On the interval of fluctuation of the singular values of random matrices

Olivier Guédon[1] Alexander E. Litvak[2] Alain Pajor
Nicole Tomczak-Jaegermann[3]

Abstract

Let A be a matrix whose columns $X_1, \ldots, X_N$ are independent random vectors in $\mathbb{R}^n$. Assume that the tails of the 1-dimensional marginals decay as $\mathbb{P}(|\langle X_i, a\rangle| \geq t) \leq Ct^{-p}$ uniformly in $a \in S^{n-1}$ and $i \leq N$. Then for $p > 4$ we prove that with high probability $A/\sqrt{n}$ has the Restricted Isometry Property (RIP) provided that Euclidean norms $|X_i|$ are concentrated around $\sqrt{n}$. We also show that the covariance matrix is well approximated by the empirical covariance matrix. As consequence, we establish a good rate of convergence when N is proportional to n, provided that $\max_i |X_i| \leq C\sqrt{n}$ with high probability. Moreover, we give precise estimates for both problems when the decay is of the type $\exp(-t^\alpha)$, with $\alpha \in (0, 2]$.

AMS 2010 Classification: primary 60B20, 46B06, 15B52; secondary 46B09, 60D05

Keywords: random matrices, norm of random matrices, approximation of covariance matrices, compressed sensing, restricted isometry property, log-concave random vectors, concentration inequalities, deviation inequalities, heavy tails, spectrum, singular values, order statistics.

---

[1]The research is partially supported by the ANR GeMeCoD, ANR 2011 BS01 007 01
[2]Research partially supported by the E.W.R. Steacie Memorial Fellowship.
[3]This author holds the Canada Research Chair in Geometric Analysis.

1 Introduction and main results

Fix positive integers n, N and let A be an $n \times N$ random matrix whose columns $X_1, \dots, X_N$ are independent random vectors in $\mathbb{R}^n$. For a subset $I \subset \{1, \dots, N\}$ of cardinality m, denote by A^I the $n \times m$ matrix whose columns are $X_i, i \in I$. We are interested in estimating the interval of fluctuation of the spectrum of some matrices related to A when the random vectors X_i, $i \le N$ have heavy tails; firstly, uniform estimates of the spectrum of $(A^I)^\top A^I$ which is the set of squares of the singular values of A^I, where I runs over all subsets of cardinality m for some fixed parameter m and secondly estimates for the spectrum of $AA^\top$. The first problem is related to the notion of Restricted Isometry Property (RIP) with m a parameter of sparsity whereas the second is about approximation of a covariance matrix by empirical covariance matrices.

These questions have been substantially developed over recent years and many papers devoted to these notions were written. In this work, we say that a random vector X satisfies the tail behavior $\mathbf{H}(\phi)$, if

$$\mathbf{H}(\phi): \quad \forall a \in S^{n-1}\ \forall t > 0 \quad \mathbb{P}\left(|\langle X, a\rangle| \ge t\right) \le 1/\phi(t) \tag{1}$$

for a certain function ϕ and we assume that X_i satisfies $\mathbf{H}(\phi)$ for all $i \le N$. We will focus on two choices of the function ϕ, namely $\phi(t) = t^p$, with $p > 4$, which means heavy tail behavior for marginals, and $\phi(t) = (1/2)\exp(t^\alpha)$, with $\alpha \in (0, 2]$, which corresponds to an exponential power type tail behavior and makes the link to the known subexponential case ($\alpha = 1$, see [2, 3]).

The concept of the Restricted Isometry Property was introduced in [10] in order to study an exact reconstruction problem by ℓ_1 minimization algorithm, classical in compressed sensing. Although it provided only a sufficient condition for the reconstruction, it played a decisive role in the development of the theory, and it is still an important property. This is mostly due to the fact that a large number of important classes of random matrices have RIP. It is also noteworthy that the problem of reconstruction can be reformulated in terms of convex geometry, namely in terms of neighborliness of the symmetric convex hull of $X_1, \dots, X_N$, as was shown in [12].

Let us recall the intuition of RIP (for the definition see (6) below). For an $n \times N$ matrix T and $1 \le m \le N$, the *isometry constant of order* m of T is the parameter $0 < \delta_m(T) < 1$ which says that the square of Euclidean norms $|Tz|$ and $|z|$ are approximately equal, up to $1 + \delta_m(T)$, for all m-sparse vectors $z \in \mathbb{R}^N$ (that is, $|\mathrm{supp}(z)| \le m$). Equivalently, this means that for every $I \subset \{1, \ldots, N\}$ with $|I| \leq m$, the spectrum of $(T^I)^\top T^I$ is contained in the interval $[1 - \delta_m(T), 1 + \delta_m(T)]$. In particular when $\delta_m(T) < \theta$ for small θ, then the squares of singular values of the matrices T^I belong to $[1-\theta, 1+\theta]$. Note that in compressed sensing for the reconstruction of vectors by ℓ_1 minimization, one does not need RIP for all $\theta > 0$ (see [12] and [11]). The RIP contains implicitly a normalization, in particular it implies that the Euclidean norms of the columns belong to an interval centered around one.

Let A be an $n \times N$ random matrix whose columns are $X_1, \ldots, X_N$. In view of the example of matrices with i.i.d. entries, centered and with variance one, for which $\mathbb{E}|X_i|^2 = n$, we normalized the matrix by considering $A/\sqrt{n}$ and we introduce the concentration function

$$P(\theta) := \mathbb{P}\left(\max_{i \leq N} \left|\frac{|X_i|^2}{n} - 1\right| \geq \theta\right). \tag{2}$$

Untill now the only known cases of random matrices satisfying a RIP were the cases of subgaussian [9, 10, 12, 22] and subexponential [4] matrices. Our first main theorem says that matrices we consider have the RIP of order m, with "large" m of the form $m = n\psi(n/N)$ with ψ depending on ϕ and possibly on other parameters. In particular, when N is proportional to n, then m is proportional to n. We present a simplified version of our result, for the detailed version see Theorem 3.1 below.

Theorem 1.1 *Let $0 < \theta < 1$. Let A be a random $n \times N$ matrix whose columns $X_1, \ldots, X_N$ are independent and satisfy hypothesis $\mathbf{H}(\phi)$ for some ϕ. Assume that n, N are large enough. Then there exists a function ψ depending on ϕ and θ such that with high probability (depending on the concentration function $P(\theta)$) the matrix $A/\sqrt{n}$ has RIP of order $m = [n\psi(n/N)]$ and $\delta_m(A/\sqrt{n}) \leq \theta$.*

The second problem we investigate goes back to a question of Kannan, Lovász and Simonovits (KLS). Let X_i's and A be as above and assume additionally that X_i's are identically distributed as a centered random vector X. KLS question asks how fast the empirical covariance matrix $U := (1/N)AA^\top$ converges to the covariance matrix $\Sigma := (1/N)\mathbb{E}AA^\top = \mathbb{E}U$. Of course this depends on assumptions on X. In particular, is it true that with high probability the operator norm $\|U - \Sigma\| \leq \varepsilon\|\Sigma\|$ for N being proportional to n? Originally this was asked for log-concave random vectors but the general question of approximating the covariance matrix by sample covariance matrices is an important subject in Statistics as well as on its own right. The corresponding question in Random Matrix Theory is about the limit behavior of smallest and largest singular values. In the case of Wishart matrices, that is when the coordinates of X are i.i.d. centered random variables of variance one, the Bai-Yin theorem [6] states that under assumption of boundedness of fourth moments the limits of minimal and maximal singular numbers of U are $(1 \pm \sqrt{\beta})^2$ as $n, N \to \infty$ and $n/N \to \beta \in (0,1)$. Moreover, it is known [7, 28] that boundedness of fourth moment is necessary in order to have the convergence of the largest singular value. The asymptotic non-limit behavior (also called "non-asymptotic" in Statistics), i.e., sharp upper and lower bounds for singular values in terms of n and N, when n and N are sufficiently large, was studied in several works. To keep the notation more compact and clear we put

$$M := \max_{i \le N} |X_i|, \qquad S := \sup_{a \in S^{n-1}} \left| \frac{1}{N} \sum_{i=1}^{N} \left(\langle X_i, a \rangle^2 - \mathbb{E}\langle X_i, a \rangle^2 \right) \right|. \tag{3}$$

Note that if $\mathbb{E}\langle X, a\rangle^2 = 1$ for every $a \in S^{n-1}$ (that is, X is isotropic), then the bound $S \le \varepsilon$ is equivalent to the fact that the singular values of U belong to the interval $[1-\varepsilon, 1+\varepsilon]$. For Gaussian matrices it is known ([13, 30]) that with probability close to one

$$S \le C\sqrt{n/N}, \tag{4}$$

where C is a positive absolute constant. In [2, 3] the same estimate was obtained for a large class of random matrices, which in particular did not require that entries of the columns are independent, or that X_i's are identically distributed. In particular this solved the original KLS problem. More precisely, (4) holds with high probability under the assumptions that the X_i's satisfy hypothesis $\mathbf{H}(\phi)$ with $\phi(t) = e^t/2$ and that $M \le C(Nn)^{1/4}$ with high probability. Both conditions hold for log-concave random vectors.

Until recent time, quite strong conditions on the tail behavior of the one dimensional marginals of the X_i were imposed, typically of subexponential type. Of course, in view of Bai-Yin theorem, it is a natural question whether one can replace the function $\phi(t) = e^t/2$ by the function $\phi(t) = e^{t^\alpha}/2$ with $\alpha \in (0,1)$ or $\phi(t) = t^p$, for $p \ge 4$. The first attempt in this direction was done in [32], where the bound $S \le C(p,K)(n/N)^{1/2-2/p}(\ln\ln n)^2$ was obtained for every $p > 4$ provided that $M \leq K\sqrt{n}$. Clearly, $\ln \ln n$ is a "parasitic" term, which, in particular, does not allow to solve the KLS problem with N proportional to n. This problem was solved in [23, 29] under strong assumptions and in particular when $M \leq K\sqrt{n}$ and X has i.i.d. coordinates with bounded p-th moment with $p > 4$. Very recently, in [24], the "right" upper bound $S \leq C(n/N)^{1/2}$ was proved for $p > 8$ provided that $M \leq C(Nn)^{1/4}$. The methods used in [24] play an influential role in the present paper.

The problems of estimating the smallest and the largest singular values are quite different. One expects weaker assumption for estimating the smallest singular value. This already appeared in the work [29] and was pushed further in [18] and in [31]. See also related work [19].

In this paper we solve the KLS problem for $4 < p \leq 8$, in Theorem 1.2. Our argument works also in other cases and makes the bridge between the known cases $p > 8$ and the exponential case.

Theorem 1.2 *Let $X_1, \ldots, X_N$ be independent random vectors in $\mathbb{R}^n$ satisfying hypothesis $\mathbf{H}(\phi)$ with $\phi(t) = t^p$ for some $p \in (4, 8]$. Let $\varepsilon \in (0, 1)$ and $\gamma = p - 4 - 2\varepsilon > 0$. Then*

$$S \leq C \left(\left(\frac{M^2}{n} \right) \left(\frac{n}{N} \right) + C(p, \varepsilon) \left(\frac{n}{N} \right)^{\gamma/p} \right),$$

with probability larger than $1 - 8e^{-n} - 2\varepsilon^{-p/2} \max\{N^{-3/2}, n^{-(p/4-1)}\}$.

In particular, if N is proportional to n and M^2/n is bounded by a constant with high probability, which is the case for large classes of random vectors, then with high probability

$$S \leq C \left(n/N\right)^{\gamma/p}.$$

Let X have i.i.d. coordinates distributed as a centered random variable with finite p-th moment. Then by Rosenthal's inequality ([27], see also [16] and Lemma 6.3 below), X satisfies hypothesis $\mathbf{H}(\phi)$ with $\phi(t) = t^p$. Let X_1, ..., X_N be independent random vectors distributed as X. It is known ([7], [28], see also [21] for a quantitative version) that when N is proportional to n and in the absence of fourth moment, $M^2/n \to \infty$ as $n \to \infty$. Hence, concerning KLS question, an upper bound for S involving M^2/n is interesting only when $p \geq 4$ and therefore the only case left open is the case $p = 4$.

The main novelty of our proof is a delicate analysis of behavior of norms of submatrices, namely quantities A_k and B_k, $k \le N$, defined in (5) below. This analysis is done in Theorem 2.1, which is in the heart of the technical part of the paper and it will be presented in the next section. The estimates for B_k are responsible for RIP, Theorem 1.1, while the estimates for A_k are responsible for KLS problem, Theorem 1.2.

As usual in this paper C, C_0, C_1, ..., c, c_0, c_1, ... always denote absolute positive constants whose values may vary from line to line.

The paper is organized as follows. In Section 2, we formulate the main technical result. For the reader convenience, we postpone its proof till Section 5. In Section 3, we discuss the results on RIP. The fully detailed formulation of the main result in this direction is Theorem 3.1, while Theorem 1.1 is its very simplified corollary. In Section 4, we prove Theorem 1.2 as a consequence of Theorem 4.4. The case $p > 8$ and the exponential cases are proved in Theorem 4.6 using the same argument. Symmetrization and formulas for sums of the k smallest order statistics of independent non-negative random variables with heavy tails allow to reduce the problem on hand to estimates for A_k. In the last Section 6, we discuss optimality of the results.

An earlier version of the main results of this paper was announced in [14].

Acknowledgment. A part of this research was performed while the authors were visiting at several universities. Namely, the first named author visited University of Alberta at Edmonton in April 2013 and the second and the forth named author visited University Paris-Est at Paris in June 2013 and June 2014. The authors would like to thank all these universities for their support and hospitality.

2 Norms of submatrices

We start with a few general preliminaries and notations. We denote by B_2^n and S^{n-1} the standard unit Euclidean ball and the unit sphere in $\mathbb{R}^n$ and by $|\cdot|$ and $\langle \cdot, \cdot \rangle$ the corresponding Euclidean norm and inner product. Given a set $E \subset \{1, ..., N\}$, $|E|$ denotes its cardinality and B_2^E denotes the unit Euclidean ball in $\mathbb{R}^E$, with the convention $B_2^{\emptyset} = \{0\}$.

A standard volume argument implies that for every integer n and for every $\varepsilon \in (0, 1)$ there exists an ε-net $\Lambda \subset B_2^n$ of B_2^n of cardinality not exceeding

$(1+2/\varepsilon)^n$; that is, for every $x \in B_2^n$, $\min_{y\in\Lambda} |x-y| < \varepsilon$. In particular, if $\varepsilon \le 1/2$ then the cardinality of Λ is not larger than $(2.5/\varepsilon)^n$.

