Non-Gaussian Random Matrices on Sets: Optimal Tail Dependence and Applications

Halyun Jeong*, Xiaowei Li†, Yaniv Plan‡ and Özgür Yılmaz§
Department of Mathematics, The University of British Columbia
Email: *hajeong@math.ubc.ca, †xli@math.ubc.ca, ‡yaniv@math.ubc.ca, §oyilmaz@math.ubc.ca

Abstract—Random linear mappings play a large role in modern signal processing and machine learning. For example, multiplication by a Gaussian matrix can preserve the geometry of a set while reducing the dimension.

Non-gaussian random mappings are attractive in practice for several reasons, including improved computational cost. On the other hand, mappings of interest often have heavier tails than Gaussian, which can lead to worse performance, i.e., less accurate preservation of the geometry of the set. In the sub-gaussian case, the size of the tail is measured with the sub-gaussian parameter, but the dependency has not been fully understood yet.

We present the optimal tail dependency on the sub-gaussian parameter and prove it through a new version of Bernstein's inequality. We also illustrate popular applications whose theoretical guarantees can be improved by our results.

I. INTRODUCTION

Random linear mappings play a central role in dimension reduction, compressed sensing, and numerical linear algebra (NLA) due to their propensity to preserve the geometry of a given set. The performance of a random linear mapping $A \in \mathbb{R}^{m \times n}$ is often determined by the uniform deviation bound of $\frac{1}{\sqrt{m}}\|Ax\|_2$ from $\|x\|_2$ for all vectors in the set (in other words, how good the map $\frac{1}{\sqrt{m}}A$ is as an isometry on the set). This is now well-understood by the standard techniques in the Gaussian random matrix case [1], [2].

However, in many applications, non-gaussian random mappings are more useful because of their computational/storage benefits or simply the difficulty to generate Gaussian matrices from sampling devices [3]. For example, sparse or structured random matrices are preferred in both dimension reduction [4] and random sketching in NLA [5]–[8] since they provide more efficient matrix multiplications than dense and unstructured matrices such as Gaussian ones. Certain formulations in compressed sensing also naturally require random matrices such as randomly subsampled Fourier measurements [9] or Bernoulli random matrices [10].

There has been a series of recent works [4], [11], [12] to demonstrate the effectiveness of random mappings outside the Gaussian setup. Unlike the Gaussian case in which we have a rotation invariance property, these papers use more sophisticated arguments since there are new technical challenges.

In this line of research, Liaw et al. [11] showed the following uniform deviation bound. Let $T \subset \mathbb{R}^n$, then with high probability we have

$$\sup_{x \in T}\left|\frac{1}{\sqrt{m}}\|Ax\|_2 - \|x\|_2\right| \leq \frac{K^2 \cdot O(w(T) + \mathrm{rad}(T))}{\sqrt{m}} \quad (1)$$

where K is the sub-gaussian parameter of A as in Definition I.2; $\mathrm{rad}(T) := \sup_{y \in T}\|y\|_2$, which is the radius when T is symmetric; $w(T)$ is the Gaussian width, defined by

$$w(T) := \mathbb{E}\sup_{y \in T}\langle g, y\rangle \text{ where } g \sim \mathrm{Normal}(0, I_n).$$

The sub-gaussian parameter roughly measures how fast the tail of a distribution decays; usually the bigger K is, the heavier the tail. Gaussian width measures the complexity of a set, in particular, $w^2(\mathrm{cone}(T) \cap \mathbb{S}^{n-1})$ is a meaningful approximation for dimension [12], [13].

When it comes to the dependency on the sub-gaussian parameter K in (1), other important works regarding this type of bound are either not explicit [12], [14] or at least of the same order K^2 [4], [15], [16].

In this article we refine these bounds by improving the dependency on the sub-gaussian parameter from K^2 to $K\sqrt{\log K}$. This enhances the deviation bound substantially when the sub-gaussian mapping is not well-behaved, for example, when K increases together with the signal dimension. We also prove that this dependency on K is optimal. Our bound is broadly applicable since it requires only the isotropic and sub-gaussian properties of the random matrices without any other assumptions.

