On the gap between restricted isometry properties and sparse recovery conditions

Sjoerd Dirksen, Guillaume Lecué, and Holger Rauhut

***Abstract*—We consider the problem of recovering sparse vectors from underdetermined linear measurements via ℓ_p-constrained basis pursuit. Previous analyses of this problem based on generalized restricted isometry properties have suggested that two phenomena occur if $p \neq 2$. First, one may need substantially more than $s\log(en/s)$ measurements (optimal for $p = 2$) for uniform recovery of all s-sparse vectors. Second, the matrix that achieves recovery with the optimal number of measurements may not be Gaussian (as for $p = 2$). We present a new, direct analysis which shows that in fact neither of these phenomena occur. Via a suitable version of the null space property we show that a standard Gaussian matrix provides ℓ_q/ℓ_1-recovery guarantees for ℓ_p-constrained basis pursuit in the optimal measurement regime. Our result extends to several heavier-tailed measurement matrices. As an application, we show that one can obtain a consistent reconstruction from uniform scalar quantized measurements in the optimal measurement regime.**

***Index Terms*—Restricted isometry property, compressive sensing, ℓ_p-constrained basis pursuit, Gaussian random matrix, quantized compressive sensing.**

I. Introduction

COMPRESSIVE sensing [1], [2], [3] has established itself in the recent years as a rapidly growing research area with various promising signal and image processing applications, and which has triggered many developments on the theoretical side. The theory predicts that (approximately) sparse signals can be accurately recovered from incomplete and perturbed linear measurements. The measurement process is described by a measurement matrix $A \in \mathbb{C}^{m\times n}$ with $m < n$. While the naïve reconstruction approach via ℓ_0-minimization is NP-hard [4], several tractable recovery methods have been proposed including basis pursuit (ℓ_1-minimization), iterative hard thresholding and greedy methods. For all these methods rigorous recovery guarantees have been shown, see [3] for details and further references.

The restricted isometry property (RIP) is a well-established tool to analyze the performance of sparse recovery methods. The standard version defines the restricted isometry constant of order s of $A \in \mathbb{C}^{m\times n}$ as the smallest number δ_s such that

$$(1-\delta_s)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1+\delta_s)\|x\|_2^2 \text{ for all } x \in \Sigma_s, \quad (1)$$

where Σ_s is the set of all s-sparse vectors in $\mathbb{C}^n$ and $\|\cdot\|_2$ denotes the usual ℓ_2-norm. If δ_s is sufficiently small we say that A satisfies the RIP. If $\delta_s < \delta_*$ for some suitably small δ_*, then given measurements $y = A\hat{x} + e$ with $\|e\|_2 \leq \varepsilon$, the ℓ_2-constrained ℓ_1-minimization program (also known as basis pursuit denoising)

$$\min_{z\in\mathbb{C}^n} \|z\|_1 \quad \text{subject to} \quad \|Az - y\|_2 \leq \varepsilon$$

recovers a vector $x^\sharp$ which satisfies

$$\|\hat{x} - x^\sharp\|_2 \lesssim s^{-1/2}\sigma_s(\hat{x})_1 + \varepsilon, \quad (2)$$

where $\sigma_s(\hat{x})_1 = \inf_{\|z\|_0\leq s} \|\hat{x} - z\|_1$ is the error of best s-term approximation to $\hat{x}$ in ℓ_1. An (appropriately scaled) Gaussian random matrix satisfies the RIP with high probability provided that

$$m \geq Cs\log(en/s), \quad (3)$$

where $C > 0$ is an absolute constant. This bound is optimal, see also below.

In certain cases it is of interest to measure the level of noise in ℓ_p-norms with $1 \leq p \leq \infty$ different from 2 and to study the corresponding ℓ_p-constrained basis pursuit denoising program

$$\min_{z\in\mathbb{C}^n} \|z\|_1 \quad \text{subject to} \quad \|y - Az\|_p \leq \varepsilon. \quad (\text{BPDN}_p)$$

The case $p = \infty$ appears, for instance, in quantized compressed sensing [5], where ℓ_∞-constrained basis pursuit can ensure consistent reconstruction, see also Section IV below. The program for $p = 1$ is more robust to outliers than standard basis pursuit denoising, since the ℓ_1-norm gives less weight to large residuals than the ℓ_2-norm (see also the discussion in [6, Section 6.1.2]). Also, when considering random measurement noise, different values of p are appropriate depending on the distribution of the noise (see e.g. [5],[7]). For example, $p = 1$ is well-suited for double-exponential noise, whereas $p = 2$ is appropriate for Gaussian noise.

Previous attempts in analyzing (BPDN_p) have used restricted isometry properties of the form

$$c\|x\|_q \leq \|Ax\|_p \leq C\|x\|_q, \quad \text{for all } x \in \Sigma_s. \quad (\text{RIP}_{p,q})$$

It is part of the folklore in compressive sensing that ($\text{RIP}_{p,q}$) implies stable and robust recovery via (BPDN_p), with an ℓ_q-bound on the reconstruction error (see e.g. [5], [8] for special cases). Unfortunately, all available results on the number of required measurements for Gaussian and other random matrices ensuring ($\text{RIP}_{p,q}$) scale significantly worse than (3) when $p \neq 1, 2$. For certain values of p and q, there are even negative results available which state that no matrix whatsoever can satisfy ($\text{RIP}_{p,q}$) in the optimal parameter regime (3). A more detailed overview is given below.

S. Dirksen and H. Rauhut are with Lehrstuhl C für Mathematik (Analysis), RWTH Aachen University, Pontdriesch 10, 52062 Aachen, Germany, email: {dirksen,rauhut}@mathc.rwth-aachen.de.

G. Lecué is with CNRS and Ecole Polytechnique, CMAP, Route de Saclay, 91120 Palaiseau, France, email: guillaume.lecue@cmap.polytechnique.fr

The purpose of this short note is to illuminate the discrepancy between on the one hand, the requirements needed for a matrix to satisfy a restricted isometry property of the form (RIP$_{p,q}$) and on the other hand, the conditions under which one can stably and robustly recover any s-sparse (or approximately s-sparse) vector $\hat{x} \in \mathbb{C}^n$ from noisy linear measurements $y = A\hat{x}+e$ via the generalized basis pursuit denoising program (BPDN$_p$). Our results show that a study of the statistical properties of (BPDN$_p$) via the ℓ_q-robust null space property yields better results than via (RIP$_{p,q}$), both in terms of the required number of measurements as well as the allowed distribution of the random measurements. In particular, one can achieve stable and robust reconstruction with Gaussian random matrices in the optimal parameter regime (3) for any $1 \leq p \leq \infty$. This result extends to various random matrices with heavier-tailed entries such as exponential matrices, see Section V for more information. Our proof relies on the small ball method developed in [9], [10], [11], [12]. In the last section we illustrate our results with numerical simulations.

Notation. The usual ℓ_p-norm on $\mathbb{C}^n$ is denoted by $\|x\|_p = (\sum_{j=1}^n |x_j|^p)^{1/p}$ for $1 \leq p < \infty$ and $\|x\|_\infty = \max_{j=1,\dots,n} |x_j|$. We let $B_{\ell_p^n}$ and $S_{\ell_p^n}$ denote the associated unit ball and unit sphere, respectively. The expression $\|x\|_0 := \#\{j : x_j \neq 0\}$ counts the number of nonzero coefficients of $x \in \mathbb{C}^n$. We use x^* to denote the nonincreasing rearrangement of x. The expectation of a random variable Z is written $\mathbb{E}Z$ and the probability of an event E is denoted by $\mathbb{P}(E)$. The L_p-norm of a measurable function f with respect to a measure μ is denoted by $\|f\|_{L_p(\mu)}$. A Rademacher random variable ε satisfies $\mathbb{P}(\varepsilon = 1) = \mathbb{P}(\varepsilon = -1) = 1/2$ and a Rademacher sequence is a sequence of independent Rademacher random variables. For $t \in \mathbb{R}$, $\lfloor t \rfloor$ is the largest integer smaller than t and $\lceil t \rceil$ is the smallest integer larger than t. Finally, we write $A \lesssim B$ if $A \leq cB$ for a universal constant $c > 0$.

II. The relation between (RIP$_{p,q}$) and (BPDN$_p$)

In this section we summarize the known results on (RIP$_{p,q}$) for $p, q \geq 1$ related to sparse recovery via (BPDN$_p$). Let us mention that outside of this setting, there is work available on (RIP$_{p,2}$) for $p < 1$, which is motivated by the analysis of exact sparse recovery via (non-convex) ℓ_p-minimization (see e.g. [13], [14]). Also, as part of a construction of an isomorphic embedding of ℓ_q^n into ℓ_p^{m+n}, the authors of [15] construct a random matrix satisfying (RIP$_{p,q}$) for $0 < p < q < 2$, $p \leq 1$ (see [15, Corollary 12]).

