

Random matrices acting on sets: Independent columns^{*}

Yaniv Plan[†]Roman Vershynin[‡]

Abstract

We study random matrices with independent subgaussian columns. Assuming each column has a fixed Euclidean norm, we establish conditions under which such matrices act as near-isometries when restricted to a given subset of their domain. We show that, with high probability, the maximum distortion caused by such a matrix is proportional to the Gaussian complexity of the subset, scaled by the subgaussian norm of the matrix columns. This linear dependence on the subgaussian norm is a new phenomenon, as random matrices with independent rows or independent entries typically exhibit superlinear dependence. As a consequence, normalizing the columns of random sparse matrices leads to stronger embedding guarantees.

Keywords: random matrix theory; restricted isometries; dimension reduction; random embeddings; sparse matrices.

MSC2020 subject classifications: 15A42; 60D05.

1 Introduction

Let $A \in \mathbb{R}^{m \times n}$ be a random matrix and $T \subset \mathbb{R}^n$ a fixed set. We study the question: *Under what circumstances is A well conditioned when restricted to T ?* More technically, we study the quantity

$$\delta := \sup_{x \in T} \left| \|Ax\|_2 - \|x\|_2 \right|.$$

Observe that the *isometry constant* δ is the smallest constant such that

$$\|x\|_2 - \delta \leq \|Ax\|_2 \leq \|x\|_2 + \delta \quad \text{for all } x \in T.$$

If δ is small, A embeds T into \mathbb{R}^m without significantly distorting Euclidean norms. If T is replaced by the Minkowski difference $T - T$, then the embedding approximately

^{*}R.V. is partially supported by NSF Grant DMS 2451011 and NSF+Simons Research Collaborations on the Mathematical and Scientific Foundations of Deep Learning; Y.P. is partially supported by an NSERC Discovery Grant (GR009284), an NSERC Discovery Accelerator Supplement (GR007657), and a Tier II Canada Research Chair in Data Science (GR009243).

[†]University of British Columbia, Canada. E-mail: yaniv@math.ubc.ca

[‡]University of California, Irvine, USA. E-mail: rvershyn@uci.edu

preserves Euclidean distances between points in T , so we can think of A as a *near isometry* on T . Random matrices that are near isometries are useful in a variety of applications, including compressed sensing [12, 11], dimension reduction as in the Johnson-Lindenstrauss lemma [14], and randomized sketching [27, 22]. The subset T of interest varies by application. When T belongs to the sphere and $\delta < 1$, the null space of A is a random subspace that must not intersect T , and thus the study of δ also has interesting implications regarding the behaviour of random subspaces [13, 19].

If A is a random Gaussian matrix, the size of δ is well understood [23]. It is controlled by the Gaussian width or Gaussian complexity of T [26].

Definition 1.1 (The Gaussian width and Gaussian complexity of a set). *The Gaussian width $w(T)$ and the Gaussian complexity $\gamma(T)$ of a set $T \subset \mathbb{R}^n$ are defined as follows:*

$$w(T) := \mathbb{E} \sup_{x \in T} \langle g, x \rangle, \quad \gamma(T) := \mathbb{E} \sup_{x \in T} |\langle g, x \rangle|$$

where g is a standard normal random vector in \mathbb{R}^n , i.e. $g \sim N(0, I_n)$.

These two quantities are identical for symmetric sets, and in general we have $w(T) + \text{rad}(T) \asymp \gamma(T)$, where $\text{rad}(T) := \sup_{x \in T} \|x\|_2$. For more properties of Gaussian width and complexity, see [26].

To extend our discussion to more general random matrix ensembles, recall that the *subgaussian norm* of a random variable Z is defined as

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

and Z is called *subgaussian* if $\|Z\|_{\psi_2} < \infty$. Further, the subgaussian norm of a random vector $V \in \mathbb{R}^n$ is defined as

$$\|V\|_{\psi_2} := \sup_{x \in S^{n-1}} \|\langle V, x \rangle\|_{\psi_2} \quad (1.1)$$

i.e. it is the largest subgaussian norm of one-dimensional marginals of V . For basic introduction to subgaussian distributions, see [26].

Matrices with subgaussian entries (or rows or columns) represent a notable generalization beyond the Gaussian case. In particular, this generalization includes discrete and/or sparse random matrices, which can be used to accelerate computations [16] and are essential in applications like group testing [8] and wireless network activity detection [18].

When A has independent, isotropic,¹ and subgaussian rows, the quantity δ has similar guarantees to the Gaussian case as shown in the following theorem.

Theorem 1.2 (Corollary 1.2 in [13]). *Let A be an $m \times n$ matrix whose rows are independent, isotropic, subgaussian random vectors in \mathbb{R}^n with subgaussian norms bounded by K . Then, for any subset $T \subset \mathbb{R}^n$, we have*

$$\mathbb{E} \sup_{x \in T} \left| \|Ax\|_2 - \sqrt{m} \|x\|_2 \right| \leq CK \sqrt{\log K} \gamma(T).$$

Moreover, for any $u \geq 0$, with probability at least $1 - 3e^{-u^2}$, we have

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

Here and throughout the paper, $C > 0$ denotes an absolute constant.

¹Recall that a random vector $V \in \mathbb{R}^n$ is called *isotropic* if $\mathbb{E} V V^\top = I_n$ where I_n is the identity matrix.

This theorem builds directly on the work in [19], which presents the same result but with $K\sqrt{\log K}$ replaced by K^2 . In fact, the $K\sqrt{\log K}$ dependence is known to be optimal in this setting [13]. Similar preceding results also appear in the papers [17, 21, 10]; see [19, Section 3] for a discussion of their relationship.

In this paper, we develop an analogous theory for matrices with independent subgaussian columns. While there are several studies of random matrices with independent subgaussian columns restricted to specific subsets—such as the set of sparse vectors [2], finite sets [9], or the entire sphere [25]—we are only aware of one result that allows arbitrary sets T , due to D. Bartl and S. Mendelson [4]. We now state our main theorem and corollary, and then compare to the results in [4] in Remark 1.7 below.

Theorem 1.3. *Let A be an $m \times n$ matrix whose columns A_i are independent, mean zero, subgaussian random vectors in \mathbb{R}^m satisfying $\|A_i\|_2 = 1$ almost surely and $\|A_i\|_{\psi_2} \leq K$. Then, for any subset $T \subset \mathbb{R}^n$, we have*

$$\mathbb{E} \sup_{x \in T} \left| \|Ax\|_2 - \|x\|_2 \right| \leq CK\gamma(T).$$

Moreover, for any $u \geq 0$, with probability at least $1 - 3e^{-u^2}$, we have

$$\sup_{x \in T} \left| \|Ax\|_2 - \|x\|_2 \right| \leq CK [w(T) + u \cdot \text{rad}(T)].$$

A trivial rescaling argument yields the following:

Corollary 1.4. *Let A be an $m \times n$ matrix whose columns A_i are independent, mean zero, subgaussian random vectors in \mathbb{R}^m satisfying $\|A_i\|_2 = \lambda$ almost surely and $\|A_i\|_{\psi_2} \leq K$. Then, for any subset $T \subset \mathbb{R}^n$, we have*

$$\mathbb{E} \sup_{x \in T} \left| \|Ax\|_2 - \lambda \|x\|_2 \right| \leq CK\gamma(T).$$

Moreover, for any $u \geq 0$, with probability at least $1 - 3e^{-u^2}$, we have

$$\sup_{x \in T} \left| \|Ax\|_2 - \lambda \|x\|_2 \right| \leq CK [w(T) + u \cdot \text{rad}(T)].$$

Remark 1.5 (Why is column normalization needed?). In contrast to the previous result (Corollary 1.2), here we require a hard normalization of columns:

$$\|A_i\|_2 = \lambda \quad \text{a.s. for all } i. \tag{1.2}$$

One cannot remove this assumption even for $n = 1$, or replace it with a weaker assumption of isotropy as in Theorem 1.2. This is because a mean-zero isotropic subgaussian random vector A_1 might vanish with probability $1/2$. For example, if $\delta \sim \text{Bern}(1/2)$ and X is (independently) chosen uniformly from $\sqrt{m}S^{m-1}$, then δX is a subgaussian vector with $O(1)$ subgaussian norm. If the columns of A are i.i.d. copies of δX and $T = \{e_1\}$, the first canonical basis vector, then

$$\sup_{x \in T} \left| \|Ax\|_2 - \lambda \|x\|_2 \right| = |\delta\sqrt{m} - \lambda| \geq \sqrt{m}/2$$

with probability $1/2$ regardless of the value of λ . In other words, both the probabilistic bound and bound in expectation become substantially dependent on m without the normalization assumption on the columns.

