THE SMALLEST SINGULAR VALUE OF A RANDOM RECTANGULAR MATRIX

MARK RUDELSON AND ROMAN VERSHYNIN

Abstract. We prove an optimal estimate of the smallest singular value of a random subgaussian matrix, valid for all dimensions. For an $N \times n$ matrix A with independent and identically distributed subgaussian entries, the smallest singular value of A is at least of the order $\sqrt{N} - \sqrt{n-1}$ with high probability. A sharp estimate on the probability is also obtained.

1. Introduction

1.1. **Singular values of subgaussian matrices.** Extreme singular values of random matrices has been of considerable interest in mathematical physics, geometric functional analysis, numerical analysis and other fields. Consider an $N \times n$ real matrix A with $N \geq n$. The *singular values* $s_k(A)$ of A are the eigenvalues of $|A| = \sqrt{A^t A}$ arranged in nonincreasing order. Of particular significance are the largest and the smallest singular values

$$s_1(A) = \sup_{x:\ \|x\|_2=1} \|Ax\|_2, \qquad s_n(A) = \inf_{x:\ \|x\|_2=1} \|Ax\|_2. \tag{1.1}$$

A natural matrix model is given by matrices whose entries are independent real random variables with certain moment assumptions. In this paper, we shall consider *subgaussian random variables* ξ – those whose tails are dominated by that of the standard normal random variable. Namely, a random variable ξ is called subgaussian if there exists $B > 0$ such that

$$\mathbb{P}(|\xi| > t) \leq 2\exp(-t^2/B^2) \qquad \text{for all } t > 0. \tag{1.2}$$

The minimal B in this inequality is called the *subgaussian moment* of ξ. Inequality (1.2) is often equivalently formulated as the moment condition

$$(\mathbb{E}|\xi|^p)^{1/p} \leq CB\sqrt{p} \qquad \text{for all } p \geq 1, \tag{1.3}$$

where C is an absolute constant. The class of subgaussian random variables includes many random variables that arise naturally in applications, such as normal, symmetric ± 1 and general bounded random variables.

Date: August 3, 2009.

M.R. was supported by NSF DMS grants 0556151 and 0652684. R.V. was supported by the Alfred P. Sloan Foundation and by NSF DMS grants 0401032 and 0652617.

In this paper, we study $N \times n$ real random matrices A whose entries are independent and identically distributed mean zero subgaussian random variables. The asymptotic behavior of the extreme singular values of A is well understood. If the entries have unit variance and the dimension n grows to infinity while the aspect ratio n/N converges to a constant $\lambda \in (0,1)$, then

$$\frac{s_1(A)}{\sqrt{N}} \to 1 + \sqrt{\lambda}, \qquad \frac{s_n(A)}{\sqrt{N}} \to 1 - \sqrt{\lambda}$$

almost surely. This result was proved in [21] for Gaussian matrices, and in [2] for matrices with independent and identically distributed entries with finite fourth moment. In other words, we have asymptotically

$$s_1(A) \sim \sqrt{N} + \sqrt{n}, \qquad s_n(A) \sim \sqrt{N} - \sqrt{n}. \tag{1.4}$$

Considerable efforts were made recently to establish non-asymptotic estimates similar to (1.4), which would hold for arbitrary fixed dimensions N and n; see the survey [13] on the largest singular value, and the discussion below on the smallest singular value.

Estimates in fixed dimensions are essential for many problems of geometric functional analysis and computer science. Most often needed are upper bounds on the largest singular value and lower bounds on the smallest singular value, which together yield that A acts as a nice isomorphic embedding of $\mathbb{R}^n$ into $\mathbb{R}^N$. Such bounds are often satisfactory even if they are known to hold up to a constant factor independent of the dimension.

The largest singular value is relatively easy to bound above, up to a constant factor. Indeed, a standard covering argument shows that $s_1(A)$ is at most of the optimal order $\sqrt{N}$ for all fixed dimensions, see Proposition 2.3 below. The smallest singular value is significantly harder to control. The efforts to prove optimal bounds on $s_n(A)$ have a long history, which we shall now outline.

1.2. **Tall matrices.** A result of [3] provides an optimal bound for tall matrices, those with aspect ratio $\lambda = n/N$ satisfies $\lambda < \lambda_0$ for some sufficiently small constant $\lambda_0 > 0$. Recalling (1.4), one should expect that tall matrices satisfy

$$s_n(A) \ge c\sqrt{N} \qquad \text{with high probability.} \tag{1.5}$$

It was indeed proved in [3] that for tall ± 1 matrices one has

$$\mathbb{P}\big(s_n(A) \le c\sqrt{N}\big) \le e^{-cN} \tag{1.6}$$

where $\lambda_0 > 0$ and $c > 0$ are absolute constants.

1.3. **Almost square matrices.** As we move toward square matrices, thus making the aspect ratio $\lambda = n/N$ arbitrarily close to 1, the problem of estimating the smallest singular value becomes harder. One still expects (1.5) to be true as long as $\lambda < 1$ is any constant. Indeed, this was proved in [16] for

arbitrary aspect ratios $\lambda < 1 - c/\log n$ and for general random matrices with independent subgaussian entries. One has

$$\mathbb{P}\big(s_n(A) \le c_\lambda \sqrt{N}\big) \le e^{-cN}, \tag{1.7}$$

where $c_\lambda > 0$ depends only on λ and the maximal subgaussian moment of the entries.

In subsequent work [1], the dependence of c_λ on the aspect ratio in (1.7) was improved for random ± 1 matrices; however the probability estimate there was weaker than in (1.7). An estimate for subgaussian random matrices of all dimensions was obtained in [19]. For any $\varepsilon \ge CN^{-1/2}$, it was shown that

$$\mathbb{P}\big(s_n(A) \le \varepsilon(1-\lambda)(\sqrt{N} - \sqrt{n})\big) \le (C\varepsilon)^{N-n} + e^{-cN}.$$

However, because of the factor $(1-\lambda)$, this estimate is suboptimal and does not correspond to the expected asymptotic behavior (1.4).

1.4. **Square matrices.** The extreme case for the problem of estimating the singular value is for the square matrices, where $N = n$. Asymptotic (1.4) is useless for square matrices. However, for "almost" square matrices, those with constant defect $N - n = O(1)$, the quantity $\sqrt{N} - \sqrt{n}$ is of order $1/\sqrt{N}$, so asymptotics (1.4) heuristically suggests that these matrices should satisfy

$$s_n(A) \ge \frac{c}{\sqrt{N}} \qquad \text{with high probability.} \tag{1.8}$$

This conjecture was proved recently in [20] for all square subgaussian matrices:

$$\mathbb{P}\Big(s_n(A) \le \frac{\varepsilon}{\sqrt{N}}\Big) \le C\varepsilon + e^{-cN}. \tag{1.9}$$

1.5. **New result: bridging all classes of matrices.** In this paper, we prove the conjectural bound for $s_n(A)$ valid for all subgaussian matrices in all fixed dimensions N, n. The bound is optimal for matrices with all aspect ratios we encountered above.

Theorem 1.1. *Let A be an $N \times n$ random matrix, $N \ge n$, whose elements are independent copies of a mean zero subgaussian random variable with unit variance. Then, for every $\varepsilon > 0$, we have*

$$\mathbb{P}\Big(s_n(A) \le \varepsilon\big(\sqrt{N} - \sqrt{n-1}\big)\Big) \le (C\varepsilon)^{N-n+1} + e^{-cN} \tag{1.10}$$

where $C, c > 0$ depend (polynomially) only on the subgaussian moment B.

For *tall matrices*, Theorem 1.1 clearly amounts to the known estimates (1.5), (1.6). For *square matrices* ($N = n$), the quantity $\sqrt{N} - \sqrt{N-1}$ is of order $1/\sqrt{N}$, so Theorem 1.1 amounts to the known estimates (1.8), (1.9). Finally,

for matrices that are arbitrarily *close to square*, Theorem 1.1 yields the new optimal estimate

$$s_n(A) \geq c(\sqrt{N} - \sqrt{n}) \qquad \text{with high probability.} \tag{1.11}$$

This is a version of the asymptotics (1.4), now valid for all fixed dimensions. This bound was explicitly conjectured e.g. in [24].

Theorem 1.1 seems to be new even for Gaussian matrices. Some early progress was made by Edelman [5] and Szarek [22] who in particular proved (1.9) for Gaussian matrices, see also the subsequent work by Edelman and Sutton [6]. Gordon's inequality [10] can be used to prove that, for Gaussian matrices, $\mathbb{E}s_n(A) \geq \sqrt{N} - \sqrt{n}$, see Theorem II.13 in [4]. One can further use the concentration of measure inequality on the Euclidean sphere to estimate the probability as

$$\mathbb{P}\big(s_n(A) \leq \sqrt{N} - \sqrt{n} - t\big) \leq e^{-t^2/2}, \qquad t > 0.$$

However, this bound is not optimal, and it becomes useless for matrices that are close to square, when $N - n = o(\sqrt{n})$.

The form of estimate (1.10) may be expected if one recalls the classical ε-net argument, which underlies many proofs in geometric functional analysis. By (1.1), we are looking for a lower bound on $\|Ax\|$ that would hold uniformly for all vectors x on the unit Euclidean sphere S^{n-1}. For every fixed $x \in S^{n-1}$, the quantity $\|Ax\|_2^2$ is the sum of N independent random variables (the squares of the coordinates of Ax). Therefore, the deviation inequalities make us to expect that $\|Ax\|_2$ is of the order $\sqrt{N}$ with probability exponential in N, i.e. $1-e^{-cN}$. We can run this argument separately for each vector x in a small net $\mathcal{N}$ of the sphere S^{n-1}, and then take the union bound to make the estimate uniform over $x \in \mathcal{N}$. It is known how to choose a net $\mathcal{N}$ of cardinality exponential in the dimension $n-1$ of the sphere, i.e. $|\mathcal{N}| \leq e^{C(n-1)}$. Therefore, with probability $1 - e^{C(n-1)}e^{-cN}$, we have a good lower bound on $\|Ax\|_2 \sim \sqrt{N}$ for all vectors x in the net $\mathcal{N}$. Finally, one transfers this estimate from the net to the whole sphere S^{n-1} by approximation.

The problem with this argument is that the constants C and c are not the same. Therefore, our estimate on the probability $1 - e^{C(n-1)}e^{-cN}$ is positive only for tall matrices, when $N \geq (C/c)n$. To reach out to matrices of arbitrary dimensions, one needs to develop much more sensitive versions of the ε-net arguments. Nevertheless, the end result stated in Theorem 1.1 exhibits the same two forces played against one another – the probability quantified by the dimension N and the complexity of the sphere S^{n-1} quantified by its dimension $n-1$.

1.6. **Small ball probabilities, distance problems, and additive structure.** Our proof of Theorem 1.1 is a development of our method in [20] for square matrices. Dealing with rectangular matrices is in several ways considerably harder. Several new tools are developed in this paper, which may be of independent interest.

One new key ingredient is a *small ball probability* bound for sums of independent random vectors in $\mathbb{R}^d$. We consider the sum $S = \sum_k a_k X_k$ where X_k are i.i.d. random variables and a_k are real coefficients. We then estimate the probability that such sum falls into a given small Euclidean ball in $\mathbb{R}^d$. Useful upper bounds on the small ball probability must depend on the additive structure of the coefficients a_k. The less structure the coefficients carry, the more spread the distribution of S is, so the smaller is the small ball probability. Our treatment of small ball probabilities is a development of the Littlewood-Offord theory from [20], which is now done in arbitrary dimension d as opposed in $d = 1$ in [20]. While this paper was being written, Friedland and Sodin [8] proposed two different ways to simplify and improve our argument in [20]. With their kind permission, we include in Section 3 a multi-dimensional version of an unpublished argument of Friedland and Sodin [9], which is considerably simpler than our original proof.