By $\mathcal{M}$ we denote the class of increasing functions $\phi : [0,\infty) \to [0,\infty)$ such that the function $\ln \phi\,(1/\sqrt{x})$ is convex on $(0,\infty)$. The examples of such functions considered in this paper are $\phi(x) = x^p$ for some $p > 0$ and $\phi(x) = (1/2)\exp(x^\alpha)$ for some $\alpha > 0$.

Recall that the hypothesis $\mathbf{H}(\phi)$ has been defined in the introduction by (1). Note that this hypothesis is satisfied if

$$\sup_{a\in S^{n-1}} \mathbb{E}\, \phi(|\,\langle X, a\rangle\,|) \le 1.$$

For $k \le N$ and random vectors X_1, ..., X_N in $\mathbb{R}^n$ we define A_k and B_k by

$$A_k := \sup_{\substack{a\in S^{N-1}\\ |\mathrm{supp}(a)|\le k}} \left| \sum_{i=1}^N a_i X_i \right|, \quad B_k^2 := \sup_{\substack{a\in S^{N-1}\\ |\mathrm{supp}(a)|\le k}} \left| \left| \sum_{i=1}^N a_i X_i \right|^2 - \sum_{i=1}^N a_i^2 |X_i|^2 \right|. \quad (5)$$

We would like to note that A_k is supremum of norms of submatrices consisting of k columns of A, while B_k is related to a concentration.

Recall also a notation from the introduction $M = \max_{i\le N} |X_i|$.

We formulate now the main technical result, Theorem 2.1, which is the key result for both bounds for A_k and for B_k. We postpone its proof to Section 5.

Theorem 2.1 *Let $X_1, \ldots, X_N$ be independent random vectors in $\mathbb{R}^n$ satisfying hypothesis $\mathbf{H}(\phi)$ for some ϕ. Let $p > 4$, $\sigma \in (2, p/2)$, $\alpha \in (0, 2]$, $t > 0$, and $\lambda \ge 1$. For $k \le N$ we define M_1, β and C_ϕ in two cases.*
Case 1. *$\phi(x) = x^p$. We assume that $\lambda \le p$ and we let $C_\phi = e^4$,*

$$M_1 := C_1(\sigma, \lambda, p)\sqrt{k}\left(\frac{N}{k}\right)^{\sigma/p} \quad \textit{and} \quad \beta := C_2(\sigma,\lambda) N^{-\lambda} + C_3(\sigma, \lambda)\frac{N^2}{t^p},$$

where

$$C_1(\sigma, \lambda, p) = 32\, e^4 \sqrt{\frac{\sigma+\lambda}{1+\lambda/2}} \left(\frac{2p}{p-2\sigma}\right)^{1+2\sigma/p} \left(\frac{\sigma+\lambda}{\sigma-2}\right)^{2\sigma/p} (20e)^{\sigma/p},$$

$$C_2(\sigma, \lambda) := \left(\frac{2(\sigma+\lambda)}{5e(\sigma-2)}\right)^{\lambda} \frac{1}{2\lambda - 1} \quad \textit{and} \quad C_3(\sigma,\lambda) := \frac{(\sigma+\lambda)^p}{4(2(\sigma-2))^p}.$$

Case 2. *$\phi(x) = (1/2)\exp(x^\alpha)$. We assume that $\lambda \geq 2$ and we let $C_\phi = C^{1/\alpha}$, where C is an absolute positive constant,*

$$M_1 := (C\lambda)^{1/\alpha}\sqrt{k}\left(\ln\frac{2N}{k} + \frac{1}{\alpha}\right)^{1/\alpha}$$

and

$$\beta := \frac{1}{(10N)^\lambda}\exp\left(-\frac{\lambda k^{\alpha/2}}{(3.5\ \ln(2k))^{2\alpha}}\right) + \frac{N^2}{2\exp((2t)^\alpha)}.$$

In both cases we also assume that $\beta < 1/32$. Then with probability at least $1 - \sqrt{\beta}$ one has

$$A_k \leq (1 - 4\sqrt{\beta})^{-1}\left(M + 2\sqrt{C_\phi\, t\, M} + M_1\right)$$

and

$$B_k^2 \leq (1 - 4\sqrt{\beta})^{-2}\left(4\sqrt{\beta}M^2 + (8C_\phi\, t + M_1)\, M + 2M_1^2\right).$$

We would like to emphasize that A_k and B_k are of different nature. In particular, Theorem 2.1 in the case $\phi(t) = t^p$ has to be applied with different choices of the parameter σ. We summarize those choices in the following remark.

Remark. In the case $\phi(x) = x^p$ we will use the following two choices for σ:
1. Choosing $\sigma = p/4$ and assuming $p > 8$ we get

$$M_1 \leq C\sqrt{\frac{p}{\lambda}}\sqrt{\frac{p}{p-8}}\sqrt{k}\left(\frac{N}{k}\right)^{1/4}$$

and

$$\beta \leq \left(\frac{2(p+4\lambda)}{5eN(p-8)}\right)^\lambda \frac{1}{2\lambda - 1} + \frac{N^2(p+4\lambda)^p}{4(2t(p-8))^p}.$$

2. Choosing $\sigma = 2 + \varepsilon$ with $\varepsilon \leq \min\{1, (p-4)/4\}$, we get

$$M_1 \leq C\left(\frac{p}{p-4}\right)^{1+(4+2\varepsilon)/p}\left(\frac{\lambda}{\varepsilon}\right)^{2(2+\varepsilon)/p}\sqrt{k}\left(\frac{N}{k}\right)^{(2+\varepsilon)/p}$$

and

$$\beta = \left(\frac{2(3+\lambda)}{5e\varepsilon N}\right)^\lambda \frac{1}{2\lambda - 1} + \frac{N^2(3+\lambda)^p}{4(2\varepsilon t)^p}.$$

3 Restricted Isometry Property

We need more definitions and notations.

Let T be an $n \times N$ matrix and let $1 \le m \le N$. The isometry constant of T is defined as the smallest number $\delta_m = \delta_m(T)$ so that

$$(1-\delta_m)|z|^2 \le |Tz|^2 \le (1+\delta_m)|z|^2 \tag{6}$$

holds for all vectors $z \in \mathbb{R}^N$ with $|\mathrm{supp}(z)| \le m$. For $m = 0$, we put $\delta_0(T) = 0$. Let $\delta \in (0,1)$. The matrix T is said to satisfy the Restricted Isometry Property of order m with parameter δ, in short $\mathrm{RIP}_m(\delta)$, if $0 \le \delta_m(T) \le \delta$.

Recall that a vector $z \in \mathbb{R}^N$ is called m-sparse if $|\mathrm{supp}(z)| \le m$. The subset of m-sparse unit vectors in $\mathbb{R}^N$ is denoted by

$$U_m = U_m(\mathbb{R}^N) := \{z \in \mathbb{R}^N \ : \ |z| = 1, |\mathrm{supp}(z)| \le m\}.$$

Let X_1, ..., X_N be random vectors in $\mathbb{R}^n$ and let A be the $n \times N$ matrix whose columns are the X_i's. By the definition of B_m (see (5)) we clearly have

$$\max_{i \le N} \left| \frac{|X_i|^2}{n} - 1 \right| = \delta_1\left(\frac{A}{\sqrt{n}}\right) \le \delta_m\left(\frac{A}{\sqrt{n}}\right) = \sup_{z \in U_m} \left| \frac{|Az|^2}{n} - 1 \right|$$
$$\le \frac{B_m^2}{n} + \max_{i \le N} \left| \frac{|X_i|^2}{n} - 1 \right|. \tag{7}$$

Thus, in order to have a good bound on $\delta_m\left(A/\sqrt{n}\right)$ we require a strong concentration of each $|X_i|$ around $\sqrt{n}$ and we need to estimate B_m.

To control the concentration of $|X_i|$ we consider the function $P(\theta)$, defined in the introduction by (2). Note that this function estimates the concentration of the maximum. Therefore, when it is small, we have much better concentration of each $|X_i|$ around $\sqrt{n}$.

We are now ready to state the main result about RIP. Theorem 1.1, announced in the introduction, is a very simplified form of it.

Theorem 3.1 *Let $1 \le n \le N$. Let $X_1, \ldots, X_N$ be random vectors in $\mathbb{R}^n$ satisfying hypothesis $\mathbf{H}(\phi)$ for some $\phi \in \mathcal{M}$ and let $P(\cdot)$ be as in (2). Let $\theta \in (0,1)$. Assume that ϕ satisfies one of the two cases.*
Case 1. *Let $p > 4$ and $\phi(x) = x^p$. Let $\varepsilon \le \min\{1, (p-4)/4\}$. Assume that*

$$2^8/(\varepsilon\,\theta) \le N \le c\,\theta\,(c\,\varepsilon\,\theta)^{p/2}\; n^{p/4}$$

and set

$$m = \left[C(\theta,\varepsilon,p)\, n \left(\frac{N}{n}\right)^{-2(2+\varepsilon)/(p-4-2\varepsilon)} \right] \quad \textit{and} \quad \beta = \frac{4}{3e^2\,\varepsilon^2\,N^2} + \frac{5^p\,N^2}{4(2c\,\varepsilon\,\theta)^p\,n^{p/2}},$$

where

$$C(\theta,\varepsilon,p) = c\left(\frac{p-4}{p}\right)^{2(p+4+2\varepsilon)/(p-4-2\varepsilon)} \varepsilon^{4(2+\varepsilon)/(p-4-2\varepsilon)}\,\theta^{2p/(p-4-2\varepsilon)}, \tag{8}$$

c and C are absolute positive constants.

Case 2. *Let $\alpha \in (0,2]$ and $\phi(x) = (1/2)\exp(x^\alpha)$. Assume that*

$$\max\left\{2^{1/\alpha}, 4/\theta\right\} \le N \le c\,\theta\, \exp\left((1/2)\left(c\,\theta\sqrt{n}\right)^\alpha\right)$$

and set

$$m = \left[C^{-2/\alpha}\,\theta^2\, n \left(\ln(C^{2/\alpha}\,N/(\theta^2\,n))\right)^{-2/\alpha} \right]$$

and

$$\beta = \frac{1}{100N^2}\exp\left(\frac{-2m^{\alpha/2}}{(3.5\ln(2m))^{2\alpha}}\right) + \frac{N^2}{2}\exp\left(-c\left(\theta\sqrt{n}\right)^\alpha\right),$$

where c and C are absolute positive constants.
Then in both cases the matrix $A/\sqrt{n}$ has RIP of order m satisfying

$$\mathbb{P}\left(\delta_m(A/\sqrt{n}) \le \theta\right) \ge 1 - \sqrt{\beta} - P(\theta/2).$$

Remarks. 1. Note that for instance in case 1, the constraint $N \le c(\theta,\varepsilon,p)n^{p/4}$ is not important because for $N \gg n^{p/4}$ one has

$$m = \left[C(\theta,\varepsilon,p)\, n \left(\frac{N}{n}\right)^{-2(2+\varepsilon)/(p-4-2\varepsilon)} \right] = 0.$$

A similar remark is valid in the second case.
2. In most applications $P(\theta) \to 0$ very fast as $n, N \to \infty$. For example, for so-called isotropic log-concave random vectors it follows from results of Paouris ([25, 26], see also [17, 15] or Lemma 3.3 of [4]). As another example consider the model when X_i's are i.i.d. and moreover the coordinates of X_1 are i.i.d. random variables distributed as a random variable ξ. In the case when ξ is of variance one and has finite p-th moment, $p > 4$, then by

Rosenthal's inequality $P(\theta)$ is well bounded (for a precise bound see Corollary 6.4 below, see also Proposition 1.3 of [29]). Another case is when ξ is the Weibull random variable of variance one, that is consider ξ_0 such that $\mathbb{P}\left(|\xi_0| > t\right) = \exp\left(-t^{\alpha}\right)$ for $\alpha \in (0, 2]$ and let $\xi = \xi_0/\sqrt{\mathbb{E}\xi_0^2}$. By Lemma 3.4 from [4] (see also Theorem 1.2.8 in [11]), $P(\theta)$ satisfies (34) below.

3. Taking ε in the Case 1 of order $(p-4)^2/\ln(N/n)$ and assuming that it satisfies the condition of the theorem, we observe that in Case 1

$$m = \left[C(\theta, p)\, n \left(\frac{N}{n}\right)^{-4/(p-4)} \left(\ln \frac{N}{n}\right)^{-8/(p-4)} \right].$$

Proof. We first pass to the subset Ω_0 of our initial probability space where

$$\max_{i \le N} \left| \frac{|X_i|^2}{n} - 1 \right| \le \theta/2.$$

Note that by (2) the probability of this event is at least $1 - P(\theta/2)$ and if this event occurs then we also have

$$\max_{i \le N} |X_i| \le 3\sqrt{n}/2.$$

We will apply Theorem 2.1 with $k = m$, $t = \theta\sqrt{n}/(100 C_\phi)$, where C_ϕ is the constant from Theorem 2.1. Additionally we assume that $\beta \le 2^{-9}\theta^2$ and $M_1 \le t$. Then with probability at least $1 - \sqrt{\beta} - P(\theta/2)$ we have

$$B_m^2 \le (16\sqrt{\beta} + \theta/4) n \le \theta\, n/2.$$

Together with (7) this proves $\delta_m(A/\sqrt{n}) \le \theta$. Thus we only need to check when the estimates for β and M_1 are satisfied.

Case 1. $\phi(x) = x^p$**.** We start by proving the estimate for M_1. We let $\sigma = 2 + \varepsilon$, $\varepsilon \le \min\{1, (p-4)/4\}$ and $\lambda = 2$. Then by Theorem 2.1 (see also the Remark following it), for some absolute constant C we have

$$M_1 \le C \left(\frac{p}{p-4}\right)^{1+(4+2\varepsilon)/p} \left(\frac{1}{\varepsilon}\right)^{2(2+\varepsilon)/p} \sqrt{m} \left(\frac{N}{m}\right)^{(2+\varepsilon)/p}.$$

Therefore the estimate $M_1 \le c\,\theta\sqrt{n}$ with $c = 1/(100e^4)$ is satisfied provided that

$$m = \left[C(\theta, \varepsilon, p)\, n \left(\frac{N}{n}\right)^{-2(2+\varepsilon)/(p-4-2\varepsilon)} \right],$$

with $C(\theta, \varepsilon, p)$ defined in (8) and the absolute constants properly adjusted.