Remark 1.6 (Column normalization improves the dependence on the subgaussian norm). The hard normalization of columns has an interesting beneficial effect on the dependence on the subgaussian norm. In the previous result (Theorem 1.2) the dependence is $O(K\sqrt{\log K})$, and this dependence is optimal even for matrices with all independent entries [13]. But as soon as we normalize the columns as in Corollary 1.4, the dependence becomes linear, i.e. $O(K)$. We see in Section 2 that this small improvement leads to completely different asymptotic behaviour for the embedding guarantees of sparse matrices. We also observe in Remark 2.5 that this improvement only occurs for matrices with independent *columns* of a fixed norm, not for matrices with independent *rows* of a fixed norm.

Remark 1.7 (Comparison to [4]). We now compare to the results in [4]. Therein, the authors require the columns of A to satisfy: 1) a “thin-shell condition”, i.e., to have concentrated norms, and 2) a moment condition on the marginals. This allows a broad class of distributions, including subgaussian. We specialize their results to the setting of Corollary 1.4 with $\lambda = \sqrt{m}$, and also assume that a) the columns of A are isotropic, b) $T \subset S^{n-1}$, and c) $\gamma(T) \geq \sqrt{\log n}$ to allow a simpler comparison. In this case, the main result in [4] (Theorem 1.5), yields the following. For $u \geq 1$, with probability at least $1 - 2\exp(-c\gamma(T)^2) - n^{-u}$,

$$\sup_{x \in T} \left| \|Ax\|_2^2 - m\|x\|_2^2 \right| \leq C(K, u) \cdot \sqrt{m} [\gamma(T)(1 + \gamma(T)/\sqrt{m})] \log(en/\gamma^2(T)).$$

Note that this is a uniform bound on the squared difference $\|Ax\|_2^2 - m\|x\|_2^2 = (\|Ax\|_2^2 - \sqrt{m}\|x\|_2)(\|Ax\|_2 + \sqrt{m}\|x\|_2)$ thus explaining the extra factor of \sqrt{m} . Further, in typical cases of interest $\gamma(T) < \sqrt{m}$. Thus, it is natural to compare $C(K, \mu)\gamma(T) \log(en/\gamma^2(T))$ to the error bound given in Corollary 1.4, which takes the form $CK\gamma(T)$. In other words, by concentrating on the subgaussian case, we are able to improve on the results in [4] by giving the explicit and optimal *linear* dependence on the subgaussian norm K , and also removing the $\log(en/\gamma(T)^2)$ factor — thereby showing that there is no dependence on the ambient dimension.

We now specify some notations, discuss applications to sparse matrices in Section 2, show that column normalization can improve embedding guarantees for i.i.d. matrices in Section 3, prove our main result in Section 4, then our auxiliary results in Section 5, and then conclude with an open question in Section 6.

1.1 Notation

We use $\|\cdot\|_2$ for the Euclidean norm of vectors, and S^{n-1} for the set of vectors in \mathbb{R}^n with Euclidean norm 1. We say that $f \lesssim g$ if there exists an absolute constant C such that $f \leq Cg$ everywhere. The symbols C, c generally refer to absolute constants which may vary from instance to instance. We say $f \asymp g$ if $cf \leq g \leq Cf$, i.e., f and g are equivalent up to constants.

2 An example: sparse matrices

Sparse random matrices are a central object of study in the context of sparse Johnson Lindenstrauss embeddings [1] and sparse subspace embeddings [7]. These give a method to facilitate computation by embedding data into a lower dimension near isometrically via a sparse (and thus low complexity) transform. The central question is: *How should we choose the distribution of A to (a) minimize the number of nonzero entries and (b) ensure it behaves like a near-isometry?*

An interesting phenomenon has been observed (and understood) in this field: Choosing A with i.i.d. entries limits its ability to near-isometrically embed sparse vectors

[20, 3]. However, choosing the columns from a distribution with exactly s non-zero entries per column can give a significant improvement. The literature in this area generally considers embeddings of singletons, finite sets, or subspaces, with the notable exception in [5], discussed further below. We extend the understanding of this phenomenon to arbitrary bounded sets in the following. We begin with an illustrative example to give an intuition for this point.

2.1 Approximate vs. exact sparsity

Consider two different models to generate a sparse $m \times n$ matrix with (about) s nonzero entries per column. In the first, “approximate” model, we sample each entry independently from the $\text{Ber}(s/m)$ distribution. In the second, “exact” model, we put s ones uniformly at random in each column of the matrix. Although individual entries in both models have the same $\text{Ber}(s/m)$ distribution, it turns out that these two models have diverging behaviour as m increases. In the first model, the number of nonzero entries in a column converges to a $\text{Poisson}(s)$ distribution. Thus, the column norms will concentrate around \sqrt{s} , but their deviations will not decrease to zero as m increases. Equivalently, the matrix will randomly distort the norm of a standard basis vector, and *that distortion will never decrease to zero*. In contrast, in the second model, the number of nonzero entries per column is exactly s and so the column norms are exactly \sqrt{s} . Further, as m increases, the columns become more and more likely to have disjoint support and thus become orthogonal. Thus, the resulting matrix (multiplied by $1/\sqrt{s}$ for normalization) asymptotically becomes an *exact isometry* with probability approaching 1 as m increases.

The following two results make these observations quantitative and general.

Corollary 2.1 (Approximately s -sparse columns). *Let A be an $m \times n$ matrix whose entries are independent random variables which take the value 1 or -1 with probability $\frac{s}{2m}$ each, and 0 with probability $1 - \frac{s}{m}$. Then, for any subset $T \subset \mathbb{R}^n$, we have*

$$\mathbb{E} \sup_{x \in T} \left| \|Ax\|_2 - \sqrt{s} \|x\|_2 \right| \leq C\gamma(T).$$

Moreover, this bound is optimal in general. For any integers $1 \leq s \leq m/2$ and $n \geq 1$ there exists a non-empty subset $T \subset \mathbb{R}^n$ such that

$$\mathbb{E} \sup_{x \in T} \left| \|Ax\|_2 - \sqrt{s} \|x\|_2 \right| \geq c\gamma(T).$$

Note that the right-hand side does not decrease when m grows.

Proof. Let A_i denote the columns of A . It is not hard to check (see Lemma 5.1) that

$$\|A_i\|_{\psi_2} \lesssim \frac{1}{\sqrt{\log(2/p)}} \quad \text{where } p = \frac{s}{m}.$$

If we normalize A_i by setting $Z_i = A_i/\sqrt{p}$, then all coordinates of Z_i are independent random variables with mean zero and variance 1. Hence Z_i are independent and isotropic random vectors with subgaussian norms bounded by $K = 1/\sqrt{p \log(2/p)}$. Thus, applying Theorem 1.2 for the matrix A/\sqrt{p} , we obtain

$$\mathbb{E} \sup_{x \in T} \left| \frac{1}{\sqrt{p}} \|Ax\|_2 - \sqrt{m} \|x\|_2 \right| \lesssim K \sqrt{\log K} \gamma(T) \lesssim \frac{\gamma(T)}{\sqrt{p}}.$$

Multiply by \sqrt{p} on both sides to obtain the first part of the theorem.

