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

R. Adamczak  A. E. Litvak  A. Pajor  N. Tomczak-Jaegermann

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 (RIP) 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 this property RIP with overwhelming probability. We show that such “sensing” matrices are valid for the exact reconstruction process of \( m \)-sparse vectors via \( \ell_1 \) 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 \( m \sim n/\log^2(cN/n) \).

AMS Classification:
primary 52A20, 94A12, 52B12, 46B09
secondary 15A52, 41A45, 94B75

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

Let $1 \leq m \leq n \leq N$ be integers and let $X_1, \ldots, X_N \in \mathbb{R}^n$. Denote by $A$ the $n \times N$ matrix with $X_1, \ldots, X_N$ as columns and by $K(A) = K(X_1, \ldots, X_N)$ the convex hull of $\pm X_1, \ldots, \pm X_N$. 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 [16] and [31]).

The connection between the neighborliness of $K(A)$ and sparse solutions of underdetermined linear equations was discovered in [11], 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:

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

Here the $\ell_1$-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 $\ell_1$ 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 [12].

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 $(\langle X_1, x \rangle, \ldots, \langle X_N, x \rangle)$. 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

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

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 Restricted Isometry Property (RIP) defined by the following parameter.

**Definition.** Let $M$ be a $n \times N$ matrix. For any $1 \leq m \leq \min(n, N)$, 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 $\delta$, shortly $\text{RIP}_m(\delta)$, if $0 \leq \delta_m(M) < \delta$.

The relevance of this parameter for the reconstruction property $\text{ii)}$ 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 $\text{ii)}$ (see also [6], [9], [18]). 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 $\text{i)}$ and $\text{ii)}$ are satisfied. In other words, if $M$ has $\text{RIP}_{2m}(\sqrt{2} - 1)$ then $M$ has the reconstruction property $\text{ii)}$. This approach gives the strategy of our paper.

Recall that no general construction of centrally symmetric polytopes is known to produce 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 $m$-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 Random Matrix Theory. Due to the normalization, we shall consider the isometry constant of $A/\sqrt{n}$. The plan consists in specializing 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 $1 \leq i \leq N$, all $y \in S^{n-1}$, and $t > 0$,

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

where $C, c > 0$. 

3
The Euclidean norms of $X_1, \ldots, X_N$ are concentrated around their average:

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

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

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

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

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 $\psi_1$ random variables), contains a subclass of matrices with i.i.d. $\psi_r$ entries for $1 < r \leq 2$ (see Definition 2.1 below of $\psi_r$ 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 $1 \leq r \leq 2$.

Sub-gaussian matrices with independent $\psi_2$ 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 [12], [8], [27]); the case when the entries of $A$ are i.i.d. $(\pm 1)$ 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 3.2 will employ methods from [1]. A crucial new ingredient consists of an analysis of the quantity

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

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 3.2 to estimate the isometry constant (Theorem 3.3). 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 \( \psi_r \) random coordinates, and polytopes generated by independent \( \psi_r \) random vectors on a sphere. Section 5 is devoted to the proof of Theorem 3.2 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 \( \psi_r \)-norm by

\[ \|Y\|_{\psi_r} = \inf \{ C > 0 ; \mathbb{E} \exp \left( |Y|/C \right)^r \leq 2 \} . \]

It is well known that the \( \psi_r \)-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 \( p \geq 1, \|Y\|_p \leq p^{1/r} K \), for some \( K > 0 \), then \( \|Y\|_{\psi_r} \leq cK \) where \( c > 0 \) is a numerical constant.

Definition 2.2. Let \( X \in \mathbb{R}^n \) be a centered random vector and \( r > 0 \). We say that \( X \) is \( \psi_r \) or a \( \psi_r \) 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 $\|X\|_{\psi}$ for the *weak* $\psi_r$ norm of a random vector $X$ should not be confused with the standard convention in the probability theory that this notation stands for the $\psi_r$ 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 $\psi_1$ estimate ([30]).

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

$$
\mathbb{P}\left(\left|\sum_{i=1}^n Y_i\right| > t\right) \leq 2 \exp\left(-\frac{t^2}{4\sum_{i=1}^n \|Y_i\|_{\psi_1}^2 + 2t\psi}\right).
$$

Given a set $E \subset \{1, \ldots, N\}$ 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 supp $z$. A vector $z \in \mathbb{R}^N$ is called $m$-sparse if $|\text{supp } z| \leq m$. The subset of $m$-sparse unit vectors in $\mathbb{R}^N$ is denoted by

$$
U_m = U_m(\mathbb{R}^N) := \{z \in \mathbb{R}^N : |z| = 1, |\text{supp } z| \leq m\}. \tag{2.1}
$$

