Sparse Recovery with Pre-Gaussian Random Matrices

Simon Foucart[*]
Laboratoire Jacques-Louis Lions
Université Pierre et Marie Curie
Paris, 75013, France

Ming-Jun Lai[†]
Department of Mathematics
University of Georgia
Athens, GA 30602, USA

February 25, 2010

Abstract

For an $m \times N$ underdetermined system of linear equations with independent pre-Gaussian random coefficients satisfying simple moment conditions, it is proved that the s-sparse solutions of the system can be found by ℓ_1-minimization under the optimal condition $m \geq c\, s \ln(eN/s)$. The main ingredient of the proof is a variation of a classical Restricted Isometry Property, where the inner norm becomes the ℓ_1-norm and where the outer norm depends on the probability distributions.

Key Words: compressed sensing, sparse recovery, ℓ_1-minimization, random matrices, pre-Gaussian random variables.

1 Introduction

The field of Compressed Sensing, which has generated a wealth of research activity in recent years, asks for some concrete protocols that make it possible to reconstruct sparse vectors $\mathbf{x} \in \mathbb{R}^N$ from the mere knowledge of measurement vectors $\mathbf{y} = A\mathbf{x} \in \mathbb{R}^m$ with $m \ll N$. In other words, one seeks $m \times N$ measurement matrices A and recovery algorithms that enable to find the sparsest solutions of the underdetermined linear system $A\mathbf{x} = \mathbf{y}$. The groundbreaking works of Donoho [9] and of Candès and Tao [8] successfully tackled these questions. The problem of choosing suitable matrices was settled using probabilistic arguments, with the conclusion that most matrices chosen at random allow for an efficient reconstruction of sparse vectors. The reconstruction in question consists in solving the computationally tractable convex optimization problem

$$\underset{\mathbf{z} \in \mathbb{R}^N}{\text{minimize}} \; \|\mathbf{z}\|_1 \quad \text{subject to } A\mathbf{z} = \mathbf{y}, \tag{P$_1$}$$

---

[*] foucart@ann.jussieu.fr, This work was initiated when the author was at Vanderbilt University and was later supported by the French National Research Agency (ANR) through the project ECHANGE (ANR-08-EMER-006).

[†] mjlai@math.uga.edu, This research is supported by the National Science Foundation under grant #DMS 0713807.

in place of the unpractical combinatorial problem

$$\underset{\mathbf{z}\in\mathbb{R}^N}{\text{minimize}} \;\; \|\mathbf{z}\|_0 \quad \text{subject to } A\mathbf{z} = \mathbf{y}. \tag{P$_0$}$$

Here $\|\mathbf{z}\|_1 = \sum_{j=1}^N |z_j|$ stands for the usual ℓ_1-norm of a vector $\mathbf{z} \in \mathbb{R}^N$, while $\|\mathbf{z}\|_0$ represents its sparsity, i.e., the number of its nonzero components. A much favored tool in the study of the equivalence between (P_0) and (P_1) was introduced by Candès and Tao in [8]. It is said that a matrix $A \in \mathbb{R}^{m\times N}$ has the kth order Restricted Isometry Property if there is a constant $\delta \geq 0$ such that

$$(1-\delta)\|\mathbf{x}\|_2^2 \leq \|A\mathbf{x}\|_2^2 \leq (1+\delta)\|\mathbf{x}\|_2^2 \qquad \text{for all } k\text{-sparse } \mathbf{x} \in \mathbb{R}^N. \tag{1}$$

The smallest such constant, denoted by δ_k, is called the kth order Restricted Isometry Constant of A. There are many conditions on the δ_k's that guarantee the recovery of all s-sparse vectors $\mathbf{x} \in \mathbb{R}^N$ as solutions of (P_1) with $\mathbf{y} = A\mathbf{x}$. The arguably most natural ones are only in terms of δ_{2s}. For instance, Candès established the sufficient condition $\delta_{2s} < \sqrt{2}-1 \approx 0.4142$ in [7]. This was later improved in [11, 6] to arrive at the sufficient condition $\delta_{2s} < 3/(4+\sqrt{6}) \approx 0.4652$ in [10]. Regardless of the sufficient condition called upon, the crucial point is that it is met with overwhelming probability for certain random matrices whose number m of rows scales like the sparsity s times a power of the log factor $\ln(eN/s)$, with N denoting the number of columns. An important example for practical applications is the case of partial Fourier matrices, where $m \geq c\, s\, \ln^4(N)$ rows of an $N\times N$ Fourier matrix are drawn uniformly at random, see [21]. For Gaussian random matrices, i.e., matrices whose entries are independent copies of a zero-mean Gaussian random variable, the number of measurements can be reduced to $m \geq c\, s\, \ln(eN/s)$, see [8]. This bound cannot be reduced further if searching for stable sparse recovery algorithms, as shown by considerations about Gelfand widths, see e.g. [12]. Sparse recovery by ℓ_1-minimization, deduced from the Restricted Isometry Property (1), is also possible with $m \geq c\, s\, \ln(eN/s)$ when considering random matrices satisfying a concentration inequality, see [2] for a simple proof, or sub-Gaussian random matrices, see [18]. For pre-Gaussian random matrices, sparse recovery by ℓ_1-minimization was established in [1] under the stronger condition $m \geq c\, s\, \ln^2(eN/s)$. This was again deduced from the Restricted Isometry Property (1). The latter actually necessitates such a strong condition on the number of measurements in the pre-Gaussian setting, see [1].