We use small the ball probability estimates to prove an optimal bound for the distance problem: *how close is a random vector from an independent random subspace?* Consider a vector X in $\mathbb{R}^N$ with independent identically distributed coordinates and a subspace H spanned by $N - m$ independent copies of X. In Section 4, we show that the distance is at least of order $\sqrt{m}$ with high probability, and we obtain the sharp estimate on this probability:

$$\mathbb{P}\big(\mathrm{dist}(X, H) < \varepsilon\sqrt{m}\big) \le (C\varepsilon)^m + e^{-cN}. \tag{1.12}$$

This bound is easy for a standard normal vector X in $\mathbb{R}^N$, since $\mathrm{dist}(X, H)$ is in this case the Euclidean norm of the standard normal vector in $\mathbb{R}^m$. However, for discrete distributions, such as for X with ± 1 random coordinates, estimate (1.12) is non-trivial. In [20], it was proved for $m = 1$; in this paper we extend the distance bound to all dimensions.

To prove (1.12), we first use the small ball probability inequalities to compute the *distance to an arbitrary subspace* H. This estimate necessarily depends on the additive structure of the subspace H; the less structure, the better is our estimate, see Theorem 4.2. We then prove the intuitively plausible fact that *random subspaces have no arithmetic structure*, see Theorem 4.3. This together leads to the desired distance estimate (1.12).

The distance bound is then used to prove our main result, Theorem 1.1. Let X be some column of the random matrix A and H be the span of the other columns. The simple rank argument shows that the smallest singular

value $s_n(A) = 0$ if and only if $X \in H$ for some column. A simple quantitative version of this argument is that a lower estimate on $s_n(A)$ yields a lower bound on $\mathrm{dist}(X, H)$.

In Section 6, we show how to reverse this argument for random matrices – *deduce a lower bound on the smallest singular value* $s_n(A)$ *from lower bound* (1.12) *on the distance* $\mathrm{dist}(X, H)$. Our reverse argument is harder than its version for square matrices from [20], where we had $m = 1$. First, instead of one column X we now have to consider all linear combinations of $d \sim m/2$ columns; see Lemma 6.2. To obtain a distance bound that would be uniformly good for all such linear combinations, one would normally use an ε-net argument. However, the distance to the $(N - m)$-dimensional subspace H is not sufficiently stable for this argument to be useful for small m (for matrices close to square). We therefore develop a decoupling argument in Section 7 to bypass this difficulty.

Once this is done, the proof is quickly completed in Section 8.

Acknowledgement. We are grateful to Shuheng Zhou, Nicole Tomczak-Jaegermann, Radoslaw Adamczak, and the anonymous referee for pointing out several inaccuracies in our argument. The second named author is grateful for his wife Lilia for her love and patience during the years this paper was being written.

2. Notation and preliminaries

Throughout the paper, positive constants are denoted $C, C_1, C_2, c, c_1, c_2, \ldots$ Unless otherwise stated, these are absolute constants. In some of our arguments they may depend (polynomially) on specified parameters, such as the subgaussian moment B.

The canonical inner product on $\mathbb{R}^n$ is denoted $\langle \cdot, \cdot \rangle$, and the Euclidean norm on $\mathbb{R}^n$ is denoted $\| \cdot \|_2$. The Euclidean distance from a point a to a subset D in $\mathbb{R}^n$ is denoted $\mathrm{dist}(a, D)$. The Euclidean ball of radius R centered at a point a is denoted $B(a, R)$. The unit Euclidean sphere centered at the origin is denoted S^{n-1}. If E is a subspace of $\mathbb{R}^n$, its unit Euclidean sphere is denoted $S(E) := S^{n-1} \cap E$.

The orthogonal projection in $\mathbb{R}^n$ onto a subspace E is denoted P_E. For a subset of coordinates $J \subseteq \{1, \ldots, n\}$, we sometimes write P_J for $P_{\mathbb{R}^J}$ where it causes no confusion.

2.1. **Nets.** Consider a subset D of $\mathbb{R}^n$, and let $\varepsilon > 0$. Recall that an ε-net of D is a subset $\mathcal{N} \subseteq D$ such that for every $x \in D$ one has $\mathrm{dist}(x, \mathcal{N}) \le \varepsilon$.

The following Lemma is a variant of the well known volumetric estimate.

Proposition 2.1 (Nets)**.** *Let S be a subset of S^{n-1}, and let $\varepsilon > 0$. Then there exists an ε-net of S of cardinality at most*

$$2n\Big(1+\frac{2}{\varepsilon}\Big)^{n-1}.$$

The published variants of his lemma (e.g. [17], Lemma 2.6) have exponent n rather than $n-1$. Since the latter exponent will be crucial for our purposes, we include the proof of this lemma for the reader's convenience.

Proof. Without loss of generality we can assume that $\varepsilon < 2$, otherwise any single point forms a desired net. Let $\mathcal{N}$ be an ε-separated subset of S of maximal cardinality. By maximality, $\mathcal{N}$ is an ε-net of S. Since $\mathcal{N}$ is ε-separated, the balls $B(x, \varepsilon/2)$ with centers $x \in \mathcal{N}$ are disjoint. All these balls have the same volume, and they are contained in the spherical shell $B(0, 1+\varepsilon/2)\setminus B(0, 1-\varepsilon/2)$. Therefore, comparing the volumes, we have

$$|\mathcal{N}| \cdot \mathrm{vol}(B(0,\varepsilon/2)) \le \mathrm{vol}\big(B(0,1+\varepsilon/2) \setminus B(0,1-\varepsilon/2)\big).$$

Dividing both sides of this inequality by $\mathrm{vol}(B(0,1))$, we obtain

$$|\mathcal{N}| \cdot (\varepsilon/2)^n \le (1+\varepsilon/2)^n - (1-\varepsilon/2)^n.$$

Using the inequality $(1+x)^n - (1-x)^n \le 2nx(1+x)^{n-1}$ valid for $x \in (0,1)$, we conclude that $|\mathcal{N}|$ is bounded as desired. This completes the proof. □

The following well known argument allows one to compute the norm of a linear operator using nets. We have not found a published reference to this argument, so we include it for the reader's convenience.

Proposition 2.2 (Computing norm on nets)**.** *Let $\mathcal{N}$ be a ε-net of S^{n-1} and $\mathcal{M}$ be a δ-net of S^{m-1}. Then for any linear operator $A : \mathbb{R}^n \to \mathbb{R}^m$*

$$\|A\| \le \frac{1}{(1-\varepsilon)(1-\delta)} \sup_{x\in\mathcal{N},\, y\in\mathcal{M}} |\langle Ax, y\rangle|.$$

Proof. Every $z \in S^{n-1}$ has the form $z = x + h$, where $x \in \mathcal{N}$ and $\|h\|_2 \le \varepsilon$. Since $\|A\| = \sup_{z\in S^{n-1}} \|Az\|_2$, the triangle inequality yields

$$\|A\| \le \sup_{x\in\mathcal{N}} \|Ax\|_2 + \max_{\|h\|_2\le\varepsilon} \|Ah\|_2.$$

The last term in the right hand side is bounded by $\varepsilon\|A\|$. Therefore we have shown that

$$(1-\varepsilon)\|A\| \le \sup_{x\in\mathcal{N}} \|Ax\|_2.$$

Fix $x \in \mathcal{N}$. Repeating the above argument for $\|Ax\|_2 = \sup_{y\in S^{m-1}} |\langle Ax, y\rangle|$ yields the bound

$$(1-\delta)\|Ax\|_2 \le \sup_{y\in\mathcal{M}} |\langle Ax, y\rangle|.$$

The two previous estimates complete the proof. □

Using nets, one easily proves the well known basic bound $O(\sqrt{N})$ on the norm of a random subgaussian matrix:

Proposition 2.3 (Norm)**.** *Let A be an $N \times n$ random matrix, $N \geq n$, whose elements are independent copies of a subgaussian random variable. Then*

$$\mathbb{P}\big(\|A\| > t\sqrt{N}\big) \leq e^{-c_0 t^2 N} \qquad \text{for } t \geq C_0,$$

where $C_0, c_0 > 0$ depend only on the subgaussian moment B.

Proof. Let $\mathcal{N}$ be a (1/2)-net of S^{N-1} and M be a (1/2)-net of S^{n-1}. By Proposition 2.1, we can choose these nets such that

$$|\mathcal{N}| \leq 2N \cdot 5^{N-1} \leq 6^N, \quad |\mathcal{M}| \leq 2n \cdot 5^{n-1} \leq 6^n.$$

For every $x \in \mathcal{N}$ and $y \in \mathcal{M}$, the random variable $\langle Ax, y\rangle$ is subgaussian (see Fact 2.1 in [16]), thus

$$\mathbb{P}\big(|\langle Ax, y\rangle| > t\sqrt{N}\big) \leq C_1 e^{-c_1 t^2 N} \qquad \text{for } t > 0,$$

where $C_1, c_1 > 0$ depend only on the subgaussian moment B. Using Lemma 2.2 and taking the union bound, we obtain

$$\mathbb{P}\big(\|A\| > t\sqrt{N}\big) \leq 4|\mathcal{N}||\mathcal{M}| \max_{x\in N,\, y\in M} \mathbb{P}\big(|\langle Ax, y\rangle| > t\sqrt{d}\big) \leq 4\cdot 6^N \cdot 6^N \cdot C_1 e^{-c_1 t^2 N}.$$

This completes the proof. □

2.2. **Compressible and incompressible vectors.** In our proof of Theorem 1.1, we will make use of a partition of the unit sphere S^{n-1} into two sets of compressible and incompressible vectors. These sets were first defined in [20] as follows.

Definition 2.4 (Compressible and incompressible vectors)**.** Let $\delta, \rho \in (0, 1)$. A vector $x \in \mathbb{R}^n$ is called *sparse* if $|\mathrm{supp}(x)| \leq \delta n$. A vector $x \in S^{n-1}$ is called *compressible* if x is within Euclidean distance ρ from the set of all sparse vectors. A vector $x \in S^{n-1}$ is called *incompressible* if it is not compressible. The sets of compressible and incompressible vectors will be denoted by $Comp = Comp(\delta, \rho)$ and $Incomp = Incomp(\delta, \rho)$ respectively.

We now recall without proof two simple results. The first is Lemma 3.4 from [20]:

Lemma 2.5 (Incompressible vectors are spread)**.** *Let $x \in Incomp(\delta, \rho)$. Then there exists a set $\sigma = \sigma(x) \subseteq \{1, \dots, n\}$ of cardinality $|\sigma| \geq \frac{1}{2}\rho^2\delta n$ and such that*

$$\frac{\rho}{\sqrt{2n}} \leq |x_k| \leq \frac{1}{\sqrt{\delta n}} \qquad \text{for all } k \in \sigma. \tag{2.1}$$

The other result is a variant of Lemma 3.3 from [20], which establishes the invertibility on compressible vectors, and allows us to focus on incompressible vectors in our proof of Theorem 1.1. While Lemma 3.3 was formulated in [20] for a square matrix, the same proof applies to $N \times n$ matrices, provided that $N \geq n/2$.

Lemma 2.6 (Invertibility for compressible vectors)**.** *Let A be an $N \times n$ random matrix, $N \geq n/2$, whose elements are independent copies of a subgaussian random variable. There exist $\delta, \rho, c_3 > 0$ depending only on the subgaussian moment B such that*

$$\mathbb{P}\big(\inf_{x \in Comp(\delta,\rho)} \|Ax\|_2 \leq c_3\sqrt{N}\big) \leq e^{-c_3 N}.$$

□

3. Small ball probability and the arithmetic structure

Starting from the works of Lévy [14], Kolmogorov [12] and Esséen [7], a number of results in probability theory was concerned with the question how spread the sums of independent random variables are. It is convenient to quantify the spread of a random variable in the following way.

Definition 3.1. The *Lévy concentration function* of a random vector S in $\mathbb{R}^m$ is defined for $\varepsilon > 0$ as

$$\mathcal{L}(S, \varepsilon) = \sup_{v \in \mathbb{R}^m} \mathbb{P}(\|S - v\|_2 \leq \varepsilon).$$

An equivalent way of looking at the Lévy concentration function is that it measures the *small ball probabilities* – the likelihood that the random vector S enters a small ball in the space. An exposition of the theory of small ball probabilities can be found in [15].