To prove the second part, choose $T = \{e_1\}$, the first vector of the standard basis in \mathbb{R}^n . Then

$$\mathbb{E} \sup_{x \in T} \left| \|Ax\|_2 - \sqrt{s} \|x\|_2 \right| = \mathbb{E} \left| \sqrt{Z} - \sqrt{s} \right| \quad (\text{where } Z \sim \text{Binom}(m, p)).$$

Since $\gamma(T) \asymp 1$, it suffices to prove that $\mathbb{E} \left| \sqrt{Z} - \sqrt{s} \right| \gtrsim 1$. A quick computation² gives

$$\mathbb{E}(Z - s)^2 = mp(1 - p) \geq \frac{s}{2} \quad \text{and} \quad \mathbb{E}(Z - s)^4 \lesssim s^2.$$

Then using Paley-Zygmund inequality for $(Z - s)^2$, we get

$$\mathbb{P} \left\{ |Z - s| \geq \frac{\sqrt{s}}{2} \right\} = \mathbb{P} \left\{ (Z - s)^2 \geq \frac{1}{2} \mathbb{E}(Z - s)^2 \right\} > c. \quad (2.1)$$

Furthermore, since $\mathbb{E} Z = mp = s$, Markov inequality gives

$$\mathbb{P}\{Z \leq 2s/c\} \geq 1 - c/2.$$

Combining this with (2.1) by the union bound, we conclude that with probability at least $c/2$ we have both

$$|Z - s| \gtrsim \sqrt{s} \quad \text{and} \quad Z \lesssim s.$$

If this event occurs, we have

$$\left| \sqrt{Z} - \sqrt{s} \right| = \frac{|Z - s|}{\sqrt{Z} + \sqrt{s}} \gtrsim 1.$$

Thus $\mathbb{E} \left| \sqrt{Z} - \sqrt{s} \right| \gtrsim (c/2) \cdot 1 \gtrsim 1$, as required to complete the proof.

The reader may be wondering whether a lower bound can also be shown with sets T for which $\gamma(T)$ is not constant. Indeed, observe that $\sqrt{Z} - \sqrt{s}$ converges to a $N(0, 1/4)$ random variable as $m \rightarrow \infty$ and $s \rightarrow \infty$ provided $s/m \rightarrow 0$ (this can be shown via the Berry-Esseen central limit theorem together with a Taylor approximation). This can be used as a first step to verify our matching upper and lower bounds for the set $T = \{e_1, e_2, \dots, e_n\}$ when m and s are appropriately large. \square

Corollary 2.2 (Exactly s -sparse columns). *Let A be an $m \times n$ matrix whose columns are independent random vectors drawn uniformly from the set of vectors in \mathbb{R}^m that have all $0, -1, 1$ entries and exactly s nonzero entries. Then, for any subset $T \subset \mathbb{R}^n$, we have*

$$\mathbb{E} \sup_{x \in T} \left| \|Ax\|_2 - \sqrt{s} \|x\|_2 \right| \leq \frac{C\gamma(T)}{\sqrt{\log(2m/s)}}.$$

Moreover, this bound is optimal in general. For any integers $1 \leq s \leq m$ there exists $n = n(m, s)$ and a non-empty subset $T \subset \mathbb{R}^n$ such that

$$\mathbb{E} \sup_{x \in T} \left| \|Ax\|_2 - \sqrt{s} \|x\|_2 \right| \geq \frac{c\gamma(T)}{\sqrt{\log(2m/s)}}.$$

Note that the right-hand side decreases when m grows.

²To prove the second identity, express $Z - s = \sum_{i=1}^m X_i$ where X_i are independent centered $\text{Ber}(p)$ random variables. Expand the fourth power of the sum and use that $\mathbb{E} X_i = 0$ to get $\mathbb{E}(Z - s)^4 \lesssim m \mathbb{E} X_1^4 + m^2 (\mathbb{E} X_1^2)^2 \lesssim mp + m^2 p^2 \lesssim s^2$.

Proof. It is not hard to check (see Lemma 5.1) that the columns A_i of A satisfy

$$\|A_i\|_2 = \sqrt{s} \text{ a.s. and } \|A_i\|_{\psi_2} \lesssim \frac{1}{\sqrt{\log(2m/s)}}.$$

Apply Corollary 1.4 to obtain the first part of the theorem.

To prove the second part, choose

$$T = \{e_1 - e_2, e_1 - e_3, \dots, e_1 - e_n\}$$

where e_1, \dots, e_n is the standard basis of \mathbb{R}^n . Then

$$\begin{aligned} \mathbb{P}\{Ax = 0 \text{ for some } x \in T\} &= \mathbb{P}\{A_1 = A_i \text{ for some } i = 2, \dots, n\} \\ &= 1 - (1 - \mathbb{P}\{A_1 = A_2\})^{n-1}. \end{aligned}$$

$$\begin{aligned} \mathbb{P}\{A_1 = A_2\} &= \frac{1}{2^s \binom{m}{s}} \\ &\geq \left(\frac{2em}{s}\right)^{-s} \quad (\text{by a standard bound [26, Exercise 0.0.5]}) \\ &\geq \frac{1}{n} \quad (\text{if we choose } n = \left\lceil \left(\frac{2em}{s}\right)^{3s} \right\rceil). \end{aligned}$$

Then

$$\mathbb{P}\{Ax = 0 \text{ for some } x \in T\} \geq 1 - \left(1 - \frac{1}{n}\right)^{n-1} \geq c.$$

When $Ax = 0$ for some $x \in T$, we have $\left|\|Ax\|_2 - \sqrt{s}\|x\|_2\right| = \sqrt{2s}$. It follows that

$$\mathbb{E} \sup_{x \in T} \left|\|Ax\|_2 - \sqrt{s}\|x\|_2\right| \geq c\sqrt{2s}. \quad (2.2)$$

On the other hand, denoting by g_i independent $N(0, 1)$ random variables, we have

$$\gamma(T) = \mathbb{E} \max_{i=2, \dots, n} |g_1 - g_i| \lesssim \sqrt{\log n} \lesssim \sqrt{s \log(2m/s)}. \quad (2.3)$$

where the first inequality follows from the standard bound on the maximum of subgaussian random variables [26, Exercise 2.5.10], and the second inequality follows by our choice of n . Comparing (2.2) and (2.3), we complete the proof. \square

Remark 2.3 (Diverging asymptotic behaviour). As shown in Corollaries 2.1 and 2.2, in general as $m \rightarrow \infty$ the isometry constant for sparse matrices with i.i.d. entries remains lower bounded, whereas the isometry constant for sparse matrices with exactly s non-zeros per column decreases to 0.

Remark 2.4 (Alternative treatments of isometry constants of sparse matrices). We point out three different treatments of the isometry constants of sparse matrices depending on the structure of T :

1. There is a rich understanding of the isometric properties of sparse matrices in the literature on sparse Johnson Lindenstrauss embeddings (see e.g., [16]) when T is a singleton (or a finite set).
2. When T is a d -dimensional subspace, again there is a thorough understanding in the literature on sparse (oblivious) subspace embeddings. Here it is shown that one only needs $m = O(d)$ to have an isometry constant lower than 1, even if the sparsity is polylogarithmic (see e.g., [6]) – a result which does not follow from Corollary 2.2.

3. We are aware of one paper, [5], which addresses the isometric properties of sparse matrices for arbitrary subsets of the sphere. The main theorem in [5] controls the isometry constant by a new quantity, which is akin to a sparsified gaussian complexity. We don't repeat that theorem here because the conditions are relatively complex, but note that it recovers the results on sparse subspace embeddings (up to log factors) as a special case. Thus, the machinery in [5] allows tighter control of the isometry constant than Corollary 2.2 for some sets and parameter regimes. However, the results in [5] do not imply that the restricted isometry constant converges to 0 as $m \rightarrow \infty$ and thus do not imply the diverging asymptotic behaviour of "approximately sparse" and "exactly sparse" matrices shown in our work and discussed in the remark above. Further, we note that our work does not require T to be a subset of the sphere, and is oblivious to every property of T aside from its gaussian complexity.