This paper aims to show that sparse recovery using pre-Gaussian random matrices is still possible under the optimal condition on the number of measurements. Indeed, we show in Theorem 6.1 that, given an $m\times N$ matrix populated by independent pre-Gaussian random variables obeying simple moment conditions, it is overwhelmingly probable that every s-sparse vector is recovered by ℓ_1-minimization, provided $m \geq c\, s\, \ln(eN/s)$. Note that we are following the terminology of [5] in calling a random variable pre-Gaussian when it has a subexponential tail decay. This meaning is made precise in Definition 2.1, where an alternative view in terms of the moment growth $\mathbf{E}\big(|\xi|^{2k}\big) \leq (2k)!\,\theta^{2k}$ is given. For instance, the Laplace random variables η, whose probability density functions are $\exp(-|t|/\lambda)/(2\lambda)$ for $\lambda > 0$, are pre-Gaussian since $\mathbf{E}\big(|\eta|^r\big) = \Gamma(r+1)\lambda^r$ for all $r > 0$.

Pre-Gaussian random variables are also often called ψ_1 random variables because they are characterized by the finiteness of their Orlicz norm

$$\|\xi\|_{\psi_1} := \inf\big\{t > 0 : \mathbf{E}\exp(|\xi|/t) \le 2\big\}.$$

The arguments of this paper rely on a variation of the classical Restricted Isometry Property (1). Its formulation involves the quantity

$$/\!\!/\mathbf{x}/\!\!/_{\boldsymbol{\nu}} := \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \Big| \sum_{j=1}^{N} t_j x_j \Big| d\nu_1(t_1) \cdots d\nu_N(t_N), \qquad \mathbf{x} \in \mathbb{R}^N,$$

relative to a vector $\boldsymbol{\nu}$ of centered probability measures. It is easy to verify that such an expression, appearing e.g. in [4, 14, 19], defines a norm on $\mathbb{R}^N$ provided the first absolute moment of each ν_j is finite. A sum of such norms will replace the outer norm in (1), while the inner norm will be replaced by the ℓ_1-norm. Some variations of the classical Restricted Isometry Property (1) are already present in the Compressed Sensing literature — for Gaussian matrices, the inner norm is the ℓ_1-norm in Definition 4.1 of [9]; for adjacency matrices of lossless expanders, both inner and outer norms are ℓ_1-norms in [3] — but it is the dependency of the outer norm on the probability distributions that constitutes the novelty of our approach. Thus, for random matrices whose entries $a_{i,j}$ are distributed according to centered probability measures $\nu_{i,j}$, we set

$$/\!\!/\mathbf{x}/\!\!/ := \sum_{i=1}^{m} /\!\!/\mathbf{x}/\!\!/_{\boldsymbol{\nu}^i}, \qquad \boldsymbol{\nu}^i := [\nu_{i,1}, \ldots, \nu_{i,N}]. \tag{2}$$

If the entries $a_{i,j}$ were independent standard centered Gaussian random variables, for instance, an explicit computation would give $/\!\!/\mathbf{x}/\!\!/ = m\sqrt{2/\pi}\,\|\mathbf{x}\|_2$, and we would retrieve Definition 4.1 of [9]. We are interested in the Modified Restricted Isometry Constant $\delta_k^{/\!\!/\cdot/\!\!/}$ defined as the smallest constant $\delta \ge 0$ such that

$$(1-\delta)\,/\!\!/\,\mathbf{x}/\!\!/ \le \|A\mathbf{x}\|_1 \le (1+\delta)\,/\!\!/\,\mathbf{x}\,/\!\!/ \qquad \text{for all } k\text{-sparse } \mathbf{x} \in \mathbb{R}^N. \tag{3}$$

In Section 4, we prove that this Modified Restricted Isometry Constant can be made sufficiently small. This is deduced from the concentration inequality, relative to the norm (2), that we establish in Section 3. In Section 5, we then show that the Modified Restricted Isometry Property (3), relative a norm comparable to a Euclidean one, implies sparse recovery by ℓ_1-minimization. In Section 6, we finally combine the previous results to arrive at our main theorem. As a prelude to all this, we collect in Section 2 some auxiliary results needed in our arguments. Note that we chose to deal only with recovery of exactly sparse vectors from perfect measurements for the sake of clarity. However, in the spirit of Compressed Sensing, one needs to control the error between a nearly sparse vector and a vector recovered from slightly flawed measurements. A reader familiar with Compressed Sensing could easily perform the appropriate modifications in order to establish such a result here.

2 Preliminary Results

We start with the definition of pre-Gaussian random variables.

Definition 2.1 *A random variable ξ is pre-Gaussian if one of the following equivalent conditions holds:*

1. $\mathbf{E}(\xi) = 0$ *and there exist constants $b > 0$ and $c > 0$ such that*

$$\mathbf{P}(|\xi| > t) \leq b \exp(-c\,t) \qquad \textit{for all } t > 0,$$

2. $\mathbf{E}(\xi) = 0$ *and there exists a constant $\theta > 0$ such that*

$$\theta(\xi) := \sup_{k \geq 1} \left[\frac{\mathbf{E}|\xi|^{2k}}{(2k)!} \right]^{1/2k} \leq \theta.$$

The justification of the equivalence is found in Theorems 3.1 and 3.2 of [5, p 21–23]. In fact, Theorem 3.2 is stated with $\theta(\xi)$ replaced by

$$\theta'(\xi) := \sup_{k \geq 1} \left[\frac{\mathbf{E}|\xi|^{k}}{k!} \right]^{1/k},$$

but the two statements are similar in view of the inequalities

$$\theta(\xi) \leq \theta'(\xi) \leq 2\theta(\xi).$$

The lower inequality is clear, while the upper inequality is a simple consequence of $\mathbf{E}(|\xi|^k) \leq \mathbf{E}(|\xi|^{2k})^{1/2}$. The quantity $\theta(\xi)$ turns out to be of more convenient usage, because of the following result, see Theorem 3.6 of [5, p 61].