Let $B_\infty^N = \{x = (x_i) \in \mathbb{R}^N : \|x\|_\infty = \max_i |x_i| \leq 1\}$ and $B_2^N$ be the unit Euclidean ball. For every $E \subset \{1, \ldots, N\}$, $\varepsilon, \alpha \in (0, 1]$ by $\mathcal{N}(E, \varepsilon, \alpha)$ we denote an $\varepsilon$-net in $B_2^N \cap \alpha B_\infty^N \cap \mathbb{R}^E$ in the Euclidean metric. 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

$$
\mathbb{E}\langle X, y \rangle = 0, \quad \mathbb{E}|\langle X, y \rangle|^2 = |y|^2 \quad \text{for all } y \in \mathbb{R}^n, \tag{2.2}
$$

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 $x_1, x_2 \in \mathbb{R}^n$,

$$
f(\theta x_1 + (1-\theta)x_2) \geq f(x_1)^{\theta} f(x_2)^{1-\theta}.
$$
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 $y \in S^{n-1},$

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

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

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 ([20], see also [15]). We will use here the following result immediately derived from [19], Theorem 4.4.

**Lemma 2.6.** Let $1 \leq n \leq N$ 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 $N \leq \exp(c \theta^{c_0} n^{c_1}),$

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

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 $N, n \geq 1$ 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 $N \leq \exp(\sqrt{n})$ and for every $K \geq 1$ one has

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

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

In this paper, different universal positive constants may be denoted by the same letters $C, C_0, C', c, c_0, c'_1,$ etc.
3 Isometry constant

We begin this section by formulating, in Theorem 3.3, a general estimate for the isometry constant of random matrices with independent \( \psi_r \) 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 3.3 and thus provide us with models: the Log-Concave Ensemble, matrices with i.i.d. \( \psi_r \) entries, and matrices defined by independent \( \psi_r \) vectors on a sphere.

3.1 Estimating the isometry constant

Techniques of “compressed sensing” rely on properties of the sampling matrix, which should act almost isometrically on sparse vectors. This motivated the concept of Restricted Isometry Property (RIP) defined in [8]. To quantify this property of the “sensing” matrix, the authors introduced the isometry constant defined in the introduction, that we recall here for the convenience of the reader.

Definition 3.1. Let \( M \) be a \( n \times N \) matrices and let \( \delta \in (0,1) \). For any \( 1 \leq m \leq \min(n,N) \), 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 \( \delta \) if \( 0 \leq \delta_m(M) < \delta \).

Let \( X_1, \ldots, X_N \in \mathbb{R}^n \) and let \( A = A^{(m,N)} \) be the “sampling” matrix with the \( X_i \)'s as columns. We begin by a simple observation. Define the following quantity

\[
B_m = \sup_{z \in U_m} \left| \sum_{i \leq N} z_i X_i \right|^2 - \left| \sum_{i \leq N} z_i^2 X_i \right|^2 \right|^{1/2} \tag{3.2}
\]

Then, clearly

\[
\sup_{z \in U_m} \left| \frac{Az}{n} \right| - 1 \leq \frac{B_m^2}{n} + \max_{i \leq N} \left| \frac{X_i}{n} - 1 \right|. \tag{3.3}
\]
Thus the isometry constant is controlled by quantity $B_m$ and the second term, $\max_{i \leq N} \left| \frac{|X_i|^2}{n} - 1 \right|$. We begin by estimating $B_m$.

**Theorem 3.2.** Let $n \geq 1$ and $1 \leq m \leq N$ be integers. Let $1 \leq r \leq 2$ and $X_1, \ldots, X_N \in \mathbb{R}^n$ be independent $\psi_r$ random vectors with $\psi = \max_{i \leq N} \|X_i\|_{\psi_r}$. Let $\theta \in (0, 1/4)$, $K, K' \geq 1$ and assume that $m$ satisfies

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

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

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

holds with probability at least

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

where $C, c$ are absolute positive constants.

We postpone the proof of Theorem 3.2 to the last section. Combining this theorem with inequality (3.3), relating the RIP, $B_m$ and concentration of the Euclidean norm of the $X_i$’s, we immediately deduce an estimate for the isometry constant of a random matrix with independent $\psi_r$ columns.

**Theorem 3.3.** 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 $\psi_r$ 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

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

holds with probability larger than

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

where $C, c > 0$ are universal constants.
Note that such an estimate for $\delta_m \left( \frac{A}{\sqrt{n}} \right)$ is meaningful only if, firstly, its right hand side is $< 1$, and secondly, if it holds with probability $> 0$. In fact, the former condition is equivalent to the RIP 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 $\psi_r$ 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 $\psi_r$ estimate:
  $$\|\langle X_i, y \rangle\|_{\psi_r} \leq \psi \quad \text{for all } y \in S^{n-1} \text{ and } 1 \leq i \leq N.$$  (3.4)

- **Condition $H_2(\lambda)$**: $|X_i|$’s are concentrated around their average:
  $$\mathbb{P}\left( \max_{1 \leq i \leq N} \left| \frac{|X_i|^2}{n} - 1 \right| \geq \frac{\sqrt{2} - 1}{2} \right) < \lambda.$$  (3.5)

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

### 3.2 Examples

We now specialize Theorem 3.3 to some specific classes of matrices.

