

Johnson-Lindenstrauss Lemma and Restricted Isometry Property

Grigor Keropyan

GRIGORKEROPYAN@GMAIL.COM

Department of Mathematics

Seminar of Mathematics in Data Science (Prof. Dr. Massimo Fornasier, PD Dr. Peter Massopust)

Technische Universität München

München, Germany

Abstract

This report aims to explore connection between Johnson-Lindenstrauss Lemma and Restricted Isometry Property based on the existing research in that area. The key connection point of both of them is concentration inequality. On the one hand Johnson-Lindenstrauss Lemma implies Restricted Isometry Property through concentration inequality and the latter implies the former that is optimal up to a logarithmic factors.

Keywords: Johnson-Lindenstrauss Lemma, Restricted Isometry Property, Concentration Inequalities

1. Introduction

In this report we are exploring the connection between Johnson-Lindenstrauss (JL) Lemma and Restricted Isometry Property (RIP).

JL lemma is about mapping from a high dimensional space onto lower dimensional one while keeping the distances between points almost the same. For more precise definition please refer to the section 2 and 3. JL lemma provides some kind of dimensionality reduction and the property that keeping the relative distances the same is a desired property as one can refer to the mapped points almost the same as in higher dimension. Moreover, it turns out that constructing such kind of embeddings is tractable in practice and lately even more efficient constructions have been explored (Krahmer and Ward, 2010).

On the other hand RIP provides the mapping for every point in high dimensional space onto lower one by keeping the norm almost the same, but only for the case when these points are sparse. The word sparse here means most of the entries of the vector are zero. This has various applications as in practice observing high dimensional signals they actually are in lower dimensional in a sense that they are sparse, in particular in signal processing.

Note that what follows mostly based on the research papers (Baraniuk et al., 2008) and (Krahmer and Ward, 2010).

2. Basic Notations and Definitions

Throughout the work following notations are used. If A is any, we denote $\#A$ number of points in it. Let N is a natural number in \mathbb{N} and $x = (x_1, \dots, x_N)^T \in \mathbb{R}^N$ is a column vector. The ℓ_p -norm of a vector x is denoted by

$$\|x\|_2 := \left(\sum_{j=1}^N |x_j|^2 \right)^{1/2}, \quad \forall p \in [1, \infty) \text{ and } \|x\|_\infty := \max_{j=1, \dots, N} |x_j|.$$

For $m, N \in \mathbb{N}$ and a matrix $\Phi = (\Phi_{j,l}) \in \mathbb{R}^{m \times N}$, its operator norm is defined as

$$\|\Phi\| := \sup_{\|x\|_2=1} \|\Phi(x)\|_2$$

and the Frobenius norm as

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

Matrix Φ is called **random** if entries are random samples from certain distribution which will be specified. We say a vector $x \in \mathbb{R}^N$ is **k -sparse** if it has at most k non-zero entries. Furthermore, given a set of indices T with $\#T \leq k$, denote by X_T the set of all k -sparse vectors in \mathbb{R}^N that have zero entries outside of T . Evidently, the definition of X_T implies that it is k -dimensional subspace of \mathbb{R}^N . Given a index set T and matrix Φ we denote by $\Phi_{(T)}$ matrix consisting of the columns in Φ which are in index set T . For $j \in [N]$ Φ_j denotes the j th column of the matrix and the Φ^* is a transpose of the matrix. Upper integer part of real number x is denoted by $\lceil x \rceil$, which is the smallest integer number that is not less than x . Suppose $N \in \mathbb{N}$ and s is much less than N , denote $R := \lceil N/s \rceil$. Decomposition of vector $x = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N$ into blocks is denoted by $(x_{(1)}, x_{(2)}, \dots, x_{(R)})$, where blocks have size s i.e., $x_{(j)} \in \mathbb{R}^s \forall j \in [R-1]$, only the last block can have different dimension in order that sum of dimensions be N . In order to separate only first block from the others $x = (x_{(1)}, x_{(-1)})$ is denoted, where $x_{(-1)} = (x_{(2)}, \dots, x_{(R)}) \in \mathbb{R}^{N-s}$. For the set of indices in J th block we write $[Jth]$ and for $j, l \in [N]$ we write $j \sim l$ if they are in the same block and $j \not\sim l$ otherwise. For constructing diagonal matrix which has elements as given vector $x \in \mathbb{R}^N$ is denoted by $D_x \in \mathbb{R}^{N \times N}$, where $D_{j,j} = x_j$ and $D_{i,j} = 0$ for all $i, j \in [N]$ and $i \neq j$. For the vector $x = (x_1, x_2, \dots, x_N) \in \mathbb{R}^N$ we say x is in **decreasing arrangement** if $|x_i| \geq |x_j|$ for all $i < j$ and $i, j \in [N]$.

