Instance-optimality in Probability with an ℓ_1-Minimization Decoder

Ronald DeVore, Guergana Petrova, and Przemyslaw Wojtaszczyk *

August 13, 2008

Abstract

Let $\Phi(\omega)$, $\omega \in \Omega$, be a family of $n \times N$ random matrices whose entries $\phi_{i,j}$ are independent realizations of a symmetric, real random variable η with expectation $\mathbb{E}\eta = 0$ and variance $\mathbb{E}\eta^2 = 1/n$. Such matrices are used in compressed sensing to encode a vector $x \in \mathbb{R}^N$ by $y = \Phi x$. The information y holds about x is extracted by using a decoder $\Delta : \mathbb{R}^n \to \mathbb{R}^N$. The most prominent decoder is the ℓ_1-minimization decoder Δ which gives for a given $y \in \mathbb{R}^n$ the element $\Delta(y) \in \mathbb{R}^N$ which has minimal ℓ_1-norm among all $z \in \mathbb{R}^N$ with $\Phi z = y$. This paper is interested in properties of the random family $\Phi(\omega)$ which guarantee that the vector $\bar{x} := \Delta(\Phi x)$ will with high probability approximate x in ℓ_2^N to an accuracy comparable with the best k-term error of approximation in ℓ_2^N for the range $k \leq an/\log_2(N/n)$. This means that for the above range of k, for each signal $x \in \mathbb{R}^N$, the vector $\bar{x} := \Delta(\Phi x)$ satisfies

$$\|x - \bar{x}\|_{\ell_2^N} \leq C \inf_{z \in \Sigma_k} \|x - z\|_{\ell_2^N}$$

with high probability on the draw of Φ. Here, Σ_k consists of all vectors with at most k nonzero coordinates. The first result of this type was proved by Wojtaszczyk [19] who showed this property when η is a normalized Gaussian random variable. We extend this property to more general random variables, including the particular case where η is the Bernoulli random variable which takes the values $\pm 1/\sqrt{n}$ with equal probability. The proofs of our results use geometric mapping properties of such random matrices some of which were recently obtained in [14].

1 Introduction

Compressed sensing is a new paradigm in signal processing whose goal is to acquire signals with as few measurements (samples) as possible. It has its theoretical origins in the results of Kashin [13] and Gluskin-Garneev [10] on Gelfand widths from the 1970's but it was

*This research was supported by the Office of Naval Research Contracts N00014-03-1-0051, N00014-08-1-1113, N00014-03-1-0675, and N00014-05-1-0715; the ARO/DoD Contracts W911NF-05-1-0227 and W911NF-07-1-0185; the NSF Grant DMS-0810869; and Polish MNiSW grant N201 269335

recently put into the practical domain of signal processing with the work of Candés-Tao [6] and Donoho [9].

In the discrete setting of this paper, the signal is represented by a vector $x \in R^N$ where N is large. A sample of x is its inner product with a vector $v \in \mathbb{R}^N$. Taking n samples is then represented by the application of an $n \times N$ matrix Φ whose rows are the vectors with which we take inner products. Thus,

$$y := \Phi x \tag{1.1}$$

is the information we record about x. To extract this information, we apply a decoder Δ to y which is, typically, a nonlinear operator mapping from $\mathbb{R}^n$ to $\mathbb{R}^N$. The vector

$$\bar{x} := \Delta(\Phi x) \tag{1.2}$$

is viewed as an approximation to x.

There are several ways to measure the performance of an encoder-decoder pair (Φ, Δ). The finest measures choose a norm $\|\cdot\|_X$ on $\mathbb{R}^N$ and compare the error $\|x - \bar{x}\|_X$ with the corresponding error of k-term approximation. To describe the latter, let Σ_k denote the set of all vectors in $\mathbb{R}^N$ which have at most k nonzero coordinates. Then, Σ_k is a nonlinear space which is the union of the $\binom{N}{k}$ linear spaces X_T, $T \subset \{1, \ldots, N\}$, with $\#(T) \leq k$. Here X_T consists of all vectors $x \in \mathbb{R}^N$ which vanish outside of T. The error of best k-term approximation is

$$\sigma_k(x)_X := \inf\{\|x - z\|_X : \ z \in \Sigma_k\}. \tag{1.3}$$

If for a value of k there is a constant $C > 0$ such that for all $x \in \mathbb{R}^N$ we have

$$\|x - \Delta(\Phi x)\|_X \leq C\sigma_k(x)_X \tag{1.4}$$

then the pair (Φ, Δ) is said to be *instance-optimal* in X of order k with constant C.

If X is one of the ℓ_p^N spaces with the (quasi-)norm

$$\|x\|_{\ell_p^N} := \begin{cases} \left(\sum_{i=1}^N |x_j|^p\right)^{1/p}, & 0 < p < \infty, \\ \max_{j=1,\ldots,N} |x_j|, & p = \infty, \end{cases} \tag{1.5}$$

then a best k-term approximation to x is obtained by retaining the k largest coordinates of x (with ties handled in an arbitrary way) and setting all other coordinates to zero. Thus, an instance-optimal pair (Φ, Δ) of order k performs almost the same as identifying the k largest coordinates of x and using these to approximate x. The best pairs (Φ, Δ) are those which give instance-optimality of the highest order k. Note that an instance-optimal pair of order k will automatically recover exactly any vector $x \in \Sigma_k$, i.e. any k-sparse vector.

If we fix X, the dimensions n, N and the constant C, then there is a largest value of k for which we can have instance-optimality. Upper and lower bounds on the largest possible k were proved in [7] for the case when X is an ℓ_p^N space. We mention two contrasting results. If $X = \ell_1^N$, the instance-optimality holds for[1] $k \leq cn/\log(N/n)$ where c depends

[1]Here and later all logarithms are taken with respect to the base 2.

only on the constant C in (1.4). In going further, we shall refer to this range of k as the large range of k since it is known from results on Gelfand widths that instance-optimality can never hold for a larger range (save for the constant c). This result for instance-optimality in ℓ_1^N should be compared with what happens when $X = \ell_2^N$. In this case, the largest range of k for instance-optimality is $k \le cn/N$ where again c depends only on C. We see that even instance-optimality of order one will not hold unless the number of measurements n is of the same order as N. In this sense, we can say the compressed sensing systems do not perform well if we wish to measure the error in the norm of ℓ_2^N.

Returning to the case of $X = \ell_1^N$, the only systems which are provably instance-optimal for the large range of k given above, are constructed using probability. Namely, various constructions of random matrices are shown to yield a favorable Φ with high probability. No deterministic constructions are known for this large range (see [8] for a deterministic construction for a much more narrow range of k).

Given that there are no deterministic constructions of matrices for the large range of k, several authors, including those in [7], suggest that a more meaningful measure of performance of a encoder-decoder pair is instance-optimality in probability. By this we mean the following. Suppose $\Phi(\omega)$, $\omega \in \Omega$, is a random family of matrices on the probability space (Ω, ρ) and $\Delta(\omega)$, $\omega \in \Omega$, is a corresponding family of decoders. We say that the family of pairs $(\Phi(\omega), \Delta(\omega))$ is instance-optimal in probability of order k in X with constant C, if for each $x \in \mathbb{R}^N$, we have that

$$\|x - \Delta(\Phi x)\|_X \le C\sigma_k(x)_X \tag{1.6}$$

holds with high probability.

Surprisingly, it was shown in [7] that classical constructions of random families $\Phi(\omega)$ can be used with certain decoders $\Delta(\omega)$ to attain instance-optimality in probability in ℓ_2^N for the large range of k. Thus, from this new viewpoint, instance-optimality in probability performs the same for ℓ_2^N as it does for ℓ_1^N. There was, however, one dampening factor in the results of [7]. Namely, the decoders used in establishing instance-optimality in probability in ℓ_2^N were completely impractical and could never be implemented numerically. This led to the question of whether practical decoders such as ℓ_1-minimization or greedy algorithms gave this high level of performance.

Recently, Wojtaszczyk [19] proved that if the random family was given by filling out the entries in Φ using the Gaussian distribution with mean zero and variance $1/n$, then this could be coupled with ℓ_1-minimization to give a compressed sensing pair which is instance-optimal in probability in ℓ_2^N for the large range of k. Wojtaszczyk's proof rested heavily on the following geometrical property of the Gaussian family: with high probability, a draw of this matrix will map the unit ball in ℓ_1^N onto a set containing an ℓ_2^n ball about the origin of radius $c\sqrt{\log(N/n)}/\sqrt{n}$.

