Smallest singular value of sparse random matrices

Alexander E. Litvak[1]   Omar Rivasplata

Abstract

We extend probability estimates on the smallest singular value of random matrices with independent entries 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 of the entries. We also show that it is enough to require boundedness from above of the r-th moment, $r > 2$, of the corresponding entries.

AMS 2010 Classification: 46B06, 60B20, 15B52
Keywords: Random matrices, sparse matrices, singular numbers, invertibility of random matrices, subgaussian random variables, compressible and incompressible vectors, deviation inequalities.

1 Introduction and main results

Let $N \geq n$ be positive integers. In this paper we study the smallest singular value of $N \times n$ matrices $\Gamma = (\xi_{ji})$, whose entries are real-valued random variables obeying certain probability laws, and furthermore we are interested in allowing these matrices to contain some null entries (or, more generally, to contain entries with small variances). Thus we deal with *sparse* (or *dilute*) random matrices. Sparse random matrices and sparse structures play an important role, as they arise naturally 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 [14, 31]).

Understanding the properties of random matrices, in particular the behavior of their singular values (see the definitions in Section 2), is of importance in several fields, including Asymptotic Geometric Analysis, Approximation

[1]Research partially supported by the E.W.R. Steacie Memorial Fellowship.

Theory, Probability and Statistics. The study of extreme singular values in classical random matrix theory concentrates on their limiting behavior as the dimension grows to infinity. Such limiting behavior is now well understood for various kinds of random matrices whose entries are independent in aggregate, or independent up to the symmetry constraints (e.g. hermitian or unitary matrices), in many cases even with identical distribution being required. We refer to the following books, surveys, and recent papers for history, results, and open problems in this direction [4, 5, 9, 11, 12, 22, 31, 33].

In the *non-limiting asymptotic* case very little was known till very recently. In such a case one studies the rate of convergence, deviation inequalities, and the general asymptotic behavior of singular values of a matrix as functions of the dimensions, assuming that the dimensions are large enough (growing to infinity). The Gaussian case, i.e. the case when the entries of the matrix are independent $\mathcal{N}(0,1)$ Gaussian, was treated independently in [8] and [29] (see also [13] for related results, and the survey [7]). In the last decade the attention shifted to other models, like matrices with independent subgaussian entries (in particular, symmetric Bernoulli ± 1 entries), independent entries satisfying some moment conditions as well as matrices with independent columns or rows satisfying some natural restrictions. Major achievements were obtained in [2, 3, 18, 25, 26, 27, 30, 32].

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 one deals 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, making sparse random matrices more suited in modelling such partially connected systems.

The main goal of our paper is to show that one can significantly relax the condition of boundedness from below of all entries, replacing it by averaging type conditions. Thus our paper clarifies the role of the variances in the corresponding previous results (cf. e.g. [18, 25, 26, 27]). 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 subgaussian moment of entries (both conditions were crucial for deep results of [27]). 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. 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 $\frac{n}{N} \le c_0$ 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.

Since we will deal with random matrices under various conditions, for the sake of exposition clarity we list now all our conditions. For parameters $r > 2$, $\mu \ge 1$, $a_1 > 0$, $a_2 > 0$, $a_3 \in (0, \mu)$, and $a_4 \in (0, 1]$, we will consider $N \times n$ random matrices $\Gamma = (\xi_{ji})_{j \le N, i \le n}$ whose entries are *independent* real-valued *centered* random variables satisfying the following conditions:

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

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

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

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

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

Notice that these conditions allow our matrices to contain many null (or small) entries, in the sense that we don't impose any restrictions on the variance of a particular random variable. Naturally, in order for our random matrices to have entries of different kinds, we do not require that the entries are identically distributed. Our model is different from the sparse matrix models used e.g. in [14, 31], where zeros appeared randomly, i.e. starting from a random matrix whose entries have variances bounded away from 0, each entry was multiplied by another random variable of type 0/1. Our model is more similar to those considered in [9], where a condition similar to (iii) was used for square symmetric matrices.

