Compressed sensing under weak moment assumptions

Guillaume Lecure 1,3 Shahar Mendelson 2,4,5

February 11, 2014

Abstract

We prove that iid random vectors that satisfy a rather weak moment assumption can be used as measurement vectors in Compressed Sensing. In many cases, the moment assumption suffices to ensure that the number of measurements required for exact reconstruction is the same as the best possible estimate – exhibited by a random gaussian matrix.

1 Introduction and main results

In Compressed Sensing (see, e.g., [4] and [7]), one observes linear measurements \( y_i = \langle X_i, x_0 \rangle, i = 1, ..., N \) of an unknown vector \( x_0 \in \mathbb{R}^n \), and the goal is to identify \( x_0 \) using those measurements.

Given the measurements matrix \( \Gamma = N^{-1/2} \sum_{i=1}^{N} \langle X_i \cdot \rangle e_i \), a possible recovery procedure is the basis pursuit algorithm, defined by

\[
\hat{x} \in \arg\min \{ \|t\|_1 : \Gamma t = \Gamma x_0 \}.
\]

A well known question is to identify conditions on the vectors \( X_1, ..., X_N \) that ensure that if \( x_0 \) is \( s \)-sparse, that is, if it is supported on at most \( s \) coordinates, the unique minimizer of the basis pursuit algorithm is \( x_0 \) itself. The matrix \( \Gamma \) satisfies the exact reconstruction property in \( \Sigma_s \), the set of all \( s \)-sparse vectors in \( \mathbb{R}^n \), if every \( x_0 \in \Sigma_s \) has this property.

---

1 CNRS, CMAP, Ecole Polytechnique, 91120 Palaiseau, France.
2 Department of Mathematics, Technion, I.I.T, Haifa 32000, Israel.
3 Email: guillaume.lecue@cmap.polytechnique.fr
4 Email: shahar@tx.technion.ac.il
5 Supported by the Mathematical Sciences Institute – The Australian National University and by ISF grant 900/10.
A standard choice of a measurements matrix \( \Gamma \) is when \( X_1, \ldots, X_N \) are independent, isotropic and \( L \)-subgaussian random vectors. Recall that a random vector \( X \) in \( \mathbb{R}^n \) is isotropic if for every \( t \in \mathbb{R}^n \), \( \mathbb{E} \langle X, t \rangle^2 = \| t \|_2^2 \), and it is \( L \)-subgaussian if for every \( t \in \mathbb{R}^n \) and every \( p \ge 2 \), \( \| \langle X, t \rangle \|_{L_p} \le L \sqrt{p} \| \langle X, t \rangle \|_{L_2} \).

One may show that if the \( X_i \)'s are random vectors that are independent, isotropic and \( L \)-subgaussian, then with high probability \( \Gamma \) satisfies the exact reconstruction property for \( s \)-sparse vectors as long as \( N \gtrsim s \log(\frac{en}{s}) \) [13], and this number of measurements cannot be improved (see Proposition 2.2.18 in [6]).

The reason behind this result, and many others like it, is that isotropic subgaussian matrices act on \( \Sigma_s \) in an isomorphic way, when \( N \gtrsim s \log(\frac{en}{s}) \).

Such a property is called the Restricted isometry property (RIP) (see, for example [3, 5, 14]). A matrix \( \Gamma \) satisfies the RIP in \( \Sigma_s \) if for every \( t \in \Sigma_s \),

\[
(1 - \delta) \| t \|_2 \le \| \Gamma t \|_2 \le (1 + \delta) \| t \|_2,
\]

for some fixed \( 0 < \delta < 1 \).

Proving the RIP for subgaussian matrices uses the fact that tails of linear functionals \( \langle X, t \rangle \) decay faster than the corresponding gaussian variable. Thus, it seemed natural to ask whether the same type of estimates hold in cases where linear functionals exhibit a slower decay – for example, when \( X \) is sub-exponential, and the linear functionals satisfy that \( \| \langle X, t \rangle \|_{L_p} \le L p \| \langle X, t \rangle \|_{L_2} \) for every \( t \in \mathbb{R}^n \) and every \( p \ge 2 \).