As is well known and already described above, the property (RIP$_{2,2}$) was introduced in compressed sensing by Candès and Tao in [16], [17]. They showed that if A is an $m \times n$ standard Gaussian matrix, i.e., it has i.i.d. standard Gaussian entries, then $m^{-1/2}A$ has restricted isometry constant δ_s smaller than δ_* with probability $1-\eta$ if $m \gtrsim \delta_*^{-2}(s\log(en/s)+\log(\eta^{-1}))$. As a consequence, if δ_* is smaller than a certain fixed threshold and $\|y - A\hat{x}\|_2 \leq \varepsilon$, then any minimizer $x^\#$ for (BPDN$_2$) satisfies an ℓ_q/ℓ_1-guarantee of the form

$$\|\hat{x} - x^\#\|_q \lesssim s^{1/q-1}\sigma_s(\hat{x})_1 + s^{1/q-1/2}m^{-1/2}\varepsilon$$

for any $1 \leq q \leq 2$ (see also [3, Theorem 9.13] for more details). In particular, if $\hat{x}$ is exactly s-sparse (so $\sigma_s(\hat{x})_1 = 0$) and $\varepsilon = 0$, then $\hat{x}$ can be reconstructed exactly. Conversely, it is known that $m \gtrsim s\log(n/s)$ measurements are also necessary for exact reconstruction of all s-sparse vectors (see e.g. [3, Theorem 10.11]).

A very similar connection exists between (RIP$_{1,1}$) and (BPDN$_1$) [8]. Indeed, consider the adjacency matrix of a random left d-regular bipartite graph with n left vertices and m right vertices. This $m \times n$ matrix can be constructed by, for each column independently, assigning the value 1 to d entries selected uniformly at random without replacement and setting the remaining entries to 0. The resulting matrix A satisfies with probability $1-\eta$ an (RIP$_{1,1}$) condition of the form

$$(1-\delta)^{1/2}\|x\|_1 \leq d^{-1}\|Ax\|_1 \leq \|x\|_1 \qquad \text{for all } x \in \Sigma_s,$$

provided that $d = \lceil \delta^{-1}\log(en/(s\eta)) \rceil$ and $m \geq c_\delta s\log(en/(s\eta))$. As a consequence [8, Theorem 12], if $\|y - A\hat{x}\|_1 \leq \varepsilon$ then any minimizer $x^\#$ of (BPDN$_1$) satisfies the ℓ_1/ℓ_1 guarantee

$$\|\hat{x} - x^\#\|_1 \leq C(\delta)\Big(\sigma_s(\hat{x})_1 + \frac{\varepsilon}{d}\Big),$$

where $C(\delta) = O((1-2\delta)^{-1})$ for $\delta \uparrow 1/2$.

Interestingly, the rescaled adjacency matrix $d^{-1}A$ does not satisfy (RIP$_{2,2}$). In fact, any (RIP$_{2,2}$)-matrix with binary entries must satisfy $m \geq s^2\log(en/s)$ [18, Theorem 4.6.1]. Conversely, if A is standard Gaussian, then $m^{-1/2}A$ cannot satisfy (RIP$_{1,1}$) for $m \sim s\log(en/s)$ [8]. To see this, one can consider $x = e_1$, $\tilde{x} = s^{-1}\sum_{i=1}^s e_i$, where the e_i denote the standard basis vectors. Then $\|x\|_1 = \|\tilde{x}\|_1 = 1$, but $\|Ax\|_1 \sim \sqrt{s}\|A\tilde{x}\|_1$.

The two positive results for $p = q = 2$ and $p = q = 1$ have triggered further research on (BPDN$_p$) via restricted isometry properties. In [5] it was shown that a standard $m \times n$ Gaussian matrix with

$$m \gtrsim \Big(\delta^{-2}s\log(en/(s\delta)) + \delta^{-2}\log(\eta^{-1})\Big)^{p/2} + (p-1)2^{p-1} \tag{4}$$

satisfies an (RIP$_{p,2}$) property for $2 \leq p < \infty$ of the form

$$(1-\delta)^{1/2}\|x\|_2 \leq \mu_p^{-1}\|Ax\|_p \leq (1+\delta)^{1/2}\|x\|_2 \text{ for all } x \in \Sigma_s,$$

where $\mu_p = \mathbb{E}\|G\|_p$ and G is a standard m-dimensional Gaussian random vector. If A satisfies the latter property for sparsity levels $s, 2s, 3s$ with constants $\delta_s, \delta_{2s}, \delta_{3s}$ small enough (see [19, Theorem 1] for a precise statement), then for all $\hat{x} \in \mathbb{C}^n$ with $\|y - A\hat{x}\|_p \leq \varepsilon$, any minimizer $x^\sharp$ of (BPDN$_p$) satisfies an ℓ_2/ℓ_1-guarantee

$$\|\hat{x} - x^\sharp\|_2 \lesssim s^{-1/2}\sigma_s(\hat{x})_1 + \frac{\varepsilon}{\mu_p}.$$

In [20] it is shown that the $m \times n$ adjacency matrix A of a random left d-regular bipartite graph with n left vertices and m right vertices with high probability satisfies an (RIP$_{p,p}$) condition in the form

$$(1-\delta)\|x\|_p^p \leq d^{-1}\|Ax\|_p^p \leq (1+\delta)\|x\|_p^p \qquad \text{for all } x \in \Sigma_s,$$

provided that, in the case $1 < p < 2$,

$$m \geq C_p(s^p\delta^{-2} + s^{4-2/p-p}\delta^{-2/(p-1)})\log n,$$
$$d \geq \tilde{C}_p(\delta^{-1}s^{p-1} + s^{(p-1)/p}\delta^{-1/(p-1)})\log n,$$

where $C_p, \tilde{C}_p$ are singular for $p \downarrow 1$ and $p \uparrow 2$, or in the case $2 < p < \infty$,

$$m \geq p^{Cp}\delta^{-2}s^p \log^{p-1}(n),$$
$$d \geq p^{Cp}\delta^{-1}s^{p-1} \log^{p-1}(n).$$

If A satisfies an (RIP$_{p,p}$)-property for $p > 1$, then one can recover all $\hat{x} \in \mathbb{C}^n$ with $\|y - A\hat{x}\|_p \leq \varepsilon$ via (BPDN$_p$) with an ℓ_p/ℓ_1-guarantee of the form

$$\|\hat{x} - x^\sharp\|_p \lesssim s^{1/p-1}\sigma_s(\hat{x})_1 + \varepsilon,$$

see [20, Theorem A.6] for a more precise statement. Interestingly, [20] also proved a lower bound on m assuming that the $m \times n$ matrix satisfies (RIP$_{p,p}$). Their result [20, Theorem 4.1] essentially shows that one needs at least $m \gtrsim s^p$ measurements for $p \neq 2$, so that the case $p = 2$ should be considered a singularity. A straightforward modification of their argument shows that to satisfy (RIP$_{p,2}$) one needs at least $m \gtrsim s^{p/2}$, so that also the result in [5] (cf. (4)) cannot be improved significantly. We leave the verification of this implication to the interested reader.

To summarize, two important phenomena occur when moving away from the familiar (RIP$_{2,2}$). First, one may need to consider different random matrix constructions to satisfy (RIP$_{p,q}$) with the optimal number of measurements. Second, the optimal scaling of the number of measurements in terms of the signal sparsity may dramatically worsen, especially for $p > 2$.

III. Sparse recovery via (BPDN$_p$): improved results

One might think that the two phenomena concerning the properties (RIP$_{p,q}$) for $p \neq 2$ mentioned above, may carry over to recovery results via (BPDN$_p$) (see e.g. [20], [5]), in particular, that the minimal required number of measurements depends significantly worse than linear on the sparsity. We will now show that rather the contrary is true: the scaling in terms of the sparsity generally does not worsen if $p \neq 2$ and, moreover, the optimal recovery results are realized by a standard Gaussian matrix.

Let us note that earlier work already identified a looseness in the relation between the classical (RIP$_{2,2}$) and (BPDN$_2$). For example, if A has independent, isotropic, log-concave rows, then (1) is satisfied with high probability if $m \geq c(\delta)s\log^2(en/s)$ [21], and the square in the log-factor cannot be removed [22, Proposition 5.5]. On the other hand, this matrix still satisfies, with high probability, the exact reconstruction property for s-sparse vectors via ℓ_1-minimization in the optimal measurement regime $m \simeq s\log(en/s)$ ([23, Theorem 7.3] – see also [24] for the special case of measurement matrices with i.i.d. Weibull entries). More recently, near-matching necessary and sufficient conditions on the moments of the i.i.d. entries of a matrix to satisfy the exact reconstruction property (and more generally, stable and robust recovery via (BPDN$_2$)) in this regime were recently derived by the second-named author and Mendelson [25]. We recover as a special case a variation of this (sufficient) result, see Corollary V.3 below. Such a result cannot be proved via an RIP-based analysis since the right-hand side of (RIP$_{2,2}$), i.e.,

$$\|Ax\|_2 \leq C\|x\|_2 \quad \text{for all } x \in \Sigma_s$$

requires either strong concentration properties or a larger number of measurements m than the optimal number $s\log(en/s)$ (see the discussion in [25] and Section VI for more details).