Unfortunately, this geometric property does not hold for all classical random constructions. For example, if the matrix $\Phi(\omega)$ has its entries given by independent draws of $\pm 1/\sqrt{n}$, the resulting Bernoulli matrix cannot satisfy that property. Indeed, the vector with first coordinate one and all other coordinates zero can be the image of a vector $x \in \mathbb{R}^N$ under any of these matrices only if $\|x\|_{\ell_1^N} \ge \sqrt{n}$. This means the unit ball of ℓ_1^N cannot cover an ℓ_2^n ball of radius $< 1/\sqrt{n}$.

The purpose of the present paper is to point out that a weaker geometric property of random matrices, studied already by A. Litvak, A. Pajor, M. Rudelson and N. Tomczak-Jaegermann in [14], when coupled with decoding by ℓ_1-minimization will yield instance-optimality in probability in ℓ_2^N for the large range of k. This new geometric property replaces the role of the ℓ_2^n ball as an image by an ℓ_2^n ball intersected with an ℓ_∞^n ball of smaller radius.

The organization of this paper is as follows. The first sections of the paper concentrate on proving instance-optimality for Bernoulli matrices where the proofs are most transparent. In Section 3, we present a geometric mapping property of Bernoulli matrices, see Theorem 3.5, which can be derived as a special case from Theorem 4.2 in [14]. We use this property to prove in Section 4 that instance-optimality in probability in ℓ_2^N holds for Bernoulli matrices. We organize our arguments to extract the essential properties of Bernoulli matrices that are needed for the proof. In Section 5, we show that these properties hold for quite general random families and thereby obtain a broad generalization of the Bernoulli case.

2 Preliminary results and notation

In the first sections of this paper, we let $\Phi = \Phi(\omega) := (\phi_{i,j})$ denote the random family of $n \times N$ Bernoulli matrices. Here $\phi_{i,j} = \frac{1}{\sqrt{n}} r_{i,j}$, where the $r_{i,j}$ are independent Bernouli random variables which take the values ± 1 each with probability 1/2. We denote by $\Phi_j \in \mathbb{R}^n$, $j = 1, \ldots, N$, the columns of Φ and introduce the abbreviated notation $L := \log(N/n)$ since this term appears frequently.

From the fact that the random variables $\phi_{i,j}$ are independent and have zero mean, it is easy to deduce that for any $x \in \mathbb{R}^N$, the random variable $\|\Phi(\omega)x\|_{\ell_2^n}^2$ has expected value $\|x\|_{\ell_2^N}^2$, that is,

$$\mathbb{E}(\|\Phi(\omega)x\|_{\ell_2^n}^2) = \|x\|_{\ell_2^N}^2. \tag{2.1}$$

There are also standard estimates that show that this random variable is strongly concentrated about its mean. Namely, we have[2]

Concentration of Measure Property (CMP) for Bernoulli random matrices: *For any $x \in \mathbb{R}^N$ and any $0 < \delta < 1$, there is a set $\Omega_0(x, \delta)$ with*

$$\rho(\Omega_0(x,\delta)^c) \leq C_0 e^{-nc_0(\delta)}, \tag{2.2}$$

such that for each $\omega \in \Omega_0(x, \delta)$ we have

$$(1-\delta)\|x\|_{\ell_2^N}^2 \leq \|\Phi(\omega)x\|_{\ell_2^n}^2 \leq (1+\delta)\|x\|_{\ell_2^N}^2. \tag{2.3}$$

For example, this concentration of measure property is proved in [1] with $c_0(\delta) = \delta^2/4 - \delta^3/6$ and $C_0 = 2$. We will use these values for our analysis of the Bernoulli random matrices.

[2] In this paper we will use S^c to denote the complement of a set S.

There are several important consequences that can be drawn from the CMP. As a first example, we mention the Restricted Isometry Property (RIP) as introduced by Candés, Romberg, and Tao [4]. Given an $n \times N$ matrix A, it is said to have RIP of order k with constant δ if

$$(1-\delta)\|z\|_{\ell_2^N} \leq \|Az\|_{\ell_2^n} \leq (1+\delta)\|z\|_{\ell_2^N} \tag{2.4}$$

holds for all $z \in \Sigma_k$.

It was shown in [2, Th. 5.2] that any random family of matrices which satisfies a CMP as above, will automatically satisfy the RIP of order k for any $k \leq cn/L$ with high probability. In our analysis of Bernoulli matrices, we shall use the following special case. There are absolute constants $\tilde{c}_1, \tilde{C}_1 > 0$, $\tilde{a} > 0$ and sets $\tilde{\Omega}_1(k)$, with

$$\rho(\tilde{\Omega}_1(k)^c) \leq \tilde{C}_1 e^{-\tilde{c}_1 n} \tag{2.5}$$

such that for each $k \leq \tilde{a}n/L$ and each $\omega \in \tilde{\Omega}_1(k)$, the matrix $\Phi = \Phi(\omega)$ satisfies the RIP property of order $2k$ with constant $\delta = 1/4$, i.e.

$$\frac{3}{4}\|z\|_{\ell_2^N} \leq \|\Phi(\omega)z\|_{\ell_2^n} \leq \frac{5}{4}\|z\|_{\ell_2^N}, \quad z \in \Sigma_{2k}. \tag{2.6}$$

To close out this section, we wish to prove another simple fact about Bernoulli matrices which can be derived easily from the CMP.

Lemma 2.1 *For each $x \in I\!R^N$ there is a set $\Omega_1(x)$ with*

$$\rho(\Omega_1(x)^c) \leq 2e^{-n/24} + 2ne^{\frac{-n}{2L}} \tag{2.7}$$

such that for all $\omega \in \Omega_1(x)$, the $n \times N$ normalized Bernoulli matrix $\Phi = \Phi(\omega)$ satisfies

$$\|\Phi x\|_{\ell_2^n} \leq \sqrt{\frac{3}{2}}\|x\|_{\ell_2^N}, \tag{2.8}$$

and

$$\|\Phi x\|_{\ell_\infty^n} \leq \frac{1}{\sqrt{L}}\|x\|_{\ell_2^N}. \tag{2.9}$$

Proof: Without loss of generality we can assume that $\|x\|_{\ell_2^N} = 1$. Fix such an x. We already know that (2.8) holds for $\omega \in \Omega_0(x, 1/2)$ (see (2.2) and (2.3)), where $\rho(\Omega_0(x,1/2)^c) \leq 2e^{-n/24}$. We concentrate on establishing the ℓ_∞^n bound. We note that each entry y_i of y is of the form

$$y_i = \frac{1}{\sqrt{n}} \sum_{j=1}^{N} x_j r_{i,j}, \tag{2.10}$$

where the $r_{i,j}$ are independent Bernoulli random variables and $x = (x_1, \ldots, x_N)$. We shall use Hoeffding's inequality (see page 596 of [11]) which says that for independent mean zero random variables ϵ_j taking values in $[a_j, b_j]$, $j = 1, \ldots, N$, we have

$$\Pr\left(|\sum_{j=1}^{N} \epsilon_j| \geq \delta\right) \leq 2e^{\frac{-2\delta^2}{\sum_{j=1}^{N}(b_j-a_j)^2}}. \tag{2.11}$$

We apply this to the random variables $\epsilon_j := \frac{1}{\sqrt{n}} x_j r_{i,j}$, $j = 1, \ldots, N$, which take values in $\frac{1}{\sqrt{n}}[-x_j, x_j]$. Since $\sum_{j=1}^N (2x_j)^2 = 4$, we deduce that

$$\Pr\left(|y_i| \geq \delta\right) \leq 2e^{\frac{-n\delta^2}{2}}. \tag{2.12}$$

Applying a union bound, we get

$$\Pr\left(\|y\|_{\ell_\infty^n} \geq \delta\right) \leq 2ne^{\frac{-n\delta^2}{2}}. \tag{2.13}$$

We now take $\delta = 1/\sqrt{L}$ and deduce

$$\Pr\left(\|y\|_{\ell_\infty^n} \geq 1/\sqrt{L}\right) \leq 2ne^{\frac{-n}{2L}}. \tag{2.14}$$

We now can take $\Omega_1(x) := \Omega_0(x, 1/2) \cap \{\omega : \|y(\omega)\|_{\ell_\infty n} \leq 1/\sqrt{L}\}$. Then, (2.7) follows from a union bound on probabilities. The estimate (2.8) follows from the upper bound in (2.3) and (2.9) follows from the definition of $\Omega_1(x)$ and (2.14). □

Remark 2.2 *Note that in the above lemma we could require a much smaller ℓ_∞^n bound on y and still achieve this with high probability.*

3 Geometric mapping property of Bernoulli matrices

In this section, we derive a geometric mapping property of Bernoulli matrices, stated in Theorem 3.5. As we noted in the introduction this is a special case of Theorem 4.2 from [14]. We decided to include this proof (which is a simple consequence of a well known result of Montgomery-Smith [16]) in order to keep our results for the important special case of Bernoulli matrices easily accessible.

