New and improved Johnson-Lindenstrauss embeddings via the Restricted Isometry Property

Felix Krahmer and Rachel Ward

October 5, 2010

Abstract

Consider an $m \times N$ matrix Φ with the *Restricted Isometry Property* of order k and level δ, that is, the norm of any k-sparse vector in $\mathbb{R}^N$ is preserved to within a multiplicative factor of $1 \pm \delta$ under application of Φ. We show that by randomizing the column signs of such a matrix Φ, the resulting map with high probability embeds *any* fixed set of $p = O(e^k)$ points in $\mathbb{R}^N$ into $\mathbb{R}^m$ without distorting the norm of any point in the set by more than a factor of $1 \pm 4\delta$. Consequently, matrices with the Restricted Isometry Property and with randomized column signs provide optimal Johnson-Lindenstrauss embeddings up to logarithmic factors in N. In particular, our results improve the best known bounds on the necessary embedding dimension m for a wide class of structured random matrices; for partial Fourier and partial Hadamard matrices, we improve the recent bound $m \gtrsim \delta^{-4} \log(p) \log^4(N)$ appearing in Ailon and Liberty [1] to $m \gtrsim \delta^{-2} \log(p) \log^4(N)$, which is optimal up to the logarithmic factors in N. Our results also have a direct application in the area of compressed sensing for redundant dictionaries.

1 Introduction

The *Johnson-Lindenstrauss* (JL) Lemma states that any set of p points in high dimensional Euclidean space can be embedded into $O(\varepsilon^{-2} \log(p))$ dimensions, without distorting the distance between any two points by more than a factor between $1 - \varepsilon$ and $1 + \varepsilon$. In its original form, the Johnson-Lindenstrauss Lemma reads as follows.

Theorem 1.1 (Johnson-Lindenstrauss Lemma [24])**.** *Let $\varepsilon \in (0, 1/2)$ and let $x_1, ..., x_p \in \mathbb{R}^N$ be arbitrary points. Let $m = O(\varepsilon^{-2} \log(p))$ be a natural number. Then there exists a Lipschitz map $f : \mathbb{R}^N \to \mathbb{R}^m$ such that*

$$(1 - \varepsilon)\|x_i - x_j\|_2^2 \le \|f(x_i) - f(x_j)\|_2^2 \le (1 + \varepsilon)\|x_i - x_j\|_2^2 \tag{1}$$

for all $i, j \in \{1, 2, ..., p\}$. Here $\| \cdot \|_2$ stands for the Euclidean norm in $\mathbb{R}^N$ or $\mathbb{R}^m$, respectively.

As shown in [3], the bound for the size of m is tight up to an $O(\log(1/\varepsilon))$ factor. In the original paper of Johnson and Lindenstrauss, it was shown that a random orthogonal projection, suitably normalized, provides such an embedding with high probability [24]. Later, this property was also verified for Gaussian random matrices, among other random matrix constructions [18, 13]. As a consequence, the JL Lemma has become a valuable tool for dimensionality reduction in a myriad of applications ranging from computer science [22], numerical linear algebra [32, 20, 16], manifold learning [4], and compressed sensing [5], [38], [8]. In order to reduce storage space and implementation time of such embeddings, the design of structured random JL embeddings has been an active area of research in recent years [2, 35, 1, 26]; see [2] or [26] for a good overview of these efforts.

In most of these frameworks, the map f under consideration is a linear map represented by an $m \times N$ matrix Φ. In this case, one can consider the set of differences $E = \{x_i - x_j\}$; to prove the theorem, one then needs to show that

$$(1 - \varepsilon)\|y\|_2^2 \le \|\Phi y\|_2^2 \le (1 + \varepsilon)\|y\|_2^2, \quad \text{for all } y \in E. \tag{2}$$

[*] Hausdorff Center for Mathematics, Universität Bonn, Bonn, Germany

[†] Courant Institute of Mathematical Science, New York University, New York, NY, USA

When Φ is a random matrix, the proof that Φ satisfies the JL lemma with high probability boils down to showing a concentration inequality of the type

$$\mathbb{P}\big((1-\varepsilon)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1+\varepsilon)\|x\|_2^2\big) \geq 1 - 2\exp(-c_0\varepsilon^2 m), \tag{3}$$

for an arbitrary fixed $x \in \mathbb{R}^N$, where c_0 is an absolute constant in the optimal case, and in addition possibly mildly dependent on N in almost-optimal scenarios as for example in [1]. Indeed it directly follows by a union bound over E (as in the proof of Theorem 3.1 below) that (2) holds with high probability.

The Johnson-Lindenstrauss Lemma in Compressed Sensing. One of the more recent applications of the Johnson-Lindenstrauss Lemma is to the area of *compressed sensing*, which is centered around the following phenomenon: For many underdetermined systems of linear equations $\Phi x = y$, the solution of minimal ℓ_1-norm is also the sparsest solution. To be precise, a vector $x \in \mathbb{R}^N$ is k-sparse if $|\{j : |x_j| > 0\}| \leq k$. A by now classical sufficient condition on the matrix Φ for guaranteeing equivalence between the minimal ℓ_1 norm solution and sparsest solution is the so-called *Restricted Isometry Property* (RIP)[9, 11, 15].

Definition 1.2. *A matrix $\Phi \in \mathbb{R}^{m\times N}$ is said to have the Restricted Isometry Property of order k and level $\delta \in (0,1)$ (equivalently, (k,δ)-RIP) if*

$$(1-\delta)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1+\delta)\|x\|_2^2 \qquad \text{for all } k\text{-sparse } x \in \mathbb{R}^N. \tag{4}$$

The restricted isometry constant δ_k is defined as the smallest value of δ for which (4) holds.

