A Simple Proof of the Restricted Isometry Property for Random Matrices

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

Rice University

January 18, 2007

Abstract

We give a simple technique for verifying the Restricted Isometry Property (as introduced by Candès and Tao) for random matrices that underlies Compressed Sensing. Our approach has two main ingredients: (i) concentration inequalities for random inner products that have recently provided algorithmically simple proofs of the Johnson-Lindenstrauss lemma, (ii) covering numbers for finite dimensional balls in Euclidean space. This leads to an elementary proof of the Restricted Isometry Property and brings out connections between Compressed Sensing and the Johnson-Lindenstrauss lemma. As a result, we obtain simple and direct proofs of Kashin's theorems on widths of finite balls in Euclidean space (and their improvements due to Gluskin) and proofs of the existence of optimal Compressed Sensing measurement matrices. In the process we also prove that these measurements have a certain universality with respect to the sparsity inducing basis.

1 Introduction

It has recently become clear that matrices and projections generated by certain random processes provide solutions to a number of fundamental questions in signal processing [2, 8, 10]. In *Compressed Sensing* (CS) [2, 8], for example, a random projection of a high-dimensional but sparse or compressible 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.

However, random matrices and projections have also played a central role in a number of related fields. In particular, random projections have been used as a fundamental tool in the asymptotic theory of finite-dimensional normed spaces [18] and approximation theory [16] since the 1970s. For example, the results of Kashin and Gluskin on n-widths [11, 14] relied heavily on random matrix constructions. These same constructions were later applied in the study of point clouds in high-dimensional spaces. Specifically, the well-known *Johnson-Lindenstrauss (JL) lemma* [13] states

*This research was supported by Office of Naval Research grants ONR N00014-03-1-0051, ONR/DEPSCoR N00014-03-1-0675, ONR/DEPSCoR N00014-00-1-0470, and ONR N00014-02-1-0353; Army Research Office contract DAAD 19-02-1-0028; AFOSR grants UF/USAF F49620-03-1-0381 and FA9550-04-0148; DARPA grant N66001-06-1-2011; NSF grants DMS-354707, CCF-0431150, CNS-0435425, and CNS-0520280; and the Texas Instruments Leadership University Program. This research was completed while R. D. was the visiting Texas Instruments Professor at Rice University.

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. Over the years, simplified and sharpened forms of both the statement and proof of the JL lemma, as well as simpler and more efficient algorithms for constructing such embeddings, have been developed using elementary concentration inequalities for random inner products [1, 7, 9, 12].

The aims of this paper are twofold. First, we show an intimate linkage between the CS theory, classic results on n-widths, 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 the elementary concentration of measure inequalities for random inner products used in proving the JL lemma 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 through their common dependence on the concentration of measure phenomenon. 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 [14] on widths in the late 1970s. Kashin proved fundamental results on widths in finite-dimensional spaces. To describe his results, we recall that the ℓ_p^N norm of a vector $x \in \mathbb{R}^N$ is given by

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

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

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

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 ℓ_q^N, $d^n(U(\ell_p^N))_{\ell_q^N}$ are known for all $1 \le p, q \le \infty$ except for the case $p = 1$, $q = \infty$. The deepest result among these is the theorem of Kashin [14], 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 [11], states that there exists a constant $C_0 > 0$ such that for all $0 < n < N$ we have

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

Kashin proved only the upper bounds in (2.3) and did not achieve the best exponent in the logarithm. It was Gluskin and Garnaev [11] who provided the lower bounds and the correct exponent. Kashin's constructions involved $n \times N$ random matrices Φ 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) [17, 19]. The paper [17] also generalizes (2.3) to an arbitrary compact, convex body $K \subset \mathbb{R}^N$.

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 [8], who have shown the advantages of random projections for capturing information about sparse or compressible signals. Both groups have noted the connection between CS and 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

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

where Φ is an $n \times N$ matrix and $y \in \mathbb{R}^n$. The matrix Φ maps $\mathbb{R}^N$, where N is generally large, into $\mathbb{R}^n$, where n is typically much smaller than N.