Definition 1

We call $\xi \in \mathbb{R}^N$ as a Rademacher sequence if it is uniformly distributed on $\{-1, +1\}^N$, where $N \in \mathbb{N}$.

Definition 2

Let $\Phi \in \mathbb{R}^{n \times N}$ be a random matrix and for any $x \in \mathbb{R}^N$ expectation of random variable $\|\Phi(x)\|_2^2$ is $\|x\|_2^2$, that is

$$\mathbb{E}[\|\Phi(x)\|_2^2] = \|x\|_2^2,$$

where $n, N \in \mathbb{N}$. We say that random matrix Φ satisfies concentration inequality if the random variable $\|\Phi(x)\|_2^2$ is strongly concentrated about its expected value, that is

$$\mathbb{P}(|\|\Phi(x)\|_2^2 - \|x\|_2^2| \geq \epsilon \|x\|_2^2) \leq 2e^{-nc(\epsilon)}, \quad 0 \leq \epsilon \leq 1,$$

where ϵ is a constant and $c(\epsilon)$ is also a constant depends only on ϵ such that for all $\epsilon \in (0, 1)$, $c(\epsilon) \geq 0$.

Definition 3

We say that a matrix Φ satisfies *Restricted Isometry Property (RIP)* of order k and level $\delta \in (0, 1)$ (concisely, (k, δ) – RIP) if

$$(1 - \delta) \|x\|_2^2 \leq \|\Phi(x)\|_2^2 \leq (1 + \delta) \|x\|_2^2 \quad \forall k\text{-sparse } x \in \mathbb{R}^N. \quad (1)$$

The *restricted isometry constant* δ_k is defined as the smallest value of δ for which (1) holds.

3. Connection Between JL Lemma and RIP

Before stating the Johnson-Lindenstrauss (JL) Lemma and its connection to RIP, let us understand what is the main idea of JL Lemma. Suppose a set of point cloud \mathcal{Q} in \mathbb{R}^N is given with N is large. So, working with this point cloud could be challenging due to curse of dimensionality. In order to address this problem many techniques of dimensionality reduction have been discovered so far i.e., Principal Component Analysis (PCA), Autoencoder in deep learning, Non-negative Matrix Factorization (NMF) etc. It is also clear that all of the dimensionality reduction methods should through away some information about data. That is why addressing the problem can be done keeping the information that is needed. The idea of JL lemma is to embed this points into a lower-dimensional Euclidean space \mathbb{R}^n in a way that relative distances between points are almost preserved and how small n can be relative to $\#\mathcal{Q}$. The following is the original formulation in (Johnson and Lindenstrauss, 1982).

Lemma 4 (Johnson-Lindenstrauss) (Lemma 4.1 in (Baraniuk et al., 2008))

Let $\epsilon \in (0, 1)$ is given and $N \in \mathbb{N}$. For every set $\mathcal{Q} \in \mathbb{R}^N$, if n is a positive integer such that $n \geq n_0 := O(\ln(\#\mathcal{Q})/\epsilon^2)$, there exists a Lipschitz mapping $f : \mathbb{R}^N \rightarrow \mathbb{R}^n$ such that

$$(1 - \epsilon) \|u - v\|_2^2 \leq \|f(u) - f(v)\|_2^2 \leq (1 + \epsilon) \|u - v\|_2^2, \quad (2)$$

for all $u, v \in \mathcal{Q}$.

During the last three decades various forms and easier proofs have been explored for this lemma (Achlioptas, 2001; Dasgupta and Gupta, 1999; Frankl and Maehara, 1988). Particularly, in order to be able to use the lemma in practice researchers explored various methods of constructing such mappings. It turns out that mapping is possible to choose linear, which is already beneficial as it is easy and requires low complexity to work with linear functions. For the linear mapping case function f can be represented by $n \times N$ matrix Φ . Using the linearity of the mapping the equation 2 becomes

$$(1 - \epsilon) \|u - v\|_2^2 \leq \|\Phi(u - v)\|_2^2 \leq (1 + \epsilon) \|u - v\|_2^2, \quad \forall u, v \in \mathcal{Q}.$$