The proof follows from an analogous approach in Liaw et al. [11]. The novel part we develop is a new Bernstein's inequality under bounded first absolute moment condition. We believe that this new Bernstein's inequality may be interesting on its own as an application-oriented deviation inequality.

A. Definitions

Sub-gaussian random variables have tails bounded by Gaussian random variables. Similarly, sub-exponential random variables have tails bounded by exponential random variables. There are several equivalent ways to define the sub-gaussian and sub-exponential norm [2]. We will use the following definition, which is the easiest to work with for our purpose.

Definition I.1. *For a random variable Z, define its sub-gaussian norm as*

$$\|Z\|_{\psi_2} := \inf\{t > 0 : \mathbb{E}\exp(Z^2/t^2) \leq 2\},$$

and sub-exponential norm as

$$\|Z\|_{\psi_1} := \inf\{t > 0 : \mathbb{E}\exp(|Z|/t) \leq 2\}.$$

Furthermore, Z is called sub-gaussian (or sub-exponential) if $\|Z\|_{\psi_2} < \infty$ (or $\|Z\|_{\psi_1} < \infty$).

It is easy to see from the definition that X is sub-gaussian if and only if X^2 is sub-exponential, and $\|X^2\|_{\psi_1} = \|X\|_{\psi_2}^2$. The sub-gaussian norm for $\mathrm{Normal}(0, \sigma^2)$ is $\sqrt{8/3}\,\sigma$; for $\mathrm{Bernoulli}(1, p)$ it is $\log^{-1/2}(1 + p^{-1})$; for Rademacher random variable it is $\log^{-1/2}(2)$ and for any bounded (by M) random variable it is no more than $M \log^{-1/2}(2)$.

We will be focusing on isotropic, sub-gaussian random matrices. The isotropic condition guarantees that these matrices (after normalization $\frac{1}{\sqrt{m}}$) preserve Euclidean norm in expectation.

Definition I.2. *For a random vector $a \in \mathbb{R}^n$,*

- *a is sub-gaussian if*
$$\|a\|_{\psi_2} := \sup_{x \in \mathbb{S}^{n-1}} \|\langle a, x \rangle\|_{\psi_2} < \infty,$$
- *a is isotropic if*
$$\mathbb{E} a a^T = I_n.$$

A random matrix $A \in \mathbb{R}^{m \times n}$ is isotropic and sub-gaussian if its rows are independent, isotropic and sub-gaussian. The sub-gaussian parameter of A is defined to be

$$K := \max\{\|A_i\|_{\psi_2} : A_i \text{ are the rows of } A, \ 1 \le i \le m\}.$$

Some examples of isotropic and sub-gaussian matrices are matrices whose entries are independent and sub-gaussian with second moment 1, uniformly subsampled (with replacement) rows of orthonormal basis or tight frames, etc. [2]. In the cases of Bernoulli matrices or sparse ternary matrices, which is a generalization of the database-friendly mappings in [5], the sub-gaussian parameter could depend on the signal dimension n if the probability of an entry being nonzero is n-dependent.

B. Notations

We will use $\|\cdot\|_2$ for Euclidean norm, use $\circ$ for Hadamard (entrywise) product, say $f \lesssim g$ if $f \le Cg$ for some absolute constant C, and similar for $f \gtrsim g$. We will also use c and C to denote absolute constants (usually c for small ones and C for large ones), and these absolute constants may vary from line to line.

II. MAIN RESULTS

A. Random Matrices on Sets

We state our main theorem below. This result improves the dependence on K over Theorem 1.1 in [11]. Recall that $\mathrm{rad}(T) = \sup_{y \in T} \|y\|_2$ and $w(T)$ is the Gaussian width of T.

