# The Johnson-Lindenstrauss Lemma
# Meets Compressed Sensing

*Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin*[*]

Rice University

June 12, 2006

### Abstract

We show how two fundamental results in analysis related to $n$-widths and Compressed Sensing are intimately related to the Johnson-Lindenstrauss lemma. Our elementary approach is based on the same concentration inequalities for random inner products that have recently provided simple proofs of the Johnson-Lindenstrauss lemma. We show how these ideas lead to simple proofs of Kashin's theorems on widths of finite balls in Euclidean space (and their improvements due to Gluskin) as well as the existence of optimal Compressed Sensing measurement matrices. In the process we also prove that these measurement matrices are universal with respect to the sparsity-inducing basis.

## 1   Introduction

It is now well understood that matrices and projections generated by certain random processes provide solutions to fundamental questions in signal processing [2, 7, 9] and approximation ($n$-widths) [13]. In *Compressed Sensing* (CS) [2, 7], for example, a random projection of a high-dimensional but sparse signal vector onto a lower-dimensional space has been shown, with high probability, to contain enough information to enable signal reconstruction with small or zero error.

Random projections also play a fundamental role in the study of point clouds in high-dimensional spaces. The *Johnson-Lindenstrauss (JL) lemma* [12], for example, shows that with high probability the geometry of a point cloud is not disturbed by certain Lipschitz mappings onto a space of dimension logarithmic in the number of points. The statement and proofs of the JL-lemma have been simplified considerably by using random linear projections and concentration inequalities [1].

The aims of this paper are twofold. First, we show an intimate linkage between the CS theory and the JL lemma. Second, we exploit this linkage to provide simple proofs for a fundamental CS construct, the so-called *Restricted Isometry Property* (RIP). In particular, we show how elementary concentration of measure inequalities for random inner products developed for proving the JL

1

lemma in [1] together with simple covering arguments provide a simple and direct avenue to obtain core results for both $n$-widths and CS. The concentration inequalities are easy to verify for standard probability distributions such as Bernoulli, Gaussian, and other distributions [1].

This paper is organized as follows. Sections 2 and 3 recall the appropriate results on $n$-widths and CS, while Section 4 reviews the JL theory. Section 5 contains our main result relating the RIP to the JL lemma. Section 6 concludes the paper with a brief discussion of related issues.

## 2   $n$-widths

Historically, the circle of ideas of interest to us began with the work of Kashin [13] on widths in the late 1970s. Kashin proved fundamental results on widths in finite-dimensional spaces. To describe his results, we recall that the $\ell_p^N$ norm of a vector $x \in \mathbb{R}^N$ is given by

$$\|x\|_{\ell_p} := \|x\|_{\ell_p^N} := \begin{cases} \left(\sum_{j=1}^N |x_j|^p\right)^{1/p}, & 0 < p < \infty, \\ \max_{j=1,\ldots,N} |x_j|, & p = \infty. \end{cases} \tag{2.1}$$

Kashin's results pertain to Kolmogorov widths. However, they can be equivalently stated in terms of Gelfand widths, which are dual to the Kolmogorov widths. Given a compact set $K$ in the Banach space $X$, its *Gelfand width* is defined by

$$d^n(K)_X := \inf_{\text{codim}(Y)=n} \sup_{x \in K \cap Y} \|x\|_X. \tag{2.2}$$

The best spaces $Y$ in the sense of (2.2) are those that slice through $K$ in the most economical direction so as to minimize the diameter of the set $K \cap Y$.