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

To measure the performance of an encoder-decoder pair (Φ, Δ), 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 Δ 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

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

The proof of (3.5) is quite straightforward. Given any CS matrix Φ, 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 Φ for estimating $E_{n,N}(K)_X$. This correspondence between Y and Φ leads easily to the proof of (3.5) (see for example [5] for the simple details).

As an example, consider the unit ball $U(\ell_1^N)$ in ℓ_2^N. Since $U(\ell_1^N) + U(\ell_1^N) \subset 2U(\ell_1^N)$, we have from (2.3) and (3.5) that for all $0 < n < N$

$$C_0^{-1}\sqrt{\frac{\log(N/n)+1}{n}} \le E_{n,N}(U(\ell_1^N))_{\ell_2^N} \le 2C_0\sqrt{\frac{\log(N/n)+1}{n}} \tag{3.6}$$

with C_0 from (2.3). There are numerous variants of (3.6) where ℓ_1^N and ℓ_2^N are replaced by other ℓ_p^N spaces.

One of the main problems in CS is to understand which encoder-decoder pairs (Φ, Δ) provide estimates like (3.6). Independently Donoho [8] and Candès, Romberg, and Tao [3] have given sufficient conditions on the matrix Φ for (3.6) to hold. Both approaches show that certain random constructions generate matrices Φ with these properties. Moreover, in both of these settings, they show that the decoding can be accomplished by the linear program

$$\Delta(y) := \operatorname*{argmin}_{x\,:\,\Phi x = y} \|x\|_{\ell_1^N}. \tag{3.7}$$

Candès and Tao [4] introduced the following isometry condition on matrices Φ and established its important role in CS. Given a matrix Φ and any set T of column indices, we denote by Φ_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 a matrix Φ satisfies the *Restricted Isometry Property* (RIP) of order k if there exists a $\delta_k \in (0,1)$ such that

$$(1-\delta_k)\|x_T\|_{\ell_2^N}^2 \le \|\Phi_T x_T\|_{\ell_2^n}^2 \le (1+\delta_k)\|x_T\|_{\ell_2^N}^2, \tag{3.8}$$

holds for all sets T with $\#T \le k$. The condition (3.8) is equivalent to requiring that the Grammian matrix $\Phi_T^t \Phi_T$ has all of its eigenvalues in $[1-\delta_k, 1+\delta_k]$ (here Φ_T^t denotes the transpose of Φ_T).

The "good" matrices for CS should satisfy (3.8) for the largest possible k. For example, Candès and Tao [4] show that whenever Φ satisfies the RIP of order $3k$ with $\delta_{3k} < 1$, then

$$\|x - \Delta(\Phi x)\|_{\ell_2^N} \le \frac{C_2 \sigma_k(x)_{\ell_1^N}}{\sqrt{k}}, \tag{3.9}$$

where $\sigma_k(x)_{\ell_1^N}$ denotes the ℓ_1 error of the best k-term approximation, and the constant C_2 depends only on δ_{3k}. The proof of (3.9) is quite elementary; see [2] for the original proof and [5] for this particular formulation. Since $\sigma_k(x)_{\ell_1^N} \le \|x\|_{\ell_1^N}$, we see that we obtain the best possible performance (in the sense of (3.6)) if we can find matrices that satisfy the RIP for k of the order $n/(\log(N/n)+1)$.

The question now before us is how can we construct matrices Φ that satisfy the RIP for the largest possible range of k. For a fixed value of δ_k, the lower bound in (2.3) for widths combined

with (3.9) shows that the widest range possible is $k \leq C_3\, n/(\log(N/n)+1)$. 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 Φ, 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, 13], and in fact can even be viewed as a straightforward consequence of the JL lemma.

