On Extended Concentration Inequalities for Fast JL Embeddings of Infinite Sets

Edem Boahen, March T. Boedihardjo, Rafael Chiclana, Mark Iwen
Michigan State University
Email: boahened@msu.edu, boedihar@msu.edu, chiclan1@msu.edu, iwenmark@msu.edu

***Abstract*—The Johnson-Lindenstrauss (JL) lemma allows subsets of a high-dimensional space to be embedded into a lower-dimensional space while approximately preserving all pairwise Euclidean distances. This important result has inspired an extensive literature, with a significant portion dedicated to constructing structured random matrices with fast matrix-vector multiplication algorithms that generate such embeddings for finite point sets. In this paper, we briefly consider fast JL embedding matrices for *infinite* subsets of $\mathbb{R}^d$. Prior work in this direction such as [15], [14] has focused on constructing fast JL matrices $HD \in \mathbb{R}^{k\times d}$ by multiplying structured matrices with RIP(-like) properties $H \in \mathbb{R}^{k\times d}$ against a random diagonal matrix $D \in \mathbb{R}^{d\times d}$. However, utilizing RIP(-like) matrices H in this fashion necessarily has the unfortunate side effect that the resulting embedding dimension k must depend on the ambient dimension d no matter how simple the infinite set is that one aims to embed. Motivated by this, we explore an alternate strategy for removing this d-dependence from k herein: Extending a concentration inequality proven by Ailon and Liberty [1] in the hope of later utilizing it in a chaining argument to obtain a near-optimal result for infinite sets. Though this strategy ultimately fails to provide the near-optimal embedding dimension we seek, along the way we obtain a stronger-than-sub-exponential extension of the concentration inequality in [1] which may be of independent interest.**

I. Introduction

The Johnson-Lindenstrauss lemma [8] states that for $\varepsilon \in (0,1)$ and a finite set $T \subset \mathbb{R}^d$ with $n > 1$ elements, there exists a $k \times d$ matrix Φ with $k = \mathcal{O}(\varepsilon^{-2}\log n)$ such that

$$(1-\varepsilon)\|\mathbf{x}-\mathbf{y}\|_2^2 \le \|\Phi\mathbf{x}-\Phi\mathbf{y}\|_2^2 \le (1+\varepsilon)\|\mathbf{x}-\mathbf{y}\|_2^2 \quad \text{(I.1)}$$

holds $\forall\, \mathbf{x},\mathbf{y} \in T$. A matrix Φ satisfying (I.1) is called an **ε-JL embedding** of T into $\mathbb{R}^k$. Moreover, it has been shown that the dimension k of the Euclidean space where T is embedded is optimal for finite sets (see [12]). This result is a cornerstone in dimensionality reduction and has proved to be an extremely useful tool in many application domains (see, e.g., the relevant discussions in [3], [13], [9], [2], [4], [5], [7]).

If $T \subset \mathbb{R}^d$ is an infinite set one may bound the embedding dimension k in terms of its *Gaussian Width*, $w(T) := \mathbb{E}\sup_{\mathbf{x}\in T}\langle \mathbf{g},\mathbf{x}\rangle$, where $\mathbf{g}$ is a random vector with d independent and identically distributed (i.i.d.) mean 0 and variance 1 Gaussian entries (see, e.g,. [16, Definition 7.5.1]). Let $\mathrm{unit}(T-T) := \{(\mathbf{x}-\mathbf{y})/\|\mathbf{x}-\mathbf{y}\|_2 \mid \mathbf{x},\mathbf{y} \in T, \mathbf{x} \ne \mathbf{y}\}$. For any bounded set $T \subset \mathbb{R}^d$, standard upper bounds demonstrate that sub-Gaussian random matrices Φ are ε-JL embeddings of T into $\mathbb{R}^k$, where $k = \mathcal{O}(\varepsilon^{-2}w^2(\mathrm{unit}(T-T)))$, with high probability (w.h.p.) (see, e.g., [16, Theorem 9.1.1 and Exercise 9.1.8]). Similarly, [6, Theorem 9] shows that any JL embedding of a bounded set $T \subseteq \mathbb{R}^d$ into $\mathbb{R}^k$ must have $k \gtrsim w^2(T)$, which matches the prior upper bound for a large class of sets $T \subset \mathbb{R}^d$ when ϵ isn't too small. Most importantly for our discussion here, we note that all the bounds on k mentioned above are entirely independent of d.

