Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling

R. Adamczak[1] A. E. Litvak A. Pajor
N. Tomczak-Jaegermann[2]

Abstract

This paper considers compressed sensing matrices and neighborliness of a centrally symmetric convex polytope generated by vectors $\pm X_1, \ldots, \pm X_N \in \mathbb{R}^n$, $(N \geq n)$. We introduce a class of random sampling matrices and show that they satisfy a restricted isometry property with overwhelming probability. In particular, we prove that matrices with i.i.d. centered and variance 1 entries that satisfy uniformly a sub-exponential tail inequality possess the restricted isometry property with overwhelming probability. We show that such "sensing" matrices are valid for the exact reconstruction process of m-sparse vectors via ℓ_1 minimization with $m \leq Cn/\log^2(cN/n)$. The class of sampling matrices we study includes the case of matrices with columns that are independent isotropic vectors with log-concave densities. We deduce that if $K \subset \mathbb{R}^n$ is a convex body and $X_1, \ldots, X_N \in K$ are i.i.d. random vectors uniformly distributed on K, then, with overwhelming probability, the symmetric convex hull of these points is an m-centrally-neighborly polytope with $m \sim n/\log^2(cN/n)$.

2010 AMS Classification:
primary 52A23, 60B20, 94A12, 52B12, 46B06
secondary 15B52, 41A45, 94B75

[1]The research of this author was partially supported by the Foundation for Polish Science

[2]This author holds the Canada Research Chair in Geometric Analysis.

Key Words and Phrases: Centrally-neighborly polytopes, compressed sensing, random matrices, restricted isometry property, underdetermined systems of linear equations.

1 Introduction

Let $1 \le m \le n \le N$ be integers and let $X_1, \dots, X_N \in \mathbb{R}^n$. Denote by A the $n \times N$ matrix with $X_1, \dots, X_N$ as columns and by $K(A) = K(X_1, \dots, X_N)$ the convex hull of $\pm X_1, \dots, \pm X_N$. Recall that a centrally symmetric convex polytope is m-centrally-neighborly if any set of less than m vertices containing no-opposite pairs, is the vertex set of a face (see the books [15] and [31]).

The connection between the neighborliness of $K(A)$ and sparse solutions of undetermined system of linear equations was discovered in [10], Theorem 1, where it is proved that the following two statements are equivalent:

i) $K(A)$ has $2N$ vertices and is m-neighborly

ii) whenever $y = Az$ has a solution z having at most m non-zero coordinates (in other words z is m-sparse), then z is the unique solution of the program:

$$(P) \qquad \min \|t\|_{\ell_1}, \quad At = Az.$$

Here the ℓ_1-norm is defined by $\|t\|_{\ell_1} = \sum_{i=1}^N |t_i|$ for any $t = (t_i)_{i=1}^N \in \mathbb{R}^N$.

Statement *ii)* is the so-called exact reconstruction problem by ℓ_1 minimization or *basis pursuit* algorithm. For a more detailed and complete analysis of the reconstruction of sparse vectors by the basis pursuit algorithm we refer to [7] and [11].

Let us also mention in the same stream of ideas that problem *ii)* is dual to the problem of decoding by linear programming. In this latter problem a linear code is given by the matrix A^*, and thus a vector $x \in \mathbb{R}^n$ generates the vector $A^*x \in \mathbb{R}^N$ defined by measurements $\big(\langle X_1, x\rangle, \dots, \langle X_N, x\rangle\big)$. Suppose that A^*x is corrupted by a noise vector $z \in \mathbb{R}^N$ which is assumed to be m-sparse. The problem is to *reconstruct* x from the data, which is the noisy output $y = A^*x + z$. This problem is then tackled by a linear programming approach (see [8] for complete references) that consists of the following minimization problem

$$(P') \qquad \min_{t \in \mathbb{R}^n} \|y - A^*t\|_{\ell_1}.$$

Let us denote by $|\cdot|$ the natural Euclidean norm in $\mathbb{R}^n$ and $\mathbb{R}^N$. Looking for a sufficient condition for a given matrix M to satisfy condition *ii)*, the authors of [8] introduced the concept of the *restricted isometry property* defined by the following parameter.

Definition 1.1. *Let M be an $n \times N$ matrix and let $\delta \in (0,1)$. For any $1 \leq m \leq N$ the isometry constant of M is defined as the smallest number $\delta_m = \delta_m(M)$ so that*

$$(1-\delta_m)|z|^2 \leq |Mz|^2 \leq (1+\delta_m)|z|^2$$

holds for all m-sparse vectors $z \in \mathbb{R}^N$. The matrix M is said to satisfy the restricted isometry property of order m with parameter δ, shortly $\mathrm{RIP}_m(\delta)$, *if $0 \leq \delta_m(M) < \delta$.*

The relevance of this parameter for the reconstruction property *ii)* is for instance revealed in [7],[8], where it was shown that if $\delta_m(M) + \delta_{2m}(M) + \delta_{3m}(M) < 1$ then M satisfies *ii)* (see also [6], [9], [17]). In the present paper, we shall use the following sufficient condition from [5]: if a matrix M satisfies

$$\delta_{2m}(M) < \sqrt{2} - 1$$

then *i)* and *ii)* are satisfied. In other words, if M has $\mathrm{RIP}_{2m}(\sqrt{2}-1)$ then M has the reconstruction property *ii)*. This approach gives the strategy of our paper.

Recall that no constructive method to produce centrally symmetric polytopes is known to give polytopes with an optimal order of neighborliness. All known results are of randomized nature, namely, they show that for a certain probability on the space of $n \times N$ matrices, a polytope $K(A)$ is m-centrally-neighborly with overwhelming probability, for (large) m depending on n and N. Consequently, from now on, A will be a random matrix in some *Ensemble* in the sense of the Random Matrix Theory. Due to the normalization, we shall consider the isometry constant of $A/\sqrt{n}$. The plan is to specialize to some model of random matrices the condition $\delta_{2m}\left(\frac{A}{\sqrt{n}}\right) < \sqrt{2} - 1$.

Let $X_1, \ldots, X_N \in \mathbb{R}^n$ be independent random vectors normalized so that $\mathbb{E}|X_i|^2 = n$ for all $i = 1, \ldots, N$. The model we will develop here is structured by two conditions: an inequality of the tails of linear forms and an inequality of concentration of the Euclidean norm.

- Linear forms obey a uniform sub-exponential decay, that is, for all $1 \le i \le N$, all $y \in S^{n-1}$, and $t > 0$,

$$\mathbb{P}\left(|\langle X_i, y\rangle| > t\right) \le C \exp(-ct),$$

  where $C, c > 0$.

- The Euclidean norms of $X_1, \ldots, X_N$ are concentrated around their average:

$$\mathbb{P}\left(\max_{i \le N} \left|\frac{|X_i|^2}{n} - 1\right| \ge \frac{\sqrt{2}-1}{2}\right) < \lambda.$$

Note that such a concentration inequality is clearly necessary in order to have $\mathrm{RIP}_1((\sqrt{2}-1)/2)$.

One of the main results of this paper, Theorem 4.3, claims that under these conditions, whenever

$$m \le cn/\log^2(CN/n),$$

the random polytope $K(A)$ is m-centrally-neighborly with probability larger than $1-2\lambda-C\exp(-c\sqrt{n})$, where $C, c > 0$ are universal numerical constants. We will make it more precise in Section 4. This model includes the cases when

- X_i's are independent isotropic random vectors with a log-concave density;

- the entries of the matrix are independent, centered with variance one and satisfy a sub-exponential tail inequality;

- X_i's are on the sphere of radius $\sqrt{n}$ and linear forms exhibit a uniform sub-exponential tail inequality.

These examples give rise to new classes of compressed sensing matrices. The class of i.i.d. entries with sub-exponential tail behavior (that is, entries being ψ_1 random variables), contains a subclass of matrices with i.i.d. ψ_r entries for $1 < r \le 2$ (see Definition 2.1 below of ψ_r random variables). Since in this case the obtained bounds are better by a power of logarithm that may be essential in applications, we prove our results in full generality, for $1 \le r \le 2$.

Regarding the restricted isometry property, our result is optimal in the following sense. Let M be a $n \times N$ matrix with entries $X_{ij}/\sqrt{n}$, $i = 1, \ldots n$,

$j = 1, \ldots N$, where X_{ij} are i.i.d. symmetric exponential variables with variance one. There exist constants $C > 0$ and $0 < c < 1$ such that if the probability that the random matrix M satisfies the restricted isometry property of order m with some parameter $\delta \in (0, 1)$ is larger than c, then

$$m \log^2 \left(\frac{2N}{m} \right) \leq Cn.$$

Sub-gaussian matrices with independent ψ_2 entries, which correspond to $r = 2$, are by now well understood. They include for instance the Gaussian case when the matrix A is built with i.i.d. Gaussian $N(0, 1)$ random variables (see [11],[8],[27]); the case when the entries of A are i.i.d. (± 1) Bernoulli random variables ([8], [23], [3]); a general case of i.i.d. sub-gaussian entries is treated in ([23],[24], also see [25] for simpler proofs).

Results of this paper are based on concentration type inequalities for random matrices under consideration. The proof of the main technical result, Theorem 3.1, will employ methods from [1]. A crucial new ingredient consists of an analysis of the quantity

$$B_m := \sup_{z \in U_m} \left| \left| \sum_{i \leq N} z_i X_i \right|^2 - \sum_{i \leq N} z_i^2 |X_i|^2 \right|^{1/2},$$

where U_m denotes the set of norm one m-sparse vectors in $\mathbb{R}^N$. In Section 2 we present some definitions and preliminary tools. In Section 3 we apply Theorem 3.1 to estimate the isometry constant (Theorem 3.2). Then we study the m-neighborly property of random polytopes in Section 4 and give application to polytopes generated by random points from a convex body, polytopes generated by independent vectors with independent ψ_r random coordinates, and polytopes generated by independent ψ_r random vectors on a sphere. Section 5 is devoted to the proof of Theorem 3.1 and discussion of optimality of the result.