Remark 2.5 (Exactly s -sparse rows do not give the same benefit). The reader may be wondering whether the improvements gained by choosing the columns from a distribution with exactly s non-zeros could alternatively be gained by choosing the rows from such a distribution. In fact, this is not the case. For instance, suppose A is a square $m \times m$ matrix whose rows are independently drawn from the set of vectors that have all 0, -1 , 1 entries and exactly s nonzero entries. Then, the first column (or any other) of A has entries which are independent random variables which take the value 1 or -1 with probability $\frac{s}{2m}$ each, and 0 with probability $1 - \frac{s}{m}$. The lower bound in Corollary 2.1 then applies, since it is realized by the set $T = \{e_1\}$ which isolates the first column.

3 Normalizing the columns of an i.i.d. matrix

We now consider random matrices with independent, symmetric, subgaussian entries. We show that normalizing the columns improves the dependence of the isometry constant on the subgaussian norm. We assume the entries have variance 1 which implies that the rows are isotropic, thus fitting the models in [19, 13]. This causes the columns to concentrate around \sqrt{m} and so we normalize them onto $\sqrt{m}S^{n-1}$ to more easily compare to the results in [13]. We need to condition on the event that the columns have lower-bounded Euclidean norm as otherwise the normalization step could blow up the columns by an arbitrary amount. This is encoded in the event F below.

Corollary 3.1. *Let A be an $m \times n$ matrix whose entries are independent, symmetric, with unit variance, and subgaussian norm bounded by K . Assume $m \geq CK^2 \log(K) \cdot \log(n)$.³ Let A_i be the i th column of A . Let F be the event that all columns of A have Euclidean norm lower bounded by $\sqrt{m}/2$, that is,*

$$F := \left\{ \min_{i=1,2,\dots,n} \|A_i\|_2 \geq \frac{\sqrt{m}}{2} \right\}.$$

Then $\mathbb{P}\{F\} \geq 1 - 2\exp(-Cm/(K^2 \log K))$. Further, conditional on F , consider the column-normalized matrix \tilde{A} with columns

$$\tilde{A}_i := \frac{\sqrt{m}}{\|A_i\|_2} \cdot A_i, \quad i = 1, 2, \dots, n.$$

The matrix \tilde{A} satisfies

$$\mathbb{E} \left[\sup_{x \in T} \left| \|\tilde{A}x\|_2 - \sqrt{m}\|x\|_2 \right| \middle| F \right] \leq CK\gamma(T).$$

³The assumption that $m \geq CK^2 \log(K) \cdot \log(n)$ could be replaced with $m \geq CK^2 \log(K)$ at the expense of losing the lower bound on $\mathbb{P}(F)$.

Moreover, conditional on F , with probability at least $1 - 3 \exp(-u^2)$.

$$\sup_{x \in T} \left| \|\tilde{A}x\|_2 - \sqrt{m}\|x\|_2 \right| \leq CK [w(T) + u \cdot \text{rad}(T)]. \quad (3.1)$$

Note that one may generate the matrices A, \tilde{A} conditionally on F by re-sampling any columns of A whose Euclidean norms are below $\sqrt{m}/2$. Further, since $\mathbb{P}\{F\} \geq 1 - 2 \exp(-Cm/(K^2 \log K))$, the matrix \tilde{A} is well defined and Equation (3.1) holds unconditionally with probability at least $(1 - 3 \exp(-u^2)) \cdot \mathbb{P}\{F\} \geq 1 - 3 \exp(-u^2) - 2 \exp(-Cm/(K^2 \log K))$.

Remark 3.2 (Effect of normalizing columns). Observe that the rows of A are independent and isotropic. Thus, we are in the setting of [13, Corollary 1.2] (see Theorem 1.2 in Section 1), in which the authors show the unnormalized version of A satisfies Equation (3.1) but with $K\sqrt{\log K}$ dependence on the subgaussian norm, and that dependence is tight [13, Proposition 4.5]. Our corollary shows that normalizing columns removes the logarithmic factor, improving the dependence to linear.

As an application of this corollary, we show that normalizing the columns of an “approximately sparse” matrix improves the embedding guarantees to those of an “exactly sparse” matrix (see Section 2). This follows from Corollary 3.1 after rescaling the columns to have norm \sqrt{s} .

Corollary 3.3. *Let $m \geq s \geq C \log(n)$.⁴ Let A be an $m \times n$ matrix whose entries are independent random variables which take the value 1 or -1 with probability $\frac{s}{2m}$ each, and 0 with probability $1 - \frac{s}{m}$. Let F be the event that all columns of A have Euclidean norm lower bounded by $\sqrt{s}/2$, that is,*

$$F := \left\{ \min_{i=1,2,\dots,n} \|A_i\|_2 \geq \frac{\sqrt{s}}{2} \right\}.$$

Then $\mathbb{P}\{F\} \geq 1 - 2 \exp(-Cs)$. Further, conditional on F , consider the normalized matrix \tilde{A} with columns

$$\tilde{A}_i := \frac{\sqrt{s}}{\|A_i\|_2} A_i, \quad i = 1, 2, \dots, n.$$

Conditional on F , \tilde{A} satisfies

$$\mathbb{E} \left[\sup_{x \in T} \left| \|\tilde{A}x\|_2 - \sqrt{s}\|x\|_2 \right| \middle| F \right] \leq \frac{C}{\sqrt{\log(1 + m/s)}} \gamma(T). \quad (3.2)$$

Remark 3.4. By comparing to Corollary 2.1, we see that normalizing the columns has the same beneficial effect as choosing the columns to be exactly s -sparse: Without normalizing, the isometry constant may remain lower bounded as $m \rightarrow \infty$, but after normalizing the isometry constant shrinks to 0 as $m \rightarrow \infty$.

4 Proof of Theorem 1.3

As in our work [19], whose exposition can be found in [26, Section 9.1], Theorem 1.3 follows automatically from Talagrand’s majorizing measure theorem and the following result:

Theorem 4.1 (Subgaussian increments). *Let A be a random matrix satisfying the assumptions of Theorem 1.3. Then the random process*

$$Z_x := \|Ax\|_2 - \|x\|_2$$

⁴The assumption that $s \geq C \log(n)$ could be replaced with $s \geq 1$ at the expense of losing the lower bound on $\mathbb{P}(F)$.

has subgaussian increments, namely

$$\|Z_x - Z_y\|_{\psi_2} \leq CK\|x - y\|_2 \quad \text{for all } x, y \in \mathbb{R}^n.$$

To see how this leads to Theorem 1.3, let us show how to combine Theorem 4.1 with the specialized version of Talagrand's comparison inequality below. This result is from [19]; for more general versions, see [26, Section 8].

Theorem 4.2 (Comparison inequality). *Let $(Z_x)_{x \in T}$ be a random process on a bounded set $T \subset \mathbb{R}^n$. Assume that the process has subgaussian increments, that is there exists $M \geq 0$ such that*

$$\|Z_x - Z_y\|_{\psi_2} \leq M\|x - y\|_2 \quad \text{for all } x, y \in T.$$

Then

$$\mathbb{E} \sup_{x, y \in T} |Z_x - Z_y| \leq CM \mathbb{E} \sup_{x \in T} \langle g, x \rangle,$$

where $g \sim N(0, I_n)$. Moreover, for any $u \geq 1$, the event

$$\sup_{x, y \in T} |Z_x - Z_y| \leq CM \left(\mathbb{E} \sup_{x \in T} \langle g, x \rangle + u \cdot \text{diam}(T) \right)$$

holds with probability at least $1 - e^{-u^2}$. Here $\text{diam}(T) := \sup_{x, y \in T} \|x - y\|_2$.

The first part of Theorem 4.2 can be found in [24, Theorem 2.4.12] and the second part can be found in [10, Theorem 3.2]. See also [26, Theorem 8.5.5].

