Abstract. Compressive sensing predicts that sufficiently sparse vectors can be recovered from highly incomplete information using efficient recovery methods such as $\ell_1$-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 $\ell_1$-minimization. We provide conditions on the number of samples in terms of the sparsity and the signal length which guarantees that a fixed sparse signal can be recovered with a random draw of the matrix using $\ell_1$-minimization. Our proofs are short and provide explicit and good constants.

Key words. Compressed Sensing, sparse recovery, random matrices, $\ell_1$-minimization

AMS subject classifications. 94A20, 60B20

1. Introduction. Compressive sensing allows to reconstruct signals from far fewer measurements than what was considered necessary before. The seminal papers by E. Candès, J. Romberg, T. Tao [5, 7] and by D. Donoho [12] have triggered a large research activity in mathematics, engineering and computer science with a lot of potential 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 then there are infinitely many solutions to this equation. The situation dramatically changes if $x$ is sparse, that is, $\|x\|_0 := \#\{\ell, x_\ell \neq 0\}$ is small. We note that $\|\cdot\|_0$ is called $\ell_0$-norm although it is not a norm.

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

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

where $Ax = y$. Unfortunately, this problem is $\mathbf{NP}$-hard in general. It has become common to replace the $\ell_0$-minimization problem by the $\ell_1$-minimization problem

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

where $Ax = y$. This problem can be solved by efficient convex optimization techniques [3]. As a key result of compressive sensing, under appropriate conditions on $A$ and on the sparsity of $x$, $\ell_1$-minimization indeed reconstructs the original $x$. Certain random matrices $A$ are known to provide optimal recovery guarantees 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 the draw of the matrix $A$ under appropriate conditions. The difference to uniform recovery is that nonuniform recovery does not 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 $x$. 

---

*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).
Uniform recovery via $\ell_1$-minimization is for instance satisfied if the by-now classical restricted isometry property (RIP) holds for $A$ with high probability \cite{4,6}. A common choice is to take $A \in \mathbb{R}^{m \times N}$ as a Gaussian random matrix, that is, the entries of $A$ are independent standard normal distributed random variables. If, for $\varepsilon \in (0,1)$,

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

then with probability at least $1 - \varepsilon$ we have uniform recovery of all $s$-sparse vectors $x \in \mathbb{R}^N$ using $\ell_1$-minimization and $A$ as measurement matrix, see e.g. \cite{7,14,23}. The constants $C > 0$ is universal and estimates via the restricted isometry property give a value of around $C \approx 200$, which is significantly worse than what can be observed in practice. (Note that a direct analysis in \cite{25} 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 $\ell_1$-minimization. Our main results below provide non-uniform recovery guarantees with an explicit and good constant.

In contrast to other works such as \cite{13,14} we can treat also the recovery of complex vectors. We also get good constants in the subgaussian case, and in particular, for Bernoulli matrices. Moreover, our results also extend to stability of 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 \cite{17}. Moreover, even if certain applications require more structure of the measurement matrix (leading to structured random matrices \cite{24}), the empirically observed recovery performance of many types of matrices is very close to the one of (sub-)Gaussian random matrices \cite{15}, 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 normal distributed random variables, that is, having mean zero and variance 1. Our nonuniform sparse recovery result for Gaussian matrices and $\ell_1$-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

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

then with probability at least $1 - \varepsilon$ the vector $x$ is the unique solution to the $\ell_1$-minimization problem (1.1).

**Remark 2.2.** For large $N$ and $s$, Condition (2.1) roughly becomes

$$m > 2s \ln(4N/\varepsilon). \quad (2.2)$$

Comparing with (1.2) we realize that the log-term falls slightly short of the optimal one $\log(N/s)$. However, we emphasize that our proof is short, and the constant is explicit and good. Indeed, when in addition $s/N$ becomes very small (this is in fact the interesting regime) then we nevertheless reveal the conditions found by Donoho and Tanner \cite{13,14}, and in particular, the optimal constant 2. Note that Donoho and Tanner used methods from random polytopes, which are quite different from 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

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

It can be shown [27] that $X$ is subgaussian with $\mathbb{E}X = 0$ if and only if there exists a constant $c$ (depending only on $\beta$ and $\theta$) such that

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

Important special cases of subgaussian mean-zero random variables are standard Gaussians, and Rademacher (Bernoulli) variables, that is, random variables that take the values $\pm 1$ with equal probability. For both of these random variables the constant $c = 1/2$, 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 constant $c$ in (2.3), and let $\varepsilon \in (0,1)$. If

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

then with probability at least $1 - \varepsilon$ the vector $x$ is the unique solution to the $\ell_1$-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 $\pm 1$ with equal probability. We are then able to give explicit constants for the constants appearing in the result of Theorem 2.3. If $Y$ is a Bernoulli random variable, then by a Taylor series expansion

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

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

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

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 $\tilde{c} = 1/12$ 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

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

then with probability at least $1 - \varepsilon$ the vector $x$ is the unique solution to the $\ell_1$-minimization problem (1.1).