Now we estimate the probability. From Theorem 2.1 (and the Remark following it), with our choice of t and λ we have

$$\beta \leq \frac{4}{3e^2\, \varepsilon^2\, N^2} + \frac{5^p\, N^2}{4(2c\,\varepsilon\,\theta)^p\, n^{p/2}} \leq 2^{-9}\theta^2$$

provided that $2^8/(\varepsilon\,\theta) \leq N \leq 2^{-4}\theta\,(0.4c\,\varepsilon\,\theta)^{p/2}\, n^{p/4}$. This completes the proof of the first case.

Case 2. $\phi(x) = (1/2)\exp(x^\alpha)$**.** As in the first case we start with the condition $M_1 \leq t$. We choose $\lambda = 4$. Note that $N/m \geq 2^{1/\alpha}$ as $N \geq 2^{1/\alpha}n$. Therefore for some absolute constant C,

$$M_1 \leq \sqrt{m}\,(C\ln(2N/m))^{1/\alpha}.$$

Therefore the condition $M_1 \leq t$ is satisfied provided that

$$m \leq C_1^{-2/\alpha}\,\theta^2\, n\left(\ln(C_1^{2/\alpha}\, N/(\theta^2\, n))\right)^{-2/\alpha}$$

for an absolute positive constant C_1. This justifies the choice of m.

Now we estimate the probability. From Theorem 2.1 with our choice of t and λ we have

$$\beta \leq \frac{1}{100N^2}\exp\left(\frac{-2m^{\alpha/2}}{(3.5\ln(2m))^{2\alpha}}\right) + \frac{N^2}{2}\exp\left(-c\left(\theta\sqrt{n}\right)^\alpha\right),$$

provided that $4/\theta \leq N \leq 2^{-5}\theta\exp\left(c\,(\theta\sqrt{n})^\alpha\right)$. This completes the proof. □

4 Approximating the covariance matrix

The following technical lemma emphasizes the role of the parameter A_k in estimates of the distance between the covariance matrix and the empirical one. This role was first recognized in [8] and [2]. Other versions of the lemma appeared in [4, 5].

We use below the symmetrization method as in [24]. For a sequence of real numbers $(s_i)_i$ we denote by $(s_i^*)_i$ a non-increasing rearrangement of $(|s_i|)_i$.

Lemma 4.1 *Let $1 \leq k < N$ and $X_1, \ldots, X_N$ be independent vectors in $\mathbb{R}^n$. Let $p \geq 2$, $\alpha \in (0,2]$. Let ϕ be either $\phi(t) = t^p$ in which case we set $C_\phi = 8N^{2/\min(p,4)}$ and assume*

$$\forall 1 \leq i \leq N \quad \forall a \in S^{n-1} \quad \mathbb{E}\,|\langle X_i, a\rangle|^p \leq 1,$$

or $\phi(t) = (1/2)\exp(t^\alpha)$ in which case we assume that X_i's satisfy hypothesis $\mathbf{H}(\phi)$ and set $C_\phi = 8\sqrt{C_\alpha N}$, where $C_\alpha = (2/\alpha)\,\Gamma(5/\alpha)$, $\Gamma(\cdot)$ is the Gamma function. Then, for every $A, Z > 0$,

$$\sup_{a\in S^{n-1}} \left| \sum_{i=1}^N (\langle X_i, a\rangle^2 - \mathbb{E}\langle X_i, a\rangle^2) \right| \leq 2A^2 + 6\sqrt{n}\, Z + C_\phi$$

with probability larger than

$$1 - 4\exp(-n) - 4\mathbb{P}(A_k > A) - 4 \times 9^n \sup_{a\in S^{n-1}} \mathbb{P}\left(\Big(\sum_{i>k} (\langle X_i, a\rangle^*)^4 \Big)^{1/2} > Z \right).$$

The term involving Z in the upper bound will be bounded later using general estimates in Lemma 4.3. Thus Lemma 4.1 clearly stresses the fact that in order to estimate the distance between the covariance matrix and the empirical one, it will remain to estimate A_k, to get A.

Proof: Let $\Lambda \subset \mathbb{R}^n$ be an $(1/4)$-net of the unit Euclidean ball in the Euclidean metric of cardinality not greater than 9^n. Let $(\varepsilon_i)_{1\leq i\leq N}$ be i.i.d. ± 1 Bernoulli random variables of parameter $1/2$. By Hoeffding's inequality, for every $t > 0$ and every $(s_i)_{1\leq i\leq N} \in \mathbb{R}^N$,

$$\mathbb{P}_{(\varepsilon_i)}\left(\Big| \sum_{i=1}^N \varepsilon_i s_i \Big| \geq t (\sum_{i=1}^N |s_i|^2)^{1/2} \right) \leq 2\exp(-t^2/2).$$

Fix an arbitrary $1 \leq k < N$. For every $(s_i)_{1\leq i\leq N} \in \mathbb{R}_+^N$ there exists a permutation π of $\{1, \ldots, N\}$ such that

$$\Big| \sum_{i=1}^N \varepsilon_i s_i \Big| \leq \sum_{i=1}^k s_i^* + \Big| \sum_{i=k+1}^N \varepsilon_{\pi(i)} s_i^* \Big|.$$

Also, it is easy to check using (5) that for any $a \in S^{n-1}$ and any $I \subset \{1, \ldots, N\}$ with $|I| \leq k$, $\sum_{i\in I} \langle X_i, a\rangle^2 \leq A_k^2$, Thus, for every $a \in S^{n-1}$,

$$\mathbb{P}_{(\varepsilon_i)}\left(\Big|\sum_{i=1}^N \varepsilon_i \langle X_i, a\rangle^2\Big| \le A_k^2 + t\Big(\sum_{i=k+1}^N (\langle X_i, a\rangle^*)^4\Big)^{1/2}\right) \ge 1 - 2\exp(-t^2/2).$$

Using a union bound argument indexed by Λ and Lemma 5.3 (below) we get that

$$\mathbb{P}_{(\varepsilon_i)}\left(\sup_{a\in S^{n-1}}\Big|\sum_{i=1}^N \varepsilon_i \langle X_i, a\rangle^2\Big| \le 2\Big[A_k^2 + t\sup_{a\in\Lambda}\Big(\sum_{i=k+1}^N (\langle X_i, a\rangle^*)^4\Big)^{1/2}\Big]\right)$$

$$\ge 1 - 2\times 9^n \exp(-t^2/2).$$

Using again a union bound argument and the triangle inequality to estimate the probability that the (X_i) satisfy

$$\sup_{a\in\Lambda}\Big(\sum_{i=k+1}^N (\langle X_i, a\rangle^*)^4\Big)^{1/2} > Z,$$

and choosing $t = 3\sqrt{n}$ (so that $2\cdot 9^n \exp(-t^2/2) \le e^{-n}$) we get that

$$\sup_{a\in S^{n-1}}\Big|\sum_{i=1}^N \varepsilon_i \langle X_i, a\rangle^2\Big| \le 2A^2 + 6\sqrt{n}\,Z$$

with probability larger than

$$1 - e^{-n} - \mathbb{P}(A_k > A) - 9^n \sup_{a\in S^{n-1}} \mathbb{P}\left(\Big(\sum_{i=k+1}^N (\langle X_i, a\rangle^*)^4\Big)^{1/2} > Z\right).$$

Now we transfer the result from Bernoulli random variables to centered random variables (see [20], Section 6.1). By the triangle inequality, for every $s, t > 0$, one has

$$m(s)\,\mathbb{P}\left(\sup_{a\in S^{n-1}}\Big|\sum_{i=1}^N \big(\langle X_i, a\rangle^2 - \mathbb{E}\langle X_i, a\rangle^2\big)\Big| > s + t\right)$$

$$\le 2\mathbb{P}\left(\sup_{a\in S^{n-1}}\Big|\sum_{i=1}^N \varepsilon_i \langle X_i, a\rangle^2\Big| > t\right)$$

where $m(s) = \inf_{a\in S^{n-1}} \mathbb{P}\left(\left|\sum_{i=1}^N \left(\langle X_i, a\rangle^2 - \mathbb{E}\langle X_i, a\rangle^2\right)\right| \le s\right)$.

To conclude the proof it is enough to find s so that $m(s) \ge 1/2$. To this end we will use a general Lemma 4.2 (below). First consider $\phi(t) = t^p$. For $a \in S^{n-1}$, set $Z_i = |\langle X_i, a\rangle|^2$ and $q = p/2$. Then by Lemma 4.2 we have $m(s) \ge 1/2$ for $s = 4N^{1/r}$ and $r = \min(p/2, 2)$. Now consider $\phi(t) = (1/2)\exp(t^\alpha)$. Then for every $a \in S^{n-1}$ and every $i \le N$ using hypothesis $\mathbf{H}(\phi)$ we have

$$\mathbb{E}|\langle X_i, a\rangle|^4 \le 2\int_0^\infty t^4 \exp(-t^\alpha)\, dt = \frac{2}{\alpha}\Gamma\left(\frac{5}{\alpha}\right) := C_\alpha.$$

Given $a \in S^{n-1}$, set $Z_i = |\langle X_i, a\rangle|^2/\sqrt{C_\alpha}$. Then $\mathbb{E}Z_i^2 \le 1$. Applying again Lemma 4.2 (with $q = 2$), we observe that $m(s) \ge 1/2$ for $s = 4\sqrt{C_\alpha N}$. This completes the proof. □

It remains to prove the following general lemma. For convenience of the argument above, we formulate this lemma using two powers q and r rather than just one.

Lemma 4.2 *Let $q \ge 1$ and $Z_1, \ldots, Z_N$ be independent non-negative random variables satisfying*

$$\forall 1 \le i \le N \qquad \mathbb{E}Z_i^q \le 1.$$

Let $r = \min(q, 2)$, then

$$\forall z \ge 4N^{1/r} \qquad \mathbb{P}\left(\left|\sum_{i=1}^N \left(Z_i - \mathbb{E}Z_i\right)\right| \le z\right) \ge \frac{1}{2}.$$

Proof: By definition of r, we have for all $i = 1, \ldots, N$, $\mathbb{E}Z_i^r \le 1$. Since the Z_i's are independent, we deduce by a classical symmetrization argument that

$$\mathbb{E}\left|\sum_{i=1}^N \left(Z_i - \mathbb{E}Z_i\right)\right| \le 2\mathbb{E}\mathbb{E}_{(\varepsilon_i)}\left|\sum_{i=1}^N \varepsilon_i Z_i\right| \le 2\mathbb{E}\left(\sum_{i=1}^N Z_i^2\right)^{1/2} \le 2\mathbb{E}\left(\sum_{i=1}^N Z_i^r\right)^{1/r}$$

since $r \in [1, 2]$. From $\mathbb{E}Z_i^r \le 1$, we get that

$$\mathbb{E}\left|\sum_{i=1}^N \left(Z_i - \mathbb{E}Z_i\right)\right| \le 2\mathbb{E}\left(\sum_{i=1}^N Z_i^r\right)^{1/r} \le 2\left(\sum_{i=1}^N \mathbb{E}Z_i^r\right)^{1/r} \le 2N^{1/r}.$$

By Markov's inequality we get

$$\mathbb{P}\left(\Big|\sum_{i=1}^{N}\left(Z_i-\mathbb{E}Z_i\right)\Big|\geq 4N^{1/r}\right)\leq\frac{1}{2},$$

and since $z\geq 4N^{1/r}$, this implies the required estimate. □

The following lemma is standard (cf. Lemma 5.8 in [20], which however contains a misprint).

Lemma 4.3 *Let $q>0$ and let $Z_1,\ldots,Z_N$ be independent non-negative random variables satisfying*

$$\forall 1\leq i\leq N\quad\forall t\geq 1\quad\mathbb{P}(Z_i\geq t)\leq 1/t^q.$$

Then, for every $s>1$, with probability larger than $1-s^{-k}$, one has

$$\sum_{i=k}^{N}Z_i^*\leq\begin{cases}\frac{(2es)^{1/q}}{1-q}N^{1/q}k^{1-1/q} & \text{if } 0<q<1\\ 2esN\ln\left(\frac{eN}{k}\right) & \text{if } q=1\\ \frac{12q(es)^{1/q}}{q-1}N & \text{if } q>1.\end{cases}$$

Proof: Assume first that $0<q\leq 1$. It is clear that

$$\forall 1\leq i\leq N\qquad\mathbb{P}(Z_i^*>t)\leq\binom{N}{i}t^{-iq}\leq(Ne/it^q)^i,$$

where we used the inequality $\binom{N}{i}\leq(Ne/i)^i$. Thus if $eNt^{-q}\leq 1$, then

$$\mathbb{P}(\sup_{i\geq k}i^{1/q}Z_i^*>t)\leq\sum_{i\geq k}(Ne/t^q)^i=\left(\frac{eN}{t^q}\right)^k(1-eNt^{-q})^{-1}.$$

Therefore if $eNt^{-q}\leq 1/2$, then $\mathbb{P}(\sup_{i\geq k}i^{1/q}Z_i^*>t)\leq(2eNt^{-q})^k$. Since the inequality is trivially true if $eNt^{-q}\geq 1/2$, it is proved for every $t>0$. Therefore for $q<1$ we have

$$\sum_{i=k}^{N}Z_i^*\leq t\sum_{i=k}^{\infty}i^{-1/q}\leq t\left(k^{-1/q}-\frac{k^{1-1/q}}{1-1/q}\right)\leq\frac{t}{1-q}\,k^{1-1/q}$$

with probability larger than $1-(2eN/t^q)^k$. Choosing $t=(2esN)^{1/q}$, we obtain the estimate in the case $0<q<1$.

For $q=1$ we have

$$\sum_{i=k}^{N} Z_i^* \leq t\sum_{i=k}^{N} i^{-1} \leq t\left(\frac{1}{k}+\ln(N/k)\right) \leq t\,\ln(eN/k)$$

with probability larger than $1-(2eN/t)^k$. To obtain the desire estimate choose $t=2esN$.