#### 3.2.1 The Log-Concave Ensemble

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

**Lemma 3.4.** Assume the above “log-concave setting”. There exist universal constants $\psi, C, c > 0$ such that conditions $H_1(1, \psi)$ and $H_2(C \exp(-cn^c))$ are satisfied whenever $N \leq \exp(cn^c)$, where $c_1$ is given in Lemma 2.6.

The proof is immediate from Lemmas 2.3 and 2.6.

Applying Theorem 3.3 (with $r = 1$) together with Lemmas 3.4 and 2.7 to the Log-Concave Ensemble, we get that for every $N \leq \exp(cn^c)$,

$$\delta_m \left( \frac{A}{\sqrt{n}} \right) \leq C \sqrt{\frac{m}{n}} \log \left( \frac{eN}{m^{2/3}} \right) + \frac{\sqrt{2} - 1}{2}$$  (3.6)
holds with probability larger than

$$1 - C \exp \left( -c \sqrt{m} \log \left( \frac{eN}{m \sqrt{m/n}} \right) \right) - e^{-c\sqrt{m}} - \exp(-cn^{c_1}),$$

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 (3.6) 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 $\psi_r$ entries

Consider now the “$\psi_r$ setting”, where the entries $a_{ij}$ of the matrix $A$ are independent centered, with variance one, random $\psi_r$ variables (with $r \in [1, 2]$). Set $\psi = \max_{ij} \|a_{ij}\|_{\psi_r}$.

**Lemma 3.5.** Assume the above “$\psi_r$ 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 $\psi_r$ 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 3.6.** Let $r \in [1, 2]$ and $Y_1, \ldots, Y_n$ be independent symmetric random variables satisfying $\mathbb{P}(|Y_i| \geq t) = \exp(-t^r)$. Then for every vector $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ and every $t \geq 0$,

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

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

11
The behavior of general centered $\psi_r$ 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 $Z_1, \ldots, Z_n$ 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 3.6.

We have for $t > 0$,

$$P(U_i \geq t) \leq P(|Z_i| \geq t + \beta) \leq 2 \exp(-(t + \beta)^r) \leq 2 \exp(-t^r - \beta^r) = P(|Y_i| \geq t).$$

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 $p \geq 1$,

$$\left\| \sum_{i=1}^n a_i Z_i \right\|_p \leq 2 \left\| \sum_{i=1}^n a_i \varepsilon_i Z_i \right\|_p \quad \text{(symmetrization)}$$

$$= 2 \left\| \sum_{i=1}^n a_i \varepsilon_i |Z_i| \right\|_p$$

$$\leq 2 \left\| \sum_{i=1}^n a_i \varepsilon_i (\beta + U_i) \right\|_p \quad \text{(the contraction principle)}$$

$$\leq 2 \left\| \sum_{i=1}^n a_i \varepsilon_i \beta \right\|_p + 2 \left\| \sum_{i=1}^n a_i \varepsilon_i U_i \right\|_p$$

$$\leq C \sqrt{p} \beta \|a\|_2 + 2 \left\| \sum_{i=1}^n a_i \varepsilon_i Y_i \right\|_p$$

$$\leq C \sqrt{p} \|a\|_2 + Cp^{1/r} \|a\|_{r^*},$$

where to get the last two inequalities we used Khinchine’s inequality, Lemma 3.6 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 $\|a_{ij}\|_{\psi_r} \leq \psi$, we obtain by homogeneity

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

12
since \( r \in [1, 2] \) implies that \( p^{1/r} \geq \sqrt{p} \) and \( \|a\|_{\psi^r} \leq \|a\|_2 = 1 \). The growth condition on the moments of the random variable \( \sum_{i=1}^n y_i a_{ij} \) implies that its \( \psi_r \) norm is bounded by \( \tilde{C} \psi \).

The proof of condition \( H_2 \) goes along similar lines. Instead of Lemma 3.6 we will now use the following lemma, which is an easy consequence of Theorem 6.2 in [17] 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 3.7.** 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 \( p \geq 2 \),

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

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 \( p \geq 2 \),

\[
\left\| \sum_{i=1}^n a_i Z_i \right\|_p \leq C b (\sqrt{p} \|a\|_2 + p^{1/s} \|a\|_p).
\]

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

\[
\mathbb{P}\left( \left| \sum_{i=1}^n Z_i \right| \geq C b (\sqrt{np} + p^{1/s} n^{1/p}) \right) \leq \exp(-p).
\]

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

\[
\mathbb{P}\left( \left| \sum_{i=1}^n Z_i \right| \geq C b (\sqrt{np} + p^{1/s}) \right) \leq e^3 \exp(-p)
\]

for some (new) universal constant \( C \) or equivalently

\[
\mathbb{P}\left( \left| \sum_{i=1}^n Z_i \right| \geq t \right) \leq 2 \exp \left( -c \min \left[ \frac{t^2}{nb^2}, \left( \frac{t}{b} \right)^s \right] \right). \tag{3.7}
\]
For fixed $j$ we apply this inequality with $s = r/2$ to variables $Z_i = a^2_{ij} - 1$. Note that $\mathbb{E}Z_i = 0$ and

$$\|Z_i\|_{\psi_r/2} \leq C(1 + \|a^2_{ij}\|_{\psi_r/2}) = C(1 + \|a_{ij}\|_{\psi_r}^2) \leq \tilde{C}\psi^2.$$