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

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

\[ y = Ax + e \in \mathbb{C}^m \text{ with } ||e||_2 \leq \eta \sqrt{m}. \]  

(2.7)

It it natural to work with then with the noise constrained \( \ell_1 \) minimization problem

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

(2.8)

For the formulation of the next result we also define the error of best \( s \)-term approximation of \( x \) in \( \ell_1 \) 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 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 \( \theta \in (0, 1) \),

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

(2.9)

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

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

(2.10)

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 \( \theta \) tends to 0 we almost get 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 based on the so-called weak restricted isometry property of this theorem will be given in Section 3.7. Also, a version for subgaussian random matrices can be shown.

2.5. Relation to previous work. Recently, there have been several papers 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 [14] obtain nonuniform recovery results (terminology is “weak phase transitions”) for Gaussian matrices via methods from random polytopes. They operate essentially in an asymptotic regime (although some of their results apply also for finite values of \( N, m, s \)). They consider the case that

\[ m/N \to \delta, \ s/m \to \rho, \ \log(N)/m \to 0, \ N \to \infty, \]

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

Candès and Plan give a rather general framework for nonuniform recovery in [8], 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 get 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 obtain 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 work [9] of Candès and Recht, 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 uses the duality based recovery Theorem 3.1 due to Fuchs [18] like in our approach, but then proceeds differently. They are also able to 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, unlike our paper, they do not cover stability of the reconstruction.

3. Proofs.

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

$$\text{sgn}(x)_j := \begin{cases} \frac{x_j}{\|x\|_1} & \text{if } x_j \neq 0, \\ 0 & \text{if } x_j = 0, \end{cases} \quad j \in [N].$$

The Moore-Penrose pseudo-inverse of a matrix $B$ such that $(B^* B)$ is invertible is given by $B^\dagger = (B^* B)^{-1} B^*$, so that $B^\dagger B = \text{Id}$, where $\text{Id}$ is the identity matrix.
3.2. Recovery conditions. In this section we state some theorems that were used in the proof of main theorem, 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 [18] in the real-valued case and was extended to the complex case by J. Tropp [26], see also [24, 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 $S := \text{supp}(x)$. Assume that $A_S$ is injective and that there exists a vector $h \in \mathbb{C}^m$ such that

$$A_S^* h = \text{sgn}(x_S),$$

$$|\langle A^* h \rangle_\ell| < 1, \; \ell \in [N] \setminus S.$$

Then $x$ is the unique solution to the $\ell_1$-minimization problem (1.1) with $Ax = y$.

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

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

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

then the vector $x$ is the unique solution to the $\ell_1$-minimization problem (1.1) with $y = Ax$.

3.3. Proof of the Gaussian case. We set $S := \text{supp}(x)$, which has a cardinality $s$. By Corollary 3.2, for recovery via $\ell_1$-minimization, it is sufficient to show that

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

Therefore, the failure probability for recovery is bounded by

$$P := P(\exists \ell \notin S \mid |\langle (A_S)^\dagger a_\ell, \text{sgn}(x_S) \rangle| \geq 1).$$

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

$$\nu^2 = \mathbb{E}(X^2) = \sum_{j=1}^m \mathbb{E}[(a_\ell)_j^2] \| (A_S^*)^* \text{sgn}(x_S) \|_2^2$$

$$= \| (A_S^*)^* \text{sgn}(x_S) \|_2^2 \leq \sigma_{\text{min}}^{-2} \| A_S \|_2^2 = \sigma_{\text{min}}^{-2} (A_S) s,$$

where $\sigma_{\text{min}}$ denotes the smallest singular value. The last inequality uses the fact that $\| (A_S^*)^* \|_2 = \| A_S^\dagger \|_2 = \sigma_{\text{min}}^{-1} (A_S)$. Tail estimate for a mean-zero Gaussian random variable $X$ with variance $\sigma^2$ obeys

$$P(|X| > t) \leq e^{-t^2/2\sigma^2}.$$

(See [24, Lemma 10.2].) Then it follows that

$$P \leq P \left( \exists \ell \notin S \mid |\langle (A_S)^\dagger a_\ell, \text{sgn}(x_S) \rangle| \geq 1 |\| (A_S^*)^* \text{sgn}(x_S) \|_2 < \alpha \right)$$

$$+ P \left( \| (A_S^*)^* \text{sgn}(x_S) \|_2 \geq \alpha \right)$$

$$\leq 2N \exp(-1/2\alpha^2) + P(\sigma_{\text{min}}^{-1} (A_S) \sqrt{s} \geq \alpha).$$