For our analysis we let $X_1, \ldots, X_m$ be i.i.d. copies of a random vector X in $\mathbb{C}^n$, which is defined on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Let P_m be the associated empirical measure

$$P_m = \frac{1}{m}\sum_{i=1}^m \delta_{X_i}.$$

The following observation follows immediately from the proof of Theorem 2.1 in [9], by replacing the "Chebyshev" bound

$$\|f\|^2_{L_2(P_m)} \geq u^2 P_m(|f| \geq u)$$

by

$$\|f\|^p_{L_p(P_m)} \geq u^p P_m(|f| \geq u).$$

Lemma III.1. *Fix $1 \leq p < \infty$. Let $\mathcal{F}$ be a class of functions from $\mathbb{C}^n$ into $\mathbb{C}$. Consider*

$$Q_{\mathcal{F}}(u) = \inf_{f\in\mathcal{F}} \mathbb{P}(|f(X)| \geq u)$$

and

$$R_m(\mathcal{F}) = \mathbb{E}\sup_{f\in\mathcal{F}}\Big|\frac{1}{m}\sum_{i=1}^m \varepsilon_i f(X_i)\Big|,$$

where $(\varepsilon_i)_{i\geq 1}$ is a Rademacher sequence. Let $u > 0$ and $t > 0$, then, with probability at least $1 - 2e^{-2t^2}$,

$$\inf_{f\in\mathcal{F}} \frac{1}{m}\sum_{i=1}^m |f(X_i)|^p \geq u^p\Big(Q_{\mathcal{F}}(2u) - \frac{4}{u}R_m(\mathcal{F}) - \frac{t}{\sqrt{m}}\Big).$$

Consider the following sparse recovery problem: we take m noisy linear measurements of an (approximately) s-sparse signal $\hat{x}$, i.e., we observe $y = A\hat{x}+e$ where $A \in \mathbb{C}^{m\times n}$ and we suppose that the noise satisfies $\|e\|_p \leq \varepsilon$. We aim to recover $\hat{x}$ from y via (BPDN$_p$). For the analysis we recall the following standard notion (see for instance [3, Definition 4.21]). Given $q \geq 1$, we say that A satisfies the ℓ_q-robust null space property of order s with constants $0 < \rho < 1$ and $\tau > 0$ with respect to a norm $\|\cdot\|$ if for any set $S \subset [n]$ with $|S| \leq s$ and any $x \in \mathbb{C}^n$,

$$\|x_S\|_q \leq \frac{\rho}{s^{1-1/q}}\|x_{S^c}\|_1 + \tau\|Ax\|.$$

If A has this property, then any solution $x^\#$ to

$$\min_{z\in\mathbb{C}^n} \|z\|_1 \quad \text{subject to} \quad \|y - Az\| \leq \varepsilon$$

satisfies, for any $1 \leq r \leq q$, the reconstruction error bound

$$\|\hat{x} - x^\#\|_r \leq C_\rho s^{1/r-1}\sigma_s(\hat{x})_1 + \tau D_\rho s^{1/r-1/q}\varepsilon,$$

with $C_\rho = (1+\rho)^2/(1-\rho)$ and $D_\rho = (3+\rho)/(1-\rho)$ when $\|e\| \leq \varepsilon$ [3, Theorem 4.25].

To analyze ℓ_q-robust null space properties, we introduce the cone

$$T^q_{\rho,s} = $$
$$\big\{x \in \mathbb{C}^n : \exists S \subset [n],\ |S| = s : \|x_S\|_q \geq \frac{\rho}{s^{1-1/q}}\|x_{S^c}\|_1\big\}.$$

Note that $T^q_{\rho,s}$ contains Σ_s. We use the following observation.

Lemma III.2. *Fix $1 \le q < \infty$. Set*

$$\Sigma^q_s := \{x \in \mathbb{C}^n \ : \ \|x\|_0 \le s, \ \|x\|_q = 1\}$$

and let D^q_s be its convex hull. Then D^q_s is the unit ball with respect to the norm

$$\|x\|_{D^q_s} := \sum_{\ell=1}^{\lceil n/s \rceil} \Big(\sum_{i \in I_\ell} x_i^{*q} \Big)^{1/q},$$

where $I_1, \dots, I_{\lceil n/s \rceil}$ form a uniform partition of $[n]$, i.e.,

$$I_\ell = \begin{cases} \{s(\ell-1)+1, \dots, s\ell\}, & \ell = 1, \dots, \lceil n/s \rceil - 1, \\ \{s(\lceil n/s \rceil - 1)+1, \dots, n\}, & \ell = \lceil n/s \rceil, \end{cases}$$

and x^* is the nonincreasing rearrangement of x. As a consequence,

$$T^q_{\rho,s} \cap B_{\ell^n_q} \subset (2+\rho^{-1}) D^q_s.$$

Proof. We proceed by making straightforward modifications to the proof of [26, Lemma 3] (see also [27, Lemma 4.5] or [28]), which corresponds to the case $q = 2$.

A vector $x \in D^q_s$ can be represented as $x = \sum_i \alpha_i x_i$ with $\alpha_i \ge 0$, $\sum_i \alpha_i = 1$ and $x_i \in S_{\ell^n_q}$, $\|x_i\|_0 \le s$. In particular, $\|x_i\|_{D^q_s} = \|x_i\|_q = 1$. By the triangle inequality

$$\|x\|_{D^q_s} \le \sum_i \alpha_i \|x_i\|_{D^q_s} = \sum_i \alpha_i = 1,$$

so D^q_s is contained in the $\|\cdot\|_{D^q_s}$-unit ball. To prove the reverse inclusion, suppose that $\|x\|_{D^q_s} \le 1$. We partition the index set $[n]$ into subsets S_1, S_2, ... of size s, such that S_1 corresponds to the indices of the s largest entries of x, S_2 to the next s ones, etc. Set $\alpha_i = \|x_{S_i}\|_q$. Then x can be written as

$$x = \sum_{i : \alpha_i \neq 0} \alpha_i (\alpha_i^{-1} x_{S_i}),$$

where

$$\sum_{i : \alpha_i \neq 0} \alpha_i = \sum_i \|x_{S_i}\|_q = \|x\|_{D^q_s} \le 1.$$

Clearly, for any $\alpha_i \neq 0$, $\|\alpha_i^{-1} x_{S_i}\|_q = 1$ and $\|\alpha_i^{-1} x_{S_i}\|_0 \le s$, so $x \in D^q_s$.

To prove the second statement, fix $x \in T^q_{\rho,s} \cap B_{\ell^n_q}$ and write

$$\|x\|_{D^q_s} = \Big(\sum_{i \in I_1} x_i^{*q} \Big)^{1/q} + \Big(\sum_{i \in I_2} x_i^{*q} \Big)^{1/q} + \sum_{\ell \ge 3} \Big(\sum_{i \in I_\ell} x_i^{*q} \Big)^{1/q}. \tag{5}$$

To bound the last term, note that for each $i \in I_\ell$, $\ell \ge 3$,

$$x_i^* \le \frac{1}{s} \sum_{j \in I_{\ell-1}} x_j^* \quad \text{and} \quad \Big(\sum_{i \in I_\ell} x_i^{*q} \Big)^{1/q} \le \frac{1}{s^{1-1/q}} \sum_{j \in I_{\ell-1}} x_j^*.$$

Summing up over $\ell \ge 3$ yields

$$\sum_{\ell \ge 3} \Big(\sum_{i \in I_\ell} x_i^{*q} \Big)^{1/q} \le \frac{1}{s^{1-1/q}} \sum_{\ell \ge 2} \sum_{j \in I_\ell} x_j^*.$$

Since $x \in T^q_{\rho,s} \cap B_{\ell^n_q}$, there is an $S \subset [n]$ with $|S| = s$, such that $\|x_S\|_q \ge \frac{\rho}{s^{1-1/q}} \|x_{S^c}\|_1$. Therefore,

$$\begin{aligned} \sum_{\ell \ge 2} \sum_{i \in I_\ell} x_i^* &\le \|x_{S^c}\|_1 \le \frac{s^{1-1/q}}{\rho} \|x_S\|_q \\ &\le \frac{s^{1-1/q}}{\rho} \Big(\sum_{i \in I_1} x_i^{*q} \Big)^{1/q}, \end{aligned}$$

where we used that in the worst case S consists of indices corresponding to s largest absolute coefficients of x. It follows that

$$\sum_{\ell \ge 3} \Big(\sum_{i \in I_\ell} x_i^{*q} \Big)^{1/q} \le \frac{1}{\rho} \Big(\sum_{i \in I_1} x_i^{*q} \Big)^{1/q}.$$

Since $\|x\|_q \le 1$, (5) implies that $\|x\|_{D^q_s} \le 2 + \rho^{-1}$. □

We are now prepared to prove the main result of this article. To keep our exposition accessible, we first consider the special case of a standard Gaussian random matrix, i.e., a matrix with independent normally distributed entries with mean zero and variance one. In Section V we generalize our result to a wider class of random matrices.