One can derive a simple but rather weak bound on Lévy concentration function from Paley-Zygmund inequality.

Lemma 3.2. *Let ξ be a random variable with mean zero, unit variance, and finite fourth moment. Then for every $\varepsilon \in (0,1)$ there exists $p \in (0,1)$ which depends only on ε and on the fourth moment, and such that*

$$\mathcal{L}(\xi, \varepsilon) \leq p.$$

Remark. In particular, this bound holds for subgaussian random variables, and with p that depends only on ε and the subgaussian moment.

Proof. We use Paley-Zygmund inequality, which states for a random variable Z that

$$\mathbb{P}(|Z| > \varepsilon) \geq \frac{(\mathbb{E}Z^2 - \varepsilon^2)^2}{\mathbb{E}Z^4}, \qquad \varepsilon > 0, \tag{3.1}$$

see e.g. [16], Lemma 3.5.

Let $v \in \mathbb{R}$ and consider the random variable $Z = \xi - v$. Then

$$\mathbb{E}Z^2 = 1 + v^2.$$

By Hölder inequality, we have

$$B := \mathbb{E}\xi^4 \geq (\mathbb{E}\xi^2)^2 = 1,$$

so, using Minkowski inequality, we obtain

$$(\mathbb{E}Z^4)^{1/4} \leq B^{1/4} + v \leq B^{1/4}(1+v) \leq B^{1/4}2^{1/2}(1+v^2)^{1/2}.$$

Using this in (3.1), we conclude that

$$\mathbb{P}(|\xi - v| > \varepsilon) \geq \frac{(1+v^2-\varepsilon^2)^2}{4B(1+v^2)^2} = \frac{1}{4B}\Big(1 - \frac{\varepsilon^2}{1+v^2}\Big)^2 \geq \frac{1-\varepsilon^2}{4B}.$$

This completes the proof. □

We will need a much stronger bound on the concentration function for sums of independent random variables. Here we present a multi-dimensional version of the inverse Littlewood-Offord inequality from [20]. While this paper was in preparation, Friedland and Sodin [8] proposed two different ways to simplify and improve our argument in [20]. We shall therefore present here a multi-dimensional version of one of arguments of Friedland and Sodin [9], which is considerably simpler than our original proof.

We consider the sum

$$S = \sum_{k=1}^N a_k \xi_k$$

where ξ_k are independent and identically distributed random variables, and a_k are some vectors in $\mathbb{R}^m$. The Littlewood-Offord theory describes the behavior of the Lévy concentration function of S in terms of the additive structure of the vectors a_k.

In the scalar case, when $m = 1$, the additive structure of a sequence $a = (a_1, \dots, a_N)$ of real numbers a_k can be described in terms of the shortest arithmetic progression into which it (essentially) embeds. This length is conveniently expressed as the essential *least common denominator* of a, defined as follows. We fix parameters $\alpha, \gamma \in (0,1)$, and define

$$\mathrm{LCD}_{\alpha,\gamma}(a) := \inf\Big\{\theta > 0 : \ \mathrm{dist}(\theta a, \mathbb{Z}^N) < \min(\gamma\|\theta a\|_2, \alpha)\Big\}.$$

The requirement that the distance is smaller than $\gamma\|\theta a\|_2$ forces to consider only non-trivial integer points as approximations of θa – only those in a non-trivial cone around the direction of a. One typically uses this definition with γ a small constant, and for $\alpha = c\sqrt{N}$ with a small constant $c > 0$. The inequality

$\operatorname{dist}(\theta a, \mathbb{Z}^N) < \alpha$ then yields that most coordinates of θa are within a small constant distance from integers.

The definition of the essential least common denominator carries over naturally to higher dimensions and thus allows one to control the arithmetic structure of a sequence $a = (a_1, \dots, a_N)$ of vectors $a_k \in \mathbb{R}^m$. To this end, we define the product of such multi-vector a and a vector $\theta \in \mathbb{R}^m$ as

$$\theta \cdot a = (\langle \theta, a_1 \rangle, \dots, \langle \theta, a_N \rangle) \in \mathbb{R}^N.$$

A more traditional way of looking at $\theta \cdot a$ is to regard it as the product of the matrix a with rows a_k and the vector θ.

Then we define, for $\alpha > 0$ and $\gamma \in (0,1)$,

$$\mathrm{LCD}_{\alpha,\gamma}(a) := \inf \Big\{ \|\theta\|_2 : \ \theta \in \mathbb{R}^m, \operatorname{dist}(\theta \cdot a, \mathbb{Z}^N) < \min(\gamma \|\theta \cdot a\|_2, \alpha) \Big\}.$$

The following theorem gives a bound on the small ball probability for a random sum $S = \sum_{k=1}^N a_k \xi_k$ in terms of the additive structure of the coefficient sequence a. The less structure in a, the bigger its least common denominator is, and the smaller is the small ball probability for S.

Theorem 3.3 (Small ball probability)**.** *Consider a sequence $a = (a_1, \dots, a_N)$ of vectors $a_k \in \mathbb{R}^m$, which satisfies*

$$\sum_{k=1}^N \langle a_k, x \rangle^2 \ge \|x\|_2^2 \qquad \text{for every } x \in \mathbb{R}^m. \tag{3.2}$$

Let $\xi_1, \dots, \xi_N$ be independent and identically distributed, mean zero random variables, such that $\mathcal{L}(\xi_k, 1) \le 1 - b$ for some $b > 0$. Consider the random sum $S = \sum_{k=1}^N a_k \xi_k$. Then, for every $\alpha > 0$ and $\gamma \in (0,1)$, and for

$$\varepsilon \ge \frac{\sqrt{m}}{\mathrm{LCD}_{\alpha,\gamma}(a)},$$

we have

$$\mathcal{L}(S, \varepsilon\sqrt{m}) \le \Big(\frac{C\varepsilon}{\gamma\sqrt{b}} \Big)^m + C^m e^{-2b\alpha^2}.$$

Remark. The non-degeneracy condition (3.2) is meant to guarantee that the system of vectors (a_k) is genuinely m-dimensional. It disallows these vectors to lie on or close to any lower-dimensional subspace of $\mathbb{R}^m$.

Halász [11] developed a powerful approach to bounding concentration function; his approach influenced our arguments below. Halász [11] operated under a similar non-degeneracy condition on the vectors a_k: for every $x \in S^{m-1}$, at least cN terms satisfy $|\langle a_k, x \rangle| \ge 1$. After properly rescaling a_k by the factor $\sqrt{c/N}$, Halász's condition is seen to be more restrictive than (3.2).

3.1. **Proof of the Small Ball Probability Theorem.** To estimate the Lévy concentration function we apply the Esséen Lemma, see e.g. [23], p. 290.

Lemma 3.4. *Let Y be a random vector in $\mathbb{R}^m$. Then*

$$\sup_{v\in\mathbb{R}^m} \mathbb{P}(\|Y-v\|_2 \le \sqrt{m}) \le C^m \int_{B(0,\sqrt{m})} |\phi_Y(\theta)|\, d\theta$$

where $\phi_Y(\theta) = \mathbb{E}\exp(2\pi i\langle\theta, Y\rangle)$ is the characteristic function of Y.

Applying Lemma 3.4 to the vector $Y = S/\varepsilon$ and using the independence of random variables $\xi_1, \dots, \xi_N$, we obtain

$$\mathcal{L}(S, \varepsilon\sqrt{m}) \le C^m \int_{B(0,\sqrt{m})} \prod_{k=1}^N |\phi(\langle\theta, a_k\rangle/\varepsilon)|\, d\theta, \tag{3.3}$$

where $\phi(t) = \mathbb{E}\exp(2\pi i t\xi)$ is the characteristic function of $\xi := \xi_1$. To estimate this characteristic function, we follow the conditioning argument of [18], [20]. Let ξ' be an independent copy of ξ and denote by $\bar\xi$ the symmetric random variable $\xi - \xi'$. Then

$$|\phi(t)|^2 = \mathbb{E}\exp(2\pi i t\bar\xi) = \mathbb{E}\cos(2\pi t\bar\xi).$$

Using the inequality $|x| \le \exp(-\frac{1}{2}(1-x^2))$, which is valid for all $x \in \mathbb{R}$, we obtain

$$|\phi(t)| \le \exp\Big(-\frac{1}{2}\big(1 - \mathbb{E}\cos(2\pi t\bar\xi)\big)\Big).$$

By conditioning on ξ' we see that our assumption $\mathcal{L}(\xi, 1) \le 1-b$ implies that $\mathbb{P}(|\bar\xi| \ge 1) \ge b$. Therefore

$$\begin{aligned}
1 - \mathbb{E}\cos(2\pi t\bar\xi) &\ge \mathbb{P}(|\bar\xi| \ge 1)\cdot \mathbb{E}\Big(1 - \cos(2\pi t\bar\xi) \mid |\bar\xi| \ge 1\Big) \\
&\ge b \cdot \frac{4}{\pi^2}\mathbb{E}\Big(\min_{q\in\mathbb{Z}} |2\pi t\bar\xi - 2\pi q|^2 \mid |\bar\xi| \ge 1\Big) \\
&= 16b \cdot \mathbb{E}\Big(\min_{q\in\mathbb{Z}} |t\bar\xi - q|^2 \mid |\bar\xi| \ge 1\Big).
\end{aligned}$$

Substituting of this into (3.3) and using Jensen's inequality, we get

$$\begin{aligned}
&\mathcal{L}(S, \varepsilon\sqrt{m}) \\
&\le C^m \int_{B(0,\sqrt{m})} \exp\Big(-8b\mathbb{E}\Big(\sum_{k=1}^N \min_{q\in\mathbb{Z}} |\bar\xi\langle\theta, a_k\rangle/\varepsilon - q|^2 \;\Big|\; |\bar\xi| \ge 1\Big)\Big)\, d\theta \\
&\le C^m \mathbb{E}\Big(\int_{B(0,\sqrt{m})} \exp\Big(-8b \min_{p\in\mathbb{Z}^N} \Big\|\frac{\bar\xi}{\varepsilon}\,\theta\cdot a - p\Big\|_2\Big)\, d\theta \;\Big|\; |\bar\xi| \ge 1\Big) \\
&\le C^m \sup_{z\ge 1} \int_{B(0,\sqrt{m})} \exp(-8bf^2(\theta))\, d\theta,
\end{aligned}$$

where

$$f(\theta) = \min_{p \in \mathbb{Z}^N} \Big\| \frac{z}{\varepsilon}\, \theta \cdot a - p \Big\|_2 .$$

The next and major step is to bound the size of the *recurrence set*

$$I(t) := \Big\{ \theta \in B(0, \sqrt{m}) : \ f(\theta) \le t \}.$$

Lemma 3.5 (Size of the recurrence set)**.** *We have*

$$\mathrm{vol}(I(t)) \le \Big(\frac{Ct\varepsilon}{\gamma \sqrt{m}} \Big)^m, \qquad t < \alpha/2.$$

Proof. Fix $t < \alpha/2$. Consider two points $\theta', \theta'' \in I(t)$. There exist $p', p'' \in \mathbb{Z}^N$ such that

$$\Big\| \frac{z}{\varepsilon}\, \theta' \cdot a - p' \Big\|_2 \le t, \quad \Big\| \frac{z}{\varepsilon}\, \theta'' \cdot a - p'' \Big\|_2 \le t.$$

Let

$$\tau := \frac{z}{\varepsilon}\, (\theta' - \theta''), \qquad p := p' - p''.$$

Then, by the triangle inequality,

$$\|\tau \cdot a - p\|_2 \le 2t. \tag{3.4}$$

Recall that by the assumption of the theorem,

$$\mathrm{LCD}_{\alpha,\gamma}(a) \ge \frac{\sqrt{m}}{\varepsilon}.$$

Therefore, by the definition of the least common denominator, we have that either

$$\|\tau\|_2 \ge \frac{\sqrt{m}}{\varepsilon},$$

or otherwise

$$\|\tau \cdot a - p\|_2 \ge \min(\gamma \|\tau \cdot a\|_2, \alpha). \tag{3.5}$$

In the latter case, since $2t < \alpha$, inequalities (3.4) and (3.5) together yield

$$2t \ge \gamma \|\tau \cdot a\|_2 \ge \gamma \|\tau\|_2,$$

where the last inequality follows from condition (3.2).