(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 $x \mapsto e^{xr} - 1$ so that it is convex. For $r$ away from zero, this modification changes the norm by an absolute constant). Therefore, applying (3.7) with $t = \varepsilon n$ yields

$$\mathbb{P}\left( \left| \frac{1}{n} \sum_{i=1}^{n} a^2_{ij} - 1 \right| \geq \varepsilon \right) \leq 2 \exp\left( -c \min\left[ \frac{\varepsilon^2 n}{\psi^4}, \left( \frac{\varepsilon n}{\psi^2} \right)^{r/2} \right]\right) \leq 2 \exp\left( -\frac{\varepsilon r n^{r/2}}{\psi^{2r}} \right).$$

For $r = 2$ the proof is similar, but uses Lemma 3.6 (which in this case reduces to Bernstein’s $\psi_1$ inequality) instead of Lemma 3.7 (the argument is simpler since in this case the involved norms of the vector $a$ do not depend on $p$ and we get (3.7) directly).

The lemma follows now by the union bound.

Applying Theorem 3.3 together with Lemma 3.5 to the “$\psi_r$ setting”, we get that for every $N \leq \exp(cn^{r/2}/\psi^{2r})$,

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

holds with probability at least

$$1 - C \exp( -cn^{r/2}/\psi^{2r})$$

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 (3.5) becomes empty. Let \( 1 \leq r \leq 2 \) and assume that the vectors are \( \psi_r \) and let \( \psi = \max_{i \leq N} \|X_i\|_{\psi_r} \).

Let \( K \geq 1 \) and set \( \xi = \psi K \). Then Theorem 3.3 immediately gives that

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

with probability larger than

\[
1 - C \exp \left( -cK^r \sqrt{m} \log \left( \frac{eN}{m \sqrt{m/n}} \right) \right)
\]

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 \( 1 \leq m \leq n \), 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 [16] 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 [11].

**Theorem 4.1.** ([11], Theorem 1) Let \( A \) be a \( n \times N \) matrix, \( n \leq N \). 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 \((P)\):

\[
(P) \quad \min \|t\|_{\ell_1}, \quad At = Az,
\]
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 $\ell_1$ 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 $1 \leq r \leq 2$ and $\lambda \in (0, 1/2)$. Let $X_1, \ldots, X_N$ be independent random vectors satisfying $H_1(r, \psi)$ with parameter $\psi$ and $H_2(\lambda)$ with probability $\lambda$. Let $A$ be the $n \times N$ matrix with $X_1, \ldots, X_N$ as columns. Then, with probability larger than

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

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 3.3 and the definition of property $H_1(r, \psi)$ imply that for arbitrary $\theta' \in (0, 1)$, and $K, K' \geq 1$, setting $\xi = \psi K + K'$, we have

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

holds with probability larger than

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

$$- \mathbb{P}\left( \max_{i \leq N} |X_i| \geq K'\sqrt{n} \right) - \mathbb{P}\left( \max_{i \leq N} \left| \frac{|X_i|^2}{n} - 1 \right| \geq \theta' \right). \quad (4.1)$$
In view of Lemma 4.2, we look for \( m \) and \( \theta' \) to ensure \( \delta_{2m}(A/\sqrt{n}) < \sqrt{2} - 1. \) For instance, we let \( \theta' = (\sqrt{2} - 1)/2 \) and note that (3.5) implies

\[
P\left( \max_{i \leq N} |X_i| \geq \left( \frac{\sqrt{2} - 1}{2} + 1 \right)^{1/2} \sqrt{n} \right) < \lambda. \tag{4.2}
\]

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 \( \psi. \) We shall use the fact that \( 1 \leq \xi/\psi \leq \tilde{C}, \) where \( \tilde{C} \) is a universal constant.

Now set \( m_0 = \left[ c'n/\psi^4 \log^{2/r}(C''\psi^6 N/n) \right] \) (for some new constants \( C'', c' > 0 \)). It is clearly sufficient to prove that the polytopes \( K^+(A) \) and \( K(A) \) are \( m_0 \)-neighborly and \( m_0 \)-centrally-neighborly, respectively. Thus adjusting the constants \( C'', c' > 0 \) and writing \( m \) for \( m_0 \), we obtain

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

Combining this with the choice of \( \theta' \), passing from \( m \) to \( 2m \) and adjusting the constants again if necessary, we conclude that \( \delta_m \left( \frac{4}{\sqrt{n}} \right) < \sqrt{2} - 1 \) with probability larger than

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

The last estimate follows from (4.1) by applying (3.5) and (4.2) to the last two terms, respectively; and where \( C'', c'' > 0 \) are again new constants. \( \square \)

### 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 Lemma 3.4 and bound (3.6) we get the following:
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. 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. 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

\[
m \leq cn/ \log^2(CN/n),
\]

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

Remark: It is known ([2]) that there is a universal constant \( \psi \) 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 \( m \sim cn/ \log^{2/r}(CN/n) \) for \( 1 \leq r \leq 2 \) and \( m \sim cn/ \log(CN/n) \) for \( 2 \leq r \leq \infty \).

