Electron. Commun. Probab. **23** (2018), no. 13, 1–8.
https://doi.org/10.1214/17-ECP100
ISSN: 1083-589X

ELECTRONIC COMMUNICATIONS in PROBABILITY

Column normalization of a random measurement matrix*

Shahar Mendelson†

Abstract

In this note we answer a question of G. Lecué, by showing that column normalization of a random matrix with iid entries need not lead to good sparse recovery properties, even if the generating random variable has a reasonable moment growth. Specifically, for every $2 \leq p \leq c_1 \log d$ we construct a random vector $X \in \mathbb{R}^d$ with iid, mean-zero, variance 1 coordinates, that satisfies $\sup_{t \in S^{d-1}} \|\langle X, t\rangle\|_{L_q} \leq c_2\sqrt{q}$ for every $2 \leq q \leq p$. We show that if $m \leq c_3\sqrt{p}d^{1/p}$ and $\tilde{\Gamma} : \mathbb{R}^d \to \mathbb{R}^m$ is the column-normalized matrix generated by m independent copies of X, then with probability at least $1-2\exp(-c_4 m)$, $\tilde{\Gamma}$ does not satisfy the exact reconstruction property of order 2.

Keywords: sparse recovery; column normalization.
AMS MSC 2010: 62G99.
Submitted to ECP on March 2, 2017, final version accepted on November 13, 2017.

1 Introduction

Sparse Recovery is one of the most important research topics in modern signal processing. It focuses on the possibility of identifying a sparse signal—i.e., a signal that is supported on relatively few coordinates in $\mathbb{R}^d$ relative to the standard basis—using linear measurements. We refer the reader to the books [2, 3] for extensive surveys on sparse recovery and related topics.

In a basic sparse recovery problem one pre-selects an $m \times d$ matrix Γ that generates the given data. For an unknown (sparse) vector v, the coordinates of the vector Γv are the m linear measurements of v one may use for recovery. The hope is that for a well chosen Γ, the resulting m linear measurements are enough to identify v, and because v is sparse, the number of measurements required for recovery is significantly smaller than the dimension d.

One of the main achievements of the theory of sparse recovery was the introduction of a convex optimization problem called *basis pursuit*, which is an effective recovery procedure: it selects $t \in \mathbb{R}^d$ that solves the minimization problem

$$\min \|t\|_1 \quad \text{subject to} \quad \Gamma v = \Gamma t, \tag{1.1}$$

where we denote by $\|x\|_p = (\sum_{j=1}^d |x_j|^p)^{1/p}$.

The question of finding conditions on the measurement matrix Γ that ensure the recovery of any sparse vector has been studied extensively. Specifically, one would like to guarantee that for every s-sparse vector v, the ℓ_1 minimization problem (1.1) has a unique solution—v itself.

*Supported in part by the Israel Science Foundation.
†Department of Mathematics, Technion, I.I.T., Haifa, Israel and Mathematical Sciences Institute, The Australian National University, Canberra, Australia. E-mail: shahar@tx.technion.ac.il

Definition 1.1. *Let Σ_s be the set of s-sparse vectors in $\mathbb{R}^d$. A matrix $\Gamma : \mathbb{R}^d \to \mathbb{R}^m$ satisfies the exact reconstruction property of order s if for every $v \in \Sigma_s$ there is a unique solution to the minimization problem* (1.1) *and that unique solution is v.*

Because measurements are 'expensive', one would like to find matrices Γ that satisfy the exact reconstruction property of order s with the smallest number of measurements (rows) possible. One may show that if Γ satisfies the exact reconstruction property of order s, then it must have at least $m \sim s\log(ed/s)$ rows. Moreover, typical realizations of various random matrices with $\sim s\log(ed/s)$ rows indeed satisfy the exact reconstruction property of order s (see, e.g., [3]). Therefore, the optimal number of measurements required for the exact reconstruction property of order s is $m \sim s\log(ed/s)$, and that number serves as the benchmark for an optimal measurement matrix.

The question we are interested in has to do with the normalization of the columns of the measurement matrix. It is often assumed in literature that the columns of Γ have unit Euclidean norm (see, for example, [2] and [3] and references therein); i.e., if $\{e_1, ..., e_d\}$ is the standard basis in $\mathbb{R}^d$ then $\|\Gamma e_j\|_2 = 1$ for $1 \le j \le d$. Column normalization appears frequently in various notions used in the study of the exact reconstruction property. Among these well-studied notions are *coherence* [3]; the *restricted eigenvalues condition* [1]; and the *compatibility condition* [2]. Moreover, in many real-world applications, measurement matrices with normalized columns tends to perform better than matrices whose columns have not been normalized.