In finite-dimensional geometry, the asymptotic behavior of the Gelfand widths of the unit ball $U(\ell_p^N)$ in $\ell_q^N$, $d^n(U(\ell_p^N))_{\ell_q^N}$ are known for all $1 \leq p, q \leq \infty$ except for the case $p = 1$, $q = \infty$. The deepest result among these is the theorem of Kashin [13], who determined the behavior of $d^n(U(\ell_1^N))_{\ell_2^N}$. The result, when put in the final form by Gluskin and Garnaev [10], states that there exists a constant $C_0 > 0$ such that for all $0 < n < N$ we have

$$C_0^{-1}\sqrt{\frac{\log(N/n)+1}{n}} \leq d^n(U(\ell_1^N))_{\ell_2^N} \leq C_0\sqrt{\frac{\log(N/n)+1}{n}}. \tag{2.3}$$

Kashin proved only the upper bounds in (2.3) and did not achieve the best exponent in the logarithm. It was Gluskin and Garnaev [10] who provided the lower bounds and the correct exponent. Kashin's constructions involved $n \times N$ random matrices $\Phi$ whose entries $\phi_{i,j}$ are independent realizations of a 0/1 Bernoulli process. Gluskin and Garnaev's improvement used matrices with Gaussian entries. It was not until quite recently that it was understood that the Bernoulli processes also yield the upper bound in (2.3) [15].

## 3   Compressed Sensing (CS)

Similar to $n$-widths, CS exploits the fact that many signal classes have a low-dimensional structure compared to the high-dimensional ambient space. CS has been recently brought to the forefront by the work of Candès, Romberg, and Tao [2] and Donoho [7], who have shown the advantages of

random projections for capturing information about sparse or compressible signals. Both groups have noted the connection of CS to $n$-widths.

In the discrete CS problem, we are interested in economically recording information about a vector (signal) $x \in \mathbb{R}^N$. We allocate a budget of $n$ nonadaptive questions to ask about $x$. Each question takes the form of a linear functional applied to $x$. Thus, the information we extract from $x$ is given by

$$\Phi x = y \tag{3.1}$$

where $\Phi$ is an $n \times N$ matrix and $y \in \mathbb{R}^n$. The matrix $\Phi$ maps $\mathbb{R}^N$, where $N$ is generally large, into $\mathbb{R}^n$, where $n$ is generally small.

To extract the information that $y$ holds about $x$, we use a decoder $\Delta$ that maps from $\mathbb{R}^n$ back into $\mathbb{R}^N$. The role of $\Delta$ is to provide an approximation $\bar{x} := \Delta(y) = \Delta(\Phi x)$ to $x$. The mapping $\Delta$ is typically nonlinear. The central question of CS is: what are the good encoder-decoder pairs $(\Phi, \Delta)$?

To measure the performance of an encoder-decoder pair $(\Phi, \Delta)$, we introduce a norm $\|\cdot\|_X$ in which we measure error. Then,

$$E(x, \Phi, \Delta)_X := \|x - \Delta(\Phi x)\|_X \tag{3.2}$$

is the error of the encoder-decoder on $x$. More generally, if $K$ is any closed and bounded set contained in $\mathbb{R}^N$, then the error of this encoder-decoder on $K$ is given by

$$E(K, \Phi, \Delta)_X := \sup_{x \in K} E(x, \Phi, \Delta)_X. \tag{3.3}$$

Thus, the error on the set $K$ is determined by the largest error on $K$. To address the question of what constitute good encoder-decoder pairs, we introduce $\mathcal{A}_{n,N} := \{(\Phi, \Delta) : \ \Phi \text{ is } n \times N\}$. The best-possible performance of an encoder-decoder on $K$ is given by

$$E_{n,N}(K)_X := \inf_{(\Phi, \Delta) \in \mathcal{A}_{n,N}} E(K, \Phi, \Delta)_X. \tag{3.4}$$

This is the so-called "minimax" way of measuring optimality that is prevalent in approximation theory, information-based complexity, and statistics.

The decoder $\Delta$ is important in practical applications of CS. In particular, we often seek decoders that can be efficiently implemented numerically. However, in the theoretical setting of this paper, the role of the decoder is less important, and one could always view the decoder to be the theoretically best one chosen for the task at hand.