Recalling the definition of τ, we have proved that every pair of points $\theta', \theta'' \in I(t)$ satisfies:

$$\text{either} \quad \|\theta' - \theta''\|_2 \ge \frac{\sqrt{m}}{z} =: R \quad \text{or} \quad \|\theta' - \theta''\|_2 \le \frac{2t\varepsilon}{\gamma z} =: r.$$

It follows that $I(t)$ can be covered by Euclidean balls of radii r, whose centers are R-separated in the Euclidean distance. Since $I(t) \subset B(0, \sqrt{m})$, the number of such balls is at most

$$\frac{\mathrm{vol}(B(0, \sqrt{m} + R/2))}{\mathrm{vol}(B(0, R/2))} = \Big(\frac{2\sqrt{m}}{R} + 1 \Big)^m \le \Big(\frac{3\sqrt{m}}{R} \Big)^m.$$

(In the last inequality we used that $R \le \sqrt{m}$ because $z \ge 1$). Recall that the volume of a Euclidean ball of radius r in $\mathbb{R}^m$ is bounded by $(Cr/\sqrt{m})^m$. Summing these volumes, we conclude that

$$\mathrm{vol}(I(t)) \le \Big(\frac{3Cr}{R}\Big)^m,$$

which completes the proof of the lemma. □

Proof of Theorem 3.3. We decompose the domain into two parts. First, by the definition of $I(t)$, we have

$$\begin{aligned}\int_{B(0,\sqrt{m})\setminus I(\alpha/2)} \exp(-8bf^2(\theta))\,d\theta &\le \int_{B(0,\sqrt{m})} \exp(-2b\alpha^2)\,d\theta \\ &\le C^m \exp(-2b\alpha^2). \qquad (3.6)\end{aligned}$$

In the last line, we used the estimate $|\mathrm{vol}(B(0,\sqrt{m})| \le C^m$.

Second, by the integral distribution formula and using Lemma 3.5, we have

$$\begin{aligned}\int_{I(\alpha/2)} \exp(-8bf^2(\theta))\,d\theta &= \int_0^{\alpha/2} 16bt\exp(-8bt^2)|\mathrm{vol}(I(t))|\,dt \\ &\le 16b\Big(\frac{C\varepsilon}{\gamma\sqrt{m}}\Big)^m \int_0^\infty t^{m+1}\exp(-8bt^2)\,dt \\ &\le \Big(\frac{C'\varepsilon}{\gamma\sqrt{b}}\Big)^m\sqrt{m} \le \Big(\frac{C''\varepsilon}{\gamma\sqrt{b}}\Big)^m. \qquad (3.7)\end{aligned}$$

Combining (3.6) and (3.7) completes the proof of Theorem 3.3. □

3.2. **Least common denominator of incompressible vectors.** We now prove a simple fact that the least common denominator of any incompressible vector a in $\mathbb{R}^N$ is at least of order $\sqrt{N}$. Indeed, by Lemma 2.5 such a vector has many coordinates of order $1/\sqrt{N}$. Therefore, to make a dilation θa of this vector close to an integer point, one has to scale a by at least $\theta \gtrsim \sqrt{N}$. We now make this heuristic reasoning formal.

Lemma 3.6 (LCD of incompressible vectors)**.** *For every $\delta, \rho \in (0,1)$ there exist $c_1(\delta,\rho) > 0$ and $c_2(\delta) > 0$ such that the following holds. Let $a \in \mathbb{R}^N$ be an incompressible vector: $a \in Incomp(\delta,\rho)$. Then, for every $0 < \gamma < c_1(\delta,\rho)$ and every $\alpha > 0$, one has*

$$\mathrm{LCD}_{\alpha,\gamma}(a) > c_2(\delta)\sqrt{N}.$$

Remark. The proof gives $c_1(\delta,\rho) = \frac{1}{2}\rho^2\sqrt{\delta}$ and $c_2(\delta) = \frac{1}{2}\sqrt{\delta}$.

Proof. By Lemma 2.5, there exists a set $\sigma_1 \subseteq \{1,\dots,N\}$ of size

$$|\sigma_1| \ge \frac{1}{2}\rho^2\delta N$$

and such that

$$\frac{\rho}{\sqrt{2N}} \le |a_k| \le \frac{1}{\sqrt{\delta N}} \qquad \text{for } k \in \sigma_1. \tag{3.8}$$

Let $\theta := \mathrm{LCD}_{\alpha,\gamma}(a)$. Then there exists $p \in \mathbb{Z}^N$ such that

$$\|\theta a - p\|_2 < \gamma \|\theta a\|_2 = \gamma\theta.$$

This shows in particular that $\theta > 0$; dividing by θ gives

$$\Big\| a - \frac{p}{\theta} \Big\|_2 < \gamma.$$

Then by Chebychev inequality, there exists a set $\sigma_2 \subseteq \{1, \dots, N\}$ of size

$$|\sigma_2| > N - \frac{1}{2}\rho^2 \delta N$$

and such that

$$\Big| a_k - \frac{p_k}{\theta} \Big| < \frac{\sqrt{2}}{\rho\sqrt{\delta}} \cdot \frac{\gamma}{\sqrt{N}} \qquad \text{for } k \in \sigma_2. \tag{3.9}$$

Since $|\sigma_1| + |\sigma_2| > N$, there exists $k \in \sigma_1 \cap \sigma_2$. Fix this k. By the left hand side of (3.8), by (3.9) and the assumption on γ we have:

$$\Big| \frac{p_k}{\theta} \Big| \ge \frac{\rho}{\sqrt{2N}} - \frac{\sqrt{2}}{\rho\sqrt{\delta}} \cdot \frac{\gamma}{\sqrt{N}} > 0.$$

Thus $|p_k| > 0$; since p_k is an integer, this yields $|p_k| \ge 1$. Similarly, using the right hand side of (3.8), (3.9) and the assumption on γ, we get

$$\Big| \frac{p_k}{\theta} \Big| \le \frac{1}{\sqrt{\delta N}} + \frac{\sqrt{2}}{\rho\sqrt{\delta}} \cdot \frac{\gamma}{\sqrt{N}} < \frac{2}{\sqrt{\delta N}}.$$

Since $|p_k| \ge 1$, this yields

$$|\theta| > \frac{\sqrt{\delta N}}{2}.$$

This completes the proof. □

4. The distance problem and arithmetic structure

Here we use the Small Ball Probability Theorem 3.3 to give an optimal bound for the distance problem: *how close is a random vector X in $\mathbb{R}^N$ from an independent random subspace H of codimension m?*

If X has the standard normal distribution, then the distance does not depend on the distribution of H. Indeed, for an arbitrary fixed H, the distance $\mathrm{dist}(X, H)$ is distributed identically with the Euclidean norm of a standard normal random vector in $\mathbb{R}^m$. Therefore,

$$\mathrm{dist}(X, H) \sim \sqrt{m} \qquad \text{with high probability.}$$

More precisely, standard computations give for every $\varepsilon > 0$ that

$$\mathbb{P}\big(\mathrm{dist}(X, H) < \varepsilon\sqrt{m}\big) \le (C\varepsilon)^m. \tag{4.1}$$

However, if X has a more general distribution with independent coordinates, the distance $\mathrm{dist}(X, H)$ may strongly depend on the subspace H. For example, if the coordinates of X are ± 1 symmetric random variables. then for $H = \{x : \ x_1 + x_2 = 0\}$ the distance equals 0 with probability $1/2$, while for $H = \{x : \ x_1 + \cdots + x_N = 0\}$ the distance equals 0 with probability $\sim 1/\sqrt{N}$.

Nevertheless, a version of the distance bound (4.1) remains true for general distributions if H is a random subspace. For spaces of codimension $m = 1$, this result was proved in [20]. In this paper, we prove an optimal distance bound for general dimensions.

Theorem 4.1 (Distance to a random subspace)**.** *Let X be a vector in $\mathbb{R}^N$ whose coordinates are independent and identically distributed mean zero subgaussian random variables with unit variance. Let H be a random subspace in $\mathbb{R}^N$ spanned by $N - m$ vectors, $0 < m < \tilde{c}N$, whose coordinates are independent and identically distributed mean zero subgaussian random variables with unit variance, independent of X. Then, for every $v \in \mathbb{R}^N$ and every $\varepsilon > 0$, we have*

$$\mathbb{P}\big(\mathrm{dist}(X, H + v) < \varepsilon\sqrt{m}\big) \le (C\varepsilon)^m + e^{-cN},$$

where $C, c, \tilde{c} > 0$ depend only on the subgaussian moments.

Remark. To explain the term e^{-cN}, consider ± 1 symmetric random variables. Then with probability at least 2^{-n} the random vector X coincides with one of the random vectors that span H, which makes the distance equal zero.

We will deduce Theorem 4.1 from a more general inequality that holds for *arbitrary* fixed subspace H. This bound will depend on the arithmetic structure of the subspace H, which we express using the least common denominator.

For $\alpha > 0$ and $\gamma \in (0, 1)$, the essential *least common denominator of a subspace E* in $\mathbb{R}^N$ is defined as

$$\mathrm{LCD}_{\alpha,\gamma}(E) := \inf\{\mathrm{LCD}_{\alpha,\gamma}(a) : \ a \in S(E)\}.$$

Clearly,

$$\mathrm{LCD}_{\alpha,\gamma}(E) = \inf\Big\{\|\theta\|_2 : \ \theta \in E, \ \mathrm{dist}(\theta, \mathbb{Z}^N) < \min(\gamma\|\theta\|_2, \alpha)\Big\}.$$

Then Theorem 3.3 quickly leads to the following general distance bound:

Theorem 4.2 (Distance to a general subspace)**.** *Let X be a vector in $\mathbb{R}^N$ whose coordinates are independent and identically distributed mean zero subgaussian*

random variables with unit variance. Let H be a subspace in $\mathbb{R}^N$ of dimension $N-m>0$. Then for every $v \in \mathbb{R}^N$, $\alpha > 0$, $\gamma \in (0,1)$, and for

$$\varepsilon \geq \frac{\sqrt{m}}{\mathrm{LCD}_{\alpha,\gamma}(H^\perp)},$$

we have

$$\mathbb{P}\big(\mathrm{dist}(X, H+v) < \varepsilon\sqrt{m}\big) \leq \Big(\frac{C\varepsilon}{\gamma}\Big)^m + C^m e^{-c\alpha^2}$$

where $C, c > 0$ depend only on the subgaussian moment.

Proof. Let us write X in coordinates, $X = (\xi_1, \dots, \xi_N)$. By Lemma 3.2 and the remark below it, all coordinates of X satisfy the inequality $\mathcal{L}(\xi_k, 1/2) \leq 1-b$ for some $b > 0$ that depends only on the subgaussian moment of ξ_k. Hence the random variables $\xi_k/2$ satisfy the assumption in Theorem 3.3.

Next, we connect the distance to a sum of independent random vectors:

$$\mathrm{dist}(X, H+v) = \|P_{H^\perp}(X-v)\|_2 = \Big\| \sum_{k=1}^N a_k \xi_k - w \Big\|_2, \tag{4.2}$$

where

$$a_k = P_{H^\perp} e_k, \quad w = P_{H^\perp} v,$$

and where $e_1, \dots, e_N$ denotes the canonical basis of $\mathbb{R}^N$. Therefore, the sequence of vectors $a = (a_1, \dots, a_N)$ is in the isotropic position:

$$\sum_{k=1}^N \langle a_k, x\rangle^2 = \|x\|_2^2 \qquad \text{for any } x \in H^\perp,$$

so we can use Theorem 3.3 in the space $H^\perp$ (identified with $\mathbb{R}^m$ by a suitable isometry).