Proving the RIP for a sub-exponential ensemble is a much harder task than for subgaussian ensembles (cf. [1]). Moreover, the RIP does not exhibit the same behaviour as in the gaussian case. Indeed, one may show that for sub-exponential ensembles, RIP holds with high probability only when \( N \gtrsim s \log^2(\frac{en}{s}) \), and this estimate is optimal as can be seen when \( X \) has independent, symmetric exponential random variables as coordinates [1].

On the other hand, the result in [8] (see Chapter 7 there) shows that exact reconstruction can still be achieved by isotropic sub-exponential measurement vectors when \( N \gtrsim s \log(\frac{en}{s}) \) – the same number of measurements needed for the gaussian ensemble.

Clearly, this estimate cannot be based on the RIP, and one may ask whether weaker tail assumptions on the measurement vectors may still lead to exact recovery with the ‘gaussian’ number of measurements.

The main result presented here is precisely in this direction:

**Theorem A.** There exist absolute constants \( c_0, c_1 \) and \( c_2 \) and for every
$\alpha \geq 1/2$ there exists a constant $c_3(\alpha)$ that depends only on $\alpha$ for which the following holds. Let $X = (x_i)_{i=1}^n$ be a random vector on $\mathbb{R}^n$ such that

1. There are $\kappa_1, \kappa_2, w > 1$ that satisfy that for every $1 \leq j \leq n$, $\|x_j\|_{L_2} = 1$, and for $p = \kappa_2 \log(wn)$, $\|x_j\|_{L_p} \leq \kappa_1 p^\alpha$.

2. There are $u, \beta > 0$ that satisfy for every $t \in \Sigma \cap S^{n-1}$,

$$P(\langle X, t \rangle > u) \geq \beta.$$

If

$$N \geq c_0 \max \left\{ s \log(en/s), (c_3(\alpha)\kappa_1)^2(\kappa_2 \log(wn))^{\max\{2\alpha-1,1\}} \right\}$$

and $X_1, \ldots, X_N$ are independent copies of $X$, then, with probability at least $1 - 2\exp(-c_1\beta^2N) - 1/w^{\kappa_2n^{\kappa_2-1}}$, $\Gamma = N^{-1/2} \sum_{i=1}^N \langle X_i, \cdot \rangle e_i$ satisfies the exact reconstruction property in $\Sigma_{s_1}$ for $s_1 = c_2u^2\beta s$.

It follows from Theorem A, that a random matrix with iid centered entries that have variance 1 and an $L_p$ moment bounded by $p$ for $p = 2 \log n$ can be used as a measurement matrix, and just as in the gaussian case, requires only $N \gtrsim s \log(en/s)$ measurements.

Another straightforward application of Theorem A is for measurement vectors that are absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^n$ and that have almost surely bounded coordinates. The absolute continuity suffices to show that (2) in Theorem A holds for some $u$ and $\beta$, while the fact that the coordinates are uniformly bounded clearly implies (1).

Just as noted for sub-exponential ensembles, Theorem A cannot be proved using the RIP, and its proof must take a different path.

We end this introduction with a word about notation. Throughout, absolute constants or constants that depend on other parameters are denoted by $c$, $C$, $c_1$, $c_2$, etc., (and, of course, we will specify when a constant is absolute and when it depends on other parameters). The values of these constants may change from line to line. The notation $x \sim y$ (resp. $x \lesssim y$) means that there exist absolute constants $0 < c < C$ for which $cy \leq x \leq Cy$ (resp. $x \leq Cy$). If $b > 0$ is a parameter then $x \lesssim_b y$ means that $x \leq C(b)y$ for some constant $C(b)$ that depends only on $b$.

Let $\ell_p^n$ be $\mathbb{R}^n$ endowed with the norm $\|x\|_{\ell_p^n} = \left( \sum_j |x_j|^p \right)^{1/p}$. The unit ball there is denoted by $B_p^n$ and the unit Euclidean sphere in $\mathbb{R}^n$ is $S^{n-1}$. If $A \subset \mathbb{R}^n$ then $1_A$ denotes the indicator function of $A$. 

3
2 Proof of Theorem A.

The proof of Theorem A has several components, the first of which is rather standard - and shows that the so-called null space property suffices to ensure exact reconstruction. Although the proof of this fact is well known, we present it for the sake of completeness.

Lemma 2.1 Let $\Gamma : \mathbb{R}^n \to \mathbb{R}^N$ be a given matrix and fix $0 < r < 1$. If $B_1^n \cap rS^{n-1}$ does not intersect the kernel $\ker(\Gamma)$ then $\Gamma$ satisfies the exact reconstruction property in $\Sigma_{\lceil (2r)^{-2} \rceil}^\ell$.

