Advances in Mathematics 284 (2015) 1–20

Contents lists available at ScienceDirect

Advances in Mathematics

www.elsevier.com/locate/aim

The limit of the smallest singular value of random matrices with i.i.d. entries

Konstantin Tikhomirov

Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, T6G 2G1, Canada

ARTICLE INFO

Article history:
Received 13 October 2014
Received in revised form 17 July 2015
Accepted 27 July 2015
Available online 5 August 2015
Communicated by D.W. Stroock

Keywords:
Random matrices
Smallest singular value

ABSTRACT

Let $\{a_{ij}\}$ $(1 \le i, j < \infty)$ be i.i.d. real-valued random variables with zero mean and unit variance and let an integer sequence $(N_m)_{m=1}^{\infty}$ satisfy $m/N_m \longrightarrow z$ for some $z \in (0, 1)$. For each $m \in \mathbb{N}$ denote by A_m the $N_m \times m$ random matrix (a_{ij}) $(1 \le i \le N_m,\ 1 \le j \le m)$ and let $s_m(A_m)$ be its smallest singular value. We prove that the sequence $\left(N_m{}^{-1/2} s_m(A_m)\right)_{m=1}^{\infty}$ converges to $1 - \sqrt{z}$ almost surely. Our result does not require boundedness of any moments of a_{ij}'s higher than the 2-nd and resolves a long standing question regarding the weakest moment assumptions on the distribution of the entries sufficient for the convergence to hold.

1. Introduction

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

E-mail address: ktikhomi@ualberta.ca.

http://dx.doi.org/10.1016/j.aim.2015.07.020

given by

$$s_1(B) = \sup_{y \in S^{m-1}} \|By\| = \|B\|; \quad s_m(B) = \inf_{y \in S^{m-1}} \|By\|.$$

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

The extreme singular values of random matrices attract considerable attention of researchers both in *limiting* and *non-limiting* settings. We refer the reader to surveys and monographs [2,11,16,21] for extensive information on the spectral theory of random matrices. Here, we shall focus on the following specific question: for matrices with i.i.d. entries, what are the weakest possible assumptions on the entries which are sufficient for the smallest singular value to "concentrate"?

We note that a corresponding problem for the *largest* singular value (i.e. the operator norm) was essentially resolved in the i.i.d. case, where finiteness of the fourth moment of the entries turns out to be crucial both in limiting and non-limiting settings. We refer the reader to [24] and [3] for results on a.s. convergence of the largest singular value, and [7] for the non-limiting case (see also [17,9] for some negative results on concentration of the operator norm).

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

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

Further, it is proved in [13,14] that for square $m \times m$ matrices with i.i.d. centered entries with unit variance and a bounded fourth moment, one has $s_m(A) \approx m^{-1/2}$ with a large probability.

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

A considerable progress has been made recently in the direction of weakening the moment assumptions on matrix entries. For square matrices, given a sufficiently large m and an $m \times m$ matrix with i.i.d. entries with zero mean and unit variance, its smallest singular value is bounded from below by a constant (negative) power of m with probability close to one [19, Theorem 2.1] (see also [5, Theorem 4.1] for sparse matrices).

For *tall* rectangular matrices, Srivastava and Vershynin proved in [18] that for any $\varepsilon, \eta > 0$ and an $N \times m$ random matrix A with independent isotropic rows X_i such that $\sup_{y\in S^{m-1}} \mathbb{E}|\langle X_i, y\rangle|^{2+\eta} \le C$, the singular value $s_m(A)$ satisfies $\mathbb{E}s_m(A)^2 \ge (1-\varepsilon)N$ provided that the aspect ratio N/m is bounded from below by a certain function of ε and η. This result of [18] was strengthened by Koltchinskii and Mendelson [6] who proved that, under similar assumptions on the matrix, $s_m(A) \ge (1-\varepsilon)\sqrt{N}$ with a very large probability. Moreover, another theorem of [6] states that, for a sufficiently tall $N \times m$ random matrix A with i.i.d. isotropic rows satisfying certain "spreading" condition, $s_m(A) \gtrsim \sqrt{N}$ with probability very close to one. Some further strengthening of the results of [6] is obtained in [22].

A situation when no upper bounds for moments of the matrix entries are given, was considered in [20]. It was proved that for any $\delta > 1$, $N \ge \delta m$ and for an $N \times m$ random matrix A with i.i.d. entries satisfying $\inf_{\lambda\in\mathbb{R}} \mathbb{P}\{|a_{11}-\lambda| \ge \alpha\} \ge \beta$ for some $\alpha, \beta > 0$, one has $\mathbb{P}\{s_m(A) \ge \alpha u\sqrt{N}\} \ge 1 - 2\exp(-vN)$, where $u, v > 0$ depend only on β and δ.

The result of [20] can be used to show that in the limiting setup of the Bai–Yin theorem but without the assumptions on moments higher than the 2-nd, the sequence $\big({N_m}^{-1/2} s_m(A_m)\big)_{m=1}^\infty$ satisfies

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

where r is a certain function of $z = \lim m/N_m$ and the distribution of a_{ij}'s. The same conclusion can be derived from [6, Theorem 1.4], if we additionally assume that the limiting aspect ratio z is bounded from above by a sufficiently small positive quantity (i.e. the matrices are tall). However, both [20, Theorem 1] and [6, Theorem 1.4] do not give the precise asymptotics.

This problem is resolved in our paper. The main result is the following

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

$$\big({N_m}^{-1/2} s_m(A_m)\big)_{m=1}^\infty$$

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

Theorem 1 in a strong form establishes the *asymmetry* of the limiting behaviour of the extreme singular values: whereas the fourth moment is necessary for the operator norm, the second moment is sufficient for the convergence of the smallest singular value.

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

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

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

$$\limsup_{m\to\infty} {N_m}^{-1/2}\big(s_m(\tilde{A}) - s_m(A)\big)$$

is bounded from above by a positive number which depends only on M and can be made arbitrarily small by increasing the truncation level (in a more technical form, this is stated in Theorem 15 of the note). Then, applying the Bai–Yin theorem [4] to the truncated matrices $\tilde{A}_m$, we get

$$\liminf_{m\to\infty} {N_m}^{-1/2} s_m(A) \gtrsim \liminf_{m\to\infty} {N_m}^{-1/2} s_m(\tilde{A}) \gtrsim 1 - \sqrt{z} \quad \text{almost surely,}$$

which implies the result. Thus, the argument of the paper [4] remains the crucial element of the proof, although we apply it only to the truncated variables, for which all positive moments are bounded. Let us emphasize that, whereas a truncation procedure for matrices also appears as a technical step in [4], in our approach the truncation level M is *not a function of* m.

Note that the equivalence $s_m(A_m) \approx s_m(\tilde{A}_m)$ would follow immediately if the difference $A_m - \tilde{A}_m$ had the operator norm very small compared to $\sqrt{N_m}$ with a large probability. However, the moment assumptions that we impose on a_{ij}'s are too weak to expect a good upper bound for $\|A_m - \tilde{A}_m\|$. To overcome this problem, we shall consider a special *non-convex* function of the matrix $A_m - \tilde{A}_m$ which has much better concentration properties than the norm and which shall act as a "replacement" for the norm in our calculations. This quantity and its concentration properties are discussed in Section 3 and are the main novel ingredient of the paper.

2. Preliminaries

In this section, we introduce notation and present some classical or elementary facts, which we include for an easier referencing.

We denote by $(\Omega, \Sigma, \mathbb{P})$ a probability space, and adopt the usual notations and definitions from the Probability Theory such as i.i.d. random variables, the expectation, etc. Let $\{e_i\}_{i=1}^N$ be the standard unit vector basis in $\mathbb{R}^N$, $\|\cdot\|$ and $\langle\cdot,\cdot\rangle$ be the canonical Euclidean norm and corresponding inner product, and $\|\cdot\|_\infty$ be the maximum (ℓ_∞-) norm. The unit Euclidean ball in $\mathbb{R}^n$ shall be denoted by B_2^n and the cube $[-1,1]^n$ – by B_∞^n. For a finite set I, $|I|$ is its cardinality. Universal constants are denoted by C, c_1, etc. A numerical subscript in the name of a constant determines the statement where the constant is defined. Similarly, a function defined within a statement and intended to be used further in the paper, has the statement number as a subscript.