There is a simple and direct relationship between $n$-widths and optimal CS. If $K \subset \mathbb{R}^N$ is any set for which $K = -K$ and for which there is a $C_1 > 0$ such that $K + K \subset C_1 K$, then

$$d^n(K)_X \leq E_{n,N}(K)_X \leq C_1 d^n(K)_X, \quad 1 \leq n \leq N. \tag{3.5}$$

The proof of (3.5) is quite straightforward. Given any CS matrix $\Phi$, we can use its null space $\mathcal{N}$ as a candidate for $Y$ in computing $d^n(K)_X$. Conversely, given any $Y$ of codimension $n$ used in computing $d^n(K)_X$, we can take any basis for the orthogonal complement of $Y$ and use these basis vectors as the rows of a CS matrix $\Phi$ for estimating $E_{n,N}(K)_X$. This correspondence between $Y$ and $\Phi$ leads easily to the proof of (3.5) (see e.g. [5] for the simple details).

As an example, we have from (2.3) and (3.5) that for all $0 < n < N$

$$C_0^{-1} \sqrt{\frac{\log(N/n) + 1}{n}} \leq E_{n,N}(U(\ell_1^N))_{\ell_2^N} \leq C_1 C_0 \sqrt{\frac{\log(N/n) + 1}{n}} \qquad (3.6)$$

with $C_0$ from (2.3). There are numerous variants of (3.6) where $\ell_1^N$ and $\ell_2^N$ are replaced by other $\ell_p^N$ spaces. For example, from the upper inequality in (3.6), one can easily derive that for any $p \leq 1$ we have

$$E_{n,N}(U(\ell_p^N))_{\ell_2^N} \leq C_2 \left[ \frac{\log(N/n) + 1}{n} \right]^{1/2 - 1/p}, \qquad (3.7)$$

where $C_2$ depends only on $p$.

One of the main problems in CS is to understand which encoder-decoder pairs $(\Phi, \Delta)$ provide estimates like (3.7). Independently Donoho [7] and Candès, Romberg, and Tao [3] have given sufficient conditions on the matrix $\Phi$ for (3.7) to hold. Both approaches show that certain random constructions generate matrices $\Phi$ with these properties. Moreover, in both of these settings, they show that the decoding can be accomplished by the linear program

$$\Delta(y) := \underset{x \; : \; \Phi x = y}{\operatorname{argmin}} \|x\|_{\ell_1^N}. \qquad (3.8)$$

Candès, Romberg, and Tao [2] show that the critical property needed for the matrices $\Phi$ is a form of restricted isometry. Given a matrix $\Phi$ and any set $T$ of column indices, we denote by $\Phi_T$ the $n \times \#(T)$ matrix composed of these columns. Similarly for a vector $x \in \mathbb{R}^N$, we denote by $x_T$ the vector obtained by retaining only the entries in $x$ corresponding to the column indices $T$. We say that the matrix $\Phi$ has the *Restricted Isometry Property* (RIP) of order $k$ if there exists a $0 < \delta_k < 1$ such that

$$(1 - \delta_k)\|x_T\|_{\ell_2}^2 \leq \|\Phi_T x_T\|_{\ell_2}^2 \leq (1 + \delta_k)\|x_T\|_{\ell_2}^2, \quad \#T \leq k. \qquad (3.9)$$

This is equivalent to having bounds on the spectral norm of the Grammian matrices $\Phi_T^t \Phi_T$, where superscript $^t$ denotes matrix transpose.

The "good" matrices for CS should satisfy (3.9) for the largest possible $k$. For example, Candès and Tao [4] show that whenever $\Phi$ satisfies the restricted isometry for $k$ and $\delta_k$, then

$$\|x - \Delta(\Phi x)\|_{\ell_2^N} \leq \frac{C_3 \|x\|_{\ell_1^N}}{\sqrt{k}}, \qquad (3.10)$$

where the constant $C_3$ depends only on $\delta_k$ from (3.9). The proof of (3.10) is quite elementary; see [2] for the original proof and [5] for this particular refinement. We see that we obtain the best possible performance (in the sense of (3.6)) if we can find matrices that satisfy the Restricted Isometry Property for $k$ of the order $n/(\log(N/n) + 1)$.