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 stating our main results let us comment our conditions in more detail. The first condition is a standard requirement saying that the random variables are not "too big". For the limiting case it is known that one needs boundedness of the forth moments. 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 [34], 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 an $N \times n$ 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 behavior of the largest singular value?"*) is one of the central questions in random matrix theory. Such estimates are well known for the Gaussian and subgaussian cases. We refer to [3, 16] 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 the normalization conditions (1.1) and (1.16) in [9]. Our methods are similar to those used in [18, 27], however we deal with a rather different model, and correspondingly our proofs require much more delicate computations. In particular, the proof of key Proposition 4.1, which estimates the probability that for a fixed vector x the Euclidean norm $\|\Gamma x\|_2$ is small, is much more involved (cf. the proof of [18, Proposition 3.4] or [26, Corollary 2.7]).

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 behavior 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 [18] (for uniformly *bounded above* mean zero random variables with bounded below variances 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 1.1. *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 $c_1 = c_1(r, \mu, a_3)$, $c_2 = c_2(r, \mu, a_2, a_3)$ and $\delta_0 = \delta_0(r, \mu, a_1, a_3)$.

Our next theorem is about "almost square" matrices. This theorem extends [18, Theorem 3.1]. Here both conditions (iii) and (iv) are needed in order to substitute condition (1).

Theorem 1.2. *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 $n \leq 2/\tilde{c}_2$, Theorem 1.2 is trivial for every $\delta > 0$, either by adjusting the constant c_2 (for small N) or by using Theorem 1.1 (for large N).

Let us note that in a sense our Theorems 1.1 and 1.2 are incomparable with the corresponding result of [27]. First, we don't restrict our results only to the subgaussian case. The requirement of boundedness of the subgaussian moment is much stronger, implying in particular boundedness of moments of all orders, which naturally yields stronger estimates. Second, another condition essentially used in [27] 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 [26], where the i.i.d. case was treated, and from [1, 2], 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 [26] two theorems about square matrices were proved. The first one is for random matrices whose entries have bounded fourth moment. Our Theorem 1.3 extends this result with much better probability. The second main result of [26] requires the boundedness of subgaussian moments as well as identical distributions of entries in each column, and, thus, is incomparable with Theorem 1.3.

Theorem 1.3. *Let $r > 2$, $\mu \geq 1$, $a_1, a_2, a_3, a_4 > 0$ 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 there exists a positive constant $\gamma_0 = \gamma_0(r, \mu, a_1, a_3) < 1$ such that if $a_4 > 1 - \gamma_0$ then for every $\varepsilon \geq 0$*

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

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

Finally we would like to mention that all results can be extended to the complex case in a standard way.

Acknowledgment. The authors would like to thank N. Tomczak-Jaegermann for many useful conversations. We also thank S. Spektor for showing us reference [24] and S. O'Rourke for showing us reference [9]. The second named author thanks G. Schechtman for hosting him at the Weizmann Institute of Science in Spring 2008, during which time part of this work was done.

2 Notation and preliminaries

We start this section by agreeing on the notation that we will use throughout. For $1 \leq p \leq \infty$, we write $\|x\|_p$ for the ℓ_p-norm of $x = (x_i)_{i\geq 1}$, i.e. the norm defined by

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

Then, as usual, $\ell_p^n = (\mathbb{R}^n, \|\cdot\|_p)$. The unit ball of ℓ_p^n is denoted B_p^n. Also, S^{n-1} denotes the unit sphere of ℓ_2^n, 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, by $|A|$ we denote the cardinality of A.

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

Given a subspace E of $\mathbb{R}^n$ we denote by P_E the orthogonal projection onto E. If $E = \mathbb{R}^\sigma$ 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 (see for instance [23, Lemma 2.6]) establishes that for each $\varepsilon > 0$ there is an ε-net $\mathcal{N}$ of D with cardinality $|\mathcal{N}| \le (1 + 2/\varepsilon)^n$.

2.1 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 further 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 particularly important. They may be equivalently given by the expressions

$$s_1(\Gamma) = \|\Gamma : \ell_2^n \to \ell_2^N\| = \sup\big\{|\Gamma x| \ : \ |x| = 1\big\}, \quad s_n(\Gamma) = \inf\big\{|\Gamma x| \ : \ |x| = 1\big\}.$$

In particular for every vector $x \in \mathbb{R}^n$ one has

$$s_n(\Gamma)|x| \le |\Gamma x| \le s_1(\Gamma)|x|. \tag{2}$$

Note that the estimate on the left-hand side becomes trivial 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 (2) providing an estimate for the distortion of the norms under Γ.

