Quantitative estimates of the convergence of the empirical covariance matrix in Log-concave Ensembles

Radosław Adamczak[1] Alexander E. Litvak
Alain Pajor Nicole Tomczak-Jaegermann[2]

Abstract

Let K be an isotropic convex body in $\mathbb{R}^n$. Given $\varepsilon > 0$, how many independent points X_i uniformly distributed on K are needed for the empirical covariance matrix to approximate the identity up to ε with overwhelming probability? Our paper answers this question from [12]. More precisely, let $X \in \mathbb{R}^n$ be a centered random vector with a log-concave distribution and with the identity as covariance matrix. An example of such a vector X is a random point in an isotropic convex body. We show that for any $\varepsilon > 0$, there exists $C(\varepsilon) > 0$, such that if $N \sim C(\varepsilon)\, n$ and $(X_i)_{i \leq N}$ are i.i.d. copies of X, then $\left\| \frac{1}{N} \sum_{i=1}^N X_i \otimes X_i - \mathrm{Id} \right\| \leq \varepsilon$, with probability larger than $1 - \exp(-c\sqrt{n})$.

AMS Classification: primary 52A20, 46B09, 52A21 secondary 15A52, 60E15
Keywords: convex bodies, log-concave measures, isotropic measures, random matrices, norm of random matrices, uniform laws of large numbers, approximation of covariance matrices

[1]Work on this paper began when this author held a postdoctoral position at the Department of Mathematical and Statistical Sciences, University of Alberta in Edmonton, Alberta. The position was partially sponsored by the Pacific Institute for the Mathematical Sciences.
[2]This author holds the Canada Research Chair in Geometric Analysis.

1 Introduction

Let $X \in \mathbb{R}^n$ be a centered random vector with covariance matrix Σ and consider N independent random vectors $(X_i)_{i\le N}$ distributed as X. By the law of large numbers, the empirical covariance matrix $\frac{1}{N}\sum_{i=1}^{N} X_i \otimes X_i$ converges to $\mathbb{E}\, X \otimes X = \Sigma$ as $N \to \infty$. Our aim is to give quantitative estimate of the rate of this convergence, that is, to estimate the size N of the sample for which

$$\Big\| \frac{1}{N}\sum_{i=1}^{N} X_i \otimes X_i - \Sigma \Big\| \le \varepsilon \|\Sigma\| \tag{1.1}$$

holds with high probability.

This question was investigated in [12] motivated by a problem of complexity in computing volume in high dimension. In particular the authors proved that

$$\mathbb{E}\, \Big\| \frac{1}{N}\sum_{i=1}^{N} X_i \otimes X_i - \Sigma \Big\| \le C\, \frac{n^2}{N} \|\Sigma\|,$$

where $C = \max_{i\le N} \mathbb{E}|X_i|^4/(\mathbb{E}|X_i|^2)^2$. Chebyshev's inequality yields then a first estimate: for any $\varepsilon > 0$, $\delta \in (0,1)$,

$$\mathbb{P}\left(\Big\| \frac{1}{N}\sum_{i=1}^{N} X_i \otimes X_i - \Sigma \Big\| \le \varepsilon \|\Sigma\| \right) \ge 1 - \delta \tag{1.2}$$

whenever $N \ge \frac{C}{\varepsilon\delta} n^2$.

When random vectors are standard Gaussian, the covariance matrix is the identity and it is known (see the survey [8]) that (1.1) holds with high probability whenever $N \ge 4n/\varepsilon^2$. This raises the question about the order of the best N. In particular can it be proportional to n, under reasonable assumptions? More precisely, the question in [12] was phrased in the following setting.

Let $K \subset \mathbb{R}^n$ be a convex body and let $X \in K$ be a random point uniformly distributed on K. Suppose that X is centered at 0 and that the covariance matrix of X is the identity of $\mathbb{R}^n$. In such a case we shall say that X (or K) is isotropic. Note that any convex body with non empty interior has an affine isotropic image. In this setting and under these assumptions, the question may be stated as follows:

Question: ([12]) *Let K be an isotropic convex body in $\mathbb{R}^n$. Given $\varepsilon > 0$, how many independent points X_i uniformly distributed on K are needed for* *the empirical covariance matrix to approximate the identity up to ε with overwhelming probability?*

Our main aim in this paper is to answer this question. As it is well known to specialists, a good framework for this kind of geometric probabilistic questions is given by log-concave distribution (see below for the definition). This is a stable and well structured class of measures in $\mathbb{R}^n$ that contains uniform measure on convex bodies. Thus our goal is to estimate

$$\mathbb{P}\left(\left\|\frac{1}{N}\sum_{i=1}^{N} X_i \otimes X_i - \Sigma\right\| \leq \varepsilon\|\Sigma\|\right) \tag{1.3}$$

where Σ is the covariance matrix of a centered random vector $X \in \mathbb{R}^n$ with a log-concave distribution and (X_i) are N independent random vectors distributed as X.

Since for a symmetric matrix M, one has $\|M\| = \sup_{y\in S^{n-1}}\langle My, y\rangle$, (1.1) is implied by

$$\left|\frac{1}{N}\sum_{i=1}^{N}(\langle X_i, y\rangle^2 - \mathbb{E}\langle X_i, y\rangle^2)\right| \leq \varepsilon\langle \Sigma y, y\rangle \quad \text{for all} \quad y \in \mathbb{R}^n. \tag{1.4}$$

In the case when the covariance matrix is the identity, it is equivalent to

$$1 - \varepsilon \leq \frac{1}{N}\sum_{i=1}^{N}\langle X_i, y\rangle^2 \leq 1 + \varepsilon \quad \text{for all} \quad y \in S^{n-1}. \tag{1.5}$$

Because of the linear invariance, there is no loss of generality to consider just this case when the covariance matrix is the identity.

In this framework, a breakthrough was achieved in [7] where it was proved that for any $\varepsilon, \delta \in (0,1)$, there exists $C(\varepsilon,\delta) > 0$ such that if a body K is isotropic then $N = C(\varepsilon,\delta)n\log^3 n$ i.i.d. uniformly distributed points on K satisfy (1.2). This estimate was further improved to $N = C(\varepsilon,\delta)n\log^2 n$ in [23] and to $N = C(\varepsilon,\delta)n\log n$ in [9] and [22]; the former paper treated the case when K is invariant under every reflection with respect to coordinate subspaces and the latter proved the estimate in full generality

One should note that in all these results, the probability in (1.2) does not go to 1 as n goes to infinity, as one expects in this type of high dimensional phenomena. This probability, $1-\delta$, is given by a parameter δ and $C(\varepsilon,\delta)$ depends on it. Thus letting δ tend to zero may destroy the estimate on N. To emphasize this important feature we will talk about *overwhelming probability* if the probability goes to 1 as n goes to infinity.

The first result establishing (1.1) with overwhelming probability was given in [18]. When a body K is invariant under every reflection with respect to coordinate subspaces, it is proved in [2] that for any $\varepsilon \in (0,1)$ there exist $C(\varepsilon) > 0$ such that (1.5) holds whenever $N \geq C(\varepsilon)\, n$ and with probability going to 1 as n goes to infinity. Finally, the present paper shows, as a consequence of our main results (Theorems 4.1 and 4.2), that the same is true for an arbitrary body K (in the isotropic position).

An important related direction concerns norms of random matrices with independent log-concave columns (or rows). More precisely, let $X \in \mathbb{R}^n$ be a centered random vector with a log-concave distribution such that the covariance matrix is the identity. Consider N independent random vectors $(X_i)_{i\leq N}$ distributed as X and define $A = A^{(N)}$ to be the $n \times N$ matrix with $(X_i)_{i\leq N}$ as columns. For n, N arbitrary (and N not too large, actually, $n = N$ being the central case) the question is to prove an estimate for the norm $\|A\|$ as an operator $A : \ell_2^N \to \ell_2^n$, valid with overwhelming probability. This problem can be viewed as an "isomorphic form" of an upper estimate in (1.5) (for $n = N$, say), and the papers discussed above provided some answers – with "parasitic" logarithmic factors – to this question as well. The present article gives optimal estimates for $\|A\|$ (in Theorem 3.6 and Corollaries 3.8 and 4.12); for example, for the square matrix if $n = N$, we have $\|A\| \leq C\sqrt{n}$, with overwhelming probability.

To observe a still one more point of view, for arbitrary n and N, consider again $A = A^{(N)}$. The set of $n \times n$ matrices may be equipped with the distribution of AA^* to be a matrix probability space and because of the analogy with Random Matrix Theory, in particular with Wishart Ensemble, let us call it a *Log-concave Ensemble*.

In the last decades, in Asymptotic Geometric Analysis, considerable work and progress have been achieved in understanding the properties of random vectors with log-concave distribution, and more recently, in understanding spectral properties of random matrices with independent rows (or columns) with log-concave distribution. It appears that in high dimension they behave somewhat similarly as if the coordinate would be independent. This leads by analogy with Random Matrix Theory to questions on the spectrum of AA^* similar to those of the Wishart Ensemble. One important difference is that now the entries are dependent but strongly structured by the log-concavity hypothesis.

Denote by $\lambda_1 = \lambda_1(A^{(N)}) \le \cdots \le \lambda_n = \lambda_n(A^{(N)})$ the eigenvalues of AA^* (the squares of the singular values of A). It was proved in [21] that when n/N goes to $\beta \in (0,1)$ as $n, N \to \infty$, then the empirical measures of the eigenvalues have a limit. It is the so-called Marchenko-Pastur distribution, as for the Wishart Ensemble when all entries of the matrix A are i.i.d. It is also known ([4]) in the case when all the entries of A are i.i.d. (with a finite fourth moment) and $\lim_{n\to+\infty} \frac{n}{N} = \beta \in (0,1)$ that $\lim \lambda_1/N = (1-\sqrt{\beta})^2$ and $\lim \lambda_n/N = (1+\sqrt{\beta})^2$. One could conjecture that such results are also valid in the log-concave setting. Nevertheless, these results are asymptotic and not quantitative (given fixed dimension).

Problem (1.5) is of course equivalent to quantitative estimates for $\lambda_1(A^{(N)})$ and $\lambda_n(A^{(N)})$, that is of the support of the spectrum of A. An answer is given by Proposition 4.4 where it is shown that for $n \le N \le \exp(\sqrt{n})$,

$$1 - C\sqrt{\frac{n}{N}}\log\frac{2N}{n} \le \frac{1}{N}\sum_{i=1}^{N}\langle X_i, y\rangle^2 \le 1 + C\sqrt{\frac{n}{N}}\log\frac{2N}{n} \qquad \text{for all } y \in S^{n-1}$$

holds with probability larger than $1 - \exp(-c\sqrt{n})$, where $C, c > 0$ are numerical constants. Thus, putting $\beta = \frac{n}{N} \in (0,1)$, we get

$$1 - C\sqrt{\beta}\log(2/\beta) \le \frac{\lambda_1}{N} \le \frac{\lambda_n}{N} \le 1 + C\sqrt{\beta}\log(2/\beta)$$

with overwhelming probability. As a consequence already mentioned earlier, $\|A\| \le C(\sqrt{N} + \sqrt{n})$ with overwhelming probability, where $C > 0$ is a numerical constant (Corollary 4.12).

Our general method follows an approach that can be traced back to Bourgain [7] (cf. also [10]). It relies upon a crucial new ingredient of a novel chaining argument that in an essential way depends on the distribution of coordinates of a point on the unit sphere. What makes this approach work, by rather subtle estimates, is a special structure of the sets used for the chaining.