Now, if we denote all differences of the points in the set \mathcal{Q} by $D := \{u - v\}$ for all $u, v \in \mathcal{Q}$ such that $u \neq v$ the above equation is equivalent to

$$(1 - \epsilon) \|y\|_2^2 \leq \|\Phi(y)\|_2^2 \leq (1 + \epsilon) \|y\|_2^2, \quad \forall y \in D. \quad (3)$$

In such a setting finding a deterministic matrix could be challenging. However, random matrices provide tractable way to face this problem. Indeed, when Φ is a random matrix, the equation 3 holds with high probability if the matrix satisfies concentration inequality, that is

$$\mathbb{P}(|\|\Phi(x)\|_2^2 - \|x\|_2^2| \geq \epsilon \|x\|_2^2) \leq 2e^{-nc(\epsilon)}, \quad 0 \leq \epsilon \leq 1.$$

From the above discussion we understand that the key point for proving JL Lemma is to find a random matrix which satisfies concentration inequality. Later in this section is showed how RIP and JL Lemma are connected through this inequality, until then it is worth to mention few examples satisfying concentration inequality that are well known in the literature.

1. Entries $\Phi_{i,j}$ of Φ are independent samples from Normal distribution

$$\{\Phi_{i,j}\}_{i \in [n], j \in [N]} \stackrel{i.i.d}{\sim} \mathcal{N}(0, \frac{1}{n}).$$

Here the samples from normal distribution might be very large and working with them in practice arise a difficulties. The next example is bounded and do not have such a problem.

2. Entries $\Phi_{i,j}$ of Φ are independent samples from following distribution

$$X := \begin{cases} +1/\sqrt{n} & \text{with probability } 1/2, \\ -1/\sqrt{n} & \text{with probability } 1/2. \end{cases}$$

Here samples are either $+1/\sqrt{n}$ or $-1/\sqrt{n}$ and they are bounded, however, in practice we prefer to have sparse matrices that working with them will be fast. The next example is a nice result for sparse case.

3. Entries $\Phi_{i,j}$ of Φ are independent samples from following distribution

$$X := \begin{cases} +\sqrt{3}/\sqrt{n} & \text{with probability } 1/6, \\ 0 & \text{with probability } 2/3, \\ -\sqrt{3}/\sqrt{n} & \text{with probability } 1/6. \end{cases}$$

In all the above examples the constant $c(\epsilon)$ can be chosen as $\epsilon^2/4 - \epsilon^3/6$.

In order to derive the first connection of JL Lemma and RIP let's restate the results and replicate the proves in (Baraniuk et al., 2008). As a reminder X_T is a linear subspace in \mathbb{R}^N such that vectors have zero entries outside of index set T .

Lemma 5 (Lemma 5.1 in (Baraniuk et al., 2008))

Let $n, N \in \mathbb{N}$ and $\Phi \in \mathbb{R}^{n \times N}$ be a random matrix which satisfies concentration inequality (see Definition 2). Then, for an arbitrary set T such that $k := \#T < n$ and $\forall \delta \in (0, 1)$, we have

$$(1 - \delta) \|x\|_2 \leq \|\Phi(x)\|_2 \leq (1 + \delta) \|x\|_2, \quad \forall x \in X_T, \quad (4)$$

with probability

$$\geq 1 - 2(12/\delta)^k e^{-c(\delta/2)^n}. \quad (5)$$

Proof. First we note that in case x is zero the equation 4 is trivially satisfied. In other cases we have it is equivalent to

$$(1 - \delta) \leq \|\Phi(x/\|x\|_2)\|_2 \leq (1 + \delta), \quad \forall x \in X_T,$$

where the linearity of norm and Φ is used. Thus, equation 4 is equivalent to

$$(1 - \delta) \leq \|\Phi(x)\|_2 \leq (1 + \delta), \quad \forall x \in X_T, \|x\|_2 = 1. \quad (6)$$

