniformly, and to replace the condition of finite 3-rd moments by finite r-th moments ($r > 2$).

Lemma 3.1. *Let $2 < r \leq 3$ and $\mu \geq 1$. Suppose $\xi_1, \ldots, \xi_n$ are independent centered random variables such that $\mathbb{E}|\xi_i|^r \leq \mu^r$ for every $i = 1, \ldots, n$. Let $x = (x_i) \in \ell_2$ be such that $|x| = 1$. Then for every $\lambda \geq 0$*

$$\mathbb{P}\left(\Big|\sum_{i=1}^{n} \xi_i x_i\Big| > \lambda\right) \geq \left(\frac{[\mathbb{E}\sum_{i=1}^{n} \xi_i^2 x_i^2 - \lambda^2]_+}{8\mu^2}\right)^{r/(r-2)}.$$

Proof. Define $f = \left|\sum_{i=1}^{n} \xi_i x_i\right|$. Let $\varepsilon_1, \ldots, \varepsilon_n$ be independent symmetric Bernoulli ± 1 random variables, which are also independent of $\xi_1 \ldots, \xi_n$. Using the symmetrization inequality [24, Lemma 6.3], and applying Khinchine's inequality, we obtain

$$\mathbb{E}\, f^r \leq 2^r\, \mathbb{E}\Big|\sum_{i=1}^{n} \varepsilon_i \xi_i x_i\Big|^r = 2^r\, \mathbb{E}_\xi\, \mathbb{E}_\varepsilon \Big|\sum_{i\geq 1} \varepsilon_i \xi_i x_i\Big|^r \leq 2^r 2^{r/2}\, \mathbb{E}_\xi \left(\sum_{i\geq 1} \xi_i^2 x_i^2\right)^{r/2}.$$

Now consider the set

$$\mathcal{S} := \left\{ s = (s_i) \in \ell_1 \;\middle|\; s_i \geq 0 \text{ for every } i \text{ and } \sum_{i\geq 1} s_i = 1 \right\}.$$

We define a function $\varphi : \mathcal{S} \to \mathbb{R}$ by

$$\varphi(s) = \mathbb{E}_\xi \left(\sum_{i\geq 1} \xi_i^2 s_i\right)^{r/2}.$$

This function is clearly convex, so that

$$\sup_{s\in\mathcal{S}} \varphi(s) = \sup_{i\geq 1} \varphi(e_i) = \sup_{i\geq 1} \mathbb{E}_\xi (\xi_i^2)^{r/2} \leq \mu^r.$$

Thus $\mathbb{E}\, f^r \leq 2^{3r/2}\mu^r$. On the other hand, using the independence of $\xi_1, \ldots, \xi_n$,

$$\mathbb{E}\, f^2 = \mathbb{E} \sum_{i\geq 1} \xi_i^2 x_i^2.$$

Lemma 1.1 with $p = r/2$, $q = r/(r-2)$ implies the desired estimate. ■

The next proposition, which is a consequence of Theorem 1.2, allows us to estimate the small ball probability. The proof goes along the same lines as the proof of [26, Proposition 3.2] (see also [28, Proposition 3.4]), with slight modifications to remove the assumption about variances. Recall that for a subset $\sigma \subset \{1, 2, \ldots, n\}$, P_σ denotes the coordinate projection onto $\mathbb{R}^\sigma$.

Proposition 3.2. *Let $2 < r \leq 3$ and $\mu \geq 1$. Let $(\xi_i)_{i=1}^n$ be independent centered random variables with $\mathbb{E}\,|\xi_i|^r \leq \mu^r$ for all $i = 1, 2, \ldots, n$. There is a universal constant $c > 0$ such that*

(a) For every $a < b$ and every $x = (x_i) \in \mathbb{R}^n$ satisfying $A := \sqrt{\mathbb{E}\sum_{i=1}^n \xi_i^2 x_i^2} > 0$ one has

$$\mathbb{P}\Big(a \leq \sum_{i=1}^n \xi_i x_i < b\Big) \leq \frac{b-a}{\sqrt{2\pi}A} + c\Big(\frac{\|x\|_r}{A}\mu\Big)^r.$$

(b) For every $t > 0$, every $x = (x_i) \in \mathbb{R}^n$ and every $\sigma \subset \{1, 2, \ldots, n\}$ satisfying $A_\sigma := \sqrt{\mathbb{E}\sum_{i\in\sigma} \xi_i^2 x_i^2} > 0$ one has

$$\sup_{v\in\mathbb{R}} \mathbb{P}\Big(\Big|\sum_{i=1}^n x_i\xi_i - v\Big| < t\Big) \leq \frac{2t}{\sqrt{2\pi}A_\sigma} + c\Big(\frac{\|P_\sigma x\|_r}{A_\sigma}\mu\Big)^r.$$

The next corollary gives an estimate on the small ball probability in the spirit of [39, Corollary 2.10].

Corollary 3.3. *Let $2 < r \leq 3$ and $\mu \geq 1$. Let $\xi_1, \ldots, \xi_n$ be independent centered random variables with $\mathbb{E}\,|\xi_i|^r \leq \mu^r$ for every $i = 1, \ldots, n$. Suppose $x = (x_i) \in \mathbb{R}^n$ and $\sigma \subset \{1, \ldots, n\}$ are such that $A \leq |x_i| \leq B$ and $\mathbb{E}\,\xi_i^2 \geq 1$ for all $i \in \sigma$. Then for all $t \geq 0$*

$$\sup_{v\in\mathbb{R}} \mathbb{P}\Big(\Big|\sum_{i=1}^n x_i\xi_i - v\Big| < t\Big) \leq \frac{C}{|\sigma|^{r/2-1}}\Big(\frac{t}{A} + \mu^r\Big(\frac{B}{A}\Big)^r\Big),$$

where $C > 0$ is an absolute constant.

Proof. By assumptions on coordinates of x we have

$$A_\sigma^2 := \mathbb{E}\sum_{i\in\sigma} \xi_i^2 x_i^2 \geq |\sigma| A^2$$

and

$$\|P_\sigma x\|_r^r = \sum_{i\in\sigma} |x_i|^r \leq |\sigma| B^r.$$

Then, by part (b) of Proposition 3.2

$$\begin{aligned}\sup_{v\in\mathbb{R}} \mathbb{P}\Big(\Big|\sum_{i=1}^n x_i\xi_i - v\Big| < t\Big) &\leq \sqrt{\frac{2}{\pi}}\frac{t}{A|\sigma|^{1/2}} + c\mu^r \frac{B^r|\sigma|}{A^r|\sigma|^{r/2}} \\ &\leq \frac{C}{|\sigma|^{r/2-1}}\Big(\frac{t}{A} + \mu^r\Big(\frac{B}{A}\Big)^r\Big).\end{aligned}$$

■

We need the following lemma proved in [39, Lemma 3.4].

Lemma 3.4. *Let $\gamma, \rho \in (0,1)$, and let $x \in Incomp(\gamma n, \rho)$. Then there exists a set $\sigma = \sigma_x \subset \{1,\dots,n\}$ of cardinality $|\sigma| \geq \frac{1}{2}\rho^2\gamma n$ and such that for all $k \in \sigma$*

$$\frac{\rho}{\sqrt{2n}} \leq |x_k| \leq \frac{1}{\sqrt{\gamma n}}.$$

The next lemma is a version of [39, Lemma 3.7], modified in order to remove the assumption "*variances* ≥ 1".

Lemma 3.5. *Let $2 < r \leq 3$ and $\mu \geq 1$. Let $\xi_1,\dots,\xi_n$ be independent centered random variables with $\mathbb{E}|\xi_i|^r \leq \mu^r$ for every i. Suppose $\overline{\sigma} := \{i \mid \mathbb{E}\xi_i^2 \geq 1\}$ has cardinality $|\overline{\sigma}| \geq a_4 n$. Let $\gamma, \rho \in (0,1)$, and consider a vector $x \in Incomp(\gamma n, \rho)$. Assuming that $a_4 + \frac{1}{2}\rho^2\gamma > 1$ we have for every $t \geq 0$*

$$\sup_{v\in\mathbb{R}} \mathbb{P}\Big(\Big|\sum_{i=1}^n x_i\xi_i - v\Big| < t\Big) \leq c(tn^{\frac{3-r}{2}} + \mu^r n^{\frac{2-r}{2}}),$$

where c is a positive constant which depends on γ, ρ, a_4, and r.

Proof. Let σ_x be the set of spread coefficients of x from Lemma 3.4, so that $|\sigma_x| \geq \frac{1}{2}\rho^2\gamma n$. Set $\sigma := \overline{\sigma} \cap \sigma_x$. Then

$$|\sigma| = |\overline{\sigma}| + |\sigma_x| - |\overline{\sigma} \cup \sigma_x| \geq a_4 n + \frac{1}{2}\rho^2\gamma n - n =: c_0 n.$$

By the construction, for every $i \in \sigma$ we have

$$\frac{\rho}{\sqrt{2n}} \leq |x_i| \leq \frac{1}{\sqrt{\gamma n}}.$$

Applying Corollary 3.3 we obtain

$$
\begin{aligned}
\sup_{v\in\mathbb{R}} \mathbb{P}\Big(\Big|\sum_{i=1}^{n} x_i\xi_i - v\Big| < t\Big) &\le \frac{C}{|\sigma|^{r/2-1}}\Big(\frac{\sqrt{2n}t}{\rho} + \mu^r\Big(\frac{\sqrt{2}}{\rho\sqrt{\gamma}}\Big)^r\Big) \\
&\le \frac{C}{(c_0 n)^{r/2-1}}\Big(\frac{\sqrt{2n}t}{\rho} + \mu^r\Big(\frac{\sqrt{2}}{\rho\sqrt{\gamma}}\Big)^r\Big) \\
&\le c(tn^{\frac{3-r}{2}} + \mu^r n^{\frac{2-r}{2}}).
\end{aligned}
$$

■

Chapter 4

Tall random matrices

In this chapter we prove Theorem 1, which establishes an estimate on the smallest singular value for "tall" random matrices, meaning matrices whose aspect ratio n/N is bounded above by a small positive constant (independent of n and N). It is important to notice that Theorem 1 uses only conditions (i), (ii), and (iii), i.e. no condition on the rows is required here.