To describe a very rough idea of this structure, involved in the proof of Theorem 3.6 below, assume for simplicity that $m = n = 2^s$ and let $a_k = 2^{s-k}$ for $1 \le k \le s$. For each k, first consider the subset of the Euclidean unit ball in $\mathbb{R}^N$ of all vectors that have the support of cardinality less than or equal to a_k and with the ℓ_∞ norm of the coordinates bounded by α_k, and then define $\mathcal{M}^{(k)}$ to be a preassigned ε_k net (in the Euclidean norm) of this set, where $0 < \alpha_k$, $\varepsilon_k < 1$ are judiciously fixed in advance. Using sets $\mathcal{M}^{(k)}$ in successive steps of chaining we arrive to the set $\mathcal{M}$ that consists of sums $v = \sum_k v_k$ where v_k's are mutually disjointly supported vectors from $\mathcal{M}^{(k)}$ (assuming that the Euclidean norm of v is less than 2). As can be expected the actual definition of $\mathcal{M}$ contains a number of delicate points which were omitted here and can be found at the beginning of the proof of Theorem 3.6. However it is given in just one step without discussing each individual step of the chaining.

The paper is organized as follows. In the next Section 2 we present some definitions and preliminary tools. In Section 3 we study the norm of a restriction of the matrix $A = A^{(N)}$ defined by

$$A_m = \sup_{\substack{F\subset\{1,\dots,N\}\\ |F|\le m}} \|A_{|\mathbb{R}^F}\| = \sup_{\substack{z\in S^{N-1}\\ |\operatorname{supp} z|\le m}} |Az|.$$

We show in Theorem 3.6 that with overwhelming probability,

$$A_m \le C\left(\sqrt{n} + \sqrt{m}\log\frac{2N}{m}\right).$$

In Section 4.1 we prove the result announced in the abstract, answering a question from [12]. This theorem appears as a particular case of a more general study of

$$\sup_{y\in S^{n-1}}\left|\frac{1}{N}\sum_{i=1}^{N}(\langle X_i, y\rangle^p - \mathbb{E}\langle X_i, y\rangle^p)\right|$$

defined for any $p \ge 1$. Such processes have been studied in [10], [11] and [17].

Section 4.2 describes several observations for norms of random matrices from ℓ_2 to ℓ_p, $p \ne 2$. In the final Section 4.3 we sketch a more elementary proof of the main result of Section 4.1, when $p = 2$.

2 Notation and preliminaries

We equip $\mathbb{R}^n$ and $\mathbb{R}^N$ with the natural scalar product $\langle\,\cdot, \cdot\rangle$ and the natural Euclidean norm $|\cdot|$. We also denote by the same notation $|\cdot|$ the cardinality of a set. In this paper, X will denote a random vector in $\mathbb{R}^n$ and (X_i) will be independent random vectors with the same distribution as X. By Id we shall denote the identity on $\mathbb{R}^n$ and by $\Sigma = \Sigma(X) = \mathbb{E}\, X \otimes X$, the covariance matrix of X (here $X \otimes X$ is the rank one operator defined by $X \otimes X(y) = \langle X, y \rangle X$, for all $y \in \mathbb{R}^n$). By $\|M\|$ we shall denote the operator norm of a matrix M, that is $\|M\| = \sup_{|y|=1} |My|$.

Definition 2.1. *A random vector $X \in \mathbb{R}^n$ is called isotropic if*

$$\mathbb{E}\langle X, y \rangle = 0, \quad \mathbb{E}\, |\langle X, y \rangle|^2 = |y|^2 \quad \text{for all } y \in \mathbb{R}^n, \tag{2.1}$$

in other words, if X is centered and its covariance matrix is the identity:

$$\mathbb{E}\, X \otimes X = \mathrm{Id}\,.$$

Recall that a function $f : \mathbb{R}^n \to \mathbb{R}$ is called log-concave if for any $\theta \in [0, 1]$ and any $x_1, x_2 \in \mathbb{R}^n$,

$$f\big(\theta x_1 + (1-\theta)x_2\big) \geq f(x_1)^\theta f(x_2)^{1-\theta}.$$

Definition 2.2. *A measure μ on $\mathbb{R}^n$ is log-concave if for any measurable subsets A, B of $\mathbb{R}^n$ and any $\theta \in [0, 1]$,*

$$\mu(\theta A + (1-\theta)B) \geq \mu(A)^\theta \mu(B)^{(1-\theta)}$$

whenever the set

$$\theta A + (1-\theta)B = \{\theta x_1 + (1-\theta)x_2 \,:\, x_1 \in A,\ x_2 \in B\}$$

is measurable.

The Brunn-Minkowski inequality provides examples of log-concave measures, that are the uniform Lebesgue measure on compact convex subsets of $\mathbb{R}^n$ as well as their marginals (cf. e.g., [24]). More generally, Borell's theorem [5] characterizes the log-concave measures that are not supported by any hyperplane as the absolutely continuous measures (with respect to the Lebesgue measure) with a log-concave density. Note that the distribution of an isotropic vector is not supported by any hyperplane. Moreover, it is known [6] that if a measure is log-concave then linear functionals exhibit a sub-exponential decay. To be more precise, recall that for a random variable Y, the ψ_1 norm of Y is

$$\|Y\|_{\psi_1} = \inf \left\{ C > 0\,;\ \mathbb{E} \exp \left(\frac{|Y|}{C} \right) \leq 2 \right\}.$$

A straightforward computation shows that for every integer $p \geq 1$,

$$(\mathbb{E}|Y|^p)^{1/p} \leq cp\|Y\|_{\psi_1} \tag{2.2}$$

where c is an absolute constant.

We can now state the sub-exponential decay of linear functionals in terms of ψ_1 norm [6]:

Lemma 2.3. *Let $X \in \mathbb{R}^n$ be a centered random vector with a log-concave distribution. Then for every $y \in S^{n-1}$,*

$$\| \langle X, y \rangle \|_{\psi_1} \leq \psi \, (\mathbb{E}|\langle X, y\rangle|^2)^{1/2}$$

where $\psi > 0$ is universal constant. Moreover, if X has a symmetric distribution then $\psi = 2$.

The moreover part easily follows by a direct calculation (see [20]).

Putting together (2.2) and Lemma 2.3, we get that for every $y \in S^{n-1}$,

$$(\mathbb{E}| \langle X, y \rangle |^p)^{1/p} \leq Cp \, (\mathbb{E}|\langle X, y\rangle|^2)^{1/2} \tag{2.3}$$

where C is an absolute positive constant.

3 Norm of a random matrix

In this Section $X_1, \dots, X_N$ are independent random vectors in $\mathbb{R}^n$. Mostly we work with i.i.d. random vectors, distributed according to an isotropic, log-concave probability measure on $\mathbb{R}^n$. Random $n \times N$ matrix whose columns are X_i's is denoted by A and its operator norm from ℓ_2^N to ℓ_2^n is denoted by $\|A\|$. We will also use the following related notation, for $1 \leq m \leq N$,

$$A_m = \sup_{\substack{F \subset \{1,\dots,N\} \\ |F| \leq m}} \|A_{|\mathbb{R}^F}\| = \sup_{\substack{z \in S^{N-1} \\ |\operatorname{supp} z| \leq m}} |Az|.$$

Note that A_m is increasing in m. Given a set $E \subset \{1, ..., N\}$ by P_E we denote the orthogonal projection from $\mathbb{R}^N$ onto coordinate subspace of vectors whose support is in E. Such a subspace is denoted by $\mathbb{R}^E$.

Lemma 3.1. *Let $X_1, \ldots, X_N$ be i.i.d. random vectors, distributed according to an isotropic, log-concave probability measure on $\mathbb{R}^n$. There exists an absolute positive constant C_0 such that for any $N \leq \exp(\sqrt{n})$ and for every $K \geq 1$ one has*

$$\max_{i \leq N} |X_i| \leq C_0 K \sqrt{n}$$

with probability at least $1 - \exp(-K\sqrt{n})$.

Proof By [22] we have for every $i \leq N$

$$\mathbb{P}\left\{ |X_i| \geq Ct\sqrt{n} \right\} \leq \exp(-tc\sqrt{n}),$$

where C and c are absolute positive constants. The result follows by the union bound (and adjusting absolute constants). □

Lemma 3.2. *Let $x_1, \ldots, x_N \in \mathbb{R}^n$. There exists a set $E \subset \{1, ..., N\}$, such that*

$$\sum_{i \neq j} \langle x_i, x_j \rangle \leq 4 \sum_{i \in E} \sum_{j \in E^c} \langle x_i, x_j \rangle .$$

Proof Clearly one has

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

from which the lemma follows. □

Now, given a $E \subset \{1, ..., N\}$, $\varepsilon, \alpha \in (0, 1]$, by $\mathcal{N}(E, \varepsilon, \alpha)$ we denote an ε-net of $B_2^N \cap \alpha B_\infty^N \cap \mathbb{R}^E$ in the Euclidean metric. Standard volume estimate shows that we may assume that the cardinality of $\mathcal{N}(E, \varepsilon, \alpha)$ does not exceed $(3/\varepsilon)^m$, where m is the cardinality of E.

We will need the following two lemmas.

Lemma 3.3. *Let $X_1, \ldots, X_N$ be independent random vectors in $\mathbb{R}^n$ and let $\psi > 0$ such that*

$$\sup_{i \leq N} \sup_{y \in S^{n-1}} \| \langle X_i, y \rangle \|_{\psi_1} \leq \psi.$$

Let $m \leq N$, $\varepsilon, \alpha \in (0, 1]$ and $L \geq 2m \log \frac{12eN}{m\varepsilon}$. Then

$$\mathbb{P}\left(\sup_{\substack{F \subset \{1,...,N\} \\ |F| \leq m}} \sup_{E \subset F} \sup_{z \in \mathcal{N}(F, \varepsilon, \alpha)} \sum_{i \in E} \left| \left\langle z_i X_i, \sum_{j \in F \setminus E} z_j X_j \right\rangle \right| > \psi \, \alpha L A_m \right) \leq e^{-L/2}.$$

Proof Denote the underlying probability space by Ω. For $F \subset \{1, ..., N\}$ with $|F| \leq m$, $E \subset F$, and $z \in \mathcal{N}(F, \varepsilon, \alpha)$, define the subset $\Omega(F, E, z)$ of Ω by

$$\Omega(F, E, z) = \left\{ \sum_{i \in E} \left| \left\langle z_i X_i, \sum_{j \in F \setminus E} z_j X_j \right\rangle \right| > \psi \alpha L A_m \right\}.$$

Fix F, E and z as above and set $y = \sum_{j \in F \setminus E} z_j X_j$. Clearly, y is independent of vectors X_i's, $i \in E$, and $|y| \leq A_m$. Note that $|y| > 0$ on $\Omega(F, E, z)$ (otherwise $\langle z_i X_i, y \rangle = 0$ for all $i \in E$ and the sharp inequality defining $\Omega(F, E, z)$ would be violated). Thus, using the fact that $\|z\|_\infty \leq \alpha$, we obtain

$$\sum_{i \in E} \left| \left\langle z_i X_i, \sum_{j \in F \setminus E} z_j X_j \right\rangle \right| \leq \alpha A_m \sum_{i \in E} |\langle X_i, y/|y| \rangle| ,$$

on $\Omega(F, E, z)$. Since $A_m > 0$ on $\Omega(F, E, z)$, this implies

$$\mathbb{P}\left(\Omega(F, E, z)\right) \leq \mathbb{P}\left(\sum_{i \in E} |\langle X_i, y/|y| \rangle| > \psi L \right).$$

On the other hand, by Chebyshev's inequality and the assumption on the ψ_1-norms of linear functionals, the latter probability is less than

$$e^{-L}\, \mathbb{E} \exp\left(\sum_{i \in E} \frac{|\langle X_i, y/|y| \rangle|}{\psi} \right) \leq 2^{|E|}\, e^{-L} \leq 2^m\, e^{-L}.$$

Therefore by the union bound,

$$\begin{aligned}
&\mathbb{P}\left(\sup_{\substack{F \subset \{1,\ldots,N\} \\ |F| \leq m}} \sup_{E \subset F} \sup_{z \in \mathcal{N}(F, \varepsilon, \alpha)} \sum_{i \in E} \left| \left\langle z_i X_i, \sum_{j \in F \setminus E} z_j X_j \right\rangle \right| > \psi\, \alpha L A_m \right) \\
&\leq \sum_{k=1}^{m} \binom{N}{k}\ 2^m \left(\frac{3}{\varepsilon} \right)^m \sup_{F,E,z} \mathbb{P}\left(\Omega(F, E, z)\right) \\
&\leq \sum_{k=1}^{m} \binom{N}{k}\ 2^m \left(\frac{3}{\varepsilon} \right)^m\ 2^m\, e^{-L} \leq \left(\frac{eN}{m} \right)^m \left(\frac{12}{\varepsilon} \right)^m\ e^{-L} \\
&= \exp\left(m \log \frac{12eN}{m\varepsilon} - L \right),
\end{aligned}$$

which implies the result. $\square$

We will also need another lemma of a similar type. We provide the proof for sake of completeness.