If we further demand that a $\Phi \in \mathbb{R}^{k\times d}$ satisfying (I.1) $\forall\, \mathbf{x},\mathbf{y} \in T$ also be a structured matrix with an associated fast matrix-vector multiplication algorithm, the situation complicates. In this setting state-of-the-art results [15], [14] build on Restricted Isometry Property (RIP) related results implied by, e.g., [11], [10] to produce structured ϵ-JL embeddings of infinite sets that also have fast matrix-vector multiplication algorithms. However, their dependence on the RIP has the unfortunate side effect that the bounds they obtain on the embedding dimension k must always depend (logarithmically) on d no matter how simple the set T is.

Returning to the setting of *finite* sets T, in [1] Ailon and Liberty construct ϵ-JL embeddings with fast matrix-vector multiplication algorithms that also have near-optimal embedding dimensions for sets of sufficiently small cardinality. In particular, they construct a $k \times d$ matrix A for which the mapping $\mathbf{x} \mapsto A\mathbf{x}$ can be computed in $\mathcal{O}(d\log k)$-time that also satisfies the following sub-Gaussian concentration inequality: For any $\mathbf{x} \in \mathbb{R}^d$ with $\|\mathbf{x}\|_2 = 1$ and $0 < t < 1$,

$$\mathbb{P}\left\{|\|A\mathbf{x}\|_2 - 1| > t\right\} \le c_1\exp\{-c_2kt^2\}, \qquad \text{(I.2)}$$

for some universal constants c_1, $c_2 > 0$. This then allows for optimal dimensionality reduction of finite sets T with n elements by taking $k = \mathcal{O}(\epsilon^{-2}\log n)$, where $\epsilon > 0$ is the desired distortion of the JL-embedding. Looking at (I.2) in the context of, e.g., [16, Chapter 8], one might wonder if (I.2) can be extended to hold for all $t > 0$. If so, a chaining argument could then be employed to extend the fast ϵ-JL embedding results in [1] to more arbitrary (e.g., infinite) subsets of $\mathbb{R}^d$ with embedding dimensions k that don't depend on d.

Unfortunately, the approach in [1] apparently fails to provide sub-Gaussian concentration for large distortions t. We are, however, at least able to demonstrate an extended concentration inequality herein that is better than sub-exponential.

Theorem I.1. *There is a $k \times d$ random matrix A, for which the mapping $\mathbf{x} \mapsto A\mathbf{x}$ can be computed in time $\mathcal{O}(d\log k)$,* *such that for any* $\mathbf{x} \in \mathbb{R}^d$ *with* $\|\mathbf{x}\|_2 = 1$ *and* $t > 0$

$$\mathbb{P}\{|\|A\mathbf{x}\|_2 - 1| > t\} \le c_3 \exp\{-c_4 k^{2/3} t^{4/3}\},$$

for some universal constants c_3*,* $c_4 > 0$*.*

The matrix A appearing in the above result is defined as $A = BDHD'$, where B is a 4-wise independent $k \times d$ matrix, D and D' are independent diagonal matrices with entries that are random variables taking values ± 1 with probability $1/2$, and H is a $d \times d$ Walsh-Hadamard matrix. This is a simplified version of the matrix that Ailon and Liberty constructed in [1]. While Ailon and Liberty's construction involves iterating the transformation HD' multiple times, Example III.3 demonstrates that such iterations do not lead to improvements in Theorem I.1 for large distortions. We refer to Section II for a detailed construction of the matrix A.