The question now before us is how can we construct matrices $\Phi$ that satisfy the RIP for the largest possible range of $k$. For a fixed value of $\delta_k$, the lower bound in (2.3) for widths combined with (3.10) shows that the widest range possible is $k \leq C_4 \, n/(\log(N/n) + 1)$. The only known constructions yielding matrices that satisfy the RIP for this range are based on random matrices. Basically, one shows that matrices built using random entries from certain probability distributions will have the RIP with high probability.

It is important to note that verifying the RIP may be a difficult task. First, this property requires bounded condition number for all submatrices built by selecting $k := \#(T)$ arbitrary columns. Thus, there are $\binom{N}{k}$ such matrices to check. Second, the spectral norm of a matrix is not generally easy to compute. In some cases, spectral norms can be bounded by using diagonal dominance, but this will not be the case for the random constructions we are interested in because subtle cancelations help control these norms.

The main point of this paper is to show that for certain random constructions of $\Phi$, the RIP follows in a simple way from the same concentration of measure inequalities for inner products that have been employed to prove the JL lemma [1, 12].

# 4   The Johnson-Lindenstrauss (JL) lemma

The Johnson-Lindenstrauss (JL) lemma is concerned with the following problem. We are given a set $Q$ of points in $\mathbb{R}^N$ with $N$ typically large. We would like to embed these points into a lower dimensional Euclidean space $\mathbb{R}^n$ while approximately preserving the relative distances between any two of these points. The question is how small can we make $n$ (relative to $\#(Q)$) and which types of embeddings work. The original formulation of Johnson and Lindenstrauss [12] is as follows.

**Lemma 4.1 [Johnson-Lindenstrauss]** *Let $\epsilon \in (0,1)$ be given. For every set $Q$ of $\#(Q)$ points in $\mathbb{R}^N$, if $n$ is a positive integer such that $n > n_0 = O(\ln(\#(Q))/\epsilon^2)$, there exists a Lipschitz mapping $f : \mathbb{R}^N \to \mathbb{R}^n$ such that*

$$(1 - \epsilon)\|u - v\|_{\ell_2^N}^2 \leq \|f(u) - f(v)\|_{\ell_2^n}^2 \leq (1 + \epsilon)\|u - v\|_{\ell_2^N}^2 \tag{4.1}$$

*for all $u, v \in Q$.*

In the past several years, various improvements have been made in both the statement and the proof of this lemma (see e.g. [1, 6, 8, 11]). In particular, there are now rather simple proofs of the lemma that show that $f$ can be taken as a linear mapping represented by an $n \times N$ matrix $\Phi$ whose entries are randomly drawn from certain probability distributions. A concise description of this evolution is provided in [1].

To describe these probabilistic constructions, let $(\Omega, \rho)$ be a probability measure space and $r$ be a random variable on $\Omega$. Given $n$ and $N$, we can generate random matrices $\Phi$ by choosing the entries $\phi_{i,j}$ as independent realizations of $r$. This yields the random matrices $\Phi(\omega)$, $\omega \in \Omega^{nN}$. It was shown in [1] that, starting with any random variable satisfying certain moment conditions, for most draws $\omega$, the matrix $\Phi(\omega)$ can be used as the function $f$ in the JL lemma. More precisely, given any set of points $Q$, with high probability the matrix $f = \Phi(\omega)$ will satisfy (4.1) provided $n$ satisfies the conditions of the lemma.

Without going into complete detail, let us mention how one proves the JL lemma using such random matrices. First, for any $x \in \mathbb{R}^N$, the random variable $\|\Phi(\omega)x\|_{\ell_2^n}^2$ has expected value $\|x\|_{\ell_2^N}^2$; that is,

$$\mathbb{E}(\|\Phi(\omega)x\|_{\ell_2^n}^2) = \|x\|_{\ell_2^N}^2. \tag{4.2}$$

Second, for any $x \in \mathbb{R}^N$, the random variable $\|\Phi(\omega)x\|_{\ell_2^n}^2$ is strongly concentrated about its expected value (thanks to the moment conditions); that is,

$$\Pr(|\,\|\Phi(\omega)x\|_{\ell_2^n}^2 - \|x\|_{\ell_2^N}^2| \geq \epsilon\|x\|_{\ell_2^N}^2) \leq 2e^{-\frac{n}{2}(\epsilon^2/2 - \epsilon^3/3)}, \quad 0 < \epsilon < 1, \tag{4.3}$$

where the probability is taken over all $n \times N$ matrices $\Phi(\omega)$ [1]. To prove the JL lemma one applies (4.3) to the differences of the points in $Q$.