(3.2)
The inequality in (3.2) uses the tail estimate (3.1), the union bound, and the independence of \( a_\ell \) and \( A_S \). The first term in (3.2) is bounded by \( \varepsilon/2 \) if
\[
\alpha \leq \frac{1}{\sqrt{2\ln(4N/\varepsilon)}}. \tag{3.3}
\]
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 \( N(0, 1/m) \), which was provided in [11],
\[
P(\sigma_{\min}(B) < 1 - \sqrt{s/m - r}) \leq e^{-mr^2/2}. \tag{3.4}
\]
Its proof relies on the Slepian-Gordon Lemma [19, 20] and concentration of measure for Lipschitz functions [22].

We proceed
\[
P(\sigma_{\min}(A_S)^{1/2} \geq \alpha) = P(\sigma_{\min}(A_S) \leq \sqrt{s}/\alpha) = P\left(\sigma_{\min}(A_S/\sqrt{m}) \leq \frac{1}{\sqrt{m}} \frac{\sqrt{s}}{\alpha}\right)
\]
\[
\leq \exp\left(-m\left(1 - (\alpha^{-1} + 1)\sqrt{s/m}\right)^2\right). \tag{3.5}
\]
If we choose \( \alpha \) 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
\[
m \geq s \left[ \sqrt{2\ln(4N/\varepsilon)} + \sqrt{2\ln(2/\varepsilon)/s + 1} \right]^2,
\]
which ensures recovery with probability at least \( 1 - \varepsilon \). This concludes the proof of Theorem 2.1.

3.4. Tail estimate for sums of subgaussian variables. We will use the following estimate for sums of subgaussian random variables in the proof of Theorem 2.3. It appears for instance in [27].

**Lemma 3.3.** Let \( X_1, \ldots, X_M \) be a sequence of independent mean-zero subgaussian random variables with the same parameter \( c \) as in (2.3). Let \( \alpha \in \mathbb{R}^M \) be some vector. Then \( Z := \sum_{j=1}^M a_j X_j \) is subgaussian, that is, for \( t > 0 \),
\[
P\left(\left| \sum_{j=1}^M a_j X_j \right| \geq t\right) \leq 2\exp\left(-t^2/(4c\|\alpha\|_2^2)\right).
\]

We present the proof in the Appendix.

3.5. Conditioning of subgaussian matrices. While the following lemma is well-known in principle the right scaling in \( \delta \) seemingly has not appeared elsewhere in the literature, compare with [2, 23].

**Lemma 3.4.** Let \( S \subset [N] \) with \( \text{card}(S) = s \). Let \( A \) be an \( m \times N \) random matrix with independent, isotropic, and subgaussian rows with the same parameter \( c \) as in (2.3). Then, for \( \delta \in (0, 1) \), normalized matrix \( \tilde{A} = \frac{1}{\sqrt{m}} A \) satisfies
\[
\| \tilde{A}_S^* \tilde{A}_S - \text{Id} \|_{2 \rightarrow 2} \leq \delta
\]
with probability at least \( 1 - \varepsilon \) provided
\[
m \geq C\delta^{-2}(3s + \ln(2\varepsilon^{-1})), \tag{3.6}
\]
where \( C \) depends only on \( c \).

