Olivier Guédon · Alexander E. Litvak · Alain Pajor
Nicole Tomczak-Jaegermann

JEMS

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

**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 $\mathbb{P}(|\langle X_i, a \rangle| \geq t) \leq t^{-p}$ 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 the Euclidean norms $|X_i|$ are concentrated around $\sqrt{n}$. We also show that the covariance matrix is well approximated by empirical covariance matrices and establish corresponding quantitative estimates on the rate of convergence in terms of the ratio $n/N$. Moreover, we obtain sharp bounds for both problems when the decay is of the type $\exp(-t^\alpha)$ with $\alpha \in (0, 2]$, extending the known case $\alpha \in [1, 2]$.

**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. Introduction and main results

Fix positive integers $n$, $N$ and let $A$ be an $n \times N$ random matrix whose columns $X_1, \ldots, X_N$ are independent random vectors in $\mathbb{R}^n$. For a subset $I \subset \{1, \ldots, N\}$ 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 $X_i$, $i \leq N$, 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.

O. Guédon, A. Pajor: Université Paris-Est, Laboratoire d'Analyse et Mathématiques Appliquées (UMR 8050), UPEM, F-77454 Marne-la-Vallée, France;
e-mail: olivier.guedon@u-pem.fr, alain.pajor@u-pem.fr

A. E. Litvak, N. Tomczak-Jaegermann: Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada, T6G 2G1;
e-mail: aelitvak@gmail.com, nicole.tomczak@ualberta.ca

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 $X$ in $\mathbb{R}^n$ satisfies *hypothesis* $\mathbf{H}(\phi)$ with parameter $\tau \geq 1$ if

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

for a certain function $\phi$, and we assume that $X_i$ satisfies $\mathbf{H}(\phi)$ for all $i \leq N$. We will focus on two choices of the function $\phi$, 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 extends the known subexponential case ($\alpha = 1$, see [3, 4]).

The concept of the Restricted Isometry Property was introduced in [10] in order to study an exact reconstruction problem by means of the $\ell_1$-minimization algorithm, classical in compressed sensing. Although it provided only a sufficient condition for 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 $X_1, \ldots, X_N$, as was shown in [12].

Let us recall the intuition of RIP (for the definition see (9) below). For an $n \times N$ matrix $T$ and $1 \leq m \leq N$, the *isometry constant of order $m$* of $T$ is the parameter $0 < \delta_m(T) < 1$ such that the squares of the Euclidean norms $|Tz|$ and $|z|$ are approximately equal, up to a factor $1 + \delta_m(T)$, for all $m$-sparse vectors $z \in \mathbb{R}^N$ (that is, $|\text{supp}(z)| \leq m$). 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 $\theta$, then the squares of all singular values of the matrices $T^I$ belong to $[1-\theta, 1+\theta]$. Note that in compressed sensing, for the reconstruction of vectors by $\ell_1$-minimization, one does not need RIP for all $\theta > 0$ (see [12] and [11]). The RIP implicitly contains 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 normalize the matrix by considering $A/\sqrt{n}$ and we introduce the concentration function

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

Until now the only known cases of random matrices satisfying RIP have been subgaussian [9, 10, 12, 23] and subexponential [5] matrices. Our first main theorem says that matrices we consider have RIP of order $m$, with "large" $m$ of the form $m = n\psi(n/N)$ with $\psi$ depending on $\phi$ and possibly on other parameters. In particular, when $N$ is proportional to $n$, then $m$ is proportional to $n$. We now 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 random vectors satisfying hypothesis $\mathbf{H}(\phi)$ for some $\phi$. Assume that $n, N$ are large enough. Then there exists a function $\psi$ depending on $\phi$ and $\theta$ such that with high probability (depending on the concentration function $P(\theta)$) the matrix $A/\sqrt{n}$ has RIP of order $m = [n\psi(n/N)]$ with a parameter $\theta$ (that is, $\delta_m(A/\sqrt{n}) \leq \theta$).*

The second problem we investigate goes back to a question of Kannan, Lovász and Simonovits (KLS). As before, assume that $A$ is a random $n \times N$ matrix whose columns $X_1, \ldots, X_N$ are independent random vectors satisfying hypothesis $\mathbf{H}(\phi)$ for some $\phi$. Additionally assume that $X_i$'s are identically distributed as a centered random vector $X$. The 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 satisfies $\|U - \Sigma\| \leq \varepsilon\|\Sigma\|$ for $N$ 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 in 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 [7] states that under the assumption of boundedness of the 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 [6, 30] that boundedness of the 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 set

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

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 $S \leq \varepsilon$ 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, 32] that with probability close to one,

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

where $C$ is a positive absolute constant. In [3, 4] the same estimate was obtained for a large class of random matrices, in particular without requiring that the entries of the columns are independent, or that the $X_i$'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 $M \leq C(Nn)^{1/4}$ with high probability. Both conditions hold for log-concave random vectors.

Until recently, 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 the Bai–Yin theorem, it is a natural question whether one can replace the function

$\phi(t) = e^t/2$ by $\phi(t) = e^{t^\alpha}/2$ with $\alpha \in (0, 1)$ or $\phi(t) = t^p$, for $p \geq 4$. The first attempt in this direction was done in [34], where the bound $S \leq C(p, K)(n/N)^{1/2 - 2/p}(\ln \ln n)^2$ 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 one to solve the KLS problem with $N$ proportional to $n$. This problem was solved in [24, 31] 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 [25], 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 [25] 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 assumptions for estimating the smallest singular value. This already appeared in the work [31] and was pushed further in recent works [19, 33, 35] and in [14, 20, 26] which led to new bounds on the performance of $\ell_1$-minimization methods.

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*

$$S \leq C\left(\left(\frac{M^2}{n}\right)\left(\frac{n}{N}\right) + C(p, \varepsilon)\left(\frac{n}{N}\right)^{\gamma/p}\right) \tag{5}$$

*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

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

Let $X$ have i.i.d. coordinates distributed as a centered random variable with finite $p$-th moment, $p > 2$. Then by Rosenthal's inequality [29] (see also [17] and Lemma 6.3 below), $X$ satisfies hypothesis $\mathbf{H}(\phi)$ with $\phi(t) = t^p$. Let $X_1, \ldots, X_N$ be independent random vectors distributed as $X$. It is known [6], [30] (see also [22] for a quantitative version) that when $N$ is proportional to $n$ and the fourth moment is unbounded, $M^2/n \to \infty$ as $n \to \infty$. Hence, bounds for $S$ involving the term $M^2/n$ like the bound (5) are of interest only for $p \geq 4$. We do not know if (5) holds in the case $p = 4$.