In particular, if Φ satisfies $(2k, \delta_{2k})$-RIP with $\delta_{2k} \leq 2/(3+\sqrt{\frac{7}{4}}) \approx .4627$, and if $y = \Phi x$ admits a k-sparse solution $x^{\#}$, then $x^{\#} = \arg\min_{\Phi z = y} \|z\|_1$ [17].

Gaussian and Bernoulli random matrices satisfy (k,δ)-RIP with high probability, if the embedding dimension $m \gtrsim \delta^{-2} k \log(N/k)$ [5]. Up to the constant, lower bounds for Gelfand widths of ℓ_1-balls [19, 33] show that this dependence on N and in k is optimal. The Restricted Isometry Property also holds for a rich class of structured random matrices, where usually the best known bounds for m have additional log factors in N. All known deterministic constructions of RIP matrices require that $m \gtrsim k^2$ or at least $m \gtrsim k^{2-\mu}$ for some small constant $\mu > 0$ [7].

The similarity between the expressions in (2) and (4) suggests a connection between the JL lemma and the Restricted Isometry Property. A first result in this direction was established in [5], wherein it was shown that random matrices satisfying a concentration inequality of type (3) (and hence the JL Lemma) satisfy the RIP of optimal order. More precisely, the authors prove the following theorem.

Theorem 1.3 (Theorem 5.2 in [5])**.** *Suppose that m, N, and $0 < \delta < 1$ are given. If the probability distribution generating the $m \times N$ matrices Φ satisfies the concentration inequality (3) with $\varepsilon = \delta$ and absolute constant c_0, then there exist absolute constants c_1, c_2 such that with probability $\geq 1 - 2e^{-c_2\delta^2 m}$, the RIP (4) holds for Φ with the prescribed δ and any $k \leq c_1\delta^2 m/\log(N/k)$.*

In this sense, the JL Lemma *implies* the Restricted Isometry Property.

Contribution of this work. We prove a converse result to Theorem 1.3: We show that RIP matrices, with randomized column signs, provide Johnson-Lindenstrauss embeddings that are optimal up to logarithmic factors in the ambient dimension. In particular, RIP matrices of optimal order provide Johnson-Lindenstrauss embeddings of optimal order as such, up to a logarithmic factor in N (see Proposition 3.2). Note that without randomization, such a converse is impossible as vectors in the null space of the fixed parent matrix are always mapped to zero.

This observation has several consequences in the area of compressed sensing, and also allows us to obtain improved JL embedding results for several matrix constructions with existing RIP bounds [11, 31, 28, 36, 30]. Of particular interest is the random partial Fourier or the random partial Hadamard matrix, which is formed by choosing a random subset of m rows from the $N \times N$ discrete Fourier or Hadamard matrix respectively. For these matrices, the previous best-known embedding dimension to ensure that (2) holds with probability $1 - \eta$, given by Ailon and Liberty [1], is $m \asymp \varepsilon^{-4}\log(p/\eta)\log^4(N)$. We can improve their result to have optimal dependence on the distortion, showing that $m \asymp \varepsilon^{-2}\log(p/\eta)\log^4(N)$ rows suffice for the embedding. Using results from [29], we can also improve on the necessary embedding dimension for partial circulant matrices as given in [37].

This paper is structured as follows: Section 2 introduces necessary notation. In Section 3, we state our main results, and Section 4 gives concrete examples of how these results improve on the best-known JL bounds for several matrix constructions as well as applications of our findings in compressed sensing. In Section 5 we give the relevant concentration inequalities and explicit RIP-based matrix inequalities that are needed for the proofs, which are then carried out in Section 6.

2 Notation

Before continuing, let us fix some notation to be used in the remainder. For $N \in \mathbb{N}$, we denote $[N] = \{1, \ldots, N\}$. The ℓ_p-norm of a vector $x = (x_1, \ldots, x_N) \in \mathbb{R}^N$ is defined as

$$\|x\|_p = \big(\sum_{j=1}^{N} |x_j|^p\big)^{1/p}, \quad 1 \le p < \infty,$$

and $\|x\|_\infty = \max_{j=1,\ldots,N} |x_j|$ as usual. For a matrix $\Phi = (\Phi_{j,\ell}) \in \mathbb{R}^{m\times N}$, its operator norm is $\|\Phi\| := \sup_{\|x\|_2=1} \|\Phi x\|_2$, and its Frobenius norm is defined by

$$\|\Phi\|_{\mathcal{F}} := \big(\sum_{j=1}^{m}\sum_{\ell=1}^{N} |\Phi_{j,\ell}|^2\big)^{1/2}.$$

For two functions $f, g : S \to \mathbb{R}^+$, S an arbitrary set, we write $f \gtrsim g$ if there is a constant $C > 0$ such that $f(x) \ge Cg(x)$ for all $x \in S$; we write $f \asymp g$ if $f \gtrsim g$ and $g \gtrsim f$. Let N and $s \ll N$ be given and set $R = \lceil \frac{N}{s} \rceil$. For given $x = (x_1, \ldots, x_N) \in \mathbb{R}^N$, we say that x is in *decreasing arrangement*, if one has $|x_i| \ge |x_j|$ for $i < j$. For vectors in decreasing arrangement, we decompose $x = (x_{(1)}, \ldots, x_{(J)}, \ldots, x_{(R)})$ into blocks of size $s = k/2$, i.e. $x_{(J)} \in \mathbb{R}^s$; the last block $x_{(R)}$ is potentially of smaller size. We will also consider the coarse decomposition $x = (x_{(1)}, x_{(\flat)})$, where $x_{(\flat)} = (x_{(2)}, ..., x_{(R)}) \in \mathbb{R}^{N-s}$. Denote by $]L[$ the indices corresponding to the L-th block. For $j, \ell \in [N]$ we write $j \sim l$ if the two indices are associated to the same block, and we write $j \nsim l$ otherwise. Given a matrix $\Phi \in \mathbb{R}^{m\times N}$, write Φ_j to denote the j-th column, $\Phi_{(J)} \in \mathbb{R}^{m\times s}$ to denote the matrix that is the restriction of Φ to the k columns indexed by J (again with the obvious modification for $J = R$), and $\Phi_{(\flat)}$ to denote the restriction of Φ to all but the first k columns. Finally, for a vector $x \in \mathbb{R}^N$, we denote by $D_x = (D_{i,j}) \in \mathbb{R}^{N\times N}$ the diagonal matrix $D_{j,j} = x_j$.