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

$$\mathcal{P} \leq \mathbb{P}\left( \exists \ell \not\in S \left| \langle (A_S)^*a_\ell, \text{sgn}(x_S) \rangle \right| \geq 1 \left| \left\| (A_S^*)^*\text{sgn}(x_S) \right\|_2 < \alpha \right)$$

$$(3.7) \quad + \mathbb{P}\left( \left\| (A_S^*)^*\text{sgn}(x_S) \right\|_2 \geq \alpha \right).$$

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

$$\mathbb{P}\left( \left| \langle (A_S)^*a_\ell, \text{sgn}(x_S) \rangle \right| \geq 1 \right) = \mathbb{P}\left( \sum_{j=1}^{m} (a_\ell)_j [ (A_S^*)^*\text{sgn}(x_S) ]_j \geq 1 \right) \leq 2\exp(-1/(4\alpha^2)).$$

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

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

For the second term in (3.7), we have

$$\mathbb{P}\left( \left\| (A_S^*)^*\text{sgn}(x_S) \right\|_2 \geq \alpha \right) \leq \mathbb{P}\left( \sigma_{\text{min}}^{-1}(A_S) \sqrt{s} \geq \alpha \right)$$

$$= \mathbb{P}\left( \sigma_{\text{min}}(A_S) \leq \sqrt{s}/\alpha \right) = \mathbb{P}\left( \sigma_{\text{min}}(A_S/\sqrt{m}) \leq \frac{1}{\sqrt{m}} \frac{\sqrt{s}}{\alpha} \right).$$

By Lemma 3.4 the normalized subgaussian matrix $\hat{A}_S := A_S/\sqrt{m}$ satisfy

$$\mathbb{P}(\sigma_{\text{min}}(\hat{A}_S) < 1 - \delta) < \mathbb{P}(\sigma_{\text{min}}(\hat{A}_S) < \sqrt{1 - \delta} < \mathbb{P}(\left\| \hat{A}_S^*\hat{A}_S - \text{Id} \right\|_{2\rightarrow 2} \geq \delta) < \varepsilon/2$$

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

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

Solving (3.9) for $m$ yields the condition

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

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

3.7. Stability of reconstruction. Here, we give a very brief sketch of the proof of Theorem 2.5. It uses the concept of weak restricted isometry property (weak RIP) introduced in [8].

**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

$$\left(1 - \delta_1\right)\|v\|_2^2 \leq \|Av\|_2^2 \leq \left(1 + \delta_2\right)\|v\|_2^2$$

for all $v$ supported on $S \cup R$ and all subsets $R$ in $[N] \setminus S$ with cardinality $|R| \leq r$. 
NONUNIFORM SPARSE RECOVERY WITH SUBGAUSSIAN MATRICES

The key to the proof is the following stable and robust version of the dual certificate based recovery Theorem 3.1. Its proof follows a similar strategy as in [8] and [21, 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 \( r \leq N \) and \( \delta_1, \delta_2 \in (0, 1) \) and that there exists a vector \( v \in \mathbb{C}^m \) such that, for \( \theta \in (0, 1) \),

\[
A_S^* v = \text{sgn}(x_S),
\]

(3.10)

\[
|\langle A^* v \rangle_\ell | < 1 - \theta, \ \ell \in [N] \setminus S
\]

(3.11)

\[
\|v\|_2 \leq \beta \sqrt{s}.
\]

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

\[
\min_{z \in \mathbb{C}^N} \|z\|_1 \text{ subject to } \|Az - y\| \leq \eta
\]

satisfies

(3.12)

\[
\|x - \hat{x}\|_2 \leq \frac{\sqrt{1 + \delta_2}}{1 - \delta_1} 2\eta + \left( \frac{2\sqrt{2} \max\{\delta_1, \delta_2\}}{1 - \delta_1} + \sqrt{2} \right) \left( \frac{2\beta}{\theta} \sqrt{\frac{s}{r}} \eta + \frac{2 \sigma_2(x)_1}{\theta \sqrt{r}} \right).
\]

The weak RIP is established for Gaussian random matrices by using the estimate (3.4) for the smallest singular value of a single submatrix \( A_{S|R} \) and a corresponding estimate for the largest singular value \cite{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

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

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 \( v = (A_S^*)^* \text{sgn}(x_S) \).

Condition (3.10) is analyzed then in the same way as the corresponding condition in Theorem 3.1. This leads to the appearance of \( \theta \) in (2.9). Moreover, Condition (3.11) is straightforward to verify via \( \|v\|_2 \leq \|A_S^*\|_{2 \to 2} \|\text{sgn}(x_S)\|_2 = \|A_S^*\|_{2 \to 2} \sqrt{s} \). An appropriate choice of the numbers \( \delta_1, \delta_2 \) and \( \beta \) leads to the desired result.