Let us mention some examples of distributions that satisfy these concentration inequalities. The simplest examples are the $n \times N$ random matrices $\Phi$ whose entries $\phi_{i,j}$ are independent realizations of Gaussian random variables

$$\phi_{i,j} \sim \mathcal{N}\left(0, \frac{1}{n}\right). \tag{4.4}$$

One can also use matrices where the entries are independent realizations of $\pm$ Bernoulli random variables

$$\phi_{i,j} := \begin{cases} +\frac{1}{\sqrt{n}} & \text{with probability} \quad \frac{1}{2}, \\ -\frac{1}{\sqrt{n}} & \text{with probability} \quad \frac{1}{2}, \end{cases} \tag{4.5}$$

and a related distribution yields the matrices

$$\phi_{i,j} := \begin{cases} +\sqrt{\frac{3}{n}} & \text{with probability} \quad \frac{1}{6}, \\ 0 & \text{with probability} \quad \frac{2}{3}, \\ -\sqrt{\frac{3}{n}} & \text{with probability} \quad \frac{1}{6}. \end{cases} \tag{4.6}$$

As discussed in [1], the verification of concentration in the Gaussian case is elementary (using tail bounds for Gamma random variables), and the other two cases are verified in [1] using a straightforward relation to the moments of Gaussian random variables.

## 5   RIP via JL lemma

Our main interest is to show how the recent proofs of the JL lemma provide a simple approach to constructing random matrices that satisfy the RIP. To begin, we restrict our attention to the action of random $\Phi$ matrices on fixed $k$ dimensional subspaces. Specifically, given any set of indices $T$, denote by $X_T$ the set of all vectors in $\mathbb{R}^N$ that are zero outside of $T$. This is a $\#(T)$ dimensional linear space to which we endow the $\ell_2$ norm.

**Lemma 5.1** *Let $\Phi(\omega)$, $\omega \in \Omega^{nN}$ be a random matrix of size $n \times N$ drawn according to any distribution that satisfies the concentration inequality (4.3). Then, for any set $T$ with $\#(T) = k$ and any $0 < \delta < 1$, we have*

$$(1-\delta)\|x\|_{\ell_2^N} \leq \|\Phi(\omega)x\|_{\ell_2^n} \leq (1+\delta)\|x\|_{\ell_2^N}, \quad \text{for all} \ \ x \in X_T, \tag{5.1}$$

*with probability*

$$\geq 1 - 2(12/\delta)^k \, e^{-\frac{n}{2}(\delta^2/8 - \delta^3/24)}. \tag{5.2}$$

**Proof:** First note that it is enough to prove (5.1) in the case $\|x\|_{\ell_2^N} = 1$, since $\Phi$ is linear. Next, we choose a finite set of points $Q$ such that $Q \subseteq X_T$, $\|q\|_{\ell_2^N} \leq 1$ for all $q \in Q$, and for all $x \in X_T$ with $\|x\|_{\ell_2^N} \leq 1$ we have

$$\min_{q \in Q} \|x - q\|_{\ell_2^N} \leq \delta/4. \tag{5.3}$$

It is well known from covering numbers and easy to prove (see e.g. Chapter 13 of [14]) that we can choose such a set $Q$ with $\#(Q) \leq (12/\delta)^k$. We now apply (4.3) to this set of points with $\epsilon = \delta/2$. The result is that, with probability exceeding the right side of (5.2), we have

$$(1 - \delta/2)\|q\|_{\ell_2^N}^2 \leq \|\Phi q\|_{\ell_2^n}^2 \leq (1 + \delta/2)\|q\|_{\ell_2^N}^2, \quad \text{for all } q \in Q, \tag{5.4}$$

which trivially gives us

$$(1 - \delta/2)\|q\|_{\ell_2^N} \leq \|\Phi q\|_{\ell_2^n} \leq (1 + \delta/2)\|q\|_{\ell_2^N}, \quad \text{for all } q \in Q. \tag{5.5}$$