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 use the same notation $|\cdot|$ to denote the cardinality of a set. Unless otherwise stated, $(X_i)_{i \geq 1}$ will denote independent random vectors 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. *For a random variable $Y \in \mathbb{R}$ and $r > 0$ we define the ψ_r-norm by*

$$\|Y\|_{\psi_r} = \inf \{C > 0\, ;\ \mathbb{E} \exp (|Y|/C)^r \le 2\} .$$

It is well known that the ψ_r-norm of a random variable may be estimated from the growth of the moments. More precisely if a random variable Y is such that for any $p \ge 1$, $\|Y\|_p \le p^{1/r} K$, for some $K > 0$, then $\|Y\|_{\psi_r} \le c_r K$ where c_r is a positive constant depending only on r.

Definition 2.2. *Let $X \in \mathbb{R}^n$ be a centered random vector and $r > 0$. We say that X is ψ_r or a ψ_r vector, if $\sup_{y \in S^{n-1}} \| \langle X, y\rangle \|_{\psi_r}$ is bounded and we set*

$$\|X\|_{\psi_r} = \sup_{y \in S^{n-1}} \| \langle X, y\rangle \|_{\psi_r}.$$

Remark: The above notation of $\|X\|_{\psi_r}$ for the *weak ψ_r norm* of a random vector X should not be confused with the standard convention in the probability theory that this notation stands for the ψ_r norm of the random variable $|X|$, i.e., $\| \, |X| \, \|_{\psi_r}$–this latter meaning will never be used in this paper.

We recall the well known Bernstein's inequality which we shall use in the form of a ψ_1 estimate ([30], p. 103).

Lemma 2.3. *Let $Y_1, ..., Y_n$ be independent real random variables with zero mean such that for some $\psi > 0$ and every i, $\|Y_i\|_{\psi_1} \le \psi$. Then, for any $t > 0$,*

$$\mathbb{P}\Big(\big|\sum_{i=i}^n Y_i\big| > t\Big) \le 2 \exp\left(- \frac{t^2}{4 \sum_{i \le n} \|Y_i\|_{\psi_1}^2 + 2t\psi}\right).$$

Given a set $E \subset \{1, ..., N\}$ by P_E we denote the orthogonal projection from $\mathbb{R}^N$ onto the coordinate subspace of vectors whose supports are in E. We denote this subspace by $\mathbb{R}^E$. The support of $z \in \mathbb{R}^N$ is denoted by $\operatorname{supp} z$. A vector $z \in \mathbb{R}^N$ is called m-sparse if $|\operatorname{supp} z| \le m$. The subset of m-sparse unit vectors in $\mathbb{R}^N$ is denoted by

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

As usual we denote the N-dimensional cube and the unit N-dimensional Euclidean ball by

$$B_\infty^N = \{x = (x_i)_{i=1}^N \in \mathbb{R}^N \,:\, \|x\|_\infty = \max_{i\le N} |x_i| \le 1\}$$

and

$$B_2^N = \{x = (x_i)_{i=1}^N \in \mathbb{R}^N \,:\, |x|^2 = \sum_{i=1}^N |x_i|^2 \le 1\}.$$

For every $E \subset \{1, ..., N\}$, $\varepsilon, \alpha \in (0,1]$ we select an ε-net (in the Euclidean metric) in $B_2^N \cap \alpha B_\infty^N \cap \mathbb{R}^E$ and denote it by $\mathcal{N}(E, \varepsilon, \alpha)$. Thus for every $x \in B_2^N \cap \alpha B_\infty^N$ supported by E, there exist $\bar{x} \in \mathcal{N}(E, \varepsilon, \alpha)$ supported by E such that $|x - \bar{x}| < \varepsilon$. A standard volume comparison argument 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.

Definition 2.4. *A random vector $X \in \mathbb{R}^n$ is called isotropic if*

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

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

A subset $K \subset \mathbb{R}^n$ is said to be isotropic when a random point X uniformly distributed in K is an isotropic random vector.

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

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

It is well known [4] that if a measure has a log-concave density, then linear functionals exhibit a sub-exponential decay. More precisely, we have:

Lemma 2.5. *[4]: Let $X \in \mathbb{R}^n$ be a centered random vector with a log-concave density. Then for every $y \in S^{n-1}$,*

$$\|\,\langle X, y\rangle\,\|_{\psi_1} \le c\,\left(\mathbb{E}|\langle X, y\rangle|^2\right)^{1/2},$$

where $c > 0$ is a universal constant. As a consequence, if X is an isotropic random vector with a log-concave density then $\|X\|_{\psi_1} \le c$.

The Euclidean norm of an isotropic random vector with a log-concave density highly concentrates around its expectation, this translates geometrically to the concentration of mass of an isotropic convex body within a thin Euclidean shell ([19], see also [14]). We will use here the following result immediately derived from [18], Theorem 4.4.

Lemma 2.6. *Let $1 \le n \le N$ be integers and let $X_1, \ldots, X_N \in \mathbb{R}^n$ be isotropic random vectors with log-concave densities. There exist numerical positive constants C, c_0 and $c_1 \in (0, \frac{1}{2})$ such that for all $\theta \in (0,1)$ and $N \le \exp(c\theta^{c_0} n^{c_1})$,*

$$\mathbb{P}\left(\max_{i\le N}\left|\frac{|X_i|^2}{n} - 1\right| \ge \theta\right) \le C\exp(-c\theta^{c_0} n^{c_1}).$$

Moreover, one can take $c_0 = 3.33$ and $c_1 = 0.33$.

Remark: It is conjectured that in the above theorem one can replace $\theta^{3.33} n^{0.33}$ by $c(\theta) n^{1/2}$.

We shall also use the following result from [26] as formulated in [1].

Lemma 2.7. *Let $N, n \ge 1$ be integers and let $X_1, \ldots, X_N \in \mathbb{R}^n$ be isotropic random vectors with log-concave densities. Then there exists an absolute positive constant C_0 such that for any $N \le \exp(\sqrt{n})$ and for every $K \ge 1$ one has*

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

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

In this paper, different universal positive constants may be denoted by the same letters C, C_0, C', c, c_0, c', etc.

3 Isometry constant

We begin this section by formulating, in Theorem 3.2, a general estimate for the isometry constant of random matrices with independent ψ_r columns. Then, in order to apply such an estimate, we introduce two sufficient conditions that determine large classes of random matrices. Finally, we give examples of important classes that satisfy the estimates from Theorem 3.2 and thus provide us with models: the Log-Concave Ensemble, matrices with i.i.d. ψ_r entries, and matrices defined by independent ψ_r vectors on a sphere.

3.1 Estimating the isometry constant

Techniques of "compressed sensing" rely on properties of the sampling matrix, which should act nearly isometrically on sparse vectors. This motivated the concept of restricted isometry property defined in [8]. To quantify this property of the "sensing" matrix, the authors introduced the isometry constant $\delta_m(M)$ defined in the Introduction (Definition 1.1) for any $n \times N$ matrix M and any $1 \le m \le N$. Of course if $m > n$ then $\delta_m \ge 1$.

Let $X_1, \dots, X_N \in \mathbb{R}^n$ and let $A = A^{(n,N)}$ be the "sampling" matrix with the X_i's as columns. We begin by a simple observation. For every $m \le N$ define the following quantity

$$B_m = \sup_{z \in U_m} \left| \left| \sum_{i \le N} z_i X_i \right|^2 - \sum_{i \le N} z_i^2 |X_i|^2 \right|^{1/2}. \tag{3.1}$$

Then, clearly

$$\delta_m\left(\frac{A}{\sqrt{n}}\right) = \sup_{z \in U_m} \left| \frac{|Az|^2}{n} - 1 \right| \le \frac{B_m^2}{n} + \max_{i \le N} \left| \frac{|X_i|^2}{n} - 1 \right|. \tag{3.2}$$

Thus the isometry constant is controlled by quantity B_m and the second term, $\max_{i \le N} \left| \frac{|X_i|^2}{n} - 1 \right|$. We begin by estimating B_m in the following technical theorem.

Theorem 3.1. *Let $n \ge 1$ and $1 \le m \le N$ be integers. Let $1 \le r \le 2$ and $X_1, \dots, X_N \in \mathbb{R}^n$ be independent ψ_r random vectors with $\psi = \max_{i \le N} \|X_i\|_{\psi_r}$. Let $\theta \in (0, 1/4)$, $K, K' \ge 1$ and assume that m satisfies*

$$m \log^{2/r} \frac{2N}{\theta m} \le \theta^2 n.$$

Then setting $\xi = \psi K + K'$, the inequality

$$B_m^2 \le C \xi^2 \theta n$$

holds with probability at least

$$1 - \exp\left(-K^r \sqrt{m} \log\left(\frac{2N}{\theta m}\right)\right) - \mathbb{P}\left(\max_{i \le N} |X_i| \ge K' \sqrt{n}\right),$$

where c is an absolute positive constant.

We postpone the proof of Theorem 3.1 to the last section. Combining this theorem with inequality (3.2), we immediately deduce an estimate for the isometry constant of a random matrix with independent ψ_r columns.

Theorem 3.2. *Let $n \geq 1$ and m, N be integers such that $1 \leq m \leq \min(N, n)$. Let $1 \leq r \leq 2$. Let $X_1, \ldots, X_N \in \mathbb{R}^n$ be independent ψ_r random vectors and let $\psi = \max_{i \leq N} \|X_i\|_{\psi_r}$. Let $\theta' \in (0,1)$, $K, K' \geq 1$ and set $\xi = \psi K + K'$. Then*

$$\delta_m\left(\frac{A}{\sqrt{n}}\right) \leq C\xi^2 \sqrt{\frac{m}{n}} \log^{1/r}\left(\frac{eN}{m\sqrt{\frac{m}{n}}}\right) + \theta'$$

holds with probability larger than

$$\begin{aligned} 1 \quad &- \quad \exp\left(-cK^r\sqrt{m}\log\left(\frac{eN}{m\sqrt{\frac{m}{n}}}\right)\right) \\ &- \quad \mathbb{P}\left(\max_{i \leq N} |X_i| \geq K'\sqrt{n}\right) - \mathbb{P}\left(\max_{i \leq N}\left|\frac{|X_i|^2}{n} - 1\right| \geq \theta'\right), \end{aligned}$$

where $C, c > 0$ are universal constants.