Proof. Observe that if $x \in B_1^n$ and $\|x\|_2 \geq r$ then $y = rx/\|x\|_2 \in B_1^n \cap rS^{n-1}$. Thus, $\Gamma y \neq 0$ implies that $\Gamma x \neq 0$ – and therefore, $\sup_{x \in B_1^n \cap \ker(\Gamma)} \|x\|_2 < r$.

Let $s = \lceil (2r)^{-2} \rceil$, fix $x_0 \in \Sigma_s$ and put $I$ to be the set of coordinates on which $x_0$ is supported. Given a nonzero $h \in \ker(\Gamma)$, let $h = h_I + h_{I^c}$ - the decomposition of $h$ to coordinates in $I$ and in $I^c$. Since $h/\|h\|_1 \in B_1^n \cap \ker(\Gamma)$ then $\|h\|_2 \leq r\|h\|_1$, and by the choice of $s$, $2\sqrt{s}\|h\|_2 \leq \|h\|_1$. Therefore,

$$\|x_0 + h\|_1 = \|x_0 + h_I\|_1 + \|h_{I^c}\|_1 \geq \|x_0\|_1 - \|h_I\|_1 + \|h_{I^c}\|_1$$

$$= \|x_0\|_1 - 2\|h_I\|_1 + \|h_I\|_1 \geq \|x_0\|_1 - 2\sqrt{|I|}\|h_I\|_2 + \|h_I\|_1 > \|x_0\|_1.$$  

Hence, $\|x_0 + h\|_1 > \|x_0\|_1$ and $x_0$ is the unique minimizer of the basis pursuit algorithm, proving exact reconstruction.

The main component in the proof of Theorem A is a uniform empirical small-ball estimate, along the same lines as the results in [12] and [10].

Let $\mathcal{G}$ be a class of $\{0, 1\}$-valued functions defined on a space $\mathcal{X}$. The set $\mathcal{G}$ is a VC-class if there exists an integer $V$ for which, given any points $x_1, ..., x_{V+1} \in \mathcal{X},$

$$|\{(g(x_1), ..., g(x_{V+1})) : g \in \mathcal{G}\}| < 2^{V+1}. \tag{2.1}$$

The VC-dimension of $\mathcal{G}$, denoted by $VC(\mathcal{G})$, is the smallest integer $V$ for which (2.1) holds.

Lemma 2.2 There exists absolute constants $c_1$ and $c_2$ for which the following holds. Let $\mathcal{F}$ be a class of functions and assume that there is $\beta$ and $u$ for which

$$\inf_{f \in \mathcal{F}} P\{|f| > u\} \geq \beta.$$
Let $G_u = \{1_{|f|>u} : f \in F\}$. If $VC(G_u) \leq d$ and $N \geq c_1 d/\beta^2$ then with probability at least $1 - \exp(-c_2 \beta^2 N)$,

$$\inf_{f \in F} \{|i : |f(X_i)| > u\} \geq \frac{\beta N}{2}.$$ 

**Proof.** Let $H(X_1, ..., X_N) = \sup_{g \in G_u} |N^{-1} \sum_{i=1}^N g(X_i) - \mathbb{E} g(X)|$. By the bounded differences inequality (see, for example, Theorem 6.2 in [2]), with probability at least $1 - \exp(-t)$,

$$H(X_1, ..., X_N) \leq \mathbb{E} H(X_1, ..., X_N) + c_1 \sqrt{\frac{t}{N}}.$$ 

Since $VC(G_u) \leq d$, then by standard arguments,

$$\mathbb{E} H(X_1, ..., X_N) \leq c_2 \sqrt{\frac{d}{N}} \leq \frac{\beta}{4},$$

provided that $N \geq d/\beta^2$. Therefore, taking $t = N \beta^2 / 16 c_1$, then with probability at least $1 - \exp(-c_3 \beta^2 N)$, for every $f \in F$,

$$\frac{1}{N} \sum_{i=1}^N 1_{\{|f|>u\}}(X_i) \geq P\{|f| > u\} - \frac{\beta}{2} \geq \frac{\beta}{2},$$

and on that event, $\{|i : |f(X_i)| > u\} \geq \beta N/2$ for every $f \in F$. 