To formulate this geometric property, we define the following norm on $\mathbb{R}^n$

$$\|y\|_J := \max\left\{\sqrt{n}\|y\|_{\ell_\infty^n}, \sqrt{\frac{n}{L}}\|y\|_{\ell_2^n}\right\}. \tag{3.1}$$

Notice that $\|y\|_J \leq 1$ is equivalent to $\|y\|_{\ell_\infty^n} \leq 1/\sqrt{n}$ and $\|y\|_{\ell_2^n} \leq \sqrt{\frac{L}{n}}$. A second norm on $\mathbb{R}^n$ we are interested in is

$$\|y\|_\Phi := \min\{\|x\|_{\ell_1^N} : \ \Phi x = y\}, \tag{3.2}$$

where Φ is a normalized Bernoulli matrix. It will follow from the arguments given below that with high probability Φ has full rank and thus this is a norm.

Our main goal is to compare these two norms. We want to prove that there is an absolute constant C such that with high probability on the draw of Φ, we have

$$\|y\|_\Phi \leq C\|y\|_J, \quad \text{for all } y \in \mathbb{R}^n. \tag{3.3}$$

Rather than do this directly, we will do this by duality. Given a norm $\|\cdot\|$ on $I\!R^n$, the dual space of $(I\!R^n, \|\cdot\|)$ is $I\!R^n$ with the dual norm $\|\cdot\|^*$ defined for $\lambda \in I\!R^n$ by

$$\|\lambda\|^* := \sup_{\|y\|=1} \langle \lambda, y \rangle = \sup_{\|y\|\le 1} \langle \lambda, y \rangle. \tag{3.4}$$

Of course, we have $\|\lambda\|^*_{\ell_1^n} = \|\lambda\|_{\ell_\infty^n}$ and $\|\lambda\|^*_{\ell_2^n} = \|\lambda\|_{\ell_2^n}$. In order to prove (3.3) for a given Φ, it is sufficient to show that

$$\|\lambda\|^*_J \le C\|\lambda\|^*_\Phi, \quad \forall \lambda \in I\!R^n. \tag{3.5}$$

Indeed, if we have (3.5) then we have

$$\|y\|_\Phi = \sup_{\|\lambda\|^*_\Phi = 1} \langle \lambda, y \rangle \le \sup_{\|\lambda\|^*_J \le C} \langle \lambda, y \rangle \le C\|y\|_J, \quad y \in I\!R^n. \tag{3.6}$$

So we shall now concentrate on proving that (3.5) holds with high probability on the draw of Φ. We begin by giving a description of these dual norms. For this, we define

$$\mathcal{K}(\lambda, t) := \inf_{\lambda = \lambda_1 + \lambda_2} \{\|\lambda_1\|_{\ell_1^n} + t\|\lambda_2\|_{\ell_2^n}\}, \quad t > 0, \tag{3.7}$$

which is the K-functional between ℓ_1^n and ℓ_2^n.

The following lemma is well-known (see Lemma 1 of [16]). We include its proof for completeness since we could not find a proof in the literature.

Lemma 3.1 *For any $\lambda \in I\!R^n$, we have*

$$\|\lambda\|^*_\Phi = \max_{1\le j\le N} |\langle \lambda, \Phi_j \rangle| = \frac{1}{\sqrt{n}} \max_{1\le j\le N} |\langle \lambda, \sqrt{n}\Phi_j \rangle|, \tag{3.8}$$

and

$$\|\lambda\|^*_J = \frac{1}{\sqrt{n}} \inf_{\lambda=\lambda_1+\lambda_2} \{\|\lambda_1\|_{\ell_1^n} + \sqrt{L}\|\lambda_2\|_{\ell_2^n}\} = \frac{1}{\sqrt{n}}\mathcal{K}(\lambda, \sqrt{L}). \tag{3.9}$$

Proof: We first prove (3.8). Since $\{y : \|y\|_\Phi \le 1\} = \{\Phi x : \|x\|_{\ell_1^N} \le 1\}$, for any $\lambda \in I\!R^n$, we have

$$\|\lambda\|^*_\Phi = \sup_{\|y\|_\Phi \le 1} \langle \lambda, y \rangle = \sup_{\|x\|_{\ell_1^N} \le 1} \sum_{i=1}^n \lambda_i \sum_{j=1}^N \phi_{i,j} x_j = \sup_{\|x\|_{\ell_1^N} \le 1} \langle \Phi^t \lambda, x \rangle = \|\Phi^t \lambda\|_{\ell_\infty^N} \tag{3.10}$$

which proves (3.8). Here Φ^t is the transpose of the matrix Φ.

Next, we prove (3.9). Let us first observe that for any decomposition $\lambda = \lambda_1 + \lambda_2$, we have

$$\|\lambda\|^*_J = \sup_{\|y\|_J \le 1} \langle \lambda, y \rangle = \sup_{\|y\|_J \le 1} \langle \lambda_1 + \lambda_2, y \rangle \le \frac{1}{\sqrt{n}}\|\lambda_1\|_{\ell_1^n} + \sqrt{\frac{L}{n}}\|\lambda_2\|_{\ell_2^n}, \tag{3.11}$$

because whenever $\|y\|_J \le 1$ then $\|y\|_{\ell_2^n} \le \sqrt{\frac{L}{n}}$ and $\|y\|_{\ell_\infty^n} \le 1/\sqrt{n}$. If we now take an infimum over all decompositions $\lambda = \lambda_1 + \lambda_2$, we see that the left side of (3.9) does not excede the right side.

To show that the right side of (3.9) does not exceed the left side, let us define the norm $\|(\alpha,\beta)\| = \max\{\|\alpha\|_{\ell_\infty^n}, \|\beta\|_{\ell_2^n}\}$ on the space $I\!R^n \oplus I\!R^n$. We have $\|(\alpha',\beta')\|^* = \|\alpha'\|_{\ell_1^n} + \|\beta'\|_{\ell_2^n}$. From (3.1) we see that the mapping S from $(I\!R^n, \|\cdot\|_J)$ to $(I\!R^n \oplus I\!R^n, \|\cdot\|)$ defined by $S(z) = (\sqrt{n}z, \sqrt{\frac{n}{L}}z)$ is an isometry onto its range Z which is a subspace of dimension n. That is $\|z\|_J = \|S(z)\|$. Now λ induces a functional with norm $\|\lambda\|_J^*$ on Z and using the Hahn–Banach theorem we see that there exists a pair $(\mu_1,\mu_2) \in I\!R^n \oplus I\!R^n$ such that

$$\begin{aligned} \langle z,\lambda\rangle &= \langle S(z),(\mu_1,\mu_2)\rangle \text{ for each } z \in I\!R^n, && (3.12)\\ \|\lambda\|_J^* &= \|(\mu_1,\mu_2)\|^* = \|\mu_1\|_{\ell_1^n} + \|\mu_2\|_{\ell_2^n}. && (3.13)\end{aligned}$$

From (3.12) we get $\langle z,\lambda\rangle = \langle z, \sqrt{n}\mu_1 + \sqrt{\frac{n}{L}}\mu_2\rangle$ for each $z \in I\!R^n$, which gives $\lambda = \lambda_1 + \lambda_2$ where $\lambda_1 = \sqrt{n}\mu_1$ and $\lambda_2 = \sqrt{\frac{n}{L}}\mu_2$. From (3.13) we get

$$\|\lambda\|_J^* = \frac{1}{\sqrt{n}}\{\|\lambda_1\|_{\ell_1^n} + \sqrt{L}\|\lambda_2\|_{\ell_2^n}\} \geq \frac{1}{\sqrt{n}}\mathcal{K}(\lambda,\sqrt{L}). \tag{3.14}$$

□

Our next goal is to prove that with high probability on the draw of Φ, we have for each individual $\lambda \in I\!R^n$ the estimate $\|\lambda\|_J^* \leq C\|\lambda\|_\Phi^*$ with an absolute constant C. To prove this, we need to bound from below the probability that a linear combination of Bernoulli random variables can be small. We shall call this the *Lower Bound Property* (LBP). Such a LBP was proven by Montgomery-Smith [16] who showed that for each $\lambda \in I\!R^n$, each $t > 0$ and independent Bernoulli random variables r_i, $i = 1,\ldots,n$, we have

$$\Pr\left(|\sum_{i=1}^n \lambda_i r_i| > \frac{1}{2}\mathcal{K}(\lambda,t)\right) \geq 2e^{-c_1t^2}, \quad c_1 = 4\ln 24. \tag{3.15}$$

Since for any $C \geq 1$, we have $C\mathcal{K}(\lambda,t) \geq \mathcal{K}(\lambda,Ct)$, it follows that

$$\Pr\left(2C|\sum_{i=1}^n \lambda_i r_i| \geq \mathcal{K}(\lambda,Ct)\right) \geq 2e^{-c_1t^2}, \tag{3.16}$$

for any $C \geq 1$. [3] Now we use (3.15) to prove the following lemma.