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

Lemma 2. *For any $n \in \mathbb{N}$ and $\varepsilon \in (0, 1]$ there exists an ε-net in B_2^n of cardinality at most $\left(\frac{3}{\varepsilon}\right)^n$.*

Lemma 3. *For any $n \in \mathbb{N}$ and any $T \subset S^{n-1}$ there is an $n^{-1/2}$-net in T with respect to $\|\cdot\|_\infty$ of cardinality at most $\exp(C_3 n)$. Here, $C_3 > 0$ is a universal constant.*

Remark 1. Both lemmas above follow from a well-known estimate for covering numbers for pairs of convex sets in $\mathbb{R}^n$ (see, for example, [12, Lemma 4.16]). For Lemma 3, the estimate for the pair (B_2^n, B_∞^n) yields an existence of a $(4n)^{-1/2}$-net $\bar{\mathcal{N}}$ in B_2^n with respect to $\|\cdot\|_\infty$ of cardinality at most $\exp(C_3 n)$ for an absolute constant $C_3 > 0$. Then $\mathcal{N} \subset T$ can be constructed by picking a point from every non-empty intersection of the form $(y' + (4n)^{-1/2} B_\infty^n) \cap T$, $y' \in \bar{\mathcal{N}}$.

The next statement, which is sometimes called the Bernstein (or Hoeffding's) inequality, can be derived from classical Khintchine's inequality for the sum of weighted independent signs by a symmetrization procedure:

Lemma 4. *(See, for example, [21, Proposition 5.10].) Let $n \in \mathbb{N}$, $M > 0$, $y = (y_1, y_2, \ldots, y_n)$ with $\|y\| = 1$, and let $a_1, a_2, \ldots, a_n$ be independent mean zero random variables with $|a_j| \le M$ a.s. ($j = 1, 2, \ldots, n$). Then*

$$\mathbb{P}\Big\{\Big|\sum_{j=1}^{n} a_j y_j\Big| \ge \tau\Big\} \le 2\exp(-c_4 \tau^2 / M^2), \quad \tau > 0,$$

where $c_4 > 0$ is a universal constant.

The lemma below is a law of large numbers, where instead of the arithmetic mean of a collection of random variables we consider more general weighted sums. As in the case of the classical weak LLN, the statement can be proved by applying Levy's continuity theorem for characteristic functions.

Lemma 5. *Let $a_1, a_2, \ldots$ be i.i.d. random variables with zero mean. Then for any $\varepsilon > 0$ there is $\delta > 0$ depending only on ε and the distribution of a_j's with the following property: whenever $(t_j)_{j=1}^\infty$ is a sequence of non-negative real numbers such that $\sum_{j=1}^\infty t_j = 1$ and $\max t_j \le \delta$, we have*

$$\mathbb{P}\Big\{\Big|\sum_{j=1}^{\infty} a_j t_j\Big| > \varepsilon\Big\} < \varepsilon.$$

Given an $m \times m$ random symmetric matrix T with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_m$, *the empirical spectral distribution* of T is the function on $\mathbb{R}$ given by

$$F^T(t) = \frac{1}{m}\big|\{j \le m : \lambda_j \le t\}\big|, \quad t \in \mathbb{R}.$$

Theorem 6 *(Marčenko–Pastur law). (See [10,23], [2, Theorem 3.6].) Let $\{a_{ij}\}$ $(1 \le i, j \le \infty)$ be a set of i.i.d. random variables with zero mean and unit variance and let $(N_m)_{m=1}^{\infty}$ be an integer sequence satisfying $m/N_m \longrightarrow z$ for some $z \in (0, 1)$. For every $m \in \mathbb{N}$ denote by A_m the random $N_m \times m$ matrix with entries a_{ij} $(1 \le i \le N_m$, $1 \le j \le m)$ and by T_m the matrix $\frac{1}{N_m} A_m^T A_m$. Then with probability one the sequence of empirical spectral distributions $\{F^{T_m}\}$ converges pointwise to a non-random distribution given by*

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

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

Remark 2. Note that the above theorem does not require any assumptions on moments higher than the 2-nd, and so can be applied in our setting. For our proof, we will actually need a much weaker result than Theorem 6, namely, that $\limsup\limits_{m\to\infty} \frac{s_m(A_m)}{\sqrt{N_m}} \le 1 - \sqrt{z}$ almost surely. The latter can be immediately verified with help of Theorem 6: for every fixed $t > (1 - \sqrt{z}\,)^2$, we have $\lim\limits_{m\to\infty} F^{T_m}(t) = F_{MP}(t) > 0$ with probability one, hence the smallest non-zero eigenvalues $\lambda_{\min}(T_m)$ of matrices T_m satisfy $\limsup\limits_{m\to\infty} \lambda_{\min}(T_m) \le t$ a.s. This implies $\limsup\limits_{m\to\infty} \frac{s_m(A_m)}{\sqrt{N_m}} \le \sqrt{t}$ a.s., which gives the required estimate by letting $t \to (1 - \sqrt{z}\,)^2$.

3. Norms of coordinate projections of random vectors

For any $N \in \mathbb{N}$ and a subset $I \subset \{1, 2, \dots, N\}$, let us denote by $\mathrm{Proj}_I : \mathbb{R}^N \to \mathbb{R}^N$ the coordinate projection onto the subspace spanned by $\{e_i\}_{i\in I}$. Throughout the rest of the paper, we will often use expressions of the form $\min\limits_{|I|\ge r} \|\mathrm{Proj}_I x\|$, where x is some vector in $\mathbb{R}^N$ and r is a positive real number. This notation should be interpreted as the minimum of $\|\mathrm{Proj}_I x\|$ over *all* subsets $I \subset \{1, 2, \dots, N\}$ of cardinality at least r.

The goal of this section is to show that, given a sufficiently large random $N \times n$ matrix A with i.i.d. entries with zero mean and unit variance, the quantity

$$\sup_{y\in S^{n-1}} \min_{|I|\ge N-\varepsilon N} \|\mathrm{Proj}_I Ay\| \tag{1}$$

is of order $\sqrt{N}$ with a very large probability (the probability shall depend on $\varepsilon > 0$). It shall act as a "replacement" of the matrix norm $\|A\|$ which in our setting may be greater than $\sqrt{N}$ by the order of magnitude with probability close to one. We remark here that a quantity

$$\max_{|I|=m} \|\mathrm{Proj}_I D\| = \sup_{y \in S^{n-1}} \max_{|I|=m} \|\mathrm{Proj}_I Dy\|,$$

where $m \leq N$ and D is an $N \times n$ random matrix with i.i.d. isotropic log-concave rows, played a crucial role in the paper [1] by Adamczak, Litvak, Pajor and Tomczak-Jaegermann, dealing with the problem of approximating covariance matrix of a log-concave random vector by the sample covariance matrix. In our case, however, the latter quantity is inapplicable as it may not concentrate near $\sqrt{N}$ (even for small m).

First, we prove the required estimate for (1) under the additional assumption that the entries of A are symmetrically distributed (Lemma 12). Then we generalize the result to non-symmetric distributions in Proposition 13.

Let us outline the proof of Lemma 12. A crucial observation (that will be formally justified later) is that there exist finite sets $\mathcal{N}_1 \subset 2B_2^n$ and $\mathcal{N}_2 \subset B_2^n \cap (n^{-1/4} B_\infty^n)$ with $|\mathcal{N}_1| \lesssim n^{\sqrt{n}}$ and $|\mathcal{N}_2| \leq \exp(Cn)$ such that S^{n-1} is a subset of the Minkowski sum $\mathcal{N}_1 + \mathcal{N}_2 + \frac{2}{\sqrt{n}} B_\infty^n$. In this way, estimating the supremum over the unit sphere can be reduced to considering separately

$$\sup_{y \in S} \min_{|I| \geq N - \varepsilon N/3} \|\mathrm{Proj}_I Ay\|,$$

where $S = \mathcal{N}_1, \mathcal{N}_2, \frac{2}{\sqrt{n}} B_\infty^n$. For $S = \mathcal{N}_1, \mathcal{N}_2$, we shall use estimates for individual vectors (Lemmas 7 and 10 below) and then apply the union bound. The cube is treated in Lemma 11.