For every $\theta = (\theta_1, \dots, \theta_N) \in H^\perp$ and every k we have $\langle \theta, a_k\rangle = \langle P_{H^\perp}\theta, e_k\rangle = \langle \theta, e_k\rangle = \theta_k$, so

$$\theta \cdot a = \theta$$

where the right hand side is considered as a vector in $\mathbb{R}^N$. Therefore the least common denominator of a subspace can be expressed by that of a sequence of vectors $a = (a_1, \dots, a_N)$:

$$\mathrm{LCD}_{\alpha,\gamma}(H^\perp) = \mathrm{LCD}_{\alpha,\gamma}(a).$$

The theorem now follows directly from Theorem 3.3. □

In order to deduce the Distance Theorem 4.1, it will now suffice to bound below the least common denominator of a random subspace $H^\perp$. Heuristically, the randomness should remove any arithmetic structure from the subspace, thus making the least common denominator exponentially large. Our next results shows that this is indeed true.

Theorem 4.3 (Structure of a random subspace)**.** *Let H be a random subspace in $\mathbb{R}^N$ spanned by $N-m$ vectors, $1 \le m < \tilde{c}N$, whose coordinates are independent and identically distributed mean zero subgaussian random variables with unit variance. Then, for $\alpha = c\sqrt{N}$, we have*

$$\mathbb{P}\big(\mathrm{LCD}_{\alpha,c}(H^\perp) < c\sqrt{N} e^{cN/m} \big) \le e^{-cN},$$

where $c \in (0,1)$ and $\tilde{c} \in (0,1/2)$ depend only on the subgaussian moment.

Assuming that this result holds, we can complete the proof of the Distance Theorem 4.1.

Proof of Theorem 4.1. Consider the event

$$\mathcal{E} := \big\{ \mathrm{LCD}_{\alpha,c}(H^\perp) \ge c\sqrt{N} e^{cN/m} \big\}.$$

By Theorem 4.3, $\mathbb{P}(\mathcal{E}^c) \le e^{-cN}$.

Let us condition on a realization of H in $\mathcal{E}$. By the independence of X and H, Theorem 4.2 used with $\alpha = c\sqrt{N}$ and $\gamma = c$ gives

$$\mathbb{P}\big(\mathrm{dist}(X,H) < \varepsilon\sqrt{m} \mid \mathcal{E}\big) \le (C_1\varepsilon)^m + C^m e^{-c_1 N}$$

for every

$$\varepsilon > C_2 \sqrt{\frac{m}{N}} e^{-cN/m}.$$

Since $m \le \tilde{c}N$, with an appropriate choice of $\tilde{c}$ we get

$$C_2 \sqrt{\frac{m}{N}} e^{-cN/m} \le \frac{1}{C_1} e^{-c_3 N/m} \quad \text{and } C^m e^{-c_1 N} \le e^{-c_3 N}.$$

Therefore, for every $\varepsilon > 0$,

$$\mathbb{P}\big(\mathrm{dist}(X,H) < \varepsilon\sqrt{m} \mid \mathcal{E}\big) \le (C_1\varepsilon)^m + 2e^{-c_3 N} \le (C_1\varepsilon)^m + e^{-c_4 N}.$$

By the estimate on the probability of $\mathcal{E}^c$, this completes the proof. □

4.1. **Proof of the Structure Theorem 4.3.** Note first, that throughout the proof we can assume that $N > N_0$, where N_0 is a suitably large number, which may depend only the subgaussian moment. Indeed, the assumption on m implies that $N > m/\tilde{c} \ge 1/\tilde{c}$. Choosing $\tilde{c} > 0$ suitably small depending on the subgaussian moment, we can make N_0 suitably large.

Let $X_1, \dots, X_{N-m}$ denote the independent random vectors that span the subspace H. Consider an $(N-m) \times N$ random matrix B with rows X_k. Then

$$H^\perp \subseteq \ker(B).$$

Therefore, for every set S in $\mathbb{R}^N$ we have:

$$\inf_{x \in S} \|Bx\|_2 > 0 \text{ implies } H^\perp \cap S = \emptyset. \tag{4.3}$$

This observation will help us to "navigate" the random subspace $H^\perp$ away from undesired sets S on the unit sphere.

We start with a variant of Lemma 3.6 of [20]; here we use the concept of compressible and incompressible vectors in $\mathbb{R}^N$ rather than $\mathbb{R}^n$.

Lemma 4.4 (Random subspaces are incompressible)**.** *There exist $\delta, \rho \in (0,1)$ such that*

$$\mathbb{P}\big(H^\perp \cap S^{N-1} \subseteq \mathit{Incomp}(\delta,\rho)\big) \geq 1 - e^{-cN}.$$

Proof. Let B be the $(N-m)\times N$ matrix defined above. Since $N-m > (1-\tilde{c})N$ and $\tilde{c} < 1/2$, we can apply Lemma 2.6 for the matrix B. Thus, there exist $\delta, \rho \in (0,1)$ such that

$$\mathbb{P}\big(\inf_{x\in \mathit{Comp}(\delta,\rho)} \|Bx\|_2 \geq c_3\sqrt{N}\big) \geq 1 - e^{-c_3 N}.$$

By (4.3), $H^\perp \cap \mathit{Comp}(\delta,\rho) = \emptyset$ with probability at least $1 - e^{-c_3 N}$. □

Fix the values of δ and ρ given by Lemma 4.4 for the rest of this section. We will further decompose the set of incompressible vectors into level sets S_D according to the value of the least common denominator D. We shall prove a nontrivial lower bound on $\inf_{x\in S_D} \|Bx\|_2 > 0$ for each level set up to D of the exponential order. By (4.3), this will mean that $H^\perp$ is disjoint from every such level set. Therefore, all vectors in $H^\perp$ must have exponentially large least common denominators D. This is Theorem 4.3.

Let $\alpha = \mu\sqrt{N}$, where $\mu > 0$ is a small number to be chosen later, which depends only on the subgaussian moment. By Lemma 3.6,

$$\mathrm{LCD}_{\alpha,c}(x) \geq c_0\sqrt{N} \qquad \text{for every } x \in \mathit{Incomp}.$$

Definition 4.5 (Level sets)**.** Let $D \geq c_0\sqrt{N}$. Define $S_D \subseteq S^{N-1}$ as

$$S_D := \big\{x \in \mathit{Incomp} :\ D \leq \mathrm{LCD}_{\alpha,c}(x) < 2D\big\}.$$

To obtain a lower bound for $\|Bx\|_2$ on the level set, we proceed by an ε-net argument. To this end, we first need such a bound for a single vector x.

Lemma 4.6 (Lower bound for a single vector)**.** *Let $x \in S_D$. Then for every $t > 0$ we have*

$$\mathbb{P}\big(\|Bx\|_2 < t\sqrt{N}\big) \leq \Big(Ct + \frac{C}{D} + Ce^{-c\alpha^2}\Big)^{N-m}. \tag{4.4}$$

Proof. Denoting the elements of B by ξ_{jk}, we can write the j-th coordinate of Bx as

$$(Bx)_j = \sum_{j=1}^{N} \xi_{jk} x_k =: \zeta_j, \qquad j = 1, \dots, N-m.$$

Now we can use the Small Ball Probability Theorem 3.3 in dimension $m = 1$ for each of these random sums. By Lemma 3.2 and the remark below it, $\mathcal{L}(\xi_{jk}, 1/2) \le 1 - b$ for some $b > 0$ that depends only on the subgaussian moment of ξ_{jk}. Hence the random variables $\xi_{jk}/2$ satisfy the assumption in Theorem 3.3. This gives for every j and every $t > 0$:

$$\mathbb{P}(|\zeta_j| < t) \le Ct + \frac{C}{\mathrm{LCD}_{\alpha,c}(x)} + Ce^{-c\alpha^2} \le Ct + \frac{C}{D} + Ce^{-c\alpha^2}.$$

Since ζ_j are independent random variables, we can use Tensorization Lemma 2.2 of [20] to conclude that for every $t > 0$,

$$\mathbb{P}\Big(\sum_{j=1}^{N-m} |\zeta_j|^2 < t^2(N-m) \Big) \le \Big(C''t + \frac{C''}{D} + C''e^{-c\alpha^2} \Big)^{N-m}.$$

This completes the proof, because $\|Bx\|_2^2 = \sum_{j=1}^{N-m} |\zeta_j|^2$ and $N \le 2(N-m)$ by the assumption. □

Next, we construct a small ε-net of the level set S_D. Since this set lies in S^{N-1}, Lemma 2.1 yields the existence of an $(\sqrt{N}/D)$-net of cardinality at most $(CD/\sqrt{N})^N$. This simple volumetric bound is not sufficient for our purposes, and this is the crucial step where we explore the additive structure of S_D to construct a smaller net.

Lemma 4.7 (Nets of level sets)**.** *There exists a $(4\alpha/D)$-net of S_D of cardinality at most $(C_0 D/\sqrt{N})^N$.*

Remark. Recall that α is chosen as a small proportion of $\sqrt{N}$. Hence Lemma 4.7 gives a better bound than the standard volumetric bound in Lemma 2.1.

Proof. We can assume that $4\alpha/D \le 1$, otherwise the conclusion is trivial. For $x \in S_D$, denote

$$D(x) := \mathrm{LCD}_{\alpha,c}(x).$$

By the definition of S_D, we have $D \le D(x) < 2D$. By the definition of the least common denominator, there exists $p \in \mathbb{Z}^N$ such that

$$\|D(x)x - p\|_2 < \alpha. \tag{4.5}$$

Therefore

$$\Big\| x - \frac{p}{D(x)} \Big\|_2 < \frac{\alpha}{D(x)} \le \frac{\alpha}{D} \le \frac{1}{4}.$$

Since $\|x\|_2 = 1$, it follows that

$$\Big\| x - \frac{p}{\|p\|_2} \Big\|_2 < \frac{2\alpha}{D}. \tag{4.6}$$

On the other hand, by (4.5) and using that $\|x\|_2 = 1$, $D(x) \le 2D$ and $4\alpha/D \le 1$, we obtain

$$\|p\|_2 < D(x) + \alpha \le 2D + \alpha \le 3D. \tag{4.7}$$

Inequalities (4.6) and (4.7) show that every point $x \in S_D$ is within Euclidean distance $2\alpha/D$ from the set

$$\mathcal{N} := \Big\{ \frac{p}{\|p\|_2} : \ p \in \mathbb{Z}^N \cap B(0, 3D) \Big\}.$$

A known volumetric argument gives a bound on the number of integer points in $B(0, 3D)$:

$$|\mathcal{N}| \le (1 + 9D/\sqrt{N})^N \le (C_0 D/\sqrt{N})^N$$

(where in the last inequality we used that by Definition 4.5 of the level sets, $D > c_0\sqrt{N}$). Finally, there exists a $(4\alpha/D)$-net of S_D with the same cardinality as $\mathcal{N}$, and which lies in S_D. Indeed, to obtain such a net, one selects one (arbitrary) point from the intersection of S_D with a ball of radius $2\alpha/D$ centered at each point from $\mathcal{N}$. This completes the proof. □

Lemma 4.8 (Lower bound for a level set)**.** *There exist $c_1, c_2, \mu \in (0, 1)$ such that the following holds. Let $\alpha = \mu\sqrt{N} \ge 1$ and $D \le c_1\sqrt{N}e^{c_1 N/m}$. Then*

$$\mathbb{P}\big(\inf_{x \in S_D} \|Bx\|_2 < c_2 N/D \big) \le 2e^{-N}.$$

Proof. By Lemma 2.3, there exists $K \ge 1$ that depends only on the subgaussian moment and such that

$$\mathbb{P}(\|B\| > K\sqrt{N}) \le e^{-N}.$$

Therefore, in order to complete the proof, it is enough to find $\nu > 0$ which depends only on the subgaussian moment, and such that the event

$$\mathcal{E} := \Big\{ \inf_{x \in S_D} \|Bx\|_2 < \frac{\nu N}{2D} \text{ and } \|B\| \le K\sqrt{N} \Big\}$$

has probability at most e^{-N}.

