ON SPARSE RECONSTRUCTION FROM FOURIER AND GAUSSIAN MEASUREMENTS

MARK RUDELSON AND ROMAN VERSHYNIN

ABSTRACT. This paper improves upon best known guarantees for exact reconstruction of a sparse signal f from a small universal sample of Fourier measurements. The method for reconstruction that has recently gained momentum in the Sparse Approximation Theory is to relax this highly non-convex problem to a convex problem, and then solve it as a linear program. We show that there exists a set of frequencies Ω such that one can exactly reconstruct every r-sparse signal f of length n from its frequencies in Ω, using the convex relaxation, and Ω has size $k(r,n) = O(r\log(n)\cdot\log^2(r)\log(r\log n)) = O(r\log^4 n)$. A random set Ω satisfies this with high probability. This estimate is optimal within the $\log\log n$ and $\log^3 r$ factors. We also give a relatively short argument for a similar problem with $k(r,n) \lesssim r[12+8\log(n/r)]$ Gaussian measurements. We use methods of geometric functional analysis and probability theory in Banach spaces, which makes our arguments quite short.

1. Introduction

During the last two years, the Sparse Approximation Theory benefited from a rapid development of methods based on the Linear Programming. The idea was to relax a sparse recovery problem to a convex optimization problem. The convex problem can be further be rendered as a linear program, and analyzed with all available methods of Linear Programming.

Convex relaxation of sparse recovery problems can be traced back in its rudimentary form to mid-seventies; references to its early history can be found in [31]. With the development of fast methods of Linear Programming in the eighties, the idea of convex relaxation became truly promising. It was put forward most enthusiastically and successfully by Donoho and his collaborators since the late eighties, starting from the seminal paper [17] (see Theorem 8 attributed there to Logan, and Theorem 9). There is extensive work being carried out, both in theory

Date: October 9, 2006.

The first author is partially supported by the NSF grant DMS 0245380. The second author is an Alfred P. Sloan Research Fellow. He is also partially supported by the NSF grant DMS 0401032.

and in practice, based on the convex relaxation [9, 16, 18, 19, 15, 21, 29, 30, 31, 12, 10, 11, 14, 3, 2, 5, 6, 27, 4, 7, 23].

To have theoretical guarantees for the convex relaxation method, one needs to show that *the sparse approximation problem is equivalent to its convex relaxation*. Proving this presents a mathematical challenge. Known theoretical guarantees work only for random measurements (e.g. random Gaussian and Fourier measurements).

In this paper, we improve upon the best known theoretical guarantees for universal random Fourier (and non-harmonic Fourier) measurements due to Candes and Tao [5]. Our argument is also substantially simpler. In addition, we give a relatively short argument that yields an asymptotically optimal estimate on the number of Gaussian measurements, moreover with reasonable absolute constants. Our proofs are based on methods of Geometric Functional Analysis, Such methods were recently successfully used for related problems [27], [23].

In Section 2, we state the sparse reconstruction problem and describe the convex relaxation method. A guarantee of its correctness is a general *restricted isometry condition* on the measurement ensemble, due to Candes and Tao ([6], see [4]). Under this condition, the reconstruction problem with respect to these measurements is equivalent to its convex relaxation. Section 3 then improves upon best known guarantees for universal Fourier measurements, see Theorem 3.3. Section 4 deals with Gaussian measurements, see Theorem 4.1.

2. The Sparse Reconstruction Problem and its Convex Relaxation

We want to reconstruct an unknown signal $f \in \mathbb{C}^n$ from linear measurements $\Phi f \in \mathbb{C}^k$, where Φ is some known $k \times n$ matrix, called the *measurement matrix*. In the interesting case $k < n$, the problem is underdetermined, and we are interested in the sparsest solution. We can state this as the optimization problem

$$\text{minimize } \|f^*\|_0 \text{ subject to } \Phi f^* = \Phi f, \tag{1}$$

where $\|f\|_0 = |\text{supp} f|$ is the number of nonzero coefficients of f. This problem is highly non-convex. So we will consider its *convex relaxation*:

$$\text{minimize } \|f^*\|_1 \text{ subject to } \Phi f^* = \Phi f, \tag{2}$$

where $\|f\|_p$ denotes the ℓ_p norm throughout this paper, $(\sum_{i=1}^n |f_i|^p)^{1/p}$. Problem (2) can be classically reformulated as the *linear program*

$$\text{minimize } \sum_{i=1}^n t_i \text{ subject to } -t \le f^* \le t, \ \Phi f^* = \Phi f,$$

which can be efficiently solved using general or special methods of linear programming. Then the main question is:

Under what conditions on Φ *are problems* (1) *and* (2) *equivalent?*

In this paper, we will be interested in the *exact reconstruction*, i.e. we expect that the solutions to (1) and (2) are equal to each other and to f. Results for approximate reconstruction can be derived as consequences, see [5].

For exact reconstruction to be possible at all, one has to assume that the signal f is r-sparse, that is $\operatorname{supp}(f) \le r$, and that the number of measurements $k = k(r, n)$ has to be at least twice the sparsity r. Our goal will be to find sufficient conditions (guarantees) for the exact reconstruction. The number of measurements $k(r, n)$ should be kept as small as possible. Intuitively, the number of measurements should be of the order of r, which is the 'true' dimension of f, rather than the nominal dimension n.

Various results that appeared over the last two years demonstrate that many natural measurement matrices Φ yield exact reconstruction, with the number of measurements $k(r, n) = O(r \cdot \log^a n)$ for some absolute constant a, see [3, 5, 6, 27]. In Sections 3, we improve best known estimates on k for Fourier (and, more generally, nonharmonic Fourier) matrices. In Section 4, we consider Gaussian random matrices.

A general sufficient condition for exact reconstruction is the *restricted isometry condition* on Φ, due to Candes and Tao ([6], see [4]). It roughly says that the matrix Φ acts as an almost isometry on all $O(r)$-sparse vectors. Precisely, we define the restricted isometry constant δ_r to be the smallest positive number such that the inequality

$$C(1 - \delta_r)\|x\|_2^2 \le \|\Phi_T x\|_2^2 \le C(1 + \delta_r)\|x\|_2^2 \qquad (3)$$

holds for some number $C > 0$ and for all x and all subsets $T \subset \{1, \dots, n\}$ of size $|T| \le r$, where Φ_T denotes the $k \times |T|$ matrix that consists of the columns of Φ indexed by T. The following theorem is due to Candes and Tao ([6], see [4]).

