Electronic Transactions on Numerical Analysis.
Volume 41, pp. 167-178, 2014.
Copyright © 2014, Kent State University.
ISSN 1068-9613.

ETNA
Kent State University
http://etna.math.kent.edu

NONUNIFORM SPARSE RECOVERY WITH SUBGAUSSIAN MATRICES[*]

ULAŞ AYAZ[†‡] AND HOLGER RAUHUT[‡]

Abstract. Compressive sensing predicts that sufficiently sparse vectors can be recovered from highly incomplete information using efficient recovery methods such as ℓ_1-minimization. Random matrices have become a popular choice for the measurement matrix. Indeed, near-optimal uniform recovery results have been shown for such matrices. In this note we focus on nonuniform recovery using subgaussian random matrices and ℓ_1-minimization. We provide conditions on the number of samples in terms of the sparsity and the signal length which guarantee that a fixed sparse signal can be recovered with a random draw of the matrix using ℓ_1-minimization. Our proofs are short and provide explicit and convenient constants.

Key words. compressed sensing, sparse recovery, random matrices, ℓ_1-minimization

AMS subject classifications. 94A20, 60B20

1. Introduction.

Compressive sensing allows to reconstruct signals from by far fewer measurements than what is considered necessary at first sight. The seminal papers by E. Candès, J. Romberg, and T. Tao [7, 9] and by D. Donoho [12] have triggered substantial research activities in mathematics, engineering, and computer science with a lot of possible applications.

In mathematical terms, we aim at solving the linear system of equations $y = Ax$ for $x \in \mathbb{C}^N$ when $y \in \mathbb{C}^m$ and $A \in \mathbb{C}^{m\times N}$ are given and when $m \ll N$. Clearly, in general this task is impossible since even if A has full rank, there are infinitely many solutions to this equation. The situation changes dramatically if x is sparse, that is, $\|x\|_0 := \#\{\ell, x_\ell \neq 0\}$ is small. We note that $\|\cdot\|_0$ is called the ℓ_0-norm although it is not a norm.

As a first approach, one is led to solve the optimization problem

$$\min_{z\in\mathbb{C}^N} \|z\|_0 \quad \text{subject to } Az = y,$$

where $y = Ax$. Unfortunately, this problem is NP-hard in general. It has become common to replace the ℓ_0-minimization problem by the ℓ_1-minimization problem

$$\min_{z\in\mathbb{C}^N} \|z\|_1 \quad \text{subject to } Az = y, \tag{1.1}$$

where $y = Ax$. This problem can be solved by efficient convex optimization techniques [3]. As a key result of compressive sensing and under appropriate conditions on A and on the sparsity of x, ℓ_1-minimization indeed reconstructs the original x. Certain random matrices A are known to provide optimal recovery with high probability. There are basically two types of recovery results:

- **Uniform recovery.** Such results state that with high probability on the draw of the random matrix A, *every* sparse vector can be reconstructed under appropriate conditions.
- **Nonuniform recovery.** Such results state that a given sparse vector x can be reconstructed with high probability on the draw of the matrix A under appropriate conditions. The difference to uniform recovery is that nonuniform recovery does not

---

[*] Received December 15, 2011. Accepted April 30, 2014. Published online on July 16, 2014. Recommended by G. Teschke.

[†] Hausdorff Center for Mathematics, University of Bonn, Endenicher Allee 60, 53115 Bonn, Germany (`ulas.ayaz@hcm.uni-bonn.de`).

[‡] RWTH Aachen University, Templergraben 55, 52056 Aachen, Germany (`rauhut@mathc.rwth-aachen.de`).

> imply that there is a matrix that recovers all x simultaneously. Or, in other words, the small exceptional set of matrices for which recovery fails may depend on x.

Uniform recovery via ℓ_1-minimization is, for instance, satisfied if the by now classical restricted isometry property (RIP) holds for A with high probability [4, 8]. A common choice for $A \in \mathbb{R}^{m\times N}$ are Gaussian random matrices, that is, the entries of A are independent standard normally distributed random variables. If, for $\varepsilon \in (0,1)$,

$$m \geq C(s\ln(N/s) + \ln(2/\varepsilon)), \tag{1.2}$$

then with a probability of at least $1-\varepsilon$, we have uniform recovery of all s-sparse vectors $x \in \mathbb{R}^N$ using ℓ_1-minimization and A as measurement matrix; see, e.g., [9, 15, 24]. The constant $C > 0$ is universal and estimates via the restricted isometry property give an approximate value of $C \approx 200$, which is significantly worse than what can be observed in practice. (Note that a direct analysis in [26] for the Gaussian case, which avoids the restricted isometry property, gives $C \approx 12$. This is still somewhat larger than the constants we report below in the nonuniform setting.) For this reason, this note considers nonuniform sparse recovery using Gaussian and more general subgaussian random matrices in connection with ℓ_1-minimization. Our main results below guarantee nonuniform recovery with explicit and convenient constants. In contrast to other works such as [13, 15], we can treat the recovery of complex vectors as well. We also get useful constants in the subgaussian case and, in particular, for Bernoulli matrices. Moreover, our results also establish stability of the reconstruction when the vectors are only approximately sparse and measurements are perturbed.

Gaussian and subgaussian random matrices are very important for the theory of compressive sensing because they provide a model of measurement matrices which can be analyzed very accurately (as shown in this note). They are used in real-world sensing scenarios, for instance, in the one-pixel camera [17]. Moreover, even if certain applications require more structure of the measurement matrix (leading to structured random matrices [25]), the empirically observed recovery performance of many types of matrices is very close to the one of (sub-)Gaussian random matrices [14], which underlines the importance of understanding subgaussian random matrices in compressive sensing.