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

$$s_n(\Gamma) \le \lambda \quad \Longleftrightarrow \quad \exists x \in S^{n-1} \ : \ |\Gamma x| \le \lambda. \tag{3}$$

2.2 Subgaussian random variables.

All random quantities appearing in this work are defined on the same underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$. We will present estimates for the smallest singular value of matrices whose entries are independent random variables satisfying certain assumptions. Our results are valid for a large class of matrices which includes, in particular, those whose entries are subgaussian random variables.

A (real-valued) random variable X is called *subgaussian* when there exists a positive constant b such that for every $t \in \mathbb{R}$

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

When this condition is satisfied with a particular value of $b > 0$, we also say that X is b-subgaussian, or subgaussian with parameter b. The minimal b in this capacity is called the *subgaussian moment* of X.

It is an easy consequence of this definition that if X is b-subgaussian, then $\mathbb{E}(X) = 0$ and $\mathrm{Var}(X) \leq b^2$. Thus all subgaussian random variables are centered. The next proposition presents well-known equivalent conditions for a centered random variable to be subgaussian.

Proposition 2.1. *For a centered random variable X, the following statements are equivalent:*

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

(2) $\exists b > 0,\ \forall \lambda > 0,\ \mathbb{P}(|X| \geq \lambda) \leq 2e^{-\lambda^2/b^2}$

(3) $\exists b > 0,\ \forall p \geq 1,\ (\mathbb{E}|X|^p)^{1/p} \leq b\sqrt{p}$

(4) $\exists c > 0,\ \mathbb{E}\, e^{cX^2} < +\infty$

Two important examples of subgaussian random variables are the centered Gaussian themselves and the symmetric Bernoulli ± 1 random variables. In general, any centered and bounded random variable is subgaussian.

We point out that, as consequence of the subgaussian tail estimate, the norm of a matrix whose entries are independent subgaussian random variables is of the order of $\sqrt{N}$ with high probability. Namely, the following proposition holds (see e.g. [18, Fact 2.4], where this was shown for symmetric random variables, the case of centered is essentially the same).

Proposition 2.2. *Let $N \geq n \geq 1$ be positive integers. Suppose Γ is an $N \times n$ matrix whose entries are independent subgaussian random variables with subgaussian parameters bounded above uniformly by b. Then there are positive constants c, C (depending only on b) such that for every $t > C$*

$$\mathbb{P}\big(\|\Gamma\| > t\sqrt{N}\big) \leq e^{-ct^2 N}.$$

2.3 Compressible and incompressible vectors.

As equivalence (3) suggests, to estimate the smallest singular value of 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, as usual, we use 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 will 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.

The sets of sparse, compressible, and incompressible vectors will be denoted, respectively, $Sparse(m)$, $Comp(m,\rho)$, and $Incomp(m,\rho)$. The idea to split the Euclidean sphere into two parts goes back to Kashin's work [15] 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 ([28]) in the same context. The splitting the sphere essentially as described above appeared in [18, 19] and was later used in many works (e.g. in [26, 27]).

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

$$\begin{aligned} x \in Comp(m,\rho) &\iff \exists \sigma \subset \{1,\dots,n\} \text{ with } |\sigma^c| \le m \text{ such that } |P_\sigma x| \le \rho \\ x \in Incomp(m,\rho) &\iff \forall \sigma \subset \{1,\dots,n\} \text{ with } |\sigma^c| \le m \text{ one has } |P_\sigma x| > \rho. \end{aligned} \tag{4}$$

2.4 Two more results.

Here we formulate two results, which will be used in the next section. The first one is a quantitative version of the Central Limit Theorem (CLT), called Berry-Esséen inequality. The second one is a general form of the Paley-Zygmund inequality (see e.g. [18, Lemma 3.5]).

Theorem 2.3 (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 standard form of Berry-Esséen inequality requires finite 3-rd moment (i.e., it is usually stated for $r = 3$), see e.g. [10, p. 544] or [21, p. 300]. The form used here is from [24] (see Theorem 5.7 there).
2. If $r \geq 3$, 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 $r = 3$).