Theorem II.1. *Let $A \in \mathbb{R}^{m \times n}$ be an isotropic and sub-gaussian matrix with sub-gaussian parameter K, and let $T \subset \mathbb{R}^n$ be a bounded set, then*

$$\mathbb{E} \sup_{x \in T} \left| \|Ax\|_2 - \sqrt{m}\|x\|_2 \right| \le CK\sqrt{\log K}\,[w(T) + \mathrm{rad}(T)],$$

and the event

$$\sup_{x \in T} \left| \|Ax\|_2 - \sqrt{m}\|x\|_2 \right| \le CK\sqrt{\log K}\,[w(T) + u \cdot \mathrm{rad}(T)]$$

holds with probability at least $1 - \exp(-u^2)$.

Generally $\mathrm{rad}(T)$ is dominated by $w(T)$. For example, if $0 \in T$, then $w(T) \ge \mathrm{rad}(T)/\sqrt{2\pi}$. In that case, with high probability, $\frac{1}{\sqrt{m}} A$ is a near isometry on T whenever $m \gtrsim K^2 \log K w^2(T)$.

To prove Theorem II.1, we first need to look at the case when T consists of only a single point. Let $x \in \mathbb{S}^{n-1}$ be a fixed vector on the unit sphere and define $X := Ax \in \mathbb{R}^m$ as the random vector that A maps x to. Then the coordinates X_i of this random vector are independent with

$$\mathbb{E} X_i^2 = 1 \text{ (isotropic)}, \quad \|X_i\|_{\psi_2} \le K \text{ (sub-gaussian)}. \quad (2)$$

Furthermore Lemma 5.3 in [11] states that

$$\left\| \|X\|_2 - \sqrt{m} \right\|_{\psi_2} \lesssim K^2.$$

In other words, $\|Ax\|_2$ has a sub-gaussian concentration around $\sqrt{m}$. It is worth noting that this concentration is independent of the ambient dimension m.

In [11], it is listed as an open problem whether this K^2 dependence can be improved. Here we give a complete answer to this question. The following Theorem II.2 improves this dependence from K^2 to $K\sqrt{\log K}$. Then in Proposition II.3, we construct an example that shows this $K\sqrt{\log K}$ is also tight. We state our improved dependence on K in a more general form by also allowing arbitrary weights b_i on X_i.

Theorem II.2. *Let $X = (X_1, \ldots, X_m)$ be a random vector in $\mathbb{R}^m$ with independent sub-gaussian entries satisfying $\mathbb{E} X_i^2 = 1$ and $\|X_i\|_{\psi_2} \le K$. Then for any fixed $b = (b_1, \ldots, b_m) \in \mathbb{R}^m$, we have*

$$\left\| \|b \circ X\|_2 - \|b\|_2 \right\|_{\psi_2} \le CK\sqrt{\log K}\|b\|_\infty.$$

Here $\circ$ denotes the entrywise product.

If b is the all ones vector, this upper bound is achieved (up to a constant) by scaled Bernoulli random variables.

Proposition II.3. *Let $K \ge 3$ and $X = (X_1, \ldots, X_m) \in \mathbb{R}^m$ be a random vector with independent entries such that $\frac{1}{K^2 \log K} X_i^2 \sim \mathrm{Bernoulli}\left(1, \frac{1}{K^2 \log K}\right)$, then $\|X_i\|_{\psi_2} \le K$, and for $m \ge K^2 \log K$,*

$$\left\| \|X\|_2 - \sqrt{m} \right\|_{\psi_2} \ge cK\sqrt{\log K}. \quad (3)$$

Here the assumption $m \ge K^2 \log K$ is mild because it essentially requires that X is non-zero in expectation.

B. A New Bernstein's Inequality

To prove Theorem II.2, we use a new version of Bernstein's inequality (Theorem II.4). Compared to the standard Bernstein's inequality for sub-exponential random variables [2], we impose an extra assumption on the first absolute moment ($E|Y_i| \le 2$ in Theorem II.4). This condition comes naturally form the isotropic condition of A because if we set $Y_i := X_i^2 - 1$, then $\mathbb{E}|Y_i| \le \mathbb{E} X_i^2 + 1 = 2$.