Lemma 3.4. *Let $X_1, \dots, X_N$ be independent random vectors in $\mathbb{R}^n$ and let $\psi > 0$ such that*

$$\sup_{i \le N} \sup_{y \in S^{n-1}} \| \langle X_i, y \rangle \|_{\psi_1} \le \psi.$$

Let $1 \le k, m \le N$, $\varepsilon, \alpha \in (0, 1]$, $\beta > 0$, and $L > 0$. Let $B(m, \beta)$ denote the set of vectors $x \in \beta B_2^N$ with $|\operatorname{supp} x| \le m$ and let $\mathcal{B}$ be a subset of $B(m, \beta)$ of cardinality M. Then

$$\begin{aligned} \mathbb{P}\left(\sup_{\substack{F \subset \{1,\dots,N\} \\ |F| \le k}} \sup_{x \in \mathcal{B}} \sup_{z \in \mathcal{N}(F, \varepsilon, \alpha)} \sum_{i \in F} \left| \left\langle z_i X_i, \sum_{j \notin F} x_j X_j \right\rangle \right| > \psi \alpha \beta L A_m \right) & \\ \le \ M \left(\frac{6eN}{k\varepsilon} \right)^k e^{-L}. & \end{aligned}$$

Proof The proof is analogous to the argument in Lemma 3.3. For $F \subset \{1, ..., N\}$ with $|F| \le k$, $x \in \mathcal{B}$, and $z \in \mathcal{N}(F, \varepsilon, \alpha)$ consider

$$\Omega(F, x, z) = \left\{ \sum_{i \in F} \left| \left\langle z_i X_i, \sum_{j \notin F} x_j X_j \right\rangle \right| > \psi \alpha \beta L A_m \right\}.$$

Fix F, x, z as above and set $y = \sum_{j \notin F} x_j X_j$. Clearly, y is independent of the vectors X_i's, $i \in F$, moreover, $|y| \le \beta A_m$, and, similarly as in before, $|y| > 0$ on $\Omega(F, x, z)$. Thus, using the fact that $\|z\|_\infty \le \alpha$, we obtain

$$\sum_{i \in F} \left| \left\langle z_i X_i, \sum_{j \notin F} x_j X_j \right\rangle \right| \le \alpha \beta A_m \sum_{i \in F} |\langle X_i, y/|y| \rangle| ,$$

on $\Omega(F, x, z)$. Therefore, again as in Lemma 3.3, we have

$$\begin{aligned} \mathbb{P}\left(\Omega(F, x, z)\right) \ &\le \ \mathbb{P}\left(\sum_{i \in F} |\langle X_i, y/|y| \rangle| > \psi L \right) \\ &\le \ e^{-L} \, \mathbb{E} \exp\left(\sum_{i \in F} \frac{|\langle X_i, y/|y| \rangle|}{\psi} \right) \le 2^{|F|} \, e^{-L} \le 2^k \, e^{-L}. \end{aligned}$$

By the union bound we get

$$\mathbb{P}\left(\sup_{\substack{F\subset\{1,\dots,N\}\\|F|\le k}}\ \sup_{x\in\mathcal{B}}\ \sup_{z\in\mathcal{N}(F,\varepsilon,\alpha)}\sum_{i\in F}\left|\left\langle z_iX_i,\sum_{j\notin F}x_jX_j\right\rangle\right|>\psi\alpha\beta LA_m\right)$$
$$\le\ M\sum_{l=1}^{k}\binom{N}{l}\left(\frac{3}{\varepsilon}\right)^k 2^k\,e^{-L}\le M\left(\frac{eN}{k}\right)^k\left(\frac{6}{\varepsilon}\right)^k e^{-L},$$

which proves the result. □

Remark 3.5. Observe that if X_i's are i.i.d. random vectors, distributed according to an isotropic, log-concave probability measure on $\mathbb{R}^n$, then, by Lemma 2.3, they satisfy the condition for the ψ_1-norm of Lemmas 3.3 and 3.4.

Theorem 3.6. *Let $n\ge 1$ and $1\le N\le e^{\sqrt{n}}$ be integers. Let $X_1,\dots,X_N$ are i.i.d. random vectors, distributed according to an isotropic, log-concave probability measure on $\mathbb{R}^n$. Let $K\ge 1$. Then there are absolute positive constants C and c such that*

$$\mathbb{P}\left(\exists m\le N\,:\,A_m\ge CK\left(\sqrt{n}+\sqrt{m}\log\frac{2N}{m}\right)\right)\le\exp\left(-cK\sqrt{n}\right).$$

Remark 3.7. Let $X\in\mathbb{R}^n$ be a random vector with an isotropic exponential distribution, that is with the density defined for $x=(x_i)\in\mathbb{R}^n$ by $\prod_1^n\frac{1}{\sqrt{2}}\exp(-\sqrt{2}|x_i|)$. It is clearly an isotropic vector with a log-concave distribution. Consider now the matrix $A^{(N)}$ build as before from a sample of X of size N. Since

$$\mathbb{P}(|X|\ge t\sqrt{n})\ge\int_{|s|\ge t\sqrt{n}}\frac{1}{\sqrt{2}}\exp(-\sqrt{2}|s|)\,ds=\exp(-\sqrt{2}t\sqrt{n})$$

we get that for any $1\le m\le N$,

$$\mathbb{P}(A_m\ge t\sqrt{n})\ge\exp(-\sqrt{2}t\sqrt{n}).$$

This shows that the probability estimate in Theorem 3.6 is optimal up to numerical constants. The analysis of this example shows that up to numerical constants the logarithmic term in the estimate of A_m in Theorem 3.6 is also optimal (for the details see [1]).

Letting $m = N$ we get a clearly optimal estimate for the operator norm $\|A\|$, valid with overwhelming probability.

Corollary 3.8. *In the setting of Theorem 3.6 we get, for every $K \geq 1$,*

$$\|A\| \leq CK\left(\sqrt{n} + \sqrt{N}\right), \tag{3.1}$$

with probability at least $1 - e^{-cK\sqrt{n}}$, where $C, c > 0$ are absolute constants.

Remark 3.9. The final remark of [7] states that by refining a bit the method of proof of Lemma 2 of that paper one may obtain that if $X_1, \ldots, X_n$ are n independent vectors in $\mathbb{R}^n$ distributed according to a probability measure μ on $\mathbb{R}^n$ satisfying $\|\langle x, y\rangle\|_{\psi_1} < 1/\sqrt{n}$ for all $y \in S^{n-1}$, then, with probability $1 - \delta$, the matrix A admits the bound for the operator norm

$$\|A\| \leq C(\delta)\left(\int \left(\max_{1\leq i\leq n} |X_i|\right) d\mu + 1\right).$$

By Lemmas 2.3 and 3.1, and taking into account the normalization, this would imply a version of (3.1) with $N = n$ and probability $1 - \delta$.

Remark 3.10. Note that $\sqrt{n} + \sqrt{m}\log\frac{2N}{m}$ in the formula in Theorem 3.6 can be substituted with

$$\sqrt{n} + \sqrt{m}\log\frac{2N}{\max\{n, m\}}.$$

Indeed, if $m \geq n$ there is nothing to prove, otherwise

$$\sqrt{n} + \sqrt{m}\log\frac{2N}{m} = \sqrt{n} + \sqrt{m}\log\frac{n}{m} + \sqrt{m}\log\frac{2N}{n} \leq 2\sqrt{n} + \sqrt{m}\log\frac{2N}{n}.$$

Finally, another immediate consequence.

Corollary 3.11. *There are absolute positive constants C and c such that for every $n \geq 1$, $1 \leq N \leq e^{\sqrt{n}}$, $K \geq 1$, and X_i's as in Theorem 3.6 one has*

$$\mathbb{P}\left(\exists_{E\subset\{1,\ldots,N\}}\left|\sum_{i\in E} X_i\right| \geq CK\left(\sqrt{n|E|} + |E|\log\frac{2N}{n}\right)\right) \leq \exp\left(-cK\sqrt{n}\right).$$

Proof Given E set $m = |E|$. Consider vector $z \in S^{N-1}$ defined by $z_i = 1/\sqrt{m}$ if $i \in E$ and $z_i = 0$ otherwise. We have

$$\left| \sum_{i \in E} X_i \right| = \sqrt{m}|Az| \leq \sqrt{m} A_m.$$

Therefore Theorem 3.6 and Remark 3.7 imply the result. $\square$

Proof of Theorem 3.6. As $N \leq e^{\sqrt{n}}$, it is easy to see, by applying the union bound and adjusting absolute constants, that it is sufficient to prove that for K sufficiently large and every fixed $m \leq N$, one has

$$\mathbb{P}\left(A_m \geq CK \left(\sqrt{n} + \sqrt{m} \log \frac{2N}{m} \right) \right) \leq \exp\left(-cK\sqrt{n}\right).$$

We shall define a set $\mathcal{M}$ of vectors with a special structure and supports less than or equal to m which serves simultaneously two purposes: we will be able to estimate with large probability $\sup_{x \in \mathcal{M}} |Ax|$, and we will use $\mathcal{M}$ to approximate an arbitrary vector from B_2^N of support less than or equal to m. Then a standard argument will lead to the required estimate for A_m.

First observe that if for a vector $x \in S^{N-1}$ there is a simultaneous control of the size of support and its ℓ_∞-norm (more precisely, $|\operatorname{supp} x| \sim s$ and $\|x\|_\infty \leq s^{-1/2}$, for some $s \geq 1$) then $|Ax|$ can be estimated, with large probability, directly by using Lemmas 3.2 and 3.3 (it is also a part of the estimates below). It is therefore natural to expect vectors from $\mathcal{M}$ to be sums of (disjointly supported) vectors admitting such a simultaneous control as above. Formally, the definition of $\mathcal{M}$ splits into two cases. If

$$m \ \log \frac{48eN}{m} \leq \sqrt{n}, \tag{3.2}$$

we set

$$\mathcal{M} = \bigcup_{\substack{E \subset \{1, \dots N\} \\ |E| = m}} \mathcal{N}(E, 1/4, 1).$$

Otherwise, let l be the smallest integer such that

$$\frac{m}{2^l} \log \frac{48e2^l N}{m} \leq \sqrt{n}, \tag{3.3}$$

and fix positive integers $a_0, a_1, \dots, a_l$ such that $a_k \le m\, 2^{-k+1}$ for $1 \le k \le l$ and $a_0 \le m\, 2^{-l}$, and $\sum_{k=0}^{l} a_k = m$. (We shall later set $a_k := [m\, 2^{-k+1}] - [m\, 2^{-k}]$ for $1 \le k \le l$ and $a_0 := [m\, 2^{-l}]$.)

Then set $\mathcal{M} = \mathcal{M}_0 \cap 2B_2^N$, where $\mathcal{M}_0$ consists of all vectors of the form $x = \sum_{k=0}^{l} x_k$, where x_i's have disjoint supports and

$$x_0 \in \bigcup_{\substack{E \subset \{1,\dots N\} \\ |E| \le a_0}} \mathcal{N}(E, 1/4, 1), \;\; x_k \in \bigcup_{\substack{E \subset \{1,\dots N\} \\ |E| \le a_k}} \mathcal{N}\left(E, 2^{-k}, \sqrt{\frac{2^k}{m}}\right) \;\text{ for } 1 \le k \le l.$$

Note that for every vector $x \in \mathcal{M}$ we have $|\operatorname{supp} x| \le \sum_0^l a_k = m$ and $|x| \le 2$.