Proposition 2.1 *If $\xi_1, \ldots, \xi_n$ are independent pre-Gaussian random variables, then*

$$\theta^2(x_1\xi_1 + \cdots + x_n\xi_n) \leq x_1^2\,\theta^2(\xi_1) + \cdots + x_n^2\,\theta^2(\xi_n).$$

The next result claims that the norm $/\!\!/ \cdot /\!\!/_{\boldsymbol{\nu}}$ is comparable to a Euclidean norm.

Proposition 2.2 *Suppose that $\xi_1, \ldots, \xi_N$ are independent zero-mean random variables satisfying*

$$\mathbf{E}\big(|\xi_j|\big) \geq \mu \qquad \textit{and} \qquad \mathbf{E}\big(\xi_j^2\big) \leq \sigma^2 \qquad \textit{for all } j = 1, \ldots, N.$$

Then, with $\nu_1, \ldots, \nu_N$ denoting the centered probability measures associated to $\xi_1, \ldots, \xi_N$,

$$\frac{\mu}{\sqrt{8}} \|\mathbf{x}\|_2 \leq /\!\!/ \mathbf{x} /\!\!/_{\boldsymbol{\nu}} \leq \sigma \|\mathbf{x}\|_2 \qquad \textit{for all } \mathbf{x} \in \mathbb{R}^N.$$

Proof. Let us observe first that

$$/\!\!/ \mathbf{x} /\!\!/_{\boldsymbol{\nu}} = \mathbf{E} \Big| \sum_{j=1}^{N} x_j \xi_j \Big|.$$

For the upper estimate, the independence of the ξ_j's simply yields

$$\mathbf{E} \Big| \sum_{j=1}^{N} x_j \xi_j \Big| \le \Big(\mathbf{E} \big(\sum_{j=1}^{N} x_j \xi_j \big)^2 \Big)^{1/2} = \Big(\sum_{j=1}^{N} x_j^2 \, \mathbf{E}(\xi_j)^2 \Big)^{1/2} \le \sigma \|\mathbf{x}\|_2.$$

As for the lower estimate, we use the symmetrization procedure, see Lemma 6.3 of [17], to write

$$\mathbf{E} \Big| \sum_{j=1}^{N} x_j \xi_j \Big| \ge \frac{1}{2} \mathbf{E} \Big| \sum_{j=1}^{N} \epsilon_j x_j \xi_j \Big| = \frac{1}{2} \mathbf{E}_{\boldsymbol{\xi}} \mathbf{E}_{\boldsymbol{\epsilon}} \Big| \sum_{j=1}^{N} \epsilon_j x_j \xi_j \Big|, \tag{4}$$

where $(\epsilon_1, \ldots, \epsilon_N)$ is a Rademacher sequence independent of $(\xi_1, \ldots, \xi_N)$. Next, using Khintchine's inequality with optimal constants due to Haagerup [16], we have

$$\mathbf{E}_{\boldsymbol{\epsilon}} \Big| \sum_{j=1}^{N} \epsilon_j x_j \xi_j \Big| \ge \frac{1}{\sqrt{2}} \Big(\sum_{j=1}^{N} x_j^2 \, \xi_j^2 \Big)^{1/2} = \frac{\|\mathbf{x}\|_2}{\sqrt{2}} \Big(\sum_{j=1}^{N} \frac{x_j^2}{\|\mathbf{x}\|_2^2} \, \xi_j^2 \Big)^{1/2}.$$

Then, using the concavity of the function $t \mapsto t^{1/2}$, we derive

$$\mathbf{E}_{\boldsymbol{\epsilon}} \Big| \sum_{j=1}^{N} \epsilon_j x_j \xi_j \Big| \ge \frac{\|\mathbf{x}\|_2}{\sqrt{2}} \sum_{j=1}^{N} \frac{x_j^2}{\|\mathbf{x}\|_2^2} \, |\xi_j|. \tag{5}$$

The desired estimate follows from (4), (5), and $\mathbf{E}(|\xi_j|) \ge \mu$. ■

Finally, we state Bernstein's inequality as in Lemma 2.2.11 of [22] for easy reference.