The proof depends upon an estimate on the norm $|\Gamma x|$ for a fixed vector x, which is provided by the following proposition.

Proposition 4.1. *Let $1 \le n < N$ be positive integers. Suppose Γ is a matrix of size $N \times n$ whose entries are independent centered random variables satisfying conditions (i), (ii) and (iii) for some $2 < r \le 3$, $\mu \ge 1$ and $a_1, a_2, a_3 > 0$ with $a_3 < \mu$. Then for every $x \in S^{n-1}$ we have*

$$\mathbb{P}\Big(|\Gamma x| \le b_1 \sqrt{N}\Big) \le e^{-b_2 N},$$

where $b_1, b_2 > 0$ depend only on μ, a_3 and r.

Remark. Our proof gives that

$$b_1^2 = \frac{a_3^4}{2^5 \mu^2} \Big(\frac{a_3^2}{2^5 \mu^2}\Big)^{r/(r-2)}, \qquad b_2 = \frac{a_3^2}{2^3 \mu^2} \Big(\frac{a_3^2}{2^5 \mu^2}\Big)^{r/(r-2)}.$$

In order to keep the flow of our exposition uninterrupted we present right away the proof of Theorem 1. The proof of Proposition 4.1 will come right after.

Proof of Theorem 1. Passing to $r_0 = \min\{3, r\}$ we may assume without lost of generality that $r \le 3$.

Let $t \ge 0$ and $\Omega_0 := \big\{\omega \;\big|\; \|\Gamma\| \le a_1 \sqrt{N}\big\}$. By (1.2) it is enough to estimate the probability of the event

$$E := \big\{\omega \;\big|\; \exists x \in S^{n-1} \text{ s.t. } |\Gamma x| \le t\sqrt{N}\big\}.$$

To this end we use the inclusion $E \subset (E \cap \Omega_0) \cup \Omega_0^c$ and the union bound.

To estimate $\mathbb{P}(E \cap \Omega_0)$, let $0 < \varepsilon \le 1$, and let $\mathcal{N}$ be an ε-net of S^{n-1} with cardinality $|\mathcal{N}| \le (3/\varepsilon)^n$. For any $x \in S^{n-1}$ we can find $y \in \mathcal{N}$ such that $|x - y| \le \varepsilon$. If further x satisfies $|\Gamma x| \le t\sqrt{N}$, then the corresponding y satisfies

$$|\Gamma y| \le |\Gamma x| + \|\Gamma\| \cdot |y - x| \le t\sqrt{N} + \varepsilon a_1 \sqrt{N} = (t + \varepsilon a_1)\sqrt{N}. \tag{4.1}$$

Taking $\varepsilon = \min\{t/a_1,\, 1\}$, we see that for each $x \in S^{n-1}$ satisfying $|\Gamma x| \le t\sqrt{N}$ there is a corresponding $y \in \mathcal{N}$ such that $|x - y| \le \varepsilon$ and $|\Gamma y| \le 2t\sqrt{N}$. Hence, using the union bound, setting $t = b_1/2$ and using Proposition 4.1, one has

$$\begin{aligned}\mathbb{P}(E \cap \Omega_0) &\le \sum_{y \in \mathcal{N}} \mathbb{P}\Big(|\Gamma y| \le 2t\sqrt{N}\Big) \\ &\le |\mathcal{N}| e^{-b_2 N} \le \Big(\frac{3}{\varepsilon}\Big)^n e^{-b_2 N},\end{aligned}$$

where b_1 and b_2 are as in Proposition 4.1. Thus

$$\mathbb{P}(E \cap \Omega_0) \le \exp\Big(-\frac{b_2 N}{2}\Big)$$

as long as

$$\Big(\frac{3}{\varepsilon}\Big)^n \le \exp\Big(\frac{b_2 N}{2}\Big).$$

Bearing in mind that $N = (1+\delta)n$, we can see that the last condition is satisfied if

$$\delta \ge \delta_0 := \max\left\{\frac{2}{b_2} \ln\Big(\frac{6a_1}{b_1}\Big),\ \frac{2}{b_2} \ln 3\right\}. \tag{4.2}$$

To finish, we use $\mathbb{P}(E) \le \mathbb{P}(E \cap \Omega_0) + \mathbb{P}(\Omega_0^c)$ with the estimate for $\mathbb{P}(E \cap \Omega_0)$ just obtained and the estimate $\mathbb{P}(\Omega_0^c) \le e^{-a_2 N}$ coming from condition (ii). ■

Proof of Proposition 4.1. Take an arbitrary $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ with $|x| = 1$. For $a > 0$ (a parameter whose value will be specified later), define a set of "good" rows as follows:

$$J = J(a) = \left\{ j \in \{1, \dots, N\} \;\middle|\; \mathbb{E} \sum_{i=1}^n \xi_{ji}^2 x_i^2 \ge a \right\}.$$

Suppose that the cardinality of set J is $|J| = \alpha N$ for some $\alpha \in [0,1]$. Note that for each index $j = 1, \dots, N$ we have

$$\mathbb{E} \sum_{i=1}^n \xi_{ji}^2 x_i^2 \le \max_{1 \le i \le n} \mathbb{E}\, \xi_{ji}^2 \le \max_{1 \le i \le n} (\mathbb{E}\, \xi_{ji}^r)^{2/r} \le \mu^2.$$

Then on one side we have

$$\begin{aligned}\sum_{j=1}^N \Big(\mathbb{E} \sum_{i=1}^n \xi_{ji}^2 x_i^2\Big) &= \sum_{j \in J} \Big(\mathbb{E} \sum_{i=1}^n \xi_{ji}^2 x_i^2\Big) + \sum_{j \in J^c} \Big(\mathbb{E} \sum_{i=1}^n \xi_{ji}^2 x_i^2\Big) \\ &\le \mu^2 \alpha N + a(1-\alpha)N,\end{aligned}$$

while on the other hand, using condition (iii),

$$\sum_{j=1}^{N}\Big(\mathbb{E}\sum_{i=1}^{n}\xi_{ji}^2x_i^2\Big)=\sum_{i=1}^{n}\Big(\mathbb{E}\sum_{j=1}^{N}\xi_{ji}^2\Big)x_i^2\geq\sum_{i=1}^{n}a_3^2Nx_i^2=a_3^2N.$$

Hence we have $\mu^2\alpha N+a(1-\alpha)N\geq a_3^2N$, so α satisfies

$$\alpha\geq\frac{a_3^2-a}{\mu^2-a}. \tag{4.3}$$

Note that for each $j=1,\ldots,N$, the j-th entry of Γx is $(\Gamma x)_j=\sum_{i=1}^n\xi_{ji}x_i$. Define $f_j:=\big|\sum_{i=1}^n\xi_{ji}x_i\big|$, so

$$|\Gamma x|^2=\sum_{j=1}^{N}f_j^2.$$

Clearly $f_1,\ldots,f_N$ are independent. For any $t,\tau>0$ we have

$$\begin{aligned}\mathbb{P}\big(|\Gamma x|^2\leq t^2N\big)&=\mathbb{P}\Big(\sum_{j=1}^{N}f_j^2\leq t^2N\Big)\\&=\mathbb{P}\Big(\tau N-\frac{\tau}{t^2}\sum_{j=1}^{N}f_j^2\geq 0\Big)\\&\leq\mathbb{E}\exp\Big(\tau N-\frac{\tau}{t^2}\sum_{j=1}^{N}f_j^2\Big)\\&=e^{\tau N}\prod_{j=1}^{N}\mathbb{E}\exp\Big(-\frac{\tau f_j^2}{t^2}\Big).\end{aligned} \tag{4.4}$$

From Lemma 3.1 we know that for every $j=1,\ldots,N$,

$$\mathbb{P}(f_j>\lambda)\geq\left(\frac{[\mathbb{E}\sum_{i=1}^n\xi_{ji}^2x_i^2-\lambda^2]_+}{8\mu^2}\right)^{r/(r-2)}=:\beta_j(r). \tag{4.5}$$

Note that for every $j\in J$ one has

$$\beta_j\geq\left(\frac{[a-\lambda^2]_+}{8\mu^2}\right)^{r/(r-2)}. \tag{4.6}$$

For arbitrary $t>0$, $\eta>0$ and $\lambda>0$, set $\tau:=\frac{\eta t^2}{\lambda^2}$. For each $j=1,\ldots,N$ we have

$$\begin{aligned}\mathbb{E}\exp\Big(-\frac{\tau f_j^2}{t^2}\Big)&=\int_0^1\mathbb{P}\Big(\exp\Big(-\frac{\eta f_j^2}{\lambda^2}\Big)>s\Big)ds\\&=\int_0^{e^{-\eta}}\mathbb{P}\Big(\exp\Big(\frac{\eta f_j^2}{\lambda^2}\Big)<\frac{1}{s}\Big)ds+\int_{e^{-\eta}}^1\mathbb{P}\Big(\exp\Big(\frac{\eta f_j^2}{\lambda^2}\Big)<\frac{1}{s}\Big)ds\\&\leq e^{-\eta}+\mathbb{P}(f_j<\lambda)(1-e^{-\eta}).\end{aligned}$$

Choosing $\eta=\ln 2$ and applying (4.5), we obtain

$$\mathbb{E}\exp\Big(-\frac{\tau f_j^2}{t^2}\Big)\leq e^{-\eta}+(1-\beta_j(r))(1-e^{-\eta})=1-\frac{\beta_j(r)}{2}\leq\exp\left(-\frac{\beta_j(r)}{2}\right).$$

Since $\tau < \frac{t^2}{\lambda^2}$, inequality (4.4) implies

$$\mathbb{P}\big(|\Gamma x|^2 \le t^2 N\big) \le e^{\tau N} \prod_{j=1}^{N} e^{-\beta_j(r)/2} \le e^{(t^2/\lambda^2)N} \prod_{j\in J} e^{-\beta_j(r)/2}. \tag{4.7}$$

Taking $a = a_3^2/2$ and $\lambda = a_3/2$ and using (4.6) we observe that for every $j \in J$ we have $\beta_j \ge \big(\frac{a_3^2}{32\mu^2}\big)^{r/(r-2)}$. Also note this choice of a and (4.3) imply $\alpha \ge a_3^2/(2\mu^2)$. Now let

$$t^2 := \frac{a_3^4}{2^5\mu^2}\Big(\frac{a_3^2}{2^5\mu^2}\Big)^{r/(r-2)}.$$

Then continuing from (4.7) we obtain

$$\mathbb{P}\bigg(|\Gamma x|^2 \le \frac{a_3^4}{2^5\mu^2}\Big(\frac{a_3^2}{2^5\mu^2}\Big)^{r/(r-2)} N\bigg) \le \exp\bigg\{-\frac{a_3^2}{2^3\mu^2}\Big(\frac{a_3^2}{2^5\mu^2}\Big)^{r/(r-2)} N\bigg\}.$$

This completes the proof. ■

Chapter 5

Almost square random matrices

In this chapter we prove Theorem 2. We will be using all conditions (i) through (iv). The two key ingredients for the proof of this theorem are Proposition 4.1 and Proposition 3.2.