Using Theorem I.1 one can now quickly prove the following theorem providing d-independent bounds on the embedding dimension k of A for infinite sets.

Theorem I.2. *Let* A *be the random matrix in Theorem I.1. Let* $S \subset \{x \in \mathbb{R}^d : \|x\|_2 = 1\}$ *and* $\epsilon, p \in (0,1)$*. Then* A *is an* ϵ*-JL map of* S *into* $\mathbb{R}^k$ *with probability at least* $1-p$ *provided that*

$$k \ge \frac{C}{\epsilon^4}\left(\left(\ln\frac{1}{p}\right)^{\frac{3}{4}} + \sum_{j=0}^{\infty}\frac{1}{2^j}\left(\ln N\left(S, \|\cdot\|_2, \frac{1}{2^j}\right)\right)^{\frac{3}{4}}\right)^2,$$

where $N\left(S, \|\cdot\|_2, \frac{1}{2^j}\right)$ *is the* $\frac{1}{2^j}$*-covering number of* S*, and* $C \ge 1$ *is a universal constant.*

Looking at Theorem I.2 we can see that it is indeed independent of d as desired. Unfortunately, it also provides sub-optimal dependence on the covering numbers (in this case) of the set S. It is proven in Section IV for completeness.

II. Preliminaries

In this section, we review the necessary background and results that will be applied throughout the paper.

A. Definitions

Given $\mathbf{x} \in \mathbb{R}^d$, we write $\|\mathbf{x}\|_2$ for the Euclidean norm of $\mathbf{x}$. Given a $k \times d$ matrix B, the operator norm $\|B^T\|_{2\to4}$ is defined as the maximum ratio of the ℓ_4 norm of the matrix-vector product to the ℓ_2 norm of the vector, formally expressed as:

$$\|B^T\|_{2\to4} = \sup_{\|\mathbf{x}\|_2=1} \|B^T\mathbf{x}\|_4,$$

where the ℓ_4 norm of a vector $\mathbf{x} \in \mathbb{R}^d$ is defined as

$$\|\mathbf{x}\|_4 = \left(\sum_{i=1}^d |x_i|^4\right)^{1/4}.$$

A *Rademacher sequence* $\boldsymbol{\xi} \in \mathbb{R}^d$ is a random vector whose coordinates are independent and take the value 1 or -1 with equal probability. A random variable X is said to be *sub-Gaussian* if its *sub-Gaussian norm*, defined as

$$\|X\|_{\Psi_2} = \inf\left\{c > 0 \colon \mathbb{E}e^{X^2/c^2} \le 2\right\},$$

is finite. An equivalent characterization of sub-Gaussianity is given by the tail bound:

$$\mathbb{P}\{|X| \ge t\} \le 2e^{-t^2/C_1^2} \quad \forall\, t \ge 0,$$

for some constant $C_1 > 0$. Another useful characterization is that X is sub-Gaussian if for any $p \ge 1$,

$$\mathbb{E}|X|^p \le C_2^p p^{p/2} \quad \text{for some constant } C_2 > 0.$$

A *Walsh-Hadamard* matrix H_d is a $d \times d$ orthogonal matrix defined recursively: For $d = 1$, $H_1 = [1]$, and for $d = 2^n$, it is defined as

$$H_d = \frac{1}{\sqrt{2}}\begin{bmatrix} H_{d/2} & H_{d/2} \\ H_{d/2} & -H_{d/2}\end{bmatrix}.$$

The entries of a Walsh-Hadamard matrix are given by

$$H_d(i,j) = d^{-1/2}(-1)^{\langle i,j\rangle},$$

where $\langle i,j\rangle$ denotes the dot product of the binary representations of the indices i and j. For convenience, we will omit the subscript and denote the Walsh-Hadamard matrix simply as H instead of H_d

A matrix B of size $k \times d$ is said to be *4-wise independent* if for any $1 \le i_1 < i_2 < i_3 < i_4 \le k$ and any $(b_1, b_2, b_3, b_4) \in \{1,-1\}^4$, the number of columns $B^{(j)}$ for which $(A_{i_1}^{(j)}, A_{i_2}^{(j)}, A_{i_3}^{(j)}, A_{i_4}^{(j)}) = k^{-1/2}(b_1, b_2, b_3, b_4)$ is exactly $d/16$.