Lemma 7. *For each $\varepsilon \in (0, 1]$ there is $N_7 = N_7(\varepsilon) > 0$ depending only on ε with the following property: let $N \geq N_7$ and let $X = (X_1, X_2, \ldots, X_N)$ be a random vector of independent variables, each X_i having zero mean and unit variance. Then*

$$\min_{|I| \geq N - \varepsilon N} \|\mathrm{Proj}_I X\| \leq C_7 \sqrt{N}$$

with probability at least $1 - \exp(-c_7 \varepsilon N)$, where $C_7, c_7 > 0$ are universal constants.

Proof. Fix any $\varepsilon \in (0, 1]$ and define N_7 as the smallest positive integer such that

$$\left(\frac{e}{4}\right)^{\varepsilon N} + \exp(-\varepsilon e N/4) \leq \exp(-\varepsilon N/3)$$

for all $N \geq N_7$. Choose any $N \geq N_7$ and let X be as stated above. Set $M = \frac{4}{\varepsilon}$. In view of Markov's inequality,

$$\mathbb{P}\big\{|\{i \leq N : |X_i| \geq \sqrt{M}\}| \geq 4N/M\big\} \leq \binom{N}{\lceil 4N/M \rceil}\Big(\frac{1}{M}\Big)^{\lceil 4N/M \rceil} \leq \Big(\frac{e}{4}\Big)^{\lceil 4N/M \rceil}.$$

Let $\tilde{X} = (\tilde{X}_1, \tilde{X}_2, \ldots, \tilde{X}_N)$ be a vector of truncations of X_i's, with

$$\tilde{X}_i(\omega) = \begin{cases} X_i(\omega), & \text{if } |X_i(\omega)| \leq \sqrt{M}; \\ 0, & \text{otherwise.} \end{cases}$$

Then, from the above estimate,

$$\mathbb{P}\big\{\min_{|I| \geq N - \varepsilon N} \|\mathrm{Proj}_I X\| > \|\tilde{X}\|\big\} \leq \mathbb{P}\big\{|\{i \leq N : |X_i| \geq \sqrt{M}\}| \geq \varepsilon N\big\} \leq \Big(\frac{e}{4}\Big)^{\varepsilon N}.$$

Now, let us estimate the Euclidean norm of $\tilde{X}$ using the Laplace transform. Set $\lambda = \frac{1}{M}$. We have

$$\begin{aligned}
\mathbb{E}\exp(\lambda \|\tilde{X}\|^2) &= \prod_{i=1}^{N} \mathbb{E}\exp(\lambda \tilde{X}_i^2) \\
&= \prod_{i=1}^{N} \Big(1 + \int_1^{\exp(\lambda M)} \mathbb{P}\big\{\exp(\lambda \tilde{X}_i^2) \geq \tau\big\}\, d\tau\Big) \\
&\leq \prod_{i=1}^{N} \Big(1 + \int_1^{e} \mathbb{P}\Big\{\tilde{X}_i^2 \geq \frac{\tau - 1}{e\lambda}\Big\}\, d\tau\Big) \\
&\leq \prod_{i=1}^{N} \Big(1 + e\lambda \mathbb{E}\tilde{X}_i^2\Big) \\
&\leq \big(1 + e\lambda\big)^N \\
&\leq \exp(eN/M).
\end{aligned}$$

Hence,

$$\mathbb{P}\big\{\|\tilde{X}\| \geq \sqrt{2eN}\,\big\} \leq \exp(-eN/M).$$

Finally, using the definition of N_7, we get

$$\begin{aligned}
\mathbb{P}\big\{\min_{|I| \geq N - \varepsilon N} \|\mathrm{Proj}_I X\| > \sqrt{2eN}\,\big\} &\leq \mathbb{P}\big\{\min_{|I| \geq N - \varepsilon N} \|\mathrm{Proj}_I X\| > \|\tilde{X}\|\big\} + \mathbb{P}\big\{\|\tilde{X}\| \geq \sqrt{2eN}\,\big\} \\
&\leq \Big(\frac{e}{4}\Big)^{\varepsilon N} + \exp(-\varepsilon e N/4) \\
&\leq \exp(-\varepsilon N/3). \qquad \square
\end{aligned}$$

Lemma 8. *For every $K > 0$ there is $L_8 = L_8(K) > 0$ depending only on K with the following property: Let $N, n \in \mathbb{N}$, $N \geq n$, and let $A = (a_{ij})$ be an $N \times n$ random matrix with i.i.d. symmetrically distributed entries with unit variance. For each $y = (y_1, y_2, \dots, y_n) \in S^{n-1}$ let $I_y : \Omega \to 2^{\{1,2,\dots,N\}}$ be a random subset of $\{1, 2, \dots, N\}$ defined as*

$$I_y = \Big\{ i \leq N : \sum_{j=1}^{n} a_{ij}^2 y_j^2 \leq 2 \Big\}.$$

Then for every $y \in S^{n-1}$ we have

$$\mathbb{P}\big\{ \|\mathrm{Proj}_{I_y} Ay\| \geq L_8 \sqrt{N} \,\big\} \leq \exp(-KN).$$

Proof. Fix any $K > 0$ and let N, n and $A = (a_{ij})$ be as stated above. Let r_{ij} ($1 \leq i \leq N$, $1 \leq j \leq n$) be Rademacher variables jointly independent with A, and let $\bar{A}$ denote the random $N \times n$ matrix $(r_{ij} a_{ij})$. Then, since a_{ij}'s are symmetrically distributed, for any fixed vector $y = (y_1, y_2, \dots, y_n) \in S^{n-1}$ the distribution of $\|\mathrm{Proj}_{I_y} Ay\|$ is the same as that of $\|\mathrm{Proj}_{I_y} \bar{A}y\|$. Define a subset of (non-random) $N \times n$ matrices:

$$\mathcal{M}_y = \Big\{ B = (b_{ij}) \in \mathbb{R}^{N \times n} : \sum_{j=1}^{n} b_{ij}^2 y_j^2 \leq 2 \text{ for all } i = 1, 2, \dots, N \Big\}$$

and for every $B = (b_{ij}) \in \mathcal{M}_y$ denote by $\bar{B}$ the random matrix $(r_{ij} b_{ij})$. Note that at every point ω of the probability space the matrix $\mathrm{Proj}_{I_y(\omega)} \bar{A}(\omega)$ belongs to $\mathcal{M}_y$. Then, conditioning on a_{ij}'s, we get for every $\tau > 0$:

$$\mathbb{P}\big\{ \|\mathrm{Proj}_{I_y} Ay\| \geq \tau \big\} = \mathbb{P}\big\{ \|\mathrm{Proj}_{I_y} \bar{A}y\| \geq \tau \big\} \leq \sup_{B \in \mathcal{M}_y} \mathbb{P}\big\{ \|\bar{B}y\| \geq \tau \big\}. \tag{2}$$

Note that for each $B \in \mathcal{M}_y$ and $i \leq N$, the i-th coordinate of the vector $\bar{B}y$ satisfies in view of Lemma 4:

$$\mathbb{P}\big\{ |\langle \bar{B}y, e_i \rangle| \geq \tau \big\} \leq 2 \exp\big(-c_4 \tau^2 / 2 \big), \quad \tau > 0.$$

A standard application of the Laplace transform then yields

$$\mathbb{P}\big\{ \|\bar{B}y\| \geq L_8 \sqrt{N} \,\big\} \leq \exp(-KN)$$

for some $L_8 > 0$ depending only on K. This, together with (2), proves the result. □

Lemma 9. *Let ξ be a symmetrically distributed random variable with unit variance. For every $\varepsilon > 0$ and $K > 0$ there is $\delta_9 = \delta_9(\varepsilon, K) > 0$ depending on ε, K and the distribution of ξ with the following property: whenever $N, n \in \mathbb{N}$, $N \geq n$; $A = (a_{ij})$ is an $N \times n$*

random matrix with i.i.d. entries distributed as ξ *and* $y \in S^{n-1}$ *is a vector satisfying* $\|y\|_\infty \le \delta_9$*, we have*

$$\mathbb{P}\{|I_y| \le N - \varepsilon N\} \le \exp(-KN),$$

where I_y *is defined as in Lemma 8.*

Proof. Fix any $K > 0$ and $\varepsilon \in (0, 1]$. In view of Lemma 5, there is $\delta > 0$ such that for all $y = (y_1, y_2, \dots) \in \ell_2$ with $\|y\| = 1$ and $\|y\|_\infty \le \delta$, and for a sequence of independent random variables $a_1, a_2, \dots$ distributed as ξ, we have

$$\mathbb{P}\Big\{\sum_{j=1}^{\infty} a_j^2 y_j^2 > 2\Big\} \le \varepsilon \exp\big(-1 - K/\varepsilon\big).$$