Now assume that $q>1$. Set $\ell=\lceil\log_2 k\rceil$. The same computation as before for the scale $(2^{i/q})$ instead of $(i^{1/q})$ gives that

$$\mathbb{P}(\sup_{i\geq\ell} 2^{i/q} Z_{2^i}^* > t) \leq \sum_{i\geq\ell} (Net^{-q})^{2^i} \leq \left(2eNt^{-q}\right)^{2^\ell}.$$

Note also that

$$\mathbb{P}(k^{1/q} Z_k^* > t) \leq (Net^{-q})^k.$$

Thus

$$\sum_{i=k}^{N} Z_i^* \leq kZ_k^* + \sum_{i=\ell}^{\lceil\log_2 N\rceil} 2^i Z_{2^i}^* \leq t\left(k^{1-1/q}+(4N)^{1-1/q}/(2^{1-1/q}-1)\right)$$

$$\leq t\left(k^{1-1/q}+\frac{2q}{1-q}(4N)^{1-1/q}\right) \leq t\,\frac{3q}{1-q}(4N)^{1-1/q}$$

with probability larger than $(Net^{-q})^k+(2Net^{-q})^k$. Thus, taking $t=(4esN)^{1/q}$, we obtain

$$\mathbb{P}\left(\sum_{i=k}^{N} Z_i^* \leq \frac{12q(es)^{1/q}}{q-1}N\right) \geq 1-s^{-k}.$$

□

We are now ready to tackle the problem of approximating the covariance matrix by the empirical covariance matrices, under hypothesis $H(\phi)$ with $\phi(t)=t^p$. As our proof works for all $p>4$, we also include the case $p>8$ originally solved in [24] (under additional assumption on $\max_i|X_i|$). For clarity, we split the result into two theorems. The case $4<p\leq 8$ has been stated as Theorem 1.2 in the Introduction.

Before we state our result, let us remark that $p > 2$ is a necessary condition. Indeed, let $(e_i)_{1\le i\le n}$ be an orthonormal basis of $\mathbb{R}^n$ and let Z be a random vector such that $Z = \sqrt{n}e_i$ with probability $1/n$. The covariance matrix of Z is the identity I. Let A be an $n \times N$ random matrix with independent columns distributed as Z. Note that if $\|\frac{1}{N}AA^\top - I\| < 1$ with some probability, then $AA^\top$ is invertible with the same probability. It is known (coupon collector's problem) that $N \sim n \log n$ is needed to have $\{Z_i : i \le N\} = \{\sqrt{n}e_i : i \le n\}$ with probability, say, $1/2$. Thus for vector Z, the hypothesis $H(\phi)$, $\phi(t) = t^2$ is satisfied but $N \sim n \log n$ is needed for the covariance matrix to be well approximated by the empirical covariance matrices with probability $1/2$.

Theorem 4.4 *Let $4 < p \le 8$ and $\phi(t) = t^p$. Let $X_1, \dots, X_N$ be independent random vectors in $\mathbb{R}^n$ satisfying hypothesis $\mathbf{H}(\phi)$. Let $\varepsilon \le \min\{1, (p-4)/4\}$ and $\gamma = p - 4 - 2\varepsilon$. Then with probability larger than*

$$1 - 8e^{-n} - 2\varepsilon^{-p/2} \max\left\{N^{-3/2}, n^{-(p/4-1)}\right\}$$

one has

$$\sup_{a\in S^{n-1}} \left|\frac{1}{N}\sum_{i=1}^{N}(\langle X_i, a\rangle^2 - \mathbb{E}\langle X_i, a\rangle^2)\right| \le C\left(\frac{1}{N}\max_{i\le N}|X_i|^2 + C(p,\varepsilon)\left(\frac{n}{N}\right)^{\gamma/p}\right),$$

where

$$C(p,\varepsilon) = (p-4)^{-1/2}\,\varepsilon^{-4(2+\varepsilon)/p}.$$

and C is an absolute constant.

An immediate consequence of this theorem is the following corollary.

Corollary 4.5 *Under assumptions of Theorem 4.6, assuming additionally that $\max_i |X_i|^2 \le Cn^{\gamma/p}N^{1-\gamma/p}$ with high probability, we have with high probability*

$$\sup_{a\in S^{n-1}} \left|\frac{1}{N}\sum_{i=1}^{N}(\langle X_i, a\rangle^2 - \mathbb{E}\langle X_i, a\rangle^2)\right| \le C_1\, C(p,\varepsilon)\left(\frac{n}{N}\right)^{\gamma/p},$$

where C and C_1 are absolute positive constants.

Theorem 4.6 *There exists a universal positive constant C such that the following holds. Let $p > 8$, $\alpha \in (0, 2]$. Let ϕ and C_ϕ be either $\phi(t) = t^p$ and $C_\phi = C$ or $\phi(t) = (1/2)\exp(t^\alpha)$ and $C_\phi = (C/\alpha)^{2.5/\alpha}$. Let $X_1, \dots, X_N$ be independent random vectors in $\mathbb{R}^n$ satisfying hypothesis $\mathbf{H}(\phi)$. In the case $\phi(t) = t^p$ we also define*

$$p_0 = 8e^{-n} + 2\left(\frac{3p-8}{6(p-8)}\right)^{p/2} N^{-(p-8)/8}\, n^{-p/8}$$

and in the case $\phi(t) = (1/2)\exp(t^\alpha)$, we assume $N \geq (4/\alpha)^{8/\alpha}$ and define

$$p_0 = 8e^{-n} + \frac{1}{(10N)^4}\exp\left(\frac{4n^{\alpha/2}}{(3.5\ln(2n))^{2\alpha}}\right) + \frac{N^2}{2\exp((2nN)^{\alpha/4})}.$$

Then in both cases with probability larger than $1 - p_0$ one has

$$\sup_{a \in S^{n-1}} \left| \frac{1}{N}\sum_{i=1}^{N}(\langle X_i, a\rangle^2 - \mathbb{E}\langle X_i, a\rangle^2)\right| \leq \frac{C}{N}\max_{i \leq N}|X_i|^2 + C_\phi\sqrt{\frac{n}{N}}.$$

As our argument works in all cases we prove both theorems together.

Proof of Theorems 4.4 and 4.6. We first consider the case $\phi = t^p$. Note that in this case

$$\mathbb{E}\,|\langle X_i, a\rangle|^4 \leq 1 + \int_1^\infty \mathbb{P}\left(|X_i|^4 > t\right)\, dt \leq 1 + \int_1^\infty 4s^{3-p}\, ds = \frac{p}{p-4}.$$

Thus, by Lemma 4.1 it is enough to estimate $A^2 + \sqrt{n}\, Z + \sqrt{p/(p-4)}\,\sqrt{N}$ and the corresponding probabilities. We choose $k = n$.

In the case $\phi = t^p$ we apply Lemma 4.3 with $Z_i = |\langle X_i, a\rangle|^4$, $i \leq N$, $q = p/4 > 1$ and $s = 9e$. It gives

$$\mathbb{P}\left(\left(\sum_{i>k}(\langle X_i, a\rangle^*)^4\right)^{1/2} > Z\right) \leq (9e)^{-n},$$

for

$$Z = \sqrt{\frac{12q}{q-1}}(es)^{1/2q}\sqrt{N} = \sqrt{\frac{12p}{p-4}}(3e)^{4/p}\sqrt{N}.$$

Now we estimate A_n, using Theorem 2.1.

Case 1: $4 < p \le 8$ (Theorem 4.4). We apply Theorem 2.1 (and the Remark following it), with $\sigma = 2 + \varepsilon$, where $\varepsilon < (p-4)/4$, $\lambda = 3$ and $t = 3N^{2/p}n^{\delta}$ for $\delta = 1/2 - 2/p$. Then

$$M_1 \le C(p, \varepsilon)\sqrt{n}\,(N/n)^{(2+\varepsilon)/p},$$

where

$$C_0(p, \varepsilon) = C\left(\frac{1}{p-4}\right)^{(p-4-2\varepsilon)/p}\left(\frac{1}{\varepsilon}\right)^{2(2+\varepsilon)/p}$$

and

$$\beta \le \frac{1}{5}\left(\frac{12}{5e\varepsilon N}\right)^3 + \frac{1}{4(p-4)^p n^{\delta p}} \le \varepsilon^{-p}\max\{N^{-3}, n^{-\delta p}\} \le 1/64$$

provided that n is large enough. Then, using $\delta = 1/2 - 2/p$, we obtain

$$A_n^2 \le C\left(\max_{i\le N}|X_i|^2 + N^{2/p}n^{\delta}\max_{i\le N}|X_i| + C_0^2(p, \varepsilon)\,n\,(N/n)^{2(2+\varepsilon)/p}\right)$$

$$\le 2C\left(\max_{i\le N}|X_i|^2 + C_0^2(p, \varepsilon)\,n\,(N/n)^{2(2+\varepsilon)/p}\right).$$

Combining all estimates and noticing that $(p-4)^{-\gamma} < 2$, we obtain that the desired estimate holds with probability

$$1 - 8e^{-n} - 2\varepsilon^{-p/2}\max\{N^{-3/2}, n^{-(p/4-1)}\}.$$

Case 2: $p > 8$ (Theorem 4.6). In this case we apply Theorem 2.1 (see also the Remark following it), with $\sigma = p/4$, $\lambda = (p-4)/2$, $t = 3(nN)^{1/4}$. Then $M_1 \le C\sqrt{n}(N/n)^{1/4}$ and

$$\beta \le \left(\frac{2(3p-8)}{5e(p-8)N}\right)^{(p-4)/2}\frac{1}{p-5} + \frac{(3p-8)^p}{4(6(p-8))^p N^{(p-8)/4}n^{p/4}}$$

$$\le \left(\frac{3p-8}{6(p-8)}\right)^p N^{-(p-8)/4}\,n^{-p/4} \le 1/64,$$

provided that N is large enough. Thus with probability at least $1 - \sqrt{\beta}$ we have

$$A_n^2 \le C\left(\max_{i\le N}|X_i|^2 + (nN)^{1/4}\max_{i\le N}|X_i| + \sqrt{nN}\right) \le 2C\left(\max_{i\le N}|X_i|^2 + \sqrt{nN}\right).$$

Combining all estimates we obtain that the desired estimate holds with probability

$$1 - 8e^{-n} - 2\left(\frac{3p-8}{6(p-8)}\right)^{p/2} N^{-(p-8)/8} n^{-p/8}.$$

Case 3: $\phi(t) = (1/2)\exp(t^\alpha)$ (Theorem 4.6). As in Case 2 we apply Lemma 4.1. It implies that it is enough to estimate $A^2 + \sqrt{n}\, Z + \sqrt{C(\alpha)N}$, with $C(\alpha)$ from Lemma 4.1, and the corresponding probabilities. A direct calculations show that in this case we have for $C'_\alpha = (4/\alpha)^{1/\alpha}$ and $t > 1$,

$$\mathbb{P}\left((|X|/C'_\alpha)^4 > t\right) \le 2\exp(C'_\alpha)\, t^{\alpha/4} \le \frac{1}{t^2}.$$

We apply Lemma 4.3 with $Z_i = |\langle X_i, a\rangle|^4/\sqrt{C'_\alpha}$, $i \le N$, $q = 2$ and $s = 9e$. It gives

$$\mathbb{P}\left(\left(\sum_{i>k} (\langle X_i, a\rangle^*)^4\right)^{1/2} > Z\right) \le (9e)^{-n},$$

for

$$Z = (C'_\alpha)^{1/4}\, 6\sqrt{6}e\, \sqrt{N}.$$

To estimate A_n we use Theorem 2.1 with $t = (nN)^{1/4}$ and

$$\lambda = 10\,(N/n)^{\alpha/4} \min\left\{1, (\alpha\, \ln(2N/n))^{-1}\right\}.$$

Note that

$$\max\left\{4, 10\,(N/n)^{\alpha/4}\, (\ln(2N/n))^{-1}\right\} \le \lambda \le 10\,(N/n)^{\alpha/4}.$$

Then for absolute positive constants C, C',

$$M_1 \le \sqrt{n}\, (C\lambda)^{1/\alpha} \left(\ln\frac{2N}{n} + \frac{1}{\alpha}\right)^{1/\alpha} \le \left(\frac{C'}{\alpha}\right)^{1/\alpha} (nN)^{1/4}$$

and

$$\beta \le \frac{1}{(10N)^4} \exp\left(\frac{4n^{\alpha/2}}{(3.5\ln(2n))^{2\alpha}}\right) + \frac{N^2}{2\exp((2nN)^{\alpha/4})} \le 1/64,$$

provided that $N \geq (4/\alpha)^{8/\alpha}$. Thus with probability at least $1 - \sqrt{\beta}$ we have

$$A_n^2 \leq C'' \max_{i \leq N} |X_i|^2 + \left(\frac{C'''}{\alpha}\right)^{2/\alpha} \sqrt{nN},$$

where C'' and C''' are absolute positive constants. This together with the estimate for Z completes the proof (note that $C(\alpha) \leq C(2/\alpha)^{5/\alpha}$). $\square$

5 The proof of Theorem 2.1

In this section we prove the main technical result of this paper, Theorem 2.1, which establishes upper bounds for norms of submatrices of random matrices with independent columns. Recall that for $1 \leq k \leq N$ the parameters A_k and B_k are defined by (5).

5.1 Bilinear forms of independent vectors

Let X_1,... X_N be independent random vectors and $a \in \mathbb{R}^N$. Given disjoint sets $T, S \subset \{1, ..., N\}$ we let

$$Q(a, T, S) = \left| \left\langle \sum_{i \in T} a_i X_i, \sum_{j \in S} a_j X_j \right\rangle \right|, \tag{9}$$

with the convention that $\sum_{i \in \emptyset} a_i X_i = 0$.

The following two lemmas are in the spirit of Lemma 2.3 of [24].