3 The main results

Theorem 3.1. *Fix $\eta > 0$ and $\varepsilon > 0$, and consider a finite set $E \subset \mathbb{R}^N$ of cardinality $|E| = p$. Set $k \ge 40 \log \frac{4p}{\eta}$, and suppose that $\Phi \in \mathbb{R}^{m\times N}$ satisfies the Restricted Isometry Property of order k and level $\delta \le \frac{\varepsilon}{4}$. Let $\xi \in \mathbb{R}^N$ be a Rademacher sequence, i.e., uniformly distributed on $\{-1, 1\}^N$. Then with probability exceeding $1 - \eta$,*

$$(1 - \varepsilon)\|x\|_2^2 \le \|\Phi D_\xi x\|_2^2 \le (1 + \varepsilon)\|x\|_2^2 \qquad (5)$$

uniformly for all $x \in E$.

Along the way, our method provides a direct converse to Theorem 1.3:

Proposition 3.2. *Fix $\varepsilon > 0$, and suppose that there is a constant c_3 such that for all pairs (k, m) with $k \le c_3\delta^2 m/\log(N/k)$, $\Phi = \Phi(m) \in \mathbb{R}^{m\times N}$ has the Restricted Isometry Property of order k and level $\delta \le \frac{\varepsilon}{4}$. Fix $x \in \mathbb{R}^N$ and let $\xi \in \mathbb{R}^N$ be a Rademacher sequence, i.e., uniformly distributed on $\{-1, 1\}^N$. Then there exists a constant c_4 such that for all m, ΦD_ξ satisfies the concentration inequality (3) for $c_0 = c_4 \log^{-1}\big(\frac{N}{k}\big)$.*

4 Concrete examples and applications

Using Theorem 3.1, we can improve on the best Johnson-Lindenstrauss bounds for several matrix constructions that are known to have the Restricted Isometry Property:

1. Matrices arising from bounded orthonormal systems. Consider an orthonormal system of real-valued functions φ_j, $j \in [N]$, on a measurable space $\mathcal{S}$ with respect to an orthogonalization measure $d\nu$. Such systems are called *bounded orthonormal systems* if $\sup_{j\in[N]} \sup_{x\in\mathcal{S}} |\varphi_j(x)| \leq K$ for some constant $K \geq 1$. We may associate to such a system the $m \times N$ matrix Φ with entries $\Phi_{\ell,j} = \frac{1}{\sqrt{m}}\varphi_j(x_\ell)$, where x_ℓ, $\ell \in [m]$, are drawn independently according to the orthogonalization measure $d\nu$. As shown in [11, 31, 28], matrices arising as such have (k,δ)-RIP with high probability if $m \gtrsim \delta^{-2} k \log^4(N)$. [1] By Theorem 3.1, these embeddings with randomized column signs satisfy the JL Lemma for $m \gtrsim \varepsilon^{-2} \log(p) \log^4(N)$, which is optimal up to the $\log(N)$ factors.

For measures with discrete support, such constructions are equivalent to choosing m rows at random from an $N \times N$ matrix with orthonormal rows and uniformly bounded entries. Examples include the random partial Fourier matrix or random partial Hadamard matrix, formed from the discrete Fourier matrix or discrete Hadamard matrix respectively. (In the Fourier case, we distribute the resulting real and complex parts in different coordinates, inducing an additional factor of 2.) Note that the structure of these matrices allows for fast matrix vector multiplication. Recently, Ailon and Liberty [1] verified the JL Lemma for such constructions, with column signs randomized, when $m \gtrsim \varepsilon^{-4} \log(p) \log^3(\log(p)) \log(N)$. Our result improves the factor of ε^{-4} in their result to the optimal dependence ε^{-2}. We note that while their proof also uses the RIP, it also requires arguments from [31] that are specific to discrete bounded orthonormal systems.

Examples of bounded orthonormal systems connected to continuous measures include the trigonometric system and Chebyshev system, which are associated to the uniform and Chebyshev measures, respectively. The Legendre system, while not uniformly bounded, can still be transformed via preconditioning to a bounded orthonormal system with respect to the Chebyshev measure [30]. Note that all of these constructions have an associated fast transform.

2. Partial circulant matrices. Other classes of structured random matrices known to have the RIP include *partial circulant* matrices [25, 27, 29]. In one such set-up, the first row of the $m \times N$ matrix is a Gaussian or Rademacher random vector, and each subsequent row is created by rotating one element to the right relative to the preceding row vector. Using that convolution corresponds to multiplication in the Fourier domain, these matrices have associated fast matrix-vector multiplication routines. In [29], such matrices were shown to have the RIP with high probability for $m \gtrsim \max\left(\delta^{-1} k^{\frac{3}{2}} \log^{\frac{3}{2}}(N), \delta^{-2} k \log^4(N)\right)$.

On the other hand, such a matrix composed with a diagonal matrix of random signs was shown to be a JL embedding with high probability as long as $m \gtrsim \varepsilon^{-2} \log^2(p)$ [37]. Through Theorem 3.1, the same results also obtain if $m \gtrsim \max\left(\varepsilon^{-1} \log^{3/2}\left(\frac{4p}{\eta}\right) \log^{\frac{3}{2}}(N), \varepsilon^{-2} \log\left(\frac{4p}{\eta}\right) \log^4(N)\right)$. For large p, this is an improvement compared to [37].