Lemma 2.4 (Paley-Zygmund inequality)**.** *Let $p \in (1, \infty)$, $q = p/(p-1)$. Let $f \geq 0$ be a random variable with $\mathbb{E}\, f^{2p} < \infty$. Then for every $0 \leq \lambda \leq \sqrt{\mathbb{E}\, f^2}$ we have*

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

3 Small ball probabilities for random sums

In this section we gather auxiliary results related to random sums, their small ball probabilities, etc., which are needed later. In fact, we adjust corresponding results from [18] and [26] 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 [18, 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 [17, Lemma 6.3], and applying Khinchine's inequality, we obtain

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

Now consider the set

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

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

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

This function is clearly convex, so that

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

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

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

Lemma 2.4 with $p = r/2$, $q = r/(r-2)$ implies the desired estimate. ■

The next proposition, which is a consequence of Theorem 2.3, allows us to estimate the small ball probability. The proof goes along the same lines as the proof of [18, Proposition 3.2] (see also [20, Proposition 3.4]), with slight modifications to remove the assumption about variances. Recall that for a subset $\sigma \subset \{1, 2, \dots, n\}$, P_σ denotes the coordinate projection onto $\mathbb{R}^\sigma$.

Proposition 3.2. *Let $2 < r \le 3$ and $\mu \ge 1$. Let $(\xi_i)_{i=1}^n$ be independent centered random variables with $\mathbb{E}|\xi_i|^r \le \mu^r$ for all $i = 1, 2, \dots, n$. 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

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

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

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

The next corollary gives an estimate on the small ball probability in the spirit of [26, Corollary 2.10].

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

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

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

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

■

We need the following lemma proved in [26, 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 [26, 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 $\overline{\sigma} := \{i : \mathbb{E}\xi_i^2 \geq 1\}$ 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 $t \geq 0$*

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

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

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

■

4 "Tall" matrices (proof of Theorem 1.1)

In this section we prove Theorem 1.1, 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 1.1 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

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

We postpone the proof of this technical result to the last section, so that we may keep the flow of our exposition uninterrupted.

Proof of Theorem 1.1. Passing to $r_0 = \min\{3, r\}$ we may assume without loss of generality that $r \le 3$.

Let $t \ge 0$ and $\Omega_0 := \{\omega \,:\, \|\Gamma\| \le a_1\sqrt{N}\}$. By (3) it is enough to estimate the probability of the event

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

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^{n-1} with cardinality $|\mathcal{N}| \leq (3/\varepsilon)^n$. For any $x \in S^{n-1}$ we can find $y \in \mathcal{N}$ such that $|x - y| \leq \varepsilon$. If further x satisfies $|\Gamma x| \leq t\sqrt{N}$, then the corresponding y satisfies

$$|\Gamma y| \leq |\Gamma x| + \|\Gamma\| \cdot |y - x| \leq t\sqrt{N} + \varepsilon a_1 \sqrt{N} = (t + \varepsilon a_1)\sqrt{N}. \tag{5}$$

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

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

where b_1 and b_2 are as in Proposition 4.1. Thus

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

as long as

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

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

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

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

5 "Almost square" matrices (proof of Theorem 1.2)

In this section we prove Theorem 1.2. 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 1.2. Passing to $r_0 = \min\{3, r\}$ we may assume without loss of generality that $r \leq 3$.

Consider the event

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

By equivalence (3) 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:

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

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

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

$$E \subset (E_C \cap \Omega_0) \cup (E_I \cap \Omega_0) \cup \Omega_0^c. \tag{7}$$

Our proof will require that $t \le 1$ (which will be satisfied once we choose t, see (22) below); and furthermore that t and ρ satisfy

$$\frac{2t}{a_1} \le \rho \le \frac{1}{4}. \tag{8}$$

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

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

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 (8),

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

and, by another use of (8),

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

Hence

$$\mathbb{P}(E_C \cap \Omega_0) \le \mathbb{P}\Big(\exists v \in \mathcal{N} \text{ s.t. } |\Gamma v| \le 5a_1\rho\sqrt{N}|v|\Big) \le \sum_{v\in\mathcal{N}} \mathbb{P}\Big(|\Gamma v| \le 5a_1\rho\sqrt{N}|v|\Big). \tag{9}$$

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

$$5a_1\rho \le b_1. \tag{10}$$

We choose

$$\rho := \min\left\{\frac{1}{4}, \frac{b_1}{5a_1}\right\} \tag{11}$$

so that both (10) and the right hand side of (8) are true. Now, from (9), we have