Theorem 2.1 *Let $Y_1, \ldots, Y_m$ be independent zero-mean random variables for which there exist positive constants $M, v_1, \ldots, v_m$ such that*

$$\mathbf{E}\big(|Y_i|^k\big) \le \frac{k!}{2} M^{k-2} v_i \qquad \text{for all integers } k \ge 2,$$

then, for all $t > 0$,

$$\mathbf{P}\big(|Y_1 + \ldots + Y_m| > t\big) \le 2 \exp\Big(- \frac{t^2}{2(v_1 + \ldots + v_m + tM)} \Big).$$

3 Subexponential Tail Decay

In order to establish the Modified Restricted Isometry Property (3) for all sparse vectors, we consider first individual vectors $\mathbf{x} \in \mathbb{R}^N$ and we bound the tail probability

$$\mathbf{P}\Big(\big| \|A\mathbf{x}\|_1 - /\!\!/ \mathbf{x} /\!\!/ \big| > \epsilon \, /\!\!/ \mathbf{x} /\!\!/ \Big).$$

Theorem 3.1 *Suppose the entries of a matrix $A \in \mathbb{R}^{m\times N}$ are independent pre-Gaussian random variables satisfying*

$$\mathbf{E}\big(|a_{i,j}|\big) \geq \mu \qquad \textit{and} \qquad \mathbf{E}\big(|a_{i,j}|^{2k}\big) \leq (2k)!\, \theta^{2k}, \quad k \geq 1.$$

If $/\!\!/ \cdot /\!\!/$ denotes the norm defined in (2) *for the centered probability measures $\nu_{i,j}$ associated to the entries $a_{i,j}$, then*

$$\mathbf{P}\Big(\big| \|A\mathbf{x}\|_1 - /\!\!/ \mathbf{x} /\!\!/ \,\big| > \epsilon \,/\!\!/ \,\mathbf{x}\, /\!\!/ \Big) \leq 2 \exp\left(-\kappa \epsilon^2 m\right) \tag{6}$$

for any $\mathbf{x} \in \mathbb{R}^N$ and any $\epsilon \in (0,1)$, where the constant κ depends only on θ/μ.

Proof. Setting $Y_i := |(A\mathbf{x})_i| - /\!\!/ \mathbf{x} /\!\!/_{\boldsymbol{\nu}^i}$, we observe that $Y_1, \ldots, Y_m$ are independent zero-mean random variables, and that

$$\|A\mathbf{x}\|_1 - /\!\!/ \mathbf{x} /\!\!/ = \sum_{i=1}^m Y_i.$$

Then, since $\theta(a_{i,j}) \leq \theta$, Proposition 2.1 yields

$$\theta\big((A\mathbf{x})_i\big) = \theta\big(\textstyle\sum_{j=1}^N x_j a_{i,j}\big) \leq \theta \|\mathbf{x}\|_2, \qquad \text{hence} \qquad \theta'\big((A\mathbf{x})_i\big) \leq 2\theta \|\mathbf{x}\|_2.$$

For an integer $k \geq 2$, it follows from the inequality $|Y_i| \leq \max\{|(A\mathbf{x})_i|, /\!\!/ \mathbf{x} /\!\!/_{\boldsymbol{\nu}^i}\}$ that

$$\begin{aligned} \mathbf{E}\big(|Y_i|^k\big) &\leq \max\{\mathbf{E}\big(|(A\mathbf{x})_i|^k\big), /\!\!/ \mathbf{x} /\!\!/_{\boldsymbol{\nu}^i}^k\} \leq \max\{k!\, \big(\theta'\big((A\mathbf{x})_i\big)\big)^k, /\!\!/ \mathbf{x} /\!\!/_{\boldsymbol{\nu}^i}^k\} \\ &\leq \max\{\frac{k!}{2} 2^{k/2} \big(\theta'\big((A\mathbf{x})_i\big)\big)^k, \frac{k!}{2} /\!\!/ \,\mathbf{x} /\!\!/_{\boldsymbol{\nu}^i}^k\} \leq \frac{k!}{2} \max\{\sqrt{8}\,\theta \|\mathbf{x}\|_2, /\!\!/ \mathbf{x} /\!\!/_{\boldsymbol{\nu}^i}\}^k. \end{aligned}$$

Since Proposition 2.2 implies $\mu\|\mathbf{x}\|_2/\sqrt{8} \leq /\!\!/ \mathbf{x} /\!\!/_{\boldsymbol{\nu}^i} \leq \sqrt{2}\,\theta \|\mathbf{x}\|_2$, we can apply Bernstein's inequality with

$$M = \max\{\sqrt{8}\,\theta\|\mathbf{x}\|_2, \sqrt{2}\,\theta\|\mathbf{x}\|_2\} = \sqrt{8}\,\theta\|\mathbf{x}\|_2, \quad v_i = M^2, \quad t = \epsilon \,/\!\!/ \,\mathbf{x} /\!\!/ \geq \epsilon\, m\, \mu \|\mathbf{x}\|_2/\sqrt{8}$$

to obtain

$$\begin{aligned} \mathbf{P}\Big(\big| \|A\mathbf{x}\|_1 - /\!\!/ \mathbf{x} /\!\!/ \,\big| > \epsilon m \,/\!\!/ \,\mathbf{x}\, /\!\!/_{\boldsymbol{\nu}} \Big) &\leq 2 \exp\left(- \frac{\epsilon^2 m^2 \mu^2 \,/\!\!/ \,\mathbf{x}\, /\!\!/^2 /8}{2\big(8\, m\, \theta^2 \|\mathbf{x}\|_2^2 + \epsilon\, m\, \mu\, \theta \|\mathbf{x}\|_2^2\big)} \right) \\ &= 2 \exp\left(- \frac{\epsilon^2 m}{16\big(8(\theta/\mu)^2 + \epsilon(\theta/\mu)\big)} \right). \end{aligned}$$

Since $\epsilon \leq 1$, the result follows with $\kappa := 1/\big(128(\theta/\mu)^2 + 16(\theta/\mu)\big)$. ■