3. Deterministic constructions. Several deterministic constructions of RIP matrices are known, including a recent result in [7] that requires only $m \gtrsim k^{2-\mu}$. We refer the reader to the exposition in [7] for a good overview in this direction; we highlight two such deterministic constructions here. Using finite fields, DeVore [14] provides deterministic constructs of cyclic 0-1-valued matrices with (k,δ)-RIP with $m \gtrsim \delta^{-2} k^2 \log^2(N)$. Iwen [23] provides deterministic constructions of 0-1-valued matrices whose number theoretic properties allow their products with Discrete Fourier Transform (DFT) matrices to be well approximated using a few highly sparse matrix multiplications. Both the binary-valued matrices and their products with the DFT yield (k,δ)-RIP matrices with $m \gtrsim \delta^{-2} k^2 \log^2(N)$. By Theorem 3.1, the class of matrices that results by randomizing the column signs of either of these deterministic constructions satisfies the JL Lemma with $m \gtrsim \varepsilon^{-2} \log^2(p) \log^2(N)$.

Note that the amount of randomness needed to construct such embeddings is still comparable to the first two examples, requiring N random bits. Under the model assumption that the entries of each

[1] Actually, the bound is stated slightly differently in [28], namely as $\frac{m}{\log(m)} \gtrsim \delta^{-2} k \log^2(k) \log^2(N)$. However, it is easily seen that the stated bound implies this condition with possibly a different constant.

vector $x \in E$ to be embedded has random signs, however, the required randomness in the matrix is removed completely.

In addition to their fast multiplication properties, these examples have the advantage in that the construction of the matrix embedding only uses $N + m$, $2N + m$, and N independent random bits, respectively, compared to mN bits for matrices with independent entries. We note that stronger embedding results are known with fewer bits, if one imposes restrictions on the ℓ_∞ norm of the vectors $x \in E$ to be embedded – see [26] and [12].

For each of the aforementioned examples, we summarize the number of dimensions m that are known to be sufficient (k, δ)-RIP to hold. We also list the previously best known bound for JL embedding dimension (if there is one) along with the JL bounds obtained from Theorem 3.1. Where Theorem 3.1 yields a better bound than previously known, at least for some range of parameters, we highlight the result in bold face. In each of the bounds, we list only the dependence on δ, k, and N, or ε, k, and N, omitting absolute constants.

	RIP bounds	Previous JL Bound	JL Bound from Theorem 3.1
Partial Fourier	$\delta^{-2} k \log^4(N)$	$\varepsilon^{-4} \log(\frac{p}{\eta}) \log^4(N)$	$\boldsymbol{\varepsilon^{-2} \log(\frac{p}{\eta}) \log^4(N)}$
Partial Circulant	$\max\Big(\delta^{-1} k^{\frac{3}{2}} \log^{\frac{3}{2}}(N),$ $\delta^{-2} k \log^4(N)\Big)$	$\varepsilon^{-2} \log^2(\frac{p}{\eta})$	$\boldsymbol{\max\Big(\varepsilon^{-1} \log^{\frac{3}{2}}(\frac{p}{\eta}) \log^{\frac{3}{2}}(N),}$ $\boldsymbol{\varepsilon^{-2} \log(\frac{p}{\eta}) \log^4(N)\Big)}$
Deterministic (DeVore, Iwen)	$\delta^{-2} k^2 \log^2(N)$		$\boldsymbol{\varepsilon^{-2} \log^2(\frac{p}{\eta}) \log^2(N)}$
Subgaussian	$\delta^{-2} k \log(\frac{N}{k})$	$\varepsilon^{-2} \log(\frac{p}{\eta})$	$\varepsilon^{-2} \log(\frac{p}{\eta}) \log(N)$

4. Compressed sensing in redundant dictionaries As shown recently in [8], concentration inequalities of type (3) allow for the extension of the compressed sensing methodology to redundant dictionaries – in particular, tight frames – as opposed to orthonormal bases only. Since signals with sparse representations in redundant dictionaries comprise a much more realistic model of nature, this extension of compressed sensing is fundamental. Our results show that basically all random matrix constructions arising in the standard theory of compressed sensing (i.e., based on RIP estimates) also yield compressed sensing matrices for the redundant framework.

5. Compressed sensing with cross validation Compressed sensing algorithms are designed to recover approximately sparse signals; if this assumption is violated, they may yield solutions far from the input signal. In [38], a method of *cross validation* is introduced to detect such situations, and to obtain tight bounds on the error incurred by compressed sensing reconstruction algorithms in general. There, a subset $y_1 = \Phi_1 x$ of the m measurements $y = \Phi x$ are held out from the reconstruction algorithm and only the remaining measurements $y_2 = \Phi_2 x$ are used to produce a candidate approximation $\widehat{x}$ to the unknown x. If the hold-out matrix Φ_1 satisfies the Johnson-Lindenstrauss Lemma, then the observable quantity $\|\Phi_1(x - \widehat{x})\|_2$ can be used as a reliable proxy for the unknown error $\|x - \widehat{x}\|_2$. Our work shows that any RIP matrix as in the standard compressed sensing framework can be used for cross validation up to a randomization of its column signs.