Lemma 5.1 *Let X_1,... X_N be independent random vectors in $\mathbb{R}^n$. Let $\gamma \in (1/2, 1)$, $I \subset \{1, ..., N\}$, and $a \in \mathbb{R}^N$. Let $k \geq |\text{supp}(a)|$. Then there exists $\bar{a} \in \mathbb{R}^N$ such that* $\text{supp}(\bar{a}) \subset \text{supp}\,(a)$, $|\text{supp}(\bar{a})| \leq \gamma k$, $|\bar{a}| \leq |a|$, *and*

$$Q(a, I, I^c) \leq Q(\bar{a}, I, I^c) + \max \left\{ \sum_{i=m}^{m+\ell-1} V_i^*, \sum_{i=m}^{m+\ell-1} W_i^* \right\},$$

where $\ell = \lceil (1-\gamma)k \rceil$, $m = \lceil (\gamma - 1/2)k \rceil$, and

$$V_i = \left\langle a_i X_i, \sum_{j \in I^c} a_j X_j \right\rangle \quad \textit{for } i \in I,$$

$$W_j = \left\langle \sum_{i \in I} a_i X_i, a_j X_j \right\rangle \qquad \textit{for } j \in I^c.$$

Proof. Let $E \subset \{1, ..., N\}$ be such that $\mathrm{supp}(a) \subset E$ and $|E| = k$. Everything is clear when $k = 0$ or 1, because then $Q(a, I, I^c) = 0$. Thus we may assume that $k \geq 2$. Let $F_1 = E \cap I$ and $F_2 = E \cap I^c$. First assume that $s := |F_1| \geq k/2$. Note that $(1-\gamma)k \leq k/2 \leq s$, so that $\ell \leq s$. Let $J \subset F_1$ be a set with $|J| = \ell$ such that the set $\{|V_j| : j \in J\}$ consists of ℓ smallest values among the values $\{|V_i| : i \in F_1\}$. (That is, $J \subset F_1$ is such that $|J| = \ell$ and for all $j \in J$ and $i \in F_1 \setminus J$ we have $|V_i| \geq |V_j|$.) Now we let

$$\bar{F}_1 = F_1 \setminus J \quad \text{ and } \quad \bar{F}_2 = F_2.$$

Define the vector $\bar{a} \in \mathbb{R}^N$ by the conditions

$$\bar{a}_{|\bar{F}_1} = a_{|\bar{F}_1}, \qquad \bar{a}_{|\bar{J}} = 0, \qquad \bar{a}_{|\bar{F}_2} = a_{|\bar{F}_2}.$$

Thus $\bar{a}$ differs from a only on coordinates from J; in particular its support has cardinality less than or equal to $|\mathrm{supp}(a)| - |J| = s - \ell \leq k - \ell = \gamma k$. Moreover,

$$\begin{aligned} Q(a, I, I^c) &= \left| \left\langle \sum_{i \in F_1} a_i X_i, \sum_{j \in F_2} a_j X_j \right\rangle \right| \\ &\leq \left| \left\langle \sum_{i \in J} a_i X_i, \sum_{j \in F_2} a_j X_j \right\rangle \right| + \left| \left\langle \sum_{i \in F_1 \setminus J} a_i X_i, \sum_{j \in F_2} a_j X_j \right\rangle \right| \\ &= \left| \sum_{i \in J} \left\langle a_i X_i, \sum_{j \in F_2} a_j X_j \right\rangle \right| + Q(\bar{a}, I, I^c). \end{aligned}$$

Then we have

$$\begin{aligned} Q(a, I, I^c) &\leq Q(\bar{a}, I, I^c) + \sum_{i \in J} \left| \left\langle a_i X_i, \sum_{j \in F_2} a_j X_j \right\rangle \right| \\ &\leq Q(\bar{a}, I, I^c) + \sum_{i=s-\ell+1}^{s} V_i^* \leq Q(\bar{a}, I, I^c) + \sum_{i=m}^{m+\ell-1} V_i^* \end{aligned}$$

If $|F_1| < k/2$ then $|F_2| \geq k/2$ and we proceed similarly interchanging the role of F_1 and F_2 and obtaining

$$Q(a, I, I^c) \leq Q(\bar{a}, I, I^c) + \sum_{i=m}^{m+\ell-1} W_i^*.$$

□

Lemma 5.2 *Let $X_1, \cdots, X_N$ be a sequence of random vectors in $\mathbb{R}^n$ satisfying the hypothesis $\mathbf{H}(\phi)$ for some function $\phi \in \mathcal{M}$. Let $a \in \mathbb{R}^N$ with $|a| = 1$. In the notation of Lemma 5.1, for every $t > 0$ one has*

$$\mathbb{P}\left(\sum_{i=m}^{m+\ell-1} U_i^* > tA_k\right) \leq 2^k \left(\phi\left(\frac{t\sqrt{m}}{\ell}\right)\right)^{-m} \leq 2^k \left(\phi\left(\frac{t\sqrt{\gamma_0 k}}{(1-\gamma)k+1}\right)\right)^{-\gamma_0 k},$$

where $\{U_i\}_i$ denotes either $\{V_i\}_i$ or $\{W_i\}_i$, and $\gamma_0 = \gamma - 1/2$.

Remarks. 1. Taking $\phi(t) = t^p$ for some $p > 0$, we obtain that if

$$\mathbb{P}\left(|\langle X_i, a\rangle| \geq t\right) \leq t^{-p} \tag{10}$$

then

$$\mathbb{P}\left(\sum_{i=m}^{m+\ell-1} U_i^* > tA_k\right) \leq 2^k \left(\frac{t\sqrt{m}}{\ell}\right)^{-mp}. \tag{11}$$

Note that the condition (10) is satisfied if

$$\sup_{i\leq N} \sup_{a\in S^{n-1}} \mathbb{E}|\langle X_i, a\rangle|^p \leq 1.$$

2. Taking $\phi = (1/2)\exp(x^\alpha)$ for some $\alpha > 0$, we obtain that if

$$\mathbb{P}\left(|\langle X_i, a\rangle| \geq t\right) \leq 2\exp(-t^\alpha) \tag{12}$$

then

$$\mathbb{P}\left(\sum_{i=m}^{m+\ell-1} U_i^* > tA_k\right) \leq 2^{k+m} \exp\left(-m\left(\frac{t\sqrt{m}}{\ell}\right)^\alpha\right). \tag{13}$$

Note that the condition (12) is satisfied if

$$\sup_{i\leq N} \sup_{a\in S^{n-1}} \mathbb{E}\exp\left(|\langle X_i, a\rangle|^\alpha\right) \leq 2.$$

Proof. Without loss of generality assume that $U_i = V_i$ for every i. Then

$$\sum_{i=m}^{m+\ell-1} V_i^* \leq \ell V_m^*.$$

Let $F_1 = \mathrm{supp}(a) \cap I$ and $F_2 = \mathrm{supp}(a) \cap I^c$. Note that $V_m^* > s$ means that there exists a set $F \subset F_1$ of cardinality m such that $V_i > s$ for every $i \in F$ (if cardinality of F_1 is smaller than m, the estimate for probability is trivial). Since $|F_1| \leq k$, we obtain

$$\mathbb{P}\left(\sum_{i=m}^{m+\ell-1} V_i^* > tA_k\right) \leq \mathbb{P}\left(\ell V_m^* > tA_k\right) \leq \binom{k}{m} \max_{\substack{F\subset F_1\\ |F|=m}} \mathbb{P}\left(\forall i \in F:\ |V_i| > \frac{tA_k}{\ell}\right).$$

Denote $Z := \sum_{j\in F_2} a_j X_j$. Since $|a| \leq 1$ then $|Z| \leq A_k$, and note that the X_i's, $i \in F_1$ are independent of Z. Thus, conditioning on Z we obtain

$$\begin{aligned}\mathbb{P}\left(\sum_{i=m}^{m+\ell-1} V_i^* > tA_k\right) &\leq 2^k \max_{\substack{F\subset F_1\\ |F|=m}} \prod_{i\in F} \mathbb{P}\left(|a_i|\,|\langle X_i, Z\rangle| > \frac{tA_k}{\ell}\right)\\ &\leq 2^k \max_{\substack{F\subset F_1\\ |F|=m}} \prod_{i\in F} \left(\phi\left(\frac{t}{\ell|a_i|}\right)\right)^{-1}.\end{aligned}$$

Now we show that for every $s > 0$,

$$\prod_{i\in F} \left(\phi\left(\frac{s}{|a_i|}\right)\right)^{-1} \leq \left(\phi\left(s\sqrt{m}\right)\right)^{-m},$$

Indeed, this estimate is equivalent to

$$\frac{1}{m}\sum_{i\in F} \ln\phi\left(\frac{s}{|a_i|}\right) \geq \ln\phi\left(s\sqrt{m}\right),$$

which holds by convexity of $\ln\phi(1/\sqrt{x})$, the facts that $|a| \leq 1$ and $|F| = m$, and since ϕ is increasing. Taking $s = t/\ell$, we obtain

$$\mathbb{P}\left(\sum_{i=m}^{m+\ell-1} V_i^* > tA_k\right) \leq \left(\phi\left(t\sqrt{m}/\ell\right)\right)^{-m}.$$

Finally note that $m = \lceil(\gamma - 1/2)k\rceil \geq \gamma_0 k$ and $\ell = \lceil(1-\gamma)k\rceil \leq (1-\gamma)k + 1$. Since ϕ is increasing, we obtain the last inequality, completing the proof. $\square$

An ε-net argument will be used in the following form.

Lemma 5.3 *Let $m \geq 1$ be an integer and T be an $m \times m$ matrix. Let $\varepsilon \in (0, 1/2)$ and $\mathcal{N} \subset B_2^m$ be a ε-net of B_2^m (in the Euclidean metric). Then*

$$\sup_{x \in B_2^m} |\langle Tx, x\rangle| \leq (1-2\varepsilon)^{-1} \sup_{y \in \mathcal{N}} |\langle Ty, y\rangle|.$$

Proof. Let $S = T + T^*$. For any $x, y \in \mathbb{R}^m$,

$$\langle Sx, x\rangle = \langle Sy, y\rangle + \langle Sx, x-y\rangle + \langle S(x-y), y\rangle.$$

Therefore $|\langle Sx, x\rangle| \leq |\langle Sy, y\rangle| + 2|x-y|\|S\|$. Since S is symmetric, we have

$$\|S\| = \sup_{x \in B_2^m} |\langle Sx, x\rangle|.$$

Thus, if $|x - y| \leq \varepsilon$, then

$$\|S\| \leq \sup_{y \in \mathcal{N}} |\langle Sy, y\rangle| + 2\varepsilon\|S\|$$

and

$$\sup_{x \in B_2^m} |\langle Sx, x\rangle| \leq (1-2\varepsilon)^{-1} \sup_{y \in \mathcal{N}} |\langle Sy, y\rangle|.$$

Since T is a real matrix, then for every $x \in \mathbb{R}^m$, $\langle Sx, x\rangle = 2\langle Tx, x\rangle$. This concludes the proof. □

5.2 Estimates for off-diagonal of bilinear forms

For $1 \leq k \leq N$ and $I \subset \{1, ..., N\}$ we define $Q_k(I)$ by

$$Q_k(I) = \sup_{\substack{E \subset \{1,\dots,N\} \\ |E| \leq k}} \sup_{a \in B_2^E} Q(a, E \cap I, E \cap I^c). \tag{14}$$

Lemmas 5.1, 5.2 and 5.3 imply the following proposition.

Proposition 5.4 *Let $X_1, \cdots, X_N$ be a sequence of random vectors in $\mathbb{R}^n$ satisfying the hypothesis $\mathbf{H}(\phi)$ for some function $\phi \in \mathcal{M}$. Let $\varepsilon \in (0, 1/2)$, $2 \leq k \leq N$, $I \subset \{1, ..., N\}$, $\gamma \in (1/2, 1)$, and $\gamma_0 = \gamma - 1/2$. Then for every $t > 0$ one has*

$$\mathbb{P}\left(Q_k(I) > \frac{Q_{[\gamma k]}(I) + tA_k}{1-2\varepsilon}\right) \leq \exp\left(k\left(\ln\frac{5eN}{k\varepsilon} - \gamma_0 \ln\phi\left(\frac{t\sqrt{\gamma_0 k}}{(1-\gamma)k+1}\right)\right)\right).$$

Moreover, letting $M = \max_i |X_i|$ *one has, for all* $\ell > 1$ *and* $t > 0$*,*

$$\mathbb{P}\left(Q_\ell(I) > tM\right) \le \frac{N^2}{4\phi(4t/\ell)}.$$

Proof. For every $E \subset 1, ..., N$ with $|E| = k$ let $\mathcal{N}_E$ be an ε-net in B_2^E of cardinality at most $(2.5/\varepsilon)^k$. Let $\mathcal{N}$ denote the union of $\mathcal{N}_E$'s. Lemma 5.3 yields

$$Q_k(I) \le (1-2\varepsilon)^{-1} \sup_{\substack{E\subset\{1,\dots,N\}\\ |E|\le k}} \sup_{a\in\mathcal{N}_E} Q(a, E\cap I, E\cap I^c).$$

Therefore, applying Lemmas 5.1 and 5.2, we observe that the event

$$Q_k(I) \le (1-2\varepsilon)^{-1}\left(\sup_{\substack{E\subset\{1,\dots,N\}\\ |E|\le \gamma k}} \sup_{a\in\mathcal{N}} Q(a, E\cap I, E\cap I^c) + tA_k\right)$$

occurs with probability at least

$$1 - \binom{N}{k}\left(\frac{2.5}{\varepsilon}\right)^k 2^k \left(\phi\left(\frac{t\sqrt{\gamma_0 k}}{(1-\gamma)k+1}\right)\right)^{-\gamma_0 k}$$

This implies the first estimate.