Now, fix $N, n \in \mathbb{N}$ with $N \ge n$ and $y \in S^{n-1}$ with $\|y\|_\infty \le \delta$, and let A be defined as above. Then, using the last estimate, we obtain

$$\begin{aligned}\mathbb{P}\big\{|I_y| \le N - \varepsilon N\big\} &= \mathbb{P}\Big\{\big|\big\{i \le N : \sum_j a_{ij}^2 y_j^2 > 2\big\}\big| \ge \varepsilon N\Big\} \\ &\le \binom{N}{\lceil \varepsilon N \rceil} \Big(\frac{\varepsilon}{e}\Big)^{\lceil \varepsilon N \rceil} \exp(-KN) \\ &\le \exp(-KN). \qquad \square\end{aligned}$$

As an elementary consequence of Lemmas 8 and 9 we get

Lemma 10. *Let* ξ *be a symmetrically distributed random variable with unit variance. For every* $\varepsilon > 0$ *and* $K > 0$ *there are* $\delta_{10} = \delta_{10}(\varepsilon, K) > 0$ *depending on* ε*,* K *and the distribution of* ξ*, and* $L_{10} = L_{10}(K) > 0$ *depending only on* K *such that, whenever* $N, n \in \mathbb{N}$*,* $N \ge n$*;* $A = (a_{ij})$ *is an* $N \times n$ *random matrix with i.i.d. entries distributed as* ξ*, and* $y \in S^{n-1}$ *is a vector satisfying* $\|y\|_\infty \le \delta_{10}$*, we have*

$$\mathbb{P}\Big\{\min_{|I| \ge N - \varepsilon N} \|\mathrm{Proj}_I Ay\| \ge L_{10}\sqrt{N}\Big\} \le \exp(-KN).$$

Lemma 11. *Let* ξ *be a symmetrically distributed random variable with unit variance. For every* $\varepsilon > 0$ *and* $K > 0$ *there are* $n_{11} = n_{11}(\varepsilon, K) \in \mathbb{N}$ *depending on* ε*,* K *and the distribution of* ξ*, and* $L_{11} = L_{11}(K) > 0$ *depending only on* K *such that, whenever* $N \ge n \ge n_{11}$ *and* $A = (a_{ij})$ *is an* $N \times n$ *random matrix with i.i.d. entries distributed as* ξ*, we have*

$$\mathbb{P}\Big\{\min_{|I| \ge N - \varepsilon N} \max_{y \in B_\infty^n} \|\mathrm{Proj}_I Ay\| \ge L_{11}\sqrt{nN}\Big\} \le \exp(-KN).$$

Proof. Fix any $K > 0$ and $\varepsilon > 0$ and define $n_{11} = \lceil \delta_9(\varepsilon, K+1)^{-2} \rceil$, where $\delta_9 > 0$ is taken from Lemma 9. Now, choose any $N \ge n \ge n_{11}$ and let $A = (a_{ij})$ be an $N \times n$ random matrix with i.i.d. entries distributed as ξ. Let V be the set of vertices of the cube $\frac{1}{\sqrt{n}} B_\infty^n = [-\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}]^n$. In view of Lemma 9, any $v \in V$ satisfies

$$\mathbb{P}\{|I_v| \le N - \varepsilon N\} \le \exp\bigl(-(K+1)N\bigr).$$

Next, by Lemma 8, for $L = L_8(K+2) > 0$ we have

$$\mathbb{P}\bigl\{\|\mathrm{Proj}_{I_v} Av\| \ge L\sqrt{N}\bigr\} \le \exp\bigl(-(K+2)N\bigr)$$

for all $v \in V$. Note that for any $u, v \in V$ the random sets I_u and I_v coincide everywhere on Ω. Hence, together with the above estimates, we get

$$\begin{aligned} &\mathbb{P}\Bigl\{\min_{|I| \ge N - \varepsilon N} \max_{v \in V} \|\mathrm{Proj}_I Av\| \ge L\sqrt{N}\Bigr\} \\ &\quad\le \exp\bigl(-(K+1)N\bigr) + \mathbb{P}\Bigl\{\max_{v \in V} \|\mathrm{Proj}_{I_v} Av\| \ge L\sqrt{N}\Bigr\} \\ &\quad\le \exp(-KN). \end{aligned}$$

It remains to note that for any $I \subset \{1, 2, \dots, N\}$ and $y \in B_\infty^n$ we have

$$\|\mathrm{Proj}_I Ay\| \le \sqrt{n} \max_{v \in V} \|\mathrm{Proj}_I Av\|$$

everywhere on Ω. $\square$

In the following statement, we bound the quantity (1) assuming that the matrix entries are symmetrically distributed. As we already mentioned above, to derive an estimate for the supremum over the sphere, we shall embed S^{n-1} into Minkowski sum of a multiple of B_∞^n and two specially chosen finite sets (see (3) in the proof below). This way each vector $y \in S^{n-1}$ can be "decomposed" as a sum of three vectors with particular characteristics. This approach is similar to splitting the unit sphere into sets of "close to sparse" and "far from sparse" vectors introduced in [8] and subsequently used in [15,14].

Lemma 12. *Let ξ be a symmetrically distributed random variable with unit variance, and let $\varepsilon \in (0, 1]$. Then there are $N_{12} = N_{12}(\varepsilon) \in \mathbb{N}$ depending on ε and the distribution of ξ and $w_{12} = w_{12}(\varepsilon) > 0$ depending only on ε such that, whenever $N \ge N_{12}$, $n \le N$ and $A = (a_{ij})$ is an $N \times n$ random matrix with i.i.d. entries distributed as ξ, we have*

$$\mathbb{P}\Bigl\{\sup_{y \in S^{n-1}} \min_{|I| \ge N - \varepsilon N} \|\mathrm{Proj}_I Ay\| \le C_{12}\sqrt{N}\Bigr\} \ge 1 - \exp(-w_{12}N),$$

where $C_{12} > 0$ is a universal constant.

Proof. Fix $\varepsilon \in (0,1]$ and let N_{12} be the smallest integer such that

1) $\lfloor N_{12}^{1/4} \rfloor \delta_{10}(\varepsilon/3, 2C_3) \geq 1$;
2) $N_{12} \geq \max\big(N_7(\varepsilon/3), n_{11}(\varepsilon/3, 1)\big)$;
3) for all $N \geq N_{12}$,

$$(12eN)^{\sqrt{N}} \exp(-c_7 \varepsilon N/3) + e^{-C_3 N} + e^{-N} \leq \exp\big(-\min(c_7\varepsilon/6, C_3/2, 1/2)N\big).$$

Choose $N \geq N_{12}$. Without loss of generality, we can assume that $n = N$. Let A be as stated above.

We say that a vector $y \in \mathbb{R}^N$ is m-sparse if it has at most m non-zero coordinates. It is not difficult to verify, using Lemma 2, that the set of all $\sqrt{N}$-sparse vectors in $2B_2^N$ admits an $N^{-1/2}$-net $\mathcal{N}_1$ of cardinality at most $\binom{N}{\lfloor\sqrt{N}\rfloor}(6\sqrt{N})^{\sqrt{N}} \leq (12eN)^{\sqrt{N}}$. Denote

$$T = \big\{y \in S^{N-1} : \|y\|_\infty \leq 1/\lfloor N^{1/4}\rfloor\big\}.$$

By Lemma 3, there is a finite subset $\mathcal{N}_2 \subset T$ of cardinality at most $\exp(C_3 N)$ such that for any $y \in T$ there is $y' \in \mathcal{N}_2$ with $\|y - y'\|_\infty \leq N^{-1/2}$.