Theorem III.3. *Let A be an $m \times n$ standard Gaussian matrix. Fix $1 \le p \le \infty$, $q \ge 2$ and $0 < \eta < 1$. Suppose that*

$$m \gtrsim s^{2-2/q} \log(en/s) + \log(\eta^{-1}). \tag{6}$$

Then, with probability exceeding $1 - \eta$ the following holds: for any $\hat{x} \in \mathbb{C}^n$ and $y = A\hat{x} + e$, where $\|e\|_p \le \varepsilon$, any solution $x^\#$ to (BPDN$_p$) satisfies

$$\|\hat{x} - x^\#\|_r \lesssim s^{1/r-1} \sigma_s(\hat{x})_1 + s^{1/r-1/q} \frac{\varepsilon}{m^{1/p}},$$

for any $1 \le r \le q$.

Remark III.4. The most interesting case in the above theorem is $q = 2$. Then the optimal scaling $m \ge Cs \log(en/s)$ implies that with high probability we obtain the error bound

$$\|\hat{x} - x^\#\|_2 \lesssim s^{-1/2} \sigma_s(\hat{x})_1 + m^{-1/p} \varepsilon$$

for reconstruction via ℓ_p-constrained basis pursuit.

For $q > 2$ the scaling (6) of m in terms of the sparsity is near-optimal. Indeed, it is known [29, p. 213] that for $q > 2$ and $m \le n - 1$, the Gelfand width of $B_{\ell^n_1}$ in ℓ^n_q satisfies

$$d^m(B_{\ell^n_1}, \ell^n_q) \ge d^m(B_{\ell^n_1}, \ell^n_\infty) \ge c m^{-1/2}.$$

Thus, if we want to satisfy $\|\hat{x} - x^\#\|_q \lesssim s^{1/q-1} \sigma_s(\hat{x})_1$ for all $\hat{x} \in \mathbb{C}^n$, then it is necessary (cf. [3, Theorem 10.4]) that $m^{-1/2} \lesssim s^{1/q-1}$ or $m \gtrsim s^{2-2/q}$. Thus, up to possibly a logarithmic factor we cannot improve the scaling of m in terms of the sparsity in Theorem III.3.

Proof of Theorem III.3. Suppose first that $p < \infty$. As was noted before, it suffices to show that with probability at least $1 - \eta$ the ℓ_q-robust null space property of order s holds with respect to ℓ^n_p-norm, with parameters ρ and $\tau / m^{1/p}$ for some $0 < \rho < 1$ and $\tau > 0$. Let us first observe that it suffices to show that

$$\mathbb{P}\Big(\inf_{x \in T^q_{\rho,s} \cap S_{\ell^n_q}} \|Ax\|_p \ge \frac{m^{1/p}}{\tau} \Big) \ge 1 - \eta. \tag{7}$$

Indeed, if this is true, then with probability at least $1-\eta$ the following holds: if $x \in \mathbb{C}^n$ satisfies $\|Ax\|_p < (m^{1/p}/\tau)\|x\|_q$ then $x/\|x\|_q$ is not in $T^q_{\rho,s}$. Therefore, for any $S \subset [n]$ with $|S| \le s$,

$$\|x_S\|_q \le \frac{\rho}{s^{1-1/q}}\|x_{S^c}\|_1 \le \frac{\rho}{s^{1-1/q}}\|x_{S^c}\|_1 + \frac{\tau}{m^{1/p}}\|Ax\|_p.$$

On the other hand, if $x \in \mathbb{C}^n$ satisfies $\|Ax\|_p \ge (m^{1/p}/\tau)\|x\|_q$, then trivially

$$\|x_S\|_q \le \|x\|_q \le \frac{\rho}{s^{1-1/q}}\|x_{S^c}\|_1 + \frac{\tau}{m^{1/p}}\|Ax\|_p.$$

To prove (7), we write

$$\inf_{x\in T^q_{\rho,s}\cap S_{\ell^n_q}} \frac{\|Ax\|_p}{m^{1/p}} = \inf_{x\in T^q_{\rho,s}\cap S^{n-1}_q} \Big(\frac{1}{m}\sum_{i=1}^m |\langle X_i, x\rangle|^p\Big)^{1/p},$$

where X_i denotes the i-th row of A. To apply Lemma III.1, we estimate the small ball probability $Q_{\mathcal{F}}$ and the expected Rademacher supremum $R_m(\mathcal{F})$ for the set of linear functions

$$\mathcal{F} = \{\langle\cdot, x\rangle \ : \ x \in T^q_{\rho,s} \cap S_{\ell^n_q}\}.$$

Let $V = m^{-1/2}\sum_{i=1}^m \varepsilon_i X_i$, then by Lemma III.2,

$$\begin{aligned} R_m(\mathcal{F}) &= m^{-1/2}\mathbb{E}\sup_{x\in T^q_{\rho,s}\cap S_{\ell^n_q}} \langle V, x\rangle \\ &\le (2+\rho^{-1})m^{-1/2}\mathbb{E}\sup_{x\in D^q_s}\langle V, x\rangle \\ &= (2+\rho^{-1})m^{-1/2}\mathbb{E}\sup_{x\in \Sigma^q_s}\langle V, x\rangle, \end{aligned}$$

as D^q_s is the convex hull of Σ^q_s. Since any $x \in \Sigma^q_s$ satisfies $\|x\|_2 \le s^{1/2-1/q}\|x\|_q = s^{1/2-1/q}$,

$$R_m(\mathcal{F}) \le s^{1/2-1/q}(2+\rho^{-1})m^{-1/2}\mathbb{E}\sup_{x\in\Sigma^2_s}\langle V, x\rangle.$$

Since $X_1, \dots, X_m$ are independent standard Gaussian vectors, so is V. Thus,

$$\mathbb{E}\sup_{x\in\Sigma^2_s}\langle V, x\rangle = w(\Sigma^2_s),$$

the Gaussian width of Σ^2_s. It is known that

$$w(\Sigma^2_s) \le \sqrt{2s\log(en/s)} + \sqrt{s},$$

see e.g. [26, Lemma 4], and we can conclude that

$$R_m(\mathcal{F}) \le cs^{1-1/q}(2+\rho^{-1})m^{-1/2}\sqrt{\log(en/s)}.$$

To estimate the small ball probability, note that, since $\|x\|_q \le \|x\|_2$, for any $x \in S_{\ell^n_q}$,

$$\begin{aligned} \mathbb{P}(|\langle X_i, x\rangle| \ge u) &= \mathbb{P}\Big(\Big|\Big\langle X_i, \frac{x}{\|x\|_2}\Big\rangle\Big| \ge \frac{u}{\|x\|_2}\Big) \\ &\ge \mathbb{P}\Big(\Big|\Big\langle X_i, \frac{x}{\|x\|_2}\Big\rangle\Big| \ge u\Big) = \mathbb{P}(|g| \ge u), \end{aligned}$$

where g is a standard Gaussian real-valued random variable. Therefore,

$$Q_{\mathcal{F}}(2u) \ge \mathbb{P}(|g| \ge 2u).$$

Now pick u_* small enough so that the right hand side is bigger than $1/2$, say. Pick m large enough so that

$$\max\Big\{\frac{4c(2+\rho^{-1})s^{1-1/q}\sqrt{\log(en/s)}}{u_*\sqrt{m}}, \frac{\sqrt{\log(2/\eta)}}{\sqrt{2m}}\Big\} \le 1/8.$$

By Lemma III.1 we can now conclude that (7) holds with $\tau = 4^{1/p}/u_*$.

Finally, let $p = \infty$. Since $\|Ax\|_{\log m} \le e\|Ax\|_\infty$,

$$\begin{aligned} &\mathbb{P}\Big(\inf_{x\in T^q_{\rho,s}\cap S_{\ell^n_q}} \|Ax\|_\infty \ge \frac{1}{\tau}\Big) \\ &\ge \mathbb{P}\Big(\inf_{x\in T^q_{\rho,s}\cap S_{\ell^n_q}} \|Ax\|_{\log m} \ge \frac{e}{\tau}\Big). \end{aligned}$$

Thus, in this case the result follows from our proof for $p = \log m$. □

IV. Application to quantized compressed sensing

Consider the situation where we quantize noiseless compressed sensing measurements using a uniform scalar quantization scheme. That is, we observe $y = Q_\theta(A\hat{x})$, where $Q_\theta : \mathbb{R}^m \to (\theta\mathbf{Z} + \theta/2)^m$ is the uniform quantizer with bin width θ defined by $Q_\theta(z) = \big(\theta\lfloor z_i/\theta\rfloor + \theta/2\big)_{i=1}^m$. Graphically, we divide $\mathbb{R}^m$ into hypercubes (or 'bins') with side length θ and map $A\hat{x}$ to the center of the hypercube in which it resides. We view the quantized measurements as noisy linear measurements $y = A\hat{x} + e$, by setting $e = Q_\theta(A\hat{x}) - A\hat{x}$. Since the bin width of the quantization is θ, we clearly have $\|e\|_\infty \le \theta/2$.