Theorem 1.3 then follows by observing that for any $x_0 \in T$,

$$\sup_{x \in T} |Z_x| \leq |Z_{x_0}| + \sup_{x \in T} |Z_x - Z_{x_0}| \leq |Z_{x_0}| + \sup_{x, y \in T} |Z_x - Z_y|.$$

The second summand is controlled by Theorems 4.2 and 4.1. The first summand is a single subgaussian random variable. Indeed, by Proposition 4.3 below, $\|Z_{x_0}\|_{\psi_2} \lesssim K\|x_0\|_2 \leq K \text{rad}(T)$. It is relatively straightforward to complete the proof of Theorem 1.3, see [13, Section 4] for the full detailed argument.

We now build the proof of Theorem 4.1, starting with two special cases.

4.1 Unit vector x , zero vector y

Let us first prove a partial case of Theorem 4.1 for $\|x\|_2 = 1$ and $y = 0$.

Proposition 4.3 (Concentration of norm). *Let A be a random matrix satisfying the assumptions of Theorem 1.3. Then for any $x \in S^{n-1}$ we have*

$$\left| \|Ax\|_2 - 1 \right|_{\psi_2} \lesssim K.$$

To prove this result, let us first note that Ax is a subgaussian random vector:

Lemma 4.4 (Sum of independent subgaussians). *Let X_1, \dots, X_n be independent, mean zero, subgaussian random vectors in \mathbb{R}^m . Then*

$$\left\| \sum_{i=1}^n X_i \right\|_{\psi_2}^2 \leq C \sum_{i=1}^n \|X_i\|_{\psi_2}^2.$$

In particular, under the assumptions of Theorem 1.3, for any unit vector $x \in \mathbb{R}^n$ we have

$$\|Ax\|_{\psi_2} \leq CK.$$

Proof of Lemma 4.4. The first part is an easy consequence of Definition (1.1) and [26, Proposition 2.6.1]. The second part follows if we expand $Ax = \sum_{i=1}^n x_i A_i$ and apply the first part to $X_i = x_i A_i$. \square

We now proceed with a moment generating function (MGF) bound. For any fixed $\lambda \in \mathbb{R}$, consider the function

$$\Phi_\lambda(z) := \exp(\lambda z). \quad (4.1)$$

Then the MGF of a random variable Z can be expressed as $\mathbb{E} \Phi_\lambda(Z)$.

Lemma 4.5. *Let A and x be as in Proposition 4.3. Set $S := \|Ax\|_2^2 - 1$. Then*

$$\mathbb{E} \Phi_\lambda(S) \leq \exp(C\lambda^2 K^2) \quad \text{for } |\lambda| \leq \frac{c}{K^2}.$$

Proof. Expanding the norm, we get

$$\|Ax\|_2^2 = \left\| \sum_{i=1}^n x_i A_i \right\|_2^2 = \sum_{i=1}^n x_i^2 \|A_i\|_2^2 + \sum_{i \neq j} x_i x_j \langle A_i, A_j \rangle.$$

By assumption we have $\|A_i\|_2 = 1$ and $\|x\|_2 = 1$. Thus,

$$\sum_{i=1}^n x_i^2 \|A_i\|_2^2 = \sum_{i=1}^n x_i^2 = 1$$

and so

$$S = \sum_{i \neq j} x_i x_j \langle A_i, A_j \rangle. \quad (4.2)$$

Since $\|A_i\|_2 = 1$ by assumption, using Cauchy-Schwarz we can see that S is a bounded random variable, so $\Phi_\lambda(S)$ is finite for all λ . Let's now get a quantitative bound.

Fix any $\lambda \in \mathbb{R}$ and let A' be an independent copy of A . Then

$$\begin{aligned} \mathbb{E} \Phi_\lambda(S) &= \mathbb{E} \Phi_\lambda \left(\sum_{i \neq j} x_i x_j \langle A_i, A_j \rangle \right) \\ &\leq \mathbb{E} \Phi_\lambda \left(4 \sum_{i,j=1}^n x_i x_j \langle A_i, A'_j \rangle \right) \quad (\text{decoupling [26, Exercise 6.1.4]}) \\ &= \mathbb{E} \Phi_\lambda (4 \langle Ax, A'x \rangle). \end{aligned}$$

Let us now condition on A . Recall, by Lemma 4.4, $\|A'x\|_{\psi_2} \leq CK$. Then the random variable $\langle Ax, A'x \rangle$ is (conditionally) subgaussian, and its conditional subgaussian norm is bounded by $CK\|Ax\|_2$, see [26, Definition 3.4.1]. Thus

$$\begin{aligned} \mathbb{E} \Phi_\lambda (4 \langle Ax, A'x \rangle) &= \mathbb{E}_A \mathbb{E}_{A'} \exp(4\lambda \langle Ax, A'x \rangle) \\ &\leq \mathbb{E}_A \exp(C\lambda^2 K^2 \|Ax\|_2^2) \quad (\text{by Equation [26, (2.16)]}). \end{aligned}$$

We have shown that

$$\mathbb{E} \Phi_\lambda(S) \leq \mathbb{E} \exp(C\lambda^2 K^2 \|Ax\|_2^2).$$

Recall that $S := \|Ax\|_2^2 - 1$. We now replace $\|Ax\|_2^2$ with $S + 1$ in the equation above to yield a recursive equation for the moment generating function of S . This gives

$$\mathbb{E} \Phi_\lambda(S) \leq \exp(C\lambda^2 K^2) \cdot \mathbb{E} \exp(C\lambda^2 K^2 S) \quad \text{for all } \lambda \in \mathbb{R}. \quad (4.3)$$

Now assume that

$$0 \leq \lambda \leq 1/(2CK^2).$$

Then $0 \leq C\lambda K^2 \leq 1/2$, and Jensen inequality gives

$$\mathbb{E} \exp(C\lambda^2 K^2 S) = \mathbb{E} [\exp(\lambda S)^{C\lambda K^2}] \leq [\mathbb{E} \exp(\lambda S)]^{C\lambda K^2} = [\mathbb{E} \Phi_\lambda(S)]^{C\lambda K^2}.$$

Insert this into Equation 4.3 and rearrange the terms to get

$$\mathbb{E} \Phi_\lambda(S) \leq \exp(C\lambda^2 K^2)^{1/(1-C\lambda K^2)}.$$

Since $1 - C\lambda K^2 \geq 1/2$ by assumption, we conclude that

$$\mathbb{E} \Phi_\lambda(S) \leq \exp(2C\lambda^2 K^2) \quad \text{for all } 0 \leq \lambda \leq 1/(2CK^2), \quad (4.4)$$

as desired.

Finally, let us lift the nonnegativity assumption and simply assume that

$$|\lambda| \leq 1/(2CK^2).$$

We have

$$\begin{aligned} \mathbb{E} \exp(C\lambda^2 K^2 S) &= \mathbb{E} \Phi_\mu(S) \quad (\text{where } \mu = C\lambda^2 K^2) \\ &\leq \exp(2C\mu^2 K^2) \quad (\text{using (4.4), since } 0 \leq \mu \leq 1/(2CK^2)) \\ &\leq \exp(2C\lambda^2 K^2) \quad (\text{since } \mu \leq |\lambda|). \end{aligned}$$

Substitute this into (4.3) to get

$$\mathbb{E} \Phi_\lambda(S) \leq \exp(3C\lambda^2 K^2) \quad \text{for all } |\lambda| \leq 1/(2CK^2).$$

The proof is complete. \square

Remark 4.6 (Why normalized columns give linear dependence on K). We can now point out the first place where the assumption of normalized columns allows linear dependence on K in our main result. In the proof above,

$$S = \|Ax\|_2^2 - 1 = \sum_{i=1}^n x_i^2 (\|A_i\|_2^2 - 1) + \sum_{i \neq j} x_i x_j \langle A_i, A_j \rangle.$$

Since the columns of A have norm 1, the first term disappears, leaving just a linear combination of the inner products of different columns (see Equation (4.2)). After decoupling, this gives the neat recursive equation (4.3) for S which ultimately allows a tight control of its tail. If we had not assumed normalized columns then the variations in the norms of A_i would have necessarily not allowed this (as follows from [13, Proposition 4.5]).