We claim that this holds with the following choice of parameters:

$$\nu = \frac{1}{(3CC_0)^2 e}, \quad \mu = \frac{\nu}{9K}, \quad c_1 = c\mu^2 \le \nu,$$

where $C \ge 1$ and $c \in (0, 1)$ are the constants from Lemma 4.6 and $C_0 \ge 1$ is the constant from Lemma 4.7.

By choosing $\tilde{c}$ in the statement of Theorem 4.3 suitably small, we can assume that $N > \nu^{-2}$ (this is because by the assumptions, $N > m/\tilde{c} \ge 1/\tilde{c}$). We apply Lemma 4.6 with $t = \nu\sqrt{N}/D$. Then recalling the choice of α and c_1 and our

assumption on D, one easily checks that the term Ct dominates in the right hand side of (4.4):

$$t \geq 1/D \quad \text{and} \quad t \geq e^{-c\alpha^2}.$$

This gives for arbitrary $x_0 \in S_D$:

$$\mathbb{P}\Big(\|Bx_0\|_2 < \frac{\nu N}{D}\Big) \leq \Big(\frac{3C\nu\sqrt{N}}{D}\Big)^{N-m}.$$

Now we use Lemma 4.7, which yields a small $(4\alpha/D)$-net $\mathcal{N}$ of S_D. Taking the union bound, we get

$$p := \mathbb{P}\Big(\inf_{x_0\in\mathcal{N}} \|Bx_0\|_2 < \frac{\nu N}{D}\Big) \leq \Big(\frac{C_0 D}{\sqrt{N}}\Big)^N \Big(\frac{3C\nu\sqrt{N}}{D}\Big)^{N-m}.$$

Denote $C_1 := 3CC_0$. Using the fact that $c_1 \leq \nu$ and our assumption on D, we have:

$$p \leq C_1^N \Big(\frac{D}{\sqrt{N}}\Big)^m \nu^{N-m} \leq C_1^N (\nu e^{\nu N/m})^m \nu^{N-m} \leq C_1^{2N} \nu^N = e^{-N}. \tag{4.8}$$

Assume $\mathcal{E}$ occurs. Fix $x \in S_D$ for which $\|Bx\|_2 < \frac{\nu N}{2D}$; it can be approximated by some element $x_0 \in \mathcal{N}$ as

$$\|x - x_0\|_2 \leq \frac{4\mu\sqrt{N}}{D}.$$

Therefore, by the triangle inequality we have

$$\|Bx_0\|_2 \leq \|Bx\|_2 + \|B\| \cdot \|x - x_0\|_2 \leq \frac{\nu N}{2D} + K\sqrt{N} \cdot \frac{4\mu\sqrt{N}}{D} < \frac{\nu N}{D},$$

where in the last inequality we used our choice of μ.

We have shown that the event $\mathcal{E}$ implies the event that

$$\inf_{x_0\in\mathcal{N}} \|Bx_0\|_2 < \frac{\nu N}{D},$$

whose probability is at most e^{-N} by (4.8). The proof is complete. □

Proof of Theorem 4.3. Consider $x \in S^{N-1}$ such that

$$\mathrm{LCD}_{\alpha,c}(x) < c_1\sqrt{N} e^{c_1 N/m},$$

where c_1 is the constant from Lemma 4.8. Then, by the Definition 4.5 of the level sets, either x is compressible or $x \in S_D$ for some $D \in \mathcal{D}$, where

$$\mathcal{D} := \{D : \ c_0\sqrt{N} \leq D < c_1\sqrt{N} e^{c_1 N/m}, \ D = 2^k, \ k \in \mathbb{N}\}.$$

Therefore, recalling the definition of the least common denominator of the subspace

$$\mathrm{LCD}_{\alpha,c}(H^\perp) = \inf_{x\in S(H^\perp)} \mathrm{LCD}_{\alpha,c}(x),$$

we can decompose the desired probability as follows:

$$\begin{aligned} p &:= \mathbb{P}\big(\operatorname{LCD}_{\alpha,c}(H^\perp) < c_1\sqrt{N}e^{c_1 N/m}\big) \\ &\le \mathbb{P}(H^\perp \cap \mathit{Comp} \neq \emptyset) + \sum_{D\in\mathcal{D}} \mathbb{P}(H^\perp \cap S_D \neq \emptyset). \end{aligned}$$

By Lemma 4.4, the first term in the right hand side is bounded by e^{-cN}. Further terms can be bonded using (4.3) and Lemma 4.8:

$$\mathbb{P}(H^\perp \cap S_D \neq \emptyset) \le \mathbb{P}\big(\inf_{x\in S_D} \|Bx\|_2 = 0\big) \le 2e^{-N}.$$

Since there are $|\mathcal{D}| \le C'N$ terms in the sum, we conclude that

$$p \le e^{-cN} + C'Ne^{-N} \le e^{-c'N}.$$

This completes the proof. □

5. Decomposition of the sphere

Now we begin the proof of Theorem 1.1. We will make several useful reductions first.

Without loss of generality, we can assume that the entries of A have a an absolutely continuous distribution. Indeed, we can add to each entry an independent Gaussian random variable with small variance σ, and later let $\sigma \to 0$.

Similarly, we can assume that $n \ge n_0$, where n_0 is a suitably large number that depends only on the subgaussian moment B.

We let

$$N = n - 1 + d$$

for some $d \ge 1$. We can assume that

$$1 \le d \le c_0 n, \tag{5.1}$$

with suitably small constant $c_0 > 0$ that depends only on the subgaussian moment B. Indeed, as we remarked in the Introduction, for the values of d above a constant proportion of n, Theorem 1.1 follows from (1.7). Note that

$$\sqrt{N} - \sqrt{n-1} \le \frac{d}{\sqrt{n}}.$$

Using the decomposition of the sphere $S^{n-1} = \mathit{Comp} \cup \mathit{Incomp}$, we break the invertibility problem into two subproblems, for compressible and incompressible vectors:

$$\begin{aligned}(5.2)\quad \mathbb{P}\Big(s_n(A) \le \varepsilon\big(\sqrt{N} - \sqrt{n-1}\,\big)\Big) &\le \mathbb{P}\big(s_n(A) \le \varepsilon \frac{d}{\sqrt{n}}\big) \\ &\le \mathbb{P}\big(\inf_{x \in \mathit{Comp}(\delta,\rho)} \|Ax\|_2 \le \varepsilon \frac{d}{\sqrt{n}}\big) + \mathbb{P}\big(\inf_{x \in \mathit{Incomp}(\delta,\rho)} \|Ax\|_2 \le \varepsilon \frac{d}{\sqrt{n}}\big).\end{aligned}$$

A bound for the compressible vectors follows from Lemma 2.6. Using (5.1) we get

$$\varepsilon \frac{d}{\sqrt{n}} \le c_0 \sqrt{n} \le c_0 \sqrt{N}.$$

Hence, Lemma 2.6 implies

$$(5.3)\qquad \mathbb{P}\Big(\inf_{x \in \mathit{Comp}(\delta,\rho)} \|Ax\|_2 \le \varepsilon \frac{d}{\sqrt{n}}\Big) \le e^{-c_3 N}.$$

It remains to find a lower bound on $\|Ax\|$ for the incompressible vectors x.

6. Invertibility via uniform distance bounds

In this section, we reduce the problem of bounding $\|Ax\|_2$ for incompressible vectors x to the distance problem that we addressed in Section 4.

Let $X_1, \dots, X_n \in \mathbb{R}^N$ denote the columns of the matrix A. Given a subset $J \subseteq \{1, \dots, n\}$ of cardinality d, we consider the subspace

$$H_J := \operatorname{span}(X_k)_{k \in J} \subset \mathbb{R}^N.$$

For levels $K_1, K_2 > 0$ that will only depend on δ, ρ, we define the set of totally spread vectors

$$(6.1)\qquad \mathrm{Spread}_J := \Big\{ y \in S(\mathbb{R}^J) : \ \frac{K_1}{\sqrt{d}} \le |y_k| \le \frac{K_2}{\sqrt{d}} \quad \text{for all } k \in J \Big\}.$$

In the following lemma, we let J be a random subset uniformly distributed over all subsets of $\{1, \dots, n\}$ of cardinality d. To avoid confusion, we often denote the probability and expectation over the random set J by $\mathbb{P}_J$ and $\mathbb{E}_J$, and with respect to the random matrix A by $\mathbb{P}_A$ and $\mathbb{E}_A$.

Lemma 6.1 (Total spread)**.** *For every $\delta, \rho \in (0,1)$, there exist $K_1, K_2, c_0 > 0$ which depend only on δ, ρ, and such that the following holds. For every $x \in \mathit{Incomp}(\delta, \rho)$, the event*

$$\mathcal{E}(x) := \Big\{ \frac{P_J x}{\|P_J x\|_2} \in \mathrm{Spread}_J \quad \textit{and} \quad \frac{\rho\sqrt{d}}{\sqrt{2n}} \le \|P_J x\|_2 \le \frac{\sqrt{d}}{\sqrt{\delta n}} \Big\}$$

satisfies $\mathbb{P}_J(\mathcal{E}(x)) > c_0^d$.

Remark. The proof gives $K_1 = \rho\sqrt{\delta/2}$, $K_2 = 1/K_1$, $c_0 = \rho^2\delta/2e$. In the rest of the proof, we shall use definition (6.1) of Spread_J with these values of the levels K_1, K_2.

Proof. Let $\sigma \subset \{1,\dots,n\}$ be the subset from Lemma 2.5. Recall that the parameters δ and ρ depend only on the subgaussian moment B (see Lemma 2.6). By choosing the constant c_0 in (5.1) appropriately small, we may assume that $d \le |\sigma|/2$. Then, using Stirling's approximation we have

$$\mathbb{P}_J(J \subset \sigma) = \binom{|\sigma|}{d} \Big/ \binom{n}{d} > \Big(\frac{\rho^2\delta}{2e}\Big)^d = c_0^d.$$

If $J \subset \sigma$, then summing (2.1) over $k \in J$, we obtain the required two-sided bound for $\|P_J x\|_2$. This and (2.1) yields $\frac{P_J x}{\|P_J x\|_2} \in \mathrm{Spread}_J$. Hence $\mathcal{E}(x)$ holds. □

Lemma 6.2 (Invertibility via distance)**.** *Let $\delta, \rho \in (0,1)$. There exist $C_1, c_1 > 0$ which depend only on δ, ρ, and such that the following holds. Let J be any d-element subset of $\{1,\dots,n\}$. Then for every $\varepsilon > 0$*

$$\mathbb{P}\Big(\inf_{x \in Incomp(\delta,\rho)} \|Ax\|_2 < c_1\varepsilon\sqrt{\frac{d}{n}}\Big) \le C_1^d \cdot \mathbb{P}\big(\inf_{z \in \mathrm{Spread}_J} \mathrm{dist}(Az, H_{J^c}) < \varepsilon\big). \tag{6.2}$$

Remark. The proof gives $K_1 = \rho\sqrt{\delta/2}$, $K_2 = 1/K_1$, $c_1 = \rho/\sqrt{2}$, $C_1 = 2e/\rho^2\delta$.

Proof. Let $x \in Incomp(\delta,\rho)$. For every subset J of $\{1,\dots,n\}$ we have

$$\|Ax\|_2 \ge \mathrm{dist}(Ax, H_{J^c}) = \mathrm{dist}(AP_J x, H_{J^c}).$$

In case the event $\mathcal{E}(x)$ of Lemma 6.1 holds, we use the vector $z = \frac{P_J x}{\|P_J x\|_2} \in \mathrm{Spread}_J$ to check that

$$\|Ax\|_2 \ge \|P_J x\|_2\, D(A,J),$$

where the random variable

$$D(A,J) = \inf_{z \in \mathrm{Spread}_J} \mathrm{dist}(Az, H_{J^c})$$

is independent of x. Moreover, using the estimate on $\|P_J x\|_2$ in the definition of the event $\mathcal{E}(x)$, we conclude that

$$\mathcal{E}(x) \quad \text{implies} \quad \|Ax\|_2 \ge c_1\sqrt{\frac{d}{n}}\, D(A,J). \tag{6.3}$$

Define the event

$$\mathcal{F} := \big\{A : \ \mathbb{P}_J(D(A,J) \ge \varepsilon) > 1 - c_0^d\big\},$$

where c_0 is the constant from Lemma 6.1. Chebychev inequality and Fubini theorem then yield

$$\mathbb{P}_A(\mathcal{F}^c) \le c_0^{-d}\mathbb{E}_A\mathbb{P}_J(D(A,J) < \varepsilon) = c_0^{-d}\mathbb{E}_J\mathbb{P}_A(D(A,J) < \varepsilon).$$

Since the entries of A are independent and identically distributed, the probability $\mathbb{P}_A(D(A,J)<\varepsilon)$ does not depend on J. Therefore, the right hand side of the previous inequality coincides with the right hand side of (6.2).