Remark. In fact to obtain Theorem 3.2 from Theorem 3.1 we have to choose $\theta = \sqrt{m/n}\log^{1/r}(en/(m\sqrt{m/n})$, hence formally Theorem 3.2 is a corollary of Theorem 3.1 only for m such that θ above is smaller than $1/4$. However, if $\theta \geq 1/4$, then, adjusting C (say $C = 4$), the upper bound for δ_m becomes larger than 1, so Theorem 3.2 follows easily from Theorem 3.13 in [1].

Note that bounds provided by Theorem 3.2 are interesting only if, firstly the bound on δ_m is smaller than 1 (which immediately implies the restriction $m \leq \min(N, n)$), and secondly, if it holds with positive probability. In fact, the former condition is equivalent to the restricted isometry property of order m. This leads to considerations of models of random $n \times N$ matrices that satisfy the following two conditions. Let $1 \leq r \leq 2$, $\psi > 0$ and $\lambda \in (0,1)$. Let $X_1, \ldots, X_N \in \mathbb{R}^n$ be independent ψ_r random vectors and let A be the matrix with $X_1, \ldots, X_N$ as columns.

- **Condition** $H_1(r, \psi)$**:** Linear forms obey a uniform ψ_r estimate:

$$\|X_i\|_{\psi_r} = \sup_{y \in S^{n-1}} \|\langle X_i, y\rangle\|_{\psi_r} \leq \psi \quad \text{for all} \quad 1 \leq i \leq N. \tag{3.3}$$

- **Condition $H_2(\lambda)$:** $|X_i|$'s are concentrated around their average:
$$\mathbb{P}\left(\max_{i\le N}\left|\frac{|X_i|^2}{n}-1\right|\ge\frac{\sqrt{2}-1}{2}\right)<\lambda. \tag{3.4}$$

As already mentioned in the Introduction, a condition such as $H_2(\lambda)$ is necessary to have the restricted isometry property. Indeed, if the matrix $A/\sqrt{n}$ has $\mathrm{RIP}_1((\sqrt{2}-1)/2)$ with probability λ then $H_2(\lambda)$ is satisfied.

3.2 Examples

We now specialize Theorem 3.2 to some specific classes of matrices, verifying conditions (3.3) and (3.4).

3.2.1 The Log-Concave Ensemble

We start by considering the "log-concave setting", where $X_1,\dots,X_N\in\mathbb{R}^n$ are independent isotropic vectors with log-concave densities.

Lemma 3.3. *Assume the above "log-concave setting". There exist universal constants $\psi, C, c>0$ such that conditions $H_1(1,\psi)$ and $H_2(C\exp(-cn^{c_1}))$ are satisfied whenever $N\le\exp(cn^{c_1})$, where c_1 is given in Lemma 2.6.*

The proof is immediate from Lemmas 2.5 and 2.6.

Applying Theorem 3.2 (with $r=1$) together with Lemmas 3.3 and 2.7 to the Log-Concave Ensemble, we get that for every $N\le\exp(cn^{c_1})$,
$$\delta_m\left(\frac{A}{\sqrt{n}}\right)\le C\sqrt{\frac{m}{n}}\log\left(\frac{eN}{m\sqrt{\frac{m}{n}}}\right)+\frac{\sqrt{2}-1}{2} \tag{3.5}$$
holds with probability larger than
$$1-\exp\left(-c\sqrt{m}\log\left(\frac{eN}{m\sqrt{\frac{m}{n}}}\right)\right)-e^{-c\sqrt{n}}-\exp(-cn^{c_1}),$$
where $C,c>0$ are universal constants and c_1 is given in Lemma 2.6.

It might be worthwhile to note that using directly Lemma 2.6 one can replace the second term in estimate (3.5) by a term tending to 0 when $n\to\infty$, but this would require an adjustment in probability. For example $1/n^{c_1/2c_0}$ works with the probability estimate in which $\exp(-cn^{c_1})$ is replaced by $\exp(-cn^{c_1/2})$. (Here c_0 is given in Lemma 2.6.)

3.2.2 Matrices with independent ψ_r entries

In this section we consider the "ψ_r setting", where the entries a_{ij} of the matrix A are independent centered, with variance one, random ψ_r variables (with $r \in [1,2]$). Although the argument is rather standard, we did not find it in literature, so we provide full proofs for completeness. We set $\psi = \max_{ij} \|a_{ij}\|_{\psi_r}$.

Lemma 3.4. *Assume the above "ψ_r setting" with $r \in [1,2]$. Then conditions $H_1(r, C\psi)$ and $H_2(2\exp(-cn^{r/2}/\psi^{2r}))$ are satisfied whenever $N \leq \exp(cn^{r/2}/\psi^{2r})$, where C, c are absolute positive constants.*

Proof. To prove that the columns of the matrix A are ψ_r vectors we will estimate the p-th moments of random variables $\sum_{i=1}^n y_i a_{ij}$, for any $y = (y_i) \in \mathbb{R}^n$ and any $p \geq 1$. This will be done by using Talagrand's concentration inequality for linear combinations of symmetric Weibull variables together with some symmetrization and truncation arguments.

The following Lemma is a combination of Corollaries 2.9 and 2.10 of [29].

Lemma 3.5. *Let $r \in [1,2]$ and $Y_1, \dots, Y_n$ be independent symmetric random variables satisfying $\mathbb{P}(|Y_i| \geq t) = \exp(-t^r)$. Then for every vector $a = (a_1, \dots, a_n) \in \mathbb{R}^n$ and every $t \geq 0$,*

$$\mathbb{P}\Big(\Big|\sum_{i=1}^n a_i Y_i\Big| \geq t\Big) \leq 2\exp\Big(-c\min\Big(\frac{t^2}{\|a\|_2^2}, \frac{t^r}{\|a\|_{r^*}^r}\Big)\Big),$$

where $1/r^* + 1/r = 1$ and $\|a\|_q = (|a_1|^q + \dots + |a_n|^q)^{1/q}$, for $1 \leq q < \infty$.

The behavior of general centered ψ_r variables can be easily reduced to symmetric Weibull variables. The argument is quite standard, we sketch it here for the sake of completeness.

Assume thus that $Z_1, \dots, Z_n$ are independent mean zero random variables with $\|Z_i\|_{\psi_r} \leq 1$. Let $\beta = (\log 2)^{1/r}$ and set $U_i = (|Z_i| - \beta)_+$. Let Y_i be defined as in Lemma 3.5.

We have for $t > 0$,

$$\begin{aligned}\mathbb{P}(U_i \geq t) &\leq \mathbb{P}(|Z_i| \geq t + \beta) \leq 2\exp(-(t+\beta)^r) \\ &\leq 2\exp(-t^r - \beta^r) = \mathbb{P}(|Y_i| \geq t).\end{aligned}$$

We will use the above observation together with symmetrization and the contraction principle to estimate moments of linear combinations of variables Z_i. We have for $p \geq 1$,

$$\begin{aligned}
\Big\| \sum_{i=1}^n a_i Z_i \Big\|_p &\leq 2 \Big\| \sum_{i=1}^n a_i \varepsilon_i Z_i \Big\|_p \quad \text{(symmetrization)} \\
&= 2 \Big\| \sum_{i=1}^n a_i \varepsilon_i |Z_i| \Big\|_p \\
&\leq 2 \Big\| \sum_{i=1}^n a_i \varepsilon_i (\beta + U_i) \Big\|_p \quad \text{(the contraction principle)} \\
&\leq 2 \Big\| \sum_{i=1}^n a_i \varepsilon_i \beta \Big\|_p + 2 \Big\| \sum_{i=1}^n a_i \varepsilon_i U_i \Big\|_p \\
&\leq C\sqrt{p}\beta \|a\|_2 + 2 \Big\| \sum_{i=1}^n a_i \varepsilon_i Y_i \Big\|_p \\
&\leq C\sqrt{p}\|a\|_2 + Cp^{1/r}\|a\|_{r^*},
\end{aligned}$$

where to get the last two inequalities we used Khinchine's inequality, Lemma 3.5 and integration by parts to pass from tail to moment estimates.

We are now ready to prove condition $H_1(r, C\psi)$. Fix $y \in S^{n-1}$ and consider the linear combination $\sum_{i=1}^n y_i a_{ij}$. Since $\|a_{ij}\|_{\psi_r} \leq \psi$ and $\|y\|_{r^*} \leq \|y\|_2 = 1$ for $r \in [1, 2]$, we obtain by homogeneity

$$\Big\| \sum_{i=1}^n y_i a_{ij} \Big\|_p \leq C\psi(\sqrt{p}\|y\|_2 + Cp^{1/r}\|y\|_{r^*}) \leq 2C\psi p^{1/r}.$$

The growth condition on the moments of the random variable $\sum_{i=1}^n y_i a_{ij}$ implies that its ψ_r norm is bounded by $\tilde{C}\psi$.

The proof of condition H_2 goes along similar lines. Instead of Lemma 3.5 we will now use the following lemma, which is an easy consequence of Theorem 6.2 in [16] and the observation that the p-th moment of a Weibull variable with parameter s is of order $C_s p^{1/s}$, where C_s remains bounded for s away from 0.