Proof of Theorem 2. Passing to $r_0 = \min\{3, r\}$ we may assume without loss of generality that $r \le 3$.

Consider the event

$$E := \big\{\omega \;\big|\; \exists x \in S^{n-1} \;\text{ s.t. }\; |\Gamma x| \le t\sqrt{N}\big\}.$$

By equivalence (1.2) we are to estimate $\mathbb{P}(E)$ with an appropriate value of t (which will be specified later).

We split the set E into two sets E_C and E_I defined as follows:

$$\begin{aligned} E_C &= \big\{\omega \;\big|\; \exists x \in S^{n-1} \cap Comp(m,\rho) \;\text{ s.t. }\; |\Gamma x| \le t\sqrt{N}\big\}, \\ E_I &= \big\{\omega \;\big|\; \exists x \in S^{n-1} \cap Incomp(m,\rho) \;\text{ s.t. }\; |\Gamma x| \le t\sqrt{N}\big\}, \end{aligned}$$

where $m \le n$ and $\rho \in (0,1)$ will be specified later.

Define $\Omega_0 := \big\{\omega \;\big|\; \|\Gamma\| \le a_1\sqrt{N}\big\}$. We will estimate $\mathbb{P}(E)$ using the union bound in the inclusion

$$E \subset (E_C \cap \Omega_0) \cup (E_I \cap \Omega_0) \cup \Omega_0^c. \tag{5.1}$$

Our proof will require that $t \le 1$ (which will be satisfied once we choose t, see (5.16) below); and furthermore that t and ρ satisfy

$$\frac{2t}{a_1} \le \rho \le \frac{1}{4}. \tag{5.2}$$

Case 1: Probability of $E_C \cap \Omega_0$. We work on the set $Comp(m,\rho)$, where $m \le n$ and $\rho \in (0,1)$ will be specified later.

Given $x \in S^{n-1} \cap Comp(m,\rho)$, choose $y \in Sparse(m)$ so that $|y - x| \le \rho$. It is clear that we may choose such a y in B_2^n (and thus $1 - \rho \le |y| \le 1$). Note that on Ω_0 we have $\|\Gamma\| \le a_1\sqrt{N}$. Thus if x satisfies $|\Gamma x| \le t\sqrt{N}$ then

$$|\Gamma y| \le |\Gamma x| + \|\Gamma\| \cdot |y - x| \le t\sqrt{N} + a_1\rho\sqrt{N} = (t + a_1\rho)\sqrt{N}.$$

Let $\mathcal{N}$ be a ρ-net in the set $B_2^n \cap Sparse(m)$. We may choose such a net with cardinality

$$|\mathcal{N}| \le \binom{n}{m}\Big(\frac{3}{\rho}\Big)^m \le \Big(\frac{en}{m}\Big)^m\Big(\frac{3}{\rho}\Big)^m = \Big(\frac{3en}{\rho m}\Big)^m.$$

For $y \in B_2^n \cap Sparse(m)$ chosen above, let $v \in \mathcal{N}$ be such that $|v - y| \le \rho$. We observe that, by (5.2),

$$|v| \ge |y| - \rho \ge 1 - 2\rho \ge \frac{1}{2},$$

and, by another use of (5.2),

$$\begin{aligned} |\Gamma v| &\le |\Gamma y| + \|\Gamma\| \cdot |v - y| \\ &\le (t + a_1\rho)\sqrt{N} + \rho a_1 \sqrt{N} \\ &= (t + 2a_1\rho)\sqrt{N} \\ &\le \frac{5a_1\rho}{2}\sqrt{N} \\ &\le 5a_1\rho\sqrt{N}|v|. \end{aligned}$$

Hence

$$\begin{aligned} \mathbb{P}(E_C \cap \Omega_0) &\le \mathbb{P}\Big(\exists v \in \mathcal{N} \text{ s.t. } |\Gamma v| \le 5a_1\rho\sqrt{N}|v|\Big) \\ &\le \sum_{v\in\mathcal{N}} \mathbb{P}\Big(|\Gamma v| \le 5a_1\rho\sqrt{N}|v|\Big). \end{aligned} \tag{5.3}$$

Using Proposition 4.1, we obtain

$$\mathbb{P}\Big(|\Gamma v| \le 5a_1\rho\sqrt{N}|v|\Big) \le e^{-b_2 N},$$

provided that

$$5a_1\rho \le b_1. \tag{5.4}$$

We choose

$$\rho := \min\left\{\frac{1}{4}, \frac{b_1}{5a_1}\right\} \tag{5.5}$$

so that both (5.4) and the right hand side of (5.2) are true. Now, from (5.3), we have

$$\begin{aligned} \mathbb{P}(E_C \cap \Omega_0) &\le |\mathcal{N}| e^{-b_2 N} \\ &\le \Big(\frac{3en}{\rho m}\Big)^m e^{-b_2 N}. \end{aligned}$$

Thus, if

$$m \ln\Big(\frac{3en}{\rho m}\Big) \le \frac{b_2 N}{2} \tag{5.6}$$

then

$$\mathbb{P}(E_C \cap \Omega_0) \le e^{-\frac{b_2 N}{2}}. \tag{5.7}$$

Writing $m = \gamma n$, we see that inequality (5.6) is satisfied if

$$\gamma \ln\Big(\frac{3e}{\rho\gamma}\Big) \le \frac{b_2}{2},$$

so we choose

$$\gamma = \frac{b_2}{4\ln\left(\frac{6e}{\rho b_2}\right)}. \tag{5.8}$$

<u>*Case 2: Probability of $E_I \cap \Omega_0$.*</u> We work on the set $Incomp(m, \rho)$, where ρ is defined in (5.5) and $m = \gamma n$ with γ chosen in (5.8).

For convenience we set $a := t^{1/(r-2)}/a_1$. Since $t \le 1$ and in view of (5.2), we observe that $a \le \rho/2$. Recall also that that on Ω_0 we have $\|\Gamma\| \le a_1\sqrt{N}$.

Let $\mathcal{N}$ be an a-net of the sphere S^{n-1} with cardinality $|\mathcal{N}| \le (3/a)^n$. Let $x \in S^{n-1} \cap Incomp(m, \rho)$ be such that $|\Gamma x| \le t\sqrt{N}$. Recall that by (1.3) one has $|P_\sigma x| \ge \frac{\rho}{2}$ for every $\sigma \subset \{1, \dots, n\}$ with $|\sigma^c| \le m$. Then there is $v \in \mathcal{N}$ such that $|\Gamma v| \le 2t\sqrt{N}$ and with the additional property $|P_\sigma v| \ge \frac{\rho}{2}$ for each $\sigma \subset \{1, \dots, n\}$ with $|\sigma^c| \le m$. Indeed, choosing $v \in \mathcal{N}$ such that $|x - v| \le a$ and using the relation $a_1 a = t^{1/(r-2)} \le t$ (which holds by the choice of a), we have

$$|\Gamma v| \le |\Gamma x| + \|\Gamma\| \cdot |v - x| \le t\sqrt{N} + a_1\sqrt{N}a \le 2t\sqrt{N}$$

and

$$|P_\sigma v| \ge |P_\sigma x| - |P_\sigma(v - x)| \ge \rho - a \ge \frac{\rho}{2},$$

where we used the condition $2a \le 2t/a_1 \le \rho$, required in (5.2).

Denote by $\mathcal{A}$ the set of all vectors $v \in \mathcal{N}$ with the property that for each set $\sigma \subset \{1, \dots, n\}$ with $|\sigma^c| \le m$ we have $|P_\sigma v| \ge \frac{\rho}{2}$. Then

$$\mathbb{P}(E_I \cap \Omega_0) \le \mathbb{P}\Big(\exists v \in \mathcal{A} \text{ s.t. } |\Gamma v| \le 2t\sqrt{N}\Big). \tag{5.9}$$

Now, for each fixed $v = (v_i) \in \mathcal{A}$ we have

$$\begin{aligned}
\mathbb{P}\Big(|\Gamma v|^2 \le 4t^2 N\Big) &= \mathbb{P}\Big(N - \frac{1}{4t^2}|\Gamma v|^2 \ge 0\Big) \\
&\le \mathbb{E}\exp\Big\{N - \frac{1}{4t^2}|\Gamma v|^2\Big\} \\
&= e^N\,\mathbb{E}\exp\Big\{-\frac{1}{4t^2}\sum_{j=1}^N\Big|\sum_{i=1}^n \xi_{ji}v_i\Big|^2\Big\} \\
&= e^N\prod_{j=1}^N\mathbb{E}\exp\Big\{-\frac{1}{4t^2}\Big|\sum_{i=1}^n \xi_{ji}v_i\Big|^2\Big\},
\end{aligned} \tag{5.10}$$

and our goal is to make this last expression small. To estimate the expectations we use the distribution formula:

$$\begin{aligned}
\mathbb{E}\exp\Big\{-\frac{1}{4t^2}\Big|\sum_{i=1}^n \xi_{ji}v_i\Big|^2\Big\} &= \int_0^1 \mathbb{P}\Big(\exp\Big\{-\frac{1}{4t^2}\Big|\sum_{i=1}^n \xi_{ji}v_i\Big|^2\Big\} > s\Big)ds \\
&= \int_0^\infty ue^{-u^2/2}\,\mathbb{P}\Big(\Big|\sum_{i=1}^n \xi_{ji}v_i\Big| < \sqrt{2}tu\Big)du.
\end{aligned} \tag{5.11}$$

It is now apparent that we need to estimate the quantities

$$f_j(\lambda) := \mathbb{P}\Big(\Big|\sum_{i=1}^n \xi_{ji}v_i\Big| < \lambda\Big), \quad j \le N.$$

To this end, note that for each row $j \in \{1, \dots, N\}$ there exists $\sigma_j \subset \{1, \dots, n\}$ with cardinality $|\sigma_j| \geq a_4 n$ such that $\mathbb{E}\,\xi_{ji}^2 \geq 1$ for all $i \in \sigma_j$ (this is condition (iv)). Also, for each fixed v, set

$$\sigma_v := \{i \mid |v_i| > a\}.$$

Since $v \in S^{n-1}$ we have $|\sigma_v| \leq 1/a^2$.