4.2.2 Matrices with independent \( \psi_r \) entries

In a similar way as above, Lemma 3.3 and bound (3.8) imply the following theorem (note that its conclusion becomes empty if \( N \geq \exp(cn^{c_2}/\psi^{2r}) \) and \( \psi \geq 1 \)).

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 \( \psi_r \) norms of the entries are bounded by some constant \( \psi \). Then, for any \( N \leq \exp(cn^{r/2}/\psi^{2r}) \), with probability at least \( 1 - C \exp(-cn^{r/2}/\psi^{2r}) \), the polytopes \( K^+(A) \) and \( K(A) \) are \( m \)-neighborly and \( m \)-centrally-neighborly, respectively, whenever \( 1 \leq m \leq n \) satisfies

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

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 \( \sqrt{n} \). From bound (3.9) we obtain:
Theorem 4.6. Let $1 \leq m \leq n \leq N$ be integers. Let $1 \leq r \leq 2$. Let $X_1, \ldots, X_N$ be independent vectors on a sphere of radius $\sqrt{n}$ and satisfying $H_1(r, \psi)$ for some parameter $\psi > 0$. Let $K \geq 1$ and set $\xi = \psi K$. Then, with probability at least $1 - C \exp\left(-K \sqrt{n}/\psi^2\right)$, the polytopes $K^+(A)$ and $K(A)$ are $m$-neighborly and $m$-centrally-neighborly, respectively, whenever

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

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 [12],[8],[27],[24]). The precise asymptotic dependence of $m$ on $n$ and $N$ has been well studied in [13] when $n/N \rightarrow \delta \in (0, 1)$ and in [14] when $n/N \rightarrow 0$.

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 $X_i = \sqrt{2 \delta_i (\varepsilon_{1i}, \ldots, \varepsilon_{mi})}$, where $\delta_i$ are independent random variables, $\mathbb{P}(\delta_i = 1) = \mathbb{P}(\delta_i = 0) = 1/2$ and $\varepsilon_{ji}$ are independent Bernoulli variables, independent of $\delta_i$’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 RIP property are not satisfied.

5 Main technical result

In this Section, $X_1, \ldots, X_N \in \mathbb{R}^n$ are independent $\psi_r$ 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 (3.2) as

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


and define the other two quantities as follows:

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

We clearly have

\[
|A_m^2 - B_m^2| \leq C_m^2.
\]

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

The main purpose of this Section is to prove Theorem 3.2. In fact we will prove a stronger technical result, Theorem 5.1, from which Theorem 3.2 will follow.

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

\[
P \left( B_m^2 \geq \max \{ B^2, C_m B, 24 \theta C_m^2 \} \right) \leq (1 + 3 \log m) \exp \left( -2K^r m^{(1+s)/2} \log \frac{2N}{\theta m} \right),
\]

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

\[\begin{align*}
B &= C_{0}^{1/r} \psi K \left( \log \frac{2N}{\theta m} \right)^{1/r},
\end{align*}\]

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 \( 1 \leq m \leq N \), \( \theta \in (0, 1/4) \), and \( K \geq 1 \), and for every \( 0 \leq \ell \leq \log_2 m \), one has

\[
P \left( B_m^2 \geq \max \{ \mathcal{B}^2, C_m \mathcal{B}, 24 \theta C_m^2 \} \right) \leq (1 + 2 \ell) \exp \left( -2K^p m \frac{\ell \log 12 eN^{2^\ell}}{\theta m} \right),
\]

where

\[
\mathcal{B} = C_{0}^{1/r} \psi K \left( \frac{m}{2^\ell} \right)^q \left( \log \frac{2N^{2^\ell}}{\theta m} \right)^{1/r} + m^{q-1/2} \left( \log \frac{2N}{\theta m} \right)^{1/r},
\]
$C$ is an absolute constant and $q = \max\{1, 1/r\}$.

Before starting the proof of the theorem we show how it implies Theorem 3.2 stated in Section 3.