Theorem 2.1 (Restricted Isometry Condition)**.** Let Φ be a measurement matrix whose restricted isometry constant satisfies

$$\delta_{3r} + 3\delta_{4r} \le 2. \qquad (4)$$

Let f be an r-sparse signal. Then the solution to the linear program (2) is unique and is equal to f.

This theorem says that under the restricted isometry condition (4) on the measurement matrix Φ, the reconstruction problem (1) is equivalent to its convex relaxation (2) for all r-sparse functions f.

The restricted isometry condition (4) is usually difficult to check. Indeed, the number of sets T involved in this condition is exponential in r. As a result, no explicit construction of a measurement matrix is presently known that obeys the restricted isometry condition (4). All known constructions of measurement matrices are randomized.

3. Reconstruction from Fourier measurements

Our goal will be to reconstruct an r-sparse signal $f \in \mathbb{C}^n$ from its discrete Fourier transform evaluated at $k = k(r,n)$ points. These points will be chosen at random and uniformly in $\{0, \dots, n-1\}$, forming a set Ω.

The Discrete Fourier transform $\hat{f} = \Psi f$ is defined by the DFT matrix Ψ with entries

$$\Psi_{\omega,t} = \frac{1}{\sqrt{n}} \exp(-i2\pi\omega t/n), \quad \omega, t \in \{0, \dots, n-1\}.$$

So, our measurement matrix Φ is the submatrix of Ψ consisting of random rows (with indices in Ω). The restricted isometry (4) is thus a condition on the random set Ω. We will thus often say that Ω, rather than Φ, satisfies or does not satisfy the restricted isometry condition. Let us read Theorem 2.1 in this light for Fourier ensemble:

Theorem 3.1 (Restricted Isometry Condition for Fourier measurements)**.** *Assume that a set $\Omega \in \{0, \dots, n-1\}$ satisfies the restricted isometry condition* (4)*. Let f be an r-sparse function in $\mathbb{C}^n$. Then f can be exactly reconstructed from the values of its Fourier transform on Ω as a solution to the linear program*

$$\textit{minimize } \|f^*\|_1 \textit{ subject to } \hat{f}^*(\omega) = \hat{f}(\omega), \quad \omega \in \Omega.$$

We shall apply Theorem 3.1 to a random subset Ω. To this end we define Ω by *random selectors*. Let $\delta_1, \dots, \delta_n$ be independent Bernoulli random variables taking the value 1 with probability $\delta = k/n$. Set

$$\Omega = \{j \in \{1, \dots, n\} \mid \delta_j = 1\}.$$

Since $\mathbb{E}|\Omega| = k$, we say that Ω is a set of average cardinality k. Bernstein's inequality implies that $|\Omega|$ is close to k with high probability (see, e.g. [1], Appendix A). In particular, ([1], Theorems A.1.12 and A.1.13): for any $\nu > 0$

$$\mathbb{P}(|\Omega| > (1+\nu)k) \le \left(\frac{e^\nu}{(1+\nu)^{1+\nu}} \right)^k$$

and

$$\mathbb{P}(|\Omega| < (1-\nu)k) \le \exp(-k\nu^2/2).$$

Theorem 3.1 raises a major problem, which is still open:

Problem 3.2. *What is the smallest $k = k(r,n)$ such that a random subset $\Omega \in \{0, \dots, n-1\}$ of average cardinality k satisfies the restricted isometry condition* (4) *with high probability?*

The conjectured answer to Problem 3.2 is $k(r,n) = O(r \log n)$. In a breakthrough paper [3], Candes, Romberg and Tao proved a weaker (non-uniform) statement: for any given r-sparse function f in $\mathbb{C}^n$, a random set Ω of cardinality $k = O(r \log n)$ satisfies with high probability that f can be reconstructed as in Theorem 3.1. However, this does not guarantee that there exists *one* set of frequencies Ω which is good for the reconstruction of *all* functions f as required in Theorem 3.1.

The best result on Problem 3.2 has been due to Candes and Tao [5]:

$$k(r,n) = O(r \log^6 n). \tag{5}$$

More precisely, it is proved in [5] that for any $\tau > 1$ and for all n sufficiently big (depending on τ), a random subset Ω of cardinality

$$k \geq C\tau r \log^6 n$$

satisfies the restricted isometry condition (4) with probability at least $1 - n^{-c\tau}$.

The main result of this section improves upon the best known bound (5), lowering the exponent of the logarithm. Our argument is also substantially simpler than that of Candes and Tao [5].

Theorem 3.3 (Fourier measurements)**.** *For any $t > 1$ and any $n, r > 2$, a random subset Ω of average cardinality*

$$k = (Ctr \log n) \log(Ctr \log n) \log^2 r$$

satisfies the restricted isometry condition (4) *with probability at least $1 - 5e^{-ct}$.*

Remarks. 1. If one is interested in any fixed probability of success such as 0.99, then Theorem 3.3 yields a new best known bound on the number of Fourier measurements in Problem 3.2:

$$k(r,n) = O(r \log(n) \log(r \log n) \log^2 r) = O(r \log^4 n).$$

The dependence on of $k(r,n)$ on n is thus optimal within the $\log \log n$ factor and the dependence on r is optimal within the $\log^3 r$ factor. So, our estimate is especially good really sparse functions (with small values of r).

2. To compare Theorem 3.3 to the estimate (5) due to Candes and Tao [5], if one is interested in polynomial probability of success $1 - n^{-O(1)}$, then (with $t = O(\log n)$) the number of measurements is

$$k = O(r \log^2(n) \log(r \log n) \log^2 r) = O(r \log^5 n).$$

3. Our results hold for transforms more general than the discrete Fourier transform. One can replace the DFT matrix Ψ by any orthogonal matrix with entries of magnitude $O(1/\sqrt{n})$. Theorems 3.1 and 3.3 hold for any such matrix.

In the remainder of this section, we prove Theorem 3.3. For convenience, we choose $t = 1/\varepsilon^2$ for some $\varepsilon \in (0,1)$. Let Ω be a random subset of $\{0,\dots,n\}$ of size

$$k \geq C\Big(\frac{r\log n}{\varepsilon^2}\Big)\log\Big(\frac{r\log n}{\varepsilon^2}\Big)\log^2 r, \tag{6}$$

where C is some absolute constant to be chosen later.

Recall that the measurement matrix Φ consists of the rows of Ψ whose indices are in Ω. The restricted isometry constant δ_r is non-decreasing in r. So the restricted isometry condition (4) would follow from the bound

$$\delta_{4r} \leq \frac{1}{2}.$$

Replacing r by $4r$ in the condition (6) on k and adjusting the constant C if necessary, the proof will be complete if we show that

$$\mathbb{P}(\delta_r > \frac{1}{2}) \leq 5e^{-ct}, \tag{7}$$

where we look at $\delta_r = \delta_r(\Omega)$ as a random variable. Our argument consists of two steps. First we will prove a bound on the expectation $\mathbb{E}\delta_r$. Then we will use a concentration inequality to bounds the deviation of δ_r from its expectation.