Set $\overline{\sigma}_j = \sigma_j \setminus \sigma_v$, and note that

$$|\overline{\sigma}_j| \geq a_4 n - \frac{1}{a^2}.$$

It follows that $|\overline{\sigma}_j^c| \leq (1 - a_4)n + \frac{1}{a^2}$, so to have $|\overline{\sigma}_j^c| \leq m$ it suffices to require

$$(1 - a_4)n + \frac{1}{a^2} \leq m. \tag{5.12}$$

Note that (5.12), in particular, implies $1/a^2 \leq a_4 n \leq n$. Recall that $m = \gamma n$, where γ was chosen in (5.8). Then inequality (5.12) is satisfied if $a_4 > 1 - \gamma$ (which is the condition on γ in our Theorem) and

$$t \geq \left(\frac{a_1}{\sqrt{(\gamma + a_4 - 1)n}} \right)^{r-2}. \tag{5.13}$$

Now, since $|\overline{\sigma}_j^c| \leq m$, we have $|P_{\overline{\sigma}_j} v| \geq \rho/2$, and hence

$$A_j^2 := \mathbb{E} \sum_{i \in \overline{\sigma}_j} \xi_{ji}^2 v_i^2 \geq \frac{\rho^2}{4}$$

(where we have used the property $\mathbb{E}\,\xi_{ji}^2 \geq 1$ for $i \in \sigma_j$). Consequently, using Proposition 3.2, and keeping in mind $|v_i| \leq a$ for $i \in \overline{\sigma}_j$, we get

$$\begin{aligned} f_j(\lambda) &\leq c\Big(\frac{\lambda}{\rho} + \frac{\mu^r}{\rho^r}\|P_{\overline{\sigma}_j} v\|_r^r\Big) \\ &\leq c\Big(\frac{\lambda}{\rho} + \frac{\mu^r}{\rho^r}\|P_{\overline{\sigma}_j} v\|_\infty^{r-2} \cdot |P_{\overline{\sigma}_j} v|^2\Big) \\ &\leq c\Big(\frac{\lambda}{\rho} + \frac{\mu^r a^{r-2}}{\rho^r}\Big) \end{aligned}$$

for some absolute constant $c \geq 1$. Then, continuing from (5.11) we have

$$\begin{aligned} \mathbb{E}\exp\Big\{-\frac{1}{4t^2}\Big|\sum_{i=1}^n \xi_{ji} v_i\Big|^2\Big\} &= \int_0^\infty u e^{-u^2/2} f_j(\sqrt{2}tu)\,du \\ &\leq c\int_0^\infty u e^{-u^2/2}\Big(\frac{\sqrt{2}tu}{\rho} + \frac{\mu^r a^{r-2}}{\rho^r}\Big)du \\ &= \frac{c\sqrt{2}t}{\rho}\int_0^\infty u^2 e^{-u^2/2}du + \frac{c\mu^r a^{r-2}}{\rho^r}\int_0^\infty u e^{-u^2/2}du \\ &= \frac{c\sqrt{\pi}t}{\rho} + \frac{c\mu^r t}{\rho^r a_1^{r-2}} = c_3 t, \end{aligned}$$

where

$$c_3 := c\Big(\frac{\sqrt{\pi}}{\rho} + \frac{\mu^r}{\rho^r a_1^{r-2}}\Big). \tag{5.14}$$

Therefore, from (5.10), we get (for each fixed $v \in \mathcal{A}$)

$$\mathbb{P}\Big(|\Gamma v| \le 2t\sqrt{N}\Big) \le e^N (c_3 t)^N = (c_3 e t)^N,$$

and from this, in (5.9) we get

$$\begin{aligned}
\mathbb{P}(E_I \cap \Omega_0) &\le |\mathcal{A}|(c_3 e t)^N \\
&\le \Big(\frac{3}{a}\Big)^n (c_3 e t)^N \\
&= \Big(\frac{3a_1}{t}\Big)^n (c_3 e t)^N.
\end{aligned}$$

Then we can make

$$\mathbb{P}(E_I \cap \Omega_0) \le e^{-N} \tag{5.15}$$

provided that

$$t \le \Big(\frac{1}{c_3 e^2}\Big)\Big(\frac{1}{3a_1 c_3 e^2}\Big)^{1/\delta}. \tag{5.16}$$

Choose t to satisfy equality in (5.16). Note

$$\frac{t}{a_1} \le \frac{1}{c_3 e^2 a_1} \le \frac{\rho}{ce^2\sqrt{\pi}}, \tag{5.17}$$

so the left hand side of (5.2) holds. Finally note that (5.13) is satisfied whenever

$$\delta \ge \frac{\frac{2}{r-2}\ln(3a_1 c_3 e^2)}{\ln\Big(\frac{(\gamma+a_4-1)n}{a_1^2 (c_3 e^2)^{2/(r-2)}}\Big)} \ =: \ \frac{\tilde{c}_1}{\ln(\tilde{c}_2 n)}.$$

To finish, we take probabilities in (5.1) and we use the estimates for $\mathbb{P}(E_C \cap \Omega_0)$ and $\mathbb{P}(E_I \cap \Omega_0)$ we have found in (5.7) and (5.15), respectively, combined with the estimate $\mathbb{P}(\Omega_0^c) \le e^{-a_2 N}$ coming from condition (ii). This shows that, with the chosen t, we have $\mathbb{P}(E) \le e^{-b_2 N/2} + e^{-N} + e^{-a_2 N}$, which completes the proof. ■

Chapter 6

Square random matrices

In this chapter our goal is to prove Theorem 3. We are going to use two lemmas from [39]. The first one is [39, Lemma 3.5]. Note that the proof given there works for any random matrix.

Lemma 6.1. *Let Γ be any random matrix of size $m \times n$. Let $X_1, \ldots, X_n$ denote the columns of Γ and let H_k denote the span of all column vectors except the k-th. Then for every $\gamma, \rho \in (0,1)$ and every $\varepsilon > 0$ one has*

$$\mathbb{P}\Big(\inf_{x \in F} |\Gamma x| \le \varepsilon \rho n^{-1/2}\Big) \le \frac{1}{\gamma n} \sum_{k=1}^{n} \mathbb{P}\big(\mathrm{dist}(X_k, H_k) < \varepsilon\big),$$

where $F = S^{n-1} \cap Incomp(\gamma n, \rho)$.

The next lemma is similar to [39, Lemma 3.8]. To prove it one would repeat the proof of that lemma, replacing [39, Lemma 3.7] used there with our Lemma 3.5. Observe that in applying our Lemma 3.5 we use condition (iv) on the rows of the matrices.

Lemma 6.2. *Let $r \in (2,3]$ and Γ be a random matrix as in Theorem 3. Let $X_1, \ldots, X_n$ denote its column vectors, and let $H_n = \mathrm{span}(X_1, \ldots, X_{n-1})$ be the subspace spanned by $X_1, \ldots, X_{n-1}$. Then for every $\varepsilon \ge 0$ one has*

$$\mathbb{P}\Big(\mathrm{dist}(X_n, H_n) < \varepsilon \quad \text{and} \quad \|\Gamma\| \le a_1 n^{1/2}\Big) \le c(\varepsilon n^{\frac{3-r}{2}} + \mu^r n^{\frac{2-r}{2}}),$$

where c depends only on μ and a_1.

Now we are ready for the proof of Theorem 3.

Proof of Theorem 3. Without loss of generality we assume $\varepsilon \le a_1/2$ (otherwise choose $C = 2/a_1$ and we are done). We also assume that $r \le 3$ (otherwise we pass to $r_0 = \min\{3, r\}$).

Consider the event

$$E := \big\{\omega \;\big|\; \exists x \in S^{n-1} \text{ s.t. } |\Gamma x| \le t n^{-1/2}\big\}.$$

By equivalence (1.2) we are to estimate $\mathbb{P}(E)$ with an appropriate value of t (which will be specified later).

As in the proof of Theorem 2, we split the set E into the sets E_C and E_I defined as follows:

$$E_C = \big\{\omega \;\big|\; \exists x \in S^{n-1} \cap Comp(m,\rho) \;\text{ s.t. }\; |\Gamma x| \le tn^{-1/2}\big\},$$
$$E_I = \big\{\omega \;\big|\; \exists x \in S^{n-1} \cap Incomp(m,\rho) \;\text{ s.t. }\; |\Gamma x| \le tn^{-1/2}\big\}.$$

Define $\Omega_0 := \big\{\omega \;\big|\; \|\Gamma\| \le a_1\sqrt{n}\big\}$. We will estimate $\mathbb{P}(E)$ using the union bound in the inclusion

$$E \subset (E_C \cap \Omega_0) \cup E_I \cup \Omega_0^c. \tag{6.1}$$

Case 1: Probability of $E_C \cap \Omega_0$. The proof of this case is almost a line by line repetition of the corresponding proof in Theorem 2 (see Case 1 there). Let $m \le n$ and $\rho \in (0,1)$, to be specified later. Using approximation argument and the union bound as in the proof of Case 1 in Theorem 2, and choosing

$$\rho := \min\left\{\frac{1}{4}, \frac{b_1}{5a_1}\right\}, \qquad \gamma := \frac{b_2}{4\ln\left(\frac{6e}{\rho b_2}\right)}, \qquad m = \gamma n, \tag{6.2}$$

we obtain

$$\mathbb{P}(E_C \cap \Omega_0) \le e^{-b_2 n/2}, \tag{6.3}$$

provided that

$$\frac{2t}{a_1} \le \rho. \tag{6.4}$$

Case 2: Probability of E_I. We work on the set $Incomp(m,\rho)$, where $m = \gamma n$ and γ, ρ are as chosen in (6.2).