6. Optimal asymptotics in δ for RIP to hold As mentioned above, it can be shown using a Gelfand width argument that $m \asymp k\log(\frac{N}{k})$ is the optimal asymptotics (in N and k) of the embedding dimension for a matrix with the restricted isometry property (4). Our results – combined with the known optimality of the asymptotics $m = \varepsilon^{-2}\log(p)$ for the embedding dimension in the Johnson-Lindenstrauss Lemma (1.1) – imply that up to a factor of $\log\left(\frac{1}{\delta}\right)$, $m \asymp \delta^{-2}$ is the optimal asymptotics in the restricted isometry constant δ for fixed N and k as $\delta \to 0$. Recall that this rate is realized by many of the above examples, such as Gaussian random matrices.

5 Proof Ingredients

The proof of Theorem 3.1 relies on concentration inequalities for Rademacher sequences and explicit RIP-based norm estimates. The first concentration result is a classical inequality by Hoeffding [21].

Proposition 5.1 (Hoeffding's Inequality)**.** *Let $x \in \mathbb{R}^N$, and let $\xi = (\xi_j)_{j=1}^N$ be a Rademacher sequence. Then, for any $t > 0$,*

$$\mathbb{P}\Big(|\sum_j \xi_j x_j| > t\Big) \le 2\exp\Big(-\frac{t^2}{2\|x\|_2^2}\Big). \tag{6}$$

The second concentration of measure result is a deviation bound for Rademacher chaos. There are many such bounds in the literature; the following inequality dates back to [34], but appeared with explicit constants and with a much simplified proof as Theorem 17 in [6].

Proposition 5.2. *Let X be the $N \times N$ matrix with entries $x_{i,j}$ and assume that $x_{i,i} = 0$ for all $i \in [N]$. Let $\xi = (\xi_j)_{j=1}^N$ be a Rademacher sequence. Then, for any $t > 0$,*

$$\mathbb{P}\Big(|\sum_{i,j} \xi_i\xi_j x_{i,j}| > t\Big) \le 2\exp\Big(-\frac{1}{64}\min\Big(\frac{\frac{96}{65}t}{\|X\|}, \frac{t^2}{\|X\|_{\mathcal{F}}^2}\Big)\Big). \tag{7}$$

We also need the following basic estimate for RIP matrices (see for instance Proposition 2.5 in [28]).

Proposition 5.3. *Suppose that $\Phi \in \mathbb{R}^{m\times N}$ has the Restricted Isometry Property of order $2s$ and level δ. Then for any two disjoint subsets $J, L \subset [N]$ of size $|J| \le s, |L| \le s$,*

$$\|\Phi_{(J)}^*\Phi_{(L)}\| \le \delta.$$

The proof of our norm estimate for RIP-matrices uses Proposition 5.3, and relies on the observation commonly used in the theory of compressed sensing (see for example [10]) that for z in decreasing arrangement and $\|z\|_2 = 1$, for $J \ge 2$ one has $\|z_{(J)}\|_\infty \le \frac{1}{\sqrt{s}}\|z_{(J-1)}\|_2$ and thus $\|z_{(\flat)}\|_\infty \le 1/\sqrt{s}$.

Proposition 5.4. *Let $R = \lceil N/s \rceil$. Let $\Phi = (\Phi_j) = (\Phi_{(1)}, \Phi_{(2)}, ..., \Phi_{(R)}) = (\Phi_{(1)}, \Phi_{(\flat)}) \in \mathbb{R}^{m\times N}$ have the $(2s,\delta)$-Restricted Isometry Property, let $x = (x_j) = (x_{(1)}, x_{(2)}, ..., x_{(R)}) = (x_{(1)}, x_{(\flat)}) \in \mathbb{R}^N$ be in decreasing arrangement with $\|x\|_2 \le 1$, and consider the symmetric matrix*

$$C \in \mathbb{R}^{N\times N}, \quad C_{j,\ell} = \begin{cases} x_j\Phi_j^*\Phi_\ell x_\ell, & j \nsim \ell, \quad j,\ell > s, \\ 0, & \textit{else}. \end{cases}$$

and, for $b \in \{-1,1\}^s$, the vector

$$v \in \mathbb{R}^N, \quad v = D_{x_{(\flat)}}\Phi_{(\flat)}^*\Phi_{(1)}D_{x_{(1)}}b.$$

The following bounds hold: $\|C\| \le \frac{\delta}{s}$, $\|C\|_{\mathcal{F}} \le \frac{\delta}{\sqrt{s}}$, *and* $\|v\|_2 \le \frac{\delta}{\sqrt{s}}$.

Proof.

$$\begin{aligned}
\|C\| &= \sup_{\|y\|_2=1} |\langle y, Cy\rangle| \\
&\leq \sup_{\|y\|_2=1} \sum_{\substack{J,L=2\\ J\neq L}}^{R} \left|\left\langle y_{(J)}, D_{x_{(J)}}\Phi^*_{(J)}\Phi_{(L)}D_{x_{(L)}}y_{(L)}\right\rangle\right| \\
&\leq \sup_{\|y\|_2=1} \sum_{\substack{J,L=2\\ J\neq L}}^{R} \|y_{(J)}\|_2\|y_{(L)}\|_2\|D_{x_{(J)}}\Phi^*_{(J)}\Phi_{(L)}D_{x_{(L)}}\| \\
&\leq \sup_{\|y\|_2=1} \sum_{\substack{J,L=2\\ J\neq L}}^{R} \|y_{(J)}\|_2\|y_{(L)}\|_2\|x_{(J)}\|_\infty\|x_{(L)}\|_\infty\delta \qquad (8)\\
&\leq \sup_{\|y\|_2=1} \sum_{J,L=2}^{R} \|y_{(J)}\|_2\|y_{(L)}\|_2\frac{1}{\sqrt{s}}\|x_{(J-1)}\|_2\frac{1}{\sqrt{s}}\|x_{(L-1)}\|_2\delta \\
&\leq \sup_{\|y\|_2=1} \frac{\delta}{s} \sum_{J,L=2}^{R} \left(\frac{1}{2}\|x_{(J-1)}\|_2^2 + \frac{1}{2}\|y_{(J)}\|_2^2\right)\left(\frac{1}{2}\|x_{(L-1)}\|_2^2 + \frac{1}{2}\|y_{(L)}\|_2^2\right) \\
&\leq \frac{\delta}{s}.
\end{aligned}$$

To obtain (8), we used Proposition 5.3.