B. Supporting Results

We will utilize several supporting theorems in our analysis. We start with the following classic tool that provides concentration bounds for sums of bounded independent random variables.

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

$$\mathbb{P}\left\{\left|\sum_{j=1}^d \xi_j x_j\right| > t\right\} \le 2\exp\left(-\frac{t^2}{2\|\mathbf{x}\|_2^2}\right).$$

We will also need some results from the work of Ailon and Liberty [1], particularly the following lemmas:

Lemma II.2 (Corollary 5.1 from [1])**.** *Let* B *be a* $k \times d$ *matrix with Euclidean unit length columns, and let* D *be a random* $\{\pm 1\}$ *diagonal matrix. Given* $\mathbf{x} \in \mathbb{R}^d$ *with* $\|\mathbf{x}\|_2 = 1$*, let* $Y = \|BD\mathbf{x}\|_2$*. Then, for any* $t \ge 0$*,*

$$\mathbb{P}\{|Y - 1| > t\} \le c_5 \exp\left\{-c_6 \frac{t^2}{\|\mathbf{x}\|_4^2 \|B^T\|_{2\to4}^2}\right\},$$

for some universal constants $c_5, c_6 > 0$*.*

Lemma II.3 (Lemma 4.1 from [1])**.** *There exists a* 4*-wise independent code matrix of size* $k \times f_{\mathrm{BCH}}(k)$*, where* $f_{\mathrm{BCH}}(k) = \Theta(k^2)$.

Lemma II.4 (Lemma 5.1 from [1])**.** *Assume that B is a $k \times d$* 4*-wise independent matrix. Then,*

$$\|B^T\|_{2\to 4} \leq (3d)^{1/4} k^{-1/2}.$$

III. Proofs

Consider a $d \times d$ diagonal matrix D', whose diagonal entries are independent variables that take the value 1 or -1 with probability $1/2$.

The norm in ℓ_4^d will play an important role in our analysis. Specifically, there exist vectors with $\|\mathbf{x}\|_2 = 1$ but very small $\|\mathbf{x}\|_4$, which occurs when the vector is "flat". Roughly speaking, a vector is considered flat if a substantial portion of its coordinates have similar absolute values. Our first result demonstrates that we can make a unit vector flat w.h.p. by applying HD'. This result provides a sharp version of inequality (5.6) in [1] and serves as a fundamental element of the argument for constructing the desired embedding.

Lemma III.1. *Let $\mathbf{x} \in \mathbb{R}^d$ with $\|\mathbf{x}\|_2 = 1$. Then, for any $t > 0$ we have*

$$\mathbb{P}\left\{\|HD'\mathbf{x}\|_4 \geq td^{-1/4}\right\} \leq e^{1-c_7 t^4},$$

where $c_7 > 0$ is a universal constant.

Proof. Fix $\mathbf{x} \in \mathbb{R}^d$ with $\|\mathbf{x}\|_2 = 1$. Write $Z = \sqrt{d}HD'\mathbf{x}$ and consider random variables $Z_1, \dots, Z_d$ so that $Z = (Z_1, \dots, Z_d)^T$. First, since H is orthogonal observe that

$$\sum_{j=1}^d Z_j^2 = \|Z\|_2^2 = d\|HD'\mathbf{x}\|_2^2 = d\|D'\mathbf{x}\|_2^2 = d\|\mathbf{x}\|_2^2 = d. \tag{III.1}$$

Next, we will show that each Z_i is a sub-Gaussian random variable for $i = 1, \dots, d$. To be more precise, we claim that

$$\mathbb{P}\{|Z_i| > t\} \leq 2e^{-t^2/2} \tag{III.2}$$

Fix $i = 1, \dots, d$. Then, $Z_i = d^{1/2} H_{(i)} D'\mathbf{x}$, where $H_{(i)}$ is the i^{th} row of H. Observe that $H_{(i)}D'$ is a vector whose entries are independent with each entry being $d^{-1/2}$ or $-d^{-1/2}$ with probability $1/2$. Consequently, Z_i has the same distribution as $\langle \boldsymbol{\xi}, \mathbf{x}\rangle$, where $\boldsymbol{\xi}$ is a Rademacher sequence. Thus, we can apply Proposition II.1 to obtain the claim.