We shall consider the details of the case $m \log(48eN/m) > \sqrt{n}$ (the other case, when (3.2) holds, can be treated similarly, actually, it is even simpler, since the construction of $\mathcal{M}$ is simpler). Fix $x \in \mathcal{M}$ of the form $x = \sum_{k=0}^{l} x_k$ and let F_k be the support of x_k (if there are more than one such representations, we fix one of them). Denote the coordinates of x by $x(i)$, $i \le N$, then

$$\begin{aligned} |Ax|^2 &= \left\langle \sum_{i \le N} x(i) X_i, \sum_{i \le N} x(i) X_i \right\rangle = \sum_{i \le N} x(i)^2 |X_i|^2 + \sum_{i \ne j} \langle x(i) X_i, x(j) X_j \rangle \\ &\le 2 \max_i |X_i|^2 + D_x \le 2 \max\{2 \max_i |X_i|^2, D_x\}, \end{aligned} \tag{3.4}$$

where

$$D_x = \sum_{i \ne j} \langle x(i) X_i, x(j) X_j \rangle .$$

Note that by Lemma 3.1, $\max_i |X_i| \le C_0 K \sqrt{n}$ with probability larger than $1 - e^{-K\sqrt{n}}$, and we would like to get a similar estimate for D_x.

To this aim we split D_x according to the structure of x. Namely we let

$$D'_x := \sum_{k=0}^{l} \sum_{\substack{i,j \in F_k \\ i \ne j}} \langle x(i) X_i, x(j) X_j \rangle ,$$

and

$$\begin{aligned} D''_x : \ &= \sum_{k=0}^{l} \sum_{\substack{i \in F_k \\ j \notin F_k}} \langle x(i)X_i, x(j)X_j \rangle \\ &= 2 \sum_{k=1}^{l} \sum_{i \in F_k} \sum_{r \in G_k} \left\langle x(i)X_i, \sum_{j \in F_r} x(j)X_j \right\rangle, \end{aligned}$$

where $G_k = \{0, k+1, k+2, \dots, l\}$. Note that

$$D_x = D'_x + D''_x.$$

We first estimate D'_x. By Lemma 3.2 we obtain that for every k there exists a subset $\bar{F}_k$ of F_k such that

$$\begin{aligned} D'_x \ &\leq \ 4 \sum_{k=0}^{l} \sum_{\substack{i \in \bar{F}_k \\ j \in F_k \setminus \bar{F}_k}} \langle x(i)X_i, x(j)X_j \rangle \\ &\leq \ 4 \sup_{\substack{F \subset \{1,\dots,N\} \\ |F| \leq m/2^l}} \ \sup_{E \subset F} \ \sup_{v \in \mathcal{N}(F,1/4,1)} \sum_{i \in E} \left| \left\langle v_i X_i, \sum_{j \in F \setminus E} v_j X_j \right\rangle \right| \\ &+ \ 4 \sum_{k=1}^{l} \sup_{\substack{F \subset \{1,\dots,N\} \\ |F| \leq 2m/2^k}} \ \sup_{E \subset F} \ \sup_{v \in \mathcal{N}(F, 2^{-k}, \sqrt{2^k/m})} \sum_{i \in E} \left| \left\langle v_i X_i, \sum_{j \in F \setminus E} v_j X_j \right\rangle \right|. \end{aligned}$$

We now apply Lemma 3.3 to each summand in the sum above with $L = 2K\sqrt{n}$, $\varepsilon = 1/4$, $\alpha = 1$ for the first summand (note that such an L satisfies the condition) and with $L = \frac{4m}{2^k} K \log \frac{12eN4^k}{m}$, $\varepsilon = 2^{-k}$, $\alpha = \sqrt{\frac{2^k}{m}}$ for $k \geq 1$. By the union bound we obtain

$$\begin{aligned} \mathbb{P} \left(\sup_{x \in \mathcal{M}} D'_x \right. \ &> \ \left. 8\psi K A_m \sqrt{n} + 2\psi K A_m \sum_{k=1}^{l} \sqrt{\frac{2^k}{m} \frac{8m}{2^k} \log \frac{12eN4^k}{m}} \right) \\ &\leq \ \exp\left(-K\sqrt{n}\right) + \sum_{k=1}^{l} \exp\left(-K \frac{2m}{2^k} \log \frac{12eN4^k}{m}\right) \\ &\leq \ \exp\left(-K\sqrt{n}\right) + l \exp\left(-K \frac{2m}{2^l} \log \frac{12eN4^l}{m}\right), \end{aligned}$$

where ψ is the absolute constant from Lemma 2.3.

Therefore, the choice of l implies the following bound, with some absolute positive constant C,

$$\begin{aligned}\mathbb{P}\left(\sup_{x\in\mathcal{M}} D'_x\right. &> \left. A_m K\left(8\psi\sqrt{n}+C\psi\sqrt{m}\log\frac{2N}{m}\right)\right)\\ &\le \exp\left(-K\sqrt{n}\right)+l\exp\left(-K\sqrt{n}\right)\le(2\sqrt{n}+1)\exp\left(-K\sqrt{n}\right).\end{aligned}$$

(We also used the estimate $l\le 2\sqrt{n}$, valid when $m\le N\le e^{\sqrt{n}}$.)

The estimate for D''_x essentially follows the same lines. In a sense it is simpler, since we don't need to apply Lemma 3.2. For every $1\le k\le l$ we consider $\mathcal{M}_k=\mathcal{M}'_k\cap 2B_2^N$, where $\mathcal{M}'_k$ consists of all vectors of the form $x=x_0+\sum_{s=k+1}^{l}x_s$, where x_i's ($i=0,k=1,\dots,l$) have pairwise disjoint supports and

$$x_0\in\bigcup_{\substack{E\subset\{1,\dots N\}\\|E|\le a_0}}\mathcal{N}(E,1/4,1),\ x_s\in\bigcup_{\substack{E\subset\{1,\dots N\}\\|E|\le a_s}}\mathcal{N}\left(E,2^{-s},\sqrt{\frac{2^s}{m}}\right)\ \text{ for } s\ge k+1.$$

Then $\mathcal{M}_k\subset 2B_2^N$ and

$$\begin{aligned}|\mathcal{M}_k| &\le 12^{a_0}\prod_{s=k+1}^{l}(3\cdot 2^s)^{a_s}\binom{N}{a_s}\le 12^{a_0}\prod_{s=k+1}^{l}\left(\frac{3\cdot 2^s eN}{a_s}\right)^{a_s}\\ &\le \exp\left(\frac{m}{2^l}\log 12+\sum_{s=k+1}^{l}\frac{2m}{2^s}\log\frac{3e4^sN}{2m}\right)\le\exp\left(\sum_{s=k+1}^{l+1}\frac{2m}{2^s}\log\frac{3e4^sN}{2m}\right)\\ &\le \exp\left(\frac{m}{2^k}\left(\log\frac{6e4^kN}{m}\sum_{s=0}^{l-k}\frac{1}{2^s}+\log 4\sum_{s=1}^{l-k}\frac{s}{2^s}\right)\right)\le\exp\left(\frac{4m}{2^k}\log\frac{6e4^kN}{m}\right).\end{aligned}$$

We also observe that

$$\begin{aligned}D''_x &= 2\sum_{k=1}^{l}\sum_{i\in F_k}\left\langle x(i)X_i,\sum_{r\in G_k}\sum_{j\in F_r}x(j)X_j\right\rangle\\ &\le 2\sum_{k=1}^{l}\sup_{\substack{F\subset\{1,\dots,N\}\\|F|\le 2m/2^k}}\ \sup_{u\in\mathcal{N}(F,2^{-k},\sqrt{2^k/m})}\ \sup_{v\in\mathcal{M}_k}\sum_{i\in F}\left|\left\langle u_iX_i,\sum_{j\notin F}v_jX_j\right\rangle\right|.\end{aligned}$$

Now we apply Lemma 3.4 to each summand with

$$L = L(k) = \frac{12m}{2^k} K \log \frac{12e4^k N}{m},$$

$$\varepsilon = \varepsilon_k = 2^{-k}, \ \alpha = \alpha_k = \sqrt{2^k/m}, \ \beta = 2, \ \mathcal{B} = \mathcal{B}_k = \mathcal{M}_k.$$

Using the union bound we obtain

$$\begin{aligned}
&\mathbb{P}\left(D''_x > 48\psi A_m K \sum_{k=1}^{l} \sqrt{\frac{2^k}{m}\,\frac{m}{2^k} \log \frac{12e4^k N}{m}} \right) \\
&\le \sum_{k=1}^{l} \exp\left(\frac{4m}{2^k} \log \frac{12e4^k N}{m} + \frac{2m}{2^k} \log \frac{3e4^k N}{m} - K\frac{12m}{2^k} \log \frac{12e4^k N}{m} \right) \\
&\le \sum_{k=1}^{l} \exp\left(-K\frac{6m}{2^k} \log \frac{12e4^k N}{m} \right) \le l \exp\left(-K\frac{6m}{2^l} \log \frac{12e4^l N}{m} \right).
\end{aligned}$$

As in the case for D'_x it follows that

$$\mathbb{P}\left(\sup_{x\in\mathcal{M}} D''_x > 3C\psi A_m K \sqrt{m} \log \frac{2N}{m} \right) \le 2\sqrt{n} \exp\left(-K\sqrt{n}\right),$$

where C is the same absolute constant as above. Since $D_x = D'_x + D''_x$, then

$$\mathbb{P}\left(\sup_{x\in\mathcal{M}} D_x > KA_m \left(8\psi\sqrt{n} + 4C\psi\sqrt{m} \log \frac{2N}{m} \right) \right) \le (4\sqrt{n}+1)e^{-K\sqrt{n}}. \quad (3.5)$$

Passing now to the approximation argument, pick an arbitrary $z \in S^{N-1}$ with $|\operatorname{supp} z| \le m$. Define the following subsets of $\{1, \dots, N\}$ depending on z. Denote the coordinates of z by z_i ($i = 1, \dots, N$). Let $n_1, \dots, n_N$ be such that $|z_{n_1}| \ge |z_{n_2}| \ge \dots \ge |z_{n_N}|$, so that $z_{n_i} = 0$ for $i > m$ (since $|\operatorname{supp} z| \le m$). If condition (3.2) holds we denote the support of z by E_0 and consider only this E_0. Otherwise we set

$$E_0 = \{n_i\}_{1\le i\le m/2^l}$$

and

$$E_1 = \{n_i\}_{m/2<i\le m}, \ E_2 = \{n_i\}_{m/4<i\le m/2}, \ \dots, \ E_l = \{n_i\}_{m/2^l<i\le m/2^{l-1}},$$

where l is the smallest integer satisfying (3.3) (as before). (For small values of n it can happen that E_0 is empty, but it does not create any difficulty in the proof below.) Clearly, we have

$$a_0 := |E_0| \le m/2^l, \quad a_k := |E_k| \le m/2^k + 1 \le m/2^{k-1} \ \text{ for every } 1 \le k \le l,$$

and $\sum_{i=0}^{l} a_i = m$. Note that the numbers a_k's do not depend on z, although the sets E_k's do. Finally, since $z \in S^{N-1}$, we also observe that for every $k \ge 1$,

$$\|P_{E_k} z\|_\infty \le |z_{n_s}| \le \sqrt{\frac{2^k}{m}},$$

where $s = [m/2^k]$.