Theorem II.4. *Let $a = (a_1, \ldots, a_m)$ be a fixed non-zero vector and let $Y_1, \ldots, Y_m$ be independent, mean zero sub-exponential random variables satisfying $\mathbb{E}|Y_i| \leq 2$ and $\|Y_i\|_{\psi_1} \leq K_i^2$ $(K_i \geq 2)$. Then for every $u \geq 0$, we have*

$$\mathbb{P}\left(\left|\sum_{i=1}^m a_i Y_i\right| \geq u\right) \leq$$

$$2\exp\left[-c\min\left(\frac{u^2}{\sum_{i=1}^m a_i^2 K_i^2 \log K_i}, \frac{u}{\|a\|_\infty K^2 \log K}\right)\right],$$

where $K = \max_i K_i$ and c is an absolute constant.

The improvement of Theorem II.4 over standard Bernstein's inequality is the parameter in the sub-gaussian regime, from $\sum a_i^2 K_i^4$ to $\sum a_i^2 K_i^2 \log K_i$.

Remark II.5. *In the assumption $\mathbb{E}|Y_i| \leq 2$, we can replace the 2 by any arbitrary constant and it would only lead to a different absolute constant c.*

III. APPLICATIONS

A. Johnson-Lindenstrauss Lemma

One immediate application of our result is a guarantee for all isotropic and sub-gaussian matrices as Johnson-Lindenstrauss (JL) embeddings for dimension reduction. We state this JL lemma below. It follows directly form Theorem II.2.

Lemma III.1. *Let $A \in \mathbb{R}^{m \times n}$ be an isotropic and sub-gaussian matrix with sub-gaussian parameter K. If*

$$m \geq CK^2 \log K \epsilon^{-2} \log(1/\delta), \tag{4}$$

then for any $x, y \in \mathbb{R}^n$, with probability at least $1 - \delta$ we have

$$(1 - \epsilon)\|x - y\|_2 \leq \frac{1}{\sqrt{m}}\|A(x - y)\|_2 \leq (1 + \epsilon)\|x - y\|_2.$$

It is known that the dependence on ϵ and δ in (4) is optimal for linear mappings [17]. Based on Proposition II.3 we can now say that the dependence on sub-gaussian parameter K is also optimal. Similar results have appeared in [4], [18], but to the best of our knowledge, the previous known dependence on K was K^4.

B. Random Sketches

In sketching [7], we have a hugely over-determined system $Bx = y$ and the goal is to solve the least square problem on some convex set $\mathcal{C}$, given by

$$\min_{x \in \mathcal{C}} f(x) := \|Bx - y\|_2^2. \tag{5}$$

To save computation time, we could use an isotropic, sub-gaussian matrix $A \in \mathbb{R}^{m \times n}$ to reduce the dimension and solve the sketched problem instead.

$$\min_{x \in \mathcal{C}} g(x) := \|A(Bx - y)\|_2^2. \tag{6}$$

We say a solution $\hat{x}$ to the sketched problem (6) is δ-optimal to the original optimal solution x^* of (5) if

$$f(\hat{x}) \leq (1 + \delta)^2 f(x^*).$$

With Theorem II.1, II.4 and a similar argument as in [7, Theorem 1, Lemma 2, Lemma 3], we can find that if

$$m \gtrsim K^2 \log K \frac{w^2(B\mathcal{T} \cap \mathbb{S}^{n-1})}{\delta^2},$$

then with high probability, the solution to the sketched problem is δ-optimal. Here $\mathcal{T}$ is the tangent cone of $\mathcal{C}$ at optimum x^*. This guarantee improves the dependence on K from K^4 to $K^2 \log K$ when compared to the result in [7].