Note that

$$\delta_r = \inf_{C>0}\sup_{|T|\leq r}\Big\| id_{\mathbb{C}^T} - C\sum_{i\in\Omega} y_i^T \otimes y_i^T\Big\|.$$

Here and thereafter, for vectors $x, y \in \mathbb{C}^n$ the tensor $x \otimes y$ is defined as the rank-one linear operator given by $(x\otimes y)(z) = \langle x, z\rangle y$, where $\langle\cdot\rangle$ is the canonical inner product on $\mathbb{C}^n$. The notation x^T stands for the restriction of a vector x on the coordinates in the set T. The operator $id_{\mathbb{C}^T}$ in (8) is the identity on $\mathbb{C}^T$, and the norm is the operator norm for operators on ℓ_2^T.

The orthogonality of Ψ can be expressed as $id_{\mathbb{C}^n} = \sum_{i=0}^{n-1} y_i \otimes y_i$. We shall re-normalize the vectors y_i, letting $x_i = \sqrt{n}\ y_{i-1}$. Now we have $\|x_i\|_\infty = 1$ for all i. The expectation of δ_r can now be estimated using the following probabilistic statement, which we interpret as a law of large numbers for random operators.

Theorem 3.4 (Uniform Operator Law of Large Numbers)**.** *Let $x_1,\dots,x_n$ be vectors in $\mathbb{C}^n$ with uniformly bounded entries: $\|x_i\|_\infty \leq K$ for all i. Assume that $id_{\mathbb{C}^n} = \frac{1}{n}\sum_{i=1}^n x_i \otimes x_i$. Let Ω be a random subset of $\{1,\dots,n\}$ of average size k. Then*

$$\mathbb{E}\sup_{|T|\leq r}\Big\| id_{\mathbb{C}^T} - \frac{1}{k}\sum_{i\in\Omega} x_i^T \otimes x_i^T\Big\| \leq \varepsilon \tag{8}$$

provided k satisfies (6) *(with constant C that may depend on K).*

Remark 3.5. It follows from Theorem 3.4 that

$$\mathbb{E}\delta_r \le \varepsilon. \tag{9}$$

Theorem 3.4 is proved by the techniques developed in Probability in Banach spaces. The general roadmap is similar to [25], [26].

We first observe that

$$\mathbb{E}\,\frac{1}{k}\sum_{i\in\Omega} x_i^T \otimes x_i^T = \frac{1}{n}\sum_{i=1}^{n} x_i^T \otimes x_i^T = id_{\mathbb{C}^n},$$

so the random operator whose norm we estimate in (8) has mean zero. Then a standard symmetrization technique (see [22] Lemma 6.3) implies that the left-hand side of (8) does not exceed

$$2\,\mathbb{E}\sup_{|T|\le r}\Big\|\frac{1}{k}\sum_{i\in\Omega}\varepsilon_i\, x_i^T\otimes x_i^T\Big\|,$$

where (ε_i) are independent symmetric $\{-1,1\}$-valued random variables; also (jointly) independent of Ω. Then the conclusion of Theorem 3.4 will be easily deduced from the following lemma.

Lemma 3.6. *Let $x_1,\ldots,x_k$, $k\le n$, be vectors in $\mathbb{C}^n$ with uniformly bounded entries, $\|x_i\|_\infty \le K$ for all i. Then*

$$\mathbb{E}\sup_{|T|\le r}\Big\|\sum_{i=1}^{k}\varepsilon_i\, x_i^T\otimes x_i^T\Big\| \le k_1 \sup_{|T|\le r}\Big\|\sum_{i=1}^{k} x_i^T\otimes x_i^T\Big\|^{\frac{1}{2}} \tag{10}$$

where $k_1 \le C_1(K)\sqrt{r}\log(r)\sqrt{\log n}\sqrt{\log k}$.

Let us show how Lemma 3.6 implies Theorem 3.4. We first condition on a choice of Ω and apply Lemma 3.6 for x_i, $i\in\Omega$. Then we take the expectation with respect to Ω. Denote the left hand side of (8) by E. Using Cauchy–Schwartz inequality, we obtain:

$$E \le \frac{C_1(K)\sqrt{r}\log(r)}{\sqrt{k}}\cdot\Big(\mathbb{E}\log|\Omega|\Big)^{1/2}\cdot\Bigg(\mathbb{E}\sup_{|T|\le n}\Big\|\frac{1}{k}\sum_{i\in\Omega}x_i^T\otimes x_i^T\Big\|\Bigg)^{\frac{1}{2}}.$$

By Jensen's inequality, $\mathbb{E}\log|\Omega|\le\log k$, so

$$E \le \frac{k_1}{\sqrt{k}}\Bigg(\mathbb{E}\sup_{|T|\le n}\Big\|\frac{1}{k}\sum_{i\in\Omega}x_i^T\otimes x_i^T\Big\|\Bigg)^{\frac{1}{2}} \le \frac{k_1}{\sqrt{k}}(E+1)^{1/2},$$

where the last estimate follows from the triangle inequality. This implies that

$$E \le \frac{2k_1}{\sqrt{k}},$$

provided that $\frac{k_1}{\sqrt{k}}\le 1/2$. Theorem 3.4 now follows from our choice of k made in (6). ■

Now we prove Lemma 3.6. Throughout this argument, B_p^n and B_p^T denote the unit ball of the norm $\|\cdot\|_p$ on $\mathbb{C}^n$. To this end, we first replace Bernoulli r.v.'s ε_i by standard independent normal random variables g_i, using a comparison principle (inequality (4.8) in [22]). Then our problem becomes to bound the Gaussian process, indexed by the union of the unit Euclidean balls B_2^T in $\mathbb{C}^T$ for all subsets I of $\{1, \dots, n\}$ of size at most r. We apply Dudley's inequality (Theorem 11.17 in [22]), which is a general upper bound on Gaussian processes. Let us denote the left hand side of (8) by E_1. We obtain:

$$E_1 \le C_3 \mathbb{E} \sup_{|T| \le r} \Big\| \sum_{i=1}^k g_i \; x_i^T \otimes x_i^T \Big\| = C_3 \mathbb{E} \sup_{\substack{|T|\le r \\ x \in B_2^T}} \Big| \sum_{i=1}^k g_i \langle x_i, x\rangle^2 \Big|$$
$$\le C_4 \int_0^\infty \log^{1/2} N\big(\cup_{|T|\le r} B_2^T, \delta, u\big) \, du,$$