The main novelty of our proof is a delicate analysis of the behavior of norms of submatrices, namely quantities $A_k$ and $B_k$, $k \leq N$, defined in (6) below. This analysis is done in Theorem 2.1, which is the heart of the technical part of the paper. The estimates for $B_k$ are responsible for RIP, Theorem 1.1, while the estimates for $A_k$ are responsible for the KLS problem, Theorem 1.2.

As usual, $C, C_0, C_1, \ldots, c, c_0, c_1, \ldots$ 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's 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 a very simplified corollary. In Section 4, we deduce Theorem 1.2 from Theorem 4.5. The case $p > 8$ and the exponential cases are proved in Theorem 4.7 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 us to reduce the problem at 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 [15].

## 2. Norms of submatrices

We start with a few general preliminaries and notation. 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 $E \subset \{1, \dots, N\}$, $|E|$ 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 every $\varepsilon \in (0, 1)$ there exists an $\varepsilon$-net $\Lambda$ in $B_2^n$ 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 $\Lambda$ is not larger than $(2.5/\varepsilon)^n$.

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

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

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

$$A_k := \sup_{\substack{a \in S^{N-1} \\ |\mathrm{supp}(a)| \le k}} \left| \sum_{i=1}^N a_i X_i \right|, \qquad B_k^2 := \sup_{\substack{a \in S^{N-1} \\ |\mathrm{supp}(a)| \le k}} \left| \left| \sum_{i=1}^N a_i X_i \right|^2 - \sum_{i=1}^N a_i^2 |X_i|^2 \right|. \quad (6)$$

Note that $A_k$ is the supremum of the norms of submatrices consisting of $k$ columns of $A$, while $B_k$ plays a crucial role for RIP estimates. We provide more details on the role of $A_k$ and $B_k$ in the next section.

Recall also a notation from the introduction:

$$M = \max_{i \le N} |X_i|.$$

We now formulate the main technical result, Theorem 2.1, which is the key result for bounds of both $A_k$ and $B_k$. The role of $A_k$ and $B_k$ in RIP estimates will be explained in the next section. We postpone the proof to Section 5.

**Theorem 2.1.** *Let $p > 4$, $\sigma \in (2, p/2)$, $\alpha \in (0, 2]$, $t > 0$, and $\tau, \lambda \geq 1$. Let $X_1, \ldots, X_N$ be independent random vectors in $\mathbb{R}^n$ satisfying hypothesis $\mathbf{H}(\phi)$ with parameter $\tau$ either for $\phi(x) = x^p$ or for $\phi(x) = (1/2) \exp(x^\alpha)$. For $k \leq N$ define $M_1$, $\beta$ and $C_\phi$ in the following two cases.*

**Case 1:** $\phi(x) = x^p$. *Assume that $\lambda \leq p$ and let $C_\phi := e^4$,*

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

*where*

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

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

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

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

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

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

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

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

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

**Remark.** In the case $\phi(x) = x^p$ we will use the following two choices for $\sigma$:

**1.** Choosing $\sigma = p/4$ and assuming $p > 8$ we get

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

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

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

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

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

## Remarks on optimality

**1. The case** $\phi(x) = x^p$, $p > 4$. Let $\tau \ge 1$, $N \ge (64C_2(\sigma, \lambda))^{1/\lambda}$ and $t = (64N^2C_3(\sigma, \lambda, p))^{1/p}$. Then $\beta \le 1/32$ and $\sqrt{t}M \le C_4(\sigma, \lambda, p)(M + M_1)$. Hence with probability larger than 3/4 we have

$$A_k \le C(\sigma, \lambda, p)\left(M + \sqrt{k}(N\tau/k)^{\sigma/p}\right).$$

In Proposition 6.5 below we show that there exist independent random vectors $X_i$ satisfying the conditions of Theorem 2.1 with $\tau = 1$ and such that

$$A_k \ge C(p)\sqrt{k}(N/k)^{1/p}(\ln(2N/k))^{-1/p}$$

with probability at least 1/2. Note that $M = A_1 \le A_k$. Therefore

$$\max\{M, C(p)\sqrt{k}(N/k)^{1/p}(\ln(2N/k))^{-1/p}\} \le A_k \le C(\sigma, \lambda, p)\left(M + \sqrt{k}(N/k)^{\sigma/p}\right) \tag{7}$$

with probability at least 1/4.

**2. The case** $\phi(x) = (1/2)\exp(x^\alpha)$, $\alpha \in [1, 2]$. Let $\lambda = 2$ and $t = (\ln N)^{1/\alpha}$. Then $\beta \le 1/32$. Hence with probability larger than 3/4 we have

$$A_k \le C\left(M + C^{1/\alpha}\sqrt{k}(\ln(6^{1/\alpha}\tau N/k))^{1/\alpha}\right).$$

In Proposition 6.7 below we show that there exist independent random vectors $X_i$'s satisfying the conditions of Theorem 2.1 with $\tau$ bounded by an absolute constant and such that

$$A_k \ge \sqrt{k/2}\,(\ln(N/(k+1)))^{1/\alpha}$$

with probability at least 1/2. Using again $M = A_1 \le A_k$ we observe that

$$\max\{M, \sqrt{k/2}(\ln(N/(k+1)))^{1/\alpha}\} \le A_k \le C\left(M + C^{1/\alpha}\sqrt{k}(\ln(6^{1/\alpha}N/k))^{1/\alpha}\right) \tag{8}$$

with probability at least 1/4.

## 3. Restricted Isometry Property

We need more definitions and notation. Let $T$ be an $n \times N$ matrix and let $1 \leq m \leq N$. The $m$-th *isometry constant* of $T$ is defined as the smallest number $\delta_m = \delta_m(T)$ such that

$$(1 - \delta_m)|z|^2 \leq |Tz|^2 \leq (1 + \delta_m)|z|^2 \tag{9}$$

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

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

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

Let $X_1, \ldots, X_N$ 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 (6)) we clearly have

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

$$\leq \frac{B_m^2}{n} + \max_{i \leq N} \left| \frac{|X_i|^2}{n} - 1 \right|. \tag{10}$$

Thus, in order to have a good bound on $\delta_m(A/\sqrt{n})$ 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 $p > 4$, $\alpha \in (0, 2]$, $\tau \geq 1$ and $1 \leq n \leq N$. Let $X_1, \ldots, X_N$ be independent random vectors in $\mathbb{R}^n$ satisfying hypothesis $\mathbf{H}(\phi)$ with parameter $\tau$ either for $\phi(x) = x^p$ or for $\phi(x) = (1/2)\exp(x^\alpha)$. Let $P(\cdot)$ be as in (2) and $\theta \in (0, 1)$.*

**Case 1:** *$\phi(x) = x^p$. Let $\varepsilon \leq \min\{1, (p - 4)/4\}$. Assume that*

$$\frac{2^8}{\varepsilon\theta\tau} \leq N \leq c\theta(c\varepsilon\theta)^{p/2} n^{p/4} \sqrt{\tau}$$