Note that for every $k \ge 1$ the vector $P_{E_k} z$ can be approximated by a vector from $\mathcal{N}\left(E_k, 2^{-k}, \sqrt{\frac{2^k}{m}}\right)$ and the vector $P_{E_0} z$ can be approximated by a vector from $\mathcal{N}(E_0, 1/4, 1)$. Thus there exists $x \in \mathcal{M}$, with a suitable representation $x = \sum_{k=0}^{l} x_k$, such that

$$|z - x|^2 \le \sum_{k=0}^{l} |P_{E_k} z - x_k|^2 \le 2^{-4} + \sum_{k=1}^{l} 2^{-2k} < 0.4.$$

Moreover, x is chosen to have the same support as z, and thus $w = z - x$ has the support $|\operatorname{supp} w| \le m$.

Considering all $z \in S^{N-1}$ with $|\operatorname{supp} z| \le m$ it follows that

$$A_m = \sup_{\substack{z \in S^{N-1} \\ |\operatorname{supp} z| \le m}} |Az| \le \sup_{x \in \mathcal{M}} |Ax| + \sqrt{0.4} \sup_{\substack{w \in S^{N-1} \\ |\operatorname{supp} w| \le m}} |Aw| = \sup_{x \in \mathcal{M}} |Ax| + \sqrt{0.4} A_m,$$

which implies

$$A_m \le 3 \sup_{x \in \mathcal{M}} |Ax|.$$

Recall that by (3.4) for every $x \in \mathcal{M}$ we have

$$|Ax|^2 \le 2 \max\{2 \max_i |X_i|^2, D_x\},$$

so passing to the supremum

$$A_m^2 \le 9 \sup_{x \in \mathcal{M}} |Ax|^2 \le 9 \max\{4 \max_i |X_i|^2, 2 \sup_{x \in \mathcal{M}} D_x\}. \tag{3.6}$$

Applying Lemma 3.1 and (3.5) we get

$$A_m \le K\,(6\,C_0 + 144\psi)\,\sqrt{n} + 72\,C\psi K\sqrt{m}\log\frac{2N}{m}$$

with probability larger than

$$1 - (4\sqrt{n}+2)\exp\left(-K\sqrt{n}\right) \ge 1 - \exp\left(-cK\sqrt{n}\right),$$

where c is an absolute positive constant. (In fact this estimate for probability requires that n is sufficiently large, but, as $K \ge 1$ was arbitrary, we can adjust the constants.) This concludes the proof. □

Remark 3.12. Consider now a more general situation in which $X_1, X_2, \dots X_N$ – the columns of the matrix A – are still i.i.d. centered and log-concave, but not necessarily isotropic. Then there exists an $n \times n$ matrix T, such that $(X_i)_{i=1}^N$ has the same distribution as $(TY_i)_{i=1}^N$, where $Y_1, \dots, Y_N$ are isotropic log-concave random vectors in $\mathbb{R}^n$. For the purpose of computing probabilities we may assume that $X_i = TY_i$. Therefore, with probability at least $1 - \exp(-cK\sqrt{n})$, we have for all $m \le N$,

$$\begin{aligned} A_m &= \sup_{y\in S^{n-1}} \sup_{\substack{z\in S^{N-1}\\ |\operatorname{supp} z|\le m}} \Big|\sum_{i=1}^N \langle X_i z_i, y\rangle\Big| = \sup_{y\in S^{n-1}} \sup_{\substack{z\in S^{N-1}\\ |\operatorname{supp} z|\le m}} \Big|\sum_{i=1}^N \langle Y_i z_i, T^* y\rangle\Big| \\ &\le \|T^*\| CK\left(\sqrt{n} + \sqrt{m}\log\frac{2N}{m}\right) = CK\kappa\left(\sqrt{n} + \sqrt{m}\log\frac{2N}{m}\right), \end{aligned}$$

where $\kappa = \|T^*\| = \sqrt{\|\Sigma\|}$ (note that $\Sigma = TT^*$).

We conclude this section with a more technical variant of Theorem 3.6. Note that in particular it requires weaker conditions on X_i's and does not require any bounds on N.

Theorem 3.13. *Let $1 \le n$ and $1 \le N$. Let $X_1, \dots, X_N$ be independent random vectors in $\mathbb{R}^n$ such that*

$$\sup_{i\le N} \sup_{y\in S^{n-1}} \|\,\langle X_i, y\rangle\,\|_{\psi_1} \le \psi.$$

Let A be a random $n \times N$ matrix whose columns are X_i's, and A_m, $m \le N$, is defined as before. Then for every $1 \le m \le N$, every $0 \le l \le \log m$, and every $K \ge 1$ one has

$$\mathbb{P}\left(A_m \ge C\psi K \left(\frac{m}{2^l} \log \frac{48eN2^l}{m} + \sqrt{m} \log \frac{2N}{m} \right) + 6 \max_{i \le N} |X_i| \right)$$
$$\le (1+2l) \exp\left(-2K \frac{m}{2^l} \log \frac{12eN2^l}{m} \right),$$

where C is an absolute constant. In particular, choosing $0 \le l \le \log m$ to be the largest integer satisfying

$$\frac{2m}{2^l} \log \frac{12eN2^l}{m} \ge \sqrt{m} \log \frac{2N}{m}$$

we obtain that for every $K \ge 1$

$$\mathbb{P}\left(A_m \ge C\psi K \sqrt{m} \log \frac{2N}{m} + 6 \max_{i \le N} |X_i| \right) \le (1+2\log m) \exp\left(-K \sqrt{m} \log \frac{2N}{m} \right).$$

Remark 3.14. Note that from the definitions we immediately have

$$A_m \ge A_1 \ge \max_{i \le N} |X_i|.$$

For completeness we outline a proof of Theorem 3.13.

Proof (Sketch.) We proceed as in the proof of Theorem 3.6. So first we construct $\mathcal{M}$. If $l = 0$ we define $\mathcal{M}$ exactly as after formula (3.2), otherwise it will be constructed in the same way as it was constructed after formula (3.3) (note that now l is a fixed number). Then we estimate $D_x = D'_x + D''_x$. As before we use Lemmas 3.3 and 3.4.

The only difference is that for the first summand in the formula for D'_x we use Lemma 3.3 with $L = 4K \frac{m}{2^l} \log \frac{48eN2^l}{m}$ instead of $L = 2K\sqrt{n}$. It will give us that

$$\begin{aligned} \mathbb{P}\left(\sup_{x \in \mathcal{M}} D'_x \right. & > \left. 16 A_m K \psi \frac{m}{2^l} \log \frac{48eN2^l}{m} + C A_m K \psi \sqrt{m} \log \frac{2N}{m} \right) \\ & \le \exp\left(-2K \frac{m}{2^l} \log \frac{48eN2^l}{m} \right) + l \exp\left(-2K \frac{m}{2^l} \log \frac{12eN4^l}{m} \right) \end{aligned}$$

and

$$\mathbb{P}\left(\sup_{x\in\mathcal{M}} D''_x > 3C\psi A_m K\sqrt{m}\log\frac{2N}{m}\right) \le l\exp\left(-K\frac{6m}{2^l}\log\frac{12e4^lN}{m}\right).$$

Thus, with another absolute positive constant C we have

$$\begin{aligned}\mathbb{P}\left(\sup_{x\in\mathcal{M}} D_x \right. &> \left. CA_mK\psi\left(\frac{m}{2^l}\log\frac{48eN2^l}{m}+\sqrt{m}\log\frac{2N}{m}\right)\right)\\ &\le (1+2l)\exp\left(-K\frac{2m}{2^l}\log\frac{12eN2^l}{m}\right).\end{aligned}$$

Finally we apply the same approximation procedure. By (3.4) and approximation we get formula (3.6)

$$A_m^2 \le \max\{36\max_i |X_i|^2, 18\sup_{x\in\mathcal{M}} D_x\},$$

which implies the result, by adjusting constants, if necessary. The "in particular" part of the Theorem is trivial. $\square$

Remark 3.15. It is possible to extend Theorem 3.13 to a ψ_p-setting, similar to the one considered in [10]. Let $p\in[1,2]$ and let X be a random vector such that for some $\psi_p>0$ one has

$$\mathbb{E}\exp\left((|\langle X,y\rangle|/\psi_p)^p\right)\le 2$$

for every $y\in S^{n-1}$. Then, adjusting Lemmas 3.3 and 3.4, and repeating the proof of Theorem 3.13 we can get

$$\begin{aligned}\mathbb{P}&\left(A_m \ge C\psi_pK\sqrt{m}\left(\log\frac{2N}{m}\right)^{1/p}+6\max_{i\le N}|X_i|\right)\\ &\le (1+2\log m)\exp\left(-K^p\sqrt{m}\log\frac{2N}{m}\right).\end{aligned}$$

However we will not pursue this direction here.

4 Kannan-Lovász-Simonovits question

In this section, we answer the question presented in the introduction: *Let K be an isotropic convex body in $\mathbb{R}^n$. Given $\varepsilon > 0$, how many independent points X_i uniformly distributed on K are needed for the empirical covariance matrix to approximate the identity up to ε with overwhelming probability?*

Let $X \in \mathbb{R}^n$ be a centered random vector with covariance matrix Σ and consider N independent random vectors $(X_i)_{i\le N}$ distributed as X. Using empirical processes tools, we first prove a more general statement (Proposition 4.4) and then give applications to approximation of the empirical covariance matrix and to estimates of different norms of the matrix $A = A^{(N)}$. In a final subsection we give a more elementary proof of the case ($p = 2$) that corresponds to the original question in [12].

4.1 Approximation of covariance matrix

First note that because of the linear invariance, (1.5) implies

$$\Big\| \frac{1}{N}\sum_{i=1}^N X_i \otimes X_i - \Sigma \Big\| \le \varepsilon \|\Sigma\|.$$

Therefore without loss of generality we restrict ourselves to the case when the covariance matrix is the identity.

Theorem 4.1. *Let $X_1, \ldots, X_N$ be i.i.d. random vectors, distributed according to an isotropic, log-concave probability measure on $\mathbb{R}^n$. For every $\varepsilon \in (0,1)$ and $t \ge 1$, there exists $C(\varepsilon, t) > 0$, such that if $C(\varepsilon,t)n \le N$, then with probability at least $1 - e^{-ct\sqrt{n}}$,*

$$\Big\| \frac{1}{N}\sum_{i=1}^N X_i \otimes X_i - \mathrm{Id} \Big\| \le \varepsilon, \tag{4.1}$$

where $c > 0$ is an absolute constant. Moreover, one can take $C(\varepsilon,t) = Ct^4\varepsilon^{-2}\log^2(2t^2\varepsilon^{-2})$, where $C > 0$ is an absolute constant.

Since for a symmetric matrix M, one has $\|M\| = \sup_{y\in S^{n-1}} \langle My, y\rangle$ and $\mathbb{E}\langle X_i, y\rangle^2 = |y|^2$, one can rewrite (4.1) as

$$\sup_{y\in S^{n-1}} \Big| \frac{1}{N}\sum_{i=1}^N (\langle X_i, y\rangle^2 - \mathbb{E}\langle X_i, y\rangle^2) \Big| \le \varepsilon.$$

This way approximating the covariance matrix becomes a special case of a more general problem, concerning the uniform approximation of the moments of one dimensional marginals of an isotropic log-concave measure by their empirical counterparts. In particular, Theorem 4.1 is implied by the following result.

Theorem 4.2. *Let $X_1, \ldots, X_N$ be i.i.d. random vectors, distributed according to an isotropic, log-concave probability measure on $\mathbb{R}^n$. For any $p \geq 2$ and for every $\varepsilon \in (0,1)$ and $t \geq 1$, there exists $C(\varepsilon, t, p) > 0$, such that if $C(\varepsilon, t, p) n^{p/2} \leq N$, then with probability at least $1 - e^{-c_p t \sqrt{n}}$ (where $c_p > 0$ depends only on p),*

$$\sup_{y \in S^{n-1}} \left| \frac{1}{N} \sum_{i=1}^{N} (|\langle X_i, y\rangle|^p - \mathbb{E}|\langle X_i, y\rangle|^p) \right| \leq \varepsilon. \tag{4.2}$$

Moreover, one can take $C(\varepsilon, t, p) = C_p t^{2p} \varepsilon^{-2} \log^{2p-2}(2t^2\varepsilon^{-2})$, where C_p depends only on p.