While column normalization seems a natural idea, it adds substantial technical difficulties to the analysis of basis pursuit when random measurement matrices are used: normalizing the columns of a matrix with independent rows introduces additional dependencies. Despite the added difficulties, the results of [5] highlight the possibility that column normalization may still have a significant role to play in the context of random measurement matrices, particularly in heavy-tailed situations.

Before we formulate the results of [5], let us introduce some notation. Throughout, absolute constants are denoted by $c, c_1, ...$; their value may change from line to line. $c(L)$ denotes a constant that depends only on the parameter L. We write $a \sim b$ if there are absolute constants c and C such that $ca \le b \le Ca$. Finally, B_2^d denotes the Euclidean unit ball in $\mathbb{R}^d$ and S^{d-1} is the corresponding unit sphere.

Definition 1.2. *Let x be a random variable. Given an integer $m \le d$, let (x_{ij}), $1 \le i \le m$, $1 \le j \le d$ be md independent copies of x. The random matrix generated by x is $\Gamma = (x_{ij})_{1 \le i \le m, 1 \le j \le d}$. Also, we denote by $X = (x_j)_{j=1}^d$ a vector with d independent copies of x; thus the rows of Γ are m independent copies of X.*

The following result from [5] (Theorem C' there) is a construction of random matrices generated by seemingly nice random variables, but despite that the matrices exhibit poor reconstruction properties.

Theorem 1.3. *There exist absolute constants c_1, c_2 and c_3 for which the following holds. For every $2 < p \le c_1 \log d$ there is a mean-zero, variance one random variable x that satisfies*

- *For every $2 \le q \le p$ and every $t \in S^{d-1}$,*

$$\|\langle X, t\rangle\|_{L_q} \le c_2\sqrt{q}\|\langle X, t\rangle\|_{L_2} = c_2\sqrt{q}.$$

- *If $m \le c_3\sqrt{p}(d/\log d)^{1/p}$ then with probability $1/2$, Γ does not satisfy the exact reconstruction property of order 1.*

Theorem 1.3 implies that without assuming that each $\langle X, t\rangle$ has a subgaussian moment growth[1] up to the p-moment for p close to $\log d$, the resulting measurement matrix is suboptimal. Indeed, under a modest assumption, say that $\|\langle X,t\rangle\|_{L_4} \le c\|\langle X,t\rangle\|_{L_2}$ for every $t \in \mathbb{R}^d$, the recovery of 1-sparse vectors requires at least $\sim (d/\log d)^{1/4}$ measurements. And, if $p = (\log d)/(\beta \log\log d)$ for β large enough, then the number of measurements required for the recovery of 1-sparse vectors is at least $\sim \log^{c\beta} d$, which is suboptimal when $c\beta > 1$.

To put Theorem 1.3 in some perspective, it is complemented by a positive result, once linear forms have enough subgaussian moments. Indeed, the following is an immediate corollary of Theorem A from [5].

Theorem 1.4. *Let x be a mean-zero, variance one random variable. Assume that for every $2 \le q \le c_1 \log d$ and every $t \in S^{d-1}$,*

$$\|\langle X,t\rangle\|_{L_q} \le L\sqrt{q}\|\langle X,t\rangle\|_{L_2} = L\sqrt{q}. \tag{1.2}$$

If

$$m \ge c_2 s \log(ed/s),$$

then with probability at least $1-1/d^{c_3} - 2\exp(-c_4 m)$, Γ satisfies the exact reconstruction property of order s. Here, c_1 in an absolute constant and c_2, c_3 and c_4 are constants that depend only on L.

Theorem 1.4 implies that if X has a slightly better moment growth condition than in Theorem 1.3—a subgaussian growth up to $p \sim \log d$—the random measurement matrix generated by x satisfies the exact reconstruction property of order s, for the optimal number of measurements $m \sim s\log(ed/s)$.

The connection with column-normalization arises from the main observation used in the proof of Theorem 1.4 (see Theorem B in [5]).