IV. PROOFS

A. A New Bernstein's Inequality

Lemma IV.1. *Let Y be a sub-exponential random variable satisfying $\mathbb{E}|Y| \leq 2$ and $\|Y\|_{\psi_2} \leq K^2$ $(K \geq 2)$. Then*

$$\mathbb{E}|Y|^p \leq C^p p^p \left(K^2 \log K\right)^{p-1}, \; \forall p \geq 1/3.$$

Proof. Define $f(t) := \mathbb{P}(|Y| \geq t) e^{t/K^2}$.
From $\mathbb{E}|Y| \leq 2$ we have

$$\int_0^\infty f(t)e^{-t/K^2} dt = \int_0^\infty \mathbb{P}(|Y| \geq t)dt \leq 2. \tag{7}$$

Also, from $\|Y\|_{\psi_1} \leq K^2$ and with a change of variable $s = e^{t/K^2}$ we have

$$2 \geq \mathbb{E}\exp(|Y|/K^2) = \int_0^\infty \mathbb{P}\left(e^{|Y|/K^2} \geq s\right) ds$$
$$= \int_{-\infty}^0 K^{-2}e^{t/K^2} dt + \int_0^\infty K^{-2}f(t)dt.$$

The first integral on the right hand side is 1, so

$$\int_0^\infty f(t)\,dt \leq K^2. \tag{8}$$

To bound the p-th moment of $|Y|$, first notice that

$$\mathbb{E}|Y|^p = \int_0^\infty \mathbb{P}(|Y|^p \geq s)ds = \int_0^\infty f(u)e^{-u/K^2}pu^{p-1}du.$$

Set $T = 6pK^2 \log K$ and split this integral into two parts:

$$\int_0^T f(u)e^{-u/K^2}pu^{p-1}du \leq pT^{p-1}\int_0^T f(u)e^{-u/K^2}du$$
$$\overset{(7)}{\leq} 2p\left(6pK^2 \log K\right)^{p-1};$$

$$\int_T^\infty f(u)e^{-u/K^2}pu^{p-1}du \leq pT^{p-1}e^{-T/K^2}\int_T^\infty f(u)du$$
$$\overset{(8)}{\leq} pT^{p-1}K^{-6p}K^2$$
$$\leq p\left(6pK^2 \log K\right)^{p-1}.$$

Here we used the fact that $u^{p-1}e^{-u/K^2}$ monotonically decreases on $[T, \infty)$. Combining these two parts completes the proof with $C \leq 6$. $\square$

With Lemma IV.1, the proof for Theorem II.4 essentially relies on the same argument for Bernstein's inequality for sub-exponential random variables [2, Theorem 2.8.1].

Proof outline for Theorem II.4. First bound the moment generating functions for Y_i through Taylor series and Lemma IV.1 to get

$$\mathbb{E}\exp(\lambda Y_i) \leq \exp\left(C\lambda^2 K_i^2 \log K_i\right) \text{ when } |\lambda|K_i^2 \log K_i \leq c.$$

Then for $u \geq 0$, apply Markov's inequality to get

$$\mathbb{P}\left(\sum_{i=1}^m a_i Y_i \geq u\right) \leq e^{-\lambda u}\mathbb{E}\exp\left(\lambda\sum_{i=1}^m a_i Y_i\right)$$

$$\leq \exp\left(-\lambda u + \lambda^2 C \sum_{i=1}^m a_i^2 K_i^2 \log K_i\right).$$

Finally, optimize λ over $\left[0, \frac{c}{\|a\|_\infty K^2 \log K}\right]$ to obtain one side of the bound. The bound for $\mathbb{P}\left(\sum a_i Y_i < -u\right)$ is similarly obtained by considering $-Y_i$ instead of Y_i. $\square$

B. Random Matrices on Sets

Proof of Theorem II.2. We will assume b is non-zero. Let

$$Y := \|b \circ X\|_2^2 - \|b\|_2^2 = \sum_{i=1}^m \left(b_i^2 X_i^2 - b_i^2\right).$$