Similarly,

$$\begin{aligned}
\|v\|_2 &\leq \sup_{\|y\|_2=1} \sum_{L=2}^{R} \left\langle y_{(L)}, D^*_{x_{(L)}}\Phi^*_{(L)}\Phi_{(1)}D_{(b)}x_{(1)}\right\rangle \\
&\leq \sup_{\|y\|_2=1} \sum_{L=2}^{R} \|y_{(L)}\|_2\|x_{(L)}\|_\infty\|\Phi^*_{(L)}\Phi_{(1)}\|\|b\|_\infty\|x_{(1)}\|_2 \\
&\leq \sup_{\|y\|_2=1} \sum_{L=2}^{R} \|y_{(L)}\|_2\frac{1}{\sqrt{s}}\|x_{(L-1)}\|_2\|\Phi^*_{(L)}\Phi_{(1)}\|\|b\|_\infty \\
&\leq \frac{\delta}{\sqrt{s}} \sup_{\|y\|_2=1} \sum_{L=2}^{R} \left(\frac{1}{2}\|y_{(L)}\|_2^2 + \frac{1}{2}\|x_{(L-1)}\|_2^2\right) \\
&\leq \frac{\delta}{\sqrt{s}}.
\end{aligned}$$

Let $\widehat{x}_{(L)} \in \mathbb{R}^N$ results from x by setting all indices except those in the Lth block to 0. Similarly, let $\widehat{\Phi}_{(L)}$ extend $\Phi_{(L)}$ by setting all columns with indices outside L to 0. Let $e_\ell \in \mathbb{R}^N$ denote the ℓth canonical basis vector. We have

$$\begin{aligned}\|C\|_{\mathcal{F}}^2 =& \sum_{\substack{j,l=s+1\\ j\nsim\ell}}^{N} (x_j\Phi_j^*\Phi_\ell x_\ell)^2\\ =& \sum_{L=2}^{R}\sum_{j\notin]L[} x_j^2\Phi_j^*\widehat{\Phi}_{(L)}D_{\widehat{x}_{(L)}}\left(\sum_{\ell\in]L[} e_\ell e_\ell^*\right)D_{\widehat{x}_{(L)}}\widehat{\Phi}_{(L)}^*\Phi_j\\ =& \sum_{L=2}^{R}\sum_{j\notin]L[} x_j^2\|D_{x_{(L)}}\Phi_{(L)}^*\Phi_j\|^2\\ \leq& \sum_{L=2}^{R}\sum_{j\notin]L[} x_j^2\|x_{(L)}\|_\infty^2\|\Phi_{(L)}^*\Phi_j\|^2\\ \leq& \sum_{L=2}^{R}\frac{\delta^2}{s}\|x_{(L-1)}\|_2^2\sum_{j=1}^{N}x_j^2\\ \leq& \frac{\delta^2}{s}.\end{aligned}$$

For the second equality, we used that $\sum_{\ell\in]L[} e_\ell e_\ell^*$ is the identity on the L-th block. □

6 Proof of the main results

Proof of Theorem 3.1. Without loss of generality, we assume that all $x \in E$ are normalized so that $\|x\|_2 = 1$. Furthermore, assume that $k = 2s$ is even.

We first consider a fixed $x \in E$, eventually taking a union bound over all x. We further assume that x is in decreasing arrangement. To achieve this, we reorder the entries of x, and permute the columns of Φ accordingly. This has no impact on the following estimates, as the Restricted Isometry Property of the matrix Φ is invariant under permutations of its columns. We need to estimate

$$\begin{aligned}\|\Phi D_\xi x\|_2^2 \;=&\; \|\Phi D_x\xi\|_2^2\\ =&\sum_{J=1}^{R}\|\Phi_{(J)}D_{x_{(J)}}\xi_{(J)}\|_2^2 \;+\; 2\xi_{(1)}^*D_{x_{(1)}}\Phi_{(1)}^*\Phi_{(\flat)}D_{x_{(\flat)}}\xi_{(\flat)} \;+\; \sum_{\substack{J,L=2\\ J\neq L}}^{R}\left\langle\Phi_{(J)}D_{x_{(J)}}\xi_{(J)},\Phi_{(L)}D_{x_{(L)}}\xi_{(L)}\right\rangle. \quad (9)\end{aligned}$$

We will bound the terms separately.

1. As Φ has the Restricted Isometry Property of order $k \geq s$ and level δ, it also has the RIP of order s and level δ, and each $\Phi_{(J)}$ is almost an isometry. Hence, noting that $\|D_{x_{(J)}}\xi_{(J)}\|_2 = \|D_{\xi_{(J)}}x_{(J)}\|_2 = \|x_{(J)}\|_2$, the first term can be estimated as follows.

   $$(1-\delta)\|x\|_2^2 \leq \sum_{J=1}^{R}\|\Phi_{(J)}D_{x_{(J)}}\xi_{(J)}\|_2^2 \leq (1+\delta)\|x\|_2^2.$$

   Thus, using that $\delta \leq \varepsilon/4$,

   $$\Big(1-\frac{\varepsilon}{4}\Big)\|x\|_2^2 \leq \sum_{J=1}^{R}\|\Phi_{(J)}D_{x_{(J)}}\xi_{(J)}\|_2^2 \leq \Big(1+\frac{\varepsilon}{4}\Big)\|x\|_2^2.$$