*and set*

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

*where*

$$C(\theta, \varepsilon, p) := c\left(\frac{p-4}{p}\right)^{2(p+4+2\varepsilon)/(p-4-2\varepsilon)} \varepsilon^{4(2+\varepsilon)/(p-4-2\varepsilon)} \theta^{2p/(p-4-2\varepsilon)}, \tag{11}$$

*and $c$ and $C$ are absolute positive constants.*

**Case 2:** $\phi(x) = (1/2) \exp(x^\alpha)$. *Assume that*

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

*and set*

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

$$\beta := \frac{1}{(10N\tau)^2} \exp\left( \frac{-2m^{\alpha/2}}{(3.5 \ln(2m))^{2\alpha}} \right) + \frac{N^2\tau}{2} \exp(-c(\theta\sqrt{n})^\alpha),$$

*where $c$ and $C$ are absolute positive constants.*

*Then in both cases we have*

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

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

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

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 this follows from the results of Paouris [27, 28] (see also [18, 16] or [5, Lemma 3.3]). 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 $\xi$. If $\xi$ is of variance 1 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 [31, Proposition 1.3]). Another case is when $\xi$ is the Weibull random variable of variance 1, that is, consider $\xi_0$ such that $\mathbb{P}(|\xi_0| > t) = \exp(-t^\alpha)$ for $\alpha \in (0, 2]$ and let $\xi = \xi_0/\sqrt{\mathbb{E}\xi_0^2}$. By [5, Lemma 3.4] (see also [11, Theorem 1.2.8]), $P(\theta)$ satisfies (37) below.

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

$$m = \left[ C(\theta, p)n \left( \frac{N\tau}{n} \right)^{-4/(p-4)} \left( \ln \frac{2N\tau}{n} \right)^{-8/(p-4)} \right].$$

Moreover, Proposition 6.6 below shows that for $q > p > 4$ there are independent random vectors $X_i$ satisfying hypothesis $\mathbf{H}(\phi)$ with parameter $\tau = \tau(p, q)$ and such that for $\theta = 1/2$, $N \le C(p, q)n^{p/4}(\ln(2N/n))^{-p/2}$ one cannot get a better estimate than

$$m \le 8(N/n)^{-2/(q-2)}n.$$

**4. Optimality.** In Case 2 with $\tau$ bounded by an absolute constant and $\alpha \in [1, 2]$, let $c_0 n \le N \le c_1 \exp(c_2 n^{\alpha/2})$, $\theta = 0.4$ and assume that $P(\theta/2)$ is small enough. Then

$\mathbb{P}(\delta_m \leq 1/2) \geq 1/2$ provided that $m = n(C \ln(C^{2/\alpha} N/n))^{2/\alpha}$. Proposition 6.7 below shows that the estimate for $m$ is sharp, that is, in general $m$ cannot be larger than $m = n(C \ln(2N/n))^{2/\alpha}$.

*Proof of Theorem 3.1.* We first pass to the subset $\Omega_0$ of our initial probability space where

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

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

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

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

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

Together with (10) this proves $\delta_m(A/\sqrt{n}) \leq \theta$. Thus we only need to check when the estimates for $\beta$ and $M_1$ are satisfied.

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

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

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

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

with $C(\theta, \varepsilon, p)$ defined in (11) 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 $\lambda$ we have

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

provided that $2^8/(\varepsilon\theta\tau) \leq N \leq 2^{-4}\theta\sqrt{\tau}(0.4c\varepsilon\theta)^{p/2}n^{p/4}$. 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 $N\tau/m \geq 2^{1/\alpha}$ as $N\tau \geq 2^{1/\alpha}n$. Therefore for some absolute constant $C$,

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

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

$$m \leq C_1^{-2/\alpha} \theta^2 n \big( \ln(C_1^{2/\alpha} N\tau/(\theta^2 n)) \big)^{-2/\alpha}$$

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 $\lambda$ we have

$$\beta \leq \frac{1}{(10N\tau)^2} \exp\left( \frac{-2m^{\alpha/2}}{(3.5 \ln(2m))^{2\alpha}} \right) + \frac{N^2\tau}{2} \exp(-c(\theta\sqrt{n})^\alpha) \leq 2^{-9}\theta^2$$

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

## 4. Approximating the covariance matrix

We start with the following $\varepsilon$-net argument for bilinear forms, which will be used below.

**Lemma 4.1.** *Let $m \geq 1$ be an integer and $T$ be an $m \times m$ matrix. Let $\varepsilon \in (0, 1/2)$ and $\mathcal{N}$ be an $\varepsilon$-net in $B_2^m$ (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, for every $x \in \mathbb{R}^m$ we have $\langle Sx, x \rangle = 2\langle Tx, x \rangle$. This concludes the proof. □

Now we can prove the following technical lemma, which 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 [3]. Other versions of the lemma appeared in [5, 2]. Its proof uses the symmetrization method as in [25].

Recall that $(s_i^*)_i$ denotes a non-increasing rearrangement of $(|s_i|)_i$.

**Lemma 4.2.** *Let $\tau \geq 1$, $1 \leq k < N$ and $X_1, \ldots, X_N$ be independent random vectors in $\mathbb{R}^n$. Let $p \geq 2$ and $\alpha \in (0, 2]$. Let $\phi$ be either $\phi(t) = t^p$ in which case we set $C_\phi := 8\tau^{2/p} N^{2/\min(p,4)}$ and assume*

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

*or $\phi(t) = (1/2)\exp(t^\alpha)$ in which case we assume that $X_i$'s satisfy hypothesis $\mathbf{H}(\phi)$ with parameter $\tau$ and set $C_\phi := 8\sqrt{C_\alpha N \tau}$, where $C_\alpha := (8/\alpha)\Gamma(4/\alpha)$ and $\Gamma(\cdot)$ 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*

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

The term involving $Z$ in the upper bound will be bounded later using general estimates in Lemma 4.4. Thus Lemma 4.2 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 of Lemma 4.2.* Let $\Lambda \subset \mathbb{R}^n$ be a $(1/4)$-net in 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. $\pm 1$ Bernoulli random variables of parameter $1/2$. By Hoeffding's inequality, for every $t > 0$ and every $(s_i)_{1 \leq i \leq N} \in \mathbb{R}^N$,

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

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

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

Also, it is easy to check using (6) 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 $a \in S^{n-1}$,

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

Note that $\sum_{i=1}^{N} \varepsilon_i \langle X_i, a \rangle^2 = \sum_{i \in E} \langle X_i, a \rangle^2 - \sum_{i \in E^c}^{N} \langle X_i, a \rangle^2$ for some set $E \subset \{1, \ldots, N\}$ and we can apply a union bound argument indexed by $\Lambda$ together with Lemma 4.1. We get