2. Main results.

2.1. The Gaussian case. We say that an $m \times N$ random matrix A is Gaussian if its entries are independent and standard normally distributed random variables, that is, having mean zero and variance 1. Our nonuniform sparse recovery result for Gaussian matrices and ℓ_1-minimization reads as follows.

THEOREM 2.1. *Let $x \in \mathbb{C}^N$ with $\|x\|_0 = s$. Let $A \in \mathbb{R}^{m\times N}$ be a randomly drawn Gaussian matrix, and let $\varepsilon \in (0,1)$. If*

$$m \geq s\left[\sqrt{2\ln(4N/\varepsilon)} + \sqrt{2\ln(2/\varepsilon)/s} + 1\right]^2, \tag{2.1}$$

then with probability at least $1-\varepsilon$, the vector x is the unique solution to the ℓ_1-minimization problem (1.1).

REMARK 2.2. For large N and s, condition (2.1) approximately becomes

$$m > 2s\ln(4N/\varepsilon). \tag{2.2}$$

Compared to (1.2), we realize that the logarithmic term slightly falls short of the optimal one $\ln(N/s)$. However, we emphasize that our proof is short, and the constant is explicit and of moderate size. Indeed, when in addition s/N becomes very small (this is in fact the interesting regime), then we nevertheless reproduce the conditions found by Donoho and Tanner [13, 15] and, in particular, the optimal constant 2. Note that Donoho and Tanner used methods from the theory of random polytopes, which are quite different to our proof technique.

2.2. The subgaussian case. We generalize our recovery result for matrices with entries that are independent subgaussian random variables. A random variable X is called *subgaussian* if there are constants $\beta, \theta > 0$ such that

$$\mathbb{P}(|X| \geq t) \leq \beta e^{-\theta t^2} \quad \text{for all } t > 0.$$

It can be shown [28] that X is subgaussian with $\mathbb{E}X = 0$ if and only if there exists a constant c (depending only on β and θ) such that

$$\mathbb{E}[\exp(\lambda X)] \leq e^{c\lambda^2} \text{ for all } \lambda \in \mathbb{R}. \tag{2.3}$$

Important special cases of subgaussian mean-zero random variables are standard Gaussians, and Rademacher (Bernoulli) variables, that is, random variables that take the values ± 1 with equal probability. For both of these random variables, the constant c in (2.3) satisfies $c = 1/2$; see also Section 2.3.

A random matrix with entries that are independent mean-zero subgaussian random variables with the same constant c in (2.3) is called a subgaussian random matrix. Note that the entries are not required to be identically distributed.

THEOREM 2.3. *Let $x \in \mathbb{C}^N$ with $\|x\|_0 = s$. Let $A \in \mathbb{R}^{m \times N}$ be a random draw of a subgaussian matrix with a constant c in (2.3), and let $\varepsilon \in (0, 1)$. If*

$$m \geq s \left[\sqrt{4c \ln(4N/\varepsilon)} + \sqrt{C(3 + \ln(4/\varepsilon)/s)} \right]^2, \tag{2.4}$$

then with probability at least $1 - \varepsilon$, the vector x is the unique solution to the ℓ_1-minimization problem (1.1). The constant C in (2.4) only depends on c.

More precisely, the constant $C = 1.646\tilde{c}^{-1}$, where $\tilde{c} = \tilde{c}(c)$ is the constant in (B.1).

2.3. The Bernoulli case. We specialize the previous result for subgaussian matrices to Bernoulli (Rademacher) matrices, that is, random matrices with independent entries taking the value ± 1 with equal probability. We are then able to provide explicit values for the constants appearing in the result of Theorem 2.3. If Y is a Bernoulli random variable, then by a Taylor series expansion

$$\mathbb{E}(\exp(\lambda Y)) = \frac{1}{2}\left(e^{\lambda} + e^{-\lambda}\right) \leq e^{\frac{1}{2}\lambda^2}.$$

This shows that the subgaussian constant c equals $\frac{1}{2}$ in the Bernoulli case. Furthermore, we have the following concentration inequality for a matrix $B \in \mathbb{R}^{m \times N}$ with entries that are independent realizations of $\pm 1/\sqrt{m}$,

$$\mathbb{P}\left(\left|\|Bx\|_2^2 - \|x\|_2^2\right| > t\|x\|_2^2\right) \leq 2e^{-\frac{m}{2}(t^2/2 - t^3/3)}, \tag{2.5}$$

for all $x \in \mathbb{R}^N$, $t \in (0, 1)$; see, e.g., [1, 2]. We can simply estimate $t^3 < t^2$ in (2.5) and get $\tilde{c} = \frac{1}{12}$ in (B.1) and consequently $C = 1.646\tilde{c}^{-1} = 19.76$.

COROLLARY 2.4. *Let* $x \in \mathbb{C}^N$ *with* $\|x\|_0 = s$. *Let* $A \in \mathbb{R}^{m\times N}$ *be a matrix with entries that are independent Bernoulli random variables, and let* $\varepsilon \in (0,1)$. *If*

$$m \geq 2s\left[\sqrt{\ln(4N/\varepsilon)} + \sqrt{29.64 + 9.88\ln(4/\varepsilon)/s}\right]^2, \tag{2.6}$$

then with probability at least $1-\varepsilon$, *the vector* x *is the unique solution to the* ℓ_1*-minimization problem* (1.1).

Roughly speaking, for large N and moderately large s, the second term in (2.6) can be ignored and we arrive at $m \geq 2s\ln(4N/\varepsilon)$.