Fix any realization of A for which $\mathcal{F}$ occurs, and fix any $x \in \mathit{Incomp}(\delta,\rho)$. Then

$$\mathbb{P}_J(D(A,J)\ge\varepsilon)+\mathbb{P}_J(\mathcal{E}(x))>(1-c_0^d)+c_0^d=1,$$

so we conclude that

$$\mathbb{P}_J\big(\mathcal{E}(x) \text{ and } D(A,J)\ge\varepsilon\big)>0. \tag{6.4}$$

We have proved that for every $x \in \mathit{Incomp}(\delta,\rho)$ there exists a subset $J=J(x)$ that satisfies both $\mathcal{E}(x)$ and $D(A,J)\ge\varepsilon$. Using this J in (6.3), we conclude that every matrix A for which the event $\mathcal{F}$ occurs satisfies

$$\inf_{x\in \mathit{Incomp}(\delta,\rho)}\|Ax\|_2\ge\varepsilon c_1\sqrt{\frac{d}{n}}.$$

This and the estimate of $\mathbb{P}_A(\mathcal{F}^c)$ completes the proof. □

7. The uniform distance bound

In this section, we shall estimate the distance between a random ellipsoid and a random independent subspace. This is the distance that we need to bound in the right hand side of (6.2).

Throughout this section, we let J be a fixed subset of $\{1,\dots,n\}$, $|J|=d$. We shall use the notation introduced in the beginning of Section 6. Thus, H_J denotes a random subspace, and Spread_J denotes the totally spread set whose levels K_1, K_2 depend only on δ, ρ in the definition of incompressibility.

We will denote by $K, K_0, C, c, C_1, c_1, \dots$ positive numbers that depend only on δ, ρ and the subgaussian moment B.

Theorem 7.1 (Uniform distance bound)**.** *For every $t>0$,*

$$\mathbb{P}\Big(\inf_{z\in\mathrm{Spread}_J}\mathrm{dist}(Az,H_{J^c})<t\sqrt{d}\Big)\le(Ct)^d+e^{-cN}.$$

Recall that H_{J^c} is the span of $n-d$ independent random vectors. Since their distribution is absolutely continuous (see the beginning of Section 5), these vectors are almost surely in general position, so

$$\dim(H_{J^c})=n-d. \tag{7.1}$$

Without loss of generality, in the proof of Theorem 7.1 we can assume that

$$t\ge t_0=e^{-\bar{c}N/d} \tag{7.2}$$

with a suitably small $\bar{c}>0$.

7.1. **First approach: nets and union bound.** We would like to prove Theorem 7.1 by a typical ε-net argument. Theorem 4.1 will give a useful probability bound for an individual $z \in S^{n-1}$. We might then take a union bound over all z in an ε-net of Spread_J and complete by approximation. However, the standard approximation argument will leave us with a larger error e^{-cd} on the probability, which is unsatisfactory for small d. To improve upon this step, we shall improve upon this approach using decoupling in Section 7.2.

For now, we start with a bound for an individual $z \in S^{n-1}$.

Lemma 7.2. *Let $z \in S^{n-1}$ and $v \in \mathbb{R}^N$. Then for every t that satisfies* (7.2) *we have*

$$\mathbb{P}\Big(\mathrm{dist}(Az, H_{J^c} + v) < t\sqrt{d}\Big) \le (C_1 t)^{2d-1}.$$

Proof. Denote the entries of matrix A by ξ_{ij}. Then the entries of the random vector Az,

$$\zeta_i := (Az)_i = \sum_{j=1}^n \xi_{ij} z_j, \quad j = 1, \dots, N,$$

are independent and identically distributed mean zero random variables. Moreover, since the random variables ξ_{ij} are subgaussian and $\sum_{j=1}^n z_j^2 = 1$, the random variables ζ_i are also subgaussian (see Fact 2.1 in [16]).

Therefore the random vector $X = Az$ and the random subspace $H = H_{J^c}$ satisfy the assumptions of Theorem 4.1 with $m = N - (n-d) = 2d - 1$ (we used (7.1) here). An application of Theorem 4.1 completes the proof. □

We will use this bound for every z in an ε-net of Spread_J. To extend the bound to the whole set Spread_J by approximation, we need a certain stability of the distance. This is easy to quantify and prove using the following representation of the distance in matrix form. Let P be the orthogonal projection in $\mathbb{R}^N$ onto $(H_{J^c})^\perp$, and let

$$W := PA|_{\mathbb{R}^J}. \tag{7.3}$$

Then for every $v \in \mathbb{R}^N$, the following identity holds:

$$\mathrm{dist}(Az, H_{J^c} + v) = \|Wz - w\|_2, \quad \text{where } w = Pv. \tag{7.4}$$

Since $|J| = d$ and almost surely $\dim(H_{J^c})^\perp = N - (n-d) = 2d - 1$, the random matrix W acts as an operator from a d-dimensional subspace into a $(2d-1)$-dimensional subspace. Although the entries of W are not necessarily independent, we expect W to behave as if this was the case. To this end, we condition on the realization of the subspace (H_{J^c}). Now the operator P becomes a fixed projection, and the columns of W become independent random vectors. Then W satisfies a version of Proposition 2.3:

Proposition 7.3. *Let P be an orthogonal projection in $\mathbb{R}^N$ of rank d and let $W = PA|_{\mathbb{R}^J}$ be a random matrix. Then*

$$\mathbb{P}\big(\|W\| > t\sqrt{d}\big) \le e^{-c_0 t^2 d} \qquad \text{for } t \ge C_0.$$

Proof. The argument is similar to that of Proposition 2.3. Let $\mathcal{N}$ be a (1/2)-net of $S(\mathbb{R}^J)$ and M be a (1/2)-net of $S(P\mathbb{R}^N)$. Note that for $x \in \mathcal{N}$, $y \in \mathcal{M}$, we have $\langle Wx, y\rangle = \langle Ax, y\rangle$. The proof is completed as in Proposition 2.3. □

Using Proposition7.3, we can choose a constant K_0 that depends only on the subgaussian moment, and such that

$$\mathbb{P}\big(\|W\| > K_0\sqrt{d}\big) \le e^{-d}. \tag{7.5}$$

With this bound on the norm of W, we can run the approximation argument and prove the distance bound in Lemma 7.2 uniformly over all $z \in \mathrm{Spread}_J$.

Lemma 7.4. *Let W be a random matrix as in Proposition 7.3. Then for every t that satisfies* (7.2) *we have*

$$\mathbb{P}\Big(\inf_{z\in \mathrm{Spread}_J} \|Wz\|_2 < t\sqrt{d} \text{ and } \|W\| \le K_0\sqrt{d}\Big) \le (C_2 t)^d. \tag{7.6}$$

Proof. Let $\varepsilon = t/K_0$. By Proposition 2.1, there exists an ε-net $\mathcal{N}$ of $\mathrm{Spread}_J \subseteq S(\mathbb{R}^J)$ of cardinality

$$|\mathcal{N}| \le 2d\Big(1 + \frac{2}{\varepsilon}\Big)^{d-1} \le 2d\Big(\frac{3K_0}{t}\Big)^{d-1}.$$

Consider the event

$$\mathcal{E} := \Big\{\inf_{z\in\mathcal{N}} \|Wz\|_2 < 2t\sqrt{d}\Big\}.$$

Taking the union bound and using the representation (7.4) in Lemma 7.2, we obtain

$$\mathbb{P}(\mathcal{E}) \le |\mathcal{N}| \cdot \max_{z\in\mathcal{N}} \mathbb{P}\big(\|Wz\|_2 \le 2t\sqrt{d}\big) \le 2d\Big(\frac{3K_0}{t}\Big)^{d-1}(2C_1 t)^{2d-1} \le (C_2 t)^d.$$

Now, suppose the event in (7.6) holds, i.e. there exists $z' \in \mathrm{Spread}_J$ such that

$$\|Wz'\|_2 < t\sqrt{d} \text{ and } \|W\| \le K_0\sqrt{d}.$$

Choose $z \in \mathcal{N}$ such that $\|z - z'\|_2 \le \varepsilon$. Then by the triangle inequality

$$\|Wz\|_2 \le \|Wz'\|_2 + \|W\| \cdot \|z - z'\|_2 < t\sqrt{d} + K_0\sqrt{d} \cdot \varepsilon \le 2t\sqrt{d}.$$

Therefore, $\mathcal{E}$ holds. The bound on the probability of $\mathcal{E}$ completes the proof. □

Lemma 7.4 together with (7.5) yield that

$$\mathbb{P}\Big(\inf_{z\in \mathrm{Spread}_J} \|Wz\|_2 < t\sqrt{d}\Big) \le (C_2 t)^d + e^{-d}.$$

By representation (7.4), this is a weaker version of Theorem 7.2, with e^{-d} instead of e^{-cN}. Unfortunately, this bound is too weak for small d. In particular, for square matrices we have $d = 1$, and the bound is useless.

In the next section, we will refine our current approach using decoupling.

7.2. **Refinement: decoupling.** Our problem is that the probability bound in (7.5) is too weak. We will bypass this by decomposing our event according to all possible values of $\|W\|$, and by decoupling the information about $\|Wz\|_2$ from the information about $\|W\|$.

Proposition 7.5 (Decoupling)**.** *Let W be an $N \times d$ matrix whose columns are independent random vectors. Let $\beta > 0$ and let $z \in S^{d-1}$ be a vector satisfying $|z_k| \geq \frac{\beta}{\sqrt{d}}$ for all $k \in \{1, \dots, d\}$. Then for every $0 < a < b$, we have*

$$\mathbb{P}\big(\|Wz\|_2 < a,\ \|W\| > b\big) \leq 2 \sup_{x \in S^{d-1}, w \in \mathbb{R}^N} \mathbb{P}\Big(\|Wx - w\|_2 < \frac{\sqrt{2}}{\beta} a\Big)\ \mathbb{P}\Big(\|W\| > \frac{b}{\sqrt{2}}\Big).$$

Proof. If $d = 1$ then $\|W\| = \|Wz\|_2$, so the probability in the left hand side is zero. So, let $d \geq 2$. Then we can decompose the index set $\{1, \dots, n\}$ into two disjoint subsets I and H whose cardinalities differ by at most 1, say with $|I| = \lceil d/2 \rceil$.

We write $W = W_I + W_H$ where W_I and W_H are the submatrices of W with columns in I and H respectively. Similarly, for $z \in \mathrm{Spread}_J$, we write $z = z_I + z_H$.

Since $\|W\|^2 \leq \|W_I\|^2 + \|W_H\|^2$, we have

$$\mathbb{P}\big(\|Wz\|_2 < a,\ \|W\| > b\big) = p_I + p_H,$$

where

$$\begin{aligned} p_I &= \mathbb{P}\big(\|Wz\|_2 < a,\ \|W_H\| > b/\sqrt{2}\big) \\ &= \mathbb{P}\big(\|Wz\|_2 < a \ \big|\ \|W_H\| > b/\sqrt{2}\big)\ \mathbb{P}\big(\|W_H\| > b/\sqrt{2}\big), \end{aligned}$$

and similarly for p_H. It suffices to bound p_I; the argument for p_H is similar.