2. To estimate the second term, fix $\xi_{(1)} =: b$ and consider the random variable

   $$X = b^*D_{x_{(1)}}\Phi_{(1)}^*\Phi_{(\flat)}D_{x_{(\flat)}}\xi_{(\flat)} = \left\langle v,\xi_{(\flat)}\right\rangle$$

with v as in Proposition 5.4. By Hoeffding's inequality (Proposition 5.1) combined with Proposition 5.4,

$$\mathbb{P}(|X| \geq \gamma\varepsilon) \leq 2\exp\Big(-\frac{s\gamma^2\varepsilon^2}{2\delta^2}\Big). \tag{10}$$

Taking a union bound, one obtains:

$$\mathbb{P}\left(\exists x \in E : |X| \geq \gamma\varepsilon\right) \leq \exp\left(\log p + \log 2 - \frac{\gamma^2 s\varepsilon^2}{2\delta^2}\right).$$

In order for this probability to be less than $\eta/2$, we need:

$$\log 2p - \frac{s\gamma^2\varepsilon^2}{2\delta^2} \leq \log\frac{\eta}{2},$$

that is,

$$\delta \leq \varepsilon/4\sqrt{\frac{8\gamma^2 s}{\log\left(4p/\eta\right)}}. \tag{11}$$

3. We can rewrite the third term as

$$\sum_{\substack{J,L=2\\ J\neq L}}^{R} \left\langle \Phi_{(J)} D_{x_{(J)}} \xi_{(J)}, \Phi_{(L)} D_{x_{(L)}} \xi_{(L)} \right\rangle = \langle \xi, C\xi \rangle = \sum_{j,\ell=s+1}^{N} \xi_j \xi_\ell C_{j\ell},$$

where $C \in \mathbb{R}^{N\times N}$ is the matrix as in Proposition 5.4. By Proposition 5.4, we have $\|C\| \leq \frac{\delta}{s}$ and $\|C\|_{\mathcal{F}} \leq \frac{\delta}{\sqrt{s}}$, hence by Proposition 5.2

$$\mathbb{P}\left(\left|\sum_{j,\ell=s+1}^{N} \xi_j \xi_\ell C_{j\ell}\right| \geq \tau\varepsilon\right) \leq 2\exp\left(-\frac{1}{64}\min\left(\frac{s\tau^2\varepsilon^2}{\delta^2}, \frac{96\tau s\varepsilon}{65\delta}\right)\right). \tag{12}$$

Using a union bound, one obtains:

$$\mathbb{P}\left(\exists x \in E : \left|\sum_{j,\ell=s+1}^{N} \xi_j \xi_\ell C_{jl}\right| \geq \tau\varepsilon\right) \leq 2\exp\left(\log p - \frac{1}{64} s \min\left(\frac{\tau^2\varepsilon^2}{\delta^2}, \frac{96\tau\varepsilon}{65\delta}\right)\right).$$

In order for this probability to be less than $\eta/2$, we need:

$$\log 2p - \frac{1}{64} s \min\left(\frac{\tau^2\varepsilon^2}{\delta^2}, \frac{96\tau\varepsilon}{65\delta}\right) \leq \log\left(\eta/2\right),$$

that is,

$$\delta \leq \frac{\varepsilon}{4}\min\left(\sqrt{\frac{\tau^2 s}{4\log\left(\frac{4p}{\eta}\right)}}, \frac{\frac{96}{65}\tau s}{16\log\left(\frac{4p}{\eta}\right)}\right). \tag{13}$$

By assumption, $\delta \leq \frac{\varepsilon}{4}$, so conditions (11) and (13) are satisfied by setting $\tau = .55, \gamma = .1$, and $s \geq 20\log\left(4p/\eta\right)$ (that is, $k = 2s \geq 40\log\left(4p/\eta\right)$). Then the second term is bounded by $.2\delta$ in absolute value, and the last term is bounded by $.55\delta$. Together with the deterministic RIP-based estimate for the first term, this implies the Theorem. □

Sketch of the proof of Proposition 3.2. By (12) and (10), the probability that each of second and third term in (9) is $O(\delta)$ are $\geq 1 - \exp(-c_5 k)$, as long as Φ has RIP of order k. By assumption, this condition is satisfied for k (approximately) satisfying $k = c_3\varepsilon^2 m/\log(N/k)$. Hence, Inequality (2) holds with probability $\geq 1 - \exp(c_0\delta^2 m)$, where $c_0 \gtrsim \log^{-1}\left(\frac{N}{k}\right)$, as desired. □

Remarks: Although we have stated the main result for the setting $x \in \mathbb{R}^N$ and $\Phi \in \mathbb{R}^{m\times N}$, all of the analysis holds also in the complex setting, $x \in \mathbb{C}^N$ and $\Phi \in \mathbb{C}^{m\times N}$.

As shown in [5], a random matrix Φ whose entries follow a subgaussian distribution is known to with high probability have the Restricted Isometry Property of best possible order, that is, one can choose $m \asymp \delta^{-2} k \log\left(\frac{N}{k}\right)$. When $k \geq 40 \log\left(\frac{4p}{\eta}\right)$, Φ is a JL embedding by Theorem 3.1, and our resulting bound for m is optimal up to a single logarithmic factor in N. This shows that Theorem 3.1 must also be optimal up to a single logarithmic factor in N.

Acknowledgments

The authors would like to thank Holger Rauhut, Deanna Needell, Jan Vybíral, Mark Tygert, Mark Iwen, Justin Romberg, Mark Davenport, and Arie Israel for valuable discussions on this topic. Rachel Ward gratefully acknowledges the partial support of National Science Foundation Postdoctoral Research Fellowship. Felix Krahmer gratefully acknowledges the partial support of the Hausdorff Center for Mathematics. Finally, both authors are grateful for the support of the Institute of Advanced Study through the Park City Math Institute where this project was initiated.