To obtain a satisfactory reconstruction $x^\#$ of the signal, we would like to ensure that it is *quantization consistent*. This means that we require that $y = Q_\theta(Ax^\#)$. If we define

$$B_\theta = \{z \in \mathbb{R}^m \ : \ -\theta/2 \le z_i < \theta/2, \ i = 1, \dots, m\},$$

then $x^\#$ is quantization consistent if and only if $Ax^\# - y \in B_\theta$. Thus, we should solve the following *quantization consistent basis pursuit* program

$$\min_{z\in\mathbb{R}^n} \|z\|_1 \quad \text{subject to} \quad Az - y \in B_\theta. \tag{QCBP}$$

This program is strongly related to (BPDN_∞) with $\varepsilon = \theta/2$ (which correspond to taking the closure $\overline{B_\theta}$ instead of B_θ in (QCBP)). In fact, either 1) a minimizer for (QCBP) exists, this is then also a minimizer for (BPDN_∞), or 2) no minimizer exists, in which case every minimizer of (BPDN_∞) is quantization inconsistent. In particular, Theorem III.3 implies the following statement.

Corollary IV.1. *Let A be an $m \times n$ standard Gaussian matrix and $0 < \eta < 1$. Suppose that*

$$m \gtrsim s\log(en/s) + \log(\eta^{-1}).$$

Then, with probability exceeding $1-\eta$ the following holds: for any $\hat{x} \in \mathbb{R}^n$ and quantized measurements $y = Q_\theta(A\hat{x})$, any solution $x^\#$ to (QCBP) is a quantization consistent reconstruction of $\hat{x}$ and satisfies the error bound

$$\|\hat{x} - x^\#\|_2 \lesssim s^{-1/2}\sigma_s(\hat{x})_1 + \theta.$$

Comparing Corollary IV.1 to the performance of the usual basis pursuit denoising, (BPDN_2), we can still reconstruct with the optimal number of measurements, but the reconstruction error does not decay beyond (a constant multiple of) the quantization precision θ.

Let us compare to the work in [5], where the authors introduced and analyzed (BPDN$_p$) with $2 \le p < \infty$ for the purpose of recovering a signal from quantized measurements (as described above). They did not obtain a result for $p = \infty$, but the idea is that the reconstruction becomes more consistent as $p \to \infty$. A main result in [5] shows the following, via an (RIP$_{p,2}$)-based analysis. Assume that the error vector e consists of i.i.d. $U([-\theta/2, \theta/2])$ random variables, that is, we assume that the quantization error is uniformly distributed in each bin (this is called the *high resolution assumption*). With probability at least $1 - e^{-2t^2}$,

$$\|e\|_p \le \varepsilon_p := \frac{\theta}{2(p+1)^{1/p}}(m + t(p+1)\sqrt{m})^{1/p}.$$

This suggests to try to recover $\hat{x}$ via (BPDN$_p$) with $\varepsilon = \varepsilon_p$. Let A be an $m \times n$ standard Gaussian matrix with

$$m \gtrsim (ps\log(en\sqrt{p}/s) + p\log(\eta^{-1}))^{p/2}, \tag{8}$$

then with probability at least $1 - \eta$, for any $\hat{x} \in \mathbb{R}^n$ the reconstruction $x^\#$ via (BPDN$_p$) with $y = Q_\theta(A\hat{x})$ and $\varepsilon = \varepsilon_p$ satisfies

$$\|\hat{x} - x^\sharp\|_2 \lesssim s^{-1/2}\sigma_s(\hat{x})_1 + \frac{\theta}{\sqrt{p+1}}.$$

Compared to Corollary IV.1, the reconstruction error due to quantization error shows decay with p. Note, however, that the value we can take for p is implicitly limited by (8), and in particular we cannot set $p = \infty$ so that $x^\sharp$ is not guaranteed to be quantization consistent. Moreover, when $p > 2$, the number of required measurements grows faster than linear in the sparsity. In fact, it grows exponentially in p, as opposed to the minimal number of measurements needed in Corollary IV.1.

V. Generalization to different distributions

From the proof of Theorem III.3 we extract the following statement, which allows us to generalize our recovery result (as well as Corollary IV.1) to a variety of random matrices beyond the Gaussian case, while retaining the same (optimal) recovery guarantees as for a standard Gaussian matrix.

Theorem V.1. *Let A be an $m \times n$ random matrix with i.i.d. rows $X_1, \dots, X_m$ which are distributed as X. Suppose that for some $u_* > 0$ and $\beta > 0$,*

$$\mathbb{P}\big[|\langle X, x\rangle| \ge u_*\big] \ge \beta \qquad \text{for all } x \in S_{\ell_2^n}, \tag{9}$$

and, if $V = m^{-1/2}\sum_{i=1}^m \varepsilon_i X_i$ then for some $\kappa > 0$,

$$\mathbb{E}\sup_{x\in\Sigma_s^2}\langle V, x\rangle = \mathbb{E}\Big(\sum_{i=1}^s (V_i^*)^2\Big)^{1/2} \le \kappa\sqrt{s\log(en/s)},$$

where V^* is the nonincreasing rearrangement of V. Fix $1 \le p \le \infty$ and $q \ge 2$. If

$$m \gtrsim \max\Big\{\frac{\kappa^2}{u_*^2\beta^2}s^{2-2/q}\log(en/s), \frac{\log(\eta^{-1})}{\beta^2}\Big\},$$

then with probability at least $1 - \eta$ the following holds: for any $\hat{x} \in \mathbb{C}^n$ and $y = A\hat{x} + e$, where $\|e\|_p \le \varepsilon$, any solution $x^\#$ to (BPDN$_p$) satisfies

$$\|\hat{x} - x^\#\|_r \lesssim s^{1/r-1}\sigma_s(\hat{x})_1 + s^{1/r-1/q}\frac{\varepsilon}{\beta^{1/p}u_* m^{1/p}},$$

for any $1 \le r \le q$.

To verify the small ball condition (9), it is often useful to apply the Paley-Zygmund inequality

$$\mathbb{P}(\zeta > t) \ge \frac{(\mathbb{E}\zeta - t)^2}{\mathbb{E}\zeta^2}, \qquad 0 \le t \le \mathbb{E}\zeta, \tag{10}$$

which holds for any nonnegative random variable ζ. In particular, if X is a random vector with independent, mean-zero entries $\xi_1, \dots, \xi_n$ which have variance at least σ^2 and fourth moment bounded from above by μ^4, then

$$\mathbb{P}(|\langle X, x\rangle| > t) \ge \frac{(\sigma^2 - t^2)^2}{\mu^4}, \qquad 0 \le t \le \sigma, \tag{11}$$

whenever $\|x\|_2 = 1$. We refer to [3, Lemmas 7.16 and 7.17] for details.

Let us now verify the conditions of Theorem V.1 for some concrete classes of matrices.

Corollary V.2. *Suppose that the rows of A are i.i.d. copies of X, where X is*

- *sub-isotropic, i.e., $\mathbb{E}\langle X, x\rangle^2 \ge \|x\|_2^2$ for all $x \in \mathbb{C}^n$;*
- *1-subgaussian, i.e., $\mathbb{E}\exp(t\langle X, x\rangle) \le \exp(t^2)$ for all $x \in \mathbb{C}^n$ with $\|x\|_2 \le 1$ and $t \in \mathbb{R}$.*

If $m \gtrsim s^{2-2/q}\log(en/s) + \log(\eta^{-1})$ then the conclusion of Theorem III.3 holds.

Proof. We verify the two conditions of Theorem V.1. To verify (9), we use (10) for $|\langle X, x\rangle|^2$ to get

$$\mathbb{P}(|\langle X, x\rangle| > u) \ge \frac{(\mathbb{E}|\langle X, x\rangle|^2 - u^2)^2}{\mathbb{E}|\langle X, x\rangle|^4} \gtrsim (1 - u^2)^2,$$

whenever $0 \le u \le 1$ and $\|x\|_2 = 1$. In the last inequality, we used that X is sub-isotropic and subgaussian.