Lemma 3.2 *Given n, N and any $\lambda \in I\!R^n$, there is a set $\Omega_1(\lambda)$ with*

$$\rho(\Omega_1(\lambda)^c) \leq e^{-2\sqrt{Nn}}, \tag{3.17}$$

such that for each $\Phi = \Phi(\omega)$, $\omega \in \Omega_1(\lambda)$, we have

$$\|\lambda\|_J^* \leq C_1\|\lambda\|_\Phi^*, \tag{3.18}$$

with $C_1 := 2\sqrt{2c_1}$.

[3] The estimates (3.15) and (3.16) are formulated in [16] with slightly different constants. Our constants immediately follow from arguments in [16].

Remark 3.3 *There should be no confusion between the sets* $\Omega_1(x)$ *of Lemma* 2.1 *and the sets* $\Omega_1(\lambda)$ *since* $x \in \mathbb{R}^N$ *and* $\lambda \in \mathbb{R}^n$.

Proof: We fix λ. With a view towards Lemma 3.1 we want to find a set $\Omega_1(\lambda)$ (with the favorable lower bounds on its measure) such that for $\omega \in \Omega_1(\lambda)$ we have

$$\mathcal{K}(\lambda, \sqrt{L}) \leq C_1 \max_{1\leq j\leq N} |\langle \lambda, \sqrt{n}\Phi_j \rangle|, \tag{3.19}$$

where Φ_j are the columns of $\Phi(\omega)$.

Each of the inner products appearing on the right side of (3.19) is a sum of the type appearing in (3.16). We take $t = \sqrt{\frac{L}{2c_1}}$ and $C = \sqrt{2c_1} > 1$ in (3.16) and obtain

$$\Pr\left(C_1 |\sum_{i=1}^{n} \lambda_i r_i| \geq \mathcal{K}(\lambda, \sqrt{L}) \right) \geq 2e^{-L/2} = 2\sqrt{\frac{n}{N}}, \quad C_1 = 2\sqrt{2c_1}. \tag{3.20}$$

Now define $\Omega_1(\lambda)$ as the set of all $\omega \in \Omega$ such that (3.20) holds for $\langle \sqrt{n}\Phi_j, \lambda \rangle$, $j = 1, \dots, N$, with Φ_j the columns of $\Phi(\omega)$. Each inner product of λ with $\sqrt{n}\Phi_j$ is a sum of the above form and so

$$\Pr\left(C_1 \max_{1\leq j\leq N} |\langle \lambda, \sqrt{n}\Phi_j \rangle| \leq \mathcal{K}(\lambda, \sqrt{L}) \right) \leq \left[1 - 2\sqrt{\frac{n}{N}} \right]^N. \tag{3.21}$$

Since $1 - x \leq e^{-x}$ for $x \geq 0$, we have

$$\Pr\left(C_1 \max_{1\leq j\leq N} |\langle \lambda, \sqrt{n}\Phi_j \rangle| \leq \mathcal{K}(\lambda, \sqrt{L}) \right) \leq e^{-2\sqrt{Nn}}. \tag{3.22}$$

This shows that (3.18) holds with probability $\geq 1 - e^{-2\sqrt{Nn}}$ as desired. □

The lemma shows that for each $\lambda \in \mathbb{R}^n$, we have the desired inequality with high probability. To obtain a uniform bound with high probability, we use a covering argument.

Lemma 3.4 *Let* B_Φ^n *be the unit ball in* $(\mathbb{R}^n, \|\cdot\|_\Phi^*)$. *Then for each* $\epsilon > 0$ *there is a set of points* $\Lambda_\epsilon \subset B_\Phi^n$ *with* $\#(\Lambda_\epsilon) \leq (3/\epsilon)^n$ *and for each* $\lambda \in B_\Phi^n$ *there is a* $\lambda' \in \Lambda_\epsilon$ *such that* $\|\lambda - \lambda'\|_\Phi^* \leq \epsilon$.

Proof: This is a classical result about entropy in finite dimensional spaces (see Proposition 1.3 of Chapter 15 in [15]). □

We now prove the main result of this section, which is a particular case of Theorem 4.2 from [14].

Theorem 3.5 *Let* $C_1' := 2C_1$, *where* C_1 *is the constant in Lemma* 3.2. *Then, there is a set* Ω_1 *with*

$$\rho(\Omega_1^c) \leq e^{-\sqrt{Nn}} \tag{3.23}$$

such that for each $\omega \in \Omega_1$, *any* $n \times N$ *normalized Bernoulli matrix* $\Phi := \Phi(\omega)$ *with* $N \geq [\ln 6]^2 n$ *satisfies*

$$\|y\|_\Phi \leq C_1' \|y\|_J, \quad \forall y \in \mathbb{R}^n. \tag{3.24}$$

Proof: We take $\epsilon = 1/2$ and apply Lemma 3.4 to find a set of points Λ_ϵ with cardinality $\#(\Lambda) \leq 6^n$ satisfying the lemma for this choice of ϵ. Let $\Omega_1 := \cap_{\lambda' \in \Lambda_\epsilon} \Omega_1(\lambda')$, where $\Omega_1(\lambda')$ are the sets of Lemma 3.2. Then,

$$\rho(\Omega_1^c) \leq \#(\Lambda_\epsilon) e^{-2\sqrt{Nn}} \leq e^{n \ln 6 - 2\sqrt{Nn}} \leq e^{-\sqrt{Nn}}, \tag{3.25}$$

provided $N \geq [\ln 6]^2 n$. Fix any $\omega \in \Omega_1$ and let $C^* = C^*(\omega)$ be the smallest constant such that for $\Phi = \Phi(\omega)$, we have

$$\|\lambda\|_J^* \leq C^* \|\lambda\|_\Phi^*, \quad \forall \lambda \in I\!\!R^n. \tag{3.26}$$

The existence of such C^* follows because any two norms on a finite dimensional space are equivalent. Inequality (3.26) is equivalent to $\|\lambda\|_J^* \leq C^*$ for all $\lambda \in I\!\!R^n$ such that $\|\lambda\|_\Phi^* = 1$. Let us fix such a λ. Then, by Lemma 3.4 for $\epsilon = 1/2$, there is $\lambda' \in \Lambda_{1/2}$, such that

$$\|\lambda - \lambda'\|_\Phi^* \leq \frac{1}{2}, \quad \|\lambda'\|_\Phi^* \leq 1, \tag{3.27}$$

and by Lemma 3.2, since $\omega \in \Omega_1$,

$$\|\lambda'\|_J^* \leq C_1 \|\lambda'\|_\Phi^* \leq C_1. \tag{3.28}$$

It follows then from (3.26), (3.27) and (3.28) that

$$\|\lambda\|_J^* \leq \|\lambda - \lambda'\|_J^* + \|\lambda'\|_J^* \leq C^* \|\lambda - \lambda'\|_\Phi^* + \|\lambda'\|_J^* \leq \frac{C^*}{2} + C_1.$$

From the minimality of C^*, we have $C^* \leq C_1 + C^*/2$. This implies that $C^* \leq 2C_1 = C_1'$, which proves that

$$\|\lambda\|_J^* \leq C_1' \|\lambda\|_\Phi^*, \quad \forall \lambda \in I\!\!R^n.$$

As noted at the beginning of this section, this inequality is sufficient to show (3.24). □

4 Instance Optimality for Bernoulli

In this section we show how the geometric fact established in the previous section can be used to prove instance-optimatimality in probability in ℓ_2^N for the ℓ_1-minimization decoder. Let us write

$$U_J := \left\{ y \in I\!\!R^n : \ \|y\|_{\ell_\infty^n} \leq 1/\sqrt{n}, \ \|y\|_{\ell_2^n} \leq \sqrt{\frac{L}{n}} \right\} \tag{4.1}$$

for the unit ball in $\|\cdot\|_J$.

Theorem 4.1 *Let C_1' be the constant from Theorem 3.5 and $\tilde{C}_1, \tilde{c}_1$ and $\tilde{a}$ be the constants in the RIP (2.5),(2.6). There is an absolute constant C_2 and a set Ω_2 with*

$$\rho(\Omega_2^c) \leq \tilde{C}_1 e^{-\tilde{c}_1 n} + e^{-\sqrt{Nn}}, \tag{4.2}$$

such that for each $\omega \in \Omega_2$, the $n \times N$ normalized Bernoulli matrix $\Phi(\omega)$, $N \geq [\ln 6]^2 n$, has the following property. For each $y \in U_J$ there is $z \in I\!\!R^N$, such that $y = \Phi z$, and $\|z\|_{\ell_1^N} \leq C_1'$ and $\|z\|_{\ell_2^N} \leq C_2 \frac{1}{\sqrt{k}}$, for all $k \leq \tilde{a} n / L$.