### 5.1 Proof of Theorem 3.2

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 $m$,

$$P\left( B_m^2 \geq \max\{B^2, C_m B, 24\theta C_m^2\} \right) \leq (1 + 3 \log m) \exp \left( -2K^r \sqrt{m \log \frac{2N}{\theta m}} \right) \leq \exp \left( -cK_1^r \sqrt{m \log \frac{2N}{\theta m}} \right),$$

where

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

and $c$ and $C_0$ are absolute positive constants. Thus, if $C_m \leq K_2 \sqrt{n}$ for some $K_2$, then

$$\max\{B^2, C_m B, 24\theta C_m^2\} \leq C_1 \theta n \max\{\psi^2 K_1^2, \psi K_1 K_2, K_2^2\} \leq C_1 \theta n (\psi K_1 + K_2)^2,$$

where $C_1$ is an absolute constant. This concludes the proof. \[\Box\]

### 5.2 Proof of Theorem 5.1

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

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

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

$$\mathbb{P} \left( \sup_{F \subseteq \{1, \ldots, N\}} \sup_{|F| \leq m} \mathbb{E} \left[ \sum_{i \in E} \left| \left\langle z_i X_i, \sum_{j \in F \setminus E} z_j X_j \right\rangle \right| \geq \psi \alpha LA_m \right] \right) \leq \exp \left( -\frac{1}{2} L^r m^{-(r-1)+} \right).$$

**Lemma 5.3.** Let $0 < r \leq 2$ and $X_1, \ldots, X_N \in \mathbb{R}^n$ be independent $\psi_r$ vectors with $\psi = \max_{i \leq N} \|X_i\|_{\psi_r}$. Let $1 \leq k, m \leq N$, $\varepsilon, \alpha \in (0, 1]$, $\beta > 0$, and $L > 0$. Let $B(m, \beta)$ denote the set of vectors $x \in \beta B_2^N$ with $|\text{supp} x| \leq m$ and let $\mathcal{B}$ be a subset of $B(m, \beta)$ of cardinality $M$. Then

$$\mathbb{P} \left( \sup_{F \subseteq \{1, \ldots, N\}} \sup_{|F| \leq k} \mathbb{E} \left[ \sum_{x \in \mathcal{B}} \sum_{z \in N(F, \varepsilon, \alpha)} \left| \left\langle z_i X_i, \sum_{j \in F \setminus E} z_j X_j \right\rangle \right| \geq \psi \alpha \beta LA_m \right] \right) \leq M \left( \frac{6eN}{k\varepsilon} \right)^k \exp \left( -\frac{1}{2} L^r k^{-(r-1)+} \right).$$

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

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

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

$$\mathcal{M}(\theta) = \bigcup_{E \subseteq \{1, \ldots, N\}} \mathcal{N}(E, \theta/4, 1).$$

22
Otherwise, define positive integers \( a_0, a_1, \ldots, a_\ell \) by \( a_k := \lfloor m 2^{-k+1} \rfloor - \lfloor m 2^{-k} \rfloor \) for \( 1 \leq k \leq \ell \) and \( a_0 := \lfloor m 2^{-\ell} \rfloor \). Observe that \( a_k \leq m 2^{-k+1} \) for \( 1 \leq k \leq \ell \), \( a_0 \leq m 2^{-\ell} \) 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 \((E_k)_{0 \leq k \leq \ell}\) are mutually disjoint subsets of \( \{1, \ldots, N\} \), \(|E_k| \leq a_k\), \( x_k \in \mathbb{R}^{E_k} \) for all \( 0 \leq k \leq \ell \). A \((\ell + 1)\)-tuple \(((E_0, x_0), \ldots, (E_\ell, x_\ell))\) is said to be admissible if

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

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 \( x(i), i \leq N \), then

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

So

\[
\left| |Ax|^2 - \sum_{i \leq N} x(i)^2 |X_i|^2 \right| = |D_x| \quad (5.3)
\]

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

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

and

\[
D''_x := \sum_{k=0}^\ell \sum_{i \in F_k \atop j \notin F_k} \langle x(i)X_i, x(j)X_j \rangle.
\]
so that we have

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

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

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

Thus

$$
|D'_x| \leq 4 \sum_{k=0}^{\ell} 2^{-|F_k|} \sum_{E \subseteq F_k} \left| \sum_{i \in E} \sum_{j \in F_k \setminus E} \langle x(i)X_i, x(j)X_j \rangle \right|
$$

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

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

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

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

24
for $1 \leq k \leq \ell$. By the union bound we obtain that the probability of the event

$$ \sup_{x \in \mathcal{M}(\theta)} |D'_x| \geq \psi A_m K \left( \left( \frac{m}{2^\ell} \right)^q \left( 4 \log \frac{48eN2^\ell}{\theta m} \right)^{1/r} \right) $$

$$ + \sum_{k=1}^{\ell} \left( \frac{m}{2^k} \right)^{q-1/2} \left( 4 \log \frac{12eN4^k}{\theta m} \right)^{1/r}$$

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

$$ \sup_{x \in \mathcal{M}(\theta)} |D'_x| \geq \psi A_m K \left( \left( \frac{m}{2^\ell} \right)^q \left( 4 \log \frac{48eN2^\ell}{\theta m} \right)^{1/r} \right) $$

$$ + C_1^{1/r} m^{q-1/2} \left( \log \frac{2N}{\theta m} \right)^{1/r}$$

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 $1 \leq k \leq \ell$ we consider $\mathcal{M}_k(\theta) = \mathcal{M}'_k(\theta) \cap 2B_N^2$, 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