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

Thus, if

$$m \ln\Big(\frac{3en}{\rho m}\Big) \le \frac{b_2 N}{2} \tag{12}$$

then

$$\mathbb{P}(E_C \cap \Omega_0) \le e^{-\frac{b_2 N}{2}}. \tag{13}$$

Writing $m = \gamma n$, we see that inequality (12) is satisfied if

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

so we choose

$$\gamma = \frac{b_2}{4\ln\big(\frac{6e}{\rho b_2}\big)}. \tag{14}$$

Case 2: Probability of $E_I \cap \Omega_0$. We work on the set $Incomp(m, \rho)$, where ρ is defined in (11) and $m = \gamma n$ with γ chosen in (14).

For convenience we set $a := t^{1/(r-2)}/a_1$. Since $t \le 1$ and in view of (8), 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 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 (4) 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 $a_1 a = t^{1/(r-2)} \le t$ (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 (8).

Denote by $\mathcal{A}$ the set of all $v \in \mathcal{N}$ 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

$$\mathbb{P}(E_I \cap \Omega_0) \le \mathbb{P}\Big(\exists v \in \mathcal{A} \,:\, |\Gamma v| \le 2t\sqrt{N}\Big). \tag{15}$$

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

$$\begin{aligned}\mathbb{P}\Big(|\Gamma v|^2 \le 4t^2 N\Big) &= \mathbb{P}\Big(N - \frac{1}{4t^2}|\Gamma v|^2 \ge 0\Big) \\ &\le \mathbb{E}\exp\Big\{N - \frac{1}{4t^2}|\Gamma v|^2\Big\} \\ &= e^N\, \mathbb{E}\exp\Big\{-\frac{1}{4t^2}\sum_{j=1}^N \Big|\sum_{i=1}^n \xi_{ji} v_i\Big|^2\Big\} \\ &= e^N \prod_{j=1}^N \mathbb{E}\exp\Big\{-\frac{1}{4t^2}\Big|\sum_{i=1}^n \xi_{ji} v_i\Big|^2\Big\},\end{aligned} \tag{16}$$

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

$$\begin{aligned}\mathbb{E}\exp\Big\{-\frac{1}{4t^2}\Big|\sum_{i=1}^n \xi_{ji} v_i\Big|^2\Big\} &= \int_0^1 \mathbb{P}\Big(\exp\Big\{-\frac{1}{4t^2}\Big|\sum_{i=1}^n \xi_{ji} v_i\Big|^2\Big\} > s\Big) ds \\ &= \int_0^\infty u e^{-u^2/2}\, \mathbb{P}\Big(\Big|\sum_{i=1}^n \xi_{ji} v_i\Big| < \sqrt{2}tu\Big) du.\end{aligned} \tag{17}$$

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 $|\sigma_j| \ge a_4 n$ such that $\mathbb{E}\,\xi_{ji}^2 \ge 1$ for all $i \in \sigma_j$ (this is condition (iv)). Also, for each fixed v, set

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

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

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

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

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

$$(1 - a_4)n + \frac{1}{a^2} \le m. \tag{18}$$

Note that (18), in particular, implies $1/a^2 \leq a_4 n \leq n$. Recall that $m = \gamma n$, where γ was chosen in (14). Then inequality (18) is satisfied if $a_4 > 1 - \gamma$ (which is the condition on γ in our Theorem) and

$$t \geq \left(\frac{a_1}{\sqrt{(\gamma + a_4 - 1)n}} \right)^{r-2}. \tag{19}$$

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

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

for some absolute constant $c \geq 1$. Then, continuing from (17) we have

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

where

$$c_3 := c\Big(\frac{\sqrt{\pi}}{\rho} + \frac{\mu^r}{\rho^r a_1^{r-2}} \Big). \tag{20}$$

Therefore, from (16), we get (for each fixed $v \in \mathcal{A}$)

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

and from this, in (15) we get

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

Then we can make

$$\mathbb{P}(E_I \cap \Omega_0) \leq e^{-N} \tag{21}$$

provided that

$$t \le \frac{1}{c_3 e^2} \left(\frac{1}{3a_1 c_3 e^2} \right)^{1/\delta}. \tag{22}$$