Using Lemma 6.1 with $\varepsilon = t/\rho$, and also applying Lemma 6.2, we get

$$\begin{aligned}
\mathbb{P}(E_I) &\le \frac{1}{\gamma n}\sum_{k=1}^{n} \mathbb{P}\big(\operatorname{dist}(X_k, H_k) < t/\rho\big) \\
&= \frac{1}{\gamma n}\sum_{k=1}^{n} \Big\{\mathbb{P}\big(\operatorname{dist}(X_k, H_k) < t/\rho \quad \& \quad \|\Gamma\| \le a_1\sqrt{n}\big) + \mathbb{P}\big(\|\Gamma\| > a_1\sqrt{n}\big)\Big\} \\
&\le \frac{1}{\gamma n}\sum_{k=1}^{n} \Big\{c(\varepsilon n^{\frac{3-r}{2}} + n^{\frac{2-r}{2}}) + e^{-a_2 n}\Big\} \\
&\le \frac{c}{\gamma}(\varepsilon n^{\frac{3-r}{2}} + n^{\frac{2-r}{2}}) + \frac{1}{\gamma}e^{-a_2 n}.
\end{aligned} \tag{6.5}$$

Also notice that our choice $t = \varepsilon\rho$ and our assumption $\varepsilon \le a_1/2$ guarantee that t satisfies (6.4).

To finish the proof, we take probabilities in (6.1), and we use the estimates for $\mathbb{P}(E_C \cap \Omega_0)$ and for $\mathbb{P}(E_I)$ obtained in (6.3) and (6.5), respectively, combined with the estimate $\mathbb{P}(\Omega_0^c) \le e^{-a_2 N}$ coming from condition (ii). This way we obtain

$$\mathbb{P}(E) \le e^{-b_2 n/2} + \frac{c}{\gamma}(\varepsilon n^{\frac{3-r}{2}} + n^{\frac{2-r}{2}}) + \frac{1}{\gamma}e^{-a_2 n} + e^{-a_2 N} \le C(\varepsilon n^{\frac{3-r}{2}} + n^{\frac{2-r}{2}})$$

for a suitable constant C. ■

Chapter 7

Random matrices with complex entries

In this chapter we consider matrices Γ of size $N \times n$ (where $N \geq n$) whose entries are complex-valued random variables subject to certain conditions. We will extend Theorem 1 to these kinds of matrices.

We will use the notation $\sqrt{-1}$ for the imaginary unit, which is a bit unusual but will help to avoid confusion as we are using the label i for positive integers.

A complex-valued random variable is a random variable ξ of the form

$$\xi = \mathrm{Re}(\xi) + \mathrm{Im}(\xi)\sqrt{-1},$$

where $\mathrm{Re}(\xi)$ and $\mathrm{Im}(\xi)$ are independent real-valued random variables. Such a random variable ξ is said to be *centered* if its real and imaginary parts are centered, that is, $\mathbb{E}(\mathrm{Re}(\xi)) = \mathbb{E}(\mathrm{Im}(\xi)) = 0$.

The conditions are the same as in the case of real-valued entries: For parameters $r > 2$, $\mu \geq 1$, $a_1 > 0$, $a_2 > 0$, $a_3 \in (0, \mu)$, and $a_4 \in (0, 1]$, we will consider $N \times n$ random matrices $\Gamma = (\xi_{ji})_{j \leq N, i \leq n}$ whose entries are *independent* complex-valued *centered* random variables satisfying the following conditions:

(i) Moments: $\quad \mathbb{E}\,|\xi_{ji}|^r \leq \mu^r \quad$ for all j and i.

(ii) Norm: $\quad \mathbb{P}\Big(\|\Gamma\| > a_1\sqrt{N}\Big) \leq e^{-a_2 N}$.

(iii) Columns: $\quad \mathbb{E}\|(\xi_{ji})_{j=1}^N\|_2^2 = \sum_{j=1}^N \mathbb{E}|\xi_{ji}|^2 \geq a_3^2 N \quad$ for each i.

As in the real case, the parameters μ, r, a_1, a_2, a_3, a_4 should be regarded as constants which do not depend on the dimensions n, N.

Remark. Observe that condition (i) implies bounds for the r-th moment of the real and imaginary parts of the entries. That is, if ξ is a complex-valued random variable such that $\mathbb{E}|\xi|^r \leq \mu^r$, then $\mathbb{E}|\mathrm{Re}(\xi)|^r \leq \mu^r$ and $\mathbb{E}|\mathrm{Im}(\xi)|^r \leq \mu^r$.

We will write $S_{\mathbb{C}}^{n-1}$ to denote the unit sphere of $\mathbb{C}^n$, that is, the set of all $z = (z_1, \ldots, z_n) \in \mathbb{C}^n$ such that $\sum |z_i|^2 = 1$.

In this chapter we verify the following result.

Theorem 7.1. *Let $r > 2$, $\mu \geq 1$, $a_1, a_2, a_3 > 0$ with $a_3 < \mu$. Let $1 \leq n < N$ be integers, and write N in the form $N = (1+\delta)n$. Suppose Γ is an $N \times n$ matrix whose entries are independent centered complex-valued random variables such that conditions (i), (ii) and (iii) are satisfied. There exist positive constants c_1, c_2 and δ_0 (depending only on the parameters r, μ, a_1, a_2, a_3) such that whenever $\delta \geq \delta_0$, then*

$$\mathbb{P}\Big(s_n(\Gamma) \leq c_1\sqrt{N}\Big) \leq e^{-c_2 N}.$$

We start with the analogue of Proposition 4.1 for complex entries.

Proposition 7.2. *Let $1 \leq n < N$ be positive integers. Suppose Γ is a matrix of size $N \times n$ whose entries are independent centered complex-valued random variables satisfying conditions (i), (ii) and (iii) for some $2 < r \leq 3$, $\mu \geq 1$ and $a_1, a_2, a_3 > 0$ with $a_3 < \mu$. Then for every $z \in S_{\mathbb{C}}^{n-1}$ we have*

$$\mathbb{P}\Big(|\Gamma z| \leq b_1\sqrt{N}\Big) \leq e^{-b_2 N},$$

where $b_1, b_2 > 0$ depend only on μ, a_3 and r.

Remark. In fact, as in the real case, our proof gives that

$$b_1^2 = \frac{a_3^4}{2^5\mu^2}\Big(\frac{a_3^2}{2^5\mu^2}\Big)^{r/(r-2)}, \qquad b_2 = \frac{a_3^2}{2^3\mu^2}\Big(\frac{a_3^2}{2^5\mu^2}\Big)^{r/(r-2)}.$$

We present now the proof of Theorem 7.1. The proof of Proposition 7.2 will come right after.

Proof of Theorem 7.1. We may assume without loss of generality that $r \leq 3$ (otherwise pass to $r_0 = \min\{3, r\}$).

Let $t \geq 0$ and $\Omega_0 := \big\{\omega \;\big|\; \|\Gamma\| \leq a_1\sqrt{N}\big\}$. By (1.2) it is enough to estimate the probability of the event

$$E := \big\{\omega \;\big|\; \exists z \in S_{\mathbb{C}}^{n-1} \text{ s.t. } |\Gamma z| \leq t\sqrt{N}\big\}.$$

To this end we use the inclusion $E \subset (E \cap \Omega_0) \cup \Omega_0^c$ and the union bound.

To estimate $\mathbb{P}(E \cap \Omega_0)$, let $0 < \varepsilon \leq 1$, and let $\mathcal{N}$ be an ε-net of $S_{\mathbb{C}}^{n-1}$ with cardinality $|\mathcal{N}| \leq (3/\varepsilon)^{2n}$ (this estimate follows from the usual argument in the real case $\mathbb{R}^{2n}$). For any $z \in S_{\mathbb{C}}^{n-1}$ we can find $y \in \mathcal{N}$ such that $|z - y| \leq \varepsilon$. If further z satisfies $|\Gamma z| \leq t\sqrt{N}$, then the corresponding y satisfies

$$|\Gamma y| \leq |\Gamma z| + \|\Gamma\| \cdot |y - z| \leq t\sqrt{N} + \varepsilon a_1\sqrt{N} = (t + \varepsilon a_1)\sqrt{N}. \tag{7.1}$$

Taking $\varepsilon = \min\{t/a_1,\, 1\}$, we see that for each $z \in S_{\mathbb{C}}^{n-1}$ satisfying $|\Gamma z| \le t\sqrt{N}$ there is a corresponding $y \in \mathcal{N}$ such that $|z - y| \le \varepsilon$ and $|\Gamma y| \le 2t\sqrt{N}$. Hence, using the union bound and setting $t = b_1/2$ with b_1 as in Proposition 7.2, one has

$$\begin{aligned}\mathbb{P}(E \cap \Omega_0) &\le \sum_{y \in \mathcal{N}} \mathbb{P}\Big(|\Gamma y| \le 2t\sqrt{N}\Big) \\ &\le |\mathcal{N}| e^{-b_2 N} \le \Big(\frac{3}{\varepsilon}\Big)^{2n} e^{-b_2 N},\end{aligned}$$

where b_1 and b_2 are as in Proposition 7.2. Thus

$$\mathbb{P}(E \cap \Omega_0) \le \exp\Big(-\frac{b_2 N}{2}\Big)$$

as long as

$$\Big(\frac{3}{\varepsilon}\Big)^{2n} \le \exp\Big(\frac{b_2 N}{2}\Big).$$

Bearing in mind that $N = (1+\delta)n$, we can see that the last condition is satisfied if

$$\delta \ge \delta_0 := \max\left\{\frac{4}{b_2} \ln\Big(\frac{6a_1}{b_1}\Big),\ \frac{4}{b_2} \ln 3\right\}. \tag{7.2}$$

To finish, we use $\mathbb{P}(E) \le \mathbb{P}(E \cap \Omega_0) + \mathbb{P}(\Omega_0^c)$ with the estimate for $\mathbb{P}(E \cap \Omega_0)$ just obtained and the estimate $\mathbb{P}(\Omega_0^c) \le e^{-a_2 N}$ coming from condition (ii). ■

It remains to see the proof of Proposition 7.2. We will use the next lemma, which is a variant of Lemma 3.1 for the case of complex random variables. Its proof is an easy modification of the proof of Lemma 3.1.