Proof: We define $\Omega_2 := \tilde{\Omega}_1(k) \cap \Omega_1$, where $\tilde{\Omega}_1(k)$ is the set of the RIP of order $2k$ with constant $\delta = 1/4$ and Ω_1 is the set from Theorem 3.5. Because of (3.23) and (2.5), we have that

$$\rho(\Omega_2^c) \le \tilde{C}_1 e^{-\tilde{c}_1 n} + e^{-\sqrt{Nn}}$$

and so (4.2) is satisfied.

Now let $y \in U_J$. Then $\|y\|_J \le 1$ and by Theorem 3.5, for each $\omega \in \Omega_2 \subset \Omega_1$ the matrix $\Phi := \Phi(\omega)$ has the property that there is $z \in I\!R^N$, such that $y = \Phi z$ and $\|z\|_{\ell_1^N} \le C_1'$.

We have left to prove that $\|z\|_{\ell_2^N} \le C_2 \frac{1}{\sqrt{k}}$. For this, we follow the argument from [5], used also in [19]. Consider the vector z^* which is the decreasing rearrangement of z, i.e. $z^* = (z_{i_1}, \ldots, z_{i_N})$, where $|z_{i_1}| \ge \ldots \ge |z_{i_N}|$. Let $T_0 = \{i_1, \ldots, i_k\}$ be the set of the first k indices, T_1 be the set of the next k indices and so on. The last set T_s may have $< k$ elements. Clearly $\{1, \ldots, N\} = T_0 \cup T_1 \ldots \cup T_s$, $z = z_{T_0} + \ldots + z_{T_s}$, and

$$\|z_{T_{\ell+1}}\|_{\ell_2^N} \le \frac{1}{\sqrt{k}} \|z_{T_\ell}\|_{\ell_1^N}, \quad \ell = 0, 1, \ldots, s-1,$$

since $k|z_{i_j}| \le \|z_{T_\ell}\|_{\ell_1^N}$ for all $i_j \in T_{\ell+1}$. Then we have

$$\begin{aligned} \|z_{T_1 \cup \ldots \cup T_s}\|_{\ell_2^N} &= \|z_{T_1} + \ldots + z_{T_s}\|_{\ell_2^N} \le \sum_{\ell=0}^{s-1} \|z_{T_{\ell+1}}\|_{\ell_2^N} \\ &\le \frac{1}{\sqrt{k}} \sum_{\ell=0}^{s-1} \|z_{T_\ell}\|_{\ell_1^N} \le \frac{1}{\sqrt{k}} \|z\|_{\ell_1^N} \le \frac{C_1'}{\sqrt{k}}. \end{aligned} \tag{4.3}$$

We have left to bound $\|z_{T_0}\|_{\ell_2^N}$. Let us first note that the reasoning in (4.3), together with condition (2.6), which holds for $\omega \in \Omega_2 \subset \tilde{\Omega}_1(k)$, gives

$$\begin{aligned} \|\Phi(z_{T_1 \cup \ldots \cup T_s})\|_{\ell_2^n} &= \|\sum_{\ell=1}^{s} \Phi(z_{T_\ell})\|_{\ell_2^n} \le \sum_{\ell=1}^{s} \|\Phi(z_{T_\ell})\|_{\ell_2^n} \\ &\le \frac{5}{4} \sum_{\ell=1}^{s} \|z_{T_\ell}\|_{\ell_2^N} \le \frac{5C_1'}{4\sqrt{k}}. \end{aligned} \tag{4.4}$$

Therefore by again employing (2.6) and (4.4), we obtain

$$\begin{aligned} \|z_{T_0}\|_{\ell_2^N} &\le \frac{4}{3} \|\Phi(z_{T_0})\|_{\ell_2^n} = \frac{4}{3} \|\Phi z - \Phi(z_{T_1 \cup \ldots \cup T_s})\|_{\ell_2^n} \\ &\le \frac{4}{3} \left[\|\Phi z\|_{\ell_2^n} + \|\Phi(z_{T_1 \cup \ldots \cup T_s})\|_{\ell_2^n} \right] \le \frac{4}{3} \|y\|_{\ell_2^n} + \frac{5C_1'}{3\sqrt{k}} \\ &\le \frac{4}{3} \sqrt{\frac{L}{n}} + \frac{5C_1'}{3\sqrt{k}} \le \frac{4\sqrt{\tilde{a}}}{3\sqrt{k}} + \frac{5C_1'}{3\sqrt{k}} = \frac{4\sqrt{\tilde{a}} + 5C_1'}{3\sqrt{k}}, \end{aligned} \tag{4.5}$$

where the next to last inequality uses the definition of U_J and the last inequality uses that $k \le \tilde{a}n/L$. This is the bound we want for $\|z_{T_0}\|_{\ell_2^N}$.

Combining (4.3) and (4.5), we obtain

$$\|z\|_{\ell_2^N} \leq \|z_{T_0}\|_{\ell_2^N} + \|z_{T_1\cup\ldots\cup T_s}\|_{\ell_2^N} \leq \frac{4\sqrt{\tilde{a}}+5C_1'}{3\sqrt{k}} + \frac{C_1'}{\sqrt{k}} = \frac{4\sqrt{\tilde{a}}+8C_1'}{3\sqrt{k}},$$

which proves the theorem with $C_2 = \frac{4\sqrt{\tilde{a}}+8C_1'}{3}$. □

Now we are ready to show that Bernoulli random matrices coupled with decoding by ℓ_1-minimization give instance-optimality in ℓ_2^N with high probability for the large range of k. The ℓ_1-minimization decoder Δ is defined by

$$\Delta(y) := \operatorname*{argmin}_{\Phi w = y} \|w\|_{\ell_1^N}. \tag{4.6}$$

Given that $y = \Phi x$, we shall also use the notation

$$\bar{x} := \Delta(\Phi x). \tag{4.7}$$

In particular, if $\Phi = \Phi(\omega)$ is the random Bernoulli matrix then $\bar{x} = \bar{x}(\omega)$ will depend on the draw $\omega \in \Omega$. We shall also use the abbreviated notation

$$\sigma_k(x) := \sigma_k(x)_{\ell_2^N}, \quad x \in I\!R^N. \tag{4.8}$$

The main result of this section is the following theorem.

Theorem 4.2 *For an absolute constant $C_3 > 0$, we have the following. For each $x \in I\!R^N$ and each $k \leq \tilde{a}n/\log(N/n)$, $N \geq [\ln 6]^2 n$, there is a set $\Omega(x,k)$ with*

$$\rho(\Omega(x,k)^c) \leq \tilde{C}_1 e^{-\tilde{c}_1 n} + e^{-\sqrt{Nn}} + 2e^{-n/24} + 2ne^{\frac{-n}{2\log(N/n)}}, \tag{4.9}$$

such that for each $\omega \in \Omega(x,k)$, we have

$$\|x - \bar{x}\|_{\ell_2^N} \leq C_3 \sigma_k(x), \tag{4.10}$$

where $\tilde{a}$, $\tilde{c}_1$, $\tilde{C}_1$ are the constants from (2.5)*.*

Proof: We will prove the theorem for the largest k satisfying $k \leq \tilde{a}n/L$. The theorem follows for all other k from the monotonicity of σ_k. Let x_k be a best approximation to x from Σ_k, so $\|x - x_k\|_{\ell_2^N} = \sigma_k(x)$, and let $y' = \Phi(x - x_k)$. We shall take $\Omega(x,k) = \Omega_2 \cap \Omega_1(x - x_k)$ where Ω_2 is the set from Theorem 4.1 and $\Omega_1(x - x_k)$ is the set from Lemma 2.1. Then the estimate (4.9) follows from (4.2) and (2.7).