2.4. Stable and robust recovery. In this section we state some extensions of our results for nonuniform recovery with Gaussian matrices that show stability of the reconstruction when passing from sparse signals to only approximately sparse ones and robustness under perturbations of the measurements. In this context we assume the noisy model

$$y = Ax + e \in \mathbb{C}^m \quad \text{with } \|e\|_2 \leq \eta\sqrt{m}. \tag{2.7}$$

It it natural to work then with the noise constrained ℓ_1-minimization problem

$$\min_{z\in\mathbb{C}^N} \|z\|_1 \quad \text{subject to } \|Az - y\|_2 \leq \eta\sqrt{m}. \tag{2.8}$$

For the formulation of the next result, we define the error of the best s-term approximation of x in the ℓ_1-norm by

$$\sigma_s(x)_1 := \inf_{\|z\|_0\leq s} \|x - z\|_1.$$

THEOREM 2.5. *Let* $x \in \mathbb{C}^N$ *be an arbitrary but fixed vector, and let* $S \subset \{1,2,\ldots,N\}$ *denote the index set corresponding to its* s *largest absolute entries. Let* $A \in \mathbb{R}^{m\times N}$ *be a draw of a Gaussian random matrix. Suppose we take noisy measurements as in* (2.7). *If, for* $\theta \in (0,1)$,

$$m \geq s\left[\frac{\sqrt{2\ln(12N/\varepsilon)}}{1-\theta} + \sqrt{2\ln(6/\varepsilon)/s} + \sqrt{2}\right]^2, \tag{2.9}$$

then with probability at least $1-\varepsilon$, *the solution* $\hat{x}$ *to the minimization problem* (2.8) *satisfies*

$$\|x - \hat{x}\|_2 \leq \frac{C_1}{\theta}\eta + \frac{C_2}{\theta}\frac{\sigma_s(x)_1}{\sqrt{s}}. \tag{2.10}$$

Here, the constants $C_1, C_2 > 0$ *are universal.*

Condition (2.9) on the number of required measurements is very similar to (2.1) in the exact sparse and noiseless case. When θ tends to 0, we almost obtain the same condition, but then the right hand side of the stability estimate (2.10) blows up. In other words, we need to take slightly more measurements than required for exact recovery in order to ensure stability and robustness of the reconstruction.

A sketch of the proof of this theorem based on the so-called weak restricted isometry property is given in Section 3.7. We note that a version of this result for subgaussian random matrices can be shown as well.

2.5. Relation to previous work. Recently, several papers appeared dealing with nonuniform recovery. Most of these papers only consider the Gaussian case while our results extend to subgaussian and, in particular, to Bernoulli matrices.

As already mentioned, Donoho and Tanner [15] obtain nonuniform recovery results (the terminology is "weak phase transitions") for Gaussian matrices via methods from the theory of random polytopes. The authors there operate essentially in an asymptotic regime (although some of their results apply also to finite values of N, m, s). They consider the case when

$$m/N \to \delta,\ s/m \to \rho,\ \ln(N)/m \to 0,\ N \to \infty,$$

where ρ, δ are some fixed values. The recovery conditions are then expressed in terms of ρ and δ in this asymptotic regime. In particular, the authors find a (weak) transition curve $\rho_W(\delta)$ such that $\rho < \rho_W(\delta)$ implies recovery with high probability and $\rho > \rho_W(\delta)$ means failure with high probability (as $N \to \infty$). Moreover, they show that $\rho_W(\delta) \sim 2\ln(\delta^{-1})$ as $\delta \to 0$. Translated back into the quantities N, m, s, this gives $m \geq 2s\ln(N)$ in an asymptotic regime, which is essentially the condition in (2.2).

Candès and Plan give a rather general framework for nonuniform recovery in [5], which applies to measurement matrices with independent rows having bounded entries. In fact, they prove a recovery condition for such random matrices of the form $m \geq Cs\ln(N)$ for some constant C. However, they do not obtain explicit and good constants. Dossal et al. [16] derive a recovery condition for Gaussian matrices of the form $m \geq cs\ln(N)$, where c approaches 2 in an asymptotic regime. These two papers also contain stability results for noisy measurements.

Chandrasekaran et al. [10] use convex geometry in order to obtain nonuniform recovery results. They develop a rather general framework that applies also to low-rank recovery and further setups. However, they can only treat Gaussian measurements. They approach the recovery problem via Gaussian widths of certain convex sets. In particular, they estimate the number of Gaussian measurements needed in order to recover an s-sparse vector by $m \geq 2s(\ln(N/s - 1) + 1)$, which is essentially the optimal result. Their method heavily relies on properties of Gaussian random vectors, and therefore, it does not seem possible to extend it to more general subgaussian random matrices such as Bernoulli matrices.

Shortly before finishing this work, we became aware of the article of Candès and Recht [6], who derived closely related results. For Gaussian measurement matrices, they show that, for any $\beta > 1$, an s-sparse vector can be recovered with probability at least $1 - 2N^{-f(\beta,s)}$ if

$$m \geq 2\beta s \ln N + s,$$

where

$$f(\beta, s) = \left[\sqrt{\frac{\beta}{2s} + \beta - 1} - \sqrt{\frac{\beta}{2s}}\right]^2.$$

Their method of proof uses the duality based recovery Theorem 3.1 due to Fuchs [19] like in our approach, but then proceeds differently. They derive a similar recovery condition for subgaussian matrices but only state it for the special case of Bernoulli matrices. Furthermore, they also work out recovery results in the context of block-sparsity and low-rank recovery. However, differently to our paper, they do not cover the stability of the reconstruction.