Remark 3.1 *The advantage of taking the ℓ_1-norm rather than the ℓ_2-norm as the inner norm in the Modified Restricted Isometry Property* (3) *is apparent at this point. If we had used the ℓ_2-norm, we would have considered the random variable $Y_i = (A\mathbf{x})_i^2 - \mathbf{E}\big((A\mathbf{x})_i^2\big)$, and we would have tried to bound the kth moment of $(A\mathbf{x})_i^2$ by $k!M^k$ for some $M > 0$ in order to apply Bernstein's inequality. However, it proves difficult to obtain more than $\mathbf{E}\big((A\mathbf{x})_i^{2k}\big) \leq (2k)!\big(\theta(A\mathbf{x})_i\big)^{2k} \leq (2k)!\big(\theta\|\mathbf{x}\|_2\big)^{2k}$.*

Remark 3.2 *The methods and results of Sections 2 and 3 are standard in geometry of Banach spaces, see e.g. [13]. They have been spelled out with the Compressed Sensing reader in mind.*

4 Modified Restricted Isometry Property

In this section, we show how to pass from the concentration inequality (6) for individual vectors to the Modified Restricted Isometry Property (3) for all sparse vectors. In fact, we prove that (3) fails with exponentially small probability. We essentially follow the ideas of [2]. The details are included for the reader's convenience.

Theorem 4.1 *Let $A \in \mathbb{R}^{m \times N}$ be a random matrix and let $/\!\!/ \cdot /\!\!/$ be a norm on $\mathbb{R}^N$. Suppose that, for any $\mathbf{x} \in \mathbb{R}^N$ and any $\epsilon \in (0,1)$,*

$$\mathbf{P}\Big(\big| \|A\mathbf{x}\|_1 - /\!\!/ \mathbf{x} /\!\!/ \big| > \epsilon /\!\!/ \mathbf{x} /\!\!/ \Big) \leq 2 \exp\big(- \kappa \epsilon^2 m \big). \tag{7}$$

Then there exist constants $c_1, c_2 > 0$ depending only on κ such that, for any $\delta \in (0,1)$,

$$\mathbf{P}\Big(\big| \|A\mathbf{x}\|_1 - /\!\!/ \mathbf{x} /\!\!/ \big| > \delta /\!\!/ \mathbf{x} /\!\!/ \quad \textit{for some } s\textit{-sparse } \mathbf{x} \in \mathbb{R}^N \Big) \leq 2 \exp\big(- c_1 \delta^2 m \big)$$

provided

$$m \geq \frac{c_2}{\delta^3} \, s \, \ln\Big(\frac{eN}{s} \Big).$$

Proof. We start by considering a fixed index set $S \subseteq \{1, \ldots, N\}$ of cardinality s. Let $\mathcal{S}$ denote the unit sphere of the space $\mathbb{R}^S$ of vectors supported on S embedded with the norm $/\!\!/ \cdot /\!\!/$. According to Lemma 4.10 of [20], we can find a subset $\mathcal{U}$ of $\mathcal{S}$ such that

$$\min_{\mathbf{u} \in \mathcal{U}} /\!\!/ \mathbf{x} - \mathbf{u} /\!\!/ \leq \gamma := \frac{\delta}{3} \quad \text{for all } \mathbf{x} \in \mathcal{S} \qquad \text{and} \qquad \mathrm{card}(\mathcal{U}) \leq \Big(1 + \frac{2}{\gamma} \Big)^s.$$

The concentration inequality (7), together with a union bound, gives

$$\begin{aligned} \mathbf{P}\Big(\big| \|A\mathbf{u}\|_1 - /\!\!/ \mathbf{u} /\!\!/ \big| > \gamma /\!\!/ \mathbf{u} /\!\!/ \ \text{ for some } \mathbf{u} \in \mathcal{U} \Big) &\leq \Big(1 + \frac{2}{\gamma} \Big)^s 2 \exp\big(- \kappa \gamma^2 m \big) \\ &\leq 2 \exp\Big(- \kappa \gamma^2 m + \frac{2s}{\gamma} \Big) = 2 \exp\Big(- \frac{\kappa \delta^2 m}{9} + \frac{6s}{\delta} \Big). \end{aligned}$$

This means that the matrix A is drawn with high probability in such a way that

$$(1 - \gamma) /\!\!/ \mathbf{u} /\!\!/ \leq \|A\mathbf{u}\|_1 \leq (1 + \gamma) /\!\!/ \mathbf{u} /\!\!/ \quad \text{for all } \mathbf{u} \in \mathcal{U}. \tag{8}$$

Let $\tilde{\delta}$ be the smallest positive constant such that

$$\|A\mathbf{x}\|_1 \leq (1 + \tilde{\delta}) /\!\!/ \mathbf{x} /\!\!/ \qquad \text{for all } \mathbf{x} \in \mathcal{S}. \tag{9}$$

Given $\mathbf{x} \in \mathcal{S}$, picking $\mathbf{u} \in \mathcal{U}$ with $/\!\!/ \mathbf{x} - \mathbf{u} /\!\!/ \leq \gamma$, we derive

$$\|A\mathbf{x}\|_1 \leq \|A\mathbf{u}\|_1 + \|A(\mathbf{x} - \mathbf{u})\|_1 \leq 1 + \gamma + (1 + \tilde{\delta}) /\!\!/ \mathbf{x} - \mathbf{u} /\!\!/ \leq 1 + \gamma + (1 + \tilde{\delta})\gamma.$$