Remark 4.3. Proofs of both Theorems, 4.1 and 4.2, use Theorem 3.6 which requires the condition $N \leq \exp(\sqrt{n})$. For larger N, however, the result follows by a formal argument. Assume that the statement has been proved for $N \leq \exp(\sqrt{n})$ and assume that $N > \exp(\sqrt{n})$. Let $X_i = \{X_i(k)\}_{k=1}^n \in \mathbb{R}^n$, $i \leq N$, be the random vectors under consideration. Pick the smallest m such that $N \leq \exp(\sqrt{m})$. Clearly, $m > n$. Now consider random vectors $Y_i = \{Y_i(k)\}_{k=1}^m \in \mathbb{R}^m$, $i \leq N$, defined by $Y_i(k) = X_i(k)$ for $k \leq n$ and $Y_i(k) = g_{ik}$ for $k > n$, where g_{ik} are independent Gaussian $\mathcal{N}(0,1)$ random variables. Then Y_i's are isotropic log-concave random vectors to which the result can be applied. Identifying $y = \{y(k)\}_{k=1}^n \in S^{n-1}$ with $z = \{z(k)\}_{k=1}^m \in S^{m-1}$, defined by $z(k) = y(k)$ for $k \leq n$, $z(k) = 0$ for $k > n$, we get

$$\begin{aligned} &\sup_{y \in S^{n-1}} \left| \frac{1}{N} \sum_{i=1}^{N} (|\langle X_i, y\rangle|^p - \mathbb{E}|\langle X_i, y\rangle|^p) \right| \\ \leq \quad &\sup_{y \in S^{m-1}} \left| \frac{1}{N} \sum_{i=1}^{N} (|\langle Y_i, y\rangle|^p - \mathbb{E}|\langle Y_i, y\rangle|^p) \right| \leq \varepsilon \end{aligned}$$

with probability even higher than claimed. Thus in the proofs of both theorems we may assume without loss of generality that $N \leq \exp(\sqrt{n})$.

In the first step of the proof of Theorem 4.2 we shall use some tools from the probability in Banach spaces, in particular classical symmetrization and contraction methods as in [11] and [17]. These tools work for general empirical processes and are not necessary in our setting since we are dealing more specifically with powers of linear forms. We choose this approach, though, as it requires less computations and leads to a unified, simpler and more transparent presentation.

Theorem 4.2 is an easy consequence of the following technical proposition applied with $s = t$.

Proposition 4.4. *In the setting of Theorem 4.2, if $n \leq N \leq e^{\sqrt{n}}$, then for any $s, t \geq 1$, the estimate*

$$\sup_{y \in S^{n-1}} \Big| \frac{1}{N} \sum_{i=1}^{N} (|\langle X_i, y \rangle|^p - \mathbb{E}|\langle X_i, y \rangle|^p) \Big| \\ \leq C^{p-1} t s^{p-1} p \log^{p-1}\Big(\frac{2N}{n}\Big) \sqrt{\frac{n}{N}} + \frac{C^p s^p n^{p/2}}{N} + C^p p^p \Big(\frac{n}{2N}\Big)^s \tag{4.3}$$

holds with probability at least

$$1 - \exp(-cs\sqrt{n}) - \exp\big(-c_p \min\{u, v\}\big)$$

where $u = t^2 s^{2p-2} n \log^{2p-2}(2N/n)$, $v = ts^{-1}\sqrt{Nn}/\log(2N/n)$, $C, c > 0$ are absolute constants and $c_p > 0$ depends on p only.

Remark 4.5. The two parameters s and t play different role in the proof and reflect different asymptotic behavior of the probability with which (4.4) holds. The first parameter s is related to a level of truncation of linear forms whereas the second is a factor in the deviation when one deals only with the truncated part. For instance, by taking $s = t^{1/2}$, it allows us to get a probability converging to one as $t \to \infty$, if both dimensions are fixed.

Before we proceed to the proof of the above proposition, let us introduce some tools from the classical theory of probability in Banach spaces. Below, $\varepsilon_1, \ldots, \varepsilon_N$ will always denote a sequence of independent Rademacher variables, independent of the sequence $X_1, \ldots, X_N$.

Lemma 4.6 (Contraction principle, see [16], Theorem 4.12)**.** *Let $F\colon \mathbb{R}^+ \to \mathbb{R}_+$ be convex and increasing. Let further $\varphi_i\colon \mathbb{R} \to \mathbb{R}$, $i \le N$ be 1-Lipschitz with $\varphi_i(0) = 0$. Then, for any bounded set $T \subset \mathbb{R}^N$,*

$$\mathbb{E}F\Big(\frac{1}{2}\sup_{t\in T}\Big|\sum_{i=1}^N \varepsilon_i\varphi_i(t_i)\Big|\Big) \le \mathbb{E}F\Big(\sup_{t\in T}\Big|\sum_{i=1}^N \varepsilon_i t_i\Big|\Big).$$

Using standard symmetrization inequalities for sums of independent random variables (see e.g., Chapter 2.3. of [26]) and applying the lemma with $F \equiv 1$, and $\varphi_i(s) = \frac{|s|^p \wedge B^p}{pB^{p-1}}$ for $s \in \mathbb{R}$, we obtain the following corollary.

Corollary 4.7. *Let $\mathcal{F}$ be a family of functions, uniformly bounded by $B > 0$. Then for any independent random variables $X_1, \ldots, X_N$ and any $p \ge 1$, we have*

$$\mathbb{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^N(|f(X_i)|^p - \mathbb{E}|f(X_i)|^p)\Big| \le 4pB^{p-1}\mathbb{E}\sup_{f\in\mathcal{F}}\Big|\sum_{i=1}^N \varepsilon_i f(X_i)\Big|$$

We will also use the celebrated Talagrand's concentration inequality for suprema of bounded empirical processes [25]. The version from [13] presented below, provides the best known constants in this inequality (we will however not take advantage of explicit constants). For a simple proof (with worse constants) we refer the reader to [14, 15]

Lemma 4.8 ([13], Theorem 1.1)**.** *Let $X_1, X_2, \ldots, X_N$ be independent random variables with values in a measurable space $(\mathcal{S}, \mathcal{B})$ and let $\mathcal{F}$ be a countable class of measurable functions $f\colon \mathcal{S} \to [-a, a]$, such that for all i, $\mathbb{E}f(X_i) = 0$. Consider the random variable*

$$Z = \sup_{f\in\mathcal{F}}\sum_{i=1}^N f(X_i).$$

Then, for all $t \ge 0$,

$$\mathbb{P}(Z \ge \mathbb{E}Z + t) \le \exp\Big(-\frac{t^2}{2(\sigma^2 + 2a\mathbb{E}Z) + 3at}\Big),$$

where

$$\sigma^2 = \sup_{f\in\mathcal{F}}\sum_{i=1}^N \mathbb{E}f(X_i)^2.$$

Proof of Proposition 4.4 For simplicity, throughout this proof we will use the letter C to denote absolute constants, whose values may change from line to line.

For $B > 1$ (to be specified later) consider

$$\begin{aligned}
&\mathbb{E} \sup_{y \in S^{n-1}} \Big| \sum_{i=1}^N \Big((|\langle X_i, y\rangle| \wedge B)^p - \mathbb{E}(|\langle X_i, y\rangle| \wedge B)^p \Big) \Big| \\
&\quad \le 4pB^{p-1} \mathbb{E} \sup_{y \in S^{n-1}} \Big| \sum_{i=1}^N \varepsilon_i (|\langle X_i, y\rangle| \wedge B) \Big|,
\end{aligned}$$

where the last line follows from Corollary 4.7. The function $t \mapsto |t| \wedge B$ is a contraction, so

$$\begin{aligned}
&\mathbb{E} \sup_{y \in S^{n-1}} \Big| \sum_{i=1}^N \Big((|\langle X_i, y\rangle| \wedge B)^p - \mathbb{E}(|\langle X_i, y\rangle| \wedge B)^p \Big) \Big| \\
&\quad \le 8pB^{p-1} \mathbb{E} \sup_{y \in S^{n-1}} \Big| \sum_{i=1}^N \varepsilon_i \langle X_i, y\rangle \Big| \le 8pB^{p-1} \mathbb{E} \Big| \sum_{i=1}^N \varepsilon_i X_i \Big| \\
&\quad \le 8pB^{p-1} \sqrt{Nn}.
\end{aligned}$$

Since by (2.3), $\mathbb{E}(|\langle X_i, y\rangle| \wedge B)^{2p} \le C^{2p} p^{2p}$, Lemma 4.8 implies that for $t \ge 1$, with probability at least

$$\begin{aligned}
1 - \exp\Big(- \frac{64 B^{2p-2} t^2 Nn}{2NC^{2p}p^{2p} + 32pB^{2p-1}\sqrt{Nn} + 24pB^{2p-1} t \sqrt{Nn}} \Big) \\
\ge 1 - \exp(-c_p \min(t^2 n B^{2p-2}, t\sqrt{Nn}/B)),
\end{aligned} \tag{4.4}$$

one has

$$\sup_{y \in S^{n-1}} \Big| \sum_{i=1}^N \Big((|\langle X_i, y\rangle| \wedge B)^p - \mathbb{E}(|\langle X_i, y\rangle| \wedge B)^p \Big) \Big| \le 16tpB^{p-1}\sqrt{Nn}. \tag{4.5}$$

Observe that

$$\sup_{y\in S^{n-1}}\Big|\frac{1}{N}\sum_{i=1}^{N}(|\langle X_i,y\rangle|^p-\mathbb{E}|\langle X_i,y\rangle|^p)\Big|$$
$$\leq \sup_{y\in S^{n-1}}\Big|\sum_{i=1}^{N}\frac{1}{N}(|\langle X_i,y\rangle|\wedge B)^p-\mathbb{E}(|\langle X_i,y\rangle|\wedge B)^p)\Big|$$
$$+\sup_{y\in S^{n-1}}\frac{1}{N}\sum_{i=1}^{N}(|\langle X_i,y\rangle|^p-B^p)\mathbf{1}_{\{|\langle X_i,y\rangle|\geq B\}}$$
$$+\sup_{y\in S^{n-1}}\frac{1}{N}\mathbb{E}\sum_{i=1}^{N}(|\langle X_i,y\rangle|^p-B^p)\mathbf{1}_{\{|\langle X_i,y\rangle|\geq B\}},$$

Each of the obtained three terms is estimated separately, with the first term already discussed in (4.5) and (4.4). By (2.3) and Chebyshev's inequality we have

$$\mathbb{E}|\langle X_i,y\rangle|^p\mathbf{1}_{\{|\langle X_i,y\rangle|\geq B\}}\leq \|\langle X_i,y\rangle\|_{2p}^p\sqrt{\mathbb{P}(|\langle X_i,y\rangle|\geq B)}\leq C^pp^pe^{-B/C}.$$

Together with the previous inequalities this implies that

$$\sup_{y\in S^{n-1}}\Big|\frac{1}{N}\sum_{i=1}^{N}(|\langle X_i,y\rangle|^p-\mathbb{E}|\langle X_i,y\rangle|^p)\Big|$$
$$\leq 16tpB^{p-1}\sqrt{\frac{n}{N}}+\sup_{y\in S^{n-1}}\frac{1}{N}\sum_{i=1}^{N}|\langle X_i,y\rangle|^p\mathbf{1}_{\{|\langle X_i,y\rangle|\geq B\}}+C^pp^pe^{-B/C}, \tag{4.6}$$

with probability at least

$$1-\exp(-c_p\min(t^2nB^{2p-2},t\sqrt{Nn}/B)).$$

Thus it remains to estimate $\sup_{y\in S^{n-1}}\sum_{i=1}^{N}|\langle X_i,y\rangle|^p\mathbf{1}_{\{|\langle X_i,y\rangle|\geq B\}}$. To this end we use Theorem 3.6 and Remark 3.10. It follows that for $s\geq 1$, with probability at least $1-e^{-cs\sqrt{n}}$, we have, for all $m\leq N$ and all $z\in S^{N-1}$ with $|\operatorname{supp} z|=m$,

$$\Big|\sum_{i=1}^{N}z_iX_i\Big|\leq Cs\Big(\sqrt{n}+\sqrt{m}\log\Big(\frac{2N}{n}\Big)\Big). \tag{4.7}$$