Now, we claim that

$$S^{N-1} \subset \mathcal{N}_1 + \mathcal{N}_2 + \frac{2}{\sqrt{N}} B_\infty^N, \tag{3}$$

i.e. any vector $y = (y_1, y_2, \ldots, y_N) \in S^{N-1}$ can be represented as $y = y^1 + y^2 + y^3$ for some $y^1 \in \mathcal{N}_1$, $y^2 \in \mathcal{N}_2$ and $y^3 \in \frac{2}{\sqrt{N}} B_\infty^N$. Indeed, we can always find a subset $J \subset \{1, 2, \ldots, N\}$ of cardinality $\lfloor\sqrt{N}\rfloor$ such that $|y_j| \leq 1/\lfloor N^{1/4}\rfloor$ whenever $j \notin J$. Denote $r = \sqrt{1 - \|y - \mathrm{Proj}_J y\|^2}$ and $\tilde{y} = \mathrm{Proj}_J y - r|J|^{-1/2} \sum_{j\in J} e_j$. Note that $\tilde{y}$ is $\sqrt{N}$-sparse and has the Euclidean norm at most 2, so there is $y^1 \in \mathcal{N}_1$ such that $\|\tilde{y} - y^1\|_\infty \leq \|\tilde{y} - y^1\| \leq N^{-1/2}$. Next, the vector $y - \tilde{y}$ satisfies $\|y - \tilde{y}\| = 1$ and $\|y - \tilde{y}\|_\infty \leq 1/\lfloor N^{1/4}\rfloor$, i.e. $y - \tilde{y} \in T$. Hence there is $y^2 \in \mathcal{N}_2$ such that $\|y - \tilde{y} - y^2\|_\infty \leq N^{-1/2}$. Finally, for the vector $y^3 = y - y^1 - y^2$ we get

$$\|y - y^1 - y^2\|_\infty \leq \|\tilde{y} - y^1\|_\infty + \|y - \tilde{y} - y^2\|_\infty \leq \frac{2}{\sqrt{N}},$$

so $y^3 \in \frac{2}{\sqrt{N}} B_\infty^N$. This proves (3).

For each $y^1 \in \mathcal{N}_1$, in view of Lemma 7 and the condition $N \geq N_7(\varepsilon/3)$, we have

$$\mathbb{P}\Big\{\min_{|I| \geq N - \varepsilon N/3} \|\mathrm{Proj}_I A y^1\| > 2C_7\sqrt{N}\Big\} \leq \exp(-c_7\varepsilon N/3).$$

Next, for every $y^2 \in \mathcal{N}_2$, Lemma 10 together with the inequalities $\lfloor N^{1/4}\rfloor \delta_{10}(\varepsilon/3, 2C_3) \geq 1$ and $\|y^2\|_\infty \leq 1/\lfloor N^{1/4}\rfloor$ implies that

$$\mathbb{P}\Big\{ \min_{|I|\geq N-\varepsilon N/3} \|\mathrm{Proj}_I A y^2\| \geq L_{10}\sqrt{N} \Big\} \leq \exp(-2C_3 N)$$

for some constant $L_{10} > 0$. Finally, by Lemma 11 and in view of the condition $N \geq n_{11}(\varepsilon/3, 1)$ we have

$$\mathbb{P}\Big\{ \min_{|I|\geq N-\varepsilon N/3} \max_{y\in \frac{1}{\sqrt{N}}B_\infty^N} \|\mathrm{Proj}_I A y\| \geq L_{11}\sqrt{N} \Big\} \leq \exp(-N),$$

where $L_{11} > 0$ is a universal constant. Let $\mathcal{E}$ denote the event

$$\begin{aligned}
\mathcal{E} = \Big\{\omega \in \Omega :\ & \text{for every } y^1 \in \mathcal{N}_1 \text{ there is a set } I_1 = I_1(y^1) \text{ with } |I_1| \geq N - \varepsilon N/3 \\
& \text{such that } \|\mathrm{Proj}_{I_1} A(\omega) y^1\| \leq 2C_7\sqrt{N} \textbf{ AND} \\
& \text{for every } y^2 \in \mathcal{N}_2 \text{ there is a set } I_2 = I_2(y^2) \text{ with } |I_2| \geq N - \varepsilon N/3 \\
& \text{such that } \|\mathrm{Proj}_{I_2} A(\omega) y^2\| \leq L_{10}\sqrt{N} \textbf{ AND} \\
& \text{there is a set } I_3 \text{ with } |I_3| \geq N - \varepsilon N/3 \\
& \text{such that } \max_{y\in \frac{2}{\sqrt{N}}B_\infty^N} \|\mathrm{Proj}_{I_3} A(\omega) y\| \leq 2L_{11}\sqrt{N} \Big\}.
\end{aligned}$$

Then from the above probability estimates and the definition of N_{12} we obtain

$$\mathbb{P}(\mathcal{E}) \geq 1 - (12eN)^{\sqrt{N}} \exp(-c_7\varepsilon N/3) - \exp(-C_3 N) - \exp(-N) \geq 1 - \exp(-w_{12}N),$$

where $w_{12} = \min\big(\frac{c_7\varepsilon}{6}, \frac{C_3}{2}, \frac{1}{2}\big)$.

Finally, take any $\omega \in \mathcal{E}$ and any $y \in S^{N-1}$, and let $y^1 \in \mathcal{N}_1$, $y^2 \in \mathcal{N}_2$ and $y^3 \in \frac{2}{\sqrt{N}}B_\infty^N$ satisfy $y = y^1 + y^2 + y^3$. Then, by the definition of $\mathcal{E}$, there are sets $I_1, I_2, I_3 \subset \{1, 2, \dots, N\}$ with $|I_\ell| \geq N - \varepsilon N/3$ $(\ell = 1, 2, 3)$ such that

$$\begin{aligned}
\|\mathrm{Proj}_{I_1} A(\omega) y^1\| &\leq 2C_7\sqrt{N}; \\
\|\mathrm{Proj}_{I_2} A(\omega) y^2\| &\leq L_{10}\sqrt{N}; \\
\|\mathrm{Proj}_{I_3} A(\omega) y^3\| &\leq 2L_{11}\sqrt{N}.
\end{aligned}$$

Note that the intersection $I = I_1 \cap I_2 \cap I_3$ necessarily satisfies $|I| \geq N - \varepsilon N$, and from the last inequalities we get $\|\mathrm{Proj}_I A(\omega) y\| \leq (2C_7 + L_{10} + 2L_{11})\sqrt{N}$. Since our choice of $y \in S^{N-1}$ and $\omega \in \mathcal{E}$ was arbitrary, we get

$$\begin{aligned}
&\mathbb{P}\Big\{ \sup_{y\in S^{N-1}} \min_{|I|\geq N-\varepsilon N} \|\mathrm{Proj}_I A y\| \leq (2C_7 + L_{10} + 2L_{11})\sqrt{N} \Big\} \\
&\geq \mathbb{P}(\mathcal{E}) \geq 1 - \exp\big(-w_{12}N\big). \qquad \square
\end{aligned}$$

Finally, we can state the main result of the section.

Proposition 13. *Let ξ be a random variable with zero mean and unit variance, and let $\varepsilon \in (0,1]$. Then there are $N_{13} = N_{13}(\varepsilon) \in \mathbb{N}$ depending on ε and the distribution of ξ and $w_{13} = w_{13}(\varepsilon) > 0$ depending only on ε such that, whenever $N \geq N_{13}$, $n \leq N$ and $A = (a_{ij})$ is an $N \times n$ random matrix with i.i.d. entries distributed as ξ, we have*

$$\mathbb{P}\Big\{ \sup_{y \in S^{n-1}} \min_{|I| \geq N - \varepsilon N} \|\mathrm{Proj}_I Ay\| \leq C_{13}\sqrt{N} \Big\} \geq 1 - \exp(-w_{13}N),$$

where $C_{13} > 0$ is a universal constant.