Choose t to satisfy equality in (22). Note

$$\frac{t}{a_1} \le \frac{1}{c_3 e^2 a_1} \le \frac{\rho}{ce^2\sqrt{\pi}}, \tag{23}$$

so the left hand side of (8) holds. Finally note that (19) is satisfied whenever

$$\delta \ge \frac{\frac{2}{r-2}\ln(3a_1c_3e^2)}{\ln\Big(\frac{(\gamma+a_4-1)n}{a_1^2(c_3e^2)^{2/(r-2)}}\Big)} =: \frac{\tilde{c}_1}{\ln(\tilde{c}_2 n)}.$$

To finish, we take probabilities in (7) 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 (13) and (21), 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. ■

6 Square matrices (proof of Theorem 1.3)

In this section our goal is to prove Theorem 1.3. We are going to use two lemmas from [26]. The first one is [26, 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 $X_1, \dots, X_n$ 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 [26, Lemma 3.8]. To prove it one would repeat the proof of that lemma, replacing [26, Lemma 3.7] used there with our Lemma 3.5.

Lemma 6.2. *Let $r \in (2,3]$ and Γ be a random matrix as in Theorem 1.3. Let $X_1, \dots, X_n$ denote its column vectors, and consider the subspace $H_n = \mathrm{span}(X_1, \dots, X_{n-1})$. Then there exists a positive constant $\gamma_0 = \gamma_0(r, \mu, a_1, a_3) < 1$ such that if $a_4 > 1 - \gamma_0$ 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 on r, μ, a_1, a_3, and a_4.

Now we are ready for the proof of Theorem 1.3.

Proof of Theorem 1.3. 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

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

By equivalence (3) we are to estimate $\mathbb{P}(E)$ with an appropriate value of t (which will be specified later).

As in the proof of Theorem 1.2, we split the set E into the sets E_C and E_I defined as follows:

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

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

$$E \subset (E_C \cap \Omega_0) \cup E_I \cup \Omega_0^c. \tag{24}$$

Case 1: Probability of $E_C \cap \Omega_0$. The proof of this case is almost line to line repetition of the corresponding proof in Theorem 1.2 (see Case 1 there). Let $m \le n$ and $\rho \in (0, 1)$ be specified later. Using approximation argument and the union bound as in the proof of Case 1 in Theorem 1.2, and choosing

$$\rho := \min\left\{\frac{1}{4}, \frac{b_1}{5a_1}\right\}, \qquad \gamma := \frac{b_2}{4\ln\left(\frac{6e}{\rho b_2}\right)}, \qquad m = \gamma n, \tag{25}$$

we obtain

$$\mathbb{P}(E_C \cap \Omega_0) \le e^{-b_2 n/2}, \tag{26}$$

provided that

$$\frac{2t}{a_1} \le \rho. \tag{27}$$

Case 2: Probability of E_I. We work on the set $Incomp(m, \rho)$, where $m = \gamma n$ and γ, ρ chosen in (25).

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

$$\begin{aligned}
\mathbb{P}(E_I) &\le \frac{1}{\gamma n} \sum_{k=1}^{n} \mathbb{P}\big(\mathrm{dist}(X_k, H_k) < t/\rho\big) \\
&\le \frac{1}{\gamma n} \sum_{k=1}^{n} \Big\{\mathbb{P}\big(\mathrm{dist}(X_k, H_k) < t/\rho \quad \& \quad \|\Gamma\| \le a_1\sqrt{n}\big) + \mathbb{P}\big(\|\Gamma\| > a_1 \sqrt{n}\big)\Big\} \\
&\le \frac{1}{\gamma n} \sum_{k=1}^{n} \Big\{c(\varepsilon n^{\frac{3-r}{2}} + n^{\frac{2-r}{2}}) + e^{-a_2 n}\Big\} \\
&\le \frac{c}{\gamma}(\varepsilon n^{\frac{3-r}{2}} + n^{\frac{2-r}{2}}) + \frac{1}{\gamma} e^{-a_2 n}.
\end{aligned} \tag{28}$$

Also notice that our choice $t = \varepsilon\rho$ and our assumption $\varepsilon \le a_1/2$ guarantee that t satisfies (27).