References

[1] N. Ailon and E. Liberty. Almost optimal unrestricted fast Johnson-Lindenstrauss transform. *Preprint*, 2010.

[2] N. Ailon, E. Liberty, and A. Singer. Dense fast random projections and Lean Walsh transforms. *Proceedings of the 12th International Workshop on Randomization and Computation (RANDOM)*, pages 512–522, 2008.

[3] N. Alon. Problems and results in extremal combinatorics. *Discrete Math*, 273:31–53, 2003.

[4] R. Baraniuk and M. Wakin. Random projections of smooth manifolds. In *Foundations of Computational Mathematics*, pages 941–944, 2006.

[5] R. G. Baraniuk, M. Davenport, R. A. DeVore, and M. Wakin. A simple proof of the Restricted Isometry Property for random matrices. *Constr. Approx.*, 28(3):253–263, 2008.

[6] S. Boucheron, G. Lugosi, and P. Massart. Concentration inequalities using the entropy method. *Ann. Probab.*, 31(3):1583–1614, 2003.

[7] J. Bourgain, S. Dilworth, K. Ford, S. Konyagin, and D. Kutzarova. Explicit constructions of RIP matrices and related problems. *Preprint*, 2010.

[8] E. Candès, Y. Eldar, and D. Needell. Compressed sensing with coherent and redundant dictionaries. *Preprint*, 2010.

[9] E. J. Candès, J., T. Tao, and J. Romberg. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

[10] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.

[11] E. J. Candès and T. Tao. Near optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.

[12] A. Dasgupta, R. Kumar, and T. Sarlos. A sparse Johnson-Lindenstrauss transform. *STOC*, page 341350, 2010.

[13] S. Dasgupta and A. Gupta. An elementary proof of a theorem of Johnson and Lindenstrauss. *Random Structures and Algorithms*, 22:60–65, 2003.

[14] R. DeVore. Deterministic constructions of compressed sensing matrices. *J. Complexity*, 23:918–925, 2007.

[15] D. L. Donoho. For most large underdetermined systems of linear equations the minimal l^1 solution is also the sparsest solution. *Commun. Pure Appl. Anal.*, 59(6):797–829, 2006.

[16] Edo Liberty, Franco Woolfe, Per-Gunnar Martinsson, Vladimir Rokhlin, and Mark Tygert. Randomized algorithms for the low-rank approximation of matrices. *Proceedings of the National Academy of Sciences*, 104(51):20167–20172, 2007.

[17] S. Foucart. A note on guaranteed sparse recovery via ℓ_1-minimization. *Appl. Comput. Harmon. Anal.*, to appear.

[18] P. Frankl and H. Maehara. The Johnson-Lindenstrauss Lemma and the sphericity of some graphs. *Journal of Combinatorial Theory B*, 44:355–362, 1988.

[19] A. Garnaev and E. Gluskin. The widths of euclidean balls. *Doklady An. SSSR.*, 277:1048–1052, 1984.

[20] N. Halko, P. Martinsson, and J. Tropp. Finding structure with randomness: Stochastic algorithms for constructing approximate matrix decompositions. *Preprint*, 2009.

[21] W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963.

[22] P. Indyk. Algorithmic applications of low-distortion embeddings. *Proc. 42nd IEEE Symposium on Foundations of Computer Science*, 2001.

[23] M. Iwen. Simple deterministically constructible RIP matrices with sublinear Fourier sampling requirements. *Information Sciences and Systems, 2009. CISS 2009. 43rd Annual Conference on*, pages 870–875, 2009.

[24] W. B. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. *Contemp. Math*, 26:189–206, 1984.

[25] Justin Romberg. Compressive Sampling by Random Convolution. *SIAM Journal on Imaging Science*, 2009.

[26] D. Kane and J. Nelson. A derandomized sparse Johnson-Lindenstrauss transform. *Preprint*, 2010.

[27] H. Rauhut. Circulant and Toeplitz matrices in compressed sensing. In *Proc. SPARS'09*, Saint-Malo, France, 2009.

[28] H. Rauhut. Compressive sensing and structured random matrices. In M. Fornasier, editor, *Theoretical Foundations and Numerical Methods for Sparse Recovery*, volume 9 of *Radon Series Comp. Appl. Math.*, pages 1–92. deGruyter, 2010.

[29] H. Rauhut, J. Romberg, and J. Tropp. Restricted isometries for partial random circulant matrices. *In preparation*, 2010.

[30] H. Rauhut and R. Ward. Sparse Legendre expansions via ℓ_1 minimization. *Preprint*, 2010.

[31] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, 61:1025–1045, 2008.

[32] T. Sarlos. Improved approximation algorithms for large matrices via random projections, 2006. Unpublished.

[33] Simon Foucart, Alain Pajor, Holger Rauhut, and Tino Ullrich. The Gelfand widths of ℓ_p-balls for $0 < p \leq 1$. *preprint*, 2010.

[34] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126(3):505–563, 1996.

[35] Thong Do, Lu Gan, Yi Chen, Nam Nguyen, and Trac Tran. Fast and efficient dimensionality reduction using structurally random matrices. *Proc. of ICASSP*, 2009.

[36] J. Tropp, J. Laska, M. Duarte, J. Romberg, and R. G. Baraniuk. Beyond Nyquist: Efficient sampling of sparse, bandlimited signals. *IEEE Trans. Inform. Theory*, 56(1):520–544, 2010.

[37] J. Vybíral. A variant of the Johnson-Lindenstrauss lemma for circulant matrices. *Preprint*, 2010.

[38] R. Ward. Compressed sensing with cross validation. *IEEE Trans. Inform. Theory*, 55:5773–5782, 2009.