Lemma 1.5. *Recall that Σ_s denotes the set of s-sparse vectors in $\mathbb{R}^d$. Let $\Gamma : \mathbb{R}^d \to \mathbb{R}^m$. If*

(a) $\|\Gamma x\|_2 \ge \alpha\|x\|_2$ for every $x \in \Sigma_s$,

(b) $\|\Gamma e_j\|_2 \le \beta$ for every $j \in \{1, ..., d\}$,

and $s_1 = \lfloor \alpha^2(s-1)/(4\beta^2)\rfloor - 1$, then Γ satisfies the exact reconstruction property of order s_1.

Lemma 1.5 gives a clear motivation for considering column-normalized random measurement matrices, and that motivation grows stronger when one examines the proof of Theorem 1.4. It turns out that the 'bottleneck' in the proof is the upper bound on $\max_{1\le j\le d}\|\Gamma e_j\|_2$, while guaranteeing (a) requires a rather minimal *small-ball* assumption. Therefore, the seemingly more restrictive condition (a) is almost universally true (see [7, 5] for more details) whereas (b) is the only place in which the moment growth assumption is used in the proof of Theorem 1.4.

Clearly, column normalization resolves the issue of an upper estimate on $\max_{1\le j\le d}\|\Gamma e_j\|_2$. That, and the fact that (a) is true under minimal assumptions has led G. Lecué [4] to ask whether with column normalization, the moment growth condition (1.2) can be relaxed significantly, leading to a much stronger version of Theorem 1.4.

Question 1.6. Let x be a mean-zero, variance 1 random variable, set Γ to be the $m \times d$ matrix generated by x and let $\tilde{\Gamma}$ be the column-normalized matrix generated by x; thus,

[1] Recall that a characterization of an L-subgaussian random variable is that $\|z\|_{L_p} \le L\sqrt{p}\|z\|_{L_2}$ for every $p \ge 2$.

the entries of $\tilde{\Gamma}$ are

$$\tilde{\Gamma}_{ij} = \frac{x_{ij}}{\left(\sum_{\ell=1}^m x_{\ell j}^2\right)^{1/2}} = \frac{\Gamma_{ij}}{\|\Gamma e_j\|_2}.$$

If $\|\langle X,t\rangle\|_{L_4} \leq L\|\langle X,t\rangle\|_{L_2}$ for every $t \in \mathbb{R}^d$ and $m = c(L)s\log(ed/s)$, does $\tilde{\Gamma}$ satisfy the exact reconstruction property of order s with high probability?

Our main result is a version of Theorem 1.3 for a column-normalized matrix generated by well chosen random variable, showing that the answer to Question 1.6 is negative.

Theorem 1.7. *There exist absolute constants c_1, c_2 and c_3 for which the following holds. For every $2 \leq p \leq \log d$ there is a symmetric, variance 1 random variable x with the following properties:*

- *If $x_1, ..., x_d$ are independent copies of x and $X = (x_j)_{j=1}^d$, then for every $t \in S^{d-1}$ and every $2 \leq q \leq p$, $\|\langle X,t\rangle\|_{L_q} \leq c_1\sqrt{q}\|\langle X,t\rangle\|_{L_2}$.*
- *If $m \leq c_2\sqrt{p}d^{1/p}$, then with probability at least $1 - 2\exp(-c_3 m)$, the $m \times d$ column-normalized matrix generated by x does not satisfy the exact reconstruction property of order 2.*

Theorem 1.7 implies that column normalization does not improve the poor behaviour described in Theorem 1.3. Indeed, for $p = 4$, linear forms $\langle X,t\rangle$ satisfy an $L_2 - L_4$ norm equivalence, but the recovery of 2-sparse vectors using $\tilde{\Gamma}$ requires at least $m \sim d^{1/4}$ measurements — significantly larger than the optimal number of measurements, $m \sim \log d$. Moreover, if $\beta > 1$ and $p = (\log d)/\beta \log\log d$, then although $\|\langle X,t\rangle\|_{L_q} \lesssim \sqrt{q}\|\langle X,t\rangle\|_{L_2}$ for every $2 \leq q \leq p$, the recovery of 2-sparse vectors using $\tilde{\Gamma}$ requires at least $m \sim \log^{c\beta} d$ measurements, which, again, is suboptimal when $c\beta > 1$.

Remark 1.8. Theorem 1.7 actually improves the estimates established in [5]: a logarithmic factor in the bound on the number of measurements is removed, and the probability estimate is significantly better: $1 - 2\exp(-cm)$ rather than constant probability.