We now prove (4.10). According to Lemma 2.1, we have for all $\omega \in \Omega(x,k) \subset \Omega_1(x - x_k)$ that

$$\|y'\|_{\ell_2^n} \leq \sqrt{\frac{3}{2}}\|x - x_k\|_{\ell_2^N} = \sqrt{\frac{3}{2}}\sigma_k(x),$$

and

$$\|y'\|_{\ell_\infty^n} \leq \frac{1}{\sqrt{L}}\|x - x_k\|_{\ell_2^N} = \frac{1}{\sqrt{L}}\sigma_k(x).$$

Hence the vector $\frac{\sqrt{2}}{\sqrt{3}\sigma_k(x)}\sqrt{\frac{L}{n}}y' \in U_J$. For $\omega \in \Omega(x,k) \subset \Omega_2$, Theorem 4.1 says that there is a vector $z' \in \mathbb{R}^N$, such that $\Phi(x - x_k) = y' = \Phi z'$ and

$$\|z'\|_{\ell_2^N} \le \sqrt{\frac{3}{2}} C_2 \sigma_k(x) \sqrt{\frac{n}{L}} \frac{1}{\sqrt{k}}, \quad \text{and } \|z'\|_{\ell_1^N} \le \sqrt{\frac{3}{2}} C_1' \sqrt{\frac{n}{L}} \sigma_k(x). \tag{4.11}$$

Note that

$$\sigma_k(x_k + z')_{\ell_1^N} := \inf_{\tilde{x} \in \Sigma_k} \|x_k + z' - \tilde{x}\|_{\ell_1^N} = \inf_{\tilde{x} \in \Sigma_k} \|z' - (\tilde{x} - x_k)\|_{\ell_1^N} \le \|z'\|_{\ell_1^N},$$

and therefore using (4.11) it follows that

$$\sigma_k(x_k + z')_{\ell_1^N} \le \sqrt{\frac{3}{2}} C_1' \sqrt{\frac{n}{L}} \sigma_k(x). \tag{4.12}$$

Since $\Phi x = \Phi(x_k + z')$, we have that $\bar{x} = \Delta(\Phi(x_k + z'))$. For any $\omega \in \Omega(x,k) \subset \Omega_2 \subset \tilde{\Omega}_1(k)$, the Bernoulli matrix $\Phi(\omega)$ satisfies the RIP of order $2k$ and constant $\delta = 1/4$. Under these conditions, Candés showed [3] (improving the result from [5]) that there is an absolute constant $\tilde{C}$ such that

$$\|x_k + z' - \bar{x}\|_{\ell_2^N} \le \frac{\tilde{C}}{\sqrt{k}} \sigma_k(x_k + z')_{\ell_1^N}.$$

This inequality and (4.12) give

$$\|x_k + z' - \bar{x}\|_{\ell_2^N} \le \sqrt{\frac{3}{2}} C_1' \tilde{C} \sigma_k(x) \sqrt{\frac{n}{L}} \frac{1}{\sqrt{k}} \le C' \sigma_k(x), \tag{4.13}$$

where the last inequality uses the definition of k to conclude that $k \ge a'n/L$ for an absolute constant $a' > 0$. Therefore, it follows from (4.11) and (4.13) that

$$\begin{aligned}
\|x - \bar{x}\|_{\ell_2^N} &\le \|x - x_k - z'\|_{\ell_2^N} + \|x_k + z' - \bar{x}\|_{\ell_2^N} \\
&\le \|x - x_k\|_{\ell_2^N} + \|z'\|_{\ell_2^N} + \|x_k + z' - \bar{x}\|_{\ell_2^N} \\
&\le \left[1 + \sqrt{\frac{3}{2a'}} C_2 + C'\right] \sigma_k(x) = C_3 \sigma_k(x),
\end{aligned} \tag{4.14}$$

which proves the theorem. □

5 Generalizations to other random matrices

In this section, we shall extend the above results to more general random families of matrices. We assume in this section that the random matrix $\Phi = \Phi(\omega)$ has entries given by independent realizations of a fixed symmetric random variable η with expectation $\mathbb{E}\eta = 0$ and variance $\mathbb{E}\eta^2 = 1/n$. The columns Φ_j, $j = 1, \dots, N$, of Φ will be vectors in $\mathbb{R}^n$ with $\mathbb{E}\|\Phi_j\|_{\ell_2^n} = 1$. We shall show that under quite mild conditions on η, such a matrix, when coupled with the ℓ_1-minimization decoder (4.6), will give instance-optimality in probability in ℓ_2^N for the large range of k, $1 \le k \le an/\log(N/n)$. We shall use the notation $r = \sqrt{n}\eta$ to denote the random variable scaled to have variance one.

Our road map to proving instance-optimality is to follow the proof given in the previous sections for the Bernoulli case. That proof depends on four basic properties:

(a) The Concentration of Measure Property, i.e.(2.2) and (2.3) should hold with some choice of $c_0(\delta)$ and C_0 for the $n \times N$ random family $\Phi(\omega)$, whose entries $\phi_{i,j}$ are independent realizations of η for all n and N.

(b) The random family $\Phi(\omega)$ should with high probability satisfy the Restricted Isometry Property (2.5),(2.6) of order $2k$ and constant $\delta = 1/4$ for $1 \leq k \leq \tilde{a}n/\log(N/n)$ for some constants $\tilde{C}_1$, $\tilde{c}_1$ and $\tilde{a}$.

(c) The ℓ_∞^n bound for Φx given in (2.9) of Lemma 2.1 should be valid. Namely, for each $x \in I\!R^N$ there is a set $\Omega_1(x)$ with $\rho(\Omega_1(x)^c) \leq 2ne^{-\frac{\sqrt{n}}{4M\log(N/n)}}$, where $M \geq 1$ is an absolute constant, such that for all $\omega \in \Omega_1(x)$ we have $\|\Phi x\|_{\ell_\infty^n} \leq \frac{1}{\sqrt{\log(N/n)}}\|x\|_{\ell_2^N}$.

(d) The Lower Bound Property given in (3.15) (and therefore (3.16)) should be valid for the independent random variables $r_i = \sqrt{n}\eta_i$ for some constant $c_1 \geq 1/2$. The requirement $c_1 \geq 1/2$ is needed to assure that a proof similar to the one of Lemma 3.2 holds.

If all these properties are satisfied, then the proof given in the Bernoulli case carries over in an almost word for word fashion to the more general random matrices. Therefore, our discussion in this section will center on sufficient conditions on η so that **(a-d)** hold. The main point of this section is that to establish the validity of **(a-d)** it is enough to have **(a)**. Our main result is the following theorem.

Theorem 5.1 *If the symmetric random variable η satisfies property* **(a)**, *then* **(a-d)** *are valid. Using the random matrices $\Phi(\omega)$, whose entries are independent realizations of η, to encode and ℓ_1-minimization* (4.6) *to decode gives an encoder-decoder pair which is instance-optimal in probability in ℓ_2^N for the large range of k. That is, given any $x \in I\!R^N$ and any $k \leq \tilde{a}n/\log(N/n)$, for $N \geq [\ln 6]^2 n$, there is a set $\Omega_2(x,k)$ such that*

$$\rho(\Omega_2(x,k)^c) \leq \tilde{C}_1 e^{-\tilde{c}_1 n} + e^{-\sqrt{Nn}} + C_0 e^{-c_0(1/2)n} + 2ne^{-\frac{\sqrt{n}}{4M\log(N/n)}}, \tag{5.1}$$

where $\tilde{a}$, $\tilde{C}_1$, $\tilde{c}_1$, C_0, $c_0(1/2)$ are the constants from **(a-d)**, *$M \geq 1$ is an absolute constant, and for each $\omega \in \Omega_2(x,k)$ we have*

$$\|x - \Delta(\Phi(\omega)x)\|_{\ell_2^N} \leq C_4\sigma_k(x), \tag{5.2}$$

with C_4 depending only on the constants in **(a-d)**.

Remark 5.2 *According to Theorem* 5.1, *the Concentration of Measure Property* **(a)** *is the only property that a symmetric random variable η needs to satisfy so that the corresponding compressed sensing matrix $\Phi(\omega)$ coupled with the ℓ_1-minimization decoder gives a pair which is instance-optimal in probability in ℓ_2^N for the large range of k. Thus, our result covers practically all random variables used to assemble encoders appearing in the literature.*

As we have already mentioned, the proof of this theorem is the same as the proof in the Bernoulli case once we show **(a-d)**. Therefore, in the remainder of this section we discuss why **(b-d)** follow from **(a)**.

Proof (a) implies (b): It is well-known and proven in [2] that **(a)** implies the RIP of order $2k$ and constant δ for every $0 < \delta < 1$ and a range of k of the form $k \leq \tilde{a}(\delta)n/\log(N/n)$ for some constants $\tilde{C}_1$ and $\tilde{c}_1$, depending on δ. □

Before we proceed further, let us prove the following technical result.