As we are in a finite dimension, the unit ball of X_T is possible to cover having only finite number of balls of radius no more than $\delta/4$. The same logic works for covering the unit sphere of X_T by balls which have radius no more than $\delta/4$ and their centers are in X_T . So, there is finite set of points $\mathcal{Q}_T \subseteq X_T$ such that $\|q\|_2 = 1$, $\forall q \in \mathcal{Q}_T$ and $\forall x \in X_T$ with $\|x\|_2 = 1$ we have

$$\min_{q \in \mathcal{Q}_T} \|x - q\|_2 \leq \delta/4. \quad (7)$$

Moreover, it is known that such points is possible to choose such that $\#\mathcal{Q}_T \leq (12/\delta)^k$. Now, using the concentration inequality for the case of $\epsilon = \delta/2$ and the union bound of probability we obtain

$$(1 - \delta/2) \|q\|_2^2 \leq \|\Phi(q)\|_2^2 \leq (1 + \delta/2) \|q\|_2^2, \quad \forall q \in \mathcal{Q}_T,$$

with probability

$$\geq 1 - 2(12/\delta)^k e^{-c(\delta/2)^n},$$

which is the same as

$$(1 - \delta/2) \|q\|_2 \leq \|\Phi(q)\|_2 \leq (1 + \delta/2) \|q\|_2, \quad \forall q \in \mathcal{Q}_T,$$

with probability

$$\geq 1 - 2(12/\delta)^k e^{-c(\delta/2)^n}.$$

Let's denote

$$A := \operatorname{argmin}_a \{a \in \mathbb{R} : \|\Phi(x)\|_2 \leq (1 + a), \quad \forall x \in X_T, \|x\|_2 = 1\}$$

From (7) we have for every above x there is $q \in \mathcal{Q}_T$ with $\|q\|_2 = 1$ and $\|x - q\|_2 \leq \delta/4$. Using this we obtain

$$\|\Phi(x)\|_2 \leq \|\Phi(q)\|_2 + \|\Phi(x - q)\|_2 \leq 1 + \delta/2 + (1 + A)\delta/4,$$

where we used the fact that $\|\Phi(q)\|_2 \leq 1 + \delta/2$ from above and $\|\Phi(x - q)\|_2 \leq (1 + A)\|x - q\|_2 \leq (1 + A)\delta/4$ by definition of A . Having this and using the definition of A as being the smallest we have $A \leq \delta/2 + (1 + A)\delta/4 \iff A(1 - \delta/4) \leq 3\delta/4 \iff A \leq 3\delta/4(1 - \delta/4) = 3\delta/(4 - \delta) \leq \delta$. We obtained that $A \leq \delta$ which means right hand side of the (6) is satisfied. For the left hand side we have

$$\|\Phi(x)\|_2 \geq \|\Phi(q)\|_2 - \|\Phi(x - q)\|_2 \geq 1 - \delta/2 - (1 + \delta)\delta/4 \geq 1 - \delta/2 - \delta/4 - \delta^2/4 \geq 1 - \delta,$$

where we have used the fact that $\|\Phi(q)\|_2 \geq 1 - \delta/2$ and $A \leq \delta$ from above. Therefore, equation 6 is established with probability $\geq 1 - 2(12/\delta)^k e^{-c(\delta/2)n}$ which completes the proof of the lemma. ■

The following Theorem shows that JL Lemma implies Restricted Isometry Property in a sense that if concentration inequality holds (which means JL Lemma holds) then RIP holds. For the Definition of RIP see section 2.

Theorem 6 (Theorem 5.2 in (Baraniuk et al., 2008))

Let $n, N \in \mathbb{N}, \delta \in (0, 1)$ and $\Phi \in \mathbb{R}^{n \times N}$ be a random matrix which satisfies concentration inequality (see Definition 2), then $\exists c_1, c_2 > 0$ depending only on δ such that Φ satisfies Restricted Isometry Property of order $k \leq c_1 n / \log(N/k)$ and level δ with probability $\geq 1 - 2e^{-c_2 n}$.

Proof. For each index set T with $\#T = k$ Lemma 5 proves that the matrix Φ fails to satisfy (4) for the space X_T with probability

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

The number of such spaces is $\binom{N}{k}$. Using the union bound we obtain that (4) fails with probability

$$\leq \binom{N}{k} 2(12/\delta)^k e^{-c(\delta/2)n} \leq (eN/k)^k 2(12/\delta)^k e^{-c(\delta/2)n} = 2e^{-c(\delta/2)n + k[\log(eN/k) + \log(12/\delta)]} \quad (8)$$

where $\binom{N}{k} \leq (eN/k)^k$ inequality is used. Now, let's fix $c_1 > 0$ and taking $c_2 \leq c(\delta/2) - c_1[1 + (1 + \log(12/\delta))/\log(N/k)]$ gives the exponent of the right side in (8)

$$\begin{aligned} & -c(\delta/2)n + k[\log(eN/k) + \log(12/\delta)] \leq \\ & -c(\delta/2)n + c_1 n / \log(N/k) [\log(eN/k) + \log(12/\delta)] = \\ & n[-c(\delta/2) + c_1(1 + (1 + \log(12/\delta))/\log(N/k))] \leq -c_2 n, \end{aligned}$$