Now we prove the "moreover" part. For every $E \subset \{1, \dots, N\}$ of cardinality ℓ denote $F_1 = E \cap I$, $F_2 = E \cap I^c$, $m = |F_1|$ (so $|F_2| = \ell - m$). We also denote

$$M_0 := \max_{i\in I}\max_{j\in I^c} |\langle X_i, X_j\rangle| \quad \text{and} \quad M_1 := \max_{j\in I^c}|X_j|$$

Then for any $a \in B_2^E$ we have

$$\left|\left\langle \sum_{i\in F_1} a_i X_i, \sum_{j\in F_2} a_j X_j\right\rangle\right| \le \left|\sum_{i\in F_1} a_i \sum_{j\in F_2} a_j\right| M_0$$

$$\le \sqrt{m(\ell-m)}\left(\sum_{i\in F_1} a_i^2\right)^{1/2}\left(\sum_{j\in F_2} a_j^2\right)^{1/2} M_0 \le \frac{\ell}{2}\,\frac{M_0}{2}.$$

Therefore, by the union bound,

$$\begin{aligned}\mathbb{P}\left(Q_\ell(I) > tM_1\right) &\le \mathbb{P}\left(M_0 > 4tM_1/\ell\right)\\ &\le \sum_{i\in I}\sum_{j\in I^c}\mathbb{P}\left(|\langle X_i, X_j\rangle| > 4tM_1/\ell\right).\end{aligned}$$

Finally, using the fact that X_i is independent of X_j for $i \neq j$, $|X_j| \leq M_1$ for every $j \in I^c$, and using the tail behavior of variables $\langle X_i, z \rangle$, we obtain

$$\mathbb{P}\left(Q_\ell(I) > tM\right) \leq \mathbb{P}\left(Q_\ell(I) > tM_1\right) \leq \frac{|I|\,|I^c|}{\phi(4t/\ell)} \leq \frac{N^2}{4\phi(4t/\ell)}.$$

$\square$

Proposition 5.5 *Let $1 \leq k \leq N$. Let $X_1, \cdots, X_N$ be random vectors in $\mathbb{R}^n$ satisfying $\mathbf{H}(\phi)$ for some function $\phi \in \mathcal{M}$. Let $t > 0$, $\lambda \geq 1$.*

Case 1. *Let $p > 4$ and $\phi(x) = x^p$. Let $\sigma \in (2, p/2)$. Then*

$$Q_k(I) \leq e^4 \left(t \max_{i \leq N} |X_i| + C_2(\sigma, \lambda, p)\sqrt{k} \left(\frac{5eN}{k}\right)^{\sigma/p} A_k \right)$$

occurs with probability at least

$$1 - \left(\frac{2(\sigma+\lambda)}{5eN(\sigma-2)}\right)^{\lambda} \frac{1}{2\lambda-1} - \frac{N^2(\sigma+\lambda)^p}{4(2t(\sigma-2))^p} \tag{15}$$

and

$$C_2(\sigma, \lambda, p) = 8\sqrt{\frac{\sigma+\lambda}{1+\lambda/2}} \left(\frac{2p}{p-2\sigma}\right)^{1+2\sigma/p} \left(\frac{2(\sigma+\lambda)}{\sigma-2}\right)^{2\sigma/p}.$$

Case 2. *Assume that $\phi(x) = (1/2)\exp(x^\alpha)$ for some $\alpha > 0$. Then for every $t > 0$,*

$$Q_k(I) \leq C^{1/\alpha} \left(t \max_{i \leq N} |X_i| + (C\lambda)^{1/\alpha} \sqrt{k} \left(\left(\ln \frac{20eN}{k} \right)^{1/\alpha} + \left(\frac{1}{\alpha}\right)^{1/\alpha} \right) A_k \right)$$

with probability at least

$$1 - \frac{1}{(10N)^\lambda} \exp\left(-\frac{\lambda k^{\alpha/2}}{(3.5\,\ln(2k))^{2\alpha}} \right) - \frac{N^2}{2\exp((2t)^\alpha)}.$$

Proof. Let $\gamma \in (1/2, 1)$ to be chosen later. For integers $s \geq 0$ denote $k_0 = k$, $k_{s+1} = [\gamma k_s]$. Clearly, the sequence is strictly decreasing whenever $k_s \geq 1$ and $k_s \leq \gamma^s k$. Assume that $k \geq 1/(1-\gamma)$. Define m to be the largest integer $m \geq 1$ such that $k_{m-1} \geq 1/(1-\gamma)$. Note that $\gamma k_{m-1} \geq 1$. Therefore

$$1 \leq k_m < \frac{1}{1-\gamma} \leq k_{m-1}. \tag{16}$$

By Proposition 5.4 we observe that for every positive t_s and $\varepsilon_s \in (0, 1/2)$, $0 \leq s \leq m$, the event

$$Q_k(I) \leq \left(Q_{k_m}(I) + \sum_{s=0}^{m-1} t_s A_{k_s} \right) \prod_{s=0}^{m-1} (1 - 2\varepsilon_s)^{-1}$$

occurs with probability at least

$$1 - 2 \sum_{s=0}^{m-1} \exp\left(k_s \left(\ln \frac{5eN}{k_s \varepsilon_s} - \gamma_0 \ln \phi \left(\frac{t_s \sqrt{\gamma_0 k}}{(1-\gamma)k + 1} \right) \right) \right). \tag{17}$$

Let $\varepsilon > 0$ and a positive decreasing sequence $(\varepsilon_s)_s$ be chosen later and set

$$t_s = \frac{(1-\gamma)k_s + 1}{\sqrt{\gamma_0 k_s}} \, \phi^{-1} \left(\left(\frac{5eN}{k_s \varepsilon_s} \right)^{(1+\varepsilon)/\gamma_0} \right),$$

where $\phi^{-1}(s) = \min\{t \geq 0 \,:\, \phi(t) \geq s\}$.

We start estimating $Q_k(I)$. Since $\ln(1-x) \geq -2x$ on $(0, 3/4]$, we observe that for $\varepsilon_s < 3/8$,

$$\sum_{s=0}^{m-1} \ln(1 - 2\varepsilon_s) \geq \sum_{s=0}^{m-1} -4\varepsilon_s$$

so that

$$\prod_{s=0}^{m-1} (1 - 2\varepsilon_s)^{-1} \leq \exp\left(4 \sum_{s=0}^{m-1} \varepsilon_s \right)$$

Note that

$$\sum_{s=0}^{m-1} t_s A_{k_s} \leq A_k \sum_{s=0}^{m-1} t_s.$$

Thus by (17) and by our choice of t_s,

$$Q_k(I) \le \exp\left(4\sum_{s=0}^{m-1}\varepsilon_s\right)\left(Q_{k_m}(I) + A_k\sum_{s=0}^{m-1}t_s\right) \tag{18}$$

with probability at least

$$1 - 2\sum_{s=0}^{m-1}\exp\left(-k_s\,\varepsilon\,\ln\frac{5eN}{k_s\varepsilon_s}\right) \ge 1 - 2\exp\left(-k_{m-1}\,\varepsilon\,\ln\frac{5eN}{k_{m-1}}\right)\sum_{s=0}^{m-1}\varepsilon_s^{k_s\varepsilon}.$$

Since $k_{m-1} \ge 1/(1-\gamma)$, this probability is larger than

$$1 - 2\exp\left(-\frac{\varepsilon}{1-\gamma}\,\ln\left(5e(1-\gamma)N\right)\right)\sum_{s=0}^{m-1}\varepsilon_s^{k_s\varepsilon}. \tag{19}$$

Thus it is enough to choose appropriately ε_s and to estimate $\sum_{s=0}^{m-1}t_s$, $Q_{k_m}(I)$ and $\sum_{s=0}^{m-1}\varepsilon_s^{k_s\varepsilon}$. We distinguish two cases for ϕ.

Case 1: $\phi(x) = x^p$. In this case we choose $\varepsilon_s = (s+2)^{-2}$ so that

$$\sum_{s=0}^{m-1}\varepsilon_s^{k_s\varepsilon} = \sum_{s=0}^{m-1}(s+2)^{-2k_s\varepsilon} \le \sum_{s=0}^{m-1}(s+2)^{-2k_{m-1}\varepsilon} \le \frac{1}{2k_{m-1}\varepsilon - 1}.$$

Choose $\varepsilon = \lambda(1-\gamma)$. Since $\lambda \ge 1$ and $k_{m-1} \ge 1/(1-\gamma)$, we have $2k_{m-1}\varepsilon \ge 2\varepsilon/(1-\gamma) = 2\lambda$ and

$$\sum_{s=0}^{m-1}(s+2)^{-2k_s\varepsilon} \le \frac{1}{2\lambda - 1}.$$

Using again $k_{m-1} \ge (1-\gamma)^{-1}$, we conclude that the probability in (19) is larger than

$$1 - (5eN(1-\gamma))^{-\lambda}\,\frac{2}{2\lambda - 1}. \tag{20}$$

Now we estimate $\sum_{s=0}^{m-1}t_s$. We have

$$t_s = \frac{(1-\gamma)k_s + 1}{\sqrt{\gamma_0 k_s}}\phi^{-1}\left(\left(\frac{5eN}{k_s\varepsilon_s}\right)^{(1+\varepsilon)/\gamma_0}\right) = \frac{(1-\gamma)k_s + 1}{\sqrt{\gamma_0 k_s}}\left(\frac{5eN}{k_s\varepsilon_s}\right)^{(1+\varepsilon)/\gamma_0 p}.$$

Recall that $\gamma > 1/2$, $k_{m-1} \geq 1/(1-\gamma)$, so that $(1-\gamma)k_s + 1 \leq 2(1-\gamma)k_s$ for $s \leq m-1$. Thus

$$t_s \leq \frac{2(1-\gamma)\sqrt{k_s}}{\sqrt{\gamma_0}} \left(\frac{5eN}{k_s\varepsilon_s}\right)^{(1+\varepsilon)/\gamma_0 p}.$$

Let $b = (1+\varepsilon)/\gamma_0 p$. Assume that $b < 1/2$. Since $k_s \leq \gamma^s k$, we have

$$\sum_{s=0}^{m-1} t_s \leq \frac{2(1-\gamma)k^{1/2-b}(5eN)^b}{\sqrt{\gamma_0}} \sum_{s=0}^{m-1} (s+2)^{\delta b}\gamma^{s(1/2-b)}. \tag{21}$$

Since the function $h(z) = z^{2b}\gamma^{z(1/2-b)}$ on $\mathbb{R}^+$ is first increasing and then decreasing, we get

$$\sum_{s=0}^{m-1}(s+2)^{2b}\gamma^{s(1/2-b)} = \gamma^{-2(1/2-b)}\sum_{s=2}^{m+1} h(s) \leq \gamma^{-1}\left(\sup_{z>0} h(z) + \int_0^\infty h(z)\,dz\right)$$

$$\leq 2\left(\left(\frac{2b}{(1/2-b)e\ln(1/\gamma)}\right)^{2b} + \frac{\Gamma(1+2b)}{((1/2-b)\ln(1/\gamma))^{1+2b}}\right).$$

As $2b \leq 1$, $\Gamma(1+2b) \leq 1$. Using also that $\ln(1/\gamma) \geq 1-\gamma$, we observe that the previous quantity does not exceed

$$\frac{4}{((1/2-b)(1-\gamma))^{1+2b}}.$$

Coming back to (21), we get

$$\sum_{s=0}^{m-1} t_s \leq \frac{8k^{1/2-b}(5eN)^b}{(1/2-b)^{1+2b}(1-\gamma)^{2b}\sqrt{\gamma-1/2}}. \tag{22}$$

To conclude this computation, we choose the parameter

$$\gamma = \frac{1+\lambda+\sigma/2}{\sigma+\lambda}.$$

Note that $\gamma \in (1/2, 1)$ as required, since $\lambda \geq 1$ and $2 < \sigma$. With such a choice of γ, we have $b = \sigma/p < 1/2$, since $\sigma < p/2$. Thus from (22) and (20)

$$\sum_{s=0}^{m-1} t_s \leq 8\sqrt{k}\left(\frac{5eN}{k}\right)^{\sigma/p}\left(\frac{p}{p/2-\sigma}\right)^{1+2\sigma/p}\left(\frac{\sigma+\lambda}{\sigma/2-1}\right)^{2\sigma/p}\sqrt{\frac{\sigma+\lambda}{1+\lambda/2}}$$

holds with probability larger than

$$1-\left(5eN\,\frac{\sigma/2-1}{\sigma+\lambda}\right)^{-\lambda}\frac{2}{2\lambda-1}.$$

Finally, to estimate Q_{k_m}, we note that

$$k_m<\frac{1}{1-\gamma}=\frac{\sigma+\lambda}{\sigma/2-1},$$

and apply "moreover" part of Proposition 5.4 (with $\ell=k_m$). Note that at the beginning of the proof we assumed that $k\geq 1/(1-\gamma)$. In the case $k<1/(1-\gamma)$ the result trivially holds by the "moreover" part of Proposition 5.4 applied with $\ell=k$.

Case 2: $\phi(x)=(1/2)\exp(x^\alpha)$. In this case we choose $\gamma=2/3$, so that $\gamma_0=1/6$. As before we assume that $k\geq 1/(1-\gamma)=3$ (otherwise $Q_k(I)\leq Q_2(I)$). By (16) we have $k_m<3$, hence, by (18)

$$Q_k(I)\leq\exp\left(4\sum_{s=0}^{m-1}\varepsilon_s\right)\left(Q_2(I)+A_k\sum_{s=0}^{m-1}t_s\right).$$

We define k_s by

$$\varepsilon_s=\frac{1}{2}\exp\left(-\left(\frac{k}{k_s}\right)^{\alpha/2}\frac{1}{(s+2)^{2\alpha}}\right).$$

Observe that since $k_s\leq\gamma^s k$ and $\gamma=2/3$, one has

$$\varepsilon_s\leq\frac{1}{2}\exp\left(-\left(\frac{3}{2}\right)^{\alpha s/2}\frac{1}{(s+2)^{2\alpha}}\right)\leq\frac{1}{2e}\,(s+2)^{2\alpha}\left(\frac{2}{3}\right)^{s\alpha/2},$$

which implies

$$\sum_{s=0}^{m-1}\varepsilon_s\leq\frac{C}{\alpha},\tag{23}$$

for a positive absolute constant C.

We have

$$t_s=\sqrt{6}\,\frac{k_s/3+1}{\sqrt{k_s}}\phi^{-1}\left(\left(\frac{5eN}{k_s\varepsilon_s}\right)^{6(1+\varepsilon)}\right)=\sqrt{6}\,\frac{k_s/3+1}{\sqrt{k_s}}\left(\ln\left(2\left(\frac{5eN}{k_s\varepsilon_s}\right)^{6(1+\varepsilon)}\right)\right)^{1/\alpha}.$$

By (16) we have $k_m < 3 \le k_{m-1}$, hence,

$$t_s \le \sqrt{6}\,\frac{2}{3}\,\sqrt{k_s}\,\left(6\,(1+\varepsilon)\right)^{1/\alpha}\left(\ln\frac{20eN}{k_s\varepsilon_s} + \ln\frac{1}{2\varepsilon_s}\right)^{1/\alpha}$$
$$\le \sqrt{6}\,\frac{2}{3}\,2^{1/\alpha}\,\sqrt{k_s}\,\left(6\,(1+\varepsilon)\right)^{1/\alpha}\left(\left(\ln\frac{20eN}{k_s\varepsilon_s}\right)^{1/\alpha} + \left(\ln\frac{1}{2\varepsilon_s}\right)^{1/\alpha}\right).$$

By the choice of ε_s we obtain

$$\sum_{s=0}^{m-1}\sqrt{k_s}\left(\ln\frac{1}{2\varepsilon_s}\right)^{1/\alpha} \le \sqrt{k}\sum_{s=0}^{m-1}(s+2)^{-2} \le 3\sqrt{k}. \tag{24}$$

Since $3^{-s}k \le k_s \le (2/3)^s k$, we observe

$$\sum_{s=0}^{m-1}\sqrt{k_s}\left(\ln\frac{20eN}{k_s\varepsilon_s}\right)^{1/\alpha} \le \sqrt{k}\sum_{s=0}^{m-1}\left(\frac{2}{3}\right)^{s/2}\left(\ln\frac{20eN3^s}{k}\right)^{1/\alpha}$$
$$\le \sqrt{k}\left(\sum_{s=0}^{m-1}\left(\frac{2}{3}\right)^{s/2}2^{1/\alpha}\left(\ln\frac{20eN}{k}\right)^{1/\alpha} + \sum_{s=0}^{m-1}\left(\frac{2}{3}\right)^{s/2}(2s\ln 3)^{1/\alpha}\right)$$
$$\le C_1^{1/\alpha}\sqrt{k}\left(\left(\ln\frac{20eN}{k}\right)^{1/\alpha} + \Gamma(1+1/\alpha)\right),$$

where C_1 is an absolute positive constant and Γ is the Gamma function. This together with (24) implies that

$$\sum_{s=0}^{m-1} t_s \le \left(C_2(1+\varepsilon)\right)^{1/\alpha}\sqrt{k}\left(\left(\ln\frac{20eN}{k}\right)^{1/\alpha} + \Gamma(1+1/\alpha)\right), \tag{25}$$

where C_2 is an absolute positive constant.