Lemma 7.3. *Let $2 < r \le 3$ and $\mu \ge 1$. Suppose $\xi_1, \dots, \xi_n$ are independent centered complex-valued random variables such that $\mathbb{E}|\xi_i|^r \le \mu^r$ for every $i = 1, \dots, n$. Let $z = (z_i) \in S_{\mathbb{C}}^{n-1}$. Then for every $\lambda \ge 0$*

$$\mathbb{P}\Big(\Big|\sum_{i=1}^n \xi_i x_i\Big| > \lambda\Big) \ge \Big(\frac{[\mathbb{E}\sum_{i=1}^n |\xi_i|^2 |x_i|^2 - \lambda^2]_+}{8\mu^2}\Big)^{r/(r-2)}.$$

Proof of Proposition 7.2. Take an arbitrary $z = (z_1, \dots, z_n) \in S_{\mathbb{C}}^{n-1}$. For $a > 0$ (a parameter whose value will be specified later), define a set of "good" rows as follows:

$$J = J(a) = \Big\{ j \in \{1, \dots, N\} \ \Big| \ \mathbb{E}\sum_{i=1}^n |\xi_{ji}|^2 |z_i|^2 \ge a \Big\}.$$

Suppose that the cardinality of set J is $|J| = \alpha N$ for some $\alpha \in [0,1]$. Note that for each index $j = 1, \dots, N$ we have

$$\mathbb{E}\sum_{i=1}^n |\xi_{ji}|^2 |z_i|^2 \le \max_{1 \le i \le n} \mathbb{E}|\xi_{ji}|^2 \le \max_{1 \le i \le n} (\mathbb{E}|\xi_{ji}|^r)^{2/r} \le \mu^2.$$

Then on one side we have

$$\sum_{j=1}^{N}\Big(\mathbb{E}\sum_{i=1}^{n}|\xi_{ji}|^2|z_i|^2\Big) = \sum_{j\in J}\Big(\mathbb{E}\sum_{i=1}^{n}|\xi_{ji}|^2|z_i|^2\Big) + \sum_{j\in J^c}\Big(\mathbb{E}\sum_{i=1}^{n}|\xi_{ji}|^2|z_i|^2\Big)$$
$$\leq \mu^2\alpha N + a(1-\alpha)N,$$

while on the other hand, using condition (iii),

$$\sum_{j=1}^{N}\Big(\mathbb{E}\sum_{i=1}^{n}|\xi_{ji}|^2|z_i|^2\Big) = \sum_{i=1}^{n}\Big(\mathbb{E}\sum_{j=1}^{N}|\xi_{ji}|^2\Big)|z_i|^2 \geq \sum_{i=1}^{n} a_3^2 N|z_i|^2 = a_3^2 N.$$

Hence we have $\mu^2\alpha N + a(1-\alpha)N \geq a_3^2 N$, so α satisfies

$$\alpha \geq \frac{a_3^2 - a}{\mu^2 - a}. \tag{7.3}$$

Note that for each $j = 1,\dots,N$, the j-th entry of Γz is $(\Gamma z)_j = \sum_{i=1}^{n}\xi_{ji}z_i$. Define $f_j := \big|\sum_{i=1}^{n}\xi_{ji}z_i\big|$, so

$$|\Gamma z|^2 = \sum_{j=1}^{N} f_j^2.$$

Clearly $f_1,\dots,f_N$ are independent. For any $t,\tau > 0$ we have

$$\begin{aligned}
\mathbb{P}\big(|\Gamma z|^2 \leq t^2 N\big) &= \mathbb{P}\Big(\sum_{j=1}^{N} f_j^2 \leq t^2 N\Big)\\
&= \mathbb{P}\Big(\tau N - \frac{\tau}{t^2}\sum_{j=1}^{N} f_j^2 \geq 0\Big)\\
&\leq \mathbb{E}\exp\Big(\tau N - \frac{\tau}{t^2}\sum_{j=1}^{N} f_j^2\Big)\\
&= e^{\tau N}\prod_{j=1}^{N}\mathbb{E}\exp\Big(-\frac{\tau f_j^2}{t^2}\Big).
\end{aligned} \tag{7.4}$$

From Lemma 7.3 we know that for every $j = 1,\dots,N$,

$$\mathbb{P}(f_j > \lambda) \geq \left(\frac{[\mathbb{E}\sum_{i=1}^{n}|\xi_{ji}|^2|z_i|^2 - \lambda^2]_+}{8\mu^2}\right)^{r/(r-2)} =: \beta_j(r). \tag{7.5}$$

Note that for every $j \in J$ one has

$$\beta_j \geq \left(\frac{[a-\lambda^2]_+}{8\mu^2}\right)^{r/(r-2)}. \tag{7.6}$$

For arbitrary $t > 0$, $\eta > 0$ and $\lambda > 0$, set $\tau := \frac{\eta t^2}{\lambda^2}$. For each $j = 1,\dots,N$ we have

$$\begin{aligned}
\mathbb{E}\exp\Big(-\frac{\tau f_j^2}{t^2}\Big) &= \int_0^1 \mathbb{P}\Big(\exp\Big(-\frac{\eta f_j^2}{\lambda^2}\Big) > s\Big)ds\\
&= \int_0^{e^{-\eta}} \mathbb{P}\Big(\exp\Big(\frac{\eta f_j^2}{\lambda^2}\Big) < \frac{1}{s}\Big)ds + \int_{e^{-\eta}}^1 \mathbb{P}\Big(\exp\Big(\frac{\eta f_j^2}{\lambda^2}\Big) < \frac{1}{s}\Big)ds\\
&\leq e^{-\eta} + \mathbb{P}(f_j < \lambda)(1 - e^{-\eta}).
\end{aligned}$$

Choosing $\eta = \ln 2$ and applying (7.5), we obtain

$$\mathbb{E}\exp\Big(-\frac{\tau f_j^2}{t^2}\Big) \le e^{-\eta} + (1-\beta_j(r))(1-e^{-\eta}) = 1 - \frac{\beta_j(r)}{2} \le \exp\left(-\frac{\beta_j(r)}{2}\right).$$

Since $\tau < \frac{t^2}{\lambda^2}$, inequality (7.4) implies

$$\mathbb{P}\big(|\Gamma x|^2 \le t^2 N\big) \le e^{\tau N}\prod_{j=1}^{N} e^{-\beta_j(r)/2} \le e^{(t^2/\lambda^2)N}\prod_{j\in J} e^{-\beta_j(r)/2}. \tag{7.7}$$

Taking $a = a_3^2/2$ and $\lambda = a_3/2$ and using (7.6) we observe that for every $j \in J$ we have $\beta_j \ge \big(\frac{a_3^2}{32\mu^2}\big)^{r/(r-2)}$. Also note this choice of a and (7.3) imply $\alpha \ge a_3^2/(2\mu^2)$. Now let

$$t^2 := \frac{a_3^4}{2^5\mu^2}\Big(\frac{a_3^2}{2^5\mu^2}\Big)^{r/(r-2)}.$$

Then continuing from (7.7) we obtain

$$\mathbb{P}\left(|\Gamma z|^2 \le \frac{a_3^4}{2^5\mu^2}\Big(\frac{a_3^2}{2^5\mu^2}\Big)^{r/(r-2)} N\right) \le \exp\Big\{-\frac{a_3^2}{2^3\mu^2}\Big(\frac{a_3^2}{2^5\mu^2}\Big)^{r/(r-2)} N\Big\}.$$

This completes the proof. ■

Chapter 8

A remark about a recent result of Srivastava and Vershynin

In [42] the following proposition is proved: Let $p \geq 1$, and consider a random vector $X = (\xi_1, \dots, \xi_n)$, where ξ_i are independent random variables with zero means, *unit variances* and with uniformly bounded $2p$-moments. Then for every k with $1 \leq k \leq n$ and for every orthogonal projection $P : \mathbb{R}^n \to \mathbb{R}^n$ with $\operatorname{rank}(P) = k$, one has

$$\mathbb{E}\big||PX|^2 - k\big|^p \leq ck^{p/2} \tag{$\star$}$$

In the estimate ($\star$), the constant $c > 0$ depends only on p and on the uniform bound for the $2p$-moments. Let us write μ for such bound, so $\|\xi\|_{L_{2p}} \leq \mu$.

In this chapter we aim to prove a corresponding estimate without imposing the condition of *unit variances* on all the random variables $\xi_1, \dots, \xi_n$.

Proposition 8.1. *Let $p \geq 2$, and consider a random vector $X = (\xi_1, \dots, \xi_n)$, where ξ_i are independent random variables with zero means and with uniformly bounded $2p$-moments, say $\|\xi_i\|_{L_{2p}} \leq \mu$. Then for every k with $1 \leq k \leq n$ and for every orthogonal projection $P : \mathbb{R}^n \to \mathbb{R}^n$ with $\operatorname{rank}(P) = k$, one has*

$$\mathbb{E}\big|\, |PX|^2 - \mathbb{E}|PX|^2 \,\big|^p \leq (cp\mu^2\sqrt{k})^p,$$

where $c > 0$ is an absolute constant.