3. Proofs.

3.1. Notation. We start by defining some notation needed in the proofs. Let $[N]$ denote the set $\{1, 2, \dots, N\}$. The column submatrix of a matrix A consisting of the columns indexed by S is written as $A_S = (a_j)_{j \in S}$, where $S \subset [N]$ and $a_j \in \mathbb{R}^m$, $j = 1, \dots, m$, denote the columns of A. Similarly, $x_S \in \mathbb{C}^S$ denotes the vector $x \in \mathbb{C}^N$ restricted to the entries in S, and $x \in \mathbb{C}^N$ is called s-sparse if $\operatorname{supp}(x) = \{\ell : x_\ell \neq 0\} = S$ with $S \subset [N]$ and $|S| = s$, i.e., $\|x\|_0 = s$. We further need to introduce the sign vector $\operatorname{sgn}(x) \in \mathbb{C}^N$ having entries

$$\operatorname{sgn}(x)_j := \begin{cases} \frac{x}{|x_j|} & \text{if } x_j \neq 0, \\ 0 & \text{if } x_j = 0, \end{cases} \quad j \in [N].$$

The Moore-Penrose pseudo-inverse of a matrix B where (B^*B) is invertible is given by $B^\dagger = (B^*B)^{-1}B^*$ so that $B^\dagger B = \mathrm{Id}$, where Id is the identity matrix.

3.2. Recovery conditions. In this section we state some results that are used in the proof of the main theorems directly or indirectly. The proofs of Theorems 2.1 and 2.3 require a condition for sparse recovery which not only depends on the matrix A but also on the sparse vector $x \in \mathbb{C}^N$ to be recovered. The following theorem is due to J. J. Fuchs [19] in the real-valued case and was extended to the complex case by J. Tropp [27]; see also [25, Theorem 2.8] for a slightly simplified proof.

THEOREM 3.1. *Let $A \in \mathbb{C}^{m \times N}$ and $x \in \mathbb{C}^N$ with $S := \operatorname{supp}(x)$. Assume that A_S is injective and that there exists a vector $h \in \mathbb{C}^m$ such that*

$$\begin{aligned} &A_S^* h = \operatorname{sgn}(x_S), \\ &|(A^*h)_\ell| < 1, \ \ell \in [N] \setminus S. \end{aligned}$$

Then x is the unique solution to the ℓ_1-minimization problem (1.1) with $y = Ax$.

Choosing the vector $h = \left(A_S^\dagger\right)^* \operatorname{sgn}(x_S)$ leads to the following corollary.

COROLLARY 3.2. *Let $A \in \mathbb{C}^{m \times N}$ and $x \in \mathbb{C}^N$ with $S := \operatorname{supp}(x)$. If the matrix A_S is injective and if*

$$|\langle (A_S)^\dagger a_\ell, \operatorname{sgn}(x_S) \rangle| < 1 \ \text{ for all } \ \ell \in [N] \setminus S,$$

then the vector x is the unique solution to the ℓ_1-minimization problem (1.1) with $y = Ax$.

3.3. Proof of recovery in the Gaussian case. We set $S := \operatorname{supp}(x)$, which has cardinality s. By Corollary 3.2, for recovery via ℓ_1-minimization, it is sufficient to show that

$$|\langle (A_S)^\dagger a_\ell, \operatorname{sgn}(x_S) \rangle| = |\langle a_\ell, (A_S^\dagger)^* \operatorname{sgn}(x_S) \rangle| < 1 \ \text{ for all } \ \ell \in [N] \setminus S.$$

Therefore, the failure probability for recovery is bounded by

$$\mathcal{P} := \mathbb{P}\left(\exists \ell \notin S \ |\langle (A_S)^\dagger a_\ell, \operatorname{sgn}(x_S) \rangle| \geq 1\right).$$

If we condition $X := \langle a_\ell, (A_S^\dagger)^* \operatorname{sgn}(x_S) \rangle$ on A_S, it is a Gaussian random variable. Furthermore, $X = \sum_{j=1}^m (a_\ell)_j [(A_S^\dagger)^* \operatorname{sgn}(x_S)]_j$ is centered, so its variance ν^2 can be estimated by

$$\begin{aligned} \nu^2 = \mathbb{E}(X^2) &= \sum_{j=1}^m \mathbb{E}[(a_\ell)_j^2][(A_S^\dagger)^* \operatorname{sgn}(x_S)]_j^2 \\ &= \|(A_S^\dagger)^* \operatorname{sgn}(x_S)\|_2^2 \leq \sigma_{\min}^{-2}(A_S) \|\operatorname{sgn}(x_S)\|_2^2 = \sigma_{\min}^{-2}(A_S)\, s, \end{aligned}$$

where $\sigma_{\min}$ denotes the smallest singular value. The last inequality uses the fact that $\|(A_S^\dagger)^*\|_{2\to 2} = \|A_S^\dagger\|_{2\to 2} = \sigma_{\min}^{-1}(A_S)$. A tail estimate for a mean-zero Gaussian random variable X with variance σ^2 obeys the inequality

$$\mathbb{P}(|X| > t) \le e^{-t^2/2\sigma^2}; \tag{3.1}$$

see [25, Lemma 10.2]. Then it follows that

$$\begin{aligned} \mathcal{P} &\le \mathbb{P}\left(\exists \ell \notin S\ |\langle (A_S)^\dagger a_\ell, \operatorname{sgn}(x_S)\rangle| \ge 1 \,\middle|\, \|(A_S^\dagger)^* \operatorname{sgn}(x_S)\|_2 < \alpha\right) \\ &\quad + \mathbb{P}(\|(A_S^\dagger)^* \operatorname{sgn}(x_S)\|_2 \ge \alpha) \\ &\le 2N \exp(-1/2\alpha^2) + \mathbb{P}(\sigma_{\min}^{-1}(A_S)\sqrt{s} \ge \alpha). \end{aligned} \tag{3.2}$$