Writing $Wz = W_I z_I + W_H z_H$ and using the independence of the matrices W_I and W_H, we conclude that

$$\begin{aligned} p_I &\leq \sup_{w \in \mathbb{R}^N} \mathbb{P}\big(\|W_I z_I - w\|_2 < a\big)\ \mathbb{P}\big(\|W_H\| > b/\sqrt{2}\big) \\ &\leq \sup_{w \in \mathbb{R}^N} \mathbb{P}\big(\|W z_I - w\|_2 < a\big)\ \mathbb{P}\big(\|W\| > b/\sqrt{2}\big). \end{aligned} \tag{7.7}$$

(In the last line we used $W_I z_I = W z_I$ and $\|W_H\| \leq \|W\|$).

By the assumption on z and since $|I| \geq d/2$, we have

$$\|z_I\|_2 = \Big(\sum_{k \in I} |z_k|^2\Big)^{1/2} \geq \frac{\beta}{\sqrt{2}}.$$

Hence for $x := z_I/\|z_I\|_2$ and $u := w/\|z_I\|_2$, we obtain

$$\mathbb{P}\big(\|Wz_I - w\|_2 < a\big) \le \mathbb{P}\big(\|Wx - u\|_2 < \sqrt{2}a/\beta\big).$$

Together with (7.7), this completes the proof. □

We use this decoupling in the following refinement of Lemma 7.4.

Lemma 7.6. *Let W be a random matrix as in* (7.3)*, where P is the orthogonal projection of $\mathbb{R}^N$ onto the random subspace $(H_{J^c})^\perp$, defined as in Theorem 7.1. Then for every $s \ge 1$ and every t that satisfies* (7.2)*, we have*

$$\begin{aligned}(7.8)\qquad &\mathbb{P}\Big(\inf_{z\in \mathrm{Spread}_J} \|Wz\|_2 < t\sqrt{d} \text{ and } sK_0\sqrt{d} < \|W\| \le 2sK_0\sqrt{d}\Big)\\ &\le (C_3 t e^{-c_3 s^2})^d + e^{-cN}.\end{aligned}$$

Proof. Let $\varepsilon = t/2sK_0$. By Proposition 2.1, there exists an ε-net $\mathcal{N}$ of $\mathrm{Spread}_J \subseteq S(\mathbb{R}^J)$ of cardinality

$$|\mathcal{N}| \le 2d\Big(1 + \frac{2}{\varepsilon}\Big)^{d-1} \le 2d\Big(\frac{6sK_0}{t}\Big)^{d-1}.$$

Consider the event

$$\mathcal{E} := \Big\{ \inf_{z\in\mathcal{N}} \|Wz\|_2 < 2t\sqrt{d} \text{ and } \|W\| > sK_0\sqrt{d}\Big\}.$$

We condition on the realization of the subspace H_{J^c} as above to make the columns of W independent. By the definition (6.1) of Spread_J, any $z \in \mathcal{N}$ satisfies the condition of the Decoupling Proposition 7.5 with $\beta = K_1$. Taking the union bound and then using Proposition 7.5, we obtain

$$\begin{aligned}\mathbb{P}(\mathcal{E} \mid H_{J^c}) &\le |\mathcal{N}| \cdot \max_{z\in\mathcal{N}} \mathbb{P}\big(\|Wz\|_2 \le 2t\sqrt{d} \text{ and } \|W\| > sK_0\sqrt{d} \mid H_{J^c}\big)\\ &\le |\mathcal{N}| \cdot 2 \max_{z\in S(\mathbb{R}^J),\, w\in\mathbb{R}^N} \mathbb{P}\Big(\|Wz - w\|_2 < \frac{\sqrt{2}}{K_1}\cdot 2t\sqrt{d} \mid H_{J^c}\Big)\\ &\quad \cdot \mathbb{P}\Big(\|W\| > \frac{sK_0\sqrt{d}}{\sqrt{2}} \mid H_{J^c}\Big).\end{aligned}$$

Assume now that $\mathrm{LCD}_{\alpha,c}(H_{J^c}^\perp) \ge c\sqrt{N}e^{cN/m}$, where α and c are as in Theorem 4.3. Then using Proposition 7.3 and representation (7.4), we conclude as in the proof of Theorem 4.1 that

$$\mathbb{P}(\mathcal{E} \mid H_{J^c}) \le 4d\Big(\frac{6sK_0}{t}\Big)^{d-1} \cdot (C't)^{2d-1} \cdot e^{-c's^2 d}$$

for any t satisfying (7.2). Since $s \ge 1$ and $d \ge 1$, we can bound this as

$$\mathbb{P}(\mathcal{E} \mid H_{J^c}) \le (C_3 t e^{-c_3 s^2})^d.$$

Therefore, by Theorem 4.3,

$$\begin{aligned}\mathbb{P}(\mathcal{E}) &\le \mathbb{P}(\mathcal{E} \mid \mathrm{LCD}_{\alpha,c}(H_{J^c}^{\perp}) \ge c\sqrt{N}e^{cN/m}) + \mathbb{P}(\mathrm{LCD}_{\alpha,c}(H_{J^c}^{\perp}) < c\sqrt{N}e^{cN/m}) \\ &\le (C_3 t e^{-c_3 s^2})^d + e^{-cN}.\end{aligned}$$

Now, suppose the event in (7.8) holds, i.e. there exists $z' \in \mathrm{Spread}_J$ such that

$$\|Wz'\|_2 < t\sqrt{d} \text{ and } sK_0\sqrt{d} < \|W\| \le 2sK_0\sqrt{d}.$$

Choose $z \in \mathcal{N}$ such that $\|z - z'\|_2 \le \varepsilon$. Then by the triangle inequality

$$\|Wz\|_2 \le \|Wz'\|_2 + \|W\| \cdot \|z - z'\|_2 < t\sqrt{d} + 2sK_0\sqrt{d} \cdot \varepsilon \le 2t\sqrt{d}.$$

Therefore, $\mathcal{E}$ holds. The bound on the probability of $\mathcal{E}$ completes the proof. □

Proof of the Uniform Distance Theorem 7.1. Recall that, without loss of generality, we assumed that (7.2) held. Let k_1 be the smallest natural number such that

$$2^{k_1} \cdot K_0\sqrt{d} > C_0\sqrt{N}, \tag{7.9}$$

where C_0 and K_0 are constants from Lemma 2.3 and Lemma 7.6 respectively. Summing the probability estimates of Proposition 7.4 and Lemma 7.6 for $s = 2^k$, $k = 1, \dots, k_1$, we conclude that

$$\begin{aligned}&\mathbb{P}\Big(\inf_{z \in \mathrm{Spread}_J} \|Wz\|_2 < t\sqrt{d}\Big) \\ &\le (C_2 t)^d + \sum_{s=2^k,\, k=1,\dots,k_1} \Big((C_3 t e^{-c_3 s^2})^d + e^{-cN}\Big) + \mathbb{P}(\|W\| > C_0\sqrt{N}) \\ &\le (C_4 t)^d + k_1 e^{-cN} + \mathbb{P}(\|A\| > C_0\sqrt{N}).\end{aligned}$$

By (7.9) and Proposition 2.3, the last expression does not exceed $(Ct)^d + e^{-cN}$. In view of representation (7.4), this completes the proof. □

8. Completion of the proof

In Section 6, we reduced the invertibility problem for incompressible vectors to computing the distance between a random ellipsoid and a random subspace. This distance was estimated in Section 7. These together lead to the following invertibility bound:

Theorem 8.1 (Invertibility for incompressible vectors)**.** *Let $\delta, \rho \in (0,1)$. There exist $C, c > 0$ which depend only on δ, ρ, and such that the following holds. For every $t > 0$,*

$$\mathbb{P}\Big(\inf_{x \in Incomp(\delta,\rho)} \|Ax\|_2 < t\frac{d}{\sqrt{n}}\Big) \le (Ct)^d + e^{-cN}.$$

Proof. Without loss of generality, we can assume that (7.2) holds. We use Lemma 6.2 with $\varepsilon = t\sqrt{d}$ and then Theorem 7.1 to get the bound $(C't)^d$ on the desired probability. This completes the proof. □

Proof of Theorem 1.1. This follows directly from (5.2), (5.3), and Theorem 8.1. □

References

[1] S. Artstein-Avidan, O. Friedland, V.D. Milman, S. Sodin, *Polynomial bounds for large Bernoulli sections of l_1^N*, Israel J. Math. 156 (2006), 141–155.

[2] Z. D. Bai, Y. Q. Yin, *Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix*, Ann. Probab. 21 (1993), no. 3, 1275–1294.

[3] G. Bennett, L. E. Dor, V. Goodman, W. B. Johnson, C. M. Newman, *On uncomplemented subspaces of L_p, $1 < p < 2$*, Israel J. Math. 26 (1977), 178–187.

[4] K. Davidson, S. J. Szarek, *Local operator theory, random matrices and Banach spaces*, Handbook of the geometry of Banach spaces, Vol. I, 317–366, North-Holland, Amsterdam, 2001.

[5] A. Edelman, *Eigenvalues and condition numbers of random matrices*, SIAM J. Matrix Anal. Appl. 9 (1988), 543–560.

[6] A. Edelman, B. Sutton, *Tails of Condition Number Distributions*, SIMAX (2005), 547–560.

[7] C. G. Esseen, *On the Kolmogorov-Rogozin inequality for the concentration function*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 5 (1966), 210–216.

[8] O. Friedland, S. Sodin, *Bounds on the concentration function in terms of Diophantine approximation*, C. R. Math. Acad. Sci. Paris 345 (2007), 513–518.

[9] O. Friedland, S. Sodin, Private communication.

[10] Y. Gordon, *Some inequalities for Gaussian processes and applications*, Israel J. Math. 50 (1985), no. 4, 265–289.

[11] G. Halász, *Estimates for the concentration function of combinatorial number theory and probability*, Periodica Mathematica Hungarica 8 (1977), 197–211.

[12] A. Kolmogorov, *Sur les propriétés des fonctions de concentrations de M. P. Lévy*, Ann. Inst. H. Poincaré 16 (1958), 27–34.

[13] M. Ledoux, *Deviation inequalities on largest eigenvalues*, Geometric Aspects of Functional Analysis, Israel Seminar 2004-2005. Lecture Notes in Math. 1910, 167–219, Springer, 2007.

[14] P. Lévy, *Théorie de l'addition des variables aléatoires*, Gauthier-Villars, 1937.

[15] W. V. Li, Q.-M. Shao, Gaussian processes: inequalities, small ball probabilities and applications. Stochastic processes: theory and methods, 533–597, Handbook of Statistics, 19, North-Holland, Amsterdam, 2001.

[16] A. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, *Smallest singular value of random matrices and geometry of random polytopes*, Adv. Math. 195 (2005), 491–523.

[17] V. D. Milman, G. Schechtman, Asymptotic theory of finite-dimensional normed spaces. Lecture Notes in Mathematics, 1200. Springer-Verlag, Berlin, 1986.

[18] M. Rudelson, *Invertibility of random matrices: norm of the inverse*, Annals of Mathematics 168 (2008), 575–600.

[19] M. Rudelson, *Lower estimates for the singular values of random matrices*, C. R. Math. Acad. Sci. Paris 342 (2006), no. 4, 247–252.

[20] M. Rudelson, R. Vershynin, *The Littlewood-Offord Problem and invertibility of random matrices*, Advances of Mathematics 218 (2008), 600–633.

[21] J. W. Silverstein, *The smallest eigenvalue of a large dimensional Wishart matrix*, Ann. Probab. 13 (1985), 1364–1368.

[22] S. Szarek, *Condition numbers of random matrices*, J. Complexity 7 (1991), no. 2, 131–149.

[23] T. Tao, V. Vu, *Additive combinatorics.* Cambridge Studies in Advanced Mathematics, 105. Cambridge University Press, Cambridge, 2006.

[24] R. Vershynin, *Some problems in asymptotic convex geometry and random matrices motivated by numerical algorithms.* In: Banach Spaces and their applications in analysis, de Gruyter, 2007, pp. 209–218

Department of Mathematics, University of Missouri, Columbia, MO 65211, USA

E-mail address: `rudelson@math.missouri.edu`

Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA

E-mail address: `romanv@umich.edu`