Now we estimate the probability. By the choice of k_s we have

$$\sum_{s=0}^{m-1}\varepsilon_s^{\varepsilon k_s} = \sum_{s=0}^{m-1}\exp\left(-\varepsilon k_s\ln(1/\varepsilon_s)\right) = \sum_{s=0}^{m-1}\exp\left(-\varepsilon k_s\left(\ln 2 + (k/k_s)^{\alpha/2}(s+2)^{-2\alpha}\right)\right)$$
$$\le \sum_{s=0}^{m-1}\exp\left(-\varepsilon k_s^{1-\alpha/2}\,k^{\alpha/2}(s+2)^{-2\alpha}\right).$$

Since $k_s \geq k_{m-1} \geq 1/(1-\gamma)$ and $s+2 \leq m+1$ for every $s \leq m-1$, we get that

$$\sum_{s=0}^{m-1} \varepsilon_s^{\varepsilon k_s} \leq m \exp\left(-\frac{\varepsilon}{(1-\gamma)^{1-\alpha/2}} \frac{k^{\alpha/2}}{(m+1)^{2\alpha}}\right).$$

Since m is chosen such that $1/(1-\gamma) \leq k_{m-1} \leq (2/3)^{m-1}k$, we observe that

$$m - 1 \leq \frac{\ln(k(1-\gamma))}{\ln(3/2)}.$$

Therefore,

$$\sum_{s=0}^{m-1} \varepsilon_s^{\varepsilon k_s} \leq \left(1 + \frac{\ln(k/3)}{\ln(3/2)}\right) \exp\left(-\frac{\varepsilon}{(1/3)^{1-\alpha/2}} \frac{k^{\alpha/2}}{(2.5 \ln k)^{2\alpha}}\right)$$
$$\leq 2 \exp\left(-3\varepsilon \frac{k^{\alpha/2}}{3^{\alpha/2}\,(2.5 \ln k)^{2\alpha}}\right),$$

which shows that probability in (19) is at least

$$1 - \frac{4}{(15eN)^{3\varepsilon}} \exp\left(-3\varepsilon \frac{k^{\alpha/2}}{(3.5 \ln k)^{2\alpha}}\right).$$

Finally, to estimate $Q_2(I)$ we apply the "moreover" part of Proposition 5.4 (with $\ell = 2$). Choosing $\varepsilon = \lambda/3$ and combining estimates (23), and (25) with the estimate for $Q_2(I)$ we obtain the desired result. □

5.3 Estimating A_k and B_k

We are now ready to pass to the proof of Theorem 2.1. To prove the theorem we need two simple lemmas.

Lemma 5.6 *Let $\beta \in (0,1)$. Let $\mathbb{P}_1$ and $\mathbb{P}_2$ be probability measures on Ω_1 and Ω_2 respectively and let $V \subset \Omega_1 \otimes \Omega_2$ be such that*

$$\mathbb{P}_1 \otimes \mathbb{P}_2(V) \geq 1 - \beta.$$

Then there exists $W \subset \Omega_2$ such that

$$\mathbb{P}_2(W) \geq 1 - \sqrt{\beta} \quad \textit{and} \quad \forall x_2 \in W,\ \mathbb{P}_1\left(\{x_1 : (x_1, x_2) \in V\}\right) \geq 1 - \sqrt{\beta}.$$

Proof. Fix some $\delta \in (0,1)$. Let

$$W := \{x_2 \in \Omega_2 \,:\, \mathbb{P}_1\left(\{x_1 \in \Omega_1 :\; (x_1,x_2) \in V\}\right) \geq 1-\delta\}.$$

Clearly,

$$W^c = \{x_2 \in \Omega_2 \,:\, \mathbb{P}_1\left(\{x_1 \in \Omega_1 :\; (x_1,x_2) \in V^c\}\right) \geq \delta\}.$$

Then

$$\beta \geq \mathbb{P}_1 \otimes \mathbb{P}_2(V^c) = \int_{\Omega_2} \mathbb{P}_1\left(\{x_1 \in \Omega_1 :\; (x_1,x_2) \in V^c\}\right)\, d\,\mathbb{P}_2(x_2)$$

$$\geq \int_{W^c} \mathbb{P}_1\left(\{x_1 \in \Omega_1 :\; (x_1,x_2) \in V^c\}\right)\, d\,\mathbb{P}_2(x_2) \geq \delta\, \mathbb{P}_2(W^c),$$

which means $\mathbb{P}_2(W) \geq 1-\beta/\delta$. The choice $\delta = \sqrt{\beta}$ completes the proof. □

The following lemma is obvious.

Lemma 5.7 *Let $x_1,\ldots,x_N \in \mathbb{R}^n$, then*

$$\sum_{i\neq j} \langle x_i, x_j\rangle = 2^{2-N} \sum_{I\subset\{1,\ldots,N\}} \sum_{i\in I} \sum_{j\in I^c} \langle x_i, x_j\rangle\,.$$

Proof of Theorem 2.1. From Lemma 5.7 we have

$$\left| \left| \sum_{i=1}^N a_i X_i \right|^2 - \sum_{i=1}^N a_i^2\, |X_i|^2 \right| = 2^{2-N} \left| \sum_{I\subset\{1,2,\ldots,N\}} \left\langle \sum_{i\in I} a_i X_i, \sum_{j\in I^c} a_i X_i \right\rangle \right|.$$

We deduce that

$$B_k^2 \leq 2^{2-N} \sup_{a\in U_k} \sum_{I\subset\{1,2,\ldots,N\}} Q_k(a,I,I^c) \leq 2^{2-N} \sum_{I\subset\{1,2,\ldots,N\}} \sup_{a\in U_k} Q_k(a,I,I^c)$$

$$\leq 2^{2-N} \sum_{I\subset\{1,2,\ldots,N\}} Q_k(I).$$

Let $I \subset \{1,\ldots,N\}$ be fixed. Proposition 5.5 implies

$$\mathbb{P}\left(Q_k(I) \leq M_0\right) \geq 1-\beta, \tag{26}$$

where

$$M_0 := C_\phi \, t \, \max_{i \le N} |X_i| + (M_1/4)\, A_k.$$

Consider two probability spaces $\{I : I \subset \{1, ..., N\}\}$ with the normalized counting measure μ and our initial probability space $(\Omega, \mathbb{P})$, on which X_i's are defined. By (26) we observe that the $\mu \otimes \mathbb{P}$ probability of the event $V := \{Q_k(I) \le M_0\}$ is at least $1 - \beta$. Then Lemma 5.6 implies that there exists $W \subset \Omega$ such that $\mathbb{P}(W) \ge 1 - \sqrt{\beta}$ and such that for every $\omega \in W$ one has $\mu(\{Q_k(I) \le M_0\}) \ge 1 - \sqrt{\beta}$. Since $Q_k(I) \le A_k^2$, we obtain that for every $\omega \in W$,

$$B_k^2 \le 4M_0 + 4\sqrt{\beta} A_k^2.$$

Since $A_k^2 \le \max_{i \le N} |X_i|^2 + B_k^2$, we have

$$A_k^2 \le \frac{4M_0 + \max_{i\le N} |X_i|^2}{1 - 4\sqrt{\beta}} \quad \text{ and } \quad B_k^2 \le \frac{4(M_0 + \sqrt{\beta} \max_{i \le N} |X_i|^2)}{1 - 4\sqrt{\beta}}. \tag{27}$$

Therefore

$$A_k^2 \le (1 - 4\sqrt{\beta})^{-1} \left(\max_{i \le N} |X_i|^2 + 4C_\phi t \max_{i \le N} |X_i| + M_1 A_k \right).$$

Using $\sqrt{u^2 + v^2} \le u + v$, and denoting $\gamma = (1 - 4\sqrt{\beta})^{-1}$ (recall $M = \max_{i \le N} |X_i|$) we obtain

$$A_k \le \sqrt{\gamma}\, M + 2\sqrt{C_\phi \, \gamma \, t \, M} + \gamma \, M_1,$$

which proves the estimate for A_k. Plugging this into (27), we also observe

$$\begin{aligned} B_k^2 &\le \gamma \left(4\sqrt{\beta}\, M^2 + 4C_\phi t M + \gamma M_1^2 + \sqrt{\gamma}\, M\, M_1 + 2\sqrt{C_\phi \, \gamma \, tM}\, M_1 \right) \\ &\le \gamma \left(4\sqrt{\beta}\, M^2 + 8C_\phi t M + 2\gamma M_1^2 + \sqrt{\gamma}\, M\, M_1 \right). \end{aligned}$$

This completes the proof. □

6 Optimality

In this section we discuss optimality of estimates in Theorems 1.2 and 2.1.

To obtain the lower estimates on A_m we use the following observation.

Lemma 6.1 *Let $A = (X_{ij})_{i\le n, j\le N}$ be an $n\times N$ matrix with i.i.d. entries. Then*

$$\mathbb{P}(A_m \ge t) \ge \frac{1}{2} \quad \textit{whenever} \quad \mathbb{P}\left(|X_{11}| \ge \frac{t}{\sqrt{m}}\right) \ge \frac{m+1}{N}. \tag{28}$$

Proof. For every $i \le N$, let $X_j \in \mathbb{R}^n$ be the j-th columns of A. For $m \le N$ we have

$$A_m = \sup_{a\in U_m} \left|\sum_{j=1}^N a_j X_j\right| \ge \sup_{a\in U_m} \left|\sum_{j=1}^N a_j X_{1j}\right| \ge \sup_{\substack{a\in U_m\\ a_j\in\{\pm 1/\sqrt{m},0\}}} \left|\sum_{j=1}^N a_j X_{1j}\right|$$

$$= \frac{1}{\sqrt{m}} \sum_{j=1}^m X_{1j}^* \ge \sqrt{m}\, X_{1m}^*.$$

Therefore, using independence, we have

$$\mathbb{P}\left(A_m \ge t\right) \ge \mathbb{P}\left(X_{1m}^* \ge \frac{t}{\sqrt{m}}\right) = \mathbb{P}(Y \ge m),$$

where Y is a real random variable with a binomial distribution of size N and parameter $v = \mathbb{P}(|X_{11}| \ge \frac{t}{\sqrt{m}})$. It is well known that the median of Y, med(Y) satisfies

$$\lfloor Nv\rfloor \le \text{med }(Y) \le \lceil Nv\rceil.$$

Thus $\mathbb{P}(A_m \ge t) \ge \frac{1}{2}$ whenever $m \le \lfloor Nv\rfloor$. This implies the result. □

To evaluate RIP, we will use the following simple observation.

Lemma 6.2 *Let $n \le N$ and $m \le N$. Let A be an $n\times N$ random matrix satisfying*

$$\mathbb{P}(A_m \ge t\sqrt{m}) \ge \frac{1}{2}.$$

Assume also that A satisfies RIP$_m(\delta)$ for some $\delta < 1$ with probability greater than $1/2$. Then

$$mt^2 \le 2n.$$

Proof. As A satisfies RIP$_m(\delta)$ for some $\delta < 1$ with probability greater than $1/2$, then clearly

$$A_m^2 = \sup_{a\in U_m} |\sum a_i X_i|^2 \le 2n$$

with probability greater than 1/2. Therefore, with positive probability one has

$$t\sqrt{m} \le A_m \le \sqrt{2n},$$

which implies the result. □

In order to show that a matrix with i.i.d. random variables satisfies condition $\mathbf{H}(\phi)$ with $\phi(t) = t^p$ we need the Rosenthal's inequality ([27], see also [16]). As usual, by $\|\cdot\|_q$ for a random variable ξ we mean its L_q-norm and for an $a \in \mathbb{R}^n$ its ℓ_q-norm, that is

$$\|\xi\|_q = (\mathbb{E}|\xi|^q)^{1/q} \quad \text{and} \quad \|a\|_q = \left(\sum_{i=1}^n |a_i|^q\right)^{1/q}.$$

Note that originally the Rosenthal inequality was proved for symmetric random variables, but using standard symmetrization argument (i.e. passing from random variables ξ_i's to $(\xi_i - \xi_i')$'s, where (ξ_i')'s have the same distribution and are independent), one can pass to centered random variables.

Lemma 6.3 *Let $q > 2$ and $a \in \mathbb{R}^n$. Let ξ_1, ..., ξ_n be i.i.d. centered random variables with finite q-th moment. Then there exists a positive absolute constant C such that*

$$\frac{1}{2} M_q \le \left\| \sum_{i=1}^n a_i \xi_i \right\|_q \le C \frac{q}{\ln q} M_q \tag{29}$$

where $M_q := \max\{\|a\|_2\|\xi_1\|_2, \|a\|_q\|\xi_1\|_q\}$.

The following is an almost immediate corollary of Rosenthal's inequality. It should be compared with Proposition 1.3 of [29].

Corollary 6.4 *Let $p > 4$. Let ξ be a random variable of variance one and with a finite p-th moment. Let ξ_{ij}, $i \le n$, $j \le N$ be i.i.d. random variables distributed as ξ. Then for every $t > 0$,*

$$\mathbb{P}\left(\max_{j\le N}\left|\frac{1}{n}\sum_{i=1}^n \xi_{ij}^2 - 1\right| > t\right) \le \left(\frac{Cp}{t\,\ln p}\right)^{p/2} \mathbb{E}|\xi|^p \,\frac{N}{n^{p/4}},$$

where C is a positive absolute constant.