4 The Johnson-Lindenstrauss (JL) lemma and concentration of measure

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 [13] 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 for example [1, 7, 9, 12]). 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 Φ 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 (Ω, ρ) be a probability measure space and r be a random variable on Ω. Given n and N, we can generate random matrices Φ 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 ω, 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. One first proves that for any $x \in \mathbb{R}^N$, the random variable $\|\Phi(\omega)x\|_{\ell_2^n}^2$ has

expected value $\|x\|_{\ell_2^N}^2$; that is,

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

Next one must show that 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,

$$\Pr(|\,\|\Phi(\omega)x\|_{\ell_2^n}^2 - \|x\|_{\ell_2^N}^2| \geq \epsilon\|x\|_{\ell_2^N}^2) \leq 2e^{-nc_0(\epsilon)}, \quad 0 < \epsilon < 1, \tag{4.3}$$

where the probability is taken over all $n \times N$ matrices $\Phi(\omega)$ and $c_0(\epsilon)$ is a constant depending only on ϵ and such that for all $\epsilon \in (0,1)$, $c_0(\epsilon) > 0$. To prove the JL lemma one then applies (4.3) using the union bound to the set of differences between all possible pairs of points in Q.

The study of concentration of measure inequalities like (4.3) is an important subject in probability and analysis. There is now a vast literature on this subject that is described in [15]. The proof of the RIP given in the next section can begin with any random matrices that satisfy (4.3). However, in the case of CS, it is of interest to have specific examples for which one can easily establish (4.3) and for which there are reasonable bounds on the constants. We point out some examples of this type.

Perhaps the most prominent example is the $n \times N$ random matrices Φ whose entries $\phi_{i,j}$ are independent realizations of Gaussian random variables

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

The verification of (4.3) with $c_0(\epsilon) = \epsilon^2/4 - \epsilon^3/6$ is elementary using tail bounds for Gamma random variables (see [1]).

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

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

or related distributions such as

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

Again these matrices satisfy (4.3) with $c_0(\epsilon) = \epsilon^2/4 - \epsilon^3/6$ which is proved in [1] and can also be simply derived from Hoeffding's inequality (see [6]).

5 Verifying the RIP from concentration inequalities

We shall now show how the concentration of measure inequality (4.3) can be used together with covering arguments to prove the RIP for random matrices. To begin, we restrict our attention to the action of random Φ matrices on fixed k-dimensional subspaces. Specifically, given any set of indices T with $\#T \leq k$, denote by X_T the set of all vectors in $\mathbb{R}^N$ that are zero outside of T. This is a k-dimensional linear space to which we endow the ℓ_2 norm.

Our general approach will be to construct nets of points in each k-dimensional subspace, apply (4.3) to all of these points through a union bound, and then extend the result from our finite set of points to all possible k-dimensional signals. This approach is a common strategy in analysis and probability. A similar construction is used, for example, in modern proofs of Dvoretsky's Theorem (see [15, 18]). In our case, we have to add a certain bootstrapping argument to handle the fact that we do not begin with a favorable bound on the norm of Φ on these finite dimensional spaces (in fact we are trying to derive one).

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 $\#(T) = k < n$ and any $0 < \delta < 1$, we have*

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

with probability

$$\geq 1 - 2(12/\delta)^k \, e^{-c_0(\delta/2)n}. \tag{5.2}$$

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

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

It is well known from covering numbers and easy to prove (see for example Chapter 13 of [16]) that we can choose such a set Q_T with $\#(Q_T) \leq (12/\delta)^k$. We next use the union bound to apply (4.3) to this set of points with $\epsilon = \delta/2$, with the result that, with probability exceeding the right side of (5.2), we have

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

which trivially gives us

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

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 $q \in Q_T$ 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

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

which 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 $c_1, c_2 > 0$ depending only on δ such that the RIP (3.8) holds for $\Phi(\omega)$ with the prescribed δ and any $k \leq c_1 n / \log(N/k)$ with probability $\geq 1 - e^{-c_2 n}$.*