**Corollary 2.3** There exist absolute constants $c_1$ and $c_2$ for which the following holds. Let $X$ be a random vector on $\mathbb{R}^n$.

1. If there are $\beta, u > 0$ such that $P\{|\langle t, X \rangle| > u\} \geq \beta$ for any $t \in S^{n-1}$ and if $N \geq c_1 n/\beta^2$, then with probability at least $1 - \exp(-c_2 N \beta^2)$,

$$\inf_{t \in S^{n-1}} \frac{1}{N} \sum_{i=1}^N \langle X_i, t \rangle^2 \geq \frac{u^2 \beta}{2}.$$ 

2. If there are $\beta, u > 0$ such that $P\{|\langle t, X \rangle| > u\} \geq \beta$ for any $t \in \Sigma_\vartheta \cap S^{n-1}$ and if $N \geq c_1 \vartheta \log(en/\vartheta)/\beta^2$, then with probability at least $1 - \exp(-c_2 N \beta^2)$,

$$\inf_{t \in \Sigma_\vartheta \cap S^{n-1}} \frac{1}{N} \sum_{i=1}^N \langle X_i, t \rangle^2 \geq \frac{u^2 \beta}{2}.$$
Remark 2.4 Note that the first part of Corollary 2.3 gives an estimate on the smallest singular value of the random matrix $\Gamma = N^{-1/2} \sum_{i=1}^{N} \langle X_i, \cdot \rangle e_i$ along the lines of the estimate from [12], but without any assumption on the covariance structure of $X$, which is used in [12].

Proof of Corollary 2.3. To prove the first part of the claim, let $\mathcal{F} = \{ \langle t, \cdot \rangle : t \in S^{n-1} \}$. Recall that the VC dimension of a class of half-spaces in $\mathbb{R}^n$ is at most $n$, and thus, one may verify that for every $u$, the VC dimension of $G_u = \{ 1_{|f| > u} : f \in \mathcal{F} \}$ is at most $c_1 n$ for a suitable absolute constant $c_1$ (see, e.g., Chapter 2.6 in [16]). The claim now follows immediately from Theorem 2.2, because for any $t \in S^{n-1},$

$$\frac{1}{N} \sum_{i=1}^{N} \langle t, X_i \rangle^2 \geq \frac{u^2}{N} |\{ i : |\langle X_i, t \rangle| > u \}|.$$

Turning to the second part, note that $\Sigma_s \cap S^{n-1}$ is a union of $\binom{n}{s}$ spheres of dimension $s$. Applying the first part to each one of the spheres, combined with the union bound, it is evident that if $N \geq c_2 \beta^{-2} s \log(en/s)$ then with probability at least $1 - \exp(-c_3 N \beta^2),$ 

$$\inf_{t \in \Sigma_s \cap S^{n-1}} \frac{1}{N} \sum_{i=1}^{N} \langle X_i, t \rangle^2 \geq \frac{u^2 \beta}{2}.$$

Corollary 2.3 shows that the weak small ball assumption for linear functionals implies that $\Gamma$ ‘acts well’ on $s$-sparse vectors. However, to show that exact recovery is possible, one must show that it also acts well on the larger set $\sqrt{r} B_1^n \cap S^{n-1}$ for a well chosen $r$ that is proportional to $s$. Therefore, one has to use information on the way a matrix acts on $\Sigma_s$ to study the way it acts on the set $\sqrt{\kappa_0 s} B_1^n \cap S^{n-1} = \{ x \in \mathbb{R}^n : \|x\|_1 \leq \sqrt{\kappa_0 s}, \ \|x\|_2 = 1 \}.$

In the standard (RIP-based) argument, one also proves exact reconstruction by showing that the RIP holds on $\Sigma_s$. The fact that each vector in $\sqrt{\kappa_0 s} B_1^n \cap S^{n-1}$ is well approximated by vectors from $\Sigma_s$ (see, for instance, [6]) allows one to extend the RIP from $\Sigma_s$ to $\sqrt{\kappa_0 s} B_1^n \cap S^{n-1}$. However, extending the RIP requires both upper and lower estimates, and obtaining the upper part of the RIP on $\Sigma_s$ forces severe restrictions on the random vector $X$. Thus,
passing from $\Sigma_s$ to $\sqrt{n}sB_1^n \cap S^{n-1}$, with only a lower bound on $\inf_{t \in \Sigma_s} \|\Gamma_t\|_2$ at one’s disposal, requires a totally different argument.