Lemma 3.6. *If $0 < s < 1$ and $Y_1, \ldots, Y_n$ are independent symmetric random variables satisfying $\mathbb{P}(|Y_i| \geq t) = \exp(-t^s)$, then for $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$*

and $p \geq 2$,

$$\Big\| \sum_{i=1}^{n} a_i Y_i \Big\|_p \leq C\sqrt{p}\|a\|_2 + C_s p^{1/s}\|a\|_p.$$

Moreover, for $s \geq 1/2$, C_s *is bounded by some absolute constant.*

Using similar arguments as in the proof of condition H_1 we can infer from the above lemma that if $Z_1, \ldots, Z_n$ are independent mean zero random variables with $\|Z_i\|_{\psi_s} \leq b$ ($s \in [1/2, 1)$), then for $p \geq 2$,

$$\Big\| \sum_{i=1}^{n} a_i Z_i \Big\|_p \leq Cb(\sqrt{p}\|a\|_2 + p^{1/s}\|a\|_p).$$

Therefore, for any $p \geq 2$ by the Chebyshev inequality in L_p,

$$\mathbb{P}\Big(\Big|\sum_{i=1}^{n} Z_i\Big| \geq Cb(\sqrt{np} + p^{1/s}n^{1/p})\Big) \leq \exp(-p).$$

For $p \geq 3$ we have

$$\sqrt{np} + p^{1/s}n^{1/p} \leq \tilde{C}(\sqrt{np} + p^{1/s})$$

with $\tilde{C}$ universal for $s \geq 1/2$, so the above inequality yields

$$\mathbb{P}\Big(\Big|\sum_{i=1}^{n} Z_i\Big| \geq Cb(\sqrt{np} + p^{1/s})\Big) \leq e^3 \exp(-p)$$

for some (new) universal constant C or equivalently

$$\mathbb{P}\Big(\Big|\sum_{i=1}^{n} Z_i\Big| \geq t\Big) \leq 2\exp\Big(-c\min\Big[\frac{t^2}{nb^2}, \Big(\frac{t}{b}\Big)^s\Big]\Big). \tag{3.6}$$

For fixed j we apply this inequality with $s = r/2$ to variables $Z_i = a_{ij}^2 - 1$. Note that $\mathbb{E}Z_i = 0$ and

$$\begin{aligned}\|Z_i\|_{\psi_{r/2}} &\leq C(1 + \|a_{ij}^2\|_{\psi_{r/2}}) \\ &= C(1 + \|a_{ij}\|_{\psi_r}^2) \leq \tilde{C}\psi^2.\end{aligned}$$

(The additional constants appearing above stem from the fact that under the standard definition for $s < 1$, $\|\cdot\|_{\psi_s}$ is not a norm but only a quasi-norm and additionally $\|1\|_{\psi_{r/2}} \neq 1$. One can modify the function $x \mapsto e^{x^r} - 1$ so that it is convex. For r away from zero, this modification changes the norm by an absolute constant). Therefore, applying (3.6) with $t = \varepsilon n$ yields

$$\begin{aligned}\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^{n} a_{ij}^2 - 1\Big| \geq \varepsilon\Big) &\leq 2\exp\Big(-c\min\Big[\frac{\varepsilon^2 n}{\psi^4}, \Big(\frac{\varepsilon n}{\psi^2}\Big)^{r/2}\Big]\Big)\\ &\leq 2\exp\Big(-\tilde{c}\frac{\varepsilon^r n^{r/2}}{\psi^{2r}}\Big).\end{aligned}$$

For $r = 2$ the proof is similar, but uses Lemma 3.5, which in this case reduces to Bernstein's ψ_1 inequality (Lemma 2.3), instead of Lemma 3.6. The argument is simpler since in this case the involved norms of the vector a do not depend on p and we get (3.6) directly.

The lemma follows now by the union bound. □

Applying Theorem 3.2 together with Lemma 3.4 to the "ψ_r setting", we get that for every $N \leq \exp(cn^{r/2}/\psi^{2r})$,

$$\delta_m\left(\frac{A}{\sqrt{n}}\right) \leq C\psi^2\sqrt{\frac{m}{n}}\log^{1/r}\left(\frac{eN}{m\sqrt{\frac{m}{n}}}\right) + \frac{\sqrt{2}-1}{2}$$

holds with probability at least

$$1 - \exp\left(-c\sqrt{m}\log\left(\frac{eN}{m\sqrt{\frac{m}{n}}}\right)\right) - 3\exp(-cn^{r/2}/\psi^{2r}),$$

where $C, c > 0$ are universal constants.

3.2.3 Vectors on a sphere

Another interesting case is when the vectors $X_1, \ldots, X_N$ lie on a common sphere. To keep the same normalization as in the previous cases we assume that the sphere has the radius $\sqrt{n}$. Then condition (3.4) becomes empty. Let $1 \leq r \leq 2$ and assume that the vectors are ψ_r and let $\psi = \max_{i \leq N} \|X_i\|_{\psi_r}$. Let $K \geq 1$ and set $\xi = \psi K$. Then Theorem 3.2 immediately gives that

$$\delta_m\left(\frac{A}{\sqrt{n}}\right) \leq C\xi^2\sqrt{\frac{m}{n}}\log^{1/r}\left(\frac{eN}{m\sqrt{\frac{m}{n}}}\right) \tag{3.7}$$

with probability larger than

$$1 - \exp\left(-cK^r\sqrt{m}\log\left(\frac{eN}{m\sqrt{\frac{m}{n}}}\right)\right)$$

where $C, c > 0$ are universal constants.

4 The geometry of faces of random polytopes

In this Section we discuss the geometry of random polytopes. Let A be an $n \times N$ matrix. We denote by $K^+(A)$ (resp. $K(A)$) the convex hull (resp., the symmetric convex hull) of the N columns of A.

4.1 Neighborly polytopes

For an integer $1 \le m \le n$, a polytope is called *m-neighborly* if any set of less than m vertices is the vertex set of a face. In the symmetric setting, a centrally symmetric convex polytope is m-centrally-neighborly if any set of less than m vertices containing no-opposite pairs is the vertex set of a face. We refer the reader to the books [15] and [31] for classical details on neighborly polytopes. (Some new quantitative invariants related to neighborliness were recently developed in [22].)

The relation between the problem of reconstruction and neighborly polytopes was discovered in [10].

Theorem 4.1. *([10], Theorem 1) Let $1 \le m \le n \le N$ and A be an $n \times N$ matrix. The following two assertions are equivalent.*

i) *The polytope $K(A)$ has $2N$ vertices and is m-centrally-neighborly.*

ii) *Whenever $y = Az$ has a solution z having at most m non-zero coordinates, z is the unique solution of the optimization problem (P):*

$$(P) \qquad \min \|t\|_{\ell_1}, \quad At = Az,$$

We will also use the following result from [5] (which could be replaced by a similar result from [8]).

Lemma 4.2. *[5] Assume that $\delta_{2m}(A/\sqrt{n}) < \sqrt{2}-1$. Then whenever $y = Az$ has a solution z having at most m non-zero coordinates, z is the unique solution of the ℓ_1 minimization problem (P).*

We are now ready to state the main result on neighborly random polytopes.

Theorem 4.3. *Let $1 \leq m \leq n \leq N$ be integers. Let $1 \leq r \leq 2$. Let $\psi \geq 1$ and $\lambda \in (0, 1/2)$. Let $X_1, \ldots, X_N$ be independent random vectors satisfying $H_1(r, \psi)$ and $H_2(\lambda)$. Let A be the $n \times N$ matrix with $X_1, \ldots, X_N$ as columns. Then, with probability larger than*

$$1 - 2\lambda - \exp(-c\sqrt{n}/\psi^2)$$

the polytopes $K^+(A)$ and $K(A)$ are m-neighborly and m-centrally-neighborly, respectively, whenever

$$m \leq cn/\psi^4 \log^{2/r}(C\psi^6 N/n),$$

where $C, c > 0$ are universal constants.

Observe that the probability is positive for n large enough provided that $\lambda < 1/2$.

Proof. Theorem 3.2 and the definition of property $H_1(r, \psi)$ imply that for arbitrary $\theta' \in (0,1)$, and $K, K' \geq 1$, setting $\xi = \psi K + K'$, the estimate

$$\delta_m\left(\frac{A}{\sqrt{n}}\right) \leq C\xi^2 \sqrt{\frac{m}{n}} \log^{1/r}\left(\frac{eN}{m\sqrt{\frac{m}{n}}}\right) + \theta' \tag{4.1}$$

holds with probability larger than

$$\begin{aligned} 1 \quad - \quad & \exp\left(-cK^r\sqrt{m}\log\left(\frac{eN}{m\sqrt{\frac{m}{n}}}\right)\right) \\ - \quad & \mathbb{P}\left(\max_{i\leq N}|X_i| \geq K'\sqrt{n}\right) - \mathbb{P}\left(\max_{i\leq N}\left|\frac{|X_i|^2}{n} - 1\right| \geq \theta'\right). \end{aligned} \tag{4.2}$$

In view of Lemma 4.2, we look for m and θ' to ensure $\delta_{2m}(A/\sqrt{n}) < \sqrt{2}-1$. For instance, we let $\theta' = (\sqrt{2}-1)/2$ and note that (3.4) implies

$$\mathbb{P}\left(\max_{i\leq N}|X_i| \geq \left(\frac{\sqrt{2}-1}{2} + 1\right)^{1/2} \sqrt{n}\right) < \lambda. \tag{4.3}$$

So we take $K' = \left(\frac{\sqrt{2}-1}{2} + 1\right)^{1/2}$ and $K = 1$ which determines $\xi = \psi K + K'$ in terms of ψ. We shall use the fact that $1 \le \xi/\psi \le \tilde{C}$, where $\tilde{C}$ is a universal constant.

Now set $m_0 = [c'n/\psi^4 \log^{2/r}(C'\psi^6 N/n)]$ (for some new constants $C', c' > 0$). Adjusting the constants $C', c' > 0$ and writing m for m_0, we obtain

$$C\xi^2 \sqrt{\frac{m}{n}} \log^{1/r}\left(\frac{eN}{m\sqrt{\frac{m}{n}}}\right) < (\sqrt{2}-1)/2.$$

Combining this with the choice of θ', passing from m to $2m$ and adjusting the constants again if necessary, we conclude by (4.1) that $\delta_{2m}\left(\frac{A}{\sqrt{n}}\right) < \sqrt{2}-1$ with probability larger than

$$1 - \exp\left(-c''\sqrt{m}\log\left(\frac{eN}{m\sqrt{\frac{m}{n}}}\right)\right) - 2\lambda.$$

The last estimate follows from (4.2) by applying (3.4) and (4.3) to the last two terms, respectively; and where $c'' > 0$ is again a new positive constant. Then, by Lemma 4.2 and Theorem 4.1, with the same probability, $K(A)$ is m-centrally-neighborly and has $2N$ vertices.