To verify the second condition, note that by assumption, the random variable $\langle X_i, x - y\rangle$ is 2-subgaussian for any $x, y \in \Sigma_s^2$. Therefore $V = m^{-1/2}\sum_i \varepsilon_i X_i$ is a 4-subgaussian random vector (see e.g. [3, Theorem 7.27]). By Dudley's inequality (see e.g. [3, Theorem 8.23]),

$$\mathbb{E}\sup_{x\in\Sigma_s^2}\langle V, x\rangle \lesssim \int_0^1 [\log(\mathcal{N}(\Sigma_s^2, \|\cdot\|_2, u)]^{1/2}\, du.$$

Since for any $u > 0$

$$\begin{aligned}\mathcal{N}(\Sigma_s^2, \|\cdot\|_2, u) &\le \binom{n}{s}\max_{S\subset[n]:\ |S|\le s}\mathcal{N}(B_S, \|\cdot\|_2, u)\\ &\le (en/s)^s(1 + (2/u))^s,\end{aligned}$$

we conclude that

$$\begin{aligned}&\mathbb{E}\sup_{x\in\Sigma_s^2}\langle V, x\rangle\\ &\quad\lesssim \sqrt{s\log(en/s)} + \sqrt{s}\int_0^1 [\log(1 + (2/u))]^{1/2}\, du\\ &\quad\lesssim \sqrt{s\log(en/s)}.\end{aligned}$$

□

The following result concerns matrices with i.i.d. entries.

Corollary V.3. *Suppose that $X = (\xi_1, \dots, \xi_n)$, with the ξ_i independent, mean-zero and identically distributed as ξ. Suppose that for some $\lambda > 0$ and $\alpha \geq 1/2$,*

$$(\mathbb{E}|\xi|^r)^{1/r} \leq \lambda r^\alpha, \qquad \text{for all } 2 \leq r \leq \log n. \tag{12}$$

and that (9) holds. If

$$m \gtrsim \max\Big\{ \frac{\lambda^2 e^{4\alpha-2}}{u_*^2 \beta^2} s^{2-2/q} \log(en/s), \frac{\log(\eta^{-1})}{\beta^2}, (\log(n))^{2\alpha-1} \Big\},$$

then the conclusion of Theorem V.1 holds.

Specializing Corollary V.3 to $p = q = 2$, we obtain a result similar to [25, Theorem A]. Let us compare the two results. On the one hand, our result gives a better power in the $\log(n)$ factor ($2\alpha - 1$ versus $4\alpha - 1$) and improved (actually optimal) dependence on the failure probability η. On the other hand, [25, Theorem A] does not require independence of the ξ_i and needs only a small ball assumption on the set of sparse vectors Σ_s (rather than one on the larger set $T_{\rho,s}^2 \cap S_{\ell_2^n}$ used here).

Proof. We fix the randomness in the Rademacher sequence (ε_i). The random variables $V_j = m^{-1/2} \sum_{i=1}^m \varepsilon_i X_{ij}$ are then independent and mean-zero. Since X_{ij} satisfies (12), [25, Lemma 2.8] shows that if $m \geq (\log(n))^{\max\{2\alpha-1,1\}}$, then for any $2 \leq p \leq \log(n)$

$$(\mathbb{E}|V_j|^p)^{1/p} \lesssim e^{2\alpha-1} \lambda \sqrt{p},$$

i.e., the first $\log(n)$ moments show subgaussian behaviour. Therefore, (the proof of) [10, Lemma 6.5] shows that

$$\mathbb{E}\Big(\sum_{i=1}^s (V_i^*)^2 \Big)^{1/2} \lesssim e^{2\alpha-1} \lambda \sqrt{s \log(en/s)}.$$

The result is now immediate from Theorem V.1. □

Example V.4. Let A be a random matrix with i.i.d. entries A_{ij}. Below we list some instances of Corollary V.3. Note that if we measure the reconstruction error in ℓ_2 (i.e., $q = 2$), then the stated lower bounds in (i), (ii), (iv), and (v) coincide with the optimal number of measurements.

(i) If the A_{ij} are random signs (i.e. Rademachers), then $m \gtrsim s^{2-2/q} \log(en/s) + \log(\eta^{-1})$ is sufficient for the recovery guarantee in Theorem III.3. This follows from Corollary V.3 with $\lambda = 1$, $\alpha = 1/2$ and β, u_* universal constants.

(ii) If the A_{ij} are standard symmetric exponential random variables, then $m \gtrsim s^{2-2/q} \log(en/s) + \log(\eta^{-1})$ suffices for the recovery guarantee in Theorem III.3. Indeed, in this case one can apply Corollary V.3 with $\lambda = \alpha = 1$ and take for β, u_* universal constants.

(iii) Let $0 < \gamma \leq 2$. Suppose that the A_{ij} are distributed as ξ_γ, where ξ_γ is a ψ_γ-random variable which has the same distribution as $\operatorname{sign}(g)|g|^{2/\gamma}$, with g standard Gaussian. Since $(\mathbb{E}|g|^{2r/\gamma})^{1/r} \simeq (2r/\gamma)^{1/\gamma}$ for any $r \geq 1$, we can take $\alpha = 1/\gamma$ and $\lambda \simeq (2/\gamma)^{1/\gamma}$. Moreover, applying (11) with $\sigma^2 \simeq (4/\gamma)^{2/\gamma}$, $\mu^4 \simeq (8/\gamma)^{4/\gamma}$ and $t = \sigma^2/2$, we find that (9) holds with $u_* \simeq (4/\gamma)^{2/\gamma}/2$ and $\beta \simeq 2^{-4/\gamma}/4$. Using Corollary V.3 we can therefore conclude that

$$m \gtrsim \max\Big\{ (e\sqrt{8})^{4/\gamma} s^{2-2/q} \log(en/s), 2^{8/\gamma} \log(\eta^{-1}), (\log(n))^{(2/\gamma)-1} \Big\}, \tag{13}$$

measurements are sufficient for the reconstruction error bound

$$\|\hat{x} - x^\#\|_r \lesssim s^{1/r-1} \sigma_s(\hat{x})_1 + s^{1/r-1/q} \frac{2^{4/(p\gamma)} \mathcal{E}}{(4/\gamma)^{2/\gamma} m^{1/p}},$$

for any $1 \leq r \leq q$.

(iv) Suppose that the A_{ij} are distributed as a Student-t variable ξ_d of degree d. The random variable ξ_d does not have finite r-th moments for $r \geq d$. For $r < d$ we have $(\mathbb{E}|\xi_d|^r)^{1/r} \simeq \sqrt{dr}/(d-r)$ [30]. In particular, if $r < d/2$ then ξ_d has a subgaussian r-th moment. Moreover, (9) holds with constants β, u_* which do not depend on the degree d. Therefore, if $d \geq 2 \log n$, then $m \gtrsim s^{2-2/q} \log(en/s) + \log(\eta^{-1})$ suffices for the recovery guarantee in Theorem III.3.

(v) Suppose that the A_{ij} are distributed as a random variable ξ, which has probability density function

$$p(x) = \frac{\gamma - 1}{2\gamma} \min\{1, |x|^{-\gamma}\}, \qquad x \in \mathbb{R},$$

for some $\gamma > 1$. One readily calculates that

$$\mathbb{E}|\xi|^p = \frac{\gamma - 1}{\gamma} \Big(\frac{1}{\gamma - p - 1} + \frac{1}{p + 1} \Big)$$

for $p < \gamma - 1$ and $\mathbb{E}|\xi|^p = \infty$ for $p \geq \gamma - 1$. If we assume $\gamma \geq \log(n) + 2$, say, then ξ trivially satisfies the moment bound in Corollary V.3 with $\alpha = 1/2$. Moreover, if $\gamma > 5$ then $\mathbb{E}\xi^2 = (\gamma - 1)/(3\gamma - 9)$ and $\mathbb{E}\xi^4 = (\gamma - 1)/(5\gamma - 25)$ so the Paley-Zygmund inequality (11) implies that (9) holds for universal constants u_*, β if $\gamma \geq 6$, say. In conclusion, if we assume

$$\gamma \geq \max\{\log(n) + 2, 6\},$$

then $m \gtrsim s^{2-2/q} \log(en/s) + \log(\eta^{-1})$ is sufficient for the recovery guarantee in Theorem III.3.

The last two examples illustrate that only the behaviour of the first $\log n$ moments of the entries of A is important for our sparse recovery result, the higher moments need not even exist. In fact, if the degree d or the parameter γ, respectively, is large enough, then we can stably and robustly recover with the same number of measurements as in the Gaussian case. In contrast, even though the ψ_γ-random variables in (iii) of Example V.4 have moments of all orders, their first $\log n$ moments grow at an increasing rate $\alpha = 1/\gamma$ as $\gamma \downarrow 0$. As a consequence, the required number of measurements in (13) increases as γ decreases. Let us emphasize that (13) is only a sufficient condition for recovery. However, a numerical simulation in Section VII will show that the growth of m in terms of $1/\gamma$ is not an artefact of our proof: the experimentally observed reconstruction performance deteriorates as $\gamma \downarrow 0$.