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

$$\geq 1 - 2 \cdot 9^n \exp(-t^2/2).$$

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

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

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

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

with probability larger than

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

Now we transfer the result from Bernoulli random variables to centered random variables (see [21, Section 6.1]). By the triangle inequality, for every $s, t > 0$,

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

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

To conclude the proof it is enough to find $s$ such that $m(s) \geq 1/2$. To this end we will use a general Lemma 4.3 below. First consider $\phi(t) = t^p$. For $a \in S^{n-1}$, set $Z_i = |\langle X_i, a \rangle|^2 / \tau^{2/p}$ and $q = p/2$. Then by Lemma 4.3 we have $m(s) \geq 1/2$ for $s = 4\tau^{2/p} N^{1/r}$ 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 $i \leq N$, using hypothesis $\mathbf{H}(\phi)$ we have

$$\mathbb{E}|\langle X_i, a \rangle|^4 \leq 8\tau \int_0^\infty t^3 \exp(-t^\alpha) dt = \frac{8\tau}{\alpha} \Gamma \left( \frac{4}{\alpha} \right) =: \tau C_\alpha.$$

Given $a \in S^{n-1}$, set $Z_i := |\langle X_i, a \rangle|^2 / \sqrt{\tau C_\alpha}$. Then $\mathbb{E}Z_i^2 \leq 1$. Applying again Lemma 4.3 (with $q = 2$), we observe that $m(s) \geq 1/2$ for $s = 4\sqrt{C_\alpha N}$. □

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 4.3.** *Let $q \geq 1$ and $Z_1, \ldots, Z_N$ be independent non-negative random variables satisfying*

$$\forall 1 \leq i \leq N \quad \mathbb{E}Z_i^q \leq 1.$$

*Let $r := \min(q, 2)$. Then*

$$\forall z \geq 4N^{1/r} \quad \mathbb{P}\Big(\Big|\sum_{i=1}^N (Z_i - \mathbb{E}Z_i)\Big| \leq z\Big) \geq \frac{1}{2}.$$

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

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

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

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

By Markov's inequality we get

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

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

The following lemma is standard (cf. [21, Lemma 5.8], which however contains a misprint).

**Lemma 4.4.** *Let $q > 0$ and let $Z_1, \ldots, Z_N$ be independent non-negative random variables satisfying*

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

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

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

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

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

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

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

Hence if $eNt^{-q} \leq 1/2$, then $\mathbb{P}(\sup_{i \geq k} i^{1/q} Z_i^* > t) \leq (2eNt^{-q})^k$. Since this 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

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

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

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

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

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

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

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

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

We are now ready to tackle the problem of approximating the covariance matrix by empirical covariance matrices, under hypothesis $\mathbf{H}(\phi)$ with $\phi(t) = t^p$. As our proof works

for all $p > 4$, we also include the case $p > 8$ originally solved in [25] (under an additional assumption on $\max_i |X_i|$). For clarity, we split the result into two theorems. The case $4 < p \le 8$ 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 $\|(1/N)AA^\top - I\| < 1$ 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 the vector $Z$, hypothesis $\mathbf{H}(\phi)$ with $\phi(t) = t^2$ is satisfied but $N \sim n \log n$ is needed for the covariance matrix to be well approximated by empirical covariance matrices with probability $1/2$.

We also mention that we do not know how sharp the power $\gamma/p$ appearing in the bound below is. In particular, it is not clear if it can be improved to $1/2$.

**Theorem 4.5.** *Let $4 < p \le 8$ and $\phi(t) = t^p$. Let $X_1, \ldots, X_N$ 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*

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

*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 is the following corollary.

**Corollary 4.6.** *Under the assumptions of Theorem 4.5, assuming additionally that $\max_i |X_i|^2 \le C n^{\gamma/p} N^{1-\gamma/p}$ 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 4.7.** *There exists a universal positive constant $C$ such that the following holds. Let $p > 8$ and $\alpha \in (0, 2]$. Let $\phi$ and $C_\phi$ 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 $X_1, \ldots, X_N$ be independent random vectors in $\mathbb{R}^n$ satisfying hypothesis $\mathbf{H}(\phi)$. In the case $\phi(t) = t^p$ 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 $\phi(t) = (1/2)\exp(t^\alpha)$, 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 $1 - p_0$,*

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

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

*Proof of Theorems 4.5 and 4.7.* We first consider the case $\phi := t^p$. Note that in this case

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

Thus, by Lemma 4.2 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 $\phi(t) = t^p$ we apply Lemma 4.4 with $Z_i = |\langle X_i, a\rangle|^4, i \leq N, q = p/4 > 1$ and $s = 9e$. It gives

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

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:** $4 < p \leq 8$ (Theorem 4.5). 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

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

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

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

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

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

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

Combining all estimates and noticing that $(p-4)^{-\gamma} < 2$, we conclude 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 4.7). 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 $M_1 \leq C\sqrt{n}(N/n)^{1/4}$ and

$$\beta \leq \left(\frac{2(3p-8)}{5e(p-8)N}\right)^{(p-4)/2} \frac{1}{p-5} + \frac{(3p-8)^p}{4(6(p-8))^p N^{(p-8)/4} n^{p/4}}$$
$$\leq \left(\frac{3p-8}{6(p-8)}\right)^p N^{-(p-8)/4} n^{-p/4} \leq 1/64$$

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

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

Combining all estimates shows 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 4.7). As in Case 2 we apply Lemma 4.2. It implies that it is enough to estimate $A^2 + \sqrt{n}\, Z + \sqrt{C(\alpha)N}$, with $C(\alpha)$ from Lemma 4.2, and the corresponding probabilities. A direct calculation shows that in this case we have, for $C'_\alpha := (4/\alpha)^{1/\alpha}$ and $t > 1$,

$$\mathbb{P}\big((|X|/C'_\alpha)^4 > t\big) \leq 2\exp(C'_\alpha)t^{\alpha/4} \leq 1/t^2.$$

We apply Lemma 4.4 with $Z_i = |\langle X_i, a\rangle|^4/\sqrt{C'_\alpha}$, $i \leq N$, $q = 2$ and $s = 9e$. It gives

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

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

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

Note that

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

Then for absolute positive constants $C, C'$,

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

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

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

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

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

## 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 (6).

### 5.1. Bilinear forms of independent vectors

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

$$Q(a, T, S) := \left| \left\langle \sum_{i \in T} a_i X_i, \sum_{j \in S} a_j X_j \right\rangle \right|, \tag{12}$$

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

The following two lemmas are in the spirit of [25, Lemma 2.3]. Recall that $(s_i^*)_i$ denotes a non-increasing rearrangement of $(|s_i|)_i$.

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

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

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

*Proof.* Let $E \subset \{1, \ldots, N\}$ 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 $\ell$ 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

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

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

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

where $m = \lceil (\gamma - 1/2)k \rceil$ and we have used $s - \ell + 1 \geq k/2 - \lceil (1-\gamma)k \rceil + 1 > (\gamma - 1/2)k$.