where $N(Z, \delta, u)$ denotes the minimal number of balls of radius u in metric δ centered in points of Z, needed to cover the set Z. The metric δ in Dudley's inequality is defined by the Gaussian process, and in our case it is

$$\begin{aligned}
\delta(x,y) &= \Big[\sum_{i=1}^M \big(\langle x_i, x\rangle^2 - \langle x_i, y\rangle^2 \big)^2 \Big]^{\frac{1}{2}} \\
&\le \Big[\sum_{i=1}^k \big(\langle x_i, x\rangle + \langle x_i, y\rangle \big)^2 \Big]^{\frac{1}{2}} \; \max_{i \le k} |\langle x_i, x-y\rangle| \\
&\le 2 \max_{\substack{|T|\le r \\ z \in B_2^T}} \Big[\sum_{i=1}^k \langle x_i, z\rangle^2 \Big]^{\frac{1}{2}} \; \max_{i\le k} |\langle x_i, x-y\rangle| \\
&= 2R \max_{i \le k} |\langle x_i, x-y\rangle|,
\end{aligned}$$

where

$$R := \sup_{|T|\le r} \Big\| \sum_{i=1}^k x_i^T \otimes x_i^T \Big\|^{\frac{1}{2}}.$$

Hence

$$E_1 \le C_5 R \sqrt{r} \int_0^\infty \log^{1/2} N\big(\frac{1}{\sqrt{r}} D_2^{r,n}, \|\cdot\|_X, u\big) \, du. \tag{11}$$

Here

$$D_p^{r,n} = \bigcup_{|T|\le r} B_p^T, \quad \|x\|_X = \max_{i\le k} |\langle x_i, x\rangle|.$$

We will use containments

$$\frac{1}{\sqrt{r}} D_2^{r,n} \subseteq D_1^{r,n} \subseteq K B_X, \quad D_1^{r,n} \subseteq B_1^n, \tag{12}$$

where B_X denotes the unit ball of the norm $\|\cdot\|_X$. The second containment follows from the uniform boundedness of (x_i). We can thus replace $\frac{1}{\sqrt{r}}D_2^{r,n}$ in (11) by $D_1^{r,n}$. Comparing (11) to the right hand side of (10) we see that, in order to complete the proof of Lemma 3.6, it suffices to show that

$$\int_0^K \log^{1/2} N\big(D_1^{r,n}, \|\cdot\|_X, u\big)du \le C_6 \log(r)\sqrt{\log n}\sqrt{\log k}, \qquad (13)$$

with $C_6 = C_6(K)$. To this end, we will estimate the covering numbers in this integral in two different ways. For big u, we will just use the second containment in (12), which allows us to replace $D_1^{r,n}$ by B_1^n.

Lemma 3.7. *Let $x_1, \dots, x_k$, $k \le n$, be vectors as in Lemma 3.6. Then for all $u > 0$ we have*

$$N(B_1^n, \|\cdot\|_X, u) \le (2n)^m,$$

where $m = C_7 K^2 \log(k)/u^2$.

Proof. The proof essentially follows the argument of Carl [8]. We present a complete proof for readers convenience.

We use the empirical method of Maurey. Fix a vector $y \in B_1^n$. Define a random vector $Z \in \mathbb{R}^n$ that takes values $(0, \dots, 0, \operatorname{sign}(y(i)), 0, \dots, 0)$ with probability $|y(i)|$ each, $i = 1, \dots, n$ (all entries of that vector are zero except i-th). Here $\operatorname{sign}(z) = z/|z|$, whenever $z \ne 0$, and 0 otherwise. Note that $\mathbb{E}Z = y$. Let $Z_1, \dots, Z_m$ be independent copies of Z. Using symmetrization as before, we see that

$$E_3 := \mathbb{E}\Big\|y - \frac{1}{m}\sum_{j=1}^m Z_j\Big\|_X \le \frac{2}{m}\,\mathbb{E}\Big\|\sum_{j=1}^m \varepsilon_j Z_j\Big\|_X.$$

Now we condition on a choice of (Z_j) and take the expectation with respect to random signs (ε_j). Using comparison to Gaussian variables as before, we obtain

$$E_4 := \mathbb{E}\Big\|\sum_{j=1}^m \varepsilon_j Z_j\Big\|_X \le C_7\mathbb{E}\Big\|\sum_{j=1}^m g_j Z_j\Big\|_X$$
$$= C_7\mathbb{E}\max_{i\le k}\Big|\sum_{j=1}^m g_j\langle Z_j, x_i\rangle\Big|.$$

For each i, $\gamma_i := \sum_{j=1}^m g_j\langle Z_j, x_i\rangle$ is a Gaussian random variable with zero mean and with variance

$$\sigma_i = \Big(\sum_{j=1}^m |\langle Z_j, x_i\rangle|^2\Big)^{1/2} \le K\sqrt{m},$$

since $|\langle Z_j, x_i\rangle| \le \|x_i\|_\infty \le K$. Using a simple bound on the maximum of Gaussian random variables (see (3.13) in [22]), we obtain

$$E_4 \le C_7 \mathbb{E} \max_{i\le k} |\gamma_i| \le C_8 \sqrt{\log k} \max_{i \le k} \sigma_i \le C_8 \sqrt{\log k} K \sqrt{m}.$$

Taking the expectation with respect to (Z_j) we obtain

$$E_3 \le \frac{2}{m} \mathbb{E}(E_4) \le \frac{2C_8 K \sqrt{\log k}}{\sqrt{m}}.$$

With the choice of m made in the statement of the lemma, we conclude that $E_3 \le u$. We have shown that for every $y \in B_1^n$, there exists a $z \in \mathbb{C}^n$ of the form $z = \frac{1}{m}\sum_{j=1}^m Z_j$ such that $\|y - z\|_X \le u$. Each Z_j takes $2n$ values, so z takes $(2n)^m$ values. Hence B_1^n can be covered by at $(2n)^m$ balls of norm $\|\cdot\|_X$ of radius u. A standard argument shows that we can assume that these balls are centered in points of B_1^n. This completes the proof of Lemma 3.7. ■