where $k \leq c_1 n / \log(N/k)$ is used. Taking c_1 sufficiently small ensures that c_2 is positive. This proves that matrix Φ satisfies the following inequalities for every k -sparse vector x , as spaces X_T are the ones that characterize all the sparse vectors

$$(1 - \delta) \|x\|_2 \leq \|\Phi(x)\|_2 \leq (1 + \delta) \|x\|_2,$$

with probability

$$\geq 1 - 2e^{-c_2 n},$$

This concludes the proof of the Theorem. ■

Having the one implication it is interesting to think about the other direction. However, matrices that satisfy RIP are not random; so, in order to derive converse result the RIP matrix should be randomized in a natural way. In fact, the reverse implication holds by specific randomization of RIP matrices. Before stating the main reverse result some preliminary Propositions are necessary.

The following Proposition is well known Hoeffding's inequality which is a tail bound.

Proposition 7 (Hoeffding's Inequality) (Proposition 5.1 in (Krahmer and Ward, 2010))
Let $N \in \mathbb{N}$ and $x \in \mathbb{R}^N$. Taking $\xi = (\xi_1, \xi_2, \dots, \xi_N)$ as a Rademacher sequence, then, for any $t > 0$ we have

$$\mathbb{P}(|\sum_{j=1}^N \xi_j x_j| > t) \leq 2e^{-\frac{t^2}{2\|x\|_2^2}}. \quad (9)$$

Proposition 8 (Proposition 5.2 in (Krahmer and Ward, 2010))
Let $N \in \mathbb{N}$ and $X \in \mathbb{R}^N$ is a matrix such that $X_{i,i} = 0$ for all $i \in [N]$. Taking $\xi = (\xi_1, \xi_2, \dots, \xi_N)$ as a Rademacher sequence, then, for any $t > 0$ we have

$$\mathbb{P}(|\sum_{i,j=1}^N \xi_i \xi_j X_{i,j}| > t) \leq 2e^{-\frac{1}{64} \min(\frac{96t/65}{\|X\|}, \frac{t^2}{\|X\|_{\mathcal{F}}^2})} \quad (10)$$

The following is an estimate for RIP matrices.

Proposition 9 (Proposition 5.3 in (Krahmer and Ward, 2010))
Let $n, N \in \mathbb{N}$ and the matrix $\Phi \in \mathbb{R}^{n \times N}$ satisfies RIP of order $2s$ and level δ , where $\delta \in (0, 1)$ and $2s \leq N$. Then for any two disjoint subsets $J, L \subset [N]$ of size $\#J, \#L \leq s$ we have

$$\left\| |\Phi_{(J)}^* \Phi_{(L)}| \right\| \leq \delta. \quad (11)$$

Proposition 10 (Proposition 5.4 in (Krahmer and Ward, 2010))
Let $n, N \in \mathbb{N}$ and $s \leq N/2$, $R = \lceil N/s \rceil$. Suppose $\Phi = (\Phi_{(1)}, \Phi_{(2)}, \dots, \Phi_{(R)}) = (\Phi_{(1)}, \Phi_{(-1)}) \in \mathbb{R}^{n \times N}$ satisfies $(2s, \delta)$ -RIP and $x = (x_{(1)}, x_{(2)}, \dots, x_{(R)}) = (x_{(1)}, x_{(-1)}) \in \mathbb{R}^N$ be in a unit ball (i.e., $\|x\|_2 \leq 1$) and in decreasing arrangement. Consider a symmetric matrix

$$C \in \mathbb{R}^{N \times N}, \quad C_{i,j} = \begin{cases} x_j \Phi_j^* \Phi_l x_l & \text{if } j \not\sim l, j, l > s, \\ 0 & \text{otherwise} \end{cases}$$

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

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

Then, the following inequalities hold

$$\|C\| \leq \frac{\delta}{s}, \quad \|C\|_{\mathcal{F}} \leq \frac{\delta}{\sqrt{s}}, \quad \|v\|_2 \leq \frac{\delta}{\sqrt{s}}.$$

For the proofs of the above Propositions please refer to (Krahmer and Ward, 2010).

Having all the above mentioned preliminary results enables us to restate and replicate the proof of the second main results from (Krahmer and Ward, 2010). This time results

are another direction: Restricted Isometry Property implies concentration inequality which implies JL Lemma. So, RIP implies Johnson-Lindenstrauss Lemma.