If $|F_1| < k/2$ then $|F_2| \geq k/2$ and we proceed similarly interchanging the roles of $F_1$ and $F_2$ to obtain

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

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

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

*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 see that if

$$\mathbb{P}(|\langle X_i, a \rangle| \geq t) \leq t^{-p} \tag{13}$$

then

$$\mathbb{P}\left(\sum_{i=m}^{m+\ell-1} U_i^* > t A_k\right) \leq (2\tau)^k \left(\frac{t\sqrt{m}}{\ell}\right)^{-mp}. \tag{14}$$

Note that the condition (13) is satisfied if

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

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

$$\mathbb{P}(|\langle X_i, a\rangle| \geq t) \leq 2\exp(-t^\alpha) \tag{15}$$

then

$$\mathbb{P}\left(\sum_{i=m}^{m+\ell-1} U_i^* > t A_k\right) \leq (2\tau)^{k+m} \exp\left(-m\left(\frac{t\sqrt{m}}{\ell}\right)^\alpha\right). \tag{16}$$

Note that the condition (15) is satisfied if

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

*Proof of Lemma 5.2.* 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 the cardinality of $F_1$ is smaller than $m$, the estimate for probability is trivial). Since $|F_1| \leq k$, we obtain

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

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

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

$$\leq (2\tau)^k \max_{\substack{F \subset F_1 \\ |F|=m}} \prod_{i \in F} \left(\phi\left(\frac{t}{\ell|a_i|}\right)\right)^{-1}.$$

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

$$\prod_{i \in F} (\phi(s/|a_i|))^{-1} \leq (\phi(s\sqrt{m}))^{-m}.$$

Indeed, this estimate is equivalent to

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

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

$$\mathbb{P}\Big(\sum_{i=m}^{m+\ell-1} V_i^* > t A_k\Big) \leq (2\tau)^k (\phi(t\sqrt{m}/\ell))^{-m}.$$

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 $\phi$ is increasing, we obtain the last inequality, completing the proof. □

### 5.2. Estimates for the off-diagonal part of bilinear forms

For $1 \leq k \leq N$ and $I \subset \{1, \ldots, N\}$ we define

$$Q_k(I) := \sup_{\substack{E \subset \{1,\ldots,N\} \\ |E| \leq k}} \sup_{a \in B_2^E} Q(a, E \cap I, E \cap I^c). \tag{17}$$

Lemmas 5.1, 5.2 and 4.1 imply the following proposition.

**Proposition 5.3.** *Let $\tau \geq 1$ and $X_1, \ldots, X_N$ be independent random vectors in $\mathbb{R}^n$ satisfying hypothesis $\mathbf{H}(\phi)$ with parameter $\tau$ for some function $\phi \in \mathcal{M}$. Let $\varepsilon \in (0, 1/2)$, $2 \leq k \leq N$, $I \subset \{1, \ldots, N\}$, $\gamma \in (1/2, 1)$, and $\gamma_0 := \gamma - 1/2$. Then for every $t > 0$,*

$$\mathbb{P}\Big(Q_k(I) > \frac{Q_{[\gamma k]}(I) + t A_k}{1 - 2\varepsilon}\Big) \leq \exp\Big(k\Big(\ln \frac{5\tau e N}{k\varepsilon} - \gamma_0 \ln \phi\Big(\frac{t\sqrt{\gamma_0 k}}{(1 - \gamma)k + 1}\Big)\Big)\Big).$$

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

$$\mathbb{P}(Q_\ell(I) > t M) \leq \frac{N^2 \tau}{4\phi(4t/\ell)}.$$

*Proof.* For every $E \subset \{1, \ldots, N\}$ with $|E| = k$ let $\mathcal{N}_E$ be an $\varepsilon$-net in $B_2^E$ of cardinality at most $(2.5/\varepsilon)^k$. Let $\mathcal{N}$ denote the union of the $\mathcal{N}_E$'s. Lemma 4.1 yields

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

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

$$Q_k(I) \leq (1 - 2\varepsilon)^{-1}\Big(\sup_{\substack{E \subset \{1,\ldots,N\} \\ |E| \leq \gamma k}} \sup_{a \in \mathcal{N}} Q(a, E \cap I, E \cap I^c) + t A_k\Big)$$

occurs with probability at least

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

This implies the first estimate.

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

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

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

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

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

Therefore, by the union bound,

$$
\mathbb{P}(Q_\ell(I) > tM_1) \leq \mathbb{P}(M_0 > 4tM_1/\ell)
$$

$$
\leq \sum_{i \in I} \sum_{j \in I^c} \mathbb{P}(|\langle X_i, X_j \rangle| > 4tM_1/\ell).
$$

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 the variables $\langle X_i, z \rangle$, we obtain

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

**Proposition 5.4.** *Let $1 \leq k \leq N$. Let $\tau \geq 1$ and $X_1, \ldots, X_N$ be independent random vectors in $\mathbb{R}^n$ satisfying hypothesis $\mathbf{H}(\phi)$ with parameter $\tau$ for some function $\phi \in \mathcal{M}$. Let $t > 0$ and $\lambda \geq 1$.*

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

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

*occurs with probability at least*

$$
1 - \left(\frac{2(\sigma + \lambda)}{5\tau e N(\sigma - 2)}\right)^\lambda \frac{1}{2\lambda - 1} - \frac{N^2\tau(\sigma + \lambda)^p}{4(2t(\sigma - 2))^p} \tag{18}
$$

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

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

*with probability at least*

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

*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

$$1 \leq k_m < \frac{1}{1-\gamma} \leq k_{m-1}. \tag{19}$$

By Proposition 5.3 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

$$1 - 2\sum_{s=0}^{m-1}\exp\left(k_s\left(\ln\frac{5\tau eN}{k_s\varepsilon_s} - \gamma_0\ln\phi\left(\frac{t_s\sqrt{\gamma_0 k}}{(1-\gamma)k+1}\right)\right)\right). \tag{20}$$

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

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

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 (20) and by our choice of $t_s$,

$$Q_k(I) \leq \exp\left(4 \sum_{s=0}^{m-1} \varepsilon_s\right)\left(Q_{k_m}(I) + A_k \sum_{s=0}^{m-1} t_s\right) \tag{21}$$

with probability at least

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

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

$$1 - 2 \exp\left(-\frac{\varepsilon}{1-\gamma} \ln(5\tau e(1-\gamma)N)\right) \sum_{s=0}^{m-1} \varepsilon_s^{k_s \varepsilon}. \tag{22}$$

Thus it is enough to choose $\varepsilon_s$ appropriately 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 $\phi$.