To conclude this section, we extend the example of a standard symmetric exponential matrix (part (ii) of Example V.4)

to matrices with i.i.d. isotropic, unconditional, log-concave rows. In particular, we do not assume that the entries within a row are independent. Recall that a probability measure μ on $\mathbb{R}^n$ is called log-concave if for any (Borel) sets $A, B \subset \mathbb{R}^n$ and $0 \leq \theta \leq 1$,

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

A random vector Y is called *log-concave* if its probability distribution is log-concave. We call Y *isotropic* if it is mean-zero and $\mathbb{E}\langle Y, x\rangle^2 = \|x\|_2^2$ for all $x \in \mathbb{R}^n$. We say that Y is *unconditional* if, for any $\varepsilon_1, \ldots, \varepsilon_n \in \{-1, 1\}$, the vector $(\varepsilon_1 Y_1, \ldots, \varepsilon_n Y_n)$ has the same distribution as Y. A typical example of an isotropic, unconditional log-concave vector Y is a random variable uniformly distributed over the unit ball of an unconditional norm in the isotropic position.

We will use the following comparison theorem from [31] (see also Theorem 2.5 in [32]), which is based on earlier work in [33]. It will allow us to reduce the general case of matrices with i.i.d. isotropic, unconditional, log-concave rows to the special case of a standard symmetric exponential matrix.

Theorem V.5. *Let Y be an isotropic, unconditional, log-concave vector in $\mathbb{R}^d$ and E be a standard d-dimensional symmetric exponential vector, i.e., its entries are i.i.d. standard symmetric exponential variables. Let $\|\cdot\|$ be any semi-norm on $\mathbb{R}^d$. Then for any $t > 0$,*

$$\mathbb{P}\big[\, \|Y\| \geq Ct\big] \leq C\mathbb{P}\big[\, \|E\| \geq t\big],$$

where C is a universal constant.

Corollary V.6. *Let A be an $m \times n$ matrix with i.i.d. rows X_i distributed as X, where X is an isotropic, unconditional log-concave vector. If $m \gtrsim s^{2-2/q}\log(en/s) + \log(\eta^{-1})$, then the conclusion of Theorem III.3 holds.*

Proof. We verify the conditions of Theorem V.1. By a result of Borell (see e.g. [34, Proposition 2.14]), X is a sub-exponential vector. In fact, for any $p \geq 1$,

$$(\mathbb{E}|\langle X, x\rangle|^p)^{1/p} \lesssim p\mathbb{E}|\langle X, x\rangle| \qquad \text{for all } \ x \in \mathbb{R}^n.$$

Since X is isotropic, we can apply (10) for $|\langle X, x\rangle|^2$ to get

$$\mathbb{P}(|\langle X, x\rangle| > u) \geq \frac{(\mathbb{E}|\langle X, x\rangle|^2 - u^2)^2}{\mathbb{E}|\langle X, x\rangle|^4} \gtrsim (1-u^2)^2,$$

whenever $0 \leq u \leq 1$ and $\|x\|_2 = 1$. This shows that (9) holds with absolute constants $u_*, \beta > 0$.

To prove the second condition, we define a semi-norm on $\mathbb{R}^{m\times n}$ by

$$\|B\|_s = \sup_{x\in\Sigma_s^2} \Big\langle \sum_{i=1}^m B_i, x \Big\rangle$$

where the B_i are the m row vectors of $B \in \mathbb{R}^{m\times n}$. Since the X_i are unconditional,

$$\mathbb{E}\sup_{x\in\Sigma_s^2} \langle V, x\rangle = \frac{1}{\sqrt{m}}\mathbb{E}\,\|A\|_s\,.$$

Considered as a vector in $\mathbb{R}^{mn}$, A is isotropic, unconditional and log-concave. Theorem V.5 therefore implies that,

$$\mathbb{P}\big[\, \|A\|_s \geq Ct\big] \leq C\mathbb{P}\big[\, \|\mathcal{E}\|_s \geq t\big],$$

where $\mathcal{E}$ is an $m \times n$ standard symmetric exponential matrix. As a consequence, we have

$$\begin{aligned}\mathbb{E}\,\|A\|_s &= \int_0^\infty \mathbb{P}\big[\, \|A\|_s \geq t\big]dt \\ &\leq C^2 \int_0^\infty \mathbb{P}\big[\, \|\mathcal{E}\|_s \geq t\big]dt \lesssim \mathbb{E}\,\|\mathcal{E}\|_s\,.\end{aligned}$$

By the proof of Corollary V.3 (see also (ii) of Example V.4), $\mathbb{E}\,\|\mathcal{E}\|_s \lesssim \sqrt{ms\log(en/s)}$, which proves the second condition in Theorem V.1. □

As was mentioned before, Koltchinskii showed that $m \gtrsim s\log(en/s)$ isotropic, log-concave measurements suffice with high probability to recover every s-sparse vector exactly via ℓ_1-minimization [23, Theorem 7.3]. Under the additional assumption that the measurement vectors are unconditional, Corollary V.6 makes this result stable with respect to approximate sparsity and robust with respect to measurement noise, while retaining the optimal number of measurements.

VI. RIP RIP?

The classical restricted isometry property, (RIP$_{2,2}$), played a major role in the theory of compressed sensing since [16], [17]. It has proved to be an optimal tool to analyze standard basis pursuit denoising for subgaussian matrices. It has also been used to show that various other random matrices, including structured random matrices, allow for uniform sparse recovery via (BPDN$_2$) if one increases the number of measurements with additional logarithmic factors. Nevertheless, it is known that for certain ensembles (e.g. subexponential) this logarithmic increase can be avoided, establishing a gap between RIP and sparse recovery conditions.

In this work we showed that this gap becomes much more pronounced when considering (BPDN$_p$) for $p \neq 2$. An analysis of this program via an RIP condition erroneously suggests that 1) the required optimal number of measurements for uniform sparse recovery may be much larger than in the case $p = 2$, especially if $p > 2$, and 2) that one may need to consider random measurements different from Gaussian to attain this optimal number. This begs the question: does this mean that researchers interested in sparse recovery should stop considering restricted isometry properties? In this paper we showed that by proving a *lower* (RIP$_{p,q}$)-type of bound on an *extension* of the set of sparse vectors (cf. (7)), one can prove an optimal recovery result for a large class of matrices, which do not satisfy (RIP$_{p,q}$) in the optimal measurement regime. Thus, it seems the gap between restricted isometry properties and sparse recovery conditions originates in the upper bound of the RIP "for all $x \in \Sigma_s$, $\|Ax\|_p \leq C\,\|x\|_q$" – at least when considering convex optimization approaches for recovery.

To move towards a definitive answer of our question, it would be interesting to determine whether similar gaps occur between restricted isometry properties and sparse recovery conditions for other numerical methods. For example, there are several algorithms such as iterative hard thresholding and CoSamp for which convergence results are currently only known under the (classical) RIP.

VII. Simulations

Our results show that for a broad class of measurement matrices, one can stably and robustly recover sparse vectors using (BPDN$_p$) with the same number of measurements as in the Gaussian case. In the case that the measurement matrix is populated with i.i.d. copies of a mean-zero random variable, we have seen two phenomena. First, only the behaviour of the first $\log n$ moments is relevant for the reconstruction guarantees. In fact, the moments beyond level $\log n$ need not even exist. Second, the number of measurements that is sufficient for stable and robust reconstruction increases as the first $\log n$ moments grow faster (i.e., as the value of α in (12) becomes larger).

In this section we illustrate these two phenomena using numerical simulations. We consider two classes of random measurement matrices: matrices filled with i.i.d. ψ_γ-variables and matrices with i.i.d. Student-t entries (see (iii) and (iv), respectively, in Example V.4). The random variables in the first class have finite moments of every order, but these moments grow at an increasing rate $\alpha = 1/\gamma$ as $\gamma \downarrow 0$. A Student-t variable of degree d, on the other hand, only has finite r-th moment for $r < d$. However, the first few moments behave much better than those of a ψ_γ random variable: they even show subgaussian behaviour for $r < d/2$. We simulate phase transition curves [35] in the noiseless setting for these classes of random variables for various values of γ and d, respectively, as follows. At every pixel point $(m, s) \in \{1, \dots, n\}^2$ with $n = 100$ of the phase transition matrix, we construct 15 s-sparse vectors $\hat{x} \in \mathbb{R}^n$ by selecting a support set uniformly at random and filling it with 1's. For each of these vectors, we check if Basis Pursuit (solved here using the Douglas-Rachford algorithm) succeeds, meaning that the reconstruction $x^\sharp$ satisfies $\left\|x^\sharp - \hat{x}\right\|_2 \leq 0.001$. For each sparsity level s, we mark the first time that the number of successes exceeds the number of failures. This yields a phase transition curve: in the area above the curve the number of successful reconstructions exceeds the number of failures, below the curve the opposite is true. We repeat this protocol 20 times and average the phase transition curves. The result is shown in Figures 1 and 2.