Theorem 11 (Theorem 3.1 in (Krahmer and Ward, 2010))

Let $\eta > 0, \epsilon \in (0, 1), n, N \in \mathbb{N}$ be given and the set $E \subset \mathbb{R}^N$ has cardinality $p := \#E$. Suppose the matrix $\Phi \in \mathbb{R}^{n \times N}$ satisfies Restricted Isometry Property of order $k \geq 40 \log \frac{4p}{\eta}$ and level $\delta \leq \frac{\epsilon}{4}$. Taking $\xi \in \mathbb{R}^N$ as a Rademacher sequence we have

$$(1 - \epsilon) \|x\|_2^2 \leq \|\Phi D_\xi x\|_2^2 \leq (1 + \epsilon) \|x\|_2^2, \quad (12)$$

for all $x \in E$, with probability

$$\geq 1 - \eta.$$

Proof. Since (12) is equivalent to

$$(1 - \epsilon) \leq \|\Phi D_\xi x / \|x\|_2\|_2^2 \leq (1 + \epsilon), \quad (13)$$

without loss of generality we assume that $\|x\|_2 = 1$ and the above equation becomes

$$(1 - \epsilon) \leq \|\Phi D_\xi x\|_2^2 \leq (1 + \epsilon). \quad (14)$$

First, notice that rearranging the columns of Φ and doing the appropriate permutation in x does not affect the result as RIP of the matrix Φ is invariant under permutation. So, we can assume that every $x \in E$ has norm 1 and is in decreasing arrangement. Moreover, simple matrix vector multiplication shows that $D_\xi x = D_x \xi$. Now, let us fix $x \in E$ with $\|x\|_2 = 1$ and further assume that $k = 2s$ is even. Then, the quantity that we want to estimate becomes

$$\begin{aligned} \|\Phi D_\xi x\|_2^2 &= \|\Phi D_x \xi\|_2^2 \\ &= \sum_{J=1}^R \left\| \Phi_{(J)} D_{x_{(J)}} \xi_{(J)} \right\|_2^2 \\ &\quad + 2\xi_{(1)}^* D_{x_{(1)}} \Phi_{(1)}^* \Phi_{(-1)} D_{x_{(-1)}} \xi_{(-1)} \\ &\quad + \sum_{J,L=2, J \neq L}^R \langle \Phi_{(J)} D_{x_{(J)}} \xi_{(J)}, \Phi_{(L)} D_{x_{(L)}} \xi_{(L)} \rangle. \end{aligned} \quad (15)$$

In the following we will get bounds for each of the three terms in (15).

1. $\sum_{J=1}^R \left\| \Phi_{(J)} D_{x_{(J)}} \xi_{(J)} \right\|_2^2$: Since having a RIP of order $k \geq s$ and level δ means also having a RIP of order s and level δ and each $\Phi_{(J)}$ for all $J = 1, 2, \dots, R$ is almost an isometry in a sense of (1). Keeping in mind that elements of ξ are from Rademacher sequence i.e., +1 or -1, we have $\|D_{x_{(J)}} \xi_{(J)}\|_2 = \|x_{(J)}\|_2$ and using the definition of RIP of order s and level δ we obtain

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

Using the assumption of the Theorem that $\delta \leq \epsilon/4$, we have

$$(1 - \epsilon/4) \|x\|_2^2 \leq \sum_{J=1}^R \left\| \Phi_{(J)} D_{x_{(J)}} \xi_{(J)} \right\|_2^2 \leq (1 + \epsilon/4) \|x\|_2^2. \quad (16)$$

2. $2\xi_{(1)}^* D_{x_{(1)}} \Phi_{(1)}^* \Phi_{(-1)} D_{x_{(-1)}} \xi_{(-1)}$: For every fixed $b := \xi_{(1)}$ consider a random variable

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

where $v := D_{x_{(-1)}} \Phi_{(-1)}^* \Phi_{(1)} D_{x_{(1)}} b$ as in Proposition 10. Using the Hoeffding's inequality (Proposition 7) and the Proposition 10 we obtain

$$\mathbb{P}(|X| \geq \gamma\epsilon) \leq 2e^{-\frac{\gamma^2\epsilon^2}{2\|v\|_2^2}} \leq 2e^{-\frac{\gamma^2\epsilon^2}{2\delta^2/s}} = 2e^{-\frac{s\gamma^2\epsilon^2}{2\delta^2}} \quad (17)$$