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

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

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

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

Using again $k_{m-1} \geq (1-\gamma)^{-1}$, we conclude that the probability in (22) is larger than

$$1 - (5\tau eN(1-\gamma))^{-\lambda} \frac{2}{2\lambda - 1}. \tag{23}$$

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

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

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

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

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

$$\sum_{s=0}^{m-1} t_s \leq \frac{2(1-\gamma)k^{1/2-b}(5\tau eN)^b}{\sqrt{\gamma_0}} \sum_{s=0}^{m-1} (s+2)^{\delta b} \gamma^{s(1/2-b)}. \tag{24}$$

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 $\ln(1/\gamma) \geq 1 - \gamma$, we observe that the previous quantity does not exceed

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

Coming back to (24), we get

$$\sum_{s=0}^{m-1}t_s \leq \frac{8k^{1/2-b}(5\tau eN)^b}{(1/2-b)^{1+2b}(1-\gamma)^{2b}\sqrt{\gamma-1/2}}. \tag{25}$$

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 $\gamma$, we have $b = \sigma/p < 1/2$, since $\sigma < p/2$. Thus from (25) and (23),

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

with probability larger than

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

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

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

and apply the "moreover" part of Proposition 5.3 (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 5.3 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 (19) we have $k_m < 3$, hence, by (21),

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

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

$$\sum_{s=0}^{m-1} \varepsilon_s \leq \frac{C}{\alpha} \tag{26}$$

for a positive absolute constant $C$.

We have

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

By (19) we have $k_m < 3 \leq k_{m-1}$, hence

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

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

By the choice of $\varepsilon_s$ we obtain

$$\sum_{s=0}^{m-1} \sqrt{k_s}\left(\ln\frac{1}{2\varepsilon_s}\right)^{1/\alpha} \leq \sqrt{k}\sum_{s=0}^{m-1}(s+2)^{-2} \leq 3\sqrt{k}. \tag{27}$$

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

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

$$\leq \sqrt{k}\left(\sum_{s=0}^{m-1}\left(\frac{2}{3}\right)^{s/2}2^{1/\alpha}\left(\ln\frac{20\tau eN}{k}\right)^{1/\alpha} + \sum_{s=0}^{m-1}\left(\frac{2}{3}\right)^{s/2}(2s\ln 3)^{1/\alpha}\right)$$

$$\leq C_1^{1/\alpha}\sqrt{k}\left(\left(\ln\frac{20\tau eN}{k}\right)^{1/\alpha} + \Gamma(1+1/\alpha)\right),$$

where $C_1$ is an absolute positive constant. This together with (27) implies that

$$\sum_{s=0}^{m-1} t_s \leq (C_2(1+\varepsilon))^{1/\alpha}\sqrt{k}\left(\left(\ln\frac{20\tau eN}{k}\right)^{1/\alpha} + \Gamma(1+1/\alpha)\right), \tag{28}$$

where $C_2$ is an absolute positive constant.

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

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

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

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

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

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 the probability in (22) is at least

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

Finally, to estimate $Q_2(I)$ we apply the "moreover" part of Proposition 5.3 (with $\ell = 2$). Choosing $\varepsilon = \lambda/3$ and combining estimates (26) and (28) 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. We need two simple lemmas.

**Lemma 5.5.** *Let $\beta \in (0, 1)$. Let $\mathbb{P}_1$ and $\mathbb{P}_2$ be probability measures on $\Omega_1$ and $\Omega_2$ 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*

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

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

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

Clearly,

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

Then

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

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

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

The following lemma is obvious.

**Lemma 5.6.** *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 5.6 we have

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

We deduce that

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

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

Let $I \subset \{1, \ldots, N\}$ be fixed. Proposition 5.4 implies

$$\mathbb{P}(Q_k(I) \leq M_0) \geq 1 - \beta, \tag{29}$$

where

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

Consider two probability spaces, $\{I : I \subset \{1, \ldots, N\}\}$ with the normalized counting measure $\mu$, and our initial probability space $(\Omega, \mathbb{P})$, on which the $X_i$'s are defined. By (29) we observe that the $\mu \otimes \mathbb{P}$ probability of the event $V := \{Q_k(I) \leq M_0\}$ is at least $1 - \beta$. Then Lemma 5.5 implies that there exists $W \subset \Omega$ such that $\mathbb{P}(W) \geq 1 - \sqrt{\beta}$ and $\mu(\{Q_k(I) \leq M_0\}) \geq 1 - \sqrt{\beta}$ for every $\omega \in W$. Since $Q_k(I) \leq A_k^2$, we deduce that for every $\omega \in W$,

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

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

$$A_k^2 \leq \frac{4M_0 + \max_{i \leq N} |X_i|^2}{1 - 4\sqrt{\beta}} \quad \text{and} \quad B_k^2 \leq \frac{4(M_0 + \sqrt{\beta} \max_{i \leq N} |X_i|^2)}{1 - 4\sqrt{\beta}}. \tag{30}$$

Therefore

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

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

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

which proves the estimate for $A_k$. Plugging this into (30), we also observe

$$\begin{aligned} B_k^2 &\leq \gamma (4\sqrt{\beta}\, M^2 + 4C_\phi t M + \gamma M_1^2 + \sqrt{\gamma}\, M M_1 + 2\sqrt{C_\phi \gamma t M}\, M_1) \\ &\leq \gamma (4\sqrt{\beta}\, M^2 + 8C_\phi t M + 2\gamma M_1^2 + \sqrt{\gamma}\, M M_1). \end{aligned} \qquad \square$$

## 6. Optimality

In this section we discuss optimality of the estimates in Theorems 2.1 and 3.1. In Propositions 6.5–6.7 we will prove results justifying the remarks on optimality following those theorems.

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

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

$$\mathbb{P}(A_m \geq t) \geq \frac{1}{2} \quad \text{whenever} \quad \mathbb{P}\left( |X_{11}| \geq \frac{t}{\sqrt{m}} \right) \geq \frac{m+1}{N}. \tag{31}$$

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

$$\begin{aligned} A_m = \sup_{a \in U_m} \left| \sum_{j=1}^N a_j X_j \right| &\geq \sup_{a \in U_m} \left| \sum_{j=1}^N a_j X_{1j} \right| \geq \sup_{\substack{a \in U_m \\ a_j \in \{\pm 1/\sqrt{m}, 0\}}} \left| \sum_{j=1}^N a_j X_{1j} \right| \\ &= \frac{1}{\sqrt{m}} \sum_{j=1}^m X_{1j}^* \geq \sqrt{m}\, X_{1m}^*. \end{aligned}$$

Therefore, using independence, we have

$$\mathbb{P}(A_m \geq t) \geq \mathbb{P}(X_{1m}^* \geq t/\sqrt{m}) = \mathbb{P}(Y \geq m),$$

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

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

Thus $\mathbb{P}(A_m \geq t) \geq 1/2$ whenever $m \leq \lfloor Nv \rfloor$. This implies the result. $\square$

