LECTURE NOTES ON NON-ASYMPTOTIC THEORY OF RANDOM MATRICES

MARK RUDELSON

AMS SHORT COURSE ON RANDOM MATRICES

San Diego, January 2013

Partially supported by NSF DMS grant DMS 1161372.

1. Introduction

The classical random matrix theory is concerned with asymptotic of various spectral characteristics of families of random matrices, when the dimensions of the matrices tend to infinity. There are many examples when these characteristics, which are random variables themselves, converge to certain limit laws. This includes the celebrated Wigner semicircle law for the empirical measures of eigenvalues of random symmetric matrices, Marchenko–Pastur law, which is the limit of empirical measures of sample covariance matrices, Tracy–Widom distribution describing the limit of the first singular values of a sequence of random matrices, etc. [1]. These limits are of paramount importance, yet in applications one usually needs information about the behavior of such characteristics for large, but fixed n. For instance in problems in convex geometry one constructs a random section of an N-dimensional convex body