The minimality of $\tilde{\delta}$ implies that

$$1 + \tilde{\delta} \leq 1 + \gamma + (1 + \tilde{\delta})\gamma \leq 1 + 2\gamma + \tilde{\delta}/3, \qquad \text{so that} \quad \tilde{\delta} \leq 3\gamma = \delta.$$

Substituting into (9), we obtain the upper estimate

$$\|A\mathbf{x}\|_1 \le (1+\delta) \,/\!\!/\, \mathbf{x} \,/\!\!/ \qquad \text{for all } \mathbf{x} \in \mathbb{R}^S. \tag{10}$$

Subsequently, for $\mathbf{x} \in \mathcal{S}$ and $\mathbf{u} \in \mathcal{U}$ with $/\!\!/\mathbf{x}-\mathbf{u}/\!\!/ \le \gamma$, we have

$$\begin{aligned} \|A\mathbf{x}\|_1 &\ge \|A\mathbf{u}\|_1 - \|A(\mathbf{x}-\mathbf{u})\|_1 \ge 1-\gamma-(1+\delta) \,/\!\!/\, \mathbf{x}-\mathbf{u}/\!\!/ \ge 1-\gamma-(1+\delta)\gamma \\ &\ge 1-3\gamma = 1-\delta. \end{aligned}$$

Thus, we obtain the lower estimate

$$\|A\mathbf{x}\|_1 \ge (1-\delta) \,/\!\!/\, \mathbf{x} \,/\!\!/ \qquad \text{for all } \mathbf{x} \in \mathbb{R}^S. \tag{11}$$

Since both upper and lower estimates (10) and (11) hold as soon as (8) holds, we obtain

$$\mathbf{P}\Big(\big| \|A\mathbf{x}\|_1 - /\!\!/\mathbf{x} \,/\!\!/ \,\big| > \delta \,/\!\!/\, \mathbf{x} \,/\!\!/ \quad \text{for some } \mathbf{x} \in \mathbb{R}^S \Big) \le 2\exp\Big(-\frac{\kappa\delta^2 m}{9} + \frac{6s}{\delta} \Big).$$

We now take into account that the set of s-sparse vectors is the union of $\binom{N}{s} \le (eN/s)^s$ spaces $\mathbb{R}^S$ to deduce, using a union bound, that

$$\begin{aligned} &\mathbf{P}\Big(\big| \|A\mathbf{x}\|_1 - /\!\!/\mathbf{x}/\!\!/ \big| > \delta \,/\!\!/\, \mathbf{x} \,/\!\!/ \quad \text{for some } s\text{-sparse } \mathbf{x} \in \mathbb{R}^N \Big) \\ &\qquad\qquad \le \binom{N}{s} 2\exp\Big(-\frac{\kappa\delta^2 m}{9} + \frac{6s}{\delta} \Big) \le 2\exp\Big(-\frac{\kappa\delta^2 m}{9} + \frac{6s}{\delta} + s\ln\Big(\frac{eN}{s}\Big) \Big) \\ &\qquad\qquad \le 2\exp\Big(-\frac{\kappa\delta^2 m}{9} + \frac{7s}{\delta}\ln\Big(\frac{eN}{s}\Big) \Big). \end{aligned}$$

By imposing, say,

$$\frac{7s}{\delta}\ln\Big(\frac{eN}{s}\Big) \le \frac{\kappa\delta^2 m}{18}, \qquad \text{i.e.,} \qquad m \ge \frac{126}{\kappa\delta^3} s \ln\Big(\frac{eN}{s}\Big),$$

we ensure that

$$\mathbf{P}\Big(\big| \|A\mathbf{x}\|_1 - /\!\!/\mathbf{x} \,/\!\!/ \,\big| > \delta \,/\!\!/\, \mathbf{x} \,/\!\!/ \quad \text{for some } s\text{-sparse } \mathbf{x} \in \mathbb{R}^N \Big) \le 2\exp\Big(-\frac{\kappa\delta^2 m}{18} \Big).$$

This is the desired result with $c_1 = \kappa/18$ and $c_2 = 126/\kappa$. ■

5 Sparse Recovery

In this section, we verify that the Modified Restricted Isometry Property (3) implies sparse recovery by ℓ_1-minimization.

Theorem 5.1 *Let* $/\!\!/ \cdot /\!\!/$ *be a norm on* $\mathbb{R}^N$ *satisfying*

$$c\,\|\mathbf{x}\|_2 \le /\!\!/\mathbf{x}/\!\!/ \le C\,\|\mathbf{x}\|_2 \qquad \text{for all } \mathbf{x} \in \mathbb{R}^N.$$

If a matrix $A \in \mathbb{R}^{m \times N}$ *has a Modified Restricted Isometry Constant*

$$\delta_{s+t}^{/\!\!/\cdot/\!\!/} < \frac{\sqrt{t/s} - C/c}{\sqrt{t/s} + C/c} \qquad \text{for some integer } t, \tag{12}$$

then any s*-sparse vector* $\mathbf{x} \in \mathbb{R}^N$ *is exactly recovered as a solution of* (P_1) *with* $\mathbf{y} = A\mathbf{x}$*.*

Proof. As is well known, see e.g. [15], it is necessary and sufficient to establish the null space property in the form

$$\|\mathbf{v}_S\|_1 < \|\mathbf{v}_{\overline{S}}\|_1 \quad \text{for all } \mathbf{v} \in \ker A \setminus \{0\} \text{ and all } S \subseteq \{1, \ldots, N\} \text{ with } \operatorname{card}(S) = s. \tag{13}$$