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

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

$$\mathbb{P}(A_m \geq t\sqrt{m}) \geq 1/2.$$

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

$$mt^2 \leq 2n.$$

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

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

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

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

which implies the result. $\qquad\square$

In order to show that a matrix with i.i.d. entries satisfies condition $\mathbf{H}(\phi)$ with $\phi(t) = t^p$ we need Rosenthal's inequality [29] (see also [17]). As usual, by $\|\cdot\|_q$ for a random variable $\xi$ we mean its $L_q$-norm and for an $a \in \mathbb{R}^n$ its $\ell_q$-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 the standard symmetrization argument (i.e., passing from random variables $\xi_i$ to $(\xi_i - \xi_i')$'s, where $(\xi_i')$'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 $\xi_1, \dots, \xi_n$ be i.i.d. centered random variables with finite $q$-th moment. Then there exists a positive absolute constant $C$ such that*

$$\frac{1}{2} M_q \leq \left\| \sum_{i=1}^{n} a_i \xi_i \right\|_q \leq C \frac{q}{\ln q} M_q \tag{32}$$

*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 [31, Proposition 1.3].

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

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

*where $C$ is a positive absolute constant.*

*Proof.* Let $\xi_1, \ldots, \xi_n$ be i.i.d. random variables distributed as $\xi$. We apply Rosenthal's inequality to the random variables $\xi_i^2 - 1$ with $q = p/2$ and $a = (1, \ldots, 1)$. Then

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

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

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

The result follows by the union bound. $\qquad\square$

As is mentioned in the remarks on optimality following Theorem 2.1, the next proposition gives a lower bound for $A_m$ to be compared with Case 1 of Theorem 2.1.

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

$$\forall 1 \le i \le N \,\, \forall a \in S^{n-1} \quad \mathbb{E}|\langle X_i, a \rangle|^p \le 1 \tag{33}$$

*and such that*

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

*where $C$ is an absolute positive constant.*

*Proof.* Let $\lambda \ge 1$ to be set later and define

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

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 $\xi/a_p$. 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 (31) shows that $\mathbb{P}(A_m \geq t) \geq 1/2$ 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 $\lambda$ with $\lambda^p - 1 = N/(m+1)$, we obtain $\mathbb{P}(A_m \geq t) \geq 1/2$ provided that

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

Finally, to satisfy condition (33), we pass from a matrix $A$ to $A' := A/c_p = (X_{ij}/c_p)_{ij}$, where $c_p \leq Cp/\ln p$ is the constant in Rosenthal's inequality (32). By Rosenthal's inequality, the sequence of columns of $A'$ satisfies condition (33). □

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

**Proposition 6.6.** *Let $q > p > 2$, $n \leq N$ and $m \leq N$. There exist an absolute positive constant $C$ and an $n \times N$ matrix $A$ whose columns $X_1, \ldots, X_N$ are independent random vectors satisfying*

$$\forall 1 \leq i \leq N \; \forall a \in S^{n-1} \quad \mathbb{E}|\langle X_i, a \rangle|^p \leq \left(\frac{Cp}{\ln p}\right)^p \frac{q}{q - p}\left(\frac{q - 2}{q}\right)^{p/2}, \tag{34}$$

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

$$\mathbb{P}\left(\max_{i \leq N}\left|\frac{|X_i|^2}{n} - 1\right| \geq t\right) \leq t^{p/2} \tag{35}$$

*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 $\mathrm{RIP}_m(\delta)$ 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

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

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 $\xi/a_2$. 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 (32) implies (34), and Corollary 6.4 implies (35).

Now we estimate $A_m$ for the matrix $A$ whose columns are $(X_{ij})_i$, $j \leq N$. Since $\mathbb{P}(|\xi| > s) = s^{-q}$ for $s \geq 1$, by (31) we obtain $\mathbb{P}(A_m \geq t) \geq 1/2$ 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 (see Remark 4 following that theorem) as well as optimality of the bounds for $A_m$ in Case 2 of Theorem 2.1 (see remarks on optimality following that theorem).

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

$$\forall 1 \leq i \leq N \; \forall a \in S^{n-1} \quad \mathbb{E}\exp(|\langle X_i, a\rangle|^\alpha) \leq C \tag{36}$$

*and*

$$\mathbb{P}\left(\max_{i \leq N}\left|\frac{|X_i|^2}{n} - 1\right| \geq \frac{\sqrt{2}-1}{2}\right) \leq 2\exp(-cn^{\alpha/2}), \tag{37}$$

*and such that*

$$\mathbb{P}\left(A_m \geq \sqrt{\frac{m}{2}}\left(\ln\frac{N}{m+1}\right)^{1/\alpha}\right) \geq \frac{1}{2}. \tag{38}$$

*Additionally, if $n \leq N$ and if $A$ satisfies $\mathrm{RIP}_m(\delta)$ for some $\delta < 1$ with probability greater than $1/2$, then*

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

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

$$\mathbb{E}\exp(|\xi|^\alpha/2) = 2 \quad \text{and} \quad a := \mathbb{E}\xi^2 = \Gamma(2/\alpha + 1) \in [1, 2].$$

Let $X_{ij}$, $i \le n$, $j \le N$, be i.i.d. copies of $\xi/\sqrt{a}$, $A = (X_{ij})_{ij}$, and let $X_j$'s be its columns. Applying [5, Lemma 3.4] (see also [11, Theorem 1.2.8]) we observe that the $X_i$'s satisfy conditions (36) and (37). By (31) we observe that $\mathbb{P}(A_m \ge t) \ge 1/2$ provided that

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

Thus it is enough to take

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

This proves (38).

Finally, the "additionally" part follows by Lemma 6.2. $\qquad\square$

# References

[1] Adamczak, R., Guédon, O., Litvak, A. E., Pajor, A., Tomczak-Jaegermann, N.: Smallest singular value of random matrices with independent columns. C. R. Math. Acad. Sci. Paris **346**, 853–856 (2008) Zbl 1148.15003 MR 2441920

[2] Adamczak, R., Latała, R., Litvak, A., Pajor, A., Tomczak-Jaegermann, N.: Geometry of log-concave ensembles of random matrices and approximate reconstruction. C. R. Math. Acad. Sci. Paris **349**, 783–786 (2011) Zbl 1225.60013 MR 2825941

[3] Adamczak, R., Litvak, A. E., Pajor, A., Tomczak-Jaegermann, N.: Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. J. Amer. Math. Soc. **23**, 535–561 (2010) Zbl 1206.60006 MR 2601042

[4] Adamczak, R., Litvak, A. E., Pajor, A., Tomczak-Jaegermann, N.: Sharp bounds on the rate of convergence of the empirical covariance matrix. C. R. Math. Acad. Sci. Paris **349**, 195–200 (2011) Zbl 1208.60006 MR 2769907