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

25
Then $\mathcal{M}_k(\theta) \subset 2B_2^N$ and (similarly as in \cite{1}) we can estimate the cardinality

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

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

$$|D''_x| = \left|2 \sum_{k=1}^\ell \sum_{i \in F_k} x(i)X_i, \sum_{r \in G_k} \sum_{j \in F_r} x(j)X_j\right| \leq 2 \sum_{k=1}^\ell \sup_{P \subset \{1, \ldots, N\}} \sum_{u \in P} \sum_{v \in \mathcal{M}_k} \left|u(i)X_i, \sum_{j \not\in F} v(j)X_j\right|.$$

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

$$\varepsilon = \theta 2^{-k}, \alpha = \sqrt{\frac{2^k}{m}}, \beta = 2, \mathcal{B} = \mathcal{M}_k(\theta) \text{ and } L = K \left(\frac{m}{2^k}\right)^q \left(12 \log \frac{12e4^k N}{\theta m}\right)^{1/r}.$$

Using the union bound we obtain

$$\mathbb{P}\left(|D''_x| \geq 2\psi A_m K \sum_{k=1}^\ell \left(\frac{m}{2^k}\right)^{q-1/2} \left(12 \log \frac{12e4^k N}{\theta m}\right)^{1/r}\right) \leq \sum_{k=1}^\ell \exp \left(\frac{4m}{2^k} \log \frac{12e4^k N}{\theta m} + \frac{2m}{2^k} \log \frac{3e4^k N}{\theta m} - K^r \frac{12m}{2^k} \log \frac{12e4^k N}{\theta m}\right) \leq \sum_{k=1}^\ell \exp \left(-K^r \frac{6m}{2^k} \log \frac{12e4^k N}{\theta m}\right) \leq K^r \frac{6m}{2^k} \log \frac{12e4^k N}{\theta m}. $$

Thus

$$\mathbb{P}\left(\sup_{x \in \mathcal{M}(\theta)} |D''_x| \geq (2)_2 \psi A_m K m^{q-1/2} \left(\log \frac{2N}{\theta m}\right)^{1/r}\right) \leq \ell \exp \left(-K^r \frac{6m}{2^k} \log \frac{12e4^k N}{\theta m}\right),$$

26
where $C_2$ is the an absolute constant.

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

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

where

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

for some absolute constant $C_3 > 0$.

Passing now to the approximation argument, pick an arbitrary $z \in S^{N-1}$ with $|\text{supp} \, z| \leq m$. Define the following subsets of $\{1, \ldots, N\}$ depending on $z$. Denote the coordinates of $z$ by $z(n_i)$ ($i = 1, \ldots, N$). Let $n_1, \ldots, n_N$ 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 $|\text{supp} \, z| \leq m$). If $\ell = 0$, we denote the support of $z$ by \( \tilde{E}_0 \) and consider only this \( \tilde{E}_0 \). Otherwise we set

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

and

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

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

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

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

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

where $s = \lfloor m/2^k \rfloor$.

For every $k \geq 1$ the vector $P_{\tilde{E}_k} z$ can be approximated by a vector from $\mathcal{N}(\tilde{E}_k, \theta 2^{-k}, 2_k^r \sqrt{\frac{2^k}{m}})$ and the vector $P_{\tilde{E}_0} z$ can be approximated by a vector

27
from $N(\widetilde{E}_0, \theta/4, 1)$. Thus there exists $x \in \mathcal{M}(\theta)$, with a suitable representation $x = \sum_{k=0}^{\ell} x_k$, such that

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

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

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

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

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

$$\mathbb{P}\left((1 - 2\theta)B_m^2 \geq 4\theta C_m^2 + C_m\gamma + B_m\gamma\right) \leq (1 + 2\ell) \exp\left(-K^r \frac{2}{2^\ell} \log \frac{12cN2^\ell}{\theta m}\right).$$

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

$$\mathbb{P}\left(B_m^2 \geq \max\{24\theta C_m^2, 6C_m\gamma, 6\gamma^2\}\right) \leq (1 + 2\ell) \exp\left(-K^r \frac{2}{2^\ell} \log \frac{12cN2^\ell}{\theta m}\right),$$

which completes the proof.

\[ \Box \]

### 5.3 Optimality of estimates

We conclude this section by an example showing optimality, in a certain sense, of estimates in Theorem 3.2. We will limit ourselves to the $\psi_1$ 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 \( N \leq \exp(c\sqrt{n}) \) and any \( K \geq 1 \),

\[
P\left( A_m \geq CK \left( \sqrt{n} + \sqrt{m \log \frac{2N}{m}} \right) \right) \leq \exp \left( -cK\sqrt{n} \right) \tag{5.5}
\]

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

**Proposition 5.5.** For any \( 1 \leq m \leq N \) and \( t \geq 1 \),

\[
P\left( A_m \geq c \left( \sqrt{n} + \sqrt{m \log \left( \frac{2N}{m} \right)} + t \right) \right) \geq c \land e^{-t},
\]

where \( c > 0 \) is an absolute constant.