For small u, we will use a simple volumetric estimate. The diameter of B_1^r considered as a set in $\mathbb{C}^n$ is at most K with respect to the norm $\|\cdot\|_X$ (this was stated as the last containment in (12)). It follows that $N(B_1^r, \|\cdot\|, u) \le (1 + 2K/u)^r$ for all $r > 0$, see (5.7) in [24]. The set $D_1^{r,n}$ consists of $d(r,n) = \sum_{j=1}^r \binom{n}{i}$ balls of form B_1^T, thus

$$N\big(D_1^{r,n}, \|\cdot\|_X, u\big) \le d(n,r)(1 + 2K/u)^r. \tag{14}$$

Now we combine the estimate of the covering number

$$N(u) = \log^{1/2} N\big(D_1^{r,n}, \|\cdot\|_X, u\big)$$

of Lemma 3.6, and the volumetric estimate (14), to bound the integral in (13). Using Stirling's approximation, we see that $d(r,n) \le (C_9 n/r)^r$. Thus

$$N(u) \le C_{10}\sqrt{r}\big[\sqrt{\log(n/r)} + \sqrt{\log(1+2/u)}\big] =: N_1(u),$$
$$N(u) \le \frac{C_{10}}{u}\sqrt{\log k}\sqrt{\log n} =: N_2(u),$$

where $C_{10} = C_{10}(K)$. Then we bound the integral in (13) as

$$\begin{aligned}\int_0^K N(u)\,du &\le \int_0^A N_1(u)\,du + \int_A^K N_2(u)\,du \\ &\le C_{11} A\sqrt{r}\big[\sqrt{\log(n/r)} + \log(1+2/A)\big] \\ &\quad + C_{11}\log(1/A)\sqrt{\log k}\sqrt{\log n},\end{aligned}$$

where $C_{11} = C_{11}(K)$. Choosing $A = 1/\sqrt{r}$, we conclude that the integral in (13) is at most

$$\sqrt{\log(n/r)} + \log r + \log(r)\sqrt{\log k}\sqrt{\log n}.$$

This proves (13), which completes the proof of Lemma 3.6 and thus of Theorems 3.4 and 3.3. ■

Now that we proved the bound (9) on the expectation of δ_r , we shall use a concentration inequality of Ledoux and Talagrand [22] to conclude the concentration result (7).

The following result is a particular case of Theorem 6.17 [22] (used along with the inequality (6.19) of [22] for $s = Rl$):

Theorem 3.8. *Let $Y_1, \dots, Y_n$ be independent symmetric random variables taking values in some Banach space. Assume that $\|Y_j\| \le R$ for all j. Then for any integers $l \ge q$, and any $t > 0$, the random variable*

$$Y = \Big\| \sum_{j=1}^{n} Y_j \Big\|$$

satisfies

$$\mathbb{P}\big(Y \ge 8q\mathbb{E}(Y) + 2Rl + t\big) \le \Big(\frac{C_{12}}{q}\Big)^l + 2\exp\Big(-\frac{t^2}{256q\mathbb{E}(Y)^2}\Big).$$

We can now prove a tail bound for the uniform operator law of large numbers, Theorem 3.4:

Theorem 3.9 (Uniform Operator LLN: Tail Bound)**.** *Under the assumptions of Theorem 3.4, the random variable*

$$X = \sup_{|T|\le r} \Big\| id_{\mathbb{C}^T} - \frac{1}{k}\sum_{j\in\Omega} x_j^T \otimes x_j^T \Big\|$$

satisfies for any $s > 1$:

$$\mathbb{P}\big(X > C_{13}s\varepsilon\big) \le 3\exp\big(-c_{13}s\varepsilon k/r\big) + 2\exp(-s^2), \tag{15}$$

where C_{13} is an absolute constant and $c_{13} > 0$ depends on K only.

Proof. Let Υ be the space of linear operators $V : \mathbb{C}^n \to \mathbb{C}^n$ equipped with the norm

$$\|V\|_\Upsilon = \sup_{|T|\le r} \|V^T\|,$$

where $V^T = P_T V P_T$ is the operator whose matrix is the submatrix of the matrix of V with rows and columns indexed in T. Let $\delta_1, \dots, \delta_n$ be independent Bernoulli random variables taking value 1 with probability $\delta = k/n$. For $j = 1, \dots, n$ define random variables

$$X_j = \frac{1}{n} id_{\mathbb{C}^n} - \frac{1}{k}\delta_j x_j \otimes x_j.$$

Let Y_j be the symmetrization of X_j:

$$Y_j = \frac{1}{k}(\delta_j - \delta_j') \cdot x_j \otimes x_j,$$

where $\delta_1', \dots, \delta_n'$ are independent copies of $\delta_1, \dots, \delta_n$. Then

$$X = \Big\| \sum_{j=1}^n X_j \Big\|_{\Upsilon}, \quad \text{and we define} \quad Y = \Big\| \sum_{j=1}^n Y_j \Big\|_{\Upsilon}.$$

We shall use standard symmetrization inequalities, which hold for any pair of independent identically distributed random variables Z, Z' which take values in some Banach space, and all $u > 0$:

$$\mathbb{E}\|Z - \mathbb{E}Z\| \le \mathbb{E}\|Z - Z'\| \le 2\mathbb{E}\|Z - \mathbb{E}Z\|,$$
$$\mathbb{P}(\|Z\| > 2\mathbb{E}Z + u) \le 2\mathbb{P}(\|Z - Z'\| > u).$$

These inequalities can be obtained by a simple application of Fubini and the triangle inequalities, see e.g. (2.5) and (6.1) in [22]. In our situation, the symmetrization inequalities give for all $u > 0$:

$$\mathbb{E}(X) \le \mathbb{E}(Y) \le 2\mathbb{E}(X), \tag{16}$$
$$\mathbb{P}(X > 2\mathbb{E}(X) + u) \le 2\mathbb{P}(Y > u). \tag{17}$$

It thus remains to bound $\mathbb{P}(Y > u)$, and we shall do so using Theorem 3.8. To estimate $R = \max_j \|Y_j\|_\Upsilon$, note that

$$\|x_j \otimes x_j\|_\Upsilon = \sup\{|\langle x_j, z\rangle|^2 : \ z \in \mathbb{C}^n, \|z\|_2 \le 1, |\mathrm{supp}(z)| \le r\},$$

and for such z we have by Hölder's inequality that

$$|\langle x_j, z\rangle| \le \|x_j\|_\infty \|z\|_1 \le \|x_j\|_\infty \cdot \sqrt{r}\|z\|_2 \le K\sqrt{r}.$$

Thus

$$\|x_j \otimes x_j\|_\Upsilon \le K^2 r,$$

and

$$R = \max_j \|Y_j\|_\Upsilon \le \max_j \frac{2}{k}\|x_j \otimes x_j\|_\Upsilon \le 2K^2 r/k. \tag{18}$$

Now we can use Theorem 3.8. Due to the symmetrization inequalities (16) and (17), we can replace X by Y in the conclusion of Theorem 3.8 with appropriate changes to the absolute constants. Recall also that, due to Theorem 3.4, $\mathbb{E}(X) \le \varepsilon$. We thus obtain

$$\mathbb{P}\big(X > (2 + 16q)\varepsilon + 2Rl + t\big) \le \Big(\frac{C_{12}}{q}\Big)^l + 2\exp\Big(- \frac{t^2}{512q\varepsilon^2}\Big)$$

for all integers $l \le q$ and all $t > 0$.