The method we present below is based on Maurey’s empirical method and has been recently used in [15].

**Lemma 2.5** Let $A : \mathbb{R}^n \to \mathbb{R}^n$ be a matrix, set $V = (\langle Ae_i, Ae_i \rangle)_{i=1}^n$ and put $s > 1$. Assume that for every $x \in \Sigma_s$, $\|Ax\|_2 \geq \lambda \|x\|_2$, and for every nonzero $y \in \mathbb{R}^n$, set $\mu_i = |y_i|/\|y\|_1$. Then,

$$\|Ay\|_2^2 \geq \lambda^2 \|y\|_2^2 - \frac{\|y\|_2^2}{s-1} \left( \sum_{i=1}^n V_i \mu_i - \lambda^2 \right).$$

**Proof.** Fix $y \in \mathbb{R}^n$, let $Y$ be a random vector in $\mathbb{R}^n$ defined by $P(Y = \|y\|_1 \text{sgn}(y_i)e_i) = |y_i|/\|y\|_1$, and observe that $EY = y$.

Let $Y_1, \ldots, Y_s$ be independent copies of $Y$ and set $Z = s^{-1} \sum_{j=1}^s Y_j$; therefore, $Z \in \Sigma_s$ for every realization of $Y_1, \ldots, Y_s$.

By the assumption, $\|AZ\|_2^2 \geq \lambda^2 \|Z\|_2^2$, and thus,

$$E\|AZ\|_2^2 \geq \lambda^2 E\|Z\|_2^2. \tag{2.2}$$

It is straightforward to verify that if $i \neq j$ then $E\langle AY_i, AY_j \rangle = \langle Ay, Ay \rangle$; that for every $1 \leq i \leq s$,

$$E\langle AY_i, AY_i \rangle = \|y\|_1 \sum_{j=1}^n |y_j| \langle Ae_j, Ae_j \rangle;$$

and that $E\langle Y, Y \rangle = \|y\|_2^2$.

Therefore, setting $\mu_i = |y_i|/\|y\|_1$, and $W = \sum_{i=1}^n \langle Ae_i, Ae_i \rangle \mu_i$,

$$E\|AZ\|_2^2 = \frac{1}{s^2} \sum_{i,j=1}^s \langle AY_i, AY_j \rangle = \left( 1 - \frac{1}{s} \right) \|Ay\|_2^2 + \frac{\|y\|_1}{s} \sum_{j=1}^n |y_j| \langle Ae_j, Ae_j \rangle$$

$$= \left( 1 - \frac{1}{s} \right) \|Ay\|_2^2 + W \frac{\|y\|_2^2}{s},$$

and taking $A$ to be the identity in the last argument shows that

$$E\|Z\|_2^2 = \left( 1 - \frac{1}{s} \right) \|y\|_2^2 + \frac{\|y\|_2^2}{s},$$

Combining these two estimates with (2.2),

$$\left( 1 - \frac{1}{s} \right) \|Ay\|_2^2 \geq \lambda^2 \left( \left( 1 - \frac{1}{s} \right) \|y\|_2^2 + \frac{\|y\|_2^2}{s} \right) - W \frac{\|y\|_2^2}{s},$$

proving the claim. 

\[\blacksquare\]
The matrix \( A \) in question will be the square-root of \( \frac{1}{N} \sum_{i=1}^{N} X_i \otimes X_i \). Thus, for every \( t \in \mathbb{R}^n \), \( \|At\|_2^2 = \frac{1}{N} \sum_{i=1}^{N} \langle X_i, t \rangle^2 \), and if \( X_j = (x_{i,j})_{i=1}^n \) then
\[
V_j = \frac{1}{N} \sum_{i=1}^{N} x_{i,j}^2,
\]
which is an average of \( N \) iid random variables (though \( V_1, ..., V_n \) need not be independent!).

The next and final component needed for the proof of Theorem A is information on the sum of iid random variables, which will be used to analyze the random variables \( V_j \).