Lemma 5.3 *Let η be a random variable that satisfies* **(a)** *and $r = \sqrt{n}\eta$. Then*

$$\Pr(|r| > t) \leq C_5 e^{-c_5 t^2}, \quad t > 0, \tag{5.3}$$

where c_5 and C_5 depend only on the constants $c_0(1/2)$ and C_0 in **(a)**. *Furthermore, if a random variable satisfies* (5.3), *then there is a constant $M \geq 1$ such that*

$$I\!\!E|r|^k \leq \frac{k! M^{k-2}}{2}, \quad \textit{for all integers } k \geq 2. \tag{5.4}$$

Proof: To show (5.3) we apply the CMP to the vector $x = (1, 0, \dots, 0)$ and obtain for all $n \in I\!\!N$

$$\Pr(r^2 > 3n/2) \leq \Pr\left(\frac{1}{n}\sum_{j=1}^{n} r_{1,j}^2 > 3/2\right) = \Pr\left(\sum_{j=1}^{n} \eta_{1,j}^2 > 3/2\right) \leq C_0 e^{-\gamma n}, \tag{5.5}$$

with $\gamma := c_0(1/2)$. From this and monotonicity, we have $\Pr(|r| > t) \leq C_0 e^{\gamma} e^{-\frac{2}{3}\gamma t^2}$, $t \geq \sqrt{\frac{3}{2}}$. Since $\Pr(|r| > t) \leq 1$, we find $\Pr(|r| > t) \leq C_5 e^{-c_5 t^2}$, for all $t > 0$, with

$$C_5 := \max\{1, C_0\} e^{\gamma}, \quad c_5 := \frac{2}{3}\gamma.$$

To prove (5.4), we use (5.3) and obtain

$$I\!\!E|r|^k = k\int_0^\infty t^{k-1} \Pr(|r| > t)\, dt \leq k\int_0^\infty t^{k-1} C_5 e^{-c_5 t^2}\, dt = \frac{C_5}{2c_5^{k/2}} k\Gamma\left(\frac{k}{2}\right) \leq \frac{1}{2} k! M^{k-2},$$

for some $M = M(C_5, c_5)$. If $M < 1$, we can just set $M = 1$ in the above inequality. □

We shall need the following Bernstein-type inequality (see [17] Ch.II Th 17).

Theorem 5.4 *Let $\{X_j\}_{j=1}^m$ be a collection of independent random variables with finite second moments and define $\sigma_0^2 := \sum_{j=1}^m I\!\!E X_j^2$. If there is a constant M_0 such that*

$$\sum_{j=1}^{m} I\!\!E|X_j|^k \leq \frac{1}{2} k! \sigma_0^2 M_0^{k-2}, \quad \textit{for all integers } k \geq 3, \tag{5.6}$$

then, for any $\delta \geq 0$,

$$\Pr\left(|\sum_{j=1}^{m}(X_j - I\!\!E X_j)| > \delta\right) \leq 2e^{-\frac{\delta^2}{2(\sigma_0^2 + M_0\delta)}}. \tag{5.7}$$

Using Lemma 5.3 and Theorem 5.4, we prove the following result.

Lemma 5.5 *Let η be a random variable that satisfies* **(a)**. *Then, for each $x \in I\!R^N$ and $N \geq 3n$, there is a set $\Omega_1(x)$ with*

$$\rho(\Omega_1(x)^c) \leq C_0 e^{-c_0(1/2)n} + 2ne^{-\frac{\sqrt{n}}{4M\log(N/n)}}, \tag{5.8}$$

such that for all $\omega \in \Omega_1(x)$,

$$\|\Phi x\|_{\ell_2^n} \leq \sqrt{\frac{3}{2}}\|x\|_{\ell_2^N}, \tag{5.9}$$

and

$$\|\Phi x\|_{\ell_\infty^n} \leq \frac{1}{\sqrt{\log(N/n)}}\|x\|_{\ell_2^N}. \tag{5.10}$$

Here, C_0 and $c_0(1/2)$ are the constants in the CMP (see (2.2)*) and M is the constant from* (5.4).

Proof: As in the proof of Lemma 2.1, we can assume that $\|x\|_{\ell_2^N} = 1$ and we fix such $x \in I\!R^N$. Property **(a)** guarantees (see (2.2) and (2.3)) that there is a set $\Omega_0(x, 1/2)$ with

$$\rho(\Omega_0(x, 1/2)^c) \leq C_0 e^{-c_0(1/2)n} \tag{5.11}$$

such that for each $\omega \in \Omega_0(x, 1/2)$ we have (5.9).

To prove (5.10), we note that each entry y_i, $i = 1, \ldots, n$, in $y = \Phi x$ takes the form

$$y_i = \sum_{j=1}^{N} x_j \eta_{i,j}. \tag{5.12}$$

We let $X_j := x_j \eta_{i,j}$. Since $I\!E X_j^2 = \frac{x_j^2}{n}$,

$$\sigma_0^2 := \sum_{j=1}^{N} I\!E X_j^2 = \|x\|_{\ell_2^N}^2 I\!E\eta^2 = 1/n. \tag{5.13}$$

Similarly, for any $k \geq 3$, we have

$$\begin{aligned} \sum_{j=1}^{N} I\!E|X_j|^k &= \sum_{j=1}^{N} |x_j|^k n^{-k/2} I\!E|r|^k \leq \frac{1}{2}k!\left[\frac{M}{\sqrt{n}}\right]^{k-2}\frac{1}{n}\sum_{j=1}^{N}|x_j|^k \\ &\leq \frac{1}{2}k!\left[\frac{M}{\sqrt{n}}\right]^{k-2}\frac{1}{n} = \frac{1}{2}k!\left[\frac{M}{\sqrt{n}}\right]^{k-2}\sigma_0^2, \end{aligned} \tag{5.14}$$

where we have used (5.4) from Lemma 5.3 and $[\sum_{j=1}^N |x_j|^k]^{1/k} \leq \|x\|_{\ell_2^N} = 1$ for $k \geq 2$. This means that the conditions of Theorem 5.4 are satisfied and (5.7) gives

$$\Pr(|y_i| > \delta) = \Pr\left(|\sum_{j=1}^{N} x_j \eta_{i,j}| > \delta\right) \leq 2e^{-\frac{\delta^2}{2(1/n + M\delta/\sqrt{n})}} \leq 2e^{-\frac{\sqrt{n}\delta^2}{4M}}, \quad \text{for } \delta \leq 1. \tag{5.15}$$

The last inequality follows from the fact that $(2-\delta)M\sqrt{n} \geq 1$ for $M \geq 1$ and $\delta \leq 1$.

We now take $\delta = 1/\sqrt{L} < 1$ for $N \geq 3n$ and define $\Omega_1(x)$ as the intersection of $\Omega_0(x, 1/2)$ with the sets $\{\omega : |y_i| \leq 1/\sqrt{L}\}$, $i = 1, \ldots, n$. Then, (5.9) and (5.10) are both valid for $\omega \in \Omega_1(x)$. The estimate (5.8) follows from a union bound. □

Proof (a) implies (c): This is (5.10) of Lemma 5.5. □

For the proof that **(a)** implies **(d)** we shall use the following lemma.

Lemma 5.6 *Let X be a symmetric random variable with finite fourth moment $\mathbb{E}X^4 = M_1$, $\mathbb{E}X^2 = 1$ and $\mathbb{E}X = 0$. Let X_j, $j = 1, \ldots, m$, be a sequence of independent random variables with distribution such as X. Then for any sequence of numbers $\lambda = (\lambda_1, \ldots, \lambda_m) \in \ell_2^m$ and $t > 0$, we have*

$$\Pr\left(|\sum_{j=1}^{m} \lambda_j X_j| > \frac{1}{2}\mathcal{K}(\lambda, t)\right) \geq 2e^{-c_1 t^2}, \quad c_1 = 4\ln(\max\{24, 8M_1\}). \tag{5.16}$$

Proof: This lemma can be proved using the arguments of [16] given for the Bernoulli case. However, one needs to use the full strength of Theorem 3 in Chapter 3 from [12] and not its special case as in Lemma 3 from [16]. A more general version of this lemma is proved in detail in Lemma 4.3 from [14]. □

Remark 5.7 *Note that in the Bernoulli case, $M_1 = 1$ and therefore the constant in* (3.15) *is $c_1 = 4\ln 24$.*

Proof of (a) implies (d): If η satisfies the CMP **(a)**, then by Lemma 5.3 $\mathbb{E}|r|^4$ is finite and **(d)** follows from Lemma 5.6. □

6 Remarks

Finally, we wish to make some remarks concerning the relationship between the Concentration of Measure Property **(a)** and the subgaussian distribution estimate (5.3) of Lemma 5.3. First, note that the proof of Lemma 5.3 shows that if we assume that (2.2) and the upper estimate in (2.3) hold for all n and some $N = N(n)$, then we have shown that the random variable $r = \sqrt{n}\eta$ satisfies the distributional inequality (5.3). Now we will show the coverse.