We now proceed to estimate the norm of $HD'\mathbf{x}$ in ℓ_4^d. Observe that for any $p \geq 1$ we have

$$\begin{aligned}
\|Z\|_4^4 = \sum_{i=1}^d Z_i^4 &= \sum_{i=1}^d Z_i^{(2p+2)\frac{1}{p} + 2(1-\frac{1}{p})} \\
&\leq \left(\sum_{i=1}^d Z_i^{2p+2}\right)^{1/p} \left(\sum_{i=1}^d Z_i^2\right)^{1-\frac{1}{p}} \\
&= d^{1-\frac{1}{p}} \left(\sum_{i=1}^d Z_i^{2p+2}\right)^{1/p},
\end{aligned}$$

where the inequality follows from applying Hölder's inequality, and the last equality follows from (III.1). Taking expectation and using the fact that each Z_i is a sub-Gaussian random variable gives a universal constant $C > 0$ for which

$$\begin{aligned}
\mathbb{E}(\|Z\|_4^{4p}) &\leq d^{p-1} \sum_{i=1}^d \mathbb{E}|Z_i|^{2p+2} \\
&\leq d^{p-1} \sum_{i=1}^d C^{2p+2}(2p+2)^{p+1} \\
&= C^{2p+2} d^p (2p+2)^{p+1}.
\end{aligned}$$

Consequently, we deduce that

$$\mathbb{E}(\|HD'\mathbf{x}\|_4^{4p}) \leq C^{2p+2} d^{-p} (2p+2)^{p+1}.$$

Finally, Markov's inequality gives

$$\begin{aligned}
\mathbb{P}\left\{\|HD'\mathbf{x}\|_4 \geq td^{-1/4}\right\} &\leq \frac{\mathbb{E}(\|HD'\mathbf{x}\|_4^{4p})}{t^{4p} d^{-p}} \\
&\leq \frac{C^{2p+2}(2p+2)^{p+1}}{t^{4p}} \\
&\leq \left(\frac{C_1 p}{t^4}\right)^p,
\end{aligned}$$

where $C_1 > 0$ is another universal constant. Take $c_7 = \frac{1}{eC_1}$. When $t^4 \geq eC_1$, the result follows by taking $p = \frac{t^4}{eC_1}$ since it shows that

$$\mathbb{P}\left\{\|HD'\mathbf{x}\|_4 \geq td^{-1/4}\right\} \leq \exp\left(-\frac{t^4}{eC_1}\right) = e^{-c_7 t^4}.$$

When $t^4 \leq eC_1$, the result is trivial. $\square$

The following example shows that Lemma III.1 is sharp.

Example III.2. *Let $\mathbf{x} = d^{-1/2}(1, 1, \dots, 1)^T$. Then, with probability 2^{-d} we have $D'\mathbf{x} = \mathbf{x}$. In such a case, we have that $HD'\mathbf{x} = H\mathbf{x} = (1, 0, \dots, 0)^T$. Therefore, $\|HD'\mathbf{x}\|_4 = 1$. Taking $t = d^{1/4}$, this argument shows that*

$$\mathbb{P}\left\{\|HD'\mathbf{x}\|_4 \geq td^{-1/4}\right\} \geq 2^{-d} \geq e^{-t^4}.$$

As Lemma III.1 shows, applying the transformation HD' to a vector $x \in \mathbb{R}^d$ reduces its ℓ_4-norm with high probability. In [1], the embedding they consider is an iterative version of ours. Specifically, they apply the transformation HD_i, where D_i are independent copies of D', multiple times to further "flatten" the vector. While this approach works well for small distortions, the following example shows that for larger distortions iterating the transformation does not offer any additional improvement.