In general, $K(A)$ being m-centrally-neighborly does not imply $K^+(A)$ being m-neighborly. However this is true when $K(A)$ has $2N$ vertices. Indeed, this means that no two columns of A form a pair of opposite vertices. So by m-central-neighbourliness, every set of m column vectors is a face of $K(A)$. Consequently there is a supporting hyperplane so that all other vertices of $K(A)$ are on one side. In particular this plane is a supporting hyperplane for $K^+(A)$, which is our witness that $K^+(A)$ is m-neighbourly. □

4.2 Examples

We will now apply Theorem 4.3 in the three different settings introduced in the previous section.

4.2.1 The Log-Concave Ensemble

Applying Theorem 4.3 and Lemma 3.3 we obtain the following theorem for independent isotropic vectors with log-concave densities (this is for instance the case if $X_1, \ldots, X_N$ are i.i.d. random vectors uniformly distributed on an isotropic convex body).

Theorem 4.4. *Let $1 \leq m \leq n \leq N$ be integers. Let $X_1, \ldots, X_N$ be independent isotropic vectors with log-concave densities. Then, for any $N \leq \exp(cn^{c_1/2})$, with probability at least $1 - C\exp(-cn^{c_1/2})$, the polytopes $K^+(A)$ and $K(A)$ are m-neighborly and m-centrally-neighborly, respectively, whenever*

$$m \leq cn \big/ \log^2(CN/n),$$

where $C, c > 0$ are universal constants and c_1 is given in Lemma 2.6.

Remark 1. We believe that the estimate of the degree of neighborliness in Theorem 4.4 is not optimal and we conjecture that the $\log^2$ may be replaced by log as it is in the Gaussian case (see Remark 1 following Theorem 4.6).

Remark 2. It is known ([2]) that there is a universal constant ψ such that the uniform probability measure on the ball $\{x \in \mathbb{R}^n : \sum_{i=1}^n |x_i|^r \leq 1\}$ satisfies $H_1(r, \psi)$ for $1 \leq r \leq 2$ and satisfies $H(2, \psi)$ for $r \geq 2$. Of course, since it is log-concave, the concentration property H_2 is also satisfied. Applying Theorem 4.3 to these examples, we get a better estimate of the level of neighborliness than in Theorem 4.4. We get now $m \sim cn \big/ \log^{2/r}(CN/n)$ for $1 \leq r \leq 2$ and $m \sim cn \big/ \log(CN/n)$ for $2 \leq r \leq \infty$.

4.2.2 Matrices with independent ψ_r entries

In a similar way as above, Theorem 4.3 and Lemma 3.4 imply the following theorem (note that its conclusion becomes empty if $N \geq \exp(cn^{r/2}/\psi^{2r})$ and that $\psi \geq 1$, since variances are 1).

Theorem 4.5. *Let A be a matrix with entries that are independent centered variance one random variables. Let $1 \leq r \leq 2$ and assume that the ψ_r norms of the entries are bounded by some constant ψ. Then, for any $N \leq \exp(cn^{r/2}/\psi^{2r})$, with probability at least $1 - C\exp(-c\sqrt{n}/\psi^2)$, the polytopes $K^+(A)$ and $K(A)$ are m-neighborly and m-centrally-neighborly, respectively, whenever $1 \leq m \leq n$ satisfies*

$$m \leq cn \big/ \psi^4 \log^{2/r}(C\psi^6 N/n),$$

where $C, c > 0$ are universal constants.

4.2.3 Vectors on a sphere

Finally assume that the vectors are on a sphere of radius $\sqrt{n}$. Using bound (3.7) and repeating the proof of Theorem 4.3 with obvious modifications we obtain:

Theorem 4.6. *Let $1 \leq m \leq n \leq N$ be integers. Let $1 \leq r \leq 2$ and $\psi \geq 1$. Let $X_1, \ldots, X_N$ be independent vectors on a sphere of radius $\sqrt{n}$ and satisfying $H_1(r, \psi)$. Then, with probability at least $1 - \exp(-c\sqrt{n}/\psi^2)$, the polytopes $K^+(A)$ and $K(A)$ are m-neighborly and m-centrally-neighborly, respectively, whenever*

$$m \leq cn/\psi^4 \log^{2/r}(C\psi^6 N/n),$$

where $C, c > 0$ are universal constants.

Remark 1. For the matrix A with i.i.d. Gaussian $N(0,1)$ entries (the case considered in Section 3.2.2 above when $r = 2$), it is known that with overwhelming probability, $K(A)$ is m-centrally-neighborly, whenever $1 \leq m \leq n$ satisfies

$$m \leq cn/\log(CN/n),$$

where $C, c > 0$ are universal constants, (see [8], [11],[21],[24],[27]). The precise asymptotic dependence of m on n and N has been well studied in [12] when $n/N \to \delta \in (0,1)$ and in [13] when $n/N \to 0$.

Remark 2. The restricted isometry property was proved in [24] for matrices with independent rows (rather than columns), under a sub-gaussian hypothesis. It is worth noting that the corresponding result for matrices with independent isotropic sub-gaussian columns is not true in general. One can see it by considering the matrix with columns $X_i = \sqrt{2}\delta_i(\varepsilon_{1i}, \ldots, \varepsilon_{ni})$, where δ_i are independent random variables, $\mathbb{P}(\delta_i = 1) = \mathbb{P}(\delta_i = 0) = 1/2$ and ε_{ji} are independent Bernoulli variables, independent of δ_i's. The vectors X_i are then isotropic and sub-gaussian, but $\mathbb{P}(X_i = 0) = 1/2$. As a consequence, the concentration hypothesis and thus the restricted isometry property are not satisfied.

5 Main technical result

In this Section, $X_1, \ldots, X_N \in \mathbb{R}^n$ are independent ψ_r random vectors for some (fixed) $0 < r \leq 2$. Let $1 \leq m \leq N$. We shall consider three quantities

A_m, B_m and C_m depending on $X_1, \ldots, X_N$. Recall that B_m has been defined in (3.1) as

$$B_m = \sup_{z \in U_m} \left| \left| \sum_{i \le N} z_i X_i \right|^2 - \sum_{i \le N} z_i^2 |X_i|^2 \right|^{1/2}$$

and define the other two quantities as follows:

$$A_m = \sup_{z \in U_m} \left| \sum_{i \le N} z_i X_i \right|, \quad C_m = \max_{i \le N} |X_i|.$$

We clearly have

$$|A_m^2 - B_m^2| \le C_m^2.$$

Given a real number s, we will denote $\max(s, 0)$ by s_+.

The main purpose of this Section is to prove Theorem 3.1. In fact we will prove a stronger technical result, Theorem 5.1, from which Theorem 3.1 will follow.

Theorem 5.1. *Let $0 < r \le 2$. Let $n \ge 1$ and $1 \le m \le N$ be integers. Let $X_1, \ldots, X_N \in \mathbb{R}^n$ be independent ψ_r vectors with $\psi = \max_{i \le N} \|X_i\|_{\psi_r}$. For every $1 \le m \le N$, $\theta \in (0, 1/4)$, and $K \ge 1$ one has*

$$\begin{aligned} &\mathbb{P}\left(B_m^2 \ge \max\{B^2, C_m B, 24\,\theta\, C_m^2\}\right) \\ &\le \ (1 + 3\log m) \exp\left(-2K^r m^{(1+s)/2} \log \frac{2N}{\theta m}\right), \end{aligned} \tag{5.1}$$

with $s = (1 - r)_+$ and

$$B = C_0^{1/r} \psi K \ m^{q - 1/2} \left(\log \frac{2N}{\theta m}\right)^{1/r},$$

where C_0 is an absolute constant and $q = \max\{1, 1/r\}$.

Remark. In fact we shall prove a stronger statement: with the notation of Theorem 5.1, for every $1 \le m \le N$, $\theta \in (0, 1/4)$, and $K \ge 1$, and for every $0 \le \ell \le \log_2 m$, one has

$$\begin{aligned} &\mathbb{P}\left(B_m^2 \ge \max\{\overline{B}^2, C_m \overline{B}, 24\,\theta\, C_m^2\}\right) \\ &\le \ (1 + 2\ell) \exp\left(-2K^r \frac{m}{2^\ell} \log \frac{12eN2^\ell}{\theta m}\right), \end{aligned} \tag{5.2}$$

where

$$\overline{B} = C^{1/r}\psi K\left(\left(\frac{m}{2^\ell}\right)^q\left(\log\frac{2N2^\ell}{\theta m}\right)^{1/r} + m^{q-1/2}\left(\log\frac{2N}{\theta m}\right)^{1/r}\right),$$

C is an absolute constant and $q = \max\{1, 1/r\}$.

Before starting the proof of the theorem we show how it implies Theorem 3.1, stated in Section 3.

5.1 Proof of Theorem 3.1

Fix $K_1 \geq 1$ and let $K \geq K_1$ be such that

$$K^2 m \log^{2/r}\frac{2N}{\theta m} = K_1^2\theta^2 n.$$

By Theorem 5.1 with $r \geq 1$, and the condition on m,

$$\begin{aligned}
&\mathbb{P}\left(B_m^2 \geq \max\{B^2, C_m B, 24\theta C_m^2\}\right) \\
&\leq\ (1 + 3\log m)\exp\left(-2K^r\sqrt{m}\log\frac{2N}{\theta m}\right) \\
&\leq\ \exp\left(-K_1^r\sqrt{m}\log\frac{2N}{\theta m}\right),
\end{aligned}$$

where

$$B = C_0\psi K\sqrt{m}\log^{1/r}\frac{2N}{\theta m} = C_0\psi K_1\theta\sqrt{n},$$

and c is an absolute positive constant. Thus, if $C_m \leq K_2\sqrt{n}$ for some K_2, then

$$\begin{aligned}
\max\{B^2, C_m B, 24\theta C_m^2\} &\leq C_1\theta n\max\{\psi^2 K_1^2, \psi K_1 K_2, K_2^2\} \\
&\leq C_1\theta n\left(\psi K_1 + K_2\right)^2,
\end{aligned}$$

where C_1 is an absolute constant. This concludes the proof. $\square$

5.2 Proof of Theorem 5.1

We will prove the theorem in a stronger form (5.2). Then (5.1) follows by choosing $0 \le \ell \le \log_2 m$ to be the largest integer satisfying

$$\frac{1}{2^{q\ell}} \left(\log \frac{2N2^{\ell}}{\theta m} \right)^{1/r} \ge m^{-1/2} \left(\log \frac{2N}{\theta m} \right)^{1/r} .$$

The proof will use the same construction as in [1], which however requires some modifications. For completeness and the reader's convenience we provide details of the argument.