Let us mention the straightforward observation that a version of Theorem 1.4 holds for column-normalized matrices as well.

Theorem 1.9. *Let x and L be as in Theorem 1.4 and let $\tilde{\Gamma}$ be the column-normalized measurement matrix generated by x. If $m \geq c_1(L)s\log(ed/s)$, then with probability at least $1 - 1/d^{c_2(L)} - 2\exp(-c_3(L)m)$, $\tilde{\Gamma}$ satisfies the exact reconstruction property of order s.*

Theorem 1.9 is an immediate consequence of the proof of Theorem 1.4; its proof is presented in Appendix A merely for the sake of completeness.

2 Proof of Theorem 1.7

Let ε be a symmetric, $\{-1,1\}$-valued random variable, set η to be a $\{0,1\}$-valued random variable with mean δ and let $R > 0$; the values of δ and R are specified in what follows. Set

$$x = \varepsilon \cdot \max\{1, \eta R\},$$

let $x_1, ..., x_d$ be independent copies of x and put $X = (x_1, ..., x_d)$.

Let us identify conditions under which X satisfies the first part of Theorem 1.7.

Lemma 2.1. *There exists an absolute constant c_0 for which the following holds. Assume that $\delta < 1/2$ and that there is $L \geq 1$ such that for every $2 \leq q \leq p$, $R\delta^{1/q} \leq L\sqrt{q}$. Then for every $t \in \mathbb{R}^d$ and every $2 \leq q \leq p$,*

$$\|\langle X,t\rangle\|_{L_q} \leq c_0 L\sqrt{q}\|\langle X,t\rangle\|_{L_2}.$$

Moreover, for every $t \in \mathbb{R}^d$, $\|\langle X, t\rangle\|_{L_2} = c_1\|t\|_2$, and $1/\sqrt{2} \le c_1 \le 2L$.

In particular, X/c_1 is an isotropic random vector and for every $t \in \mathbb{R}^d$, $\langle X, t\rangle$ exhibits a c_0L-subgaussian moment growth up to the p-th moment.

The proof of Lemma 2.1 is based on a simple comparison argument:

Lemma 2.2. *Let $x_1, ..., x_d$ be symmetric, independent random variables and assume $z_1, ..., z_d$ are also symmetric and independent. If p is even and for every $1 \le j \le d$ and every $1 \le q \le p$, $\|x_i\|_{L_q} \le L\|z_i\|_{L_q}$, then for every $t \in \mathbb{R}^d$,*

$$\|\sum_{j=1}^d t_j x_j\|_{L_p} \le L\|\sum_{t=1}^d t_j z_j\|_{L_p}.$$

Proof. Observe that

$$\mathbb{E}(\sum_{j=1}^d t_j x_j)^p = \mathbb{E}\sum_{\vec{\beta}} c_{\vec{\beta}} \prod_{j=1}^d t_j^{\beta_j} x_j^{\beta_j} = \sum_{\vec{\beta}} c_{\vec{\beta}} \prod_{j=1}^d t_j^{\beta_j} \mathbb{E}x_j^{\beta_j},$$

with the sum taken over all choices of $\vec{\beta} = (\beta_1, ..., \beta_d) \in \{0, ..., p\}^d$, where $\sum_{j=1}^d \beta_j = p$ and $c_{\vec{\beta}}$ is the appropriate multinomial coefficient. Since $x_1, ..., x_d$ are symmetric, the only products that do not vanish are when $\beta_1, ..., \beta_d$ are even, and if $\beta_1, ..., \beta_d$ are even then

$$\prod_{j=1}^d t_j^{\beta_j} \mathbb{E}x_j^{\beta_j} \le \prod_{j=1}^d t_j^{\beta_j} L^{\beta_j} \mathbb{E}z_j^{\beta_j}.$$

Therefore,

$$\sum_{\vec{\beta}} c_{\vec{\beta}} \prod_{j=1}^d t_j^{\beta_j} \mathbb{E}x_j^{\beta_j} \le L^p \sum_{\vec{\beta}} c_{\vec{\beta}} \prod_{j=1}^d t_j^{\beta_j} \mathbb{E}z_j^{\beta_j} = L^p \mathbb{E}(\sum_{j=1}^d t_j z_j)^p. \qquad \square$$

Proof of Lemma 2.1. Note that $x = \varepsilon \max\{1, R\eta\}$ is symmetric and that $\mathbb{E}x^2 = 1 \cdot (1 - \delta) + R^2\delta$. Hence, if $\delta \le 1/2$ and $R^2\delta \le 2L^2$ then $1/2 \le \mathbb{E}x^2 \le 4L^2$—and the “moreover” part of the claim follows.