Given $\mathbf{v} \in \ker A \setminus \{0\}$, we notice that it is enough to prove the latter for an index set S_0 of s largest absolute-value components of $\mathbf{v}$. We partition the complement of S_0 in $\{1, \ldots, N\}$ as $\overline{S_0} = S_1 \cup S_2 \cup \ldots$, where

S_1 is an index set of t largest absolute-value components of $\mathbf{v}$ in $\overline{S_0}$,

S_2 is an index set of t next largest absolute-value components of $\mathbf{v}$ in $\overline{S_0}$,

etc. Setting $\delta_{s+t} := \delta_{s+t}^{/\!\!/ \cdot /\!\!/}$, we obtain from the Modified Restricted Isometry Property (3)

$$\begin{aligned}
/\!\!/ \mathbf{v}_{S_0} + \mathbf{v}_{S_1} /\!\!/ &\leq \frac{1}{1-\delta_{s+t}} \|A(\mathbf{v}_{S_0} + \mathbf{v}_{S_1})\|_1 = \frac{1}{1-\delta_{s+t}} \|A(-\sum_{k \geq 2} \mathbf{v}_{S_k})\|_1 \\
&\leq \frac{1}{1-\delta_{s+t}} \sum_{k \geq 2} \|A\mathbf{v}_{S_k}\|_1 \leq \frac{1+\delta_{s+t}}{1-\delta_{s+t}} \sum_{k \geq 2} /\!\!/ \mathbf{v}_{S_k} /\!\!/ \leq C\, \frac{1+\delta_{s+t}}{1-\delta_{s+t}} \sum_{k \geq 2} \|\mathbf{v}_{S_k}\|_2.
\end{aligned}$$

For $k \geq 2$, the inequalities $|v_i| \leq |v_j|$, $i \in S_k$, $j \in S_{k-1}$, averaged over j, raised to the power 2, and summed over i, yield

$$\|\mathbf{v}_{S_k}\|_2 \leq \frac{1}{\sqrt{t}} \|\mathbf{v}_{S_{k-1}}\|_1.$$

We therefore have

$$/\!\!/ \mathbf{v}_{S_0} + \mathbf{v}_{S_1} /\!\!/ \leq \frac{C}{\sqrt{t}} \frac{1+\delta_{s+t}}{1-\delta_{s+t}} \sum_{k \geq 2} \|\mathbf{v}_{S_{k-1}}\|_1 \leq \frac{C}{\sqrt{t}} \frac{1+\delta_{s+t}}{1-\delta_{s+t}} \|\mathbf{v}_{\overline{S_0}}\|_1. \tag{14}$$

Next, we observe that

$$\|\mathbf{v}_{S_0}\|_1 \leq \sqrt{s}\, \|\mathbf{v}_{S_0}\|_2 \leq \sqrt{s}\, \|\mathbf{v}_{S_0} + \mathbf{v}_{S_1}\|_2 \leq \frac{\sqrt{s}}{c}\, /\!\!/\, \mathbf{v}_{S_0} + \mathbf{v}_{S_1}\, /\!\!/\, . \tag{15}$$

Combining (14) and (15), we obtain

$$\|\mathbf{v}_{S_0}\|_1 \leq \frac{C}{c} \sqrt{\frac{s}{t}}\, \frac{1+\delta_{s+t}}{1-\delta_{s+t}} \|\mathbf{v}_{\overline{S_0}}\|_1.$$

The null space property (13) follows from $\big(C\sqrt{s}\,(1+\delta_{s+t})\big) \,/\, \big(c\sqrt{t}\,(1-\delta_{s+t})\big) < 1$, which is just a rewriting of Condition (12). The proof is now complete. ■

6 Main Theorem

We finally combine the results of the previous sections to prove that $m \times N$ pre-Gaussian random matrices allow for the reconstruction of all s-sparse vectors by ℓ_1-minimization with overwhelming probability provided $m \geq c\, s\, \ln(eN/s)$. Note that the distributions of the entries of the matrix need not be related, so long as they obey simple moment conditions, which are automatically fulfilled when the entries are identically distributed.

Theorem 6.1 *Suppose the entries of a matrix $A \in \mathbb{R}^{m \times N}$ are independent pre-Gaussian random variables satisfying*

$$\mathbf{E}\big(|a_{i,j}|\big) \geq \mu \qquad \text{and} \qquad \mathbf{E}\big(|a_{i,j}|^{2k}\big) \leq (2k)!\, \theta^{2k}, \quad k \geq 1.$$

Then, with probability at least

$$1 - 2\exp(-C_1 m),$$

any s-sparse vector $\mathbf{x} \in \mathbb{R}^N$ is exactly recovered as a solution of (P_1) with $\mathbf{y} = A\mathbf{x}$, provided

$$m \geq C_2\, s\, \ln(eN/s),$$

where the constants $C_1, C_2 > 0$ depend only on θ/μ.