Dualizing this estimate and using the fact that for $p \geq 2$, the ℓ_p norm is dominated by the ℓ_2 norm, we obtain, for any set $E \subset \{1, \ldots, N\}$,

$$\begin{aligned} \sup_{y \in S^{n-1}} \left(\sum_{i \in E} |\langle X_i, y \rangle|^p \right)^{1/p} &\leq \sup_{y \in S^{n-1}} \left(\sum_{i \in E} |\langle X_i, y \rangle|^2 \right)^{1/2} \\ &\leq Cs\Big(\sqrt{n} + \sqrt{|E|} \log\Big(\frac{2N}{n}\Big)\Big). \end{aligned} \tag{4.8}$$

For an arbitrary $y \in S^{n-1}$ let $E_B = E_B(y) := \{i \leq N : |\langle X_i, y \rangle| \geq B\}$. Then, by (4.8),

$$B|E_B|^{1/2} \leq \left(\sum_{i \in E_B} |\langle X_i, y \rangle|^2 \right)^{1/2} \leq Cs\Big(\sqrt{n} + \sqrt{|E_B|} \log\Big(\frac{2N}{n}\Big)\Big).$$

Thus, whenever

$$B \geq 2Cs \log\Big(\frac{2N}{n}\Big), \tag{4.9}$$

we obtain (for a different absolute constant C),

$$|E_B| \leq Cs^2 n B^{-2}.$$

This combined with (4.8) implies, after taking the p'th powers and again adjusting constants, that with probability at least $1 - e^{-cs\sqrt{n}}$, for all $y \in S^{n-1}$,

$$\begin{aligned} \sum_{i=1}^{N} |\langle X_i, y \rangle|^p \mathbf{1}_{\{|\langle X_i, y \rangle| \geq B\}} &= \sum_{i \in E_B} |\langle X_i, y \rangle|^p \\ &\leq C^p s^p \Big(n^{p/2} + n^{p/2} s^p B^{-p} \log^p\Big(\frac{2N}{n}\Big)\Big). \end{aligned}$$

Setting $B = 2Cs\log(2N/n)$, so that (4.9) is satisfied, and combining the resulting estimate with (4.6), we get

$$\begin{aligned} &\sup_{y \in S^{n-1}} \Big| \frac{1}{N} \sum_{i=1}^{N} (|\langle X_i, y \rangle|^p - \mathbb{E}|\langle X_i, y \rangle|^p) \Big| \\ &\qquad \leq 16 C^{p-1} t s^{p-1} p \log^{p-1}\Big(\frac{2N}{n}\Big) \sqrt{\frac{n}{N}} + \frac{C^p s^p n^{p/2}}{N} + C^p p^p \Big(\frac{n}{2N}\Big)^s, \end{aligned}$$

with probability at least

$$1 - \exp(-cs\sqrt{n}) - \exp\Big(- c_p \min\Big(t^2 s^{2p-2} n \log^{2p-2}(2N/n), \frac{ts^{-1}\sqrt{Nn}}{\log(2N/n)}\Big)\Big).$$

This completes the proof of Proposition 4.4, □

Remark 4.9. Let $G \in \mathbb{R}^n$ be a standard Gaussian vector with the identity as the covariance matrix and let h be a standard Gaussian random variable. Assume that h and G are independent and put $X = h\,G \in \mathbb{R}^n$. Clearly its covariance matrix is the identity and it is easy to check that $\|\langle X, y\rangle\|_{\psi_1} \le c|y|$, for every $y \in \mathbb{R}^n$, where c is a numerical constant. Nevertheless, it is known from [3] that X does not satisfy the conclusion of Lemma 3.1; in fact the density of X is not log-concave. Now let us consider the matrix $A = A^{(N)}$ with i.i.d. copies $X_i = h_i\, G_i$, $i = 1, \dots, N$ as columns with $N \le e^n$, where (h_i) are i.i.d copies of h and similarly (G_i) i.i.d copies of G, (h_i) and (G_i) independent. One can check that

$$\begin{aligned} \mathbb{E} \sup_{y\in S^{n-1}} \frac{1}{N} \sum_1^N \ |\langle X_i, y\rangle|^2 &= \mathbb{E} \sup_{y\in S^{n-1}} \frac{1}{N} \sum_1^N h_i^2 |\langle G_i, y\rangle|^2 \\ &\ge \ \mathbb{E} \sup_i \frac{1}{N} h_i^2 |G_i|^2 \ge c\frac{n}{N} \log N \end{aligned}$$

where $c > 0$ is a numerical constant. Thus $\|A\| \ge \sqrt{cn \log N}$. This example shows that the sub-exponential decay of linear forms (ψ_1 norm bounded) is not sufficient for our problem.

Remark 4.10. In comparison, a sub-gaussian decay of linear forms is sufficient. Indeed, it is known (see for instance [19]) that if there exists $c > 0$ such that $\mathbb{E} \exp(|c\langle X, y\rangle|^2) \le 2$ for every $y \in S^{n-1}$, then (1.5) holds with probability larger than $1 - \exp(-c'n)$ for some numerical constant $c' > 0$.

Remark 4.11. Another non necessarily log-concave example for which the conclusion of Theorems 3.6 and 4.1 are valid is obtained when $\|\langle X, y\rangle\|_{\psi_1} \le c|y|$, for every $y \in \mathbb{R}^n$ and $|X| \le C\sqrt{n}$ where $c, C > 0$ are numerical constants.

4.2 Additional observations

We note several observations for norms of random matrices from ℓ_2 to ℓ_p, $p \neq 2$.

Corollary 4.12. *For $1 \leq N \leq e^{\sqrt{n}}$ let Γ be a random $N \times n$ matrix with rows $X_1, \dots, X_N$. Then for $p \geq 2$, with probability at least $1 - e^{-c_p\sqrt{n}}$ (where $c_p > 0$ depends only on p),*

$$\|\Gamma\|_{\ell_2 \to \ell_p} \leq C_p(N^{1/p} + n^{1/2}), \tag{4.10}$$

with $C_p > 0$ depending only on p. Moreover

$$\tilde{c}_p N^{1/p} + c\sqrt{n} \leq \mathbb{E}\|\Gamma\|_{\ell_2 \to \ell_p} \leq \tilde{C}_p(N^{1/p} + n^{1/2}), \tag{4.11}$$

where $\tilde{C}_p, \tilde{c}_p > 0$ depend only on p and $c > 0$ is an absolute constant.

Proof Inequality (4.10) for $N \leq n$ follows from Theorem 3.6 and the comparison between ℓ_p norms. For $N \geq n$, the inequality follows from Proposition 4.4.

Since by log-concavity, moments and quantiles of $\|\Gamma\|_{\ell_2 \to \ell_p}$ are equivalent, (4.10) implies that

$$\mathbb{E}\|\Gamma\|_{\ell_2 \to \ell_p} \leq \tilde{C}_p(N^{1/p} + n^{1/2}).$$

On the other hand, a single row of Γ has expected Euclidean norm of the order of $\sqrt{n}$ and a single column of Γ has expected $\|\cdot\|_p$ norm of the order of $c(p)N^{1/p}$, so the left hand side of (4.11) follows trivially. □

Corollary 4.13. *For $1 \leq N \leq e^{\sqrt{n}}$ let Γ be a random $N \times n$ matrix with rows $X_1, \dots, X_N$. Then for $p \in [1, 2)$, with probability at least $1 - e^{-c\sqrt{n}}$ (where $c > 0$ is an absolute constant),*

$$\|\Gamma\|_{\ell_2 \to \ell_p} \leq C(N^{1/p} + N^{1/p-1/2}n^{1/2}) \tag{4.12}$$

for some absolute constant $C > 0$. Moreover

$$\tilde{c}(N^{1/p} + N^{1/p-1/2}n^{1/2}) \leq \mathbb{E}\|\Gamma\|_{\ell_2 \to \ell_p} \leq \tilde{C}(N^{1/p} + N^{1/p-1/2}n^{1/2}), \tag{4.13}$$

where $\tilde{C}, \tilde{c} > 0$ are absolute constants.

Proof Inequality (4.12) and the right-hand side of (4.13) follow from the corresponding results for $p = 2$, since

$$\|\Gamma\|_{\ell_2 \to \ell_p} \le N^{1/p-1/2} \|\Gamma\|_{\ell_2 \to \ell_2}.$$

To prove the left-hand side of (4.13), it is enough to notice that if $1/p^* + 1/p = 1$, then

$$\mathbb{E}\|\Gamma\|_{\ell_2 \to \ell_p} \ge \mathbb{E}\Big| \sum_{i=1}^{N} \frac{1}{N^{1/p^*}} X_i \Big| \ge \tilde{c} N^{1/2-1/p^*} n^{1/2} = \tilde{c} N^{1/p-1/2} n^{1/2}$$

and the expected ℓ_p norm of a single column of Γ is at least $\tilde{c} N^{1/p}$. □

One can also obtain an almost-isometric result for $p \in [1, 2)$.

Theorem 4.14. *Let $X_1, \ldots, X_N$ be i.i.d. random vectors, distributed according to an isotropic, log-concave probability measure on $\mathbb{R}^n$. For any $p \in [1, 2)$ and for every $\varepsilon \in (0, 1)$ and $t \ge 1$, there exists $C(\varepsilon, t) > 0$, such that if $C(\varepsilon) n \le N \le e^{\sqrt{n}}$, then with probability at least $1 - e^{-ct\sqrt{n}}$ (where $c > 0$ is an absolute constant),*

$$\sup_{y \in S^{n-1}} \Big| \frac{1}{N} \sum_{i=1}^{N} (|\langle X_i, y\rangle|^p - \mathbb{E}|\langle X_i, y\rangle|^p) \Big| \le \varepsilon. \tag{4.14}$$

Moreover, one can take $C(\varepsilon, t) = C t^{2p} \varepsilon^{-2} \log^{2p-2}(2t^{2p}\varepsilon^{-2})$, where $C > 0$ is an absolute constant.

Proof Since the proof differs only by technical details from the corresponding argument for $p \ge 2$, we will just indicate the necessary changes. We will use the notation from the proof of Proposition 4.4.

Just as before, we truncate at the level of $Ct \log(2N/n)$ and use the contraction principle to handle the bounded part of the process. As for the unbounded part, we also proceed as before, however now we use the comparison between the ℓ_2^k and ℓ_p^k norm for $p < 2$ and $k = |E_B| \le n$, which yields

$$\begin{aligned}
&\sup_{y \in S^{n-1}} \Big| \frac{1}{N} \sum_{i=1}^{N} (|\langle X_i, y\rangle|^p - \mathbb{E}|\langle X_i, y\rangle|^p) \Big| \\
&\qquad \le 16 C^{p-1} t^p p \log\Big(\frac{2N}{n}\Big)^{p-1} \sqrt{\frac{n}{N}} + \frac{C^p t^p n}{N} + \frac{C^p p^p n}{N},
\end{aligned}$$

with probability at least

$$1 - \exp(-ct\sqrt{n}) - \exp(-c\min(t^2 n \log^{2p-2}(2N/n), \sqrt{Nn}/\log(2N/n)))$$

(the constants in the exponents can be made independent of p, since now p runs over a bounded interval). This allows us to finish the proof. $\square$

Remark 4.15. The isomorphic result for $p = 1$ was proven in [10]. The same paper also considers $p \in (0, 1)$.

4.3 Elementary approach for $p = 2$

As announced earlier we will now briefly describe a more elementary proof of Theorem 4.1 and Theorem 4.2 for $p = 2$. In this case, the classical Bernstein inequality and a net argument on the sphere may replace the contraction principle and concentration of measure for empirical processes, that have been used – via Lemma 4.8 – to prove (4.5). The remaining part of the proof is left unchanged.