We will use this estimate for

$$q = [eC_{12}] + 1, \quad t = \sqrt{512q}s\varepsilon, \quad l = [t/R].$$

The condition $l \ge q$ is then satisfied because of estimate (18) on R and by our choice of k (where we can adjust $C = C(K)$). So with this choice, using (18) again, we conclude that

$$\mathbb{P}\big(X > (2 + 16q + 3\sqrt{512q}s)\varepsilon\big) \le \exp\Big(- \Big[\frac{\sqrt{512q}s\varepsilon k}{2K^2 r}\Big]\Big) + 2\exp(-s^2).$$

This yields the conclusion of Theorem 3.9. ■

We can now conclude the proof of Theorem 3.3, for which we need to show (7). We apply Theorem 3.9 with

$$s = \frac{1}{2C_{13}\varepsilon}.$$

Then $c_{13}s\varepsilon k/r > 1/\varepsilon^2$ by our choice of k. So the second term in (15) dominates, and we conclude that

$$\mathbb{P}(X > \frac{1}{2}) \leq 5\exp(-c_{14}/\varepsilon^2).$$

Since $\delta_r \leq X$ and $t = 1/\varepsilon^2$, this proves (7) and finishes the proof of Theorem 3.3.

4. Reconstruction from Gaussian measurements

Our goal will be to reconstruct an r-sparse signal $f \in \mathbb{R}^n$ from $k = k(r,n)$ Gaussian measurements. These are given by $\Phi f \in \mathbb{R}^k$, where Φ is a $k \times n$ random matrix ("Gaussian matrix" in the sequel), whose entries are independent $N(0,1)$ random variables. The reconstruction will be achieved by solving the linear program (2).

The problem again is to find the smallest number of measurements $k(r,n)$ for which, with high probability, we have an exact reconstruciton of every r-sparse signal f from its measurements Φf. It has recently been shown in [6, 27, 4] that

$$k(r,n) = O(r\log(n/r)), \tag{19}$$

and this was extended in [23] to sub-gaussian measurements. Estimate (19) is asymptotically optimal. However, the constant factor implicit in (19) has not been known; previous proofs of (19) yield unreasonably weak constants, which created a big gap between theoretical guarantees and good practical performance of reconstruction (2) (see e.g. [4]). Here we shall prove a first practically reasonable guarantee of the form (19).

We shall now give a quite short argument, using methods of geometric functional analysis, which yields reasonable constants in $k(r,n)$.

After the conference version of this paper [28] was presented, we learned that Donoho and Tanner have been able to compute *precise* asymptotic behavior of the constants in $k(r,n)$ in their extensive work [13].

Theorem 4.1 (Reconstruction from Gaussian measurements)**.** *Let* $0 < p < 1$. *Set*

$$k(r,n,p) = 1 + \left(18\log\frac{2.5}{1-p} + \sqrt{r\big[12 + 8\log(n/r)\big]}\cdot\alpha(r,n)\right)^2, \tag{20}$$

where

$$\alpha(r,n) = \exp\left(\frac{\log\left(1+2\log\left(\frac{en}{r}\right)\right)}{4\log\left(\frac{en}{r}\right)} + \frac{1}{24r^2\log\left(\frac{en}{r}\right)}\right).$$

If $k > k(r,n,p)$, then a $k \times n$ random Gaussian matrix Φ satisfies the following with probability greater than p. Let f be an r-sparse signal in $\mathbb{R}^n$. Then f can be exactly reconstructed from the measurements Φf as a unique solution to the linear program (2)*.*

Remarks. 1. The expression $\alpha(r,n)$, however formidable, is nicely bounded in the interesting range $n/r > 2$ and $r \geq 4$:

$$\alpha(n,r) < 1.245.$$

Also, $\alpha(r,n)$ decreases when n/r increases, and

$$\alpha(r,n) \to 1 \quad \text{when} \quad n/r \to \infty.$$

2. For large values of r, the term that contains p in (20) is negligible compared to the square root term, so in this case

$$k(r,n,p) \approx r\big[12 + 8\log(n/r)\big]\alpha^2(r,n).$$

Our proof of Theorem 4.1 is direct, we will not use the Restricted Isometry Theorem 2.1. The first part of this argument follows a general method of [23]. One interprets the exact reconstruction as the fact that the (random) kernel of Φ misses the cone generated by the (shifted) ball of ℓ_1. Then one embeds the cone in a universal set D, which is easier to handle, and proves that the random subspace does not intersect D. However, to obtain good constants as in (20), we will need to (a) improve the constant of embedding into D from [23], and (b) use Gordon's Escape Through the Mesh Theorem [20], which is tight in terms of constants. In Gordon's theorem, one measures the size of a set S in $\mathbb{R}^n$ by its *Gaussian width*

$$w(D) = \mathbb{E}\sup_{x\in S}\langle g, x\rangle,$$

where g is a random vector in $\mathbb{R}^n$ whose components are independent $N(0,1)$ random variables (Gaussian vector). The following is Gordon's theorem [20].

Theorem 4.2 (Escape Through the Mesh (Gordon))**.** *Let S be a subset of the unit Euclidean sphere S^{n-1} in $\mathbb{R}^n$. Let Y be a random $(n-k)$-dimensional subspace of $\mathbb{R}^n$, distributed uniformly in the Grassmanian with respect to the Haar measure. Assume that*

$$w(S) > \sqrt{k}.$$

Then $Y \cap S = \emptyset$ with probability at least

$$1 - 2.5\ \exp\Big(-\big(k/\sqrt{k+1} - w(S)\big)^2/18\Big).$$

Remark 4.3. Corollary 3.4 is stated in [20] with coefficient 3.5 instead of 2.5. However, the proof of Theorem 3.3 in [20] with $K = \{0\}$ yields the coefficient 2.5.