**Lemma 2.6** There exists an absolute constant \( c_0 \) for which the following holds. Let \( z \) be a mean-zero random variable and put \( z_1, ..., z_N \) to be \( N \) independent copies of \( z \). Let \( p_0 \geq 2 \) and assume that there exists \( \kappa > 0 \) and \( \alpha \geq 1/2 \) that satisfy that \( \|z\|_{L^p} \leq \kappa_1 p^\alpha \) for every \( 2 \leq p \leq p_0 \). If \( N \geq p_0^{\max\{2\alpha-1,1\}} \) then for every \( p \leq p_0 \),
\[
\| \sum_{i=1}^{N} z_i \|_{L^p} \leq c_1(\alpha) \kappa_1 \sqrt{Np}
\]
where \( c_1(\alpha) = c_0 \exp((2\alpha - 1)) \).

Lemma 2.6 shows that even under a weak moment assumption, \( \|z\|_{L^p} \lesssim p^\alpha \) for \( p \leq p_0 \) and \( \alpha \geq 1/2 \) that can be large, a normalized sum of \( N \) independent copies of \( z \) exhibits a ‘subgaussian’ moment growth up to the same \( p_0 \), provided that \( N \) is sufficiently large.

The proof of Proposition 2.6 is based on the following result due to Latała [9].

**Theorem 2.7** If \( z \) is a centred random variable and \( z_1, ..., z_N \) are independent copies of \( z \), then for any \( p \geq 2 \),
\[
\| \sum_{i=1}^{N} z_i \|_{L^p} \sim \sup \left\{ \frac{p}{s} \left( \frac{N}{p} \right)^{1/s} \|z\|_{L^s} : \max\{2, p/N\} \leq s \leq p \right\}.
\]

**Proof of Proposition 2.6.** Since \( \|z\|_{L^p} \leq \kappa_1 p^\alpha \), it follows from Theorem 2.7 that
\[
\| \sum_{i=1}^{N} z_i \|_{L^p} \leq c\kappa_1 \sup_{s} \frac{p(N/p)^{1/s} s^{-1+\alpha}}{s},
\]

8
where the supremum is for \( \max\{2, p/N\} \leq s \leq p \). It is straightforward to verify that the function \( h(s) = (N/p)^{1/s-1+\alpha} \) is decreasing when \( \alpha \leq 1 \) and attains its maximum in \( s = \max\{2, p/N\} \) or in \( s = p \) when \( \alpha > 1 \).

Therefore, if \( N \geq p \) and \( \alpha \leq 1 \), then
\[
\left\| \sum_{i=1}^{N} z_i \right\|_{L_p} \leq c_1 \kappa_1 \sqrt{Np},
\]
and if \( \alpha > 1 \),
\[
\left\| \sum_{i=1}^{N} z_i \right\|_{L_p} \leq c_1 \kappa_1 \max \left\{ \sqrt{Np}, N^{1/p}p^\alpha \right\}.
\]

Finally, if \( N \geq p^{2\alpha-1} \) then \( e^{2\alpha-1} \sqrt{Np} \geq N^{1/p}p^\alpha \), which completes the proof.

\[\Box\]

**Proof of Theorem A.** By Corollary 2.3, if \( N \geq c_1 s \log(en/s)/\beta^2 \), then with probability at least \( 1 - \exp(-c_2 N\beta^2) \),
\[
\inf_{t \in \Sigma \cap S^{n-1}} \frac{1}{N} \sum_{i=1}^{N} \langle X_i, t \rangle^2 \geq \frac{\mu^2 \beta}{2}.
\]

Moreover, by Lemma 2.5 used for \( A \) that is the squared root of \( \frac{1}{N} \sum X_i \otimes X_i \), and \( \lambda^2 = \mu^2 \beta/2 \), it follows that for \( r \geq 1 \) and on the same event as above,
\[
\inf_{t \in \sqrt{r}B^n \cap S^{n-1}} \| \Gamma t \|_2^2 \geq \lambda^2 - \frac{2r}{s} \max_{1 \leq i \leq n} V_i. \tag{2.3}
\]

Finally, fix \( w \geq 1 \) and consider \( z = x_j^2 - 1 \) - where \( x_j \) is the \( j \)-th coordinate of \( X \). Since \( z \) is a centred random variable, then by Lemma 2.6 for \( p = \kappa_2 \log(wn) \),
\[
\left\| \frac{1}{N} \sum_{i=1}^{N} z_i \right\|_{L_p} \leq c_3(\alpha)\kappa_1 \sqrt{\frac{p}{N}},
\]
where \( c_3(\alpha) \sim \exp((2\alpha - 1)) \), provided that
\[
N \geq p^{\max\{2\alpha-1,1\}} = (\kappa_2 \log(wn))^{\max\{2\alpha-1,1\}}.
\]