Turning to the first part of the claim, let $x_1, ..., x_d$ be independent copies of x, set g to be a standard gaussian random variable and let $g_1, ..., g_d$ be independent copies of g. Recall that for every $2 \le q \le p$, $R\delta^{1/q} \le L\sqrt{q}$; thus

$$(\mathbb{E}|x|^q)^{1/q} \le 1 + R\delta^{1/q} \le 2L\sqrt{q} \le c_1 L(\mathbb{E}|g|^q)^{1/q},$$

implying that $(x_1, ..., x_d)$ and $(g_1, ..., g_d)$ satisfy the conditions of Lemma 2.2 with a constant c_1L. Applying Lemma 2.2, it follows that for every $t \in S^{d-1}$ and every $2 \le q \le p$,

$$\|\sum_{j=1}^d t_j x_j\|_{L_q} \le c_1 L\|\sum_{j=1}^d t_j g_j\|_{L_q} \le c_2 L\sqrt{q}.$$

Hence, $\|\langle X, t\rangle\|_{L_q} \le c_3 L\sqrt{q}\|\langle X, t\rangle\|_{L_2}$, as claimed. $\square$

Now let η_{ij} be independent copies of η and put $x_{ij} = \varepsilon_{ij} \max\{1, \eta_{ij} R\}$ as above. The key part in the construction is the following lemma which describes the typical structure of the matrix generated by x,

$$\Gamma = (x_{ij})_{1 \le i \le m, 1 \le j \le d} : \mathbb{R}^d \to \mathbb{R}^m.$$

Lemma 2.3. *There exist absolute constants c_1, c_2, c_3 and c_4 for which the following holds. Let $\delta = c_1/d$ and $R \geq c_2 m$. Then, with probability at least $1 - 2\exp(-c_3 m)$:*

(1) there are indices $j_1 \neq j_2 \in \{1, ..., d\}$ and $1 \leq \ell \leq m$ such that $\eta_{\ell j_1} = \eta_{\ell j_2} = 1$ and for $i \neq \ell$, $\eta_{\ell j_1} = \eta_{\ell j_2} = 0$;

(2) *there is a subset $J \subset \{1, ..., d\}$ of cardinality $|J| = 2m$ such that $\eta_{ij} = 0$ for every $j \in J$ and $1 \leq i \leq m$;*

(3) *Moreover, for the subset J from (2) we have that $c_4 B_2^m \subset \Gamma B_1^J$, where $B_1^J = \{x = \sum_{j \in J} x_j e_j : \|x\|_1 \leq 1\}$.*

Corollary 2.4. *If Γ satisfies Lemma 2.3 then its column-normalized version $\tilde{\Gamma}$ does not satisfy the exact reconstruction property of order 2.*

Proof. Using the notation of Lemma 2.3 and by its first part, $\|\Gamma e_{j_1}\|_2 = \|\Gamma e_{j_2}\|_2 = (R^2 + m - 1)^{1/2}$; hence, if we denote by $\{f_1, ..., f_m\}$ the standard basis of $\mathbb{R}^m$,

$$\tilde{\Gamma} e_{j_1} = \frac{1}{(R^2 + m - 1)^{1/2}} \big(\varepsilon_{\ell j_1} R f_\ell + \sum_{i \neq \ell} \varepsilon_{i j_1} f_i\big)$$

and

$$\tilde{\Gamma} e_{j_2} = \frac{1}{(R^2 + m - 1)^{1/2}} \big(\varepsilon_{\ell j_2} R f_\ell + \sum_{i \neq \ell} \varepsilon_{i j_2} f_i\big).$$