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

$$\leq 2(12/\delta)^k e^{-c_0(\delta/2)n}. \tag{5.9}$$

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

$$\leq 2(eN/k)^k (12/\delta)^k e^{-c_0(\delta/2)n} = e^{-c_0(\delta/2)n + k[\log(eN/k) + \log(12/\delta)] + \log(2)}. \tag{5.10}$$

Thus, for a fixed $c_1 > 0$, whenever $k \leq c_1 n / \log(N/k)$, we will have that the exponent in the exponential on the right side of (5.10) is $\leq -c_2 n$ provided that $c_2 > c_0(\delta/2) - c_1[1 + (1 + \log(12/\delta))/\log(N/k)]$. Hence, we can always choose $c_1 > 0$ sufficiently small to ensure that $c_2 > 0$. This proves the theorem. □

From the validity of the theorem for the range of $k \leq c_1 n / \log(N/k)$, one can easily deduce its validity for $k \leq c_1' n / [\log(N/n) + 1]$ for $c_1' > 0$ depending only on c_1.

6 Discussion

In the proof of Lemma 5.1, we directly applied (4.3) to the set of points Q_T. An alternative way to derive this result would have been to simply add the origin to our set Q_T and then apply the JL lemma. This would have yielded a similar result with only slightly worse constants (since the JL lemma seeks to preserve not only the norms, but also the interpoint distances). While we find our direct approach to be somewhat simpler, this alternative proof clearly illustrates that the RIP can be thought of as a straightforward consequence of the JL lemma, and that any distribution that yields a satisfactory JL-embedding will also generate matrices satisfying the RIP.

Furthermore, we prove above 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 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 net of points in the k-dimensional subspaces spanned by sets of k columns of Ψ, 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 Φ with respect to the sparsity-inducing basis is an attractive feature that is known for the Gaussian distribution (4.4) (based on symmetry arguments), but to our knowledge the universality of the distributions (4.5) and (4.6) has not been previously shown. Indeed, we see that any distribution satisfying a concentration inequality analogous to that of (4.3) will provide both the canonical RIP and the universality with respect to Ψ. More generally, it follows that with high probability such a Φ will simultaneously satisfy the RIP with respect to an exponential number of fixed bases.

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 Φ satisfies the RIP could aid the design of new kinds of CS measurement systems. One promising direction is towards new CS Φ 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).

Acknowledgements

We would like to thank Boris Kashin and the referees for valuable comments on this paper.

References

[1] D. Achlioptas, *Database-friendly random projections*, Proc. 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, pp. 489–509.

[3] E. Candès, J. Romberg, and T. Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Communications on Pure and Applied Mathematics **59** (2005), no. 8, pp. 1207–1223.

[4] E. Candès and T. Tao, *Decoding by linear programing*, IEEE Trans. Inf. Theory **51** (2005), no. 12, pp. 4203–4215.

[5] A. Cohen, W. Dahmen, and R. DeVore, *Compressed sensing and best k-term approximation*, (2006), Preprint.

[6] A. Cohen, W. Dahmen, and R. DeVore, *Near optimal approximation of arbitrary signals from highly incomplete measurements*, (2007), Preprint.

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

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

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

[10] 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.

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

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

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

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

[15] M. Ledoux, *The concentration of measure phenomenon*, Amer. Math. Soc., 2001.

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

[17] V. D. Milman and A. Pajor, *Regularization of star bodies by random hyperplane cut off*, Studia Math. **159** (2003), no. 2, pp. 247–261.

[18] V. D. Milman and G. Schechtman, *Asymptotic theory of finite dimensional normed spaces*, Lecture Notes in Mathematics, vol. 1200, Springer-Verlag, Berlin, 1986.

[19] S. Mendelson, A. Pajor, and N. Tomczack-Jaegermann, *Reconstruction and subgaussian operators in Aymptotic Geometric Analysis*, (2006), Preprint.