Before we prove this proposition let us explain its relevance to Theorem 3.2. Firstly, observe that the proposition implies that with probability bounded away from zero, \( A_m \geq c \left( \sqrt{n} + \sqrt{m \log(2N/m)} \right) \). This shows that – except for allowing a change of absolute constants – one cannot obtain a better bound on \( A_m \) than (5.5), valid with overwhelming probability (i.e., with probability converging to one as \( n \to \infty \)). Secondly, assume that \( N \leq \exp(c\sqrt{n}) \). By taking \( t = cK\sqrt{n} \), we obtain that for large \( n \), \( P(A_m \geq cK\sqrt{n}) \geq \exp(-cK\sqrt{n}) \). We compare this with estimates for probabilities in (5.5). Namely, using Lemma 2.7 (noting that the density of \( X_i \)'s is log-concave), we can see that for \( m \log^2(2N/m) \leq n \), the theorem implies that \( P(A_m \geq cK\sqrt{n}) \leq \exp(-\tilde{c}K\sqrt{n}) \). So in this range of \( m \) the upper and lower bounds on probability coincide up to numerical constants in the exponent.

Regarding Theorem 3.2, again assume that \( N \leq \exp(c\sqrt{n}) \). Using again Lemma 2.7, we get with overwhelming probability that for all \( i \), \( |X_i| \leq C'\sqrt{n} \). Now assume that for some \( m \) we have with overwhelming probability \( B_m^2 \leq Cn \). Then by the obvious bound \( A_m^2 \leq B_m^2 + \sup_{z \in \mathcal{U}_m} \sum_{i \leq N} |z_i|^2 |X_i|^2 \), with probability close to one we also have \( A_m \leq C''\sqrt{n} \). On the other hand, as noted above, \( P(A_m \geq c(\sqrt{n} + \sqrt{m \log(2N/m)})) \) 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 3.2 is of the right order.
Proof of Proposition 5.5 Since
\[
A_m = \sup_{\alpha \in S^{n-1}} \sup_{|\text{supp } \alpha| \leq m} \sum_{ij} \alpha_i \beta_j X_{ij},
\]
by general tail estimates for linear combinations of exponential variables with vector valued coefficients (see e.g. Corollary 1 in [21]), we get
\[
P\left( A_m \geq c(\mathbb{E}A_m + \sqrt{t\sigma + tb}) \right) \geq c \wedge e^{-t},
\]
where
\[
\sigma^2 = \sup_{\alpha \in S^{n-1}} \sup_{|\text{supp } \alpha| \leq m} \sum_{ij} \alpha_i^2 \beta_j^2 = 1
\]
and
\[
b = \sup_{\alpha \in S^{n-1}} \sup_{|\text{supp } \alpha| \leq m} \max_{ij} |\alpha_i \beta_j| = 1.
\]
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
\[
\sqrt{m}A_m \geq \sup_{\alpha \in \{-1,0,1\}^N} \sum_{i=1}^N \alpha_i Y_i,
\]
where to simplify the notation we set \( Y_i = X_{1i} \). 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 \( |Y_1^*| \geq |Y_2^*| \geq \ldots \geq |Y_m^*| \), 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, \( T = \{ \alpha \in \{0, -1, 1\}^N : |\text{supp} \alpha| \leq m \} \), 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 \( \sqrt{m} E A_m \geq \log k \geq c m \log(2N/m) \). On the other hand we have \( E A_m \geq c \sqrt{m} \), so for \( m \geq N/2 \) it is enough to adjust the constants. \( \square \)

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, Preprint. Available at [http://arxiv.org/abs/0903.2323](http://arxiv.org/abs/0903.2323)

[2] F. Barthe, O. Guédon, S. Mendelson, A. Naor, A Probabilistic Approach to the Geometry of the \( l^p_n \) ball, Annals of Probability 33 (2005), no. 2, 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.

31
[10] D. L. Donoho, For most large systems of underdetermined equations, the minimum $\ell_1$-norm solution is the sparsest solution, Comm. Pure Appl. Math., 59 (2006), 907-934.

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

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

[13] D. L. Donoho, High-dimensional centrally-symmetric polytopes with neighborliness proportional to dimension, Disc. Comput. Geometry 35 (2006), no. 4, 617-652.

[14] 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.

[15] B. Fleury, O. Guédon and G. Paouris, A stability result for mean width of $L_p$-centroid bodies, Adv. Math., 214 (2007), no. 2, 865–877.

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

[17] P. Hitczenko, S. J. Montgomery-Smith, K. Oleszkiewicz, Moment inequalities for sums of certain independent symmetric random variables, Studia Math. 123 (1997), no. 1, 15–42.

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

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

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

[21] R. Latala, Tail and moment estimates for sums of independent random vectors with logarithmically concave tails, Studia Math. 118 (1996), 301–304.
[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, no. 5 (2006), 1021–1049.

[27] M. Rudelson, R. Vershynin, Geometric approach to error correcting codes and reconstruction of signals, Int. Math. Res. Not. 2005, no. 64, 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., Vol. 116, No. 2. (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.

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

33
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