If $\varepsilon_{\ell j_1} \neq \varepsilon_{\ell j_2}$ set $v = (e_{j_1} + e_{j_2})/2$; otherwise, set $v = (e_{j_1} - e_{j_2})/2$. In either case, v is 2-sparse. Let $w = \tilde{\Gamma} v$ and observe that the coordinates of w satisfy that

$$w_\ell = 0 \quad \text{and} \quad w_i^2 \leq \frac{1}{R^2 + m - 1} \quad \text{for} \quad i \neq \ell;$$

therefore,

$$\tilde{\Gamma} v \in \frac{\sqrt{m}}{R} B_2^m. \tag{2.1}$$

Next, let J be the set of coordinates given by the second part of Lemma 2.3. Clearly, $j_1, j_2 \notin J$ and

$$\Gamma^J = (x_{ij})_{1 \leq i \leq m, j \in J} = (\varepsilon_{ij})_{1 \leq i \leq m, j \in J}$$

is an $m \times 2m$ Bernoulli matrix. Therefore,

$$\tilde{\Gamma}^J = (\tilde{\Gamma}_{ij})_{1 \leq i \leq m, j \in J} = \frac{\Gamma^J}{\sqrt{m}}.$$

Observe that $\tilde{\Gamma} B_1^J = \tilde{\Gamma}^J B_1^J$ and by the third part of Lemma 2.3 there is an absolute constant c such that

$$c B_2^m \subset \Gamma B_1^J = \sqrt{m} \tilde{\Gamma} B_1^J. \tag{2.2}$$

Combining (2.1) with (2.2) it follows that if $\sqrt{m}/R \leq c/\sqrt{m}$ then $\tilde{\Gamma} v \in \tilde{\Gamma} B_1^J$. Since $\|v\|_1 = 1$ and $v \notin B_1^J$, it is evident that v is not the unique solution of the minimization problem

$$\min \|t\|_1 \quad \text{subject to} \quad \tilde{\Gamma} v = \tilde{\Gamma} t$$

and $\tilde{\Gamma}$ does not satisfy the exact reconstruction property of order 2. □

The proof of the first two parts of Lemma 2.3 is based on a standard fact that follows from Bernstein's inequality: if $W_1, ..., W_d$ are independent copies of a $\{0,1\}$-valued random variable W and $\mathbb{E} W = \mu$ then with probability at least $1 - 2\exp(-c\mu d)$, $\mu d/2 \leq |\{j : W_j = 1\}| \leq 3\mu d/2$.

Proof of Lemma 2.3. Let $\eta_1, ..., \eta_m$ be independent copies of η and let Y be the indicator of the event

$$\exists \ell \in \{1, ..., m\} \quad \eta_\ell = 1 \quad \text{and} \quad \eta_i = 0 \text{ for every } i \neq \ell.$$

Observe that $\mathbb{E}Y = m\delta(1-\delta)^{m-1}$ and that if $Y_1, ..., Y_d$ are independent copies of Y and $\mathbb{E}Y \geq 2m/d$ then with probability at least $1 - 2\exp(-c_1 m)$, $|\{i : Y_i = 1\}| > m$. In particular, on that event, the matrix $(\eta_{ij})_{1 \leq i \leq m, 1 \leq j \leq d}$ has at least two identical columns, each with a single entry of 1. Therefore, the first part of Lemma 2.3 holds if

$$m\delta(1-\delta)^{m-1} \geq \frac{2m}{d} \tag{2.3}$$

For the second part of the lemma, let Z be the indictor of the event

$$\eta_i = 0 \quad \text{for every } 1 \leq i \leq m$$

and note that $\mathbb{E}Z = (1-\delta)^m$. If $Z_1, ..., Z_d$ are independent copies of Z and $\mathbb{E}Z \geq 4m/d$ then with probability at least $1 - 2\exp(-c_2 m)$, $|\{i : Z_i = 1\}| \geq 2m$. Hence, if

$$(1-\delta)^m \geq \frac{4m}{d}, \tag{2.4}$$

then with probability at least $1 - 2\exp(-c_2 m)$, there is $J \subset \{1, ..., d\}$ and for every $j \in J$ and every $1 \leq i \leq m$, $\eta_{ij} = 0$.

Turning to the third part of the lemma, and by applying the second part, we have that conditioned on (η_{ij}), $x_{ij} = \varepsilon_{ij}$ for $(i,j) \in \{1, ..., m\} \times J$. Let $\Gamma^J = (\varepsilon_{ij})_{1 \leq i \leq m, j \in J}$ and recall that (ε_{ij}) are independent of (η_{ij}). Therefore, by Corollary 4.1 from [6], applied conditionally on (η_{ij}) there are absolute constants c_3 and c_4 for which, with probability at least $1 - 2\exp(-c_3 m)$,

$$c_4 B_2^m \subset \Gamma B_1^J.$$

Finally, all that remains is to see when (2.3) and (2.4) are satisfied, which clearly holds with the choice of $\delta = c/d$ and as long as $m \leq c'd$ for suitable absolute constants $c > 1$ and $0 < c' < 1$. □