We now define $A$ as the smallest number such that

$$\|\Phi x\|_{\ell_2^n} \leq (1 + A)\|x\|_{\ell_2^N}, \quad \text{for all } x \in X_T, \ \|x\|_{\ell_2^N} \leq 1. \tag{5.6}$$

Our goal is to show that $A \leq \delta$. For this, we recall that for any $x \in X_T$ with $\|x\|_{\ell_2^N} \leq 1$, we can pick a $q \in Q$ such that $\|x - q\|_{\ell_2^N} \leq \delta/4$. In this case we have

$$\|\Phi x\|_{\ell_2^n} \leq \|\Phi q\|_{\ell_2^n} + \|\Phi(x - q)\|_{\ell_2^n} \leq 1 + \delta/2 + (1 + A)\delta/4 \tag{5.7}$$

Since by definition $A$ is the smallest number for which (5.6) holds, we obtain $A \leq \delta/2 + (1 + A)\delta/4$. Therefore $A \leq \frac{3\delta/4}{1 - \delta/4} \leq \delta$, as desired. We have proved the upper inequality in (5.1). The lower inequality follows from this since

$$\|\Phi x\|_{\ell_2^n} \geq \|\Phi q\|_{\ell_2^n} - \|\Phi(x - q)\|_{\ell_2^n} \geq 1 - \delta/2 - (1 + \delta)\delta/4 \geq 1 - \delta. \tag{5.8}$$

This completes the proof. □

**Theorem 5.2** *Suppose that $n$, $N$, and $0 < \delta < 1$ are given. If the probability distribution generating the $n \times N$ matrices $\Phi(\omega)$, $\omega \in \Omega^{nN}$, satisfies the concentration inequality (4.3), then there exist constants $c_0, c_1 > 0$ such that the RIP (3.9) holds for $\Phi(\omega)$ with the prescribed $\delta$ and any $k \leq c_0 n / \log(N/k)$ with probability $\geq 1 - e^{-c_1 n}$.*

**Proof:** We know that for each of the $k$ dimensional spaces $X_T$, the matrix $\Phi(\omega)$ will fail to satisfy (5.1) with probability

$$\leq 2(12/\delta)^k e^{-\frac{n}{2}(\delta^2/8 - \delta^3/24)}. \tag{5.9}$$

There are $\binom{N}{k}$ such subspaces and $\binom{N}{k} \leq (eN/k)^k$. Hence, the RIP (5) will fail to hold with probability

$$\leq 2(eN/k)^k(12/\delta)^k e^{-\frac{n}{2}(\delta^2/8 - \delta^3/24)} = e^{-na(\delta) + k[\log(eN/k) + \log(12/\delta)] + \log(2)}, \tag{5.10}$$

with $a(\delta) = \delta^2/16 - \delta^3/48$. If we choose $c_0$ sufficiently small depending only on $\delta$, then whenever $k \leq c_0 n / \log(N/k)$, we will have that the exponent in the exponential on the right side of (5.10) is $\leq -c_1 n$ with $c_1 > 0$. This proves the theorem. □

From the validity of the theorem for the range of $k \leq c_0 n / \log(N/k)$, one can easily deduce its validity for $k \leq c_0' n / [\log(N/n) + 1]$ for a constant $c_0' > 0$ depending only on $c_0$.