[5] Adamczak, R., Litvak, A., Pajor, A., Tomczak-Jaegermann, N.: Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling. Constr. Approx. **34**, 61–88 (2011) Zbl 1222.52009 MR 2796091

[6] Bai, Z. D., Silverstein, J., Yin, Y. Q.: A note on the largest eigenvalue of a large dimensional sample covariance matrix. J. Multivariate Anal. **26**, 166–168 (1988) Zbl 0652.60040 MR 0963829

[7]  Bai, Z. D., Yin, Y. Q.: Limit of the smallest eigenvalue of a large dimensional sample covariance matrix. Ann. Probab. **21**, 1275–1294 (1993)  Zbl 0779.60026  MR 1235416

[8]  Bourgain, J.: Random points in isotropic convex sets. In: Convex Geometric Analysis (Berkeley, CA, 1996), Math. Sci. Res. Inst. Publ. 34, Cambridge Univ. Press, Cambridge, 53–58 (1999)  Zbl 0941.52003  MR 1665576

[9]  Candés, E. J., Romberg, J., Tao, T.:  Stable signal recovery from incomplete and inaccurate measurements. Comm. Pure Appl. Math. **59**, 1207–1223 (2006)  Zbl 1098.94009  MR 2230846

[10]  Candés, E. J., Tao, T.: Decoding by linear programming. IEEE Trans. Inform. Theory **51**, 4203–4215 (2005)  Zbl 1264.94121  MR 2243152

[11]  Chafaï, D., Guédon, O., Lecué, G., Pajor, A.: Interactions between Compressed Sensing Random Matrices and High Dimensional Geometry. Panoramas et Synthèses 37, Soc. Math. France, Paris (2012)  Zbl 06208972  MR 3113826

[12]  Donoho, D. L.: Compressed sensing. IEEE Trans. Information Theory **52**, 1289–1306 (2006)  Zbl 1288.94016  MR 2241189

[13]  Edelman, A.: Eigenvalues and condition numbers of random matrices. SIAM J. Matrix Anal. Appl. **9**, 543–560 (1988)  Zbl 0678.15019  MR 0964668

[14]  van de Geer, S., Muro, A.: On higher order isotropy conditions and lower bounds for sparse quadratic forms. Electron. J. Statist. **8**, 3031–3061 (2014)  Zbl 1308.62148  MR 3301300

[15]  Guédon, O., Litvak, A. E., Pajor, A., Tomczak-Jaegermann, N.: Restricted isometry property for random matrices with heavy tailed columns. C. R. Math. Acad. Sci. Paris **352**, 431–434 (2014)  Zbl 1290.15025  MR 3194251

[16]  Guédon, O., Milman, E.: Interpolating thin-shell and sharp large-deviation estimates for isotropic log-concave measures. Geom. Funct. Anal. **21**, 1043–1068 (2011)  Zbl 1242.60012  MR 2846382

[17]  Johnson, W. B., Schechtman, G., Zinn, J.: Best constants in moment inequalities for linear combinations of independent and exchangeable random variables. Ann. Probab. **13**, 234–253 (1985)  Zbl 0564.60020  MR 0770640

[18]  Klartag, B.: Power-law estimates for the central limit theorem for convex sets. J. Funct. Anal. **245**, 284–310 (2007)  Zbl 1140.52004  MR 2311626

[19]  Koltchinskii, V., Mendelson, S.: Bounding the smallest singular value of a random matrix without concentration. Int. Math. Res. Notices **2015**, 12991–13008  Zbl 1331.15027  MR 3431642

[20]  Lecué, G., Mendelson, S.: Sparse recovery under weak moment assumptions. J. Eur. Math. Soc. **19**, 881–904 (2017)  MR 3612870

[21]  Ledoux, M., Talagrand, M.: Probability in Banach Spaces. Isoperimetry and Processes. Springer, Berlin (1991)  Zbl 0748.60004  MR 1102015

[22]  Litvak, A. E., Spektor, S.: Quantitative version of a Silverstein's result. In: Geometric Aapects of Functional Analysis, Lecture Notes in Math. 2116, Springer, 335–340 (2014)  Zbl 1315.60007  MR 3364695

[23]  Mendelson, S., Pajor, A., Tomczak-Jaegermann, N.:  Reconstruction and subgaussian operators. Geom. Funct. Anal. **17**, 1248–1282 (2007)  Zbl 1163.46008  MR 2373017

[24]  Mendelson, S., Paouris, G.: On generic chaining and the smallest singular value of random matrices with heavy tails. J. Funct. Anal. **262**, 3775–3811 (2012)  Zbl 1242.60008  MR 2899978

[25]  Mendelson, S., Paouris, G.: On singular values of matrices. J. Eur. Math. Soc. **16**, 823–834 (2014)  Zbl 1290.60006  MR 3191978

[26]  Oliveira, R. I.: The lower tail of random quadratic forms with applications to ordinary least squares. Probab. Theory Related Fields **166**, 1175–1194 (2016)  Zbl 06657184  MR 3568047

[27] Paouris, G.: Concentration of mass on convex bodies. Geom. Funct. Anal. **16**, 1021–1049 (2006)  Zbl 1114.52004  MR 2276533

[28] Paouris, G.: Small ball probability estimates for log-concave measures. Trans. Amer. Math. Soc. **364**, 287–308 (2012)  Zbl 1248.60027  MR 2833584

[29] Rosenthal, H.: On the subspaces of $L_p$ ($p > 2$) spanned by sequences of independent random variables. Israel J. Math. **8**, 273–303 (1970)  Zbl 0213.19303  MR 0271721

[30] Silverstein, J.: On the weak limit of the largest eigenvalue of a large dimensional sample covariance matrix. J. Multivariate Anal. **30**, 307–311 (1989)  Zbl 0692.60022  MR 1015375

[31] Srivastava, N., Vershynin, R.: Covariance estimation for distributions with $2 + \varepsilon$ moments. Ann. Probab. **41**, 3081–3111 (2013)  Zbl 1293.62121  MR 3127875

[32] Szarek, S.: Condition numbers of random matrices. J. Complexity **7**, 131–149 (1991)  Zbl 0760.15018  MR 1108773

[33] Tikhomirov, K. E.: The smallest singular value of random rectangular matrices, with no moment assumptions on entries. Israel J. Math. **212**, 289–314 (2016)  Zbl 1346.15038  MR 3504328

[34] Vershynin, R.: How close is the sample covariance matrix to the actual covariance matrix?. J. Theoret. Probab. **25**, 655–686 (2012)  Zbl 06100921  MR 2956207

[35] Yaskov, P.: Lower bounds on the smallest eigenvalue of a sample covariance matrix. Electron. Comm. Probab. **19**, no. 83, 10 pp. (2014)  Zbl 1320.60023  MR 3291620