Example III.3. *For $i \in \mathbb{N}$, let D_i be indepentend diagonal matrices whose diagonal entries are independent and take the value 1 or -1 with equal probability. Let $x \in \mathbb{R}^d$ be a vector whose first $\sqrt{d}$ coordinates are equal to $d^{-1/4}$, and the rest are 0. Then, $\|x\|_2 = 1$. Observe that $D_1 x = x$ with probability $2^{-\sqrt{d}}$. In that case, we have $HD_1 x = Hx = x$. Repeating this argument r times, we find that with probability at least $2^{-r\sqrt{d}}$,*

$$HD_r HD_{r-1} \cdots HD_1 x = x$$

Setting the distortion $t = d^{1/8}$, we obtain

$$\mathbb{P}\left\{\|HD_r HD_{r-1} \cdots HD_1 x\|_4 \geq t d^{-1/4}\right\} \geq 2^{-r\sqrt{d}} \geq e^{-rt^4}.$$

This probability, up to a constant factor, is the same as the one achieved in Lemma III.1 using just a single iteration.

Let B be a $k \times d$ 4-wise independent code matrix. Consider D and D' independent diagonal matrices whose diagonal entries are independent and take the value ± 1 with probability $1/2$. Let H be a Walsh-Hadamard matrix of size $d \times d$. Define the matrix $A = BDHD'$.

Now we can prove our main result.

Proof of Theorem I.1. Let $u > 0$ and define the event $E_u = \{\|HD'\mathbf{x}\|_4 < u d^{-1/4}\}$. Let ${E_u}^{\mathsf{c}}$ be the complement event. Then, conditioning on E_u gives

$$\mathbb{P}\{|\|A\mathbf{x}\|_2 - 1| > t\} \leq \mathbb{P}\{|\|A\mathbf{x}\|_2 - 1| > t | E_u\} + \mathbb{P}\{{E_u}^{\mathsf{c}}\}$$

On the one hand, we can use Lemma II.2 and Lemma II.4 to obtain

$$\begin{aligned}\mathbb{P}\{|\|A\mathbf{x}\|_2 - 1| > t | E_u\} &\leq c_5 \exp\left\{-c_6 \frac{t^2}{u^2 d^{-1/2} \|B^T\|_{2\to 4}^2}\right\} \\ &\leq c_5 \exp\left\{-c_6 \frac{kt^2}{u^2\sqrt{3}}\right\}.\end{aligned}$$

On the other hand, Lemma III.1 gives

$$\mathbb{P}\{{E_u}^{\mathsf{c}}\} \leq \exp\{1 - c_7 u^4\}.$$

In conclusion, this yields

$$\mathbb{P}\{|\|A\mathbf{x}\|_2 - 1| > t\} \leq c_5 \exp\left\{-c_6 \frac{kt^2}{u^2\sqrt{3}}\right\} + \exp\{1 - c_7 u^4\}.$$

Taking $u = k^{1/6} t^{1/3}$, we conclude that for any $t > 0$

$$\mathbb{P}\{|\|A\mathbf{x}\|_2 - 1| > t\} \leq c_3 \exp\{-c_4 k^{2/3} t^{4/3}\},$$

for some universal constants c_3, $c_4 > 0$. □

IV. Proof of Theorem I.2

This section is dedicated to prove I.2.