Each X depends on the choice of $x \in E$. Using a union bound and the a little bit abuse of notation we obtain

$$\mathbb{P}(\exists x \in E : |X| \geq \gamma\epsilon) \leq 2pe^{-\frac{s\gamma^2\epsilon^2}{2\delta^2}} = e^{\log 2p - \frac{s\gamma^2\epsilon^2}{2\delta^2}}.$$

Let us choose γ such that

$$\delta \leq \frac{\epsilon}{4} \sqrt{\frac{8\gamma^2 s}{\log(4p/\eta)}},$$

which holds for the case of $\gamma = 0.1$ and from the Theorem assumption we had $\delta \leq \epsilon/4$ and $s = k/2 \geq 20 \log(4p/\eta)$. This shows that the probability in (17) is no more than $\eta/2$. Putting everything together we obtain

$$\mathbb{P}(\exists x \in E : |2X| \geq 0.2\epsilon) \leq \frac{\eta}{2} \quad (18)$$

3. $\sum_{J,L=2, J \neq L}^R \langle \Phi_{(J)} D_{x_{(J)}} \xi_{(J)}, \Phi_{(L)} D_{x_{(L)}} \xi_{(L)} \rangle$: The aim here is to use Proposition 8 and Proposition 10 to obtain bounds. Denoting

$$C \in \mathbb{R}^{N \times N}, \quad C_{i,j} = \begin{cases} x_j \Phi_j^* \Phi_l x_l & \text{if } j \not\sim l, j, l > s, \\ 0 & \text{otherwise} \end{cases}$$

as in Proposition 10, we get

$$\sum_{J,L=2, J \neq L}^R \langle \Phi_{(J)} D_{x_{(J)}} \xi_{(J)}, \Phi_{(L)} D_{x_{(L)}} \xi_{(L)} \rangle = \langle \xi, C\xi \rangle = \sum_{j,l=s+1}^N \xi_j \xi_l C_{jl}.$$

Note, this equality holds since in matrix C entries are zero where ξ_j and ξ_l does not appear together. From Proposition 10 we have $\|C\| \leq \frac{\delta}{s}$ and $\|C\|_{\mathcal{F}} \leq \frac{\delta}{\sqrt{s}}$. Fixing $\tau > 0$ and applying above inequalities in the Proposition 8 we obtain

$$\mathbb{P}\left(\left| \sum_{j,l=s+1}^N \xi_j \xi_l C_{jl} \right| \geq \tau\epsilon\right) \leq 2e^{-\frac{1}{64} \min\left(\frac{96\tau\epsilon/65}{\|C\|}, \frac{(\tau\epsilon)^2}{\|C\|_{\mathcal{F}}^2}\right)} \leq 2e^{-\frac{1}{64} \min\left(\frac{96\tau\epsilon s}{65\delta}, \frac{\tau^2\epsilon^2 s}{\delta^2}\right)}$$

Using the union bound, we get

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

Let us take τ such that

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

which is true if we take $\tau = 0.55$ and using the assumptions of the Theorem that $\delta \leq \epsilon/4$ and $s = k/2 \geq 20 \log(4p/\eta)$. Above inequality is equivalent to

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

Putting everything together we obtain

$$\mathbb{P}(\exists x \in E : \left| \sum_{j,l=s+1}^N \xi_j \xi_l C_{jl} \right| \geq \tau \epsilon) \leq \eta/2 \quad (19)$$

So,

$$\mathbb{P}(\exists x \in E : \left| \sum_{J,L=2, J \neq L}^R \langle \Phi_{(J)} D_{x_{(J)}} \xi_{(J)}, \Phi_{(L)} D_{x_{(L)}} \xi_{(L)} \rangle \right| \geq 0.55\epsilon) \leq \eta/2 \quad (20)$$