Proof. Let $P = (P_{ij})$ denote the $n \times n$ matrix of the operator P. For any vector $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ we have

$$|Px|^2 = \langle x, Px\rangle = \sum_{i,j=1}^{n} x_i x_j P_{ij}.$$

In particular this is the case for the random vector $X = (\xi_1, \dots, \xi_n)$, so

$$|PX|^2 = \langle X, PX\rangle = \sum_{i,j=1}^{n} \xi_i \xi_j P_{ij} = \sum_{i=1}^{n} \xi_i^2 P_{ii} + \sum_{i \neq j}^{n} \xi_i \xi_j P_{ij},$$

and clearly the contribution to this sum coming from the diagonal of P is the random variable

$$\Delta := \sum_{i=1}^{n} \xi_i^2 P_{ii}.$$

Write $P = D + P_0$, where D is the $n \times n$ diagonal matrix consisting of the diagonal elements of P, and $P_0 := P - D$. Thus

$$|PX|^2 = \langle X, DX \rangle + \langle X, P_0 X \rangle = \Delta + \langle X, P_0 X \rangle. \tag{8.1}$$

To estimate the term $\langle X, P_0 X \rangle$ we use a 'decoupling inequality' from [11]. Let X' be an independent copy of X, and write $\mathbb{E}_X$ and $\mathbb{E}_{X'}$ for the expectations corresponding to X and X', respectively. The inequality [11, Theorem 3.1.1] applied to our setting gives

$$\mathbb{E}\big|\langle X, P_0 X \rangle\big|^p \le 12^p \, \mathbb{E}_{X'} \, \mathbb{E}_X \big|\langle X, P_0 X' \rangle\big|^p. \tag{8.2}$$

For any $a \in \mathbb{R}^n$, using the inequality $\|\cdot\|_{L_p} \le \|\cdot\|_{L_{2p}}$, and using the standard argument of symmetrization and Khintchine's inequality, we obtain

$$\begin{aligned}\Big(\mathbb{E}\Big|\sum_{i=1}^n a_i \xi_i\Big|^p\Big)^2 &\le \mathbb{E}\Big|\sum_{i=1}^n a_i \xi_i\Big|^{2p} \le (2B_{2p})^{2p} \, \mathbb{E}\Big(\sum_{i=1}^n a_i^2 \xi_i^2\Big)^p \\ &\le \big(2\sqrt{2p}\big)^{2p} \, \mathbb{E}\Big(\sum_{i=1}^n a_i^2 \xi_i^2\Big)^p,\end{aligned} \tag{8.3}$$

where B_p denotes the constant (depending only on p) of Khintchine's inequality. If $|a| = 1$, the last expression is estimated as in the proof of our Lemma 3.1 (p. 17) with p replacing $r/2$ (see also [29, Lemma 3.1] and [26, Lemma 3.6]). This way,

$$\mathbb{E}\Big(\sum_{i=1}^n a_i^2 \xi_i^2\Big)^p \le \sup_{i \ge 1} \mathbb{E}|\xi_i|^{2p} \le \mu^{2p}.$$

This immediately implies that for any $a \in \mathbb{R}^n$,

$$\mathbb{E}\Big(\sum_{i=1}^n a_i^2 \xi_i^2\Big)^p \le |a|^{2p} \mu^{2p}.$$

Using this estimate in (8.3) we get

$$\mathbb{E}\Big|\sum_{i=1}^n a_i \xi_i\Big|^p \le \big(2\sqrt{2p}\big)^p |a|^p \mu^p = \big(2\sqrt{2p}\,\mu\big)^p |a|^p. \tag{8.4}$$

From (8.2) and (8.4), and recalling that X' is an independent copy of X, it follows that

$$\mathbb{E}\big|\langle X, P_0 X \rangle\big|^p \le \big(24\sqrt{2p}\,\mu\big)^p \, \mathbb{E}_{X'} |P_0 X'|^p = \big(24\sqrt{2p}\,\mu\big)^p \, \mathbb{E}_X |P_0 X|^p. \tag{8.5}$$

Using the triangle inequality (and noticing that $P_{ii}^2 \le P_{ii}$, since $0 \le P_{ii} \le \|P\| \le 1$), we have

$$\begin{aligned}|P_0 X| = |PX - DX| &\le |PX| + |DX| \\ &= |PX| + \Big(\sum_{i=1}^n \xi_i^2 P_{ii}^2\Big)^{1/2} \\ &\le |PX| + \Delta^{1/2} \\ &\le \sqrt{2}\,\big(|PX|^2 + \Delta\big)^{1/2}.\end{aligned}$$

Taking expectations and using Hölder inequality it follows that

$$\mathbb{E}|P_0X|^p \le \sqrt{2}^p \left(\mathbb{E}\big||PX|^2+\Delta\big|^p\right)^{1/2}. \tag{8.6}$$

From (8.1), (8.5) and (8.6) we get

$$\begin{aligned}\mathbb{E}\big||PX|^2-\Delta\big|^p &= \mathbb{E}\big|\langle X, P_0X\rangle\big|^p \\ &\le \left(48\sqrt{p}\,\mu\right)^p\left(\mathbb{E}\big||PX|^2+\Delta\big|^p\right)^{1/2}.\end{aligned}$$

This implies that the random variable $Z := |PX|^2 - \Delta$ satisfies

$$\|Z\|_{L_p}^2 \le \left(48\sqrt{p}\,\mu\right)^2 \|Z+2\Delta\|_{L_p} \le \left(48\sqrt{p}\,\mu\right)^2 \left(\|Z\|_{L_p} + 2\|\Delta\|_{L_p}\right).$$

Solving this quadratic inequality we get

$$\|Z\|_{L_p} \le \left(48\sqrt{p}\,\mu\right)^2 + \sqrt{2}\left(48\sqrt{p}\,\mu\right)\|\Delta\|_{L_p}^{1/2}. \tag{8.7}$$

Next we bound $\|\Delta\|_{L_p}$. Recall that $k = \mathrm{rank}(P)$.

$$\begin{aligned}\big\|\Delta\big\|_{L_p} &= \left\|\sum_{i=1}^n \left[\xi_i^2 P_{ii} - \mathbb{E}(\xi_i^2)P_{ii} + \mathbb{E}(\xi_i^2)P_{ii}\right]\right\|_{L_p} \\ &\le \left\|\sum_{i=1}^n \left[\xi_i^2 - \mathbb{E}(\xi_i^2)\right]P_{ii}\right\|_{L_p} + \left\|\sum_{i=1}^n \mathbb{E}(\xi_i^2)P_{ii}\right\|_{L_p}.\end{aligned} \tag{8.8}$$

Observe that the random variables $\zeta_i := \xi_i^2 - \mathbb{E}(\xi_i^2)$ appearing in the first term of (8.8) are centered and satisfy

$$\|\zeta_i\|_{L_p} \le \|\xi_i^2\|_{L_p} + \mathbb{E}(\xi_i^2) \le \|\xi_i\|_{L_{2p}}^2 + [\mathbb{E}(\xi_i^{2p})]^{1/p} \le 2\mu^2.$$

Then, using the symmetrization inequality [24, Lemma 6.3], Khintchine's inequality, and $P_{ii}^2 \le P_{ii}$, we have

$$\begin{aligned}\left\|\sum_{i=1}^n \left[\xi_i^2 - \mathbb{E}(\xi_i^2)\right]P_{ii}\right\|_{L_p}^p &= \left\|\sum_{i=1}^n \zeta_i P_{ii}\right\|_{L_p}^p \\ &\le (2B_p)^p\ \mathbb{E}\Big[\Big(\sum_{i=1}^n \zeta_i^2 P_{ii}^2\Big)^{p/2}\Big].\end{aligned}$$

The expression $\mathbb{E}\Big(\sum \zeta_i^2 P_{ii}\Big)^{p/2}$ is estimated repeating the convexity argument from the proof of our Lemma 3.1 (p. 17). This way,

$$\begin{aligned}\mathbb{E}\Big(\sum_{i=1}^n \zeta_i^2 P_{ii}\Big)^{p/2} &\le k^{p/2} \sup_{i\ge 1} \mathbb{E}|\zeta_i^2|^{p/2} \\ &= k^{p/2} \sup_{i\ge 1} (\|\zeta_i\|_p)^p \le k^{p/2}(2\mu^2)^p,\end{aligned}$$

and taking into account that $B_p \le \sqrt{p}$, we obtain

$$\Big\| \sum_{i=1}^n \big[\xi_i^2 - \mathbb{E}(\xi_i^2)\big] P_{ii} \Big\|_{L_p} \le \big(4\sqrt{p}\,\mu^2\big) k^{1/2}. \tag{8.9}$$

For the second term of (8.8), we clearly have

$$\Big\| \sum_{i=1}^n \mathbb{E}(\xi_i^2) P_{ii} \Big\|_{L_p} = \Big| \sum_{i=1}^n \mathbb{E}(\xi_i^2) P_{ii} \Big| \le \big(\max_j \mathbb{E}\,\xi_j^2\big) \Big| \sum_{i=1}^n P_{ii} \Big| \le \mu^2\, k. \tag{8.10}$$

From (8.8), (8.9) and (8.10) we obtain

$$\big\|\Delta\big\|_{L_p} \le \big(4\sqrt{p}\,\mu^2\big) k^{1/2} + \mu^2 k \le \big(5\sqrt{p}\,\mu^2\big) k.$$

Substituting this into (8.7) we get

$$\|Z\|_{L_p} \le \big(48\sqrt{p}\,\mu\big)^2 + \sqrt{2}\big(48\sqrt{p}\,\mu\big)\big(\sqrt{5}p^{1/4}\,\mu\big) k^{1/2} \le c_1 p \mu^2 \sqrt{k}, \tag{8.11}$$

where $c_1 = 48(48 + \sqrt{10})$.