Proof. Fix any $\varepsilon \in (0,1]$ and let ξ' be an independent copy of ξ. Then $\frac{1}{\sqrt{2}}(\xi - \xi')$ is symmetrically distributed and $\mathbb{E}\big(\frac{1}{\sqrt{2}}(\xi - \xi')\big)^2 = 1$. Let N_{12}, w_{12} from Lemma 12 be defined with respect to ε and the distribution of $\frac{1}{\sqrt{2}}(\xi - \xi')$, and let N_{13} be the smallest integer greater than N_{12} such that $\exp(w_{12}N_{13}/2) \geq \frac{4}{3}$. Take any $N \geq N_{13}$ and $n \leq N$ and let A be an $N \times n$ random matrix with i.i.d. entries distributed as ξ, and A' be an independent copy of A. We can find a Borel function $f : \mathbb{R}^{N \times n} \to S^{n-1}$ such that for any $B \in \mathbb{R}^{N \times n}$ we have

$$\min_{|I| \geq N - \varepsilon N} \|\mathrm{Proj}_I Bf(B)\| \geq \sup_{y \in S^{n-1}} \min_{|I| \geq N - \varepsilon N} \|\mathrm{Proj}_I By\| - 1$$

(the term "-1" above allows us to construct a piecewise constant function f, thus avoiding any measurability questions). Then we define a random vector $\tilde{Y} : \Omega \to S^{n-1}$ as $\tilde{Y}(\omega) = f(A(\omega))$. Conditioning on A, we obtain

$$\begin{aligned}
&\mathbb{P}\Big\{ \min_{|I| \geq N - \varepsilon N} \|\mathrm{Proj}_I A\tilde{Y}\| > (\sqrt{2}C_{12} + 2)\sqrt{N} \text{ and } \|A'\tilde{Y}\| \leq 2\sqrt{N} \Big\} \\
&\geq \min_{y \in S^{n-1}} \mathbb{P}\big\{ \|A'y\| \leq 2\sqrt{N} \big\} \, \mathbb{P}\Big\{ \min_{|I| \geq N - \varepsilon N} \|\mathrm{Proj}_I A\tilde{Y}\| > (\sqrt{2}C_{12} + 2)\sqrt{N} \Big\} \\
&\geq \frac{3}{4}\mathbb{P}\Big\{ \min_{|I| \geq N - \varepsilon N} \|\mathrm{Proj}_I A\tilde{Y}\| > (\sqrt{2}C_{12} + 2)\sqrt{N} \Big\},
\end{aligned}$$

where the estimate $\min_{y \in S^{n-1}} \mathbb{P}\big\{ \|A'y\| \leq 2\sqrt{N} \big\} \geq 3/4$ follows from Markov's inequality. Hence, taking into consideration that the entries of $A - A'$ are distributed as $\xi - \xi'$ and using Lemma 12, we get

$$\begin{aligned}
&\mathbb{P}\Big\{ \sup_{y \in S^{n-1}} \min_{|I| \geq N - \varepsilon N} \|\mathrm{Proj}_I Ay\| > (\sqrt{2}C_{12} + 3)\sqrt{N} \Big\} \\
&\leq \mathbb{P}\Big\{ \min_{|I| \geq N - \varepsilon N} \|\mathrm{Proj}_I A\tilde{Y}\| > (\sqrt{2}C_{12} + 2)\sqrt{N} \Big\} \\
&\leq \frac{4}{3}\mathbb{P}\Big\{ \min_{|I| \geq N - \varepsilon N} \|\mathrm{Proj}_I A\tilde{Y}\| > (\sqrt{2}C_{12} + 2)\sqrt{N} \text{ and } \|A'\tilde{Y}\| \leq 2\sqrt{N} \Big\} \\
&\leq \frac{4}{3}\mathbb{P}\Big\{ \min_{|I| \geq N - \varepsilon N} \|\mathrm{Proj}_I (A - A')\tilde{Y}\| > \sqrt{2}C_{12}\sqrt{N} \Big\}
\end{aligned}$$

$$\le \frac{4}{3}\mathbb{P}\Big\{ \sup_{y\in S^{n-1}} \min_{|I|\ge N-\varepsilon N} \|\mathrm{Proj}_I(A-A')y\| > \sqrt{2}C_{12}\sqrt{N}\,\Big\}$$
$$\le \frac{4}{3}\exp(-w_{12}N)$$
$$\le \exp(-w_{12}N/2). \qquad \square$$

4. Matrix truncation and proof of Theorem 1

In the next statement, we compare the n-th largest singular value of a random $N\times n$ matrix A with bounded entries to $s_n(\mathrm{Proj}_I A)$. Obviously,

$$s_n(\mathrm{Proj}_I A) \le s_n(A) \ \text{ for any } I\subset\{1,2,\dots,N\}.$$

We will need an inequality in the opposite direction when $|I|/N \approx 1$. A theorem of Litvak, Pajor, Rudelson and Tomczak-Jaegermann [8, Theorem 3.1] implies that for any $\delta > 1$ and $M > 0$ there are $h > 0$ and $\varepsilon > 0$ depending only on δ and M with the following property: whenever $N \ge \delta n$ and A is an $N\times n$ random matrix with i.i.d. entries with mean zero, variance one and a.s. bounded by M, we have

$$\mathbb{P}\Big\{ \min_{|I|\ge N-\varepsilon N} s_n(\mathrm{Proj}_I A) \ge h\sqrt{N}\,\Big\} \ge 1-2\exp(\varepsilon N).$$

This, together with an upper bound for $s_n(A)$, gives an estimate

$$s_n(A) \le L \min_{|I|\ge N-\varepsilon N} s_n(\mathrm{Proj}_I A)$$

with a large probability, where $L > 0$ depends only on δ and M. However, such an estimate would be insufficient for our needs, and we shall apply a more direct argument to get a stronger relation.

Proposition 14. *Let ξ be a random variable with zero mean such that $|\xi| \le M$ a.s. for some $M > 0$. For any $\eta > 0$ there are $\varepsilon_{14} = \varepsilon_{14}(\eta, M) > 0$ and $N_{14} = N_{14}(\eta, M) \in \mathbb{N}$ (both depending only on η and M) with the following property: whenever $N \ge N_{14}$, $n \le N$ and $A = (a_{ij})$ is an $N\times n$ random matrix with i.i.d. entries distributed as ξ, we have*

$$\mathbb{P}\Big\{ s_n(A) \le \min_{|I|\ge N-\varepsilon_{14}N} s_n(\mathrm{Proj}_I A) + \eta\sqrt{N}\,\Big\} \ge 1-\exp(-\varepsilon_{14}N).$$

Proof. Fix any $\eta > 0$, let $\varepsilon = \varepsilon_{14}(\eta, M)$ be the largest number in $(0, 1]$ satisfying

$$\frac{c_4}{2M^2}\eta^2 \ge \varepsilon\big(1+\ln\frac{6e}{\varepsilon}\big),$$

and $N_{14} \in \mathbb{N}$ be the smallest number such that $N - \lceil N-\varepsilon N\rceil \ge \varepsilon N/2$ for all $N \ge N_{14}$.

Let $N \ge N_{14}$, $n \le N$ and A be an $N \times n$ random matrix defined as above. We shall prove the statement by contradiction. Let us assume that

$$\mathbb{P}\Big\{s_n(A) > \min_{|I|\ge N-\varepsilon N} s_n(\mathrm{Proj}_I A) + \eta\sqrt{N}\Big\} > \exp(-\varepsilon N).$$

Cardinality of the set $T = \big\{I \subset \{1,2,\dots,N\} : |I| = \lceil N-\varepsilon N\rceil\big\}$ can be estimated as

$$|T| \le \binom{N}{\lceil N-\varepsilon N\rceil} \le \Big(\frac{eN}{N-\lceil N-\varepsilon N\rceil}\Big)^{N-\lceil N-\varepsilon N\rceil} \le \Big(\frac{2e}{\varepsilon}\Big)^{\varepsilon N}.$$

Hence, our assumption implies that there is a set $I_0 \in T$ such that

$$\mathbb{P}\big\{s_n(A) > s_n(\mathrm{Proj}_{I_0} A) + \eta\sqrt{N}\big\} > \exp(-\varepsilon N)\Big(\frac{2e}{\varepsilon}\Big)^{-\varepsilon N}. \tag{4}$$

Let $f : \mathbb{R}^{N\times n} \to S^{n-1}$ be a Borel function such that for every $B \in \mathbb{R}^{N\times n}$, $f(B) \in S^{n-1}$ is an eigenvector of $B^T B$ corresponding to its smallest eigenvalue. So, we have $\|Bf(B)\| = s_n(B)$. Then we define a random vector $\tilde{Y} : \Omega \to S^{n-1}$ as $\tilde{Y}(\omega) = f(\mathrm{Proj}_{I_0} A(\omega))$. It is not difficult to see that such a definition implies that $\tilde{Y}$ and a_{ij} ($i \notin I_0$, $1 \le j \le n$) are jointly independent. Hence,

$$\begin{aligned}\mathbb{P}\big\{s_n(A) > s_n(\mathrm{Proj}_{I_0} A) + \eta\sqrt{N}\big\} &\le \mathbb{P}\big\{\|A\tilde{Y}\| > \|\mathrm{Proj}_{I_0} A\tilde{Y}\| + \eta\sqrt{N}\big\}\\ &\le \mathbb{P}\big\{\|\mathrm{Proj}_{\{1,2,\dots,N\}\setminus I_0} A\tilde{Y}\| > \eta\sqrt{N}\big\}\\ &\le \sup_{y\in S^{n-1}} \mathbb{P}\big\{\|\mathrm{Proj}_{\{1,2,\dots,N\}\setminus I_0} Ay\| > \eta\sqrt{N}\big\}.\end{aligned}$$