The inequality in (3.2) uses the tail estimate (3.1), the union bound, and the independence of a_ℓ and A_S. The first term in (3.2) is bounded by $\varepsilon/2$ if

$$\alpha \le \frac{1}{\sqrt{2\ln(4N/\varepsilon)}}. \tag{3.3}$$

In order to estimate the second term in (3.2), we use an elegant estimate for the smallest singular value of a normalized Gaussian matrix $B \in \mathbb{R}^{m\times s}$ where the entries of B are independent and follow the normal distribution $\mathcal{N}(0, 1/m)$, which was provided in [11],

$$\mathbb{P}(\sigma_{\min}(B) < 1 - \sqrt{s/m} - r) \le e^{-mr^2/2}. \tag{3.4}$$

Its proof relies on the Slepian-Gordon Lemma [20, 21] and the concentration of measure for Lipschitz functions [23]. We proceed with

$$\begin{aligned} \mathbb{P}(\sigma_{\min}^{-1}(A_S)\sqrt{s} \ge \alpha) &= \mathbb{P}(\sigma_{\min}(A_S) \le \sqrt{s}/\alpha) = \mathbb{P}\left(\sigma_{\min}(A_S/\sqrt{m}) \le \frac{1}{\sqrt{m}}\frac{\sqrt{s}}{\alpha}\right) \\ &\le \exp\left(\frac{-m(1-(\alpha^{-1}+1)\sqrt{s/m})^2}{2}\right). \end{aligned} \tag{3.5}$$

If we choose α that makes (3.3) an equality, plug it into condition (3.5), and require that (3.5) is bounded by $\varepsilon/2$, we arrive at the condition

$$m \ge s\left[\sqrt{2\ln(4N/\varepsilon)} + \sqrt{2\ln(2/\varepsilon)/s} + 1\right]^2,$$

which ensures recovery with a probability of at least $1-\varepsilon$. This concludes the proof of Theorem 2.1. $\square$

3.4. Tail estimate for sums of subgaussian variables. We use the following estimate for sums of subgaussian random variables in the proof of Theorem 2.3. It appears for instance in [28].

LEMMA 3.3. *Let $X_1, \ldots, X_M$ be a sequence of independent mean-zero subgaussian random variables with the same parameter c in (2.3). Let $a \in \mathbb{R}^M$ be some vector. Then $Z := \sum_{j=1}^M a_j X_j$ is subgaussian, that is, for $t > 0$,*

$$\mathbb{P}(|\sum_{j=1}^M a_j X_j| \ge t) \le 2\exp(-t^2/(4c\|a\|_2^2)).$$

The proof of this Lemma is given in Appendix A.

3.5. Conditioning of subgaussian matrices. While the following lemma is well-known in principle, the correct scaling in δ seemingly has not appeared elsewhere in the literature; compare with [2, 24].

LEMMA 3.4. *Let* $S \subset [N]$ *with* $\mathrm{card}(S) = s$*. Let* A *be an* $m \times N$ *random matrix with independent, isotropic, and subgaussian rows with the same parameter* c *in* (2.3)*. Then, for* $\delta \in (0,1)$*, the normalized matrix* $\tilde{A} = \frac{1}{\sqrt{m}} A$ *satisfies*

$$\|\tilde{A}_S^* \tilde{A}_S - \mathrm{Id}\|_{2\to 2} \le \delta$$

with probability at least $1-\varepsilon$ *provided that*

$$m \ge C\delta^{-2}(3s + \ln(2\varepsilon^{-1})), \tag{3.6}$$

where C *depends only on* c*.*

The proof of this Lemma is given in Appendix B.

3.6. Proof of recovery in the subgaussian case. We follow a similar path as in the Gaussian case. We denote $S := \mathrm{supp}(x)$. We can bound the failure probability $\mathcal{P}$ by

$$\begin{aligned} \mathcal{P} \le \mathbb{P}\left(\exists \ell \notin S \; |\langle (A_S)^\dagger a_\ell, \mathrm{sgn}(x_S)\rangle| \ge 1 \,\middle|\, \|(A_S^\dagger)^* \mathrm{sgn}(x_S)\|_2 < \alpha\right) \\ + \mathbb{P}(\|(A_S^\dagger)^* \mathrm{sgn}(x_S)\|_2 \ge \alpha). \end{aligned} \tag{3.7}$$

The first term in (3.7) can be bounded by Lemma 3.3. Conditioning on A_S and $\|(A_S^\dagger)^* \mathrm{sgn}(x_S)\|_2 < \alpha$, we get

$$\mathbb{P}(|\langle (A_S)^\dagger a_\ell, \mathrm{sgn}(x_S)\rangle| \ge 1) = \mathbb{P}(|\sum_{j=1}^{m} (a_\ell)_j [(A_S^\dagger)^* \mathrm{sgn}(x_S)]_j| \ge 1) \le 2\exp(-1/(4c\alpha^2)).$$

So by the union bound, the first term in (3.7) can be estimated by $2N\exp(-1/(4c\alpha^2))$, which in turn is no larger than $\varepsilon/2$ provided that

$$\alpha \le \sqrt{1/(4c\ln(4N/\varepsilon))}. \tag{3.8}$$

For the second term in (3.7), we have

$$\begin{aligned} &\mathbb{P}(\|(A_S^\dagger)^* \mathrm{sgn}(x_S)\|_2 \ge \alpha) \le \mathbb{P}(\sigma_{\min}^{-1}(A_S)\sqrt{s} \ge \alpha) \\ &= \mathbb{P}(\sigma_{\min}(A_S) \le \sqrt{s}/\alpha) = \mathbb{P}\left(\sigma_{\min}(A_S/\sqrt{m}) \le \frac{1}{\sqrt{m}}\frac{\sqrt{s}}{\alpha}\right). \end{aligned}$$