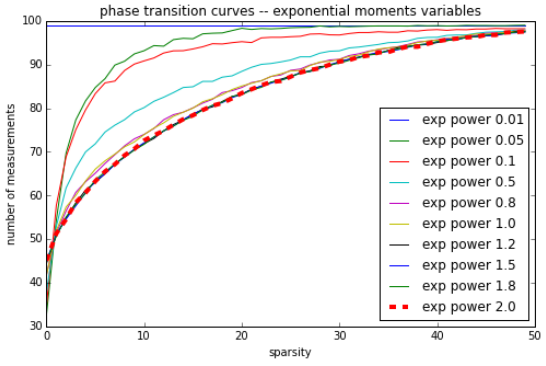

Fig. 1. Phase transition curves of the Basis Pursuit procedure for measurements matrices with i.i.d. ψ_γ entries for $\gamma = 2, 1.8, 1.5, 1.2, 1, 0.8, 0.5, 0.1, 0.05, 0.01$. The curve obtained for Gaussian measurements corresponds to $\gamma = 2$.

Except for the exponential power $\gamma = 0.5, 0.1, 0.05, 0.01$ and Student's t-distribution of degree $d = 2$, the phase transition curves are close to the one obtained for Gaussian measurements. This agrees with the theoretical recovery guarantees in Example V.4. Our results predict that in our simulation only the first $\log n \sim 4$ moments determine the reconstruction performance. In the simulation for ψ_γ-variables, our sufficient condition (13) suggested that more measurements are needed for stable and robust reconstruction as γ decreases. Our simulation shows that the growth of m in terms of $1/\gamma$ is not an artefact of our proof: the reconstruction performance deteriorates as γ decreases. The simulation for Student-t variables illustrates that if the degree d is large enough, then the first few moments are finite and grow in a subgaussian fashion and therefore the phase transition curve is almost the same as for a Gaussian matrix. If the degree becomes too small, not enough moments exist and as a consequence, the reconstruction performance deteriorates.

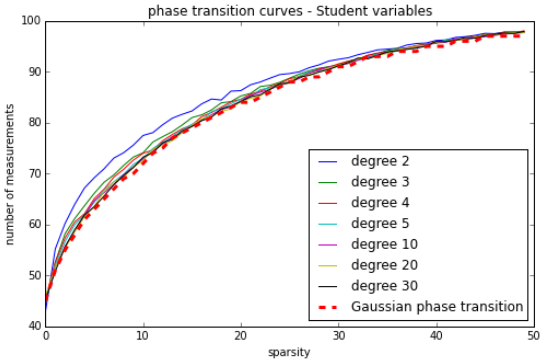

Fig. 2. Phase transition curves of the Basis Pursuit procedure for measurements matrices with i.i.d. Student t-distributed entries with degree $d = 2, 3, 4, 5, 10, 20, 30$. The curve obtained for Gaussian measurements is also drawn for comparison.

Similar phase transition curves can be observed in simulations for noisy measurements, as well as for the quantized measurements studied in Section IV. We omit a detailed description of these numerical simulations, but the interested reader can produce his/her own simulations using our code [36] for various measurements matrices, including the random matrices with correlated rows considered in Corollaries V.2 and V.6.

Acknowledgements

H. Rauhut acknowledges funding by the European Research Council through the Starting Grant StG 258926. G. Lecué acknowledges funding by Chaire Havas-Dauphine; Economie des nouvelles données. The authors thank the reviewer for his/her comments and the resulting improvements.

References

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] E. J. Candès, J., T. Tao, and J. K. Romberg, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[3] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Boston: Birkhäuser, 2013.

[4] B. K. Natarajan, "Sparse approximate solutions to linear systems," *SIAM Journal on Computing*, vol. 24, no. 2, pp. 227–234, 1995. [Online]. Available: http://dx.doi.org/10.1137/S0097539792240406

[5] L. Jacques, D. K. Hammond, and J. M. Fadili, "Dequantizing compressed sensing: when oversampling and non-Gaussian constraints combine," *IEEE Trans. Inform. Theory*, vol. 57, no. 1, pp. 559–571, 2011. [Online]. Available: http://dx.doi.org/10.1109/TIT.2010.2093310

[6] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, Cambridge, 2004. [Online]. Available: http://dx.doi.org/10.1017/CBO9780511804441

[7] J.-J. Fuchs, "Fast implementation of a ℓ_1-ℓ_1 regularized sparse representations algorithm." in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, April 2009, pp. 3329–3332.

[8] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, and M. Strauss, "Combining geometry and combinatorics: A unified approach to sparse signal recovery," in *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*, Sept 2008, pp. 798–805.

[9] V. Koltchinskii and S. Mendelson, "Bounding the smallest singular value of a random matrix without concentration," *Int. Math. Res. Notices*, to appear, arXiv:1312.3580.

[10] S. Mendelson, "Learning without concentration," *J. ACM*, to appear, arXiv:1401.0304.

[11] ——, "Learning without concentration for general loss functions," *arXiv:1410.3192*, 2014.

[12] ——, "A remark on the diameter of random sections of convex bodies," in *Geometric Aspects of Functional Analysis (GAFA Seminar Notes)*, ser. Lecture notes in Mathematics, B. Klartag and E. Milman, Eds., vol. 2116, 2014, pp. 395–404.

[13] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, vol. 24, no. 3, pp. 035 020, 14, 2008.

[14] S. Foucart and M.-J. Lai, "Sparse recovery with pre-Gaussian random matrices," *Studia Math.*, vol. 200, no. 1, pp. 91–102, 2010.

[15] O. Friedland and O. Guédon, "Sparsity and non-Euclidean embeddings," *Israel J. Math.*, vol. 197, no. 1, pp. 329–345, 2013.

[16] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[17] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies?" *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006. [Online]. Available: http://dx.doi.org/10.1109/TIT.2006.885507

[18] V. B. Chandar, "Sparse graph codes for compression, sensing, and secrecy," Ph.D. dissertation, Massachusetts Institute of Technology, 2010.

[19] P. T. Boufounos, L. Jacques, F. Krahmer, and R. Saab, "Quantization and compressive sensing," *arXiv:1405.1194*, 2014.

[20] Z. Allen-Zhu, R. Gelashvili, and I. Razenshteyn, "The restricted isometry property for the general p-norms," *arXiv:1407.2178*, 2014.

[21] R. Adamczak, R. Latała, A. Litvak, A. Pajor, and N. Tomczak-Jaegermann, "Geometry of log-concave ensembles of random matrices and approximate reconstruction," *C. R. Math. Acad. Sci. Paris*, vol. 349, no. 13-14, pp. 783–786, 2011. [Online]. Available: http://dx.doi.org/10.1016/j.crma.2011.06.025

[22] R. Adamczak, A. E. Litvak, A. Pajor, and N. Tomczak-Jaegermann, "Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling," *Constr. Approx.*, vol. 34, no. 1, pp. 61–88, 2011. [Online]. Available: http://dx.doi.org/10.1007/s00365-010-9117-4

[23] V. Koltchinskii, *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems*. Berlin: Springer, 2011.

[24] S. Foucart, "Stability and robustness of ℓ_1-minimizations with Weibull matrices and redundant dictionaries," *Linear Algebra Appl.*, vol. 441, pp. 4–21, 2014. [Online]. Available: http://dx.doi.org/10.1016/j.laa.2012.10.003

[25] G. Lecué and S. Mendelson, "Sparse recovery under weak moment assumptions," *J. Eur. Math. Soc.*, to appear, arXiv:1401.2188.

[26] M. Kabanava and H. Rauhut, "Analysis ℓ_1-recovery with frames and Gaussian measurements," *Acta Appl. Math.*, to appear, DOI:10.1007/s10440-014-9984-y.

[27] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Comm. Pure Appl. Math.*, vol. 61, no. 8, pp. 1025–1045, 2008. [Online]. Available: http://dx.doi.org/10.1002/cpa.20227

[28] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, "Reconstruction and subgaussian operators in asymptotic geometric analysis," *Geom. Funct. Anal.*, vol. 17, no. 4, pp. 1248–1282, 2007. [Online]. Available: http://dx.doi.org/10.1007/s00039-007-0618-7

[29] A. Pinkus, *n-Widths in approximation theory*. Springer-Verlag, 1985.

[30] S. Kotz and S. Nadarajah, *Multivariate t distributions and their applications*. Cambridge University Press, Cambridge, 2004. [Online]. Available: http://dx.doi.org/10.1017/CBO9780511550683

[31] R. Latała, "On weak tail domination of random vectors," *Bull. Pol. Acad. Sci. Math.*, vol. 57, no. 1, pp. 75–80, 2009. [Online]. Available: http://dx.doi.org/10.4064/ba57-1-8

[32] R. Adamczak and D. Chafaï, "Circular law for random matrices with unconditional log-concave distribution," 2014, arXiv:1303.5838.

[33] S. G. Bobkov and F. L. Nazarov, "On convex bodies and log-concave probability measures with unconditional basis," in *Geometric aspects of functional analysis*, ser. Lecture Notes in Math. Springer, Berlin, 2003, vol. 1807, pp. 53–69.

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

[35] D. Donoho and J. Tanner, "Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing," *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, vol. 367, no. 1906, pp. 4273–4293, 2009, with electronic supplementary materials available online. [Online]. Available: http://dx.doi.org/10.1098/rsta.2009.0152

[36] Notebook available at https://github.com/lecueguillaume/.