Now, for every $y = (y_1, y_2, \dots, y_n) \in S^{n-1}$, Lemma 4 and the standard procedure with the Laplace transform give for $\lambda = \frac{c_4}{2M^2}$:

$$\begin{aligned}&\mathbb{P}\big\{\|\mathrm{Proj}_{\{1,2,\dots,N\}\setminus I_0} Ay\| > \eta\sqrt{N}\big\}\\ &\quad= \mathbb{P}\Big\{\sum_{i\notin I_0}\Big(\sum_{j=1}^n a_{ij}y_j\Big)^2 > \eta^2 N\Big\}\\ &\quad\le \frac{\Big(\mathbb{E}\exp\big(\lambda\big(\sum_{j=1}^n a_{1j}y_j\big)^2\big)\Big)^{N-\lceil N-\varepsilon N\rceil}}{\exp(\lambda\eta^2 N)}\\ &\quad= \exp(-\lambda\eta^2 N)\Big(1 + \int_1^\infty \mathbb{P}\Big\{\Big|\sum_{j=1}^n a_{1j}y_j\Big| \ge \sqrt{\ln\tau/\lambda}\Big\}\,d\tau\Big)^{N-\lceil N-\varepsilon N\rceil}\\ &\quad\le \exp(-\lambda\eta^2 N)\Big(1 + 2\int_1^\infty \exp\Big(-\frac{c_4\ln\tau}{\lambda M^2}\Big)\,d\tau\Big)^{N-\lceil N-\varepsilon N\rceil}\end{aligned}$$

$$= \exp(-\lambda\eta^2 N)\, 3^{N-\lceil N-\varepsilon N\rceil}$$
$$\le \exp\bigl(-\lambda\eta^2 N + \varepsilon N\ln 3\bigr).$$

Together with (4), the last estimate implies

$$-\lambda\eta^2 + \varepsilon\ln 3 > -\varepsilon - \varepsilon\ln\frac{2e}{\varepsilon}.$$

However, this contradicts our choice of ε. Thus, the initial assumption was wrong, and the statement is proved. $\square$

Let ξ be a random variable with zero mean. Then for any $M > 0$ we call the variable

$$\xi\chi_{\{|\xi|\le M\}} - \mathbb{E}(\xi\chi_{\{|\xi|\le M\}})$$

the centered M-truncation of ξ. Here, $\chi_{\{|\xi|\le M\}}$ is the indicator of the event $\bigl\{\omega\in\Omega : |\xi(\omega)|\le M\bigr\}$.

Denote $\tilde\xi_M = \xi\chi_{\{|\xi|\le M\}} - \mathbb{E}(\xi\chi_{\{|\xi|\le M\}})$ and $\theta_M = \xi - \tilde\xi_M = \xi\chi_{\{|\xi|>M\}} + \mathbb{E}(\xi\chi_{\{|\xi|\le M\}})$. Obviously, $\mathbb{E}\tilde\xi_M = \mathbb{E}\theta_M = 0$ and $|\tilde\xi_M| \le 2M$ everywhere on Ω for any $M > 0$. Further, if the second moment of ξ is bounded then

$$\mathbb{E}\tilde\xi_M^2 = \mathbb{E}(\xi\chi_{\{|\xi|\le M\}})^2 - \bigl(\mathbb{E}(\xi\chi_{\{|\xi|\le M\}})\bigr)^2 \longrightarrow \mathbb{E}\xi^2 \quad\text{and}$$
$$\mathbb{E}{\theta_M}^2 = \mathbb{E}\xi^2 - 2\mathbb{E}\bigl(\xi^2\chi_{\{|\xi|\le M\}}\bigr) + \mathbb{E}\tilde\xi_M^2 \longrightarrow 0 \quad\text{when } M\to\infty.$$

Theorem 15. *Let ξ be a random variable with zero mean and unit variance. For any $M > 0$ and $\eta > 0$ there are $N_{15}\in\mathbb{N}$ depending on M, η and the distribution of ξ, and $w_{15} > 0$ depending only on M and η with the following property: Let $N \ge N_{15}$, $n \le N$ and let $A = (a_{ij})$ be an $N\times n$ random matrix with i.i.d. entries distributed as ξ. Further, let $\tilde A$ be an $N\times n$ matrix with the entries $\tilde a_{ij} = a_{ij}\chi_{\{|a_{ij}|\le M\}} - \mathbb{E}(a_{ij}\chi_{\{|a_{ij}|\le M\}})$ and denote $\theta = \xi\chi_{\{|\xi|>M\}} + \mathbb{E}(\xi\chi_{\{|\xi|\le M\}})$. Then*

$$\mathbb{P}\bigl\{s_n(A) \ge s_n(\tilde A) - \eta\sqrt{N} - C_{13}\sqrt{N\mathbb{E}\theta^2}\bigr\} \ge 1 - \exp(-w_{15}N).$$

Proof. Fix any $M > 0$ and $\eta > 0$ and let θ be as above. We will assume that $\mathbb{P}\{\theta = 0\} < 1$; otherwise the truncation leaves the variable unchanged and there is nothing to prove. Let $N_{14} = N_{14}(\eta, 2M)$ and $\varepsilon = \varepsilon_{14}(\eta, 2M)$ be taken from Proposition 14. Let also N_{13} and w_{13} be defined as in Proposition 13 with respect to ε and the distribution of the "normalized tail" $\theta/\sqrt{\mathbb{E}\theta^2}$. Now, let N_{15} be the smallest integer greater than $\max(N_{14}, N_{13})$ such that for all $N \ge N_{15}$ we have

$$\exp(-\varepsilon N) + \exp(-w_{13}N) \le \exp\bigl(-\min(\varepsilon/2, w_{13}/2)N\bigr).$$

Take any $N \geq N_{15}$, $n \leq N$, and let A, $\tilde{A}$ be as stated above. By Proposition 14, we have

$$\mathbb{P}\Big\{ s_n(\tilde{A}) > \min_{|I| \geq N - \varepsilon N} s_n(\mathrm{Proj}_I \tilde{A}) + \eta\sqrt{N} \Big\} \leq \exp(-\varepsilon N),$$

and, by Proposition 13,

$$\mathbb{P}\Big\{ \sup_{y \in S^{n-1}} \min_{|I| \geq N - \varepsilon N} \|\mathrm{Proj}_I(A - \tilde{A})y\| > C_{13}\sqrt{N\mathbb{E}\theta^2} \Big\} \leq \exp(-w_{13}N).$$

Combining the two relations, we get

$$\begin{aligned}
&\mathbb{P}\Big\{ s_n(A) < s_n(\tilde{A}) - \eta\sqrt{N} - C_{13}\sqrt{N\mathbb{E}\theta^2} \Big\} \\
&\quad \leq \mathbb{P}\Big\{ s_n(\tilde{A}) > \min_{|I| \geq N - \varepsilon N} s_n(\mathrm{Proj}_I \tilde{A}) + \eta\sqrt{N} \Big\} \\
&\qquad + \mathbb{P}\Big\{ s_n(A) < \min_{|I| \geq N - \varepsilon N} s_n(\mathrm{Proj}_I \tilde{A}) - C_{13}\sqrt{N\mathbb{E}\theta^2} \Big\} \\
&\quad \leq \exp(-\varepsilon N) \\
&\qquad + \mathbb{P}\Big\{ \exists y \in S^{n-1} : \min_{|I| \geq N - \varepsilon N} s_n(\mathrm{Proj}_I \tilde{A}) - \|Ay\| > C_{13}\sqrt{N\mathbb{E}\theta^2} \Big\} \\
&\quad \leq \exp(-\varepsilon N) \\
&\qquad + \mathbb{P}\Big\{ \exists y \in S^{n-1} : \min_{|I| \geq N - \varepsilon N} (\|\mathrm{Proj}_I \tilde{A}y\| - \|\mathrm{Proj}_I Ay\|) > C_{13}\sqrt{N\mathbb{E}\theta^2} \Big\} \\
&\quad \leq \exp(-\varepsilon N) + \exp(-w_{13}N) \\
&\quad \leq \exp\big(-\min(\varepsilon/2, w_{13}/2)N\big). \qquad \square
\end{aligned}$$