Above we have proved that the RIP holds for $\Phi(\omega)$ with high probability when the matrix is drawn according to one of the distributions (4.4), (4.5), or (4.6). However, we are often interested

7

in signals that are sparse or compressible in some orthonormal basis $\Psi \neq I$, in which case we would like the RIP to hold for the matrix $\Phi(\omega)\Psi$. In this setting it is easy to see that by choosing our nets of points in the $k$-dimensional subspaces spanned by columns of $\Psi$, Theorem 5.2 will establish the RIP for $\Phi(\omega)\Psi$ for each of the distributions (4.4), (4.5), and (4.6). This *universality* of $\Phi$ with respect to the sparsity-inducing basis $\Psi$ is known for the Gaussian (4.4) distribution (based on symmetry arguments), but to our knowledge the universality of the distributions (4.5) and (4.6) has not previously been shown. Indeed, we see that any distribution satisfying the concentration inequality (4.3) will provide both the canonical RIP and the universality with respect to $\Psi$. More generally, it follows that with high probability such a $\Phi$ will simultaneously satisfy the RIP with respect to an exponential number of fixed bases.

# 6    Discussion

In addition to streamlining the proofs underlying the theories of $n$-widths and CS, the bridge to the JL lemma could enable new lines of research. For instance, simplifying the verification that a random matrix $\Phi$ satisfies the RIP could aid the design of new kinds of CS measurement systems. One promising direction is towards new CS $\Phi$ that can be efficiently applied to vectors $x$, since a major application of the JL lemma has been in large database systems where efficiency is critical [1]. For example, the random matrix defined by (4.6) requires 2/3 fewer additions and multiplies than either (4.5) or (4.4).

# References

[1] D. Achlioptas, *Database-friendly random projections*, In: Proc. 20-th Annual ACM SIGACT-SIGMOD-SIGART Symp. on Principles of Database Systems, (2001), pp. 274–281.

[2] E. Candès, J. Romberg, and T. Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, IEEE Trans. Inf. Theory **52** (2006), no. 2, 489–509.

[3] E. Candès, J. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, (2005), Preprint.

[4] E. Candès and T. Tao, *Decoding by linear programing*, IEEE Trans. Inf. Theory **51**, no. 12.

[5] A. Cohen, W. Dahmen, and R. DeVore, *Remarks on compressed sensing*, (2006), Preprint.

[6] S. Dasgupta and A. Gupta, *An elementary proof of the Johnson-Lindenstrauss lemma*, Tech. Report Technical report 99-006, U.C. Berkeley, March, 1999.

[7] D. Donoho, *Compressed sensing*, IEEE Trans. Inf. Theory **52** (2006), no. 4, 1289–1306.

[8] P. Frankl and H. Maehara, *The Johnson-Lindenstrauss lemma and the sphericity of some graphs*, J. Combinatorial Theory Ser. B **44** (1988), no. 3, 355–362.

[9] A. Gilbert, S. Guha, P. Indyk, S. Muthukrishnan, and M. Strauss, *Near-optimal sparse Fourier representations via sampling*, (2005), ACM Symp. on Theoretical Computer Science, 2002.

[10] A. Garnaev and E. D. Gluskin, *The widths of Euclidean balls*, Doklady An. SSSR. **277** (1984), 1048–1052.

[11] P. Indyk and R. Motwani, *Approximate nearest neighbors: Towards removing the curse of dimenstionality*, Symp. on Theory of Computing, 1998, pp. 604–613.

[12] W. B Johnson and J. Lindenstrauss, *Extensions of Lipschitz mappings into a Hilbert space*, Conf. in Modern Analysis and Probability, 1984, pp. 189–206.

[13] B. Kashin, *The widths of certain finite dimensional sets and classes of smooth functions*, Izvestia (1977), no. 41, 334–351.

[14] G. G. Lorentz, M. von Golitschek, and Yu. Makovoz, *Constructive approximation: Advanced problems*, vol. 304, Springer Grundlehren, Berlin, 1996.

[15] A. Pajor and N. Tomczak-Jaegermann, *Subspaces of small codimension of finite dimensional banach spaces*, Proc. Amer. Math. Soc., vol. 97, 1986, pp. 637–642.