By Theorem II.4 and $\sum b_i^4 \leq \|b\|_\infty^2 \|b\|_2^2$ we have $\mathbb{P}(Y \geq u)$ is bounded above by

$$\exp\left[-c\min\left(\frac{u^2}{\|b\|_\infty^2\|b\|_2^2 K^2 \log K}, \frac{u}{\|b\|_\infty^2 K^2 \log K}\right)\right].$$

Define $h(s) := \mathbb{P}(|\|b \circ X\|_2 - \|b\|_2| \geq s)$ and consider the following two cases:

1) $0 \leq s \leq \|b\|_2$. Setting $u = s\|b\|_2 \leq \|b\|_2^2$, we have

$$h(s) = \mathbb{P}(|Y| \geq s(\|b \circ X\|_2 + \|b\|_2))$$
$$\leq \mathbb{P}(|Y| \geq s\|b\|_2)$$
$$\leq 2\exp\left(-c\frac{s^2}{\|b\|_\infty^2 K^2 \log K}\right).$$

2) $s \geq \|b\|_2$. Setting $u = s^2 \geq \|b\|_2^2$, we have

$$h(s) = \mathbb{P}((\|b \circ X\|_2 - \|b\|_2)^2 \geq s^2)$$
$$\leq \mathbb{P}(|Y| \geq s^2)$$
$$\leq 2\exp\left(-c\frac{s^2}{\|b\|_\infty^2 K^2 \log K}\right).$$

Here the first inequality uses $(\alpha - \beta)^2 \leq |\alpha^2 - \beta^2|$ for $\alpha, \beta \geq 0$.

These show that $h(s)$ is bounded by the tail of a Gaussian whose standard deviation is in the order of $\|b\|_\infty K\sqrt{\log K}$; the bound for ψ_2 norm then follows. $\square$

To prove Theorem II.1, we will use the the following lemma. This lemma is analogous to Theorem 1.3 in [11], but with optimal dependence on K.

Lemma IV.2. *Let $A \in \mathbb{R}^{m \times n}$ be an isotropic and sub-gaussian matrix with parameter K, then the random process*

$$Z_x := \|Ax\|_2 - \sqrt{m}\|x\|_2$$

has sub-gaussian increments with

$$\|Z_x - Z_y\|_{\psi_2} \leq CK\sqrt{\log K}\|x - y\|_2, \quad \forall x, y \in \mathbb{R}^n.$$

Proof. The proof of this essentially the same as the proof for Theorem 1.3 in [11], except we use Theorem II.2 and Theorem II.4 whenever possible. We would also like to point out that the random variables $\langle A_i, u\rangle\langle A_i, v\rangle$ (as in equation (5.5) from [11]) have first absolute moment

$$\mathbb{E}|\langle A_i, u\rangle\langle A_i, v\rangle| \leq \frac{1}{2}\mathbb{E}\left(\langle A_i, u\rangle^2 + \langle A_i, v\rangle^2\right) \leq \frac{1+4}{2} = \frac{5}{2}$$

since $\|u\|_2 = 1$, $\|v\|_2 \leq 2$ and A_i is isotropic. So our new Bernstein's inequality (Theorem II.4) does apply. $\square$

Proof of Theorem II.1. This follows from Lemma IV.2 and Talagrand's Majorizing Measure Theorem [16], [19]. $\square$

The proof for Proposition II.3 is long and technical, so we only present the proof idea here.

Proof idea for Proposition II.3. Let $Z = \|X\|_2 - \sqrt{m}$. Then it suffices to show $\mathbb{E}\exp\left(Z^2/(cK^2 \log K)\right) > 2$. Write this expectation as an integral and notice that

$$\mathbb{P}(|Z| \geq u) \geq \mathbb{P}\left(\|X\|_2^2 \geq (\sqrt{m} + u)^2\right) \tag{9}$$

where $\frac{1}{K^2 \log K}\|X\|_2^2 \sim \text{Binomial}\left(m, \frac{1}{K^2 \log K}\right)$. We can then prove the proposition by bounding (9) from below through a lower bound for Binomial tails [20, Lemma 4.7.2]. $\square$

V. Acknowledgements

References

[1] G. Schechtman, "Two observations regarding embedding subsets of euclidean spaces in normed spaces," *Advances in Mathematics*, vol. 200, no. 1, pp. 125–135, 2006.