We will now prove Theorem 4.1. First note that the function f is the unique solution of (2) if and only if 0 is the unique solution of the problem

$$\text{minimize } \|f - g^*\|_1 \text{ subject to } \Phi g^* \in \mathrm{Ker}(\Phi) =: Y. \tag{21}$$

Y is a $(n-k)$-dimensional subspace of $\mathbb{R}^n$. Due to the rotation invariance of the Gaussian random vectors, Y is distributed uniformly in the Grassmanian $G_{n-k,n}$ of $(n-k)$-dimensional subspaces of $\mathbb{R}^n$, with respect to the Haar measure.

Now, 0 is the unique solution to (21) if and only if 0 is the unique metric projection of f onto the subspace Y in the norm $\|\cdot\|_1$. This in turn is equivalent to the fact that 0 is the unique contact point between the subspace Y and the ball of the norm $\|\cdot\|_1$ centered at f:

$$(f + \|f\|_1 B_1^n) \cap Y = \{0\}. \tag{22}$$

(Recall that B_p^n is the unit ball of the norm $\|\cdot\|_p$.) Let $\mathcal{C}_f$ be the cone in $\mathbb{R}^n$ generated by the set $f + \|f\|_1 B_1^n$ (the cone of a set $A \in \mathbb{R}^n$ is defined as $\{ta \mid a \in A,\ t \in \mathbb{R}^+\}$). Then the statement that (22) holds for all r-sparse functions f is clearly equivalent to

$$\mathcal{C}_f \cap Y = \{0\} \quad \text{for all } r\text{-sparse functions } f. \tag{23}$$

We can represent the cone $\mathcal{C}_f$ as follows. Let

$$T^+ = \{i \mid f(i) > 0\},\ T^- = \{j \mid f(i) < 0\},\ T = T^+ \cup T^-.$$

Then

$$\mathcal{C}_f = \Big\{t \in \mathbb{R}^n \mid \sum_{i\in T^-} t(i) - \sum_{i \in T^+} t(i) + \sum_{i\in T^c} |t(i)| \le 0\Big\}.$$

We will now bound the cone $\mathcal{C}_f$ by a universal set, which does not depend on f.

Lemma 4.4. *Consider the spherical part of the cone, $K_f = \mathcal{C}_f \cap S^{n-1}$. Then $K_f \subset 2D$, where*

$$D = conv\ \{x \in S^{n-1} \mid |\mathrm{supp}(x)| \le r\}.$$

Proof. Fix a point $x \in \mathcal{C}_f \cap S^{n-1}$. Denote by $(x(1)^*, \dots, x(n)^*)$ the non-decreasing rearrangement of the sequence $(|x(1)|, \dots, |x(n)|)$. Since $|T| \le r$, we have

$$\sum_{j=1}^{r} x(j)^* \ge \sum_{i \in T} |x(i)| \ge \sum_{i \in T^c} |x(i)| \ge \sum_{j=r+1}^{n} x(j)^*$$

Combining this with

$$\sum_{j=1}^{r} x(j)^* \le \sqrt{r}$$

we obtain that

$$\sum_{j=r+1}^{n} x(j)^* \le \sqrt{r}.$$

Since $x \in S^{n-1}$, $x(i)^* \le 1/\sqrt{r}$ for any $i > r$. Assume for a moment that the absolute values of the coordinates of x form a decreasing sequence. Then the previous inequalities imply

$$(0, \dots, 0, x(r+1), \dots, x(n)) \in \sqrt{r} B_1^{\{r+1,\dots,n\}} \cap \frac{1}{\sqrt{r}} B_\infty^{\{r+1,\dots,n\}},$$

so

$$x \in B_2^{\{1,\dots,r\}} \times \left(\sqrt{r} B_1^{\{r+1,\dots,n\}} \cap \frac{1}{\sqrt{r}} B_\infty^{\{r+1,\dots,n\}} \right).$$

For a general $x \in K_f$ this means

$$x \in \bigcup_{|E|=r} B_2^E \times \left(\sqrt{r} B_1^{E^c} \cap \frac{1}{\sqrt{r}} B_\infty^{E^c} \right) := W,$$

where the union is taken over all r-element subsets of $\{1, \dots, n\}$.

The maximum of $\|x\|_D$ over $x \in W$ is attained at the extreme points of W, which have the form $x = x' + x''$, where $x' \in S^E$, and x'' has coordinates 0 and $\pm 1/\sqrt{r}$ with r non-zero coordinates. Thus, for any extreme point x of W,

$$\|x\|_D \le \|x'\|_D + \|x''\|_D = \|x'\|_2 + \|x''\|_2 \le 2.$$

The equality follows from the fact that both vectors x' and x'' are r-sparse. This completes the proof of the lemma. ∎

To use Gordon's escape through the mesh theorem, we have to estimate the Gaussian width of D.

Lemma 4.5.

$$w(D) \le \sqrt{r(3 + 2\log(n/r))} \cdot \alpha(r, n)$$

where

$$\alpha(r,n)=\exp\left(\frac{\log\left(1+2\log\left(\frac{en}{r}\right)\right)}{4\log\left(\frac{en}{r}\right)}+\frac{1}{24r^2\log\left(\frac{en}{r}\right)}\right)$$

Proof. By definition,

$$w(D)=\sup_{|J|=r}\Big(\sum_{i\in J}|g(i)|^2\Big)^{1/2}.$$

Let $p>1$ be a number to be chosen later. By Hölder's inequality, we have

$$\begin{aligned}w(D)&\le\mathbb{E}\Big(\sum_{|J|=r}\Big(\sum_{i\in J}|g(i)|^2\Big)^{p/2}\Big)^{1/p}\\&\le\binom{n}{r}^{1/p}\Big(\mathbb{E}\Big(\sum_{i=1}^{r}|g(i)|^2\Big)^{p/2}\Big)^{1/p}\\&\le\Big(\frac{en}{r}\Big)^{r/p}\Big(2^{p/2}\cdot\frac{\Gamma(p/2+r/2)}{\Gamma(r/2)}\Big)^{1/p}.\end{aligned}$$

By the Stirling's formula,

$$2^{p/2}\cdot\frac{\Gamma(p/2+r/2)}{\Gamma(r/2)}\le\left(1+\frac{p}{r}\right)^{\frac{r-1}{2}}\left(\frac{p+r}{e}\right)^{p/2}\cdot\exp\left(\frac{1}{12r}\right).$$

Therefore,

$$w(D)\le\left(\frac{en}{r}\right)^{r/p}\left(\frac{p+r}{e}\right)^{1/2}\cdot\left(1+\frac{p}{r}\right)^{\frac{r-1}{2p}}\cdot\exp\left(\frac{1}{12pr}\right).$$