The following lemma converts the bound on the moment generating function of S into a tail bound.

Lemma 4.7. *Let S be as in Lemma 4.5. Then for any $t \geq 0$*

$$\mathbb{P} \{|S| \geq t\} \leq 2 \exp \left(-\frac{C}{K^2} (t^2 \wedge t) \right).$$

Proof. Let $0 \leq \lambda \leq c/K^2$, as in Lemma 4.5. Then

$$\mathbb{P}\{S \geq t\} = \mathbb{P}\{e^{\lambda S} \geq e^{\lambda t}\} \leq e^{-\lambda t} \mathbb{E} \Phi_\lambda(S) \leq \exp(-\lambda t + C\lambda^2 K^2),$$

where the last inequality follows from Lemma 4.5.

The optimal choice of λ subject to the constraint $\lambda \leq c/K^2$ is $\lambda = t/(2CK^2) \wedge c/K^2$. With this choice, we obviously have $\lambda \leq t/(2CK^2)$, which is equivalent to $C\lambda^2 K^2 \leq \lambda t/2$. Thus,

$$\mathbb{P}\{S \geq t\} \leq \exp(-\lambda t + C\lambda^2 K^2) \leq \exp(-\lambda t/2) \leq \exp\left(-\frac{C'}{K^2}(t^2 \wedge t)\right).$$

$\mathbb{P}\{-S \geq t\}$ may be bounded similarly, and the the proof is completed via a union bound. \square

Proof of Proposition 4.3. By Lemma 4.7, for $t \geq 0$

$$\mathbb{P}\{|\|Ax\|_2^2 - 1| \geq t\} \leq 2 \exp\left(-\frac{C}{K^2}(t^2 \wedge t)\right).$$

Now we use the following elementary observation, which holds for all $z, \delta \geq 0$:

$$|z - 1| \geq \delta \quad \text{implies} \quad |z^2 - 1| \geq \delta \vee \delta^2.$$

We obtain

$$\mathbb{P}\{|\|Ax\|_2 - 1| \geq \delta\} \leq \mathbb{P}\{|\|Ax\|_2^2 - 1| \geq \delta \vee \delta^2\} \leq 2 \exp\left(-\frac{C\delta^2}{K^2}\right),$$

since if $t = \delta \vee \delta^2$ then $t^2 \wedge t = \delta^2$. The proof is complete. \square

4.2 Vectors on the sphere, squared process

We now show that $\|Ax\|_2^2 - \|Ay\|_2^2$ has a mixed tail bound for $x, y \in S^{n-1}$. We will then convert that into a subgaussian tail bound for $Z_x - Z_y$ in the next section.

Lemma 4.8 (Squared process). *Let A be a random matrix satisfying the assumptions of Theorem 1.3. Fix any pair of vectors $x, y \in S^{n-1}$. Then the random variable*

$$\Delta := \frac{\|Ax\|_2^2 - \|Ay\|_2^2}{\|x - y\|_2}$$

satisfies

$$\mathbb{P}\{|\Delta| \geq t\} \leq 2 \exp\left(-\frac{C}{K^2}(t^2 \wedge t)\right).$$

Proof. Set

$$u := x + y, \quad v := x - y.$$

Then

$$\|Ax\|_2^2 - \|Ay\|_2^2 = \langle Au, Av \rangle = \sum_i \langle A_i, A_i \rangle u_i v_i + \sum_{i \neq j} \langle A_i, A_j \rangle u_i v_j =: I + II.$$

Observe that $u_i v_i = x_i^2 - y_i^2$ and $\langle A_i, A_i \rangle = \|A_i\|_2^2 = 1$. Thus

$$I = \sum_i x_i^2 - y_i^2 = \|x\|_2^2 - \|y\|_2^2 = 1 - 1 = 0.$$

Now, set $\hat{v} = v/\|v\|_2$ so

$$\Delta = \frac{\|Ax\|_2^2 - \|Ay\|_2^2}{\|v\|_2} = \sum_{i \neq j} \langle A_i, A_j \rangle u_i \hat{v}_i.$$

Using decoupling ([26, Exercise 6.1.4]) as in the proof of Lemma 4.5, we let A' be an independent copy of A and get

$$\mathbb{E} \Phi_\lambda(\Delta) \leq \mathbb{E} \Phi_\lambda[4\langle Au, A'\hat{v} \rangle].$$

As in the proof of Lemma 4.5, taking the conditional expectation with respect to A' gives

$$\mathbb{E} \Phi_\lambda(\Delta) \leq \mathbb{E} \exp(C\lambda^2 K^2 \|Au\|_2^2).$$

Observe that $\|u\|_2 = \|x + y\|_2 \leq \|x\|_2 + \|y\|_2 = 2$ and assume $|\lambda| \leq 1/(CK^2)$. Then $C\lambda^2 K^2 \leq |\lambda|$ and by Lemma 4.5

$$\mathbb{E} \exp(C\lambda^2 K^2 \|Au\|_2^2) \leq \exp(C\lambda^2 K^2).$$

Thus,

$$\mathbb{E} \Phi_\lambda(\Delta) \leq \exp(C\lambda^2 K^2) \quad \text{for } |\lambda| \leq \frac{c}{K^2}.$$

To finish, observe that this is exactly the same bound on the MGF that we had previously for $S = \|Ax\|_2^2 - 1$ (as in Lemma 4.5). Thus, it can be converted into the same tail bound as in Lemma 4.7, but with S replaced by Δ , as desired. \square

4.3 Vectors on the sphere, original process

We now show that the random process $Z_x = |\|Ax\|_2 - 1|$ has subgaussian increments for vectors on the sphere, i.e., we show that

$$\|Z_x - Z_y\|_{\psi_2} \leq CK\|x - y\|_2$$

when $\|x\|_2 = \|y\|_2 = 1$.

This follows from the following lemma, together with the observation that

$$|Z_x - Z_y| = ||\|Ax\|_2 - 1| - |\|Ay\|_2 - 1|| \leq |\|Ax\|_2 - \|Ay\|_2|.$$

Lemma 4.9. Fix any pair of vectors $x, y \in S^{n-1}$. Then

$$||\|Ax\|_2 - \|Ay\|_2||_{\psi_2} \leq CK\|x - y\|_2.$$

Proof. This proof is essentially the same as the proof of [19, Lemma 5.4]. We repeat it here for convenience.

Assume $x \neq y$ (as otherwise the result clearly holds) and set

$$Q := \frac{||\|Ax\|_2 - \|Ay\|_2||}{\|x - y\|_2}.$$

We wish to show that

$$\mathbb{P}\{Q \geq t\} \leq C \exp\left(-C \frac{t^2}{K^2}\right) \quad \text{for } t \geq 0. \quad (4.5)$$

We show this separately for small t and large t . Set $v = x - y$ and $\hat{v} = v/\|v\|_2$.

Case 1: $t \geq 2$.

By the triangle inequality, $|\|Ax\|_2 - \|Ay\|_2| \leq \|A(x - y)\|_2 = \|Av\|_2$. Thus,

$$Q \leq \|A\hat{v}\|_2.$$

Then,

$$\mathbb{P}\{Q \geq t\} \leq \mathbb{P}\{\|A\hat{v}\|_2 \geq t\} = \mathbb{P}\{\|A\hat{v}\|_2 - 1 \geq t - 1\} \leq \mathbb{P}\{\|A\hat{v}\|_2 - 1 \geq t/2\}$$

since $t \geq 2$. Recall that $\|A\hat{v}\|_2 - 1\|_{\psi_2} \lesssim K$ by Proposition 4.3 and so

$$\mathbb{P} \left\{ \|A\hat{v}\|_2 - 1 \geq t/2 \right\} \leq 2 \exp \left(-C \frac{t^2}{K^2} \right)$$

as desired.