Finally, recall that $Z := |PX|^2 - \Delta$, so using the triangle inequality we have

$$\begin{aligned} \big\| |PX|^2 - \mathbb{E}|PX|^2 \big\|_{L_p} &\le \big\| |PX|^2 - \Delta \big\|_{L_p} + \big\| \Delta - \mathbb{E}|PX|^2 \big\|_{L_p} \\ &= \big\| Z \big\|_{L_p} + \big\| \Delta - \mathbb{E}|PX|^2 \big\|_{L_p}. \end{aligned} \tag{8.12}$$

By (8.9) and (8.11) we estimate each term of (8.12), thus obtaining

$$\big\| |PX|^2 - \mathbb{E}|PX|^2 \big\|_{L_p} \le c_1 p \mu^2 \sqrt{k} + (4\sqrt{p}\,\mu^2)\sqrt{k} \le c p \mu^2 \sqrt{k}.$$

This finishes the proof with $c = c_1 + 4$. ■

Corollary 8.2. *Let $p \ge 2$, and consider a random vector $X = (\xi_1, \dots, \xi_n)$, where ξ_i are independent random variables with zero means and with uniformly bounded $2p$-moments, say $\|\xi_i\|_{L_{2p}} \le \mu$. Let $P : \mathbb{R}^n \to \mathbb{R}^n$ be an orthogonal projection with* $\operatorname{rank}(P) = k$. *Then for every $t > 4p^2\mu^4 k$ one has*

$$\mathbb{P}\big(|PX|^2 > t\big) \le c^p t^{-p/2},$$

where $c > 0$ is the absolute constant appearing in our Proposition 8.1.

Proof. Observe that

$$\begin{aligned} \mathbb{E}|PX|^2 &= \mathbb{E}\langle X, PX \rangle \\ &= \mathbb{E} \sum_{i,j=1}^n \xi_i \xi_j P_{ij} = \sum_{j=1}^n \mathbb{E}\,\xi_j^2 P_{ij} \le \big(\max_j \mathbb{E}\,\xi_j^2\big) \sum_{i=1}^n P_{ii} \le \mu^2 k. \end{aligned}$$

Thus, by Chebyshev inequality, under the conditions of Proposition 8.1, for any $t > 4p^2\mu^4 k$ (which in particular implies $t > 2\mu^2 k$, so $t > \mathbb{E}|PX|^2 + t/2$) one has

$$\mathbb{P}\big(|PX|^2 > t\big) \le \mathbb{P}\Big(\big| |PX|^2 - \mathbb{E}|PX|^2 \big| > t/2\Big) \le \frac{2^p \cdot (cp\mu^2\sqrt{k})^p}{t^p} \le c^p t^{-p/2}.$$

■

Bibliography

[1] R. Adamczak, O. Guédon, A. E. Litvak, A. Pajor, N. Tomczak-Jaegermann, Condition number of a square matrix with i.i.d. columns drawn from a convex body. *Proc. Amer. Math. Soc.* **140** (2012), 987–998.

[2] R. Adamczak, A. E. Litvak, A. Pajor and N. Tomczak-Jaegermann, Quantitative estimates of the convergence of the empirical covariance matrix in log-concave Ensembles. *J. Amer. Math. Soc.* **234** (2010), 535–561.

[3] R. Adamczak, A. E. Litvak, A. Pajor and N. Tomczak-Jaegermann, Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling. *Constr. Approx.* **34** (2011), 61–88

[4] G.W. Anderson, A. Guionnet, O. Zeitouni, *An Introduction to Random Matrices*. Cambridge University Press, Cambridge, 2010.

[5] Z.D. Bai and J.W. Silverstein, *Spectral analysis of large dimensional random matrices*, 2nd ed. Springer Series in Statistics. Dordrecht: Springer, 2010

[6] G. Bennett, L.E. Dor, V. Goodman, W.B. Johnson, and C.M. Newman, On uncomplemented subspaces of L_p, $1 < p < 2$. *Isr. J. Math.* **26** (1977), 178–187.

[7] G. Blower, *Random Matrices: High Dimensional Phenomena*. Cambridge University Press, Cambridge, 2009.

[8] V.V. Buldygin and Yu.V. Kozachenko, Sub-Gaussian random variables. *Ukrainian Math. J.* **32** (1980), 483 - 489.

[9] Y.S. Chow, Some Convergence Theorems for Independent Random Variables. *Ann. Math. Statist.* **37** (1966), 1482 - 1493.

[10] K.R. Davidson and S.J. Szarek, Local operator Theory, Random Matrices and Banach spaces. In: W. B. Johnson, J. Lindenstrauss (Eds.), *Handbook of the Geometry of Banach Spaces, Vol. I*, pp. 317–366. Elsevier, Amsterdam, 2001.

[11] V.H. de la Peña and E. Giné, *Decoupling. From Dependence to Independence*. Springer-Verlag, New York, 1999.

[12] A. Edelman, Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.* **9** (1988), 543–560.

[13] L. Erdös, Universality of Wigner random matrices: a Survey of Recent Results. *Russ. Math. Surv.* **66** (2011), 507–626.

[14] W. Feller, *An Introduction to Probability Theory and its Applications, Vol. II.* 2nd ed. John Wiley & Sons, New York, 1971.

[15] O. Feldheim, S. Sodin, A universality result for the smallest eigenvalues of certain sample covariance matrices. *Geom. Func. Anal.* **20** (2010), 88–123.

[16] V.L. Girko, The strong circular law. Twenty years later. I. *Random Oper. Stochastic Equations*, **12** (2004), 49–104.

[17] Y. Gordon, Some inequalities for Gaussian processes and applications. *Israel J. Math.* **50** (1985), 265–289.

[18] F. Götze, A.N. Tikhomirov, The circular law for random matrices. *Ann. Probab.* **38** (2010), 1444-1491.

[19] J.P. Kahane, Propriétés locales des fonctions à séries de Fourier aléatoires. *Stud. Math.* **19** (1960), 1 - 25.

[20] J.P. Kahane, *Some random series of functions.* 2nd ed. Cambridge University Press, London, 1985.

[21] B. S. Kashin, Diameters of some finite-dimensional sets and classes of smooth functions. *Izv. Akad. Nauk SSSR*, Ser. Mat. **41** (1977), 334351.

[22] M.A. Krasnosel'skii and J.B. Rutickii, *Convex Functions and Orlicz Spaces.* Noordhoff, Groningen, 1961.

[23] R. Latała, Some estimates of norms of random matrices. *Proc. Amer. Math. Soc.* **133** (2005), 1273–1282.

[24] M. Ledoux and M. Talagrand, *Probability in Banach Spaces.* Springer-Verlag, Berlin, 1991.

[25] J. Lindenstrauss and L. Tzafriri, *Classical Banach Spaces I and II.* Reprint of the 1977/1979 editions. Springer Verlag, Berlin - Heidelberg, 1996.

[26] A.E. Litvak, A. Pajor, M. Rudelson and N. Tomczak-Jaegermann, Smallest singular value of random matrices and geometry of random polytopes. *Adv. Math.* **195** (2005), 491–523.

[27] A.E. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, R. Vershynin, Random Euclidean embeddings in spaces of bounded volume ratio. *C.R. Acad. Sci. Paris*, Ser 1, Math., **339** (2004), 33–38.

[28] A.E. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, R. Vershynin, Euclidean embeddings in spaces of finite volume ratio via random matrices. *J. Reine Angew. Math.* **589** (2005), 1–19.

[29] A.E. Litvak, O.D. Rivasplata, Smallest singular value of sparse random matrices. Preprint, arXiv:1106.0938

[30] M. Loève, *Probability Theory I.* 4th ed. Springer-Verlag, Berlin, 1977.

[31] M.B. Marcus and G. Pisier, *Random Fourier series with applications to harmonic analysis.* Princeton University Press, Princeton, 1981.

[32] M.L. Mehta, *Random Matrices.* Third Ed. Elsevier Academic Press, 2004.

[33] V.D. Milman and G. Schechtman, *Asymptotic Theory of Finite Dimensional Normed Spaces: Isoperimetric Inequalities in Riemannian Manifolds.* Lecture Notes in Mathematics 1200, Springer-Verlag, Berlin, 1986. Corrected Second Printing 2001.

[34] V.V. Petrov, *Limit Theorems of Probability Theory.* Clarendon Press, Oxford, 1995.

[35] G. Pisier, Probabilistic Methods in the Geometry of Banach Spaces. In: G. Letta, M. Pratelli, *Probability and Analysis.* Lecture Notes in Mathematics 1206, pp. 167 - 241. Springer-Verlag, Berlin, 1986.

[36] M.M. Rao, *Theory of Orlicz Spaces.* M. Dekker, New York, 1991.

[37] M. Rudelson, *Probabilistic and Combinatorial Methods in Analysis.* Unpublished lecture notes, CBMS Regional Meeting, Kent, 2006.

[38] M. Rudelson, Invertibility of random matrices: norm of the inverse. *Ann. of Math.* **168** (2008), 575–600.

[39] M. Rudelson and R. Vershynin, The Littlewood-Offord problem and invertibility of random matrices. *Adv. Math.* **218** (2008), 600–633.

[40] M. Rudelson and R. Vershynin, The smallest singular value of a random rectangular matrix. *Comm. Pure Appl. Math.* **62** (2009), 1707–1739.

[41] G. Schechtman, Special orthogonal splittings of L_1^{2k}. *Isr. J. Math.* **139** (2004), 337-347.

[42] N. Srivastava and R. Vershynin, Covariance estimation for distributions with $2+\varepsilon$ moments. Preprint.

[43] K.R. Stromberg, *Probability for Analysts.* Chapman & Hall/CRC, New York, 1994.

[44] D.W. Stroock, *Probability Theory, An analytic view.* 2nd. ed. Cambridge University Press, Cambridge, 2011.

[45] S.J. Szarek, Condition numbers of random matrices. *J. Complexity* **7** (1991), 131–149. ISSN 0885-064X

[46] T. Tao and V. Vu, On the singularity probability of random Bernoulli matrices. *J. Amer. Math. Soc.* **20** (2007), 603–628.

[47] T. Tao and V. Vu, Random matrices: the circular law. *Commun. Contemp. Math.* **10** (2008), 261-307.

[48] T. Tao and V. Vu, Inverse Littlewood-Offord theorems and the condition number of random discrete matrices. *Ann. of Math.* **169** (2009), 595–632.

[49] T. Tao and V. Vu, Random Matrices: The distribution of the smallest singular values. *Geom. Func. Anal.* **20** (2010), 260–297.

[50] R. Vershynin, Spectral norm of products of random and deterministic matrices. *Probab. Theory Related Fields* **150** (2011), 471–509.