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

\[
\mathbb{E}\exp(\theta \sum_{j=1}^M a_j X_j) = \mathbb{E} \prod_{i=1}^M \exp(\theta a_j X_j) = \prod_{i=1}^M \mathbb{E}\exp(\theta a_j X_j) \leq \prod_{i=1}^M \exp(\theta a_j X_j)
\]

\[
= \exp(c\|a\|_2^2 \theta^2).
\]

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

\[
P(Z \geq t) = \mathbb{P}(\exp(\theta Z) \geq \exp(\theta t)) \leq \mathbb{E}[\exp(\theta Z)] e^{-\theta t} \leq e^{c\|a\|_2^2 \theta^2 - \theta t}.
\]
The optimal choice \( \theta = t/(2c\|a\|^2) \) yields
\[
\mathbb{P}(Z \geq t) \leq e^{-t^2/(4c\|a\|^2)}.
\]

Repeating the above computation with \(-Z\) instead of \(Z\) shows that
\[
\mathbb{P}(-Z \geq t) \leq e^{-t^2/(4c\|a\|^2)},
\]
and the union bound yields the desired estimate \( \mathbb{P}(|Z| \geq t) \leq 2e^{-t^2/(4c\|a\|^2)} \).

**Appendix B. Proof of Lemma 3.4.** Since most available statements have an additional \(\log(\delta^{-1})\)-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, 23].

\[
\text{(B.1)} \quad \mathbb{P}\left( \|\tilde{A}x\|^2 - \|x\|^2 > t\|x\|^2 \right) \leq 2\exp(-\tilde{c}mt^2),
\]

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. [24, Proposition 10.1], there exists a finite subset \(U\) of the unit sphere \(S = \{x \in \mathbb{R}^N, \|x\|_2 = 1\}\), which satisfies
\[
|U| \leq \left(1 + \frac{2}{\rho}\right)^s \quad \text{and} \quad \min_{u \in U} \|z - u\|_2 \leq \rho \quad \text{for all } z \in S.
\]

The concentration inequality (B.1) yields
\[
\mathbb{P}\left( \|\tilde{A}u\|^2 - \|u\|^2 > t\|u\|^2 \right. \left. \text{ for some } u \in U \right)
\leq \sum_{u \in U} \mathbb{P}\left( \|\tilde{A}u\|^2 - \|u\|^2 > t\|u\|^2 \right)
\leq 2|U| \exp(-\tilde{c}t^2 m)
\leq 2 \left(1 + \frac{2}{\rho}\right)^s \exp(-\tilde{c}t^2 m).
\]

The positive number \(t\) will be set later depending on \(\delta\) and on \(\rho\). Let us assume for now that the realization of the random matrix \(\tilde{A}\) yields
\[
\text{(B.2)} \quad \|\tilde{A}u\|^2 - \|u\|^2 \leq t \quad \text{for all } u \in U.
\]

By the above, this occurs with probability exceeding
\[
1 - 2 \left(1 + \frac{2}{\rho}\right)^s \exp(-\tilde{c}t^2 m).
\]

Next we show that (B.2) implies \(\|\tilde{A}x\|^2 - \|x\|^2 \leq \delta\) for all \(x \in S\), that is \(\|\tilde{A}_S\tilde{A}_S - \text{Id}\|_{2\rightarrow 2} \leq \delta\) (when \(t\) is determined appropriately). Let \(B = \tilde{A}_S^*\tilde{A}_S - \text{Id}\), so that we have to show \(\|B\|_{2\rightarrow 2} \leq \delta\). Note that (B.2) means that \(\langle Bu, u \rangle \leq t\) for all \(u \in U\). Now consider a vector \(x \in S\), for which we choose a vector \(u \in U\) satisfying \(\|x - u\|_2 \leq \rho < \sqrt{2} - 1\). We obtain
\[
\langle Bx, x \rangle = \langle Bu + x - u, u + x - u \rangle
= \langle Bu, u \rangle + \langle B(x - u), x - u \rangle + 2 \langle Bu, x - u \rangle
\leq \langle Bu, u \rangle + \langle B(x - u), x - u \rangle + 2 \|Bu\|_2 \|x - u\|_2
\leq t + \|B\|_{2\rightarrow 2} \rho^2 + 2 \|B\|_{2\rightarrow 2} \rho.
\]
Taking the supremum over all $x \in S$, we deduce that

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

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

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

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

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

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

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

Now we choose $\rho$ such that $\ln(1 + 2/\rho) = 3$, that is, $\rho = 2/(\epsilon^3 - 1)$. Then (B.3) gives the condition

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

with $C = 1.646 \delta^{-1}$. This concludes the proof. \qed

**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 a book draft that the second author is currently preparing with him.

**REFERENCES**