Therefore, if \( N \geq (c_3(\alpha)\kappa_1)^2(\kappa_2 \log(wn))^{\max\{2\alpha-1,1\}} \), then
\[
\| V_j \|_{L_p} = \left\| \frac{1}{N} \sum_{i=1}^{N} x_{j,i}^2 \right\|_{L_p} \leq 1 + c_3(\alpha)\kappa_1 \sqrt{\frac{\kappa_2 \log(wn)}{N}} \leq 2.
\]

9
Observe that

\[
P\left( \max_{1 \leq j \leq n} V_i \geq 2e \right) \leq \sum_{j=1}^{n} P(V_j \geq 2e) \leq n \sum_{j=1}^{n} \left( \frac{\|V_j\|_{L_p}}{2e} \right)^p \leq n \left( \frac{1}{e} \right)^p = \frac{1}{w\kappa_2 n^{\kappa_2-1}}.
\]

Thus, with probability at least \(1 - \exp(-c_2 N \beta^2) - 1/(w\kappa_2 n^{\kappa_2-1})\),

\[
\inf_{t \in \sqrt{r}B_n^1 \cap S^{n-1}} \|\Gamma t\|_2^2 \geq \lambda^2 - \frac{4er}{s} \geq \lambda^2/2 \tag{2.4}
\]

provided that \(r \leq s\lambda^2/8e = su^2\beta/16e\).

Combining (2.4) with Lemma 2.1 shows that if

\[
N \gtrsim \max \left\{ s \log(en/s), (c_3(\alpha)\kappa_1)^2(\kappa_2 \log(wn))^{\max\{2\alpha-1,1\}} \right\},
\]

then on the same event as above, \(\Gamma\) satisfies the exact reconstruction property for vectors that are \(c_4(u, \beta)s\)-sparse, as claimed.

\[\square\]

**Remark 2.8** An alternative argument used to bound the Gelfand widths of convex bodies and which may be used to prove Theorem A as well can be found in [11].

**References**

[1] Radoslaw Adamczak, Alexander E. Litvak, Alain Pajor, and Nicole Tomczak-Jaegermann. Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling. *Constr. Approx.*, 34(1):61–88, 2011.

[2] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013. ISBN 978-0-19-953525-5.

[3] Emmanuel J. Candès, Justin Romberg, and Terence Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

[4] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.

[5] Emmanuel J. Candès and Terence Tao. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.
[6] Djalil Chafaï, Olivier Guédon, Guillaume Lecué, and Alain Pajor. *Interactions between compressed sensing random matrices and high dimensional geometry*, volume 37 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2012.

[7] David L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[8] Vladimir Koltchinskii. Local Rademacher complexities and oracle inequalities in risk minimization. *Ann. Statist.*, 34(6):2593–2656, 2006.

[9] Rafał Latala. Estimation of moments of sums of independent real random variables. *Ann. Probab.*, 25(3):1502–1513, 1997.

[10] Shahar Mendelson. Learning without concentration. Technical report, Technion, 2013. arXiv:1401.0304.

[11] Shahar Mendelson. A remark on the diameter of random sections of convex bodies. Technical report, Technion, 2013.

[12] Shahar Mendelson and Vladimir Koltchinskii. Bounding the smallest singular value of a random matrix without concentration. Technical report, Technion and Georgia Tech, 2013. arXiv:1312.3580.

[13] Shahar Mendelson, Alain Pajor, and Nicole Tomczak-Jaegermann. Reconstruction and subgaussian operators in asymptotic geometric analysis. *Geom. Funct. Anal.*, 17(4):1248–1282, 2007.

[14] Shahar Mendelson, Alain Pajor, and Nicole Tomczak-Jaegermann. Uniform uncertainty principle for Bernoulli and subgaussian ensembles. *Constr. Approx.*, 28(3):277–289, 2008.

[15] Roberto Imbuzeiro Oliveira. The lower tail of random quadratic forms, with applications to ordinary least squares and restricted eigenvalue properties. Technical report, IMPA, 2013. arXiv:1312.2903.

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