By Lemma 3.4, the normalized subgaussian matrix $\tilde{A}_S := A_S/\sqrt{m}$ satisfies

$$\mathbb{P}(\sigma_{\min}(\tilde{A}_S) < 1-\delta) < \mathbb{P}(\sigma_{\min}(\tilde{A}_S) < \sqrt{1-\delta}) < \mathbb{P}(\|\tilde{A}_S^*\tilde{A}_S - \mathrm{Id}\|_{2\to 2} \ge \delta) < \varepsilon/2$$

provided that $m \ge C\delta^{-2}(3s + \ln(4\varepsilon^{-1}))$ and $\delta \in (0,1)$, where C depends on the subgaussian constant c. The choice $\frac{1}{\sqrt{m}}\frac{\sqrt{s}}{\alpha} = 1-\delta$ yields $\delta = 1 - \frac{\sqrt{s}}{\alpha\sqrt{m}}$. Combining these arguments and choosing α that makes (3.8) an equality, we can bound the failure probability by ε if

$$m \ge C\left(1 - \frac{\sqrt{4cs\ln(4N/\varepsilon)}}{\sqrt{m}}\right)^{-2}(3s + \ln(4/\varepsilon)). \tag{3.9}$$

Solving (3.9) for m yields the condition

$$m \ge s\left[\sqrt{4c\ln(4N/\varepsilon)} + \sqrt{C(3 + \ln(4/\varepsilon)/s)}\right]^2.$$

This condition also implies $\delta \in (0,1)$. This concludes the proof of Theorem 2.3. $\square$

3.7. Stability of reconstruction. Here, we give a very brief sketch of the proof of Theorem 2.5. It uses the concept of the weak restricted isometry property (weak RIP) introduced in [5].

DEFINITION 3.5. *(Weak RIP) Let $S \subset [N]$ be fixed with cardinality s and fix $\delta_1, \delta_2 > 0$. Then a matrix $A \in \mathbb{R}^{m\times N}$ is said to satisfy the weak RIP with parameters $(S, r, \delta_1, \delta_2)$ if*

$$(1-\delta_1)\|v\|_2^2 \le \|Av\|_2^2 \le (1+\delta_2)\|v\|_2^2$$

for all v supported on $S \cup R$ and all subsets R in $[N] \setminus S$ with cardinality $|R| \le r$.

The key to the proof of Theorem 2.5 is the following stable and robust version of the dual certificate based recovery Theorem 3.1. Its proof follows a similar strategy as in [5] and [22, Theorem 3.1].

LEMMA 3.6. *Let $x \in \mathbb{C}^N$ and $A \in \mathbb{R}^{m\times N}$. Let S be the set of indices of the s largest absolute entries of x. Assume that A satisfies the weak RIP with parameters $(S, r, \delta_1, \delta_2)$ for $r \le N$ and $\delta_1, \delta_2 \in (0,1)$ and that there exists a vector $v \in \mathbb{C}^m$ such that, for $\theta \in (0,1)$,*

$$A_S^* v = \operatorname{sgn}(x_S),$$

$$|(A^*v)_\ell| < 1-\theta, \qquad \ell \in [N] \setminus S, \tag{3.10}$$

$$\|v\|_2 \le \beta\sqrt{s}. \tag{3.11}$$

Suppose we take noisy measurements $y = Ax + e \in \mathbb{C}^m$ with $\|e\|_2 \le \eta$. Then the solution $\hat{x}$ to

$$\min_{z\in\mathbb{C}^N} \|z\|_1 \qquad \text{subject to } \|Az - y\| \le \eta$$

satisfies

$$\|x-\hat{x}\|_2 \le \frac{\sqrt{1+\delta_2}}{1-\delta_1}2\eta + \left(\frac{2\sqrt{2}\max\{\delta_1,\delta_2\}}{1-\delta_1} + \sqrt{2}\right)\left(\frac{2\beta}{\theta}\sqrt{\frac{s}{r}}\eta + \frac{2}{\theta}\frac{\sigma_s(x)_1}{\sqrt{r}}\right). \tag{3.12}$$

The weak RIP is established for Gaussian random matrices by using the estimate (3.4) for the smallest singular value of a single submatrix $A_{S\cup R}$ and a corresponding estimate for the largest singular value [11]. Then one takes the union bound over all subsets R of $[N]\setminus S$ of cardinality r. We conclude in this way that the weak RIP holds with probability at least $1-\varepsilon$ provided that

$$m \ge \max\left\{1-\sqrt{1-\delta_1}, \sqrt{1+\delta_2}-1\right\}^2 \left[\sqrt{s+r} + \sqrt{2r\ln(eN/r)} + 2\ln(2/\varepsilon)\right]^2.$$

The number r is chosen as $s/8$ in the end so that the quotient $\sqrt{s/r}$ appearing in (3.12) becomes a constant.

We use the same ansatz for the dual vector v as before, namely $v = (A_S^\dagger)^*\operatorname{sgn}(x_S)$. Condition (3.10) is analyzed in the same way as the corresponding condition in Theorem 3.1. This leads to the appearance of θ in (2.9). Moreover, condition (3.11) is straightforward to verify via $\|v\|_2 \le \|A_S^\dagger\|_{2\to 2}\|\operatorname{sgn}(x_S)\|_2 = \|A_S^\dagger\|_{2\to 2}\sqrt{s}$. An appropriate choice of the numbers δ_1, δ_2, and β leads to the desired result.

Acknowledgement. The authors would like to thank the Hausdorff Center for Mathematics for support and acknowledge funding through the WWTF project SPORTS (MA07-004). We would also like to thank Simon Foucart. Indeed, the proof of Lemma 3.4 is taken from the book [18].