Proof. Let ξ_1, ..., ξ_n be i.i.d. random variables distributed as ξ. We apply Rosenthal's inequality to random variables $(\xi_i^2 - 1)$ with $q = p/2$ and $a = (1, 1, ..., 1)$. Then

$$\left\| \sum_{i=1}^{n} (\xi_i^2 - 1) \right\|_{p/2} \leq C_p \sqrt{n}\, \|\xi^2 - 1\|_{p/2} \leq C_p \sqrt{n}\, \left(\|\xi^2\|_{p/2} + 1 \right) \leq 2 C_p \sqrt{n}\, \|\xi\|_p^2,$$

where $C_p = Cp/\ln p$ for an absolute positive constant C. Using Chebyshev's inequality we observe

$$\mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} \xi_i^2 - 1 \right| > t \right) \leq \frac{\mathbb{E} \sum_{i=1}^{n} |\xi_i^2 - 1|^{p/2}}{(tn)^{p/2}} \leq \frac{(2C_p)^{p/2}\, \|\xi\|_p^p}{t^{p/2}\, n^{p/4}}.$$

The result follows by the union bound. □

The next proposition gives a lower bound for A_m to be compared with Case 1 of Theorem 2.1, where we got $A_m \leq C_p \sqrt{m}\, (N/m)^{2/p}$ with high probability.

Proposition 6.5 *Let $p > 2$, $1 \leq m \leq N$. There exists a sequence of random vectors $X_1, \cdots, X_N$ in $\mathbb{R}^n$ satisfying*

$$\forall 1 \leq i \leq N \ \forall a \in S^{n-1} \quad \mathbb{E} |\langle X_i, a \rangle|^p \leq 1 \tag{30}$$

and such that

$$\mathbb{P}\left(A_m \geq \frac{Cp}{\ln p} \sqrt{m} \left(\frac{N}{m} \right)^{1/p} \left(\ln \left(\frac{2N}{m} \right) \right)^{-1/p} \right) \geq \frac{1}{2},$$

where C is an absolute positive constant.

Proof. Let $\lambda \geq 1$ to be set later and let us put

$$f_p(x) = \begin{cases} \frac{p}{2(1-\lambda^{-p})|x|^{p+1}} & \text{if } 1 \leq |x| \leq \lambda \\ 0 & \text{otherwise.} \end{cases}$$

We have $\int f_p(x)\, dx = 1$ and

$$a_p^p := \int |x|^p f_p(x)\, dx = p \frac{\ln \lambda}{1 - \lambda^{-p}}.$$

Consider the random variable $\xi(\omega) = \omega$ with respect to the density f_p and let (X_{ij}) be i.i.d. copies of ξ/a_p. Clearly, $\mathbb{E}|X_{11}|^p = 1$. Since, for $s \in [1, \lambda]$

$$\mathbb{P}(|\xi| > s) = \frac{1}{1-\lambda^{-p}} \left(\frac{1}{s^p} - \frac{1}{\lambda^p} \right),$$

a short computation using (28) shows that $\mathbb{P}(A_m \geq t) \geq \frac{1}{2}$ provided that

$$t \leq \left(\frac{1-\lambda^{-p}}{p \ln \lambda} \right)^{1/p} \sqrt{m} \left(\frac{N}{(m+1)(1-\lambda^{-p}) + N\lambda^{-p}} \right)^{1/p}$$

$$= \sqrt{m} \left(\frac{1}{p \ln \lambda} \right)^{1/p} \left(\frac{N}{m+1+N/(\lambda^p - 1)} \right)^{1/p}.$$

Choosing λ from $\lambda^p - 1 = N/(m+1)$, we obtain $\mathbb{P}(A_m \geq t) \geq \frac{1}{2}$ provided that

$$t \leq \sqrt{m} \left(\frac{N}{2(m+1)\ln(2N/(m+1))} \right)^{1/p}.$$

Finally, to satisfy condition (30), we pass from matrix A to $A' = A/c_p = (X_{ij}/c_p)_{ij}$, where $c_p \leq Cp/\ln p$ is a constant in Rosenthal's inequality (29). By Rosenthal's inequality, the sequence of columns of A' satisfies the condition (30). $\square$

The next proposition gives an upper bound on the size of sparsity m in order to satisfy RIP under condition of Case 1 of Theorem 3.1 (see also Remark 3 following this theorem).

Proposition 6.6 *Let $q > p > 2$, $n \leq N$ and $m \leq N$. There exists an absolute positive constant C, an $n \times N$ matrix A, whose columns $X_1, ..., X_N$ satisfy*

$$\forall 1 \leq i \leq N \ \forall a \in S^{n-1} \quad \mathbb{E}|\langle X_i, a \rangle|^p \leq \left(\frac{Cp}{\ln p} \right)^p \frac{q}{q-p} \left(\frac{q-2}{q} \right)^{p/2}, \quad (31)$$

and for every $t \in (0,1)$,

$$\mathbb{P}\left(\max_{i \leq N} \left| \frac{|X_i|^2}{n} - 1 \right| \geq t \right) \leq t^{p/2} \quad (32)$$

provided that

$$N \leq \left(\frac{q \ln p}{C(q-2)p}\right)^{p/2} \frac{q-p}{q}\, t^p\, n^{p/4}.$$

Assume that A satisfies $RIP_m(\delta)$ for some $\delta < 1$ with probability greater than $1/2$. Then

$$m \left(\frac{N}{m+1}\right)^{2/q} \leq \frac{2(q-2)}{q}\, n.$$

Proof. Consider the density

$$f(x) = \begin{cases} \frac{q}{2|x|^{q+1}} & \text{if } |x| \geq 1 \\ 0 & \text{otherwise.} \end{cases}$$

We have $\int f(x)\, dx = 1$,

$$\int |x|^p f(x)\, dx = \frac{q}{q-p} \quad \text{and} \quad a_2^2 := \int |x|^2 f(x)\, dx = \frac{q}{q-2}.$$

Consider the random variable $\xi(\omega) = \omega$ with respect to the density f and let $(X_{ij})_{ij}$ be i.i.d. copies of ξ/a_2. Clearly,

$$\mathbb{E}|X_{11}|^2 = 1 \quad \text{and} \quad \mathbb{E}|X_{11}|^p = \frac{q}{q-p} \left(\frac{q-2}{q}\right)^{p/2}.$$

Then Rosenthal's inequality (29) implies the condition (31) and Corollary 6.4 implies (32).

Now we estimate A_m for the matrix A, whose columns are $(X_{ij})_i$, $j \leq N$. Since, for $s \geq 1$, $\mathbb{P}(|\xi| > s) = s^{-q}$, by (28), we obtain that $\mathbb{P}(A_m \geq t) \geq \frac{1}{2}$ provided that

$$t \leq \sqrt{m} \sqrt{\frac{q-2}{q}} \left(\frac{N}{m+1}\right)^{1/q}.$$

This means

$$\mathbb{P}\left(A_m \geq \sqrt{m} \sqrt{\frac{q-2}{q}} \left(\frac{N}{m+1}\right)^{1/q}\right) \geq \frac{1}{2},$$

and we complete the proof applying Lemma 6.2. □

The next proposition shows the optimality (up to absolute constants) of the sparsity parameter in Case 2 of Theorem 3.1.

Proposition 6.7 *There exist absolute positive constants c, C such that the following holds. Let $\alpha \in [1,2]$, $1 \le m \le N/2$ and n satisfies $N \le \exp(cn^{\alpha/2})$. There exists an $n \times N$ matrix A, whose columns $X_1, ..., X_N$ satisfy*

$$\forall 1 \le i \le N \ \ \forall a \in S^{n-1} \quad \mathbb{E} \exp\left(\left| \langle X_i, a \rangle \right|^{\alpha}\right) \le C \tag{33}$$

and

$$\mathbb{P}\left(\max_{i \le N} \left| \frac{|X_i|^2}{n} - 1 \right| \ge \frac{\sqrt{2}-1}{2} \right) \le 2 \exp(-cn^{\alpha/2}), \tag{34}$$

and such that

$$\mathbb{P}\left(A_m \ge \sqrt{\frac{m}{2}} \left(\ln \frac{N}{m+1} \right)^{1/\alpha} \right) \ge \frac{1}{2}. \tag{35}$$

Additionally, if $n \le N$ and if A satisfies $RIP_m(\delta)$ for some $\delta < 1$ with probability greater than $1/2$, then

$$m \left(\ln \frac{N}{m+1} \right)^{2/\alpha} \le 4n.$$

Proof. We consider a symmetric random variable ξ with the distribution defined by $\mathbb{P}\left(|\xi| > t\right) = \exp(-t^{\alpha})$. It is easy to check that

$$\mathbb{E} \exp(|\xi|^{\alpha}/2) = 2$$

and

$$a := \mathbb{E}\xi^2 = \Gamma\left(\frac{2}{\alpha} + 1 \right) \in [1,2].$$

Let X_{ij}, $i \le n$, $j \le N$ be i.i.d. copies of $\xi/\sqrt{a}$, $A = (X_{ij})_{ij}$ and X_j's be its columns. Applying Lemma 3.4 from [4] (see also Theorem 1.2.8 in [11]) we observe that X_i's satisfy conditions (33) and (34). By (28) we observe that $\mathbb{P}(A_m \ge t) \ge \frac{1}{2}$ provided that

$$\mathbb{P}\left(|\xi| \ge \frac{\sqrt{a}\, t}{\sqrt{m}} \right) = \exp\left(- \left(\sqrt{a} t / \sqrt{m} \right)^{\alpha} \right) \ge \frac{m+1}{N}.$$

Thus it is enough to take

$$t \le \sqrt{\frac{m}{a}} \left(\ln \frac{N}{m+1} \right)^{1/\alpha}.$$

This proves the estimate (35).

Finally, the "additionally" part follows by Lemma 6.2. □

References

[1] R. Adamczak, O. Guédon, A.E. Litvak, A. Pajor, N. Tomczak-Jaegermann, *Smallest singular value of random matrices with independent columns*, C. R. Math. Acad. Sci. Paris 346 (2008), no. 15-16, 853–856.

[2] R. Adamczak, A.E. Litvak, A. Pajor, N. Tomczak-Jaegermann, *Quantitative estimates of the convergence of the empirical covariance matrix in log-concave Ensembles,* J. Amer. Math. Soc. 23 (2010), no. 2, 535–561.

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

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

[5] R. Adamczak, R. Latala, A. Litvak, A. Pajor, N. Tomczak-Jaegermann, *Geometry of log-concave Ensembles of random matrices and approximate reconstruction*, C.R. Math. Acad. Sci. Paris, 349 (2011), 783–786.

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

[7] Z.D. Bai, J. Silverstein, Y.Q. Yin, *A note on the largest eigenvalue of a large dimensional sample covariance matrix*, Journal of Multivariate Analysis, 26 (1988), 166–168.

[8] J. Bourgain, *Random points in isotropic convex sets,* Math. Sci. Res. Inst. Publ., 34, 53–58, Cambridge Univ. Press, Cambridge (1999).

[9] E.J. Candés, J. Romberg, T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math., 59 (2006), 1207–1223.

[10] E.J. Candés, T. Tao, *Decoding by linear programming*, IEEE Trans. Inform. Theory, 51 (2005), 4203–4215.

[11] D. Chafaï, O. Guédon, G. Lecué, A. Pajor, *Interactions between compressed sensing random matrices and high dimensional geometry*, Panoramas et Synthèses [Panoramas and Syntheses], 37. Soc. Math. de France, Paris, 2012.

[12] D.L. Donoho, *Compressed sensing,* IEEE Trans. Information Theory 52 (2006), 1289–1306.

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

[14] O. Guédon, A.E. Litvak, A. Pajor, N. Tomczak-Jaegermann *Restricted isometry property for random matrices with heavy tailed columns,* C.R. Math. Acad. Sci. Paris, 352 (2014), 431–434.

[15] O. Guédon, E. Milman, *Interpolating thin-shell and sharp large-deviation estimates for isotropic log-concave measures*, Geom. Funct. Anal., 21 (2011), 1043–1068.

[16] W.B. Johnson, G. Schechtman, J. Zinn, *Best constants in moment inequalities for linear combinations of independent and exchangeable random variables*, Ann. Probab. 13 (1985), 234–253.

[17] B. Klartag, Power-law estimates for the central limit theorem for convex sets, *J. Funct. Anal.* 245 (2007), 284–310.

[18] V. Koltchinskii, S. Mendelson, *Bounding the smallest singular value of a random matrix without concentration*, Preprint.

[19] G. Lecué, S. Mendelson, *Sparse recovery under weak moment assumptions*, preprint.

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

[21] A.E. Litvak, S. Spektor, *Quantitative version of a Silversteins result,* GAFA, Lecture Notes in Math., to appear.

[22] S. Mendelson, A. Pajor, N. Tomczak-Jaegermann, *Reconstruction and subgaussian operators*, Geometric and Functional Analysis, 17 (2007), 1248–1282.

[23] S. Mendelson, G. Paouris, *On generic chaining and the smallest singular value of random matrices with heavy tails*, Journal of Functional Analysis, 262 (2012), 3775–3811.

[24] S. Mendelson, G. Paouris, *On singular values of matrices*, Journal of EMS, 16 (2014), 823–834.

[25] G. Paouris, *Concentration of mass on convex bodies*, Geom. Funct. Anal. 16, no. 5 (2006), 1021–1049.

[26] G. Paouris, *Small ball probability estimates for log-concave measures*, Trans. Amer. Math. Soc., 364 (2012), 287–308.

[27] H. Rosenthal, *On the subspaces of* L_p *(*$p > 2$*) spanned by sequences of independent random variables*, Israel J. Math. 8 (1970), 273–303.

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

[29] N. Srivastava, R. Vershynin, *Covariance estimation for distributions with 2+epsilon moments,* Annals of Probability 41 (2013), 3081–3111.

[30] S. Szarek, *Condition numbers of random matrices*, J. Complexity 7 (1991), no. 2, 131–149.

[31] K.E. Tikhomirov, *The smallest singular value of random rectangular matrices, with no moment assumptions on entries*, preprint.

[32] R. Vershynin, *How close is the sample covariance matrix to the actual covariance matrix?*, J. Theoret. Probab. 25 (2012), 655–686.

Olivier Guédon, and Alain Pajor
Université Paris-Est
Laboratoire d'Analyse et Mathématiques Appliquées (UMR 8050)
UPEM, F-77454, Marne-la-Vallée, France
olivier.guedon@u-pem.fr, alain.pajor@u-pem.fr

Alexander E. Litvak and Nicole Tomczak-Jaegermann,
Dept. of Math. and Stat. Sciences,
University of Alberta,
Edmonton, Alberta, Canada, T6G 2G1.
e-mail: aelitvak@gmail.com, nicole.tomczak@ualberta.ca