We require the following two lemmas proved in [1] with $r = 1$. Since the proofs for general r repeat the same arguments, we leave them for the reader. Recall that for every $E \subset \{1, ..., N\}$ of cardinality m, every $\varepsilon, \alpha \in (0, 1]$ we selected an ε-net (in the Euclidean metric) in $B_2^N \cap \alpha B_\infty^N \cap \mathbb{R}^E$ of cardinality not exceeding $(3/\varepsilon)^m$ and denoted it by $\mathcal{N}(E, \varepsilon, \alpha)$.

Lemma 5.2. *Let $0 < r \le 2$ and $X_1, \dots, X_N \in \mathbb{R}^n$ be independent ψ_r vectors with $\psi = \max_{i \le N} \|X_i\|_{\psi_r}$. Let $m \le N$, $\varepsilon, \alpha \in (0, 1]$. Let $q = \max\{1, 1/r\}$ and $L \ge m^q \left(2 \log \frac{12eN}{m\varepsilon}\right)^{1/r}$. Then*

$$\mathbb{P} \left(\sup_{\substack{F \subset \{1,\dots,N\} \\ |F| \le m}} \sup_{E \subset F} \sup_{z \in \mathcal{N}(F,\varepsilon,\alpha)} \sum_{i \in E} \left| \left\langle z_i X_i, \sum_{j \in F \setminus E} z_j X_j \right\rangle \right| \ge \psi\, \alpha L A_m \right)$$

$$\le \exp \left(-\frac{1}{2}\, L^r\, m^{-(r-1)_+} \right) .$$

Lemma 5.3. *Let $0 < r \le 2$ and $X_1, \dots, X_N \in \mathbb{R}^n$ be independent ψ_r vectors with $\psi = \max_{i \le N} \|X_i\|_{\psi_r}$. 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*

$$\mathbb{P} \left(\sup_{\substack{F \subset \{1,\dots,N\} \\ |F| \le k}} \sup_{x \in \mathcal{B}} \sup_{z \in \mathcal{N}(F,\varepsilon,\alpha)} \sum_{i \in F} \left| \left\langle z_i X_i, \sum_{j \notin F} x_j X_j \right\rangle \right| \ge \psi \alpha \beta L A_m \right)$$

$$\le \; M \left(\frac{6eN}{k\varepsilon} \right)^k \exp \left(-\frac{1}{2}\, L^r\, k^{-(r-1)_+} \right) .$$

The following formula is well known and the proof is in its statement.

Lemma 5.4. *Let $x_1, \ldots, x_N \in \mathbb{R}^n$, then*

$$\sum_{i \neq j} \langle x_i, x_j \rangle = 4 \cdot 2^{-N} \sum_{E \subset \{1,\ldots,N\}} \sum_{i \in E} \sum_{j \in E^c} \langle x_i, x_j \rangle .$$

We are now ready to start the proof of Theorem 5.1.

Proof of Theorem 5.1. As in [1], the construction splits into two cases.

If $\ell = 0$ we set

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

Otherwise, define positive integers $a_0, a_1, \ldots, a_\ell$ by $a_k := [m\, 2^{-k+1}] - [m\, 2^{-k}]$ for $1 \le k \le \ell$ and $a_0 := [m\, 2^{-\ell}]$. Observe that $a_k \le m\, 2^{-k+1}$ for $1 \le k \le \ell$, $a_0 \le m\, 2^{-\ell}$ and $\sum_{k=0}^{\ell} a_k = m$. Recall that for $E \subset \{1, \ldots, N\}$ we identify $\mathbb{R}^E$ with the subspace of vectors in $\mathbb{R}^N$ with coordinates supported by E.

We consider $(\ell + 1)$-tuples $((E_0, x_0), \ldots, (E_\ell, x_\ell))$ where $(E_k)_{0 \le k \le \ell}$ are mutually disjoint subsets of $\{1, \ldots N\}$, $|E_k| \le a_k$, $x_k \in \mathbb{R}^{E_k}$ for all $0 \le k \le \ell$. A $(\ell + 1)$-tuple $((E_0, x_0), \ldots, (E_\ell, x_\ell))$ is said to be admissible if

$$x_k \in \mathcal{N}\left(E_k, \theta 2^{-k}, \sqrt{\frac{2^k}{m}}\right) \text{ for } 1 \le k \le \ell, x_0 \in \mathcal{N}\left(E, \theta/4, 1\right), \left|\sum_{k=0}^{\ell} x_k\right| \le 2.$$

The set of all vectors $x = \sum_{k=0}^{\ell} x_k$ associated to admissible $(\ell+1)$-tuples $((E_0, x_0), \ldots, (E_\ell, x_\ell))$ will be denoted by $\mathcal{M}(\theta)$.

We shall consider the details of the case $\ell > 0$, the other case can be treated similarly.

Fix $((F_0, x_0), \ldots, (F_\ell, x_\ell))$ to be admissible and let $x = \sum_{k=0}^{\ell} x_k \in \mathcal{M}(\theta)$. Denote the coordinates of x by $x(i)$, $i \le N$, then

$$|Ax|^2 = \left\langle \sum_{i \le N} x(i) X_i, \sum_{i \le N} x(i) X_i \right\rangle = \sum_{i \le N} x(i)^2 |X_i|^2 + \sum_{i \neq j} \langle x(i) X_i, x(j) X_j \rangle .$$

So

$$\left| |Ax|^2 - \sum_{i \le N} x(i)^2 |X_i|^2 \right| = |D_x|$$

where

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

Now we split D_x according to the structure of x. Namely we let

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

and

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

so that we have

$$\left| |Ax|^2 - \sum_{i\le N} x(i)^2|X_i|^2 \right| = |D'_x + D''_x| \le |D'_x| + |D''_x|\,.$$

We first estimate D'_x. By Lemma 5.4 we have

$$\begin{aligned} D'_x &= \sum_{k=0}^{\ell} \sum_{\substack{i,j\in F_k\\ i\neq j}} \langle x(i)X_i, x(j)X_j\rangle \\ &= 4\sum_{k=0}^{\ell} 2^{-|F_k|} \sum_{E\subset F_k} \sum_{i\in E} \sum_{j\in F_k\setminus E} \langle x(i)X_i, x(j)X_j\rangle\,. \end{aligned}$$

Thus

$$\begin{aligned} |D'_x| &\le 4\sum_{k=0}^{\ell} 2^{-|F_k|} \sum_{E\subset F_k} \left| \sum_{i\in E} \sum_{j\in F_k\setminus E} \langle x(i)X_i, x(j)X_j\rangle \right| \\ &\le 4\sum_{k=0}^{\ell} \sup_{E\subset F_k} \left| \sum_{i\in E} \sum_{j\in F_k\setminus E} \langle x(i)X_i, x(j)X_j\rangle \right| \end{aligned}$$

and using the fact that $|F_k| \le a_k$ for $0\le k\le \ell$, we arrive at

$$|D'_x| \le 4\sum_{k=0}^{\ell} \sup_{\substack{F\subset\{1,\dots,N\}\\ |F|\le a_k}} \sup_{E\subset F} \sum_{i\in E} \left| \left\langle x(i)X_i, \sum_{j\in F\setminus E} x(j)X_j \right\rangle \right|.$$

We now set $q = \max\{1, 1/r\}$ and apply Lemma 5.2 to each summand in the sum above with the parameters

$$a_0, \varepsilon = \theta/4, \alpha = 1, \text{ and } L = K\left(\frac{m}{2^\ell}\right)^q \left(4\log\frac{48eN2^\ell}{\theta m}\right)^{1/r}$$

for $k=0$ and

$$a_k, \varepsilon = \theta 2^{-k}, \alpha = \sqrt{\frac{2^k}{m}}, \text{ and } L = K\left(\frac{m}{2^k}\right)^q \left(4\log\frac{12eN4^k}{\theta m}\right)^{1/r}$$

for $1 \le k \le \ell$ (note also that m from the lemma is substituted with $m/2^\ell$ and $m/2^k$ respectively). By the union bound we obtain that the probability of the event

$$\begin{aligned} \sup_{x\in\mathcal{M}(\theta)} |D'_x| \quad &\geq \quad \psi A_m K \left(\left(\frac{m}{2^\ell}\right)^q \left(4\log\frac{48eN2^\ell}{\theta m}\right)^{1/r} \right. \\ &+ \left. \sum_{k=1}^{\ell} \left(\frac{m}{2^k}\right)^{q-1/2} \left(4\log\frac{12eN4^k}{\theta m}\right)^{1/r} \right) \end{aligned}$$

is not larger than

$$\exp\left(-K^r\,\frac{2m}{2^\ell}\log\frac{48eN2^\ell}{\theta m}\right) + \sum_{k=1}^{\ell} \exp\left(-K^r\,\frac{2m}{2^k}\log\frac{12eN4^k}{\theta m}\right).$$

Therefore the probability of the event

$$\begin{aligned} \sup_{x\in\mathcal{M}(\theta)} |D'_x| \quad &\geq \quad \psi A_m K \left(\left(\frac{m}{2^\ell}\right)^q \left(4\log\frac{48eN2^\ell}{\theta m}\right)^{1/r} \right. \\ &+ \left. C_1^{1/r} m^{q-1/2} \left(\log\frac{2N}{\theta m}\right)^{1/r} \right) \end{aligned}$$

is not larger than

$$(1+\ell)\exp\left(-K^r\,\frac{2m}{2^\ell}\log\frac{12eN2^\ell}{\theta m}\right),$$

where C_1 is an absolute constant.

We now pass to the estimate for D''_x which essentially follows the same lines.