Case 2: $t < 2$. Observe that

$$\frac{|\|Ax\|_2 - \|Ay\|_2|}{\|x - y\|_2} = \frac{|\Delta|}{\|Ax\|_2 + \|Ay\|_2} \leq \frac{|\Delta|}{\|Ax\|_2}$$

with Δ defined as in Lemma 4.8.

Define the events $F = \{|\Delta| \geq t/2\}$ and $G = \{\|Ax\|_2 \leq 1/2\}$ so that

$$\left\{ \frac{|\Delta|}{\|Ax\|_2} \geq t \right\} \subset F \cup G.$$

By the union bound

$$\mathbb{P} \left\{ \frac{|\Delta|}{\|Ax\|_2} \geq t \right\} \leq \mathbb{P}(F) + \mathbb{P}(G).$$

By Lemma 4.8, $\mathbb{P}(F) \leq 2 \exp(-Ct^2/K^2)$ and by Proposition 4.3,

$$\mathbb{P}(G) = \mathbb{P} \left\{ \|Ax\|_2 - 1 \leq -\frac{1}{2} \right\} \leq 2 \exp \left(-\frac{C}{K^2} \right) \leq 2 \exp \left(-\frac{Ct^2}{K^2} \right)$$

since $t/2 \leq 1$ (we have absorbed two factors of 4 into the numeric constant C).

Thus,

$$\mathbb{P}(F) + \mathbb{P}(G) \leq 4 \exp \left(-C \frac{t^2}{K^2} \right)$$

giving the desired bound on $\mathbb{P} \{Q \geq t\}$. \square

4.4 Arbitrary vectors, original process

We now show that

$$\|Z_x - Z_y\|_{\psi_2} \lesssim K\|x - y\|_2 \quad (4.6)$$

for any $x, y \in \mathbb{R}^n$, thus completing the proof of Theorem 4.1. Again, we present this for convenience, since the argument is already made in [19] (see also [26, Section 9.1.4]).

Without loss of generality, assume $\|x\|_2 \geq \|y\|_2$. Assume $y \neq 0$ as otherwise Equation (4.6) is already shown by Proposition 4.3. By a simple rescaling argument, we may assume $\|y\|_2 = 1$. Let $\hat{x} = x/\|x\|$, i.e., the projection onto the sphere (or the convex hull of the sphere). Since projection onto a convex set is contractive, and by definition of projection, we have

$$\|\hat{x} - y\|_2 \leq \|x - y\|_2 \quad \text{and} \quad \|x - \hat{x}\|_2 \leq \|x - y\|_2. \quad (4.7)$$

By the triangle inequality, $\|Z_x - Z_y\|_{\psi_2} \leq \|Z_x - Z_{\hat{x}}\|_{\psi_2} + \|Z_{\hat{x}} - Z_y\|_{\psi_2}$. We now use the results from the previous sections to show that each term is bounded by $CK\|x - y\|_2$, thus completing the proof. Indeed, by linearity, $Z_x = \|x\|_2 \cdot Z_{\hat{x}}$. Further, $\|x\|_2 - 1 = \|x - \hat{x}\|_2$ and so

$$\|Z_x - Z_{\hat{x}}\|_{\psi_2} = \|x - \hat{x}\|_2 \cdot \|\|A\hat{x}\|_2 - 1\|_{\psi_2} \leq CK\|x - \hat{x}\|_2 \leq CK\|x - y\|_2.$$

The first inequality follows by Proposition 4.3 and the second by Equation (4.7).

Next, by Lemma 4.9 followed by Equation (4.7),

$$\|Z_{\hat{x}} - Z_y\|_{\psi_2} \leq CK\|\hat{x} - y\|_2 \leq CK\|x - y\|_2.$$

The proof is finished.

5 Proofs of the material in Sections 2, 3

We begin with a lemma controlling the subgaussian norm of the columns of “approximately sparse” and “exactly sparse” matrices, as referenced in Sections 2, 3.

Lemma 5.1. *Let $s \leq m$ be a positive integer and consider random vectors $X, Y \in \mathbb{R}^m$. The vector X is drawn uniformly from the set of vectors in \mathbb{R}^m with that have all 0, 1, −1 entries and exactly s non-zero entries. The entries of Y are independent and take the value 1 or −1 with probability $\frac{s}{2m}$ each, and 0 with probability $1 - \frac{s}{m}$. Then,*

$$\|X\|_{\psi_2} \leq \frac{C}{\sqrt{\log(1 + m/s)}} \quad \text{and} \quad \|Y\|_{\psi_2} \leq \frac{C}{\sqrt{\log(1 + m/s)}}. \quad (5.1)$$

Proof. It is not hard to show that by definition, $\|Y_i\|_{\psi_2} = 1/\sqrt{\log(1 + m/s)}$. The bound on $\|Y\|_{\psi_2}$ then follows from [26, 2.6.1] and Definition (1.1).

We now bound the moments of the marginals of X by the moments of the marginals of Y , thus controlling $\|X\|_{\psi_2}$ by $\|Y\|_{\psi_2}$. Below, given an event F and a random variable Z , we use the notation

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

i.e., it is the subgaussian norm of Z conditioned on F .

Let $u \in S^{n-1}$ and let G be the event $G := \{\|Y\|_0 \geq s\}$ where we use $\|\cdot\|_0$ for the number of nonzero entries of a vector. Since s is the median of $\|Y\|_0$, we have $\mathbb{P}\{G\} \geq 1/2$ (see [15]). Let $p \geq 1$ and observe that

$$\mathbb{E}|\langle Y, u \rangle|^p \geq \mathbb{E}\left[|\langle Y, u \rangle|^p | G\right] \cdot \mathbb{P}\{G\} \geq \mathbb{E}\left[|\langle Y, u \rangle|^p | \|Y\|_0 = s\right] \cdot \frac{1}{2} = \mathbb{E}|\langle X, u \rangle|^p \cdot \frac{1}{2}.$$

The last equality follows since conditioned on $\|Y\|_0 = s$, Y has the same distribution as X . We have shown that

$$\mathbb{E}|\langle X, u \rangle|^p \leq 2\mathbb{E}|\langle Y, u \rangle|^p.$$

It follows from the moment-based characterization of the subgaussian norm [26, Equation 2.15] that $\|X\|_{\psi_2} \leq C\|Y\|_{\psi_2}$ as desired. \square

5.1 Proof of Corollary 3.1

Our strategy is to condition on the event F that the column norms of A are lower bounded, show that conditionally on this event \tilde{A} has independent, mean-zero, subgaussian columns, and then apply Corollary 1.4.

The norm of a vector with independent subgaussian entries is well concentrated. Indeed, by [13, Theorem 4.1] the columns of A satisfy,

$$\|\|A_i\|_2 - \sqrt{m}\|_{\psi_2} \lesssim K\sqrt{\log K} \quad i = 1, 2, \dots, n$$

and thus,

$$\mathbb{P}\left\{\|A_i\|_2 \leq \frac{\sqrt{m}}{2}\right\} = \mathbb{P}\left\{\|A_i\|_2 - \sqrt{m} \leq -\frac{\sqrt{m}}{2}\right\} \leq 2\exp\left(-C\frac{m}{K^2 \log K}\right).$$

Then, by the union bound, the event

$$F := \left\{\min_{i=1,2,\dots,n} \|A_i\|_2 \geq \frac{\sqrt{m}}{2}\right\}$$

satisfies

$$\mathbb{P}\{F\} \geq 1 - 2n\exp\left(-C\frac{m}{K^2 \log K}\right) = 1 - 2\exp\left(\log(n) - C\frac{m}{K^2 \log K}\right).$$

By assumption, $m \geq CK^2 \log(K) \cdot \log(n)$, and so the $\log(n)$ term is insignificant, only changing the constant:

$$\mathbb{P}\{F\} \geq 1 - 2 \exp\left(-C \frac{m}{K^2 \log K}\right) \geq \frac{1}{2}.$$

Now condition on the event F . Clearly, the columns of A are (conditionally) non-zero. Further,

$$\|\tilde{A}_i|F\|_{\psi_2} = \left\| \frac{\sqrt{m}}{\|A_i\|_2} A_i \Big| F \right\|_{\psi_2} \leq 2\|A_i|F\|_{\psi_2} \leq C\|A_i\|_{\psi_2}. \quad (5.2)$$

The first inequality follows because $\sqrt{m}/\|A_i\|_2 \leq 2$ (conditionally). The second inequality follows because conditioning on an event with probability at least 0.5 cannot increase the subgaussian norm of A_i by more than a constant factor. For instance, this follows by checking the moment condition in [26, Proposition 2.5.2].