To finish the proof, we take probabilities in (24), and we use the estimates for $\mathbb{P}(E_C \cap \Omega_0)$ and for $\mathbb{P}(E_I)$ obtained in (26) and (28), 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. ■

7 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:

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

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 hand we have

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

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

$$\sum_{j=1}^N \left(\mathbb{E} \sum_{i=1}^n \xi_{ji}^2 x_i^2 \right) = \sum_{i=1}^n \left(\mathbb{E} \sum_{j=1}^N \xi_{ji}^2 \right) x_i^2 \ge \sum_{i=1}^n a_3^2 N x_i^2 = a_3^2 N.$$

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

$$\alpha \ge \frac{a_3^2 - a}{\mu^2 - a}. \tag{29}$$

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 $f_j := \left|\sum_{i=1}^n \xi_{ji} x_i\right|$, 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

$$\begin{aligned}\mathbb{P}\big(|\Gamma x|^2 \leq t^2 N\big) &= \mathbb{P}\left(\sum_{j=1}^N f_j^2 \leq t^2 N\right) = \mathbb{P}\left(\tau N - \frac{\tau}{t^2}\sum_{j=1}^N f_j^2 \geq 0\right) \\ &\leq \mathbb{E}\exp\left(\tau N - \frac{\tau}{t^2}\sum_{j=1}^N f_j^2\right) = e^{\tau N}\prod_{j=1}^N \mathbb{E}\exp\left(-\frac{\tau f_j^2}{t^2}\right). \end{aligned} \tag{30}$$

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

$$\mathbb{P}(f_j > \lambda) \geq \left(\frac{[\mathbb{E}\sum_{i=1}^n \xi_{ji}^2 x_i^2 - \lambda^2]_+}{8\mu^2}\right)^{r/(r-2)} =: \beta_j(r), \tag{31}$$

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

$$\beta_j \geq \left(\frac{[a-\lambda^2]_+}{8\mu^2}\right)^{r/(r-2)}. \tag{32}$$

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

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

Choosing $\eta = \ln 2$ and applying (31), 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 (30) implies

$$\mathbb{P}\big(|\Gamma x|^2 \leq t^2 N\big) \leq e^{\tau N}\prod_{j=1}^N e^{-\beta_j(r)/2} \leq e^{(t^2/\lambda^2)N}\prod_{j\in J} e^{-\beta_j(r)/2}. \tag{33}$$

Taking $a = a_3^2/2$ and $\lambda = a_3/2$ and using (32) we observe that for every $j \in J$ we have $\beta_j \geq \big(\frac{a_3^2}{32\mu^2}\big)^{r/(r-2)}$. Also note this choice of a and (29) imply $\alpha \geq a_3^2/(2\mu^2)$. 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 (33) we obtain

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

This completes the proof. ■

References

[1] R. Adamczak, A. E. Litvak, A. Pajor, N. Tomczak-Jaegermann, Sharp bounds on the rate of convergence of the empirical covariance matrix. *C. R. Math. Acad. Sci. Paris*, **349** (2011), 195–200.

[2] R. Adamczak, O. Guedon, 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. AMS*, **140** (2012), 987–998.

[3] 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.

[4] G.W. Anderson, A. Guionnet, O. Zeituni, *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] 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.

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

[9] L. Erdös, Universality of Wigner random matrices: a survey of recent results (Russian) *Uspekhi Mat. Nauk*, **66** (2011), 67–198; translation in *Russian Math. Surveys* **66** (2011), 507–626.

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

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

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

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

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

[15] B. S. Kashin, Diameters of some finite-dimensional sets and classes of smooth functions, *Izv. Akad. Nauk SSSR*, Ser. Mat. **41** (1977), 334–351.

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

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

[18] 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.

[19] 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.

[20] 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.

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

[22] M.L. Mehta, *Random Matrices.* 3rd Ed. Elsevier Academic Press, 2004.

[23] 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.

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

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

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

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

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

[29] S.J. Szarek, Condition numbers of random matrices, *J. Complexity* **7** (1991), 131–149.

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

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

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

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

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

A. E. Litvak, Dept. of Math. and Stat. Sciences, University of Alberta, Edmonton, Alberta T6G 2G1, Canada, `aelitvak@gmail.com`

O. Rivasplata, Dept. of Math. and Stat. Sciences, University of Alberta, Edmonton, Alberta T6G 2G1, Canada, `orivasplata@ualberta.ca`