Appendix A. Proof of Lemma 3.3. By independence, we have

$$\begin{aligned}\mathbb{E}\exp(\theta\sum_{j=1}^{M} a_j X_j) = \mathbb{E}\prod_{i=1}^{M}\exp(\theta a_j X_j) = \prod_{i=1}^{M}\mathbb{E}\exp(\theta a_j X_j) \le \prod_{i=1}^{M}\exp(\theta a_j X_j)\\ = \exp(c\|a\|_2^2\theta^2).\end{aligned}$$

This shows that Z is subgaussian with parameter $c\|a\|_2^2$ in (2.3). We apply Markov's inequality to get

$$\mathbb{P}(Z \ge t) = \mathbb{P}(\exp(\theta Z) \ge \exp(\theta t)) \le \mathbb{E}[\exp(\theta Z)]e^{-\theta t} \le e^{c\|a\|_2^2\theta^2 - \theta t}\,.$$

The optimal choice $\theta = t/(2c\|a\|_2^2)$ yields

$$\mathbb{P}(Z \ge t) \le e^{-t^2/(4c\|a\|_2^2)}\,.$$

Repeating the above computation with $-Z$ instead of Z shows that

$$\mathbb{P}(-Z \ge t) \le e^{-t^2/(4c\|a\|_2^2)}\,,$$

and the union bound yields the desired estimate $\mathbb{P}(|Z| \ge t) \le 2e^{-t^2/(4c\|a\|_2^2)}$. □

Appendix B. Proof of Lemma 3.4. Since most available statements have an additional $\ln(\delta^{-1})$-term in (3.6), we include the proof of this lemma for the sake of completeness.

The following concentration inequality for subgaussian random variables appears, for instance, in [1, 24].

$$\mathbb{P}(|\|\tilde{A}x\|_2^2 - \|x\|_2^2| > t\|x\|_2^2) \le 2\exp(-\tilde{c}mt^2), \tag{B.1}$$

where $\tilde{c}$ depends only on c. We will combine the above concentration inequality with the net technique. Let $\rho \in (0, \sqrt{2}-1)$ be a number to be determined later. According to a classical covering number argument, see, e.g., [25, Proposition 10.1], there exists a finite subset U of the unit sphere $\mathcal{S} = \{x \in \mathbb{R}^{\mathbb{N}}, \text{supp}(x) \subset S, \|x\|_2 = 1\}$, which satisfies

$$|U| \le \left(1 + \frac{2}{\rho}\right)^s \qquad \text{and} \qquad \min_{u\in U}\|z-u\|_2 \le \rho \quad \text{for all } z \in \mathcal{S}.$$

The concentration inequality (B.1) yields

$$\begin{aligned}&\mathbb{P}\left(\left|\|\tilde{A}u\|_2^2 - \|u\|_2^2\right| > t\,\|u\|_2^2 \quad \text{for some } u \in \mathcal{U}\right)\\ &\le \sum_{u\in U}\mathbb{P}\left(\left|\|\tilde{A}u\|_2^2 - \|u\|_2^2\right| > t\,\|u\|_2^2\right) \le 2|U|\exp\left(-\tilde{c}t^2\,m\right)\\ &\le 2\left(1+\frac{2}{\rho}\right)^s \exp\left(-\tilde{c}t^2\,m\right).\end{aligned}$$

The positive number t will be set later depending on δ and on ρ. Let us assume for now that the realization of the random matrix $\tilde{A}$ yields

$$\left|\|\tilde{A}u\|_2^2 - \|u\|_2^2\right| \le t \qquad \text{for all } u \in U. \tag{B.2}$$

By the above results, this occurs with probability exceeding

$$1 \,-\, 2\left(1+\frac{2}{\rho}\right)^{s} \exp\left(-\tilde{c}t^2\, m\right).$$

Next we show that (B.2) implies $\left| \|\tilde{A}x\|_2^2 - \|x\|_2^2 \right| \leq \delta$ for all $x \in \mathcal{S}$, that is, $\|\tilde{A}_S^*\tilde{A}_S - \mathrm{Id}\|_{2\to 2} \leq \delta$ (if t is determined appropriately). Let $B = \tilde{A}_S^*\tilde{A}_S - \mathrm{Id}$ so that we have to show $\|B\|_{2\to 2} \leq \delta$. Note that (B.2) means that $|\langle Bu, u\rangle| \leq t$ for all $u \in U$. Now consider a vector $x \in \mathcal{S}$ for which we choose a vector $u \in U$ satisfying $\|x-u\|_2 \leq \rho < \sqrt{2}-1$. We obtain

$$\begin{aligned} |\langle Bx, x\rangle| &= |\langle B(u+x-u), u+x-u\rangle| \\ &= |\langle Bu, u\rangle + \langle B(x-u), x-u\rangle + 2\,\langle Bu, x-u\rangle| \\ &\leq |\langle Bu, u\rangle| + |\langle B(x-u), x-u\rangle| + 2\,\|Bu\|_2\,\|x-u\|_2 \\ &\leq t + \|B\|_{2\to 2}\,\rho^2 + 2\,\|B\|_{2\to 2}\,\rho. \end{aligned}$$

Taking the supremum over all $x \in \mathcal{S}$, we deduce that

$$\|B\|_{2\to 2} \leq t + \|B\|_{2\to 2}\left(\rho^2 + 2\rho\right), \qquad \text{i.e.,} \qquad \|B\|_{2\to 2} \leq \frac{t}{2-(\rho+1)^2}.$$