For every $1 \le k \le \ell$ we consider $\mathcal{M}_k(\theta) = \mathcal{M}'_k(\theta) \cap 2B_2^N$, where $\mathcal{M}'_k(\theta)$ consists of all vectors of the form $v = v_0 + \sum_{s=k+1}^{\ell} v_s$, where v_i's ($i = 0, k = 1, \ldots, \ell$) have pairwise disjoint supports and

$$v_0 \in \bigcup_{\substack{E\subset\{1,\ldots N\}\\ |E|\le a_0}} \mathcal{N}(E, \theta/4, 1),\;\; v_s \in \bigcup_{\substack{E\subset\{1,\ldots N\}\\ |E|\le a_s}} \mathcal{N}\left(E, \theta\, 2^{-s}, \sqrt{\frac{2^s}{m}}\right) \;\text{ for }\; s \ge k+1.$$

Then $\mathcal{M}_k(\theta) \subset 2B_2^N$ and (similarly as in [1]) we can estimate the cardinality

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

Recalling that $x = \sum_{k=0}^{\ell} x_k \in \mathcal{M}(\theta)$ for some admissible $(\ell+1)$-tuple $((F_0, x_0), \dots, (F_\ell, x_\ell))$ and setting $G_k = \{0, k+1, k+2, \dots, \ell\}$, we observe that

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

Now we apply Lemma 5.3 to each summand $k = 1, \dots, \ell$, with parameters

$$
\varepsilon = \theta 2^{-k}, \alpha = \sqrt{\frac{2^k}{m}}, \beta = 2, \mathcal{B} = \mathcal{M}_k(\theta) \text{ and } L = K\left(\frac{m}{2^k}\right)^q \left(12 \log \frac{12eN4^k}{\theta m}\right)^{1/r}.
$$

Using the union bound we obtain

$$
\begin{aligned}
&\mathbb{P}\left(|D_x''| \ge 2\psi A_m K \sum_{k=1}^{\ell} \left(\frac{m}{2^k}\right)^{q-1/2} \left(12\log\frac{12eN4^k}{\theta m}\right)^{1/r}\right) \\
&\le \sum_{k=1}^{\ell} \exp\left(\frac{4m}{2^k}\log\frac{12e4^kN}{\theta m} + \frac{2m}{2^k}\log\frac{3e4^kN}{\theta m} - K^r\,\frac{12m}{2^k}\log\frac{12eN4^k}{\theta m}\right) \\
&\le \sum_{k=1}^{\ell} \exp\left(-K^r\,\frac{6m}{2^k}\log\frac{12eN4^k}{\theta m}\right) \le \ell \exp\left(-K^r\,\frac{6m}{2^\ell}\log\frac{12eN4^\ell}{\theta m}\right).
\end{aligned}
$$

Thus

$$\mathbb{P}\left(\sup_{x\in\mathcal{M}(\theta)}|D''_x|\geq C_2^{1/r}\psi A_m K m^{q-1/2}\left(\log\frac{2N}{\theta m}\right)^{1/r}\right)$$
$$\leq \ \ell\exp\left(-K^r\ \frac{6m}{2^\ell}\log\frac{12eN4^\ell}{\theta m}\right),$$

where C_2 is the an absolute constant.

Since $D_x = D'_x + D''_x$, then

$$\mathbb{P}\left(\sup_{x\in\mathcal{M}(\theta)}|D_x|\geq A_m\gamma\right)\leq(1+2\ell)\exp\left(-K^r\ \frac{2m}{2^\ell}\log\frac{12eN2^\ell}{\theta m}\right), \qquad (5.3)$$

where

$$\gamma = C_3^{1/r}\psi K\left(\left(\frac{m}{2^\ell}\right)^q\left(\log\frac{2N2^\ell}{\theta m}\right)^{1/r}+m^{q-1/2}\left(\log\frac{2N}{\theta m}\right)^{1/r}\right)$$

for some absolute constant $C_3 > 0$.

Passing now to the approximation argument, pick an arbitrary $z\in S^{N-1}$ with $|\operatorname{supp} z|\leq m$. 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 $|z(n_1)|\geq|z(n_2)|\geq\ldots\geq|z(n_N)|$, so that $z(n_i)=0$ for $i>m$ (since $|\operatorname{supp} z|\leq m$). If $\ell=0$, we denote the support of z by $\widetilde{E}_0$ and consider only this $\widetilde{E}_0$. Otherwise we set

$$\widetilde{E}_0=\{n_i\}_{1\leq i\leq m/2^\ell}$$

and

$$\widetilde{E}_1=\{n_i\}_{m/2<i\leq m},\ \widetilde{E}_2=\{n_i\}_{m/4<i\leq m/2},\ \dots,\ \widetilde{E}_\ell=\{n_i\}_{m/2^\ell<i\leq m/2^{\ell-1}}.$$

Recall that integers $a_0,a_1,\dots,a_\ell$ have been defined at the beginning of this proof. Then, clearly,

$$a_0=|\widetilde{E}_0|\leq m/2^\ell,\quad a_k=|\widetilde{E}_k|\leq m/2^k+1\leq m/2^{k-1}\ \text{ for every } 1\leq k\leq\ell,$$

and $\sum_{i=0}^{\ell}a_i=m$. Also observe that, since $z\in S^{N-1}$, then for every $k\geq 1$,

$$\|P_{\widetilde{E}_k}z\|_\infty\leq|z(n_s)|\leq\sqrt{\frac{2^k}{m}},$$

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

For every $k \geq 1$ the vector $P_{\widetilde{E}_k} z$ can be approximated by a vector from $\mathcal{N}\left(\widetilde{E}_k, \theta 2^{-k}, \sqrt{\frac{2^k}{m}}\right)$ and the vector $P_{\widetilde{E}_0} z$ can be approximated by a vector from $\mathcal{N}(\widetilde{E}_0, \theta/4, 1)$. Thus there exists $x \in \mathcal{M}(\theta)$, with a suitable representation $x = \sum_{k=0}^{\ell} x_k$, such that

$$|z - x|^2 \leq \sum_{k=0}^{\ell} |P_{\widetilde{E}_k} z - x_k|^2 \leq \theta^2 (2^{-4} + \sum_{k=1}^{\ell} 2^{-2k}) < \theta^2 \, (0.4).$$

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

It follows from the definitions of D_z and A that

$$D_z = D_x + \langle Aw, Ax \rangle + \langle Az, Aw \rangle - \sum_{i \leq N} w(i) \, (x(i) + z(i)) \, |X_i|^2,$$

(here $w(i)$, $x(i)$ and $z(i)$ denote the coordinates of w, x and z, respectively). Thus

$$|D_z| \leq |D_x| + |Aw|(|Ax| + |Az|) + |w| \, |x + z| \, \max_{i \leq N} |X_i|^2.$$

It follows that

$$B_m^2 = \sup_{\substack{z \in S^{N-1} \\ |\operatorname{supp} z| \leq m}} |D_z| \leq \sup_{x \in \mathcal{M}(\theta)} |D_x| + 2\theta \left(A_m^2 + C_m^2\right) \leq \sup_{x \in \mathcal{M}(\theta)} |D_x| + 2\theta \left(B_m^2 + 2C_m^2\right).$$

Thus, by (5.3) and using again $A_m \leq \sqrt{B_m^2 + C_m^2} \leq B_m + C_m$ we obtain

$$\mathbb{P}\left((1 - 2\theta) B_m^2 \geq 4\theta C_m^2 + C_m \gamma + B_m \gamma\right) \leq (1 + 2\ell) \exp\left(-K^r \, \frac{2m}{2^\ell} \log \frac{12eN2^\ell}{\theta m}\right).$$

Since $\theta \leq 1/4$, this implies

$$\mathbb{P}\left(B_m^2 \geq \max\{24\theta C_m^2, 6C_m\gamma, 6\gamma^2\}\right) \leq (1 + 2\ell) \exp\left(-K^r \, \frac{2m}{2^\ell} \log \frac{12eN2^\ell}{\theta m}\right),$$

which completes the proof. $\square$

5.3 Optimality of estimates

We conclude this section by an example showing optimality, in a certain sense, of estimates in Theorem 3.1. We will limit ourselves to the ψ_1 case, that is to $r = 1$. To this end we consider a special case when $X_i = (X_{ij})_{j=1}^n$ where X_{ij} are i.i.d. symmetric exponential variables with variance one. We begin by showing an optimal estimate for A_m.

First, from [1] (Theorem 3.5) we have that for $N \le \exp(c\sqrt{n})$ and any $K \ge 1$,

$$\mathbb{P}\left(A_m \ge CK\left(\sqrt{n} + \sqrt{m}\log\frac{2N}{m}\right)\right) \le \exp\left(-cK\sqrt{n}\right) \qquad (5.4)$$

where $C, c > 0$ are numerical constants. In the other direction, we have the following

Proposition 5.5. *For any $1 \le m \le N$ and $t \ge 1$,*

$$\mathbb{P}\Big(A_m \ge c\Big(\sqrt{n} + \sqrt{m}\log\left(\frac{2N}{m}\right) + t\Big)\Big) \ge c \wedge e^{-t},$$

where $c > 0$ is an absolute constant.

Before we prove this proposition let us explain its relevance to Theorem 3.1. Firstly, observe that the proposition implies that with probability bounded away from zero, $A_m \ge c(\sqrt{n} + \sqrt{m}\log(2N/m))$. This shows that – except for allowing a change of absolute constants – one cannot obtain a better bound on A_m than (5.4), valid with overwhelming probability (i.e., with probability converging to one as $n \to \infty$). Secondly, assume that $N \le \exp(c\sqrt{n})$. By taking $t = cK\sqrt{n}$, we obtain that for large n, $\mathbb{P}(A_m \ge cK\sqrt{n}) \ge \exp(-cK\sqrt{n})$. We compare this with estimates for probabilities in (5.4). Namely, using Lemma 2.7 (noting that the density of X_i's is log-concave), we can see that for $m\log^2(2N/m) \le n$, the theorem implies that $\mathbb{P}(A_m \ge CK\sqrt{n}) \le \exp(-\tilde{c}K\sqrt{n})$. So in this range of m the upper and lower bounds on probability coincide up to numerical constants in the exponent.