To complete the proof of Theorem 1.7, let $\delta = c_1/d$ for $c_1 > 1$ large enough as above, set $p > 2$ and put $R = \sqrt{p}d^{1/p}$—a choice that complies with the conditions of Lemma 2.3 as long as

$$m \leq c_2 \sqrt{p} d^{1/p}. \tag{2.5}$$

It follows from Corollary 2.4 that with probability at least $1 - 2\exp(-c_3 m)$, the column-normalized matrix $\tilde{\Gamma}$ generated by $x = \varepsilon \max\{1, R\eta\}$ does not satisfy the exact reconstruction property of order 2. All that remains is to show that x also satisfies the assumptions of Lemma 2.1: that $R\delta^{1/q} \leq L\sqrt{q}$ for every $2 \leq q \leq p$ and for an absolute constant L.

To that end, let $\phi(x) = \sqrt{x}(d/c_1)^{1/x}$ and observe that $\phi(x)$ is decreasing when $2 \leq x \leq 2\log(d/c_1)$; hence, $\phi(p)/\phi(q) \leq 1$ for every $2 \leq q \leq p$ as long as $p \leq 2\log(d/c_1)$. If we set $L = c_1$ then $R\delta^{1/q} \leq \sqrt{q}$ for every $q \leq p$, as required.

References

[1] Peter J. Bickel, Ya'acov Ritov, and Alexandre B. Tsybakov, *Simultaneous analysis of lasso and Dantzig selector*, Ann. Statist. **37** (2009), no. 4, 1705–1732. MR-2533469

[2] Peter Bühlmann and Sara van de Geer, *Statistics for high-dimensional data*, Springer Series in Statistics, Springer, Heidelberg, 2011, Methods, theory and applications. MR-2807761

[3] Simon Foucart and Holger Rauhut, *A mathematical introduction to compressive sensing*, Applied and Numerical Harmonic Analysis, Birkhäuser/Springer, New York, 2013. MR-3100033

[4] Guillaume Lecué, *Private communication*.

[5] Guillaume Lecué and Shahar Mendelson, *Sparse recovery under weak moment assumptions*, J. Eur. Math. Soc. (JEMS) **19** (2017), no. 3, 881–904.

[6] A. E. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann, *Smallest singular value of random matrices and geometry of random polytopes*, Adv. Math. **195** (2005), no. 2, 491–523. MR-2146352

[7] Shahar Mendelson, *Learning without concentration*, Journal of the ACM (JACM) **62** (2015), no. 3, 21.

A Proof of Theorem 1.9

The proof is a direct consequence of the argument used in the proof of Theorem 1.4. Thanks to column normalization, $\tilde{\Gamma}$ satisfies (b) in Lemma 1.5 for $\beta = 1$. All that is left to verify is (a) for α which is a constant that depends only on L.

The proof of Theorem 1.4 reveals two fact: firstly, if Γ has $m \geq c_1(L) s \log(ed/s)$ independent rows that are distributed as X then with probability at least $1 - 2\exp(-c_2(L)m)$,

$$\inf_{t \in \Sigma_s} \|\Gamma t\|_2^2 = \inf_{t \in \Sigma_s} \sum_{i=1}^m \left\langle X_i, t \right\rangle^2 \geq c_3(L) m \|t\|_2^2;$$

and secondly, with probability at least $1 - 1/d^{c_4(L)}$,

$$\max_{1 \leq j \leq d} \|\Gamma e_j\|_2 \leq c_5(L)\sqrt{m}.$$

For every $t \in \Sigma_s$, set

$$\tilde{t} = \sum_{j=1}^d \frac{t_j}{\|\Gamma e_j\|_2} e_j,$$

which is also an s-sparse vector. Observe that $\tilde{\Gamma} t = \Gamma \tilde{t}$, implying that

$$\|\tilde{\Gamma} t\|_2^2 \geq c_3 m \sum_{j=1}^d \frac{t_j^2}{\|\Gamma e_j\|_2^2} \geq \frac{c_3}{c_5^2} \|t\|_2^2,$$

and (a) from Lemma 1.5 is verified for the matrix $\tilde{\Gamma}$ for $\alpha = c_6(L)$.