The key point is the following well known observation:

Lemma 4.16. *Let x_i, $i = 1, 2, \ldots, N$, be arbitrary vectors in $\mathbb{R}^n$. Let $\varepsilon \in (0, 1)$ and let $\mathcal{N}$ be a $c\varepsilon$-net of S^{n-1}, for some constant $c \in (0, 1)$. If we have*

$$\sup_{y \in \mathcal{N}} \Big| \frac{1}{N} \sum_{i=1}^{N} (\langle x_i, y \rangle^2 - 1) \Big| \leq \varepsilon$$

then

$$\sup_{y \in S^{n-1}} \Big| \frac{1}{N} \sum_{i=1}^{N} (\langle x_i, y \rangle^2 - 1) \Big| \leq c'\varepsilon$$

where c' depends on c.

We postpone the proof of this Lemma and pass to the proof of Theorems 4.1 and 4.2.

Fix a $c\varepsilon$-net $\mathcal{N}$ of S^{n-1} of cardinality at most $(3/c\varepsilon)^n$, and $B > 0$ to be determined later. Pick an arbitrary $y \in S^{n-1}$.

For the reader's convenience recall Bernstein's inequality.

Proposition 4.17 (Bernstein's inequality, cf. e.g., [26]). *Let Z_i be independent random variables, centered and such that $|Z_i| \leq a$ for all $1 \leq i \leq N$. Put $Z = \frac{1}{N}\sum_{i=1}^N Z_i$. Then for all $\tau \geq 0$,*

$$\mathbb{P}(Z \geq \tau) \leq \exp\Big(- \frac{\tau^2 N}{2(\sigma^2 + a\tau/3)}\Big),$$

where

$$\sigma^2 = (1/N)\sum_{i=1}^N Var(Z_i).$$

In our case $Z_i = (|\langle X_i, y\rangle| \wedge B)^2 - \mathbb{E}(|\langle X_i, y\rangle| \wedge B)^2$, for $1 \leq i \leq N$, $a = B^2$. Since $\mathbb{E}(|\langle X_i, y\rangle|)^2 = 1$ then (2.3) implies

$$Var(Z_i) \leq \mathbb{E}(|\langle X_i, y\rangle| \wedge B)^4 \leq c.$$

Setting $\tau = tB\sqrt{n/N}$ we infer that

$$\Big|\frac{1}{N}\sum_{i=1}^N \Big((|\langle X_i, y\rangle| \wedge B)^2 - \mathbb{E}(|\langle X_i, y\rangle| \wedge B)^2\Big)\Big| \geq tB\sqrt{n/N}$$

with probability at most

$$\exp\Big(-c\min\big(t^2B^2n, t\sqrt{Nn}/B\big)\Big).$$

By the union bound,

$$\sup_{y\in\mathcal{N}}\Big|\frac{1}{N}\sum_{i=1}^N \Big((|\langle X_i, y\rangle| \wedge B)^2 - \mathbb{E}(|\langle X_i, y\rangle| \wedge B)^2\Big)\Big| \leq tB\sqrt{n/N}, \qquad (4.15)$$

with probability at least

$$1 - \exp\Big(n\log\Big(\frac{3}{c\varepsilon}\Big) - c\min(t^2nB^2, t\sqrt{Nn}/B)\Big).$$

This estimate corresponds to (4.5).

Using this estimate with $B = Ct\log(2N/n)$ and handling the unbounded part the same way as in Proposition 4.4 (see the argument that follows (4.5)) we obtain

$$\sup_{y\in\mathcal{N}}\Big|\frac{1}{N}\sum_{i=1}^N(|\langle X_i, y\rangle|^2 - \mathbb{E}|\langle X_i, y\rangle|^2)\Big|$$
$$\leq Ct^2\log\Big(\frac{2N}{n}\Big)\sqrt{\frac{n}{N}} + \frac{C^2t^2n}{N} + \frac{4C^2n}{N}, \qquad (4.16)$$

with probability at least

$$1 - \exp(-ct\sqrt{n}) - \exp\Big(n \log\Big(\frac{3}{c\varepsilon}\Big) - c \min\Big(t^4 n \log^2(2N/n), \frac{\sqrt{Nn}}{C\log(2N/n)}\Big)\Big).$$

This corresponds to the estimates in Proposition 4.4 (for $s = t$).

Now, for $N \geq C(\varepsilon, t)n$, and $C(\varepsilon, t)$ sufficiently large, the right hand side of (4.16) is at most ε and $5/c\varepsilon \leq 2N/n$ which leads to the probability above to be at least $1 - \exp(-ct\sqrt{n})$. So with the same probability we get

$$\sup_{y \in \mathcal{N}} \Big| \frac{1}{N} \sum_{i=1}^{N} (|\langle X_i, y\rangle|^2 - \mathbb{E}|\langle X_i, y\rangle|^2) \Big| \leq \varepsilon.$$

We can now conclude by Lemma 4.16 applied pointwise with $x_i = X_i(\omega)$ for ω from the event on which our estimates hold (recall that by the isotropicity assumption we have $\mathbb{E}|\langle X_i, y\rangle|^2 = 1$).

Proof of Lemma 4.16 Consider the semi-norm $\|\cdot\|$ on $\mathbb{R}^n$ defined by

$$\|y\| = \Big(\frac{1}{N} \sum_{i=1}^{N} |\langle x_i, y\rangle|^2\Big)^{1/2},$$

for $y \in \mathbb{R}^n$. Our assumptions imply that

$$1 - \varepsilon \leq \sqrt{1 - \varepsilon} \leq \sup_{y \in \mathcal{N}} \|y\| \leq \sqrt{1 + \varepsilon} \leq 1 + \varepsilon/2.$$

The triangle inequality and homogeneity of $\|\cdot\|$ imply, by a standard argument, that

$$\sup_{y \in S^{n-1}} \|y\| \leq (1 + \varepsilon/2)(1 - c\varepsilon)^{-1} \leq 1 + \delta,$$

where

$$\delta = \frac{1 + 5c - 3c^2}{2(1 - c)}\varepsilon$$

To get a lower estimate, write an arbitrary $y \in S^{n-1}$ in the form $y = y_1 + c\varepsilon y_2$, with $y_1 \in \mathcal{N}$ and $y_2 \in S^{n-1}$. Then $\|y\| \geq \|y_1\| - c\varepsilon\|y_2\| \geq (1 - \varepsilon) - c\varepsilon(1 + \delta) \geq 1 - \delta_1$, where

$$\delta_1 = \frac{2 + c + 3c^2 - 3c^3}{2(1 - c)}\varepsilon.$$

Thus for all $y \in S^{n-1}$, $|\|y\| - 1| \leq c_1\varepsilon$ for some c_1 depending only on c. In particular $\|y\| \in [0, 1 + c_1]$. Using the fact that the function $t \mapsto t^2$ is Lipschitz with constant $2(1+c_1)$ on the interval $[0, 1+c_1]$, we conclude that

$$\sup_{y \in S^{n-1}} \Big| \frac{1}{N} \sum_{i=1}^{N} (\langle x_i, y \rangle^2 - 1) \Big| \leq c'\varepsilon,$$

where $c' = 2c_1(1 + c_1)$ depends only on c. □

References

[1] R. Adamczak, A. E. Litvak, A. Pajor, N. Tomczak-Jaegermann, Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling, preprint; available at `http://arxiv.org/abs/0904.4723`.

[2] G. Aubrun, Sampling convex bodies: a random matrix approach. *Proc. Amer. Math. Soc.* 135 (2007), 1293–1303.

[3] G. Aubrun, Private communication.

[4] 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.

[5] C. Borell, Convex set functions in d-space, *Math. Hungar.* 6 (1975), 111–136.

[6] C. Borell, The Brunn-Minkowski inequality in Gauss space, *Invent. Math.* 30 (1975), 207–216.

[7] J. Bourgain, Random points in isotropic convex sets. *In:* "Convex geometric analysis, Berkeley, CA, 1996", Math. Sci. Res. Inst. Publ., Vol. 34, 53–58, Cambridge Univ. Press, Cambridge (1999).

[8] K. R. Davidson, and S. Szarek, Local operator theory, random matrices and Banach spaces. *In* "Handbook on the Geometry of Banach spaces," Volume 1, 317–366; W. B. Johnson, J. Lindenstrauss eds., Elsevier Science 2001.

[9] A. A. Giannopoulos, M. Hartzoulaki, and A. Tsolomitis, Random points in isotropic unconditional convex bodies. *J. London Math. Soc.* 72 (2005), 779–798.

[10] A. A. Giannopoulos, and V. D. Milman, Concentration property on probability spaces. *Adv. Math.* 156 (2000), 77–106.

[11] O. Guédon, M. Rudelson, L_p-moments of random vectors via majorizing measures. *Adv. Math.* 208, no. 2 (2007), 798–823.

[12] R. Kannan, L. Lovász, M. Simonovits, Random walks and $O^*(n^5)$ volume algorithm for convex bodies, *Random structures and algorithms*, 2(1) (1997), 1–50.

[13] T. Klein, E. Rio, Concentration around the mean for maxima of empirical processes. *Ann. Probab.* 33 (2005), 1060–1077.

[14] M. Ledoux, On Talagrand's deviation inequalities for product measures. *ESAIM: Probability and Statistics* 1 (1996), 63–87.

[15] M. Ledoux, The concentration of measure phenomenon. Mathematical Surveys and Monographs, 89. American Mathematical Society, Providence, RI, 2001.

[16] M. Ledoux, M. Talagrand, Probability in Banach Spaces. Isoperimetry and processes, Volume 23 of Ergebnisse der Mathematik und ihrer Grenzgebiete (3). Springer-Verlag, Berlin, 1991.

[17] S. Mendelson, On weakly bounded empirical processes. *Math. Ann.* 340, no. 2 (2008), 293–314.

[18] S. Mendelson, A. Pajor, On singular values of matrices with independent rows, *Bernoulli* 12 (2006), 761–773.

[19] S. Mendelson, A. Pajor, N. Tomczak-Jaegermann, Reconstruction and subgaussian operators. *Geom. Funct. Anal.*, 17 (2007), 1248–1282.

[20] V. Milman, A. Pajor, Isotropic position and inertia ellipsoids and zonoids of the unit ball of a normed n-dimensional space. *Geom. Funct. Anal.* (1987–88), 64–104, Lecture Notes in Math., 1376, Springer, Berlin, 1989.

[21] A. Pajor, L. Pastur, On the Limiting Empirical Measure of the sum of rank one matrices with log-concave distribution, *Studia Math.* to appear.

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

[23] M. Rudelson, Random vectors in the isotropic position. *J. Funct. Anal.* 164, no. 1 (1999), 60–72.

[24] R. Schneider, Convex bodies: the Brunn-Minkowski theory. Encyclopedia of Mathematics and its Applications, 44. Cambridge University Press, Cambridge, 1993.

[25] M. Talagrand, New concentration inequalities in product spaces. *Invent. Math.* 126, no. 3 (1996), 505–563.

[26] A. W. van der Vaart, J.A. Wellner, Weak convergence and empirical processes. With applications to statistics. Springer Series in Statistics. Springer-Verlag, New York, 1996.

Radosław Adamczak,
Institute of Mathematics,
University of Warsaw,
Banacha 2, 02-097 Warszawa, Poland
radamcz@mimuw.edu.pl

Alexander E. Litvak,
Dept. of Math. and Stat. Sciences,
University of Alberta,
Edmonton, Alberta, Canada, T6G 2G1.
alexandr@math.ualberta.ca

Alain Pajor,
Université Paris-Est
Équipe d'Analyse et Mathématiques Appliquées,
5, boulevard Descartes, Champs sur Marne,
77454 Marne-la-Vallée, Cedex 2, France
e-mail: Alain.Pajor@univ-mlv.fr

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