Proof of Theorem I.2. Let $j_\epsilon = \lceil \log_2 \frac{8}{\epsilon} \rceil$. For each $j \geq 0$, let $S_j \subset S$ be such that $|S_j| = N(S, \|\cdot\|_2, \frac{1}{2^j})$ and $S \subset \cup_{\mathbf{x}_0 \in S_j}\{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x} - \mathbf{x}_0\|_2 \leq \frac{1}{2^j}\}$. For each $\mathbf{x} \in S$ and each $j \geq 0$, let $\pi_j(\mathbf{x})$ be the closest point in S_j to $\mathbf{x}$. Then $\|\pi_j(\mathbf{x}) - \mathbf{x}\|_2 \leq \frac{1}{2^j}$. So

$$\|\pi_{j+1}(\mathbf{x}) - \pi_j(\mathbf{x})\|_2 \leq \|\pi_{j+1}(\mathbf{x}) - \mathbf{x}\|_2 + \|\pi_j(\mathbf{x}) - \mathbf{x}\|_2 \leq \frac{2}{2^j}.$$

For each $\mathbf{x} \in S$, since $\mathbf{x} = \pi_{j_\epsilon}(\mathbf{x}) + \sum_{j=j_\epsilon}^{\infty}(\pi_{j+1}(\mathbf{x}) - \pi_j(\mathbf{x}))$, we have

$$\begin{aligned}|\|A\mathbf{x}\|_2 - 1| &\leq |\|A\pi_{j_\epsilon}(\mathbf{x})\|_2 - 1| + \|A(\mathbf{x} - \pi_{j_\epsilon}(\mathbf{x}))\|_2 \\ &\leq |\|A\pi_{j_\epsilon}(\mathbf{x})\|_2 - 1| + \sum_{j=j_\epsilon}^{\infty} \|A(\pi_{j+1}(\mathbf{x}) - \pi_j(\mathbf{x}))\|_2.\end{aligned}$$

Since $\pi_{j_\epsilon}(\mathbf{x}) \in S_{j_\epsilon}$ and $\pi_{j+1}(\mathbf{x}) - \pi_j(\mathbf{x}) \in S_{j+1} - S_j$ for all $\mathbf{x} \in S$, by Theorem I.1, with probability at least $1 - \frac{p}{2}$, we have

$$\begin{aligned}&\sup_{\mathbf{x}\in S} |\|A\pi_{j_\epsilon}(\mathbf{x})\|_2 - 1| \\ &\leq \frac{1}{\sqrt{k}}\left(\frac{1}{c_4}\ln\left(\frac{2c_3|S_{j_\epsilon}|}{p}\right)\right)^{\frac{3}{4}} \\ &\leq \frac{1}{\sqrt{k}}\left((\ln|S_{j_\epsilon}|)^{\frac{3}{4}} + \left(\ln\frac{1}{p}\right)^{\frac{3}{4}} + \left(\frac{\ln 2c_3}{c_4}\right)^{\frac{3}{4}}\right),\end{aligned}$$

and with probability at least $1 - \frac{p}{2^{j+2}}$, we have

$$\begin{aligned}&\sup_{\mathbf{x}\in S}\left\|A\frac{\pi_{j+1}(\mathbf{x}) - \pi_j(\mathbf{x})}{\|\pi_{j+1}(\mathbf{x}) - \pi_j(\mathbf{x})\|_2}\right\|_2 \\ &\leq 1 + \frac{[\frac{1}{c_4}\ln(\frac{2^{j+2}c_3}{p}|S_{j+1}||S_j|)]^{\frac{3}{4}}}{\sqrt{k}} \\ &\leq 1 + \frac{1}{\sqrt{k}}\Bigg((\ln|S_{j+1}|)^{\frac{3}{4}} + (\ln|S_j|)^{\frac{3}{4}} + \left(\ln\frac{1}{p}\right)^{\frac{3}{4}} \\ &\quad + ((j+2)\ln 2)^{\frac{3}{4}} + \left(\frac{\ln c_3}{c_4}\right)^{\frac{3}{4}}\Bigg).\end{aligned}$$