Now set $p=2r\log(\frac{en}{r})$. Then

$$\begin{aligned}w(D)&\le(p+r)^{1/2}\cdot\left(1+\frac{p}{r}\right)^{\frac{r-1}{2p}}\cdot\exp\left(\frac{1}{12pr}\right)\\&=\sqrt{r(3+2\log(n/r)}\cdot\alpha(n,r),\end{aligned}$$

where

$$\alpha(r,n)=\exp\left(\frac{\log\left(1+2\log\left(\frac{en}{r}\right)\right)}{4\log\left(\frac{en}{r}\right)}+\frac{1}{24r^2\log\left(\frac{en}{r}\right)}\right)$$

as claimed. ■

To deduce (23) we define $S=\bigcup_f K_f$, where the union is over all r-sparse functions f. Then (23) is equivalent to

$$S\cap Y=\emptyset. \tag{24}$$

Lemma 4.4 implies that $S\subseteq 2D$. Then by Lemma 4.5,

$$w(S)\le 2w(D).$$

If $k > k(r,n,p)$, where $k(r,n,p)$ is chosen as in (20), then

$$1 - 2.5\ \exp\Big(- \big(k/\sqrt{k+1} - 2w(D)\big)^2/18\Big) > p.$$

Then (24) follows Gordon's Theorem 4.2. This completes the proof of Theorem 4.1.

Acknowledgement. We thank Emmanuel Candes for important remarks.

References

[1] N. Alon, J. H. Spencer, *The probabilistic method*, Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience [John Wiley & Sons], New York, 2000.

[2] E. Candes, J. Romberg, *Quantitative Robust Uncertainty Principles and Optimally Sparse Decompositions*, Foundations of Computational Mathematics, to appear

[3] E. Candes, J. Romberg, T. Tao, *Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information*, IEEE Inf. Theory 52 (2006) Vol 2., 489–509

[4] E. Candes, M. Rudelson, T. Tao, R. Vershynin, *Error correction via Linear Programming*, FOCS 2005 (46th Annual Symposium on Foundations of Computer Science), 295–308

[5] E. Candes, T. Tao, *Near Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?*, to appear, IEEE Inf. Theory.

[6] E. Candes, T. Tao, *Decoding by Linear Programming*, to appear, IEEE Inf. Theory.

[7] E. Candes, T. Tao, *The Dantzig selector: statistical estimation when p is much larger than n* to appear, Annals of Statistics.

[8] B. Carl, *Inequalities of Bernstein-Jackson-type and the degree of compactness of operators in Banach spaces*, Ann. Inst. Fourier (Grenoble) 35 (1985), 79–118

[9] S. Chen, D. Donoho, M. Saunders, *Atomic decomposition by basis pursuit*, SIAM J. Sci. Comput. 20 (1998), no. 1, 33–61; reprinted in: SIAM Rev. 43 (2001), no. 1, 129–159

[10] D. Donoho, *For Most Large Underdetermined Systems of Linear Equations, the minimal ℓ_1-norm solution is also the sparsest solution*, Comm. Pure Appl. Math. 59 (2006), no. 6, 797–829.

[11] D. Donoho, *For Most Large Underdetermined Systems of Linear Equations, the minimal l1-norm near-solution approximates the sparsest near-solution*, Comm. Pure Appl. Math. 59 (2006), no. 7, 907–934.

[12] D. Donoho, M. Elad, V. Temlyakov, *Stable Recovery of Sparse Overcomplete Representations in the Presence of Noise*, IEEE Trans. Inform. Theory 52 (2006), no. 1, 6–18.

[13] D. Donoho, J. Tanner, *Counting faces of randomly-projected polytopes when the projection radically lowers dimension*, submitted

[14] D. Donoho, Y. Tsaig, *Extensions of compressed sensing*, preprint

[15] D. Donoho, M. Elad, *Optimally sparse representation in general (nonorthogonal) dictionaries via ell_1 minimization*, Proc. Natl. Acad. Sci. USA 100 (2003), 2197–2202

[16] D. Donoho, X. Huo, *Uncertainty principles and ideal atomic decomposition*, IEEE Trans. Inform. Theory 47 (2001), 2845–2862

[17] D. Donoho, P. Stark, *Uncertainty principles and signal recovery*, SIAM J. Appl. Math. 49 (1989), 906–931

[18] M. Elad, A. Bruckstein, *A generalized uncertainty principle and sparse representation in pairs of bases*, IEEE Trans. Inform. Theory 48 (2002), 2558–2567

[19] A. Feuer, A. Nemirovski, *On sparse representation in pairs of bases*, IEEE Trans. Inform. Theory 49 (2003), 1579–1581

[20] Y. Gordon, *On Milman's inequality and random subspaces which escape through a mesh in* $\mathbb{R}^n$, Geometric aspects of functional analysis, Isr. Semin. 1986-87, Lect. Notes Math. 1317, 84-106 (1988).

[21] R. Gribonval, M. Nielsen, *Sparse representations in unions of bases*, IEEE Trans. Inform. Theory 49 (2003), 3320–3325.

[22] M. Ledoux and M. Talagrand, *Probability in Banach spaces*, Springer, 1991.

[23] S. Mendelson, A. Pajor, N. Tomczak-Jaegermann, *Reconstruction and subgaussian operators*, preprint

[24] G. Pisier, *The volume of convex bodies and Banach space geometry*, Cambridge Tracts in Mathematics 94, Cambridge University Press, 1989.

[25] M. Rudelson, *Approximate John's decompositions*, Operator Theory: Advances and Applications 77(1995), 245–249.

[26] M. Rudelson, *Random vectors in the isotropic position*, J. of Functional Analysis 164 (1999), 60-72.

[27] M. Rudelson, R. Vershynin, *Geometric approach to error correcting codes and reconstruction of signals*, Int. Math. Res. Not. 2005, no. 64, 4019–4041.

[28] M. Rudelson, R. Vershynin, *Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements*, CISS 2006 (40th Annual Conference on Information Sciences and Systems)

[29] J. Tropp, *Recovery of short, complex linear combinations via ℓ_1 minimization*, IEEE Trans. Inform. Theory 51 (2005), no. 4, 1568–1570.

[30] J. Tropp, *Greed is good: Algorithmic results for sparse approximation*, IEEE Trans. Inform. Theory, Vol. 50, Num. 10, October 2004, pp. 2231-2242

[31] J. Tropp, *Just relax: convex programming methods for identifying sparse signals in noise*, IEEE Trans. Inform. Theory 52 (2006), no. 3, 1030–1051.

Department of Mathematics, University of Missouri, Columbia, MO 65211, U.S.A.

E-mail address: rudelson@math.missouri.edu

Department of Mathematics, University of California, Davis, CA 95616, U.S.A.

E-mail address: vershynin@math.ucdavis.edu