Proof. Let $\nu_{i,j}$ denotes the centered probability measure associated to the entry $a_{i,j}$ and let $/\!\!/ \cdot /\!\!/$ be the norm defined in (2). According to Proposition 2.2, we have

$$m\, \frac{\mu}{\sqrt{8}}\, \|\mathbf{x}\|_2 \leq /\!\!/ \mathbf{x} /\!\!/ \leq m\, \sqrt{2}\, \theta \|\mathbf{x}\|_2 \qquad \text{for all } \mathbf{x} \in \mathbb{R}^N.$$

Theorem 5.1 then guarantees s-sparse recovery by ℓ_1-minimization as soon as

$$\delta_{s+t}^{/\!\!/ \cdot /\!\!/} < \frac{\sqrt{t/s} - 4\theta/\mu}{\sqrt{t/s} + 4\theta/\mu} \qquad \text{for some integer } t.$$

Let us choose an integer t such that $64\,(\theta/\mu)^2\, s < t \leq \big(64\,(\theta/\mu)^2 + 1\big)\, s$. Since then

$$\frac{\sqrt{t/s} - 4\theta/\mu}{\sqrt{t/s} + 4\theta/\mu} > \frac{8\theta/\mu - 4\theta/\mu}{8\theta/\mu + 4\theta/\mu} = \frac{1}{3},$$

s-sparse recovery by ℓ_1-minimization is guaranteed as soon as $\delta_{s+t}^{/\!\!/ \cdot /\!\!/} \leq 1/3$. According to Theorems 3.1 and 4.1 with $\kappa := 1/\big(128(\theta/\mu)^2 + 16(\theta/\mu)\big)$ and $\delta = 1/3$, this is guaranteed with probability at least

$$1 - 2\exp\Big(-\frac{c_1 m}{9}\Big), \qquad c_1 = \frac{\kappa}{18},$$

provided

$$m \geq 27\, c_2\, (s + t)\, \ln\Big(\frac{eN}{s+t}\Big), \qquad c_2 = \frac{126}{\kappa}.$$

This holds as soon as

$$m \geq 27\, c_2\, \big(64\,(\theta/\mu)^2 + 2\big)\, s\, \ln\Big(\frac{eN}{s}\Big).$$

The constants of the theorem are explicitly given by $C_1 = 1/\big(20736\,(\theta/\mu)^2 + 2592\,(\theta/\mu)\big)$ and $C_2 = \big(435456\,(\theta/\mu)^2 + 54432\,(\theta/\mu)\big)\big(64\,(\theta/\mu)^2 + 2\big)$. $\blacksquare$

References

[1] Adamczak, R., A. E. Litvak, A. Pajor, and N. Tomczak-Jaegermann, Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling, Preprint.

[2] Baraniuk, R., M. Davenport, R. DeVore, and M. B. Wakin, A simple proof of the restricted isometry property for random matrices, Constructive Approximation, 28(2008), 253–263.

[3] Berinde, R., A. C. Gilbert, P. Indyk, H. Karloff, and M. J. Strauss, Combining geometry and combinatorics: A unified approach to sparse signal recovery, Preprint.

[4] Bourgain, J., M. Meyer, V. Milman, and A. Pajor, On a geometric inequality, Geometric aspects of functional analysis (1986/87), 271–282, Lecture Notes in Math., 1317, Springer, Berlin, 1988.

[5] Buldygin, V. V. and Yu. V. Kozachenko, Metric Characterization of Random Variables and Random Processes, AMS Publication, Providence, 2000.

[6] Cai, T. T., L. Wang, and G. Xu, Shifting Inequality and Recovery of Sparse Signals, IEEE Transactions on Signal Processing, To appear.

[7] Candès, E. J., The restricted isometry property and its implications for compressed sensing, Comptes Rendus de l'Académie des Sciences, Série I, 346(2008), 589–592.

[8] Candès, E. and T. Tao, Decoding by linear programing, IEEE Trans. Inf. Theory 51(2005), 4203–4215.

[9] Donoho, D., For most large underdetermined systems of linear equations, the minimal ℓ_1-norm solution is also the sparsest solution, Comm. Pure Appl. Math., 59(2006), 797–829.

[10] Foucart, S., A note on guaranteed sparse recovery via ℓ_1-minimization, Applied and Comput. Harmonic Analysis, To appear.

[11] Foucart, S. and M. J. Lai, Sparsest Solutions of Underdetermined Linear Systems via ℓ_q-minimization for $0 < q \leq 1$, Applied and Comput. Harmonic Analysis, 26(2009), 395–407.

[12] Foucart, S., A. Pajor, H. Rauhut, and T. Ullrich, The Gelfand widths of ℓ_p-balls for $0 < p \leq 1$. Preprint.

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

[14] Gluskin, E. and V. Milman, Geometric probability and random cotype 2, Geometric aspects of functional analysis, 123–138, Lecture Notes in Math., 1850, Springer, Berlin, 2004.

[15] Gribonval, R. and M. Nielsen, Sparse representations in unions of bases, IEEE Trans. Inform. Theory, 49(2003), 3320–3325.

[16] Haagerup, U., The best constants in the Khinchine inequality, Studia Math., 70(1982), 231–283.

[17] Ledoux, M. and M. Talagrand, Probability in Banach Spaces, Springer–Verlag, 1991.

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

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

[20] Pisier, G., *The Volume of Convex Bodies and Banach Space Geometry*, Cambridge University Press, 1999.

[21] Rudelson, M. and R. Vershynin, On sparse reconstruction from Fourier and Gaussian measurements, Comm. Pure Appl. Math., 61(2008), 1025–1045.

[22] van der Vaart, A. W. and J. A. Wellner, Weak convergence and empirical processes, Springer, 1996.