Therefore, with probability at least $1 - p$, we have

$$\begin{aligned}&\sup_{\mathbf{x}\in S}|\|A\mathbf{x}\|_2 - 1| \\ &\leq \frac{1}{\sqrt{k}}\left((\ln|S_{j_\epsilon}|)^{\frac{3}{4}} + 3\left(\ln\frac{1}{p}\right)^{\frac{3}{4}} + C + \sum_{j\geq j_\epsilon}\frac{6}{2^j}(\ln|S_j|)^{\frac{3}{4}}\right) \\ &\quad + \frac{4}{2^{j_\epsilon}} \\ &\leq \frac{C}{\sqrt{k}}\left(\left(\ln\frac{1}{p}\right)^{\frac{3}{4}} + \frac{1}{\epsilon}\sum_{j\geq j_\epsilon}\frac{1}{2^j}(\ln|S_j|)^{\frac{3}{4}}\right) + \frac{\epsilon}{2}.\end{aligned}$$

□

References

[1] Nir Ailon and Edo Liberty. Fast dimension reduction using rademacher series on dual bch codes. *Discrete & Computational Geometry*, 42(4):615–630, September 2008.

[2] Nir Ailon and Edo Liberty. An almost optimal unrestricted fast Johnson-Lindenstrauss transform. *ACM Trans. Algorithms*, 9(3):Art. 21, 12, 2013.

[3] Richard G. Baraniuk and Michael B. Wakin. Random projections of smooth manifolds. *Found. Comput. Math.*, 9(1):51–77, 2009.

[4] Jean Bourgain, Sjoerd Dirksen, and Jelani Nelson. Toward a unified theory of sparse dimensionality reduction in Euclidean space. In *STOC'15—Proceedings of the 2015 ACM Symposium on Theory of Computing*, pages 499–508. ACM, New York, 2015.

[5] Sjoerd Dirksen. Dimensionality reduction with subgaussian matrices: a unified theory. *Found. Comput. Math.*, 16(5):1367–1396, 2016.

[6] Mark Iwen, Benjamin Schmidt, and Arman Tavakoli. Lower bounds on the low-distortion embedding dimension of submanifolds of $\mathbb{R}^n$. *Applied and Computational Harmonic Analysis*, 65:170–180, 2023.

[7] Mark A. Iwen, Benjamin Schmidt, and Arman Tavakoli. On Fast Johnson-Lindenstrauss Embeddings of Compact Submanifolds of $\mathbb{R}^N$ with Boundary. *Discrete Comput. Geom.*, 71(2):498–555, 2024.

[8] William Johnson and Joram Lindenstrauss. Extensions of lipschitz maps into a hilbert space. *Contemporary Mathematics*, 26:189–206, 01 1984.

[9] Daniel M. Kane and Jelani Nelson. Sparser Johnson-Lindenstrauss transforms. In *Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1195–1203. ACM, New York, 2012.

[10] Felix Krahmer, Shahar Mendelson, and Holger Rauhut. Suprema of chaos processes and the restricted isometry property. *Communications on Pure and Applied Mathematics*, 67(11):1877–1904, 2014.

[11] Felix Krahmer and Rachel Ward. New and improved Johnson-Lindenstrauss embeddings via the restricted isometry property. *SIAM J. Math. Anal.*, 43(3):1269–1281, 2011.

[12] Kasper Green Larsen and Jelani Nelson. Optimality of the Johnson-Lindenstrauss lemma. In *58th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2017*, pages 633–638. IEEE Computer Soc., Los Alamitos, CA, 2017.

[13] Edo Liberty, Nir Ailon, and Amit Singer. Dense fast random projections and lean Walsh transforms. *Discrete Comput. Geom.*, 45(1):34–44, 2011.

[14] Shahar Mendelson. Column randomization and almost-isometric embeddings. *Information and Inference: A Journal of the IMA*, 12(1):1–25, 2023.

[15] Samet Oymak, Benjamin Recht, and Mahdi Soltanolkotabi. Isometric sketching of any set via the restricted isometry property. *Information and Inference: A Journal of the IMA*, 7(4):707–726, 2018.

[16] Roman Vershynin. *High-dimensional probability*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2018. An introduction with applications in data science, With a foreword by Sara van de Geer.