[2] R. Vershynin, *High-Dimensional Probability.* Cambridge University Press (to appear), 2017.

[3] F. Krahmer, D. Needell, and R. Ward, "Compressive sensing with redundant dictionaries and structured measurements," *SIAM Journal on Mathematical Analysis*, vol. 47, no. 6, pp. 4606–4629, 2015.

[4] S. Dirksen, "Dimensionality reduction with subgaussian matrices: a unified theory," *Foundations of Computational Mathematics*, vol. 16, no. 5, pp. 1367–1396, 2016.

[5] D. Achlioptas, "Database-friendly random projections: Johnson-lindenstrauss with binary coins," *Journal of computer and System Sciences*, vol. 66, no. 4, pp. 671–687, 2003.

[6] D. M. Kane and J. Nelson, "Sparser johnson-lindenstrauss transforms," *Journal of the ACM (JACM)*, vol. 61, no. 1, p. 4, 2014.

[7] M. Pilanci and M. J. Wainwright, "Randomized sketches of convex programs with sharp guarantees," *IEEE Transactions on Information Theory*, vol. 61, no. 9, pp. 5096–5115, 2015.

[8] D. P. Woodruff, "Sketching as a tool for numerical linear algebra," *Foundations and Trends® in Theoretical Computer Science*, vol. 10, no. 1–2, pp. 1–157, 2014.

[9] F. Krahmer and H. Rauhut, "Structured random measurements in signal processing," *GAMM-Mitteilungen*, vol. 37, no. 2, pp. 217–238, 2014.

[10] R. Saab, R. Wang, and Ö. Yılmaz, "From compressed sensing to compressed bit-streams: practical encoders, tractable decoders," *IEEE Transactions on Information Theory*, vol. 64, no. 9, pp. 6098–6114, 2018.

[11] C. Liaw, A. Mehrabian, Y. Plan, and R. Vershynin, "A simple tool for bounding the deviation of random matrices on geometric sets," in *Geometric Aspects of Functional Analysis*. Springer, 2017, pp. 277–299.

[12] S. Oymak and J. A. Tropp, "Universality laws for randomized dimension reduction, with applications," *Information and Inference: A Journal of the IMA*, vol. 7, no. 3, pp. 337–446, 2017.

[13] E. J. Candès, "Mathematics of sparsity (and a few other things)," in *Proceedings of the International Congress of Mathematicians, Seoul, South Korea*, vol. 123, 2014.

[14] B. Klartag and S. Mendelson, "Empirical processes and random projections," *Journal of Functional Analysis*, vol. 225, no. 1, p. 229, 2005.

[15] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann, "Reconstruction and subgaussian operators in asymptotic geometric analysis," *Geometric and Functional Analysis*, vol. 17, no. 4, pp. 1248–1282, 2007.

[16] S. Dirksen, "Tail bounds via generic chaining," *Electronic Journal of Probability*, vol. 20, 2015.

[17] K. G. Larsen and J. Nelson, "The johnson-lindenstrauss lemma is optimal for linear dimensionality reduction," in *43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy*, 2016, pp. 82:1–82:11.

[18] J. Matoušek, "On variants of the johnson–lindenstrauss lemma," *Random Structures & Algorithms*, vol. 33, no. 2, pp. 142–156, 2008.

[19] M. Talagrand, *Upper and lower bounds for stochastic processes: modern methods and classical problems*. Springer Science & Business Media, 2014, vol. 60.

[20] B. Robert, *Ash. Information Theory*. Dover Publications Inc., 1990.