Proof of Theorem 1. Let $\{a_{ij}\}$ $(1 \leq i, j < \infty)$ be a two-dimensional array of i.i.d. random variables with zero mean and unit variance and let $(N_m)_{m=1}^{\infty}$ be an integer sequence satisfying $m/N_m \longrightarrow z$ for some $z \in (0,1)$. Recall that for every $m \in \mathbb{N}$, A_m denotes the random $N_m \times m$ matrix with entries a_{ij} $(1 \leq i \leq N_m,\ 1 \leq j \leq m)$. The Marčenko–Pastur law (see Theorem 6 and Remark 2) implies that

$$\limsup_{m \to \infty} \frac{s_m(A_m)}{\sqrt{N_m}} \leq 1 - \sqrt{z} \quad \text{almost surely.}$$

Thus, it suffices to prove the lower estimate

$$\liminf_{m \to \infty} \frac{s_m(A_m)}{\sqrt{N_m}} \geq 1 - \sqrt{z} \quad \text{a.s.}$$

Now, choose arbitrary $\eta > 0$ and let $M > 0$ be such that

$$\begin{aligned}
&\mathbb{E}\big(a_{11}\chi_{\{|a_{11}| \leq M\}} - \mathbb{E}(a_{11}\chi_{\{|a_{11}| \leq M\}})\big)^2 \geq (1 - \eta)^2 \quad \text{and} \\
&\mathbb{E}\big(a_{11}\chi_{\{|a_{11}| > M\}} + \mathbb{E}(a_{11}\chi_{\{|a_{11}| \leq M\}})\big)^2 \leq \eta^2.
\end{aligned}$$

For every $m \in \mathbb{N}$, let $\tilde{A}_m$ be the $N_m \times m$ matrix of truncated and centered variables $\tilde{a}_{ij} = a_{ij}\chi_{\{|a_{ij}| \le M\}} - \mathbb{E}(a_{ij}\chi_{\{|a_{ij}| \le M\}})$ $(1 \le i \le N_m, 1 \le j \le m)$. Theorem 15 and the conditions on the sequence $(N_m)_{m=1}^{\infty}$ imply that there are $m_0 \in \mathbb{N}$ and $w > 0$ such that for all $k \ge m_0$

$$\mathbb{P}\big\{s_m(A_m) \ge s_m(\tilde{A}_m) - (1 + C_{13})\eta\sqrt{N_m} \text{ for all } m \ge k\big\} \ge 1 - \sum_{m=k}^{\infty} \exp(-wN_m),$$

where the quantity on the right-hand side goes to 1 as k tends to infinity. Hence, we obtain

$$\mathbb{P}\Big\{\liminf_{m\to\infty} \frac{s_m(A_m)}{\sqrt{N_m}} \ge \liminf_{m\to\infty} \frac{s_m(\tilde{A}_m)}{\sqrt{N_m}} - (1 + C_{13})\eta\Big\} = 1.$$

On the other hand, the theorem of Bai and Yin [4] implies that

$$\lim_{m\to\infty} \frac{s_m(\tilde{A}_m)}{\sqrt{N_m}} \ge (1 - \eta)(1 - \sqrt{z}\,) \quad \text{a.s.}$$

Thus, we come to the estimate

$$\liminf_{m\to\infty} \frac{s_m(A_m)}{\sqrt{N_m}} \ge (1 - \eta)(1 - \sqrt{z}\,) - (1 + C_{13})\eta \quad \text{a.s.}$$

Since $\eta > 0$ was arbitrary, this proves the result. $\square$

Acknowledgment

The results of this paper shall form a part of author's PhD thesis. The author would like to thank his supervisor Dr. Nicole Tomczak-Jaegermann for her support and for valuable suggestions on the text.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at http://dx.doi.org/10.1016/j.aim.2015.07.020.

References

[1] R. Adamczak, A. Litvak, A. Pajor, N. Tomczak-Jaegermann, Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles, J. Amer. Math. Soc. 23 (2) (2010) 535–561.
[2] Z. Bai, J.W. Silverstein, Spectral Analysis of Large Dimensional Random Matrices, 2nd edition, Springer, New York, 2010.
[3] Z.D. Bai, J.W. Silverstein, Y.Q. Yin, A note on the largest eigenvalue of a large-dimensional sample covariance matrix, J. Multivariate Anal. 26 (2) (1988) 166–168.

[4] Z.D. Bai, Y.Q. Yin, Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix, Ann. Probab. 21 (3) (1993) 1275–1294.

[5] F. Götze, A. Tikhomirov, The circular law for random matrices, Ann. Probab. 38 (4) (2010) 1444–1491.

[6] V. Koltchinskii, S. Mendelson, Bounding the smallest singular value of a random matrix without concentration, Int. Math. Res. Not. IMRN (2015), http://dx.doi.org/10.1093/imrn/rnv096.

[7] R. Latala, Some estimates of norms of random matrices, Proc. Amer. Math. Soc. 133 (5) (2005) 1273–1282.

[8] A.E. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, Smallest singular value of random matrices and geometry of random polytopes, Adv. Math. 195 (2) (2005) 491–523.

[9] A.E. Litvak, S. Spektor, Quantitative version of a Silverstein's result, in: Geometric Aspects of Functional Analysis, in: Lecture Notes in Math., vol. 2116, 2014, pp. 335–340.

[10] V.A. Marčenko, L.A. Pastur, Distribution of eigenvalues in certain sets of random matrices, Mat. Sb. (N.S.) 72 (114) (1967) 507–536 (in Russian).

[11] L. Pastur, M. Shcherbina, Eigenvalue Distribution of Large Random Matrices, American Mathematical Society, 2011.

[12] G. Pisier, The Volumes of Convex Bodies and Banach Space Geometry, Cambridge University Press, Cambridge, 1989.

[13] M. Rudelson, R. Vershynin, The least singular value of a random square matrix is $O(n^{-1/2})$, C. R. Math. Acad. Sci. Paris 346 (15–16) (2008) 893–896.

[14] M. Rudelson, R. Vershynin, The Littlewood–Offord problem and invertibility of random matrices, Adv. Math. 218 (2) (2008) 600–633.

[15] M. Rudelson, R. Vershynin, Smallest singular value of a random rectangular matrix, Comm. Pure Appl. Math. 62 (12) (2009) 1707–1739.

[16] M. Rudelson, R. Vershynin, Non-asymptotic theory of random matrices: extreme singular values, in: Proceedings of the International Congress of Mathematicians, vol. III, Hindustan Book Agency, New Delhi, 2010, pp. 1576–1602.

[17] J.W. Silverstein, On the weak limit of the largest eigenvalue of a large-dimensional sample covariance matrix, J. Multivariate Anal. 30 (2) (1989) 307–311.

[18] N. Srivastava, R. Vershynin, Covariance estimation for distributions with $2+\varepsilon$ moments, Ann. Probab. 41 (5) (2013) 3081–3111.

[19] T. Tao, V. Vu, Random matrices: the circular law, Commun. Contemp. Math. 10 (2) (2008) 261–307.

[20] K.E. Tikhomirov, The smallest singular value of random rectangular matrices with no moment assumptions on entries, Israel J. Math. (2015), in press, arXiv:1409.7975.

[21] R. Vershynin, Introduction to the non-asymptotic analysis of random matrices, in: Y. Eldar, G. Kutyniok (Eds.), Compressed Sensing: Theory and Applications, Cambridge University Press, 2012, pp. 210–268.

[22] P. Yaskov, Lower bounds on the smallest eigenvalue of a sample covariance matrix, Electron. Commun. Probab. 19 (83) (2014) 1–10.

[23] Y.Q. Yin, Limiting spectral distribution for a class of random matrices, J. Multivariate Anal. 20 (1) (1986) 50–68.

[24] Y.Q. Yin, Z.D. Bai, P.R. Krishnaiah, On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix, Probab. Theory Related Fields 78 (4) (1988) 509–521.