Now, using (16), (18), (20) and (15) and remembering that we can assume that norm of x is 1, we obtain

$$\begin{aligned} & \mathbb{P}(\exists x \in E : \left| \|\Phi D_{\xi} x\|_2^2 / \|x\|_2^2 - 1 \right| \geq \epsilon) \\ & \leq \mathbb{P}(\exists x \in E : \left| \sum_{J=1}^R \left\| \Phi_{(J)} D_{x_{(J)}} \xi_{(J)} \right\|_2^2 \right. \\ & \quad \left. + 2\xi_{(1)}^* D_{x_{(1)}} \Phi_{(1)}^* \Phi_{(-1)} D_{x_{(-1)}} \xi_{(-1)} \right. \\ & \quad \left. + \sum_{J,L=2, J \neq L}^R \langle \Phi_{(J)} D_{x_{(J)}} \xi_{(J)}, \Phi_{(L)} D_{x_{(L)}} \xi_{(L)} \rangle \right| \geq \epsilon) \\ & \leq \mathbb{P}(\exists x \in E : \left| \sum_{J=1}^R \left\| \Phi_{(J)} D_{x_{(J)}} \xi_{(J)} \right\|_2^2 - 1 \right| \geq 0.25\epsilon) \\ & \quad + \mathbb{P}(\exists x \in E : |2\xi_{(1)}^* D_{x_{(1)}} \Phi_{(1)}^* \Phi_{(-1)} D_{x_{(-1)}} \xi_{(-1)}| \geq 0.2\epsilon) \\ & \quad + \mathbb{P}(\exists x \in E : \left| \sum_{J,L=2, J \neq L}^R \langle \Phi_{(J)} D_{x_{(J)}} \xi_{(J)}, \Phi_{(L)} D_{x_{(L)}} \xi_{(L)} \rangle \right| \geq 0.55\epsilon) \\ & \leq 0 + \eta/2 + \eta/2 = \eta. \end{aligned}$$

This concludes the proof of the Theorem.

■

The above methods also provides a way to prove the direct inverse of the Theorem 6 which is explored in the next Proposition. This means If we randomize the RIP matrix it becomes equivalent to JL Lemma which is a remarkable result as they are two different properties while being equivalent.

Proposition 12 (Proposition 3.2 in (Krahmer and Ward, 2010))

Let $n, N \in \mathbb{N}$ and $\delta > 0, \epsilon \in (0, 1)$ is fixed. Suppose that there is a constant c_1 such that for all pairs (k, n) that satisfy $k \leq c_1 \delta^2 n / \log(N/k)$, $\Phi = \Phi(n) \in \mathbb{R}^{n \times N}$ has RIP of order k and level $\delta \leq \epsilon/4$. Fixing $x \in \mathbb{R}^N$ and letting $\xi \in \mathbb{R}^N$ be a Rademacher sequence, then, there exist a constant c_2 such that for all n , ΦD_ξ satisfies the concentration inequality for $c(\epsilon) = c_2 \epsilon^2 \log^{-1} \frac{N}{k}$, for any k that satisfies the above inequality.

Proof. We can assume that $k \geq \frac{1}{2} c_1 \delta^2 / \log(N/k)$, otherwise we just increase k until it obtains the above bound. Using the inequalities in the proof of Theorem 11 from points 1, 2 and 3 and remembering that $\tau = 0.55$ and $\gamma = 0.1$ we obtain

$$\mathbb{P}(|\|\Phi D_\xi x\|_2^2 - \|x\|_2^2| \geq \epsilon \|x\|_2^2) \leq 2e^{-c_3 k} \leq 2e^{-c_2 \epsilon^2 n \log^{-1}(\frac{N}{k})} = 2e^{-nc(\epsilon)},$$

where $c_2 = c_3 c_1 / 32$.

■

References

- Dimitris Achlioptas. Database-friendly random projections. In *Proceedings of the Twentieth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, PODS '01, page 274–281, New York, NY, USA, 2001. Association for Computing Machinery. ISBN 1581133618. doi: 10.1145/375551.375608. URL <https://doi.org/10.1145/375551.375608>.
- Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, December 2008. ISSN 0176-4276. doi: 10.1007/s00365-007-9003-x.
- Sanjoy Dasgupta and Anupam Gupta. An elementary proof of the johnson-lindenstrauss lemma. Technical report, U.C. Berkeley, 1999.
- P Frankl and H Maehara. The johnson-lindenstrauss lemma and the sphericity of some graphs. *Journal of Combinatorial Theory, Series B*, 44(3):355–362, 1988. ISSN 0095-8956. doi: [https://doi.org/10.1016/0095-8956\(88\)90043-3](https://doi.org/10.1016/0095-8956(88)90043-3). URL <https://www.sciencedirect.com/science/article/pii/0095895688900433>.
- William Johnson and J. Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. *Conference in Modern Analysis and Probability*, 26:189–206, 01 1982.
- Felix Krahmer and Rachel Ward. New and improved Johnson-Lindenstrauss embeddings via the Restricted Isometry Property. *arXiv e-prints*, art. arXiv:1009.0744, September 2010.