Lemma 6.1 *Let r be a zero mean random variable that satisfies* (5.3)*. Then, the $n \times N$ random family $\Phi(\omega)$, whose entries $\phi_{i,j}$ are independent realizations of $\eta = \frac{1}{\sqrt{n}}r$ satisfies the CMP* **(a)** *for all n and N.*

Proof: Let us fix x such that $\|x\|_{\ell_2^N} = 1$ and consider the random variables $X_i := \left|\sum_{j=1}^{N} x_j r_{i,j}\right|^2$, $i = 1, \ldots, n$. Since a sum of subgaussian random variable is itself subgaussian, we have

$$\Pr(X_i > t) = \Pr\left(|\sum_{j=1}^{N} x_j r_{i,j}| > \sqrt{t}\right) \leq C_6 e^{-c_6 t}. \tag{6.1}$$

The constants C_6, c_6 can be taken independent of x since $\|x\|_{\ell_2^N} = 1$ and the $r_{i,j}$ are drawn from the same distribution (see [18]).

As in the proof of Lemma 5.3, we have for $k \geq 2$,

$$\mathbb{E}|X_i|^k = k\int_0^\infty t^{k-1}\Pr(X_i > t)\,dt \leq k\int_0^\infty t^{k-1}C_6 e^{-c_6 t}\,dt = \frac{C_6}{c_6^k}k\Gamma(k) = \frac{C_6}{c_6^k}k!. \tag{6.2}$$

Since $\mathbb{E}r^2 = 1$, by Holder's inequality it follows that $\mathbb{E}r^4 \geq 1$, and therefore

$$\begin{aligned}\mathbb{E}|X_i|^2 &= \mathbb{E}\left|\sum_{j=1}^N x_j r_{i,j}\right|^4 = \mathbb{E}r^4\sum_{j=1}^N x_j^4 + 6\sum_{1\leq j<i\leq N} x_j^2 x_i^2\\ &\geq \sum_{j=1}^N x_j^4 + 2\sum_{1\leq j<i\leq N} x_j^2 x_i^2 = \|x\|_{\ell_2^N}^4 = 1, \qquad (6.3)\end{aligned}$$

where we have used the independence of $r_{i,j}$. It follows from (6.2) for $k = 2$ that $\mathbb{E}|X_i|^2 \leq 2C_6/c_6^2$, and therefore for $\sigma_2^2 := \sum_{i=1}^n \mathbb{E}|X_i|^2$, see (6.3), we have

$$n \leq \sigma_2^2 \leq 2n\frac{C_6}{c_6^2}, \tag{6.4}$$

which combined with (6.2) results in

$$\sum_{i=1}^n \mathbb{E}|X_i|^k \leq n\frac{C_6}{c_6^k}k! \leq \frac{C_6}{c_6^k}k!\sigma_2^2 \leq \frac{1}{2}k!M_2^{k-2}\sigma_2^2, \quad \text{for } k\geq 3, \tag{6.5}$$

where $M_2 = M_2(C_6, c_6)$ is an absolute constant.

In order to prove the CMP **(a)**, we need to estimate $\Pr(|\|\Phi x\|_{\ell_2^n}^2 - \|x\|_{\ell_2^N}^2| > \delta)$ for $0 < \delta < 1$. Since

$$\|\Phi x\|_{\ell_2^n}^2 = \sum_{i=1}^n\left|\sum_{j=1}^N x_j\eta_{i,j}\right|^2 = \frac{1}{n}\sum_{i=1}^n\left|\sum_{j=1}^N x_j r_{i,j}\right|^2$$

and $\|x\|_{\ell_2^N} = 1$, we have to estimate $\Pr\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - 1\right| > \delta\right) = \Pr\left(\left|\sum_{i=1}^n X_i - n\right| > n\delta\right)$. Since $\mathbb{E}X_i = 1$, $i = 1,\dots,n$, we have that

$$\Pr\left(\left|\sum_{i=1}^n X_i - n\right| > n\delta\right) = \Pr\left(\left|\sum_{i=1}^n (X_i - \mathbb{E}X_i)\right| > n\delta\right).$$

This probability can be estimated from above using Theorem 5.4, since all conditions are satisfied, see (6.5). We obtain

$$\begin{aligned}\Pr(|\|\Phi x\|_{\ell_2^n}^2 - \|x\|_{\ell_2^N}^2| > \delta) &= \Pr\left(\sum_{i=1}^n (X_i - \mathbb{E}X_i) > n\delta\right)\\ &\leq 2e^{-\frac{\delta^2 n^2}{2(\sigma_2^2 + M_2\delta n)}}\\ &\leq 2e^{-\frac{\delta^2 n^2}{2(2nC_6/c_6^2 + M_2\delta n)}}\\ &= 2e^{-c_0(\delta)n},\end{aligned}$$

with $c_0(\delta) = \frac{\delta^2}{2(2C_6/c_6^2 + M_2\delta)}$, where we have used (6.4) in the last inequality. This is the Concentration of Measure Property **(a)**. $\square$

Using Lemma 6.1 and Lemma 5.3 we conclude that the following properties of a random variable r with $\mathbb{E}r = 0$ and $\mathbb{E}r^2 = 1$ are equivalent:

(i) r satisfies (5.3)

(ii) For each n there exists an $N := N(n)$ such that the random $n \times N$ matrix family $\Phi(\omega)$, whose entries $\phi_{i,j}$ are independent realizations of $\frac{1}{\sqrt{n}}r$, satisfies the Concentration of Measure Property with some constants C_0 and $c_0(\delta)$.

(iii) For each n and N the random $n \times N$ matrix family $\Phi(\omega)$, whose entries $\phi_{i,j}$ are independent realizations of $\frac{1}{\sqrt{n}}r$, satisfies the Concentration of Measure Property with some constants C_0 and $c_0(\delta)$.

References

[1] D. Achilioptas, *Database-friendly random projections*, Proc. ACM Symposium on the Principles of Database Systems, 2001, 274–281.

[2] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, *A simple proof of the restricted isometry property for random matrices*, Constructive Approximation, to appear.

[3] E. Candés, *The restricted isometry property and its implications for compressed sensing*, Compte Rendus de l'Academie des Sciences, Paris, Series I, **3**46(2008), 589–592.

[4] E. J. Candès, J. Romberg and T. Tao, *Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information*, IEEE Trans. Inf. Theory, **5**2(2006), 489–509.

[5] E. Candès, J. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure and Appl. Math., **5**9(2006), 1207–1223.

[6] E. Candès and T. Tao, *Decoding by linear programming*, IEEE Trans. Inf. Theory, **5**1(2005), 4203–4215.

[7] A. Cohen, W. Dahmen, and R. DeVore, *Compressed sensing and best k term approximation*, JAMS, to appear

[8] R. DeVore, *Deterministic constructions of compressed sensing matrices*, J. of Complexity, **2**3(2007), 918–925.

[9] D. Donoho, *Compressed Sensing*, EEE Trans. Information Theory, **5**2(2006), 1289–1306.

[10] A.Garnaev, E. Gluskin, *The widths of a Euclidean ball*, Doklady AN SSSR, **2**77 (1984), 1048–1052.

[11] L. Györfi, M. Kohler, A. Krzyźak, and H. Walk, *A Distribution Free Theory of Nonparametric Regression* Springer Series in Statistics, 2002.

[12] J. Kahane, *Some Random Series of Functions*, D.C Heath and Co, Kexington Ma., 1968.

[13] B. Kashin, *The widths of certain finite dimensional sets and classes of smooth functions*, Izvestia, **4**1(1977), 334–351.

[14] A. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, *Smallest singular value of random matrices and geometry of random polytopes*, Advances in Math. **1**95 (2005), 491–523.

[15] G. Lorentz, M. v. Golitschek, and Yu. Makovoz, *Constructive Approximation, vol 2*, Springer Grundlehren, vol. 304, 1996

[16] S. Montgomery-Smith, *The distribution of Rademacher sums*, Proc. AMS, **1**09(1990), 517–522.

[17] V. Petrov, *Sums of independent random variables* Nauka, Moscow 1972 (in Russian); English translation *Limit theorems of probability theory. Sequences of independent random variables*, Oxford University Press, New York, 1995.

[18] R. Taylor and T-C. Hu, *Subgaussian techniques in proving strong laws of large numbers*, Amer. Math. Monthly, **9**(1987), 295–299.

[19] P. Wojtaszczyk, *Stability and instance optimality for Gaussian measurements in compressed sensing*, preprint.

Ronald DeVore, Department of Mathematics, Texas A&M University, rdevore@math.tamu.edu

Guergana Petrova, Department of Mathematics, Texas A&M University, gpetrova@math.tamu.edu

Przemyslaw Wojtaszczyk, Institut of Aplied Mathematics, University of Warsaw, wojtaszczyk@mimuw.edu.pl