Thus, by [26, Proposition 2.6.1],

$$\|\tilde{A}_i|F\|_{\psi_2} \leq C\|A_i\|_{\psi_2} \leq CK.$$

Further, since the entries of A_i are symmetric, and thus the signs of those entries are independent from their magnitudes, the conditional entries of \tilde{A}_i are still symmetric. It follows that $\mathbb{E}[\tilde{A}_i|F] = 0$. Summarizing, conditional on F , \tilde{A} has the following properties

$$\|\tilde{A}_i\|_2 = \sqrt{m}, \quad \|\tilde{A}_i|F\|_{\psi_2} \leq CK, \quad \mathbb{E}[\tilde{A}_i|F] = 0.$$

Clearly, the columns of \tilde{A} remain independent conditional on F and so we are in the setting of Corollary 1.4 with $\lambda = \sqrt{m}$. The proposition follows.

Remark 5.2 (Replacing $m \geq CK^2 \log(K) \cdot \log(n)$ with $m \geq CK^2 \log(K)$). In Equation (5.2), conditioning on F is equivalent to conditioning on the event that $\|A_i\|_2 \geq m/\sqrt{2}$ (by independence). Thus, we only need $\mathbb{P}\{\|A_i\|_2 \geq m/\sqrt{2}\} \geq 1/2$ for Equation (5.2) to hold. This only requires $m \geq CK^2 \log(K)$, rather than $m \geq CK^2 \log(K) \cdot \log(n)$ (since we may remove the union bound involved in controlling $\mathbb{P}(F)$). Thus Corollary 3.1 holds with the milder assumption on m , but without a lower bound on $\mathbb{P}(F)$. Similar arguments allow us to take the mild assumption $s \geq 1$ in Corollary 3.3 at the expense of losing the lower bound on $\mathbb{P}(F)$.

6 Conclusion

We have shown that random matrices with normalized independent columns have similar behaviour that of random matrices with independent rows, except with better dependence on the subgaussian norm. We conclude with an open question: Can the assumption of normalized columns be removed? This question is made concrete with the following conjecture.

Conjecture 6.1. *Let A be an $m \times n$ matrix whose columns A_i are independent, mean zero, subgaussian random vectors in \mathbb{R}^m with $\|A_i\|_{\psi_2} \leq K$. Then, for any subset $T \subset \mathbb{R}^n$, we have*

$$\mathbb{E} \sup_{x \in T} \left| \|Ax\|_2 - \|Dx\|_2 \right| \leq C(K) \gamma(T)$$

where D is the diagonal matrix with $D_{i,i} = \|A_i\|_2$ and $C(K)$ depends only on K .

References

- [1] D. Achlioptas. Database-friendly random projections: Johnson-lindenstrauss with binary coins. *Journal of computer and System Sciences*, 66(4):671–687, 2003.
- [2] R. Adamczak, A. E. Litvak, A. Pajor, and N. Tomczak-Jaegermann. Restricted isometry property of matrices with independent columns and neighborly polytopes by random sampling. *Constructive Approximation*, 34:61–88, 2011.
- [3] N. Ailon and B. Chazelle. The fast johnson-lindenstrauss transform and approximate nearest neighbors. *SIAM Journal on computing*, 39(1):302–322, 2009.
- [4] D. Bartl and S. Mendelson. Random embeddings with an almost gaussian distortion. *Advances in Mathematics*, 400:108261, 2022.
- [5] J. Bourgain, S. Dirksen, and J. Nelson. Toward a unified theory of sparse dimensionality reduction in euclidean space. *Geometric and Functional Analysis*, 25:1009–1088, 2015.
- [6] S. Chenakkod, M. Dereziński, X. Dong, and M. Rudelson. Optimal embedding dimension for sparse subspace embeddings. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1106–1117, 2024.
- [7] K. L. Clarkson and D. P. Woodruff. Low rank approximation and regression in input sparsity time. In *Proceedings of the forty-fifth annual ACM symposium on Theory of Computing*, pages 81–90, 2013.
- [8] A. Cohen, N. Shlezinger, A. Solomon, Y. C. Eldar, and M. Médard. Multi-level group testing with application to one-shot pooled covid-19 tests. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1030–1034. IEEE, 2021.
- [9] M. B. Cohen, T. Jayram, and J. Nelson. Simple analyses of the sparse johnson-lindenstrauss transform. In *1st Symposium on Simplicity in Algorithms (SOSA 2018)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2018.
- [10] S. Dirksen. Tail bounds via generic chaining. *Electronic Journal of Probability*, 20, 2015.
- [11] Y. C. Eldar and G. Kutyniok. *Compressed sensing: theory and applications*. Cambridge university press, 2012.
- [12] S. Foucart and H. Rauhut. *A Mathematical Introduction to Compressive Sensing*. Birkhäuser Basel, 2013.
- [13] H. Jeong, X. Li, Y. Plan, and O. Yilmaz. Sub-gaussian matrices on sets: Optimal tail dependence and applications. *Communications on Pure and Applied Mathematics*, 75(8):1713–1754, 2022.
- [14] W. B. Johnson and J. Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. *Contemporary mathematics*, 26(189-206):1, 1984.
- [15] R. Kaas and J. M. Buhrman. Mean, median and mode in binomial distributions. *Statistica Neerlandica*, 34(1):13–18, 1980.
- [16] D. M. Kane and J. Nelson. Sparser johnson-lindenstrauss transforms. *Journal of the ACM (JACM)*, 61(1):1–23, 2014.
- [17] B. Klartag and S. Mendelson. Empirical processes and random projections. *Journal of Functional Analysis*, 225(1):229, 2005.
- [18] R. Kueng and P. Jung. Robust nonnegative sparse recovery and the nullspace property of 0/1 measurements. *IEEE Transactions on Information Theory*, 64(2):689–703, 2017.
- [19] 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*, pages 277–299. Springer, 2017.
- [20] J. Matoušek. On variants of the johnson-lindenstrauss lemma. *Random Structures & Algorithms*, 33(2):142–156, 2008.
- [21] S. Mendelson, A. Pajor, and N. Tomczak-Jaegermann. Reconstruction and subgaussian operators in asymptotic geometric analysis. *Geometric and Functional Analysis*, 17(4):1248–1282, 2007.
- [22] M. Pilanci and M. J. Wainwright. Randomized sketches of convex programs with sharp guarantees. *IEEE Transactions on Information Theory*, 61(9):5096–5115, 2015.

- [23] G. Schechtman. Two observations regarding embedding subsets of euclidean spaces in normed spaces. *Advances in Mathematics*, 200(1):125–135, 2006.
- [24] M. Talagrand. *Upper and lower bounds for stochastic processes: modern methods and classical problems*, volume 60. Springer Science & Business Media, 2014.
- [25] R. Vershynin. *Introduction to the non-asymptotic analysis of random matrices*. Cambridge Univ. Press, Cambridge, 2012.
- [26] R. Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.
- [27] Y. Yang, M. Pilanci, M. J. Wainwright, et al. Randomized sketches for kernels: Fast and optimal nonparametric regression. *The Annals of Statistics*, 45(3):991–1023, 2017.