Regarding Theorem 3.1, again assume that $N \le \exp(c\sqrt{n})$. Using again Lemma 2.7, we get with overwhelming probability that for all i, $|X_i| \le C'\sqrt{n}$. Now assume that for some m we have with overwhelming probability $B_m^2 \le Cn$. Then by the obvious bound $A_m^2 \le B_m^2 + \sup_{z\in U_m}\sum_{i\le N}|z_i|^2|X_i|^2$, with probability close to one we also have $A_m \le C''\sqrt{n}$. On the other hand, as noted above, $\mathbb{P}(A_m \geq c(\sqrt{n} + \sqrt{m}\log(2N/m)))$ is bounded away from zero. Thus, $c(\sqrt{n} + \sqrt{m}\log(2N/m)) \leq C''\sqrt{n}$, which in turn implies that for n large enough we have $m\log^2(2N/m) \leq Cn$. This shows that the factor $\log^2(2N/\theta m)$ in Theorem 3.1 is of the right order.

Proof of Proposition 5.5 Since

$$A_m = \sup_{\substack{\alpha\in S^{N-1}\\ |\operatorname{supp}\alpha|\leq m}} \sup_{\beta\in S^{n-1}} \sum_{ij} \alpha_i\beta_j X_{ij},$$

by general tail estimates for linear combinations of exponential variables with vector valued coefficients (see e.g. Corollary 1 in [20]), we get

$$\mathbb{P}\Big(A_m \geq c\big(\mathbb{E}A_m + \sqrt{t}\sigma + tb\big)\Big) \geq c \wedge e^{-t},$$

where

$$\sigma^2 = \sup_{\substack{\alpha\in S^{N-1}\\ |\operatorname{supp}\alpha|\leq m}} \sup_{\beta\in S^{n-1}} \sum_{ij} \alpha_i^2\beta_j^2 = 1$$

and

$$b = \sup_{\substack{\alpha\in S^{N-1}\\ |\operatorname{supp}\alpha|\leq m}} \sup_{\beta\in S^{n-1}} \max_{ij} |\alpha_i\beta_j| = 1.$$

Therefore, it is enough to show that $\mathbb{E}A_m \geq c(\sqrt{n} + \sqrt{m}\log(2N/m))$. Obviously, $\mathbb{E}A_m \geq c\sqrt{n}$, since a single column of the matrix A has expected Euclidean norm of the order $\sqrt{n}$. As for the other term, it is enough to consider the first row of our matrix. We have

$$\sqrt{m}A_m \geq \sup_{\substack{\alpha\in\{0,-1,+1\}^N\\ |\operatorname{supp}\alpha|=m}} \sum_{i=1}^{N} \alpha_i Y_i,$$

where to simplify the notation we set $Y_i = X_{i1}$. On the right hand side we actually have $\sum_{i=1}^m |Y_i^*|$, where Y_i^* is such a rearrangement of Y_i that $|Y_1^*| \geq |Y_2^*| \geq \ldots \geq |Y_n^*|$, which can be used to derive lower bounds on the expectation. We will however not rely on this representation, instead we will use a Sudakov type minoration principle for exponential variables proved in [28], Theorem 5.2.9, which we state here in a simplified version, adapted to our purposes.

Lemma 5.6. *Let $Y_1, \ldots, Y_N$ be independent symmetric exponential variables with variance one. Consider $T \subseteq \ell_2^N$ of cardinality k and $u \geq 1$. If for any $s, t \in T$, $t \neq s$,*

$$\sqrt{u}|t-s| + u\|t-s\|_\infty > u,$$

then $\mathbb{E}\max_{t \in T} \sum_{i=1}^N t_i Y_i \geq c\min(u, \log k)$, where $c > 0$ is an absolute constant.

In our case, $T = \{\alpha \in \{0,-1,1\}^N \colon |\operatorname{supp}\alpha| \leq m\}$, so $k \geq \binom{N}{m}$. Also, since $\|t-s\|_\infty \geq 1$ for $t, s \in T$, $t \neq s$, the condition of the lemma is trivially satisfied for any $u \geq 1$, in particular for $u = \log k$. Thus, for $m \leq N/2$, we obtain $\sqrt{m}\,\mathbb{E}A_m \geq \log k \geq cm\log(2N/m)$. On the other hand we have $\mathbb{E}A_m \geq c\sqrt{m}$, so for $m \geq N/2$ it is enough to adjust the constants. □

Let M be an $n \times N$ matrix with entries $X_{ij}/\sqrt{n}$, $i = 1, \ldots n$, $j = 1, \ldots N$, where X_{ij} are i.i.d. symmetric exponential variables with variance one. It follows from the definition 1.1 of the isometry constant of M that for any $1 \leq m \leq \min(n, N)$

$$\frac{A_m^2}{n} \leq 1 + \delta_m.$$

Consequently, if M satisfies the restricted isometry property of order m with some parameter $\delta \in (0,1)$, then $A_m^2 \leq 2n$. From Proposition 5.5 we deduce the optimality of our result regarding the restricted isometry property.

Proposition 5.7. *Let $1 \leq m \leq n \leq N$. Let M be a $n \times N$ matrix with entries $X_{ij}/\sqrt{n}$, $i = 1, \ldots n$, $j = 1, \ldots N$, where X_{ij} are i.i.d. symmetric exponential variables with variance one. There exist constants $C > 0$ and $0 < c < 1$ such that if the probability that the random matrix M satisfies the restricted isometry property of order m with some parameter $\delta \in (0,1)$ is larger than c, then*

$$m\log^2\left(\frac{2N}{m}\right) \leq Cn.$$

References

[1] R. Adamczak, A. E. Litvak, A. Pajor and N. Tomczak-Jaegermann, Quantitative estimates of the convergence of the empirical covariance matrix in log-concave Ensembles, *Journal of AMS*, 234 (2010), 535–561.

[2] F. Barthe, O. Guédon, S. Mendelson, A. Naor, A Probabilistic Approach to the Geometry of the lpn ball, *Annals of Probability*, 33 (2005), 480–513.

[3] R. Baraniuk, M. Davenport, R. DeVore, M. Wakin, A Simple Proof of the Restricted Isometry Property for Random Matrices, *Constructive Approximation*, 28 (2008), 253–263.

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

[5] E. Candes, The restricted isometry property and its implications for compressed sensing, *C. R. Acad. Sci. Paris*, Ser. I 346 (2008), 589–592.

[6] E. Candes, J. Romberg and T. Tao, Stable signal recovery from incomplete and incurable measurements, *Comm. Pure App. Math.* 59 (8), (2006), 1207–1223.

[7] E. Candes and T. Tao, Near-optimal signal recovery from random projections: universal encoding strategies. *IEEE Trans. Inform. Theory*, 52 (2006), 5406–5425.

[8] E. Candes and T. Tao, Decoding by Linear Programming. *IEEE Trans. Inform. Theory* 51 (12) (December 2005), 4203–4215.

[9] A. Cohen, W. Dahmen and R. Devore, Compressed sensing and k-term approximation, *J. Amer. Math. Soc.* 22 (2009), 211–231.

[10] D. L. Donoho, Neighborly Polytopes and Sparse solutions of underdetermined linear equations, Department of Statistics, Stanford University (2005).

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

[12] D. L. Donoho, High-dimensional centrally-symmetric polytopes with neighborliness proportional to dimension, *Disc. Comput. Geometry* 35 (2006), 617–652.

[13] D. L. Donoho and J. Tanner, Counting faces of randomly projected polytopes when the projection radically lowers dimension, *J. Amer. Math. Soc.* 22 (2009), 1–53.

[14] B. Fleury, O. Guédon and G. Paouris, A stability result for mean width of L_p-centroid bodies, *Adv. Math.* 214 (2007), 865–877.

[15] B. Grünbaum, Convex polytopes, volume 221 of Graduate Texts in Mathematics. Springer-Verlag, New York, second edition, 2003.

[16] P. Hitczenko, S. J. Montgomery-Smith, K. Oleszkiewicz, Moment inequalities for sums of certain independent symmetric random variables, *Studia Math.* 123 (1997), 15–42.

[17] B. S. Kashin and V. N. Temlyakov, A remark on compressed sensing, *Mathematical Notes* 82 (2007), 748–755.

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

[19] B. Klartag, A central limit theorem for convex sets, *Invent. Math.* 168 (2007), 91–131.

[20] R. Latała, Tail and moment estimates for sums of independent random vectors with logarithmically concave tails, *Studia Math.* 118 (1996), 301–304.

[21] N. Linial, I. Novik, How neighborly can a centrally symmetric polytope be? *Discrete Comput. Geom.* 36 (2006), 273-281.

[22] P. Mankiewicz and N. Tomczak-Jaegermann, Stability properties of neighbourly random polytopes, *Discrete and Computational Geometry* 41 (2009), 257–272.

[23] S. Mendelson, A. Pajor, N. Tomczak-Jaegermann, Reconstruction and sub-gaussian processes, *C. R. Acad. Sci. Paris* 340 (2005), 885–888.

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

[25] S. Mendelson, A. Pajor, N. Tomczak-Jaegermann, Uniform uncertainty principle for Bernoulli and sub-gaussian ensembles, *Constructive Approximation* 28 (2008), 277–289.

[26] G. Paouris, Concentration of mass on convex bodies. *Geom. Funct. Anal.* 16 (2006), 1021–1049.

[27] M. Rudelson, R. Vershynin, Geometric approach to error correcting codes and reconstruction of signals, *Int. Math. Res. Not.* 64 (2005), 4019–4041.

[28] M. Talagrand, The generic chaining. Upper and lower bounds of stochastic processes. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2005.

[29] M. Talagrand, The supremum of some canonical processes, *Amer. J. Math.* 116 (1994), 283-325.

[30] 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.

[31] G. Ziegler, Lectures on polytopes, volume 152, Graduate Texts in Mathematics. Springer-Verlag, New-York, 1995.

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

Alexander E. Litvak,
Dept. of Math. and Stat. Sciences,
University of Alberta,
Edmonton, Alberta, Canada, T6G 2G1.
e-mail: `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`