ON THE CONNECTION BETWEEN JOHNSON-LINDENSTRAUSS TRANSFORM AND RESTRICTED ISOMETRY PROPERTY

YAO ZHU∗

Abstract. In this report, we investigate the connection between the Johnson-Lindenstrauss transform (JLT) and the restricted isometry property (RIP). JLT is an important algorithm design and analysis tool in randomized numerical linear algebra, while RIP is an important property of sensing matrices to guarantee stable and robust recovery in compressive sensing. We believe studying the connection between JLT and RIP could lead to beneficial cross-talk between these two closely related communities. Specifically, we study a theorem that establishes the conversion of a RIP matrix to a JLT using a Rademacher sequence. As a result, this theorem makes it possible to construct fast JLT through analysing the RIP of structured random matrices.

Key words. Johnson-Lindenstrauss Transform, restricted isometry property, Rademacher sequence, random matrix

1. Introduction. We study the relationship between the Johnson-Lindenstrauss Transform (JLT) and the Restricted Isometry Property (RIP). As we will see in the following, techniques and results from random matrices, concentration of measure, and geometric functional analysis are critical for both constructing JLT and establishing RIP. It is this connection that motivates the study conducted here.

Definition 1.1 (Johnson-Lindenstrauss Transform). Let δ ∈ (0, 1) and let X = {x1, . . . , xp} be a set of arbitrary points in RN. A Lipschitz map f : RN → Rm is a JLT with parameter δ on X (denoted as (p, δ)-JLT) if for all xi, xj ∈ X such that

\[(1 - δ)\|x_i - x_j\|^2_2 \leq \|f(x_i) - f(x_j)\|^2_2 \leq (1 + δ)\|x_i - x_j\|^2_2\]

The Johnson-Lindenstrauss lemma [8] states that such a map f exists if m = O(δ−2 log(p)). It is further shown in [3] that

\[m = Ω(δ^{-2} \log(p)) \frac{1}{\log(1/δ)}\]

Recently it is found the proof of Johnson-Lindenstrauss lemma can be greatly simplified using random matrices (e.g. Bernoulli and Gaussian random matrices). In fact, it can be shown that a Bernoulli random matrix or a Gaussian random matrix (suitably normalized) consists a JLT with high probability (w.h.p.) provided that m = O(δ−2 log(p)) [1, 5].

We let the set E = {x_i - x_j} for all x_i, x_j ∈ X. Given that an (p, δ)-JLT can be realized by a matrix Φ ∈ Rm×N, Definition 1.1 is equivalent to

\[(1 - δ)\|x\|^2_2 \leq \|Φx\|^2_2 \leq (1 + δ)\|x\|^2_2 \text{ for all } x ∈ E\]

Definition 1.2 (Restricted Isometry Property [10]). A matrix Φ ∈ Rm×N is said to have the restricted isometry property of order s at level δ ∈ (0, 1) (denoted as (s, δ)-RIP) if

\[(1 - δ)\|x\|^2_2 \leq \|Φx\|^2_2 \leq (1 + δ)\|x\|^2_2 \text{ for all } s\text{-sparse } x ∈ RN\]

∗Department of Computer Science, Purdue University, West Lafayette, IN 47906. (yaozhu@purdue.edu).
Notice that in Definition 1.2, we adopt a real matrix $\Phi$. In the following discussions of structured random matrices, e.g., random partial Fourier matrices, we may refer to a complex matrix $\Phi$. However, the action of a complex $m \times N$ matrix on $x \in \mathbb{R}^N$ can be identified with the action of a real $2m \times N$ matrix on $x$. So nothing essentially changes if we switch to complex matrices.

2. The relationship between JLT and RIP. It is apparent from (1.2) and (1.3), the defining inequalities of JLT and RIP have the same mathematical form. The only difference lies in the quantification over $x$: For JLT, $x$ is confined to a finite set $E$ of cardinality $(\binom{k}{p})$, but without any restriction on the dimensionality of $x$. In contrast, for RIP, $x$ are confined to a union of $s$-dimensional subspaces, which is an infinite set. Given the superficially observable similarity and dissimilarity between JLT and RIP, it is interesting to seek whether there is any deeper connection between JLT and RIP.

Let $\Phi$ be a $m \times N$ random matrix sampled from an ensemble $\mathcal{M}$. For a fixed $x \in \mathbb{R}^N$, if the ensemble $\mathcal{M}$ satisfies the following concentration inequality

\begin{equation}
\mathbb{P}(\|\Phi x\|_2^2 - \|x\|_2^2 \geq \delta \|x\|_2^2) \leq 2 \exp(-\tilde{c}\delta^2 m)
\end{equation}

where $\tilde{c}$ is a constant, then it can be proved that w.h.p. $\Phi$ is a $(p, \delta)$-JLT through union bound on $E$, see Lemma 9.35 of [6]. It is first explicitly pointed out in [4] that the same concentration inequality (2.1) also implies w.h.p. the $(s, \delta)$-RIP of $\Phi$. It is in this sense that we can treat it as JLT implies RIP. It is worthy to mention that such a connection is essentially induced by the fact that any $s$-dimensional unit $\ell_2$-ball $B_s, S \subset \{1, \ldots, N\}$ admits a $\rho$-net with cardinality bounded by $\left(1 + \frac{2}{\rho}\right)^s$.

For the reverse direction, the following theorem shows that RIP implies JLT.

**Theorem 2.1 (Theorem 3.1 of [10]).** Fix $\epsilon, \delta \in (0, 1)$, and consider a finite set $E \subset \mathbb{R}^N$ of cardinality $|E| = p$. Set $s \geq 32 \log \frac{4p}{\delta}$, and suppose that $\Phi \in \mathbb{R}^{m \times N}$ satisfies $(s, \eta)$-RIP with $\eta \leq \frac{\delta}{4}$. Let $\xi \in \{-1, 1\}^N$ be a Rademacher sequence, and let $D_\xi = \text{diag}(\xi)$. Then with probability at least $1 - \epsilon$

\begin{equation}
(1 - \delta)\|x\|_2^2 \leq \|D_\xi x\|_2^2 \leq (1 + \delta)\|x\|_2^2 \quad \text{for all } x \in E
\end{equation}

**Proof.** Because the inequalities in (2.2) are invariant under scaling of $x$, we assume $\|x\|_2 = 1$ for all $x \in E$ in the proof. Since $E$ is a finite set, we can consider a fixed $x \in E$ and then take the union bound on $E$. Assuming $s = 2k$ is even, we partition $[N]$ into blocks of size $k$ according to a nonincreasing rearrangement of $x$. Let the blocks of indices be $\{S_1, S_2, \ldots, S_R\}$, where $R = \lceil \frac{N}{k} \rceil$,

\begin{align}
\|D_\xi x\|_2^2 &= \|D_\xi \sum_{j \geq 1} x_{S_j}\|_2^2 \\
&= \sum_{j \geq 1} \|D_\xi x_{S_j}\|_2^2 + 2 \langle D_\xi x_{S_j}, D_\xi x_{S_l} \rangle + \sum_{j,l \geq 2, j \neq l} \langle D_\xi x_{S_j}, D_\xi x_{S_l} \rangle
\end{align}

Because $\Phi$ satisfies $(s, \eta)$-RIP with $\eta \leq \delta/4$, and note $\|D_\xi x\|_2^2 = \|x\|_2^2$ for any $x$, we have

\begin{equation}
(1 - \delta/4)\|x_{S_j}\|_2^2 \leq \|D_\xi x_{S_j}\|_2^2 \leq (1 + \delta/4)\|x_{S_j}\|_2^2
\end{equation}
Because $\sum_{j \geq 1} \|x_{S_j}\|^2 = \|x\|^2 = 1$ we can bound the first term in (2.3) by
\[(2.4) \quad (1 - \delta/4) \leq \sum_{j \geq 1} \|\Phi D^\xi x_{S_j}\|^2 \leq (1 + \delta/4)\]

For estimating the second term in (2.3), we consider the random variable
\[X = \langle \Phi D^\xi x_{S_1}, \Phi D^\xi x_{S_2}, \ldots, \Phi D^\xi x_{S_t} \rangle = \langle \Phi D x_{S_1}, \ldots, \Phi D x_{S_t} \rangle = \sum_{i \in S_t} v_i \xi_i\]

where
\[v = D x_{S_1} e S_1 \Phi^* F \Phi D x_{S_1} \xi_{S_1}\]

Because $v$ and $\xi_{S_1}$ are independent, we can apply Hoeffding’s inequality (see Corollary 7.21 of [6]) conditionally on $v$ to get
\[(2.5) \quad \mathbb{P}(|X| \geq t \mid v) \leq 2 \exp\left(-\frac{t^2}{2\|v\|^2}\right)\]

In order to marginalize out $v$ from (2.5), we can bound $\|v\|^2$ as
\[\|v\|^2 = \sup_{\|z\| \leq 1} \langle z, v \rangle = \sup_{\|z\| \leq 1} \sum_{j \geq 2} \langle z_{S_j}, D x_{S_j} e S_j \Phi^* F \Phi D x_{S_1} \rangle\]
\[\leq \sup_{\|z\| \leq 1} \sum_{j \geq 2} \|z_{S_j}\| \|D x_{S_j} e S_j \Phi^* F \Phi D x_{S_1}\| \leq \sup_{\|z\| \leq 1} \sum_{j \geq 2} \|\Phi^* F \Phi D x_{S_1}\| \|x_{S_j}\| \|x_{S_1}\| \leq \sum_{j \geq 2} \|\Phi^* F \Phi D x_{S_1}\| \|x_{S_j}\| \|x_{S_1}\|\]

Given $\{S_1, S_2, \ldots, S_R\}$ is a partition of $[N]$ according to a nonincreasing rearrangement of $x$, we have $\|x_{S_j}\| \leq \frac{1}{\sqrt{2}} \|x_{S_{j-1}}\|$. Together with $\|x_{S_j}\| \leq \|x\| = 1$, and $\|\Phi^* F \Phi D x_{S_1}\| \|x_{S_j}\| \|x_{S_1}\| \leq \delta_{2k}$, we arrive at
\[(2.6) \quad \|v\|^2 \leq \frac{\delta_{2k}}{\sqrt{k}} \sup_{\|z\| \leq 1} \sum_{j \geq 2} \|x_{S_{j-1}}\| \|z_{S_j}\| \]

Combining (2.6) and (2.5), together with $\delta_{2k} \leq \frac{\delta}{4}$, we get the tail inequality
\[(2.7) \quad \mathbb{P}(|X| \geq t) \leq 2 \exp\left(-\frac{t^2 k}{2\delta_{2k}^2}\right) \leq 2 \exp\left(-\frac{8k t^2}{\delta^2}\right)\]

We continue to consider the third term in (2.3).
\[Y = \sum_{j,l \geq 2, j \neq l} \langle \Phi D^\xi x_{S_j}, \Phi D^\xi x_{S_l} \rangle\]
\[= \sum_{j,l \geq 2, j \neq l} \xi^* S_j D x_{S_j} e S_j \Phi^* F \Phi D x_{S_l} \xi_{S_l} = \xi^* B \xi\]
where $B$ is a matrix with zero diagonal blocks, defined as

$$B_{p,q} = \begin{cases} x_p \phi_p^* \phi_q x_q & \text{if } p \in S_j, q \in S_l \text{ for } j,l \geq 2, j \neq l, \\ 0 & \text{otherwise.} \end{cases}$$

Because $B$ is a self-adjoint matrix with zero diagonal, $Y$ is a homogeneous Rademacher chaos. Thus we can bound the tail of $Y$ through bounding $\|B\|_{2 \to 2}$ and $\|B\|_F$ (see Proposition 8.13 of [6]).

$$\|B\|_{2 \to 2} = \sup_{\|z\|_2 \leq 1} \langle Bz, z \rangle = \sup_{\|z\|_2 \leq 1} \sum_{i,j,l \geq 2, j \neq l} (z_{S_j}, D_{x_{S_j}} \phi_{S_j} \Phi_{S_l} D_{x_{S_l}} z_{S_l})$$

$$\leq \sup_{\|z\|_2 \leq 1} \sum_{i,j,l \geq 2, j \neq l} \|z_{S_j}\|_2 \|z_{S_l}\|_2 \|x_{S_l}\|_\infty \|x_{S_l}\|_\infty \|\Phi_{S_j} \Phi_{S_l}\|_2 \|z_{S_l}\|_2$$

$$\leq \delta_{2k} \sup_{\|z\|_2 \leq 1} \sum_{i,j,l \geq 2, j \neq l} \frac{1}{2} (\|z_{S_i}\|_2^2 + \|x_{S_i}\|_2^2) \|z_{S_l}\|_2$$

$$\leq \frac{\delta_{2k}}{k}$$

$$\|B\|_F^2 = \sum_{j \geq 1} \sum_{l \geq 2, j \neq l} \sum_{p \in S_j, q \in S_l} (x_p \phi_p^* \phi_q x_q)^2 = \sum_{j \geq 1} \sum_{l \geq 2, j \neq l} x_p^2 \|D_{x_{S_j}} \phi_{S_j} \Phi_{S_l} \Phi_{S_l}\|_2^2$$

$$\leq \sum_{j \geq 1} \sum_{l \geq 2, j \neq l} x_p^2 \|x_{S_j}\|_\infty^2 \|\Phi_{S_j} \Phi_{S_l} \Phi_{S_l}\|_2^2 \leq \delta_{2k}^2 \sum_{j \geq 1} \frac{1}{k} \|x_{S_j}\|_2^2 \|x_{S_l}\|_2^2$$

$$\leq \frac{\delta_{2k}}{k}$$

Applying Proposition 8.13 from [6], together with $\delta_{2k} \leq \frac{\delta}{4}$, we estimate the tail of $Y$ as

$$\mathbb{P}(\|y\| \geq r) \leq 2 \exp \left( - \min \left\{ \frac{3r^2}{128 \|B\|_F^2}, \frac{r}{32 \|B\|_{2 \to 2}} \right\} \right)$$

$$\leq 2 \exp \left( - \min \left\{ \frac{3kr^2}{128 \delta_{2k}^2}, \frac{kr}{32 \delta_{2k}} \right\} \right)$$

$$\leq 2 \exp \left( -k \min \left\{ \frac{3r^2}{8 \delta^2}, \frac{r}{8 \delta} \right\} \right)$$

From (2.3), we have

$$\|\Phi D_\xi x\|_2^2 = \sum_{j \geq 1} \|\Phi D_\xi x_{S_j}\|_2^2 + 2X + Y$$

If the event $E = \{ \omega : |X(\omega)| \leq \frac{\delta}{8} \text{ and } |Y(\omega)| \leq \frac{\delta}{2} \}$ happens, then for a fixed $x \in \mathcal{E}$ we have

$$1 - \delta \leq \|\Phi D_\xi x\|_2^2 \leq (1 + \delta)$$
By choosing $t = \frac{4}{9}$ in (2.7) and $r = \frac{4}{9}$ in (2.8), we have
\[
P(E) \geq 1 - 2 \exp(-k/8) - 2 \exp(-k \min\{3/32, 1/16\}) \geq 1 - 4 \exp(-k/16)
\]
Taking the union bound on $E$ with $|E| = p$, we have (2.9) holds for all $x \in E$ simultaneously with probability at least
\[
1 - 4p \exp(-k/16) \geq 1 - \epsilon
\]
if we choose $k \geq 16 \log \frac{4p}{\epsilon}$, i.e., $s \geq 32 \log \frac{4p}{\epsilon}$.

3. RIP of Structured Random Matrices. According to Theorem 2.1, the problem of constructing a JLT can be reduced to constructing a matrix with RIP. In Theorem 9.6 of [6], it is shown that if the random matrix $\Phi \in \mathbb{R}^{m \times N}$ is generated from a subgaussian ensemble, then it satisfies the $(s, \delta)$-RIP w.h.p. if
\[
m \gtrsim \delta^{-2} s \log(eN/s)
\]
Equation (3.1) combined with $s = O(\log(p))$ in Theorem 2.1, we conclude that a $(p, \delta)$-JLT can be constructed by randomly sampling from a subgaussian ensemble if $m \gtrsim \delta^{-2} \log(p) \log(eN)$. The downside of a random matrix from a subgaussian ensemble is that $\Phi$ is usually dense, which leads to slow matrix-vector multiplication of time $O(mN)$. In many algorithmic applications, it is desirable to have fast JLT with matrix-vector multiplication in time $O(N \log N)$. For example, if $\Phi$ is a partial Fourier matrix or partial Hadamard matrix, then $\Phi x$ can be computed in $O(N \log N)$. In [2], fast $(p, \delta)$-JLT are constructed with the embedding dimension $m \gtrsim \delta^{-4} \log(p) \log^4(N)$, which is far from the lower bound (1.1). As an alternative, Theorem 2.1 opens the door to constructing fast JLT through employing the RIP of structured random matrices, which is a nowadays active research topic [13, 14, 9, 11, 12].

3.1. RIP of random sampling matrices. Both random partial Fourier matrix and random partial Hadamard matrix are special cases of the random sampling matrix associated to a bounded orthonormal system (BOS). About RIP of the renormalized random sampling matrix, we have the following theorem (see Remark 12.33(b) of [6])

**Theorem 3.1.** Let $A \in \mathbb{C}^{m \times N}$ be the random sampling matrix associated to a BOS with constant $K \geq 1$. For $\delta \in (0, 1)$, if
\[
m \geq CK^2 \delta^{-2} s \log^3(s) \log(N)
\]
for some constant $C > 0$, then with probability at least $1 - N^{-\log^3(s)}$, $\frac{1}{\sqrt{m}} A$ has the $(s, \delta)$-RIP.

Taking the random partial Fourier matrix as an example, its associated BOS has $K = 1$. Equation (3.2) combined with $s = O(\log(p))$ in Theorem 2.1, we conclude that a fast $(p, \delta)$-JLT can be constructed from random partial Fourier matrices if the embedding dimension $m \gtrsim \delta^{-2} \log(p) \log^4 N$, which is an improvement over the bound given in [2].

3.2. RIP of partial random circulant matrices. Another type of structured random matrix that allows for fast matrix-vector multiplication is the partial random circulant matrix [13, 15, 9]. The circulant matrix $H_z$ associated with a vector $z \in \mathbb{C}^{N \times N}$ is characterized by
\[
H_z x = z * x \quad \text{for all } x \in \mathbb{C}^N
\]
where \(*\) denotes the circular convolution. It can be shown that [7]

\[
H_z = F^{-1} D(Fz) F
\]

where \(F\) is the DFT matrix. This explains why \(H_z x\) can be computed fast. Moreover, because the circular convolution is commutable, we have

\[
H_z x = z * x = x * z = H_x z
\]

**Definition 3.2 (partial random circulant matrix [9]).** Let \(\xi \in \{-1, 1\}^N\) be a Rademacher sequence, let \(T \subset [N]\) be an arbitrary fixed set of cardinality \(|T| = m\), and let \(R_T\) be the restriction operator to \(T\). The partial random circulant matrix associated with \(\xi\) is given by \(\Phi = \frac{1}{\sqrt{m}} R_T H_\xi\) with the action on \(x\) as

\[
\Phi x = \frac{1}{\sqrt{m}} R_T (\xi * x)
\]

Regarding the RIP of a partial random circulant matrix, we have the following theorem [9]

**Theorem 3.3.** Let \(\Phi \in \mathbb{R}^{m \times N}\) be a partial random circulant matrix associated with a Rademacher sequence \(\xi\). It satisfies the \((s, \delta)\)-RIP with probability at least \(1 - N^{-\log N \cdot \log^2 s}\) if

\[
m \geq C \delta^{-2} s \log^2 s \cdot \log^2 N
\]

If we combine the bound (3.6) with \(s = O(\log(p))\) in Theorem 2.1, we conclude that a fast \((p, \delta)\)-JLT can be constructed from partial random circulant matrices if the embedding dimension \(m \gtrsim \delta^{-2} \log(p) \log^4 N\).

**Proof sketch of Theorem 3.3.** Setting \(D_{s,N} = \{x \in \mathbb{C}^N : \|x\|_2 \leq 1, x \in \Sigma_s\}\), the \(s\)-order restricted isometry constant of \(\Phi\) is

\[
\delta_s = \sup_{x \in D_{s,N}} \left| \frac{1}{m} \|R_T (\xi * x)\|_2^2 - \|x\|_2^2 \right| = \sup_{x \in D_{s,N}} \left| \frac{1}{m} \|P_T (x * \xi)\|_2^2 - \|x\|_2^2 \right|
\]

where we have introduced matrices \(P_T = R_T^* R_{sT}\), and \(V_x = \frac{1}{\sqrt{m}} P_T H_x\).

\[
E\|V_x \xi\|_2^2 = \frac{1}{m} E\|P_T H_x \xi\|_2^2 = \frac{1}{m} E \sum_{l \in T} |\langle x_{-l}, \xi \rangle|^2
\]

\[
= \frac{1}{m} \sum_{l \in T} E|\langle x_{-l}, \xi \rangle|^2 = \|x\|_2^2
\]

where \(x_{-l}\) represents the \(l\)-th cyclic shift of \(x\), and the last equality follows from \(\xi\) being isotropic. Hence

\[
\delta_s = \sup_{x \in D_{s,N}} \left| \|V_x \xi\|_2^2 - E\|V_x \xi\|_2^2 \right|
\]

which says \(\delta_s\) is the supremum of a chaos process with the index set \(\mathcal{A} = \{V_x : x \in D_{s,N}\}\). It is shown in [9] (see Theorem 3.1 and Theorem 3.5 therein) that to bound the supremum of the chaos process in (3.7), it suffices to control the parameters
$$d_{2\to 2}(A) = \sup_{A \in A} \|A\|_{2\to 2}, \quad d_F(A) = \sup_{A \in A} \|A\|_F,$$

and the Talagrand’s functional
$$\gamma_2(A, \|\cdot\|_{2\to 2}).$$

For $A = \{V_x : x \in D_{s,N}\}$, we can estimate $d_F(A)$ and $d_{2\to 2}(A)$ as follows.

$$\|V_x\|_F^2 = \frac{1}{m} \|P_T H_x\|_F^2 = \frac{1}{m} \sum_{t \in T} \|x - l\|_2^2$$

$$= \|x\|_2^2 \leq 1$$

Thus

$$d_F(A) = 1$$

Using representation (3.3), we can estimate

$$\|V_x\|_{2\to 2} = \frac{1}{\sqrt{m}} \|P_T \mathcal{F}^{-1} D(F x) \mathcal{F}\|_{2\to 2} \leq \frac{\sqrt{N}}{\sqrt{m}} \|P_T \mathcal{F}^{-1}\|_{2\to 2} \|D(F x)\|_{2\to 2}$$

$$= \frac{1}{\sqrt{m}} \|D(F x)\|_{2\to 2} = \frac{1}{\sqrt{m}} \|\mathcal{F} x\|_\infty \leq \frac{1}{\sqrt{m}} \|x\|_1 \leq \frac{\sqrt{s}}{\sqrt{m}} \|x\|_2 \leq \sqrt{\frac{s}{m}}$$

Thus

$$d_{2\to 2}(A) \leq \sqrt{\frac{s}{m}}$$

$\gamma_2(A, \|\cdot\|_{2\to 2})$ can be bounded by Dudley’s integral inequality, which is omitted here.

4. Conclusion. In this report, we study the connection between the JLT and RIP. On the one hand, both JLT and RIP can be derived from the concentration inequality (2.1). On the other hand, Theorem 2.1 establishes that we can convert a RIP matrix to a JLT, which is particularly useful for constructing fast JLT from structured RIP matrices. We have seen that using either the random sampling matrices associated to a BOS or the partial random circulant matrices, we can construct fast JLT with improved embedding dimensionality. The RIP analysis of structured random matrices are achieved by bounding the suprema of chaos processes, which is a more powerful and general tool than the concentration inequality (2.1). It is in this hindsight that we consider RIP to be more fundamental than JLT.

REFERENCES