Note that the division by $2-(\rho+1)^2$ is justified by the assumption that $\rho < \sqrt{2}-1$. Then we choose

$$t = t_{\delta,\rho} := \left(2-(\rho+1)^2\right)\delta$$

so that $\|B\|_{2\to 2} \leq \delta$, and with our definition of t,

$$\mathbb{P}\left(\|\tilde{A}_S^*\tilde{A}_S - \mathrm{Id}\|_{2\to 2} > \delta\right) \leq 2\left(1+\frac{2}{\rho}\right)^{s} \exp\left(-\tilde{c}\delta^2(2-(\rho+1)^2)^2 m\right).$$

Hence, $\|\tilde{A}_S^*\tilde{A}_S - \mathrm{Id}\|_{2\to 2} \leq \delta$ with probability at least $1-\varepsilon$ provided that

$$m \geq \frac{1}{\tilde{c}(2-(\rho+1)^2)^2}\delta^{-2}\left(\ln(1+2/\rho)s + \ln(2\varepsilon^{-1})\right). \tag{B.3}$$

Now we take ρ such that $\ln(1+2/\rho) = 3$, that is, $\rho = 2/(e^3-1)$. Then (B.3) yields the condition

$$m \geq C\delta^{-2}\left(3s + \ln(2\varepsilon^{-1})\right)$$

with $C = 1.646\,\tilde{c}^{-1}$. This concludes the proof. □

REFERENCES

[1] D. Achlioptas, *Database-friendly random projections: Johnson-Lindenstrauss with binary coins*, J. Comput. System Sci., 66 (2003), pp. 671–687.

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

[3] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.

[4] E. J. Candès, *The restricted isometry property and its implications for compressed sensing*, C. R. Math. Acad. Sci. Paris, 346 (2008), pp. 589–592.

[5] E. J. Candès and Y. Plan, *A probabilistic and RIPless theory of compressed sensing*, IEEE Trans. Inform. Theory, 57 (2011), pp. 7235–7254.
[6] E. J. Candès and B. Recht, *Simple bounds for recovering low-complexity models*, Math. Program., 141 (2013), pp. 577–589.
[7] E. J. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inform. Theory, 52 (2006), pp. 489–509.
[8] ———, *Stable signal recovery from incomplete and inaccurate measurements*, Comm. Pure Appl. Math., 59 (2006), pp. 1207–1223.
[9] E. J. Candès and T. Tao, *Near-optimal signal recovery from random projections: universal encoding strategies?*, IEEE Trans. Inform. Theory, 52 (2006), pp. 5406–5425.
[10] V. Chandrasekaran, B. Recht, P. A. Parrilo, and A. S. Willsky, *The convex geometry of linear inverse problems*, Found. Comput. Math., 12 (2012), pp. 805–849.
[11] K. R. Davidson and S. J. Szarek, *Local operator theory, random matrices and Banach spaces*, in Handbook of the Geometry of Banach Spaces, Vol. 1, W. B. Johson and J. Lindenstrauss, eds., North-Holland, Amsterdam, 2001, pp. 317–366.
[12] D. L. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.
[13] D. L. Donoho and J. Tanner, *Thresholds for the recovery of sparse solutions via ℓ_1 minimization*, in 2006 IEEE Conference on Information Sciences and Systems, IEEE Conference Proceedings, Los Alamitos, CA, 2006, pp. 202–206.
[14] ———, *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., 367 (2009), pp. 4273–4293.
[15] ———, *Counting faces of randomly projected polytopes when the projection radically lowers dimension*, J. Amer. Math. Soc., 22 (2009), pp. 1–53.
[16] C. Dossal, M.-L. Chabanol, G. Peyré, and J. Fadili, *Sharp support recovery from noisy random measurements by ℓ_1 -minimization*, Appl. Comput. Harmonic Anal., 33 (2012), pp. 24–43.
[17] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. G. Baraniuk, *Single-pixel imaging via compressive sampling*, IEEE Signal Process. Mag., 25 (2008), pp. 83–91.
[18] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*, Birkhäuser, New York, 2013.
[19] J.-J. Fuchs, *On sparse representations in arbitrary redundant bases*, IEEE Trans. Inform. Theory, 50 (2004), pp. 1341–1344.
[20] Y. Gordon, *Some inequalities for Gaussian processes and applications*, Israel J. Math., 50 (1985), pp. 265–289.
[21] ———, *Elliptically contoured distributions*, Probab. Theory Related Fields, 76 (1987), pp. 429–438.
[22] M. Hügel, H. Rauhut, and T. Strohmer, *Remote sensing via ℓ_1-minimization*, Found. Comput. Math., 14 (2014), pp. 115–150.
[23] M. Ledoux, *The Concentration of Measure Phenomenon*, AMS, Providence, 2001.
[24] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, *Uniform uncertainty principle for Bernoulli and subgaussian ensembles*, Constr. Approx., 28 (2008), pp. 277–289.
[25] H. Rauhut, *Compressive sensing and structured random matrices*, in Theoretical Foundations and Numerical Methods for Sparse Recovery, Vol. 9 of Radon Ser. Comput. Appl. Math., M. Fornasier, ed., de Gruyter, Berlin, 2010, pp. 1–92.
[26] M. Rudelson and R. Vershynin, *On sparse reconstruction from Fourier and Gaussian measurements*, Comm. Pure Appl. Math., 61 (2008), pp. 1025–1045.
[27] J. A. Tropp, *Recovery of short, complex linear combinations via l_1 minimization*, IEEE Trans. Inform. Theory, 51 (2005), pp. 1568–1570.
[28] R. Vershynin, *Introduction to the non-asymptotic analysis of random matrices*, in Compressed Sensing, Theory and Applications, Y. Eldar and G. Kutyniok, eds., Cambridge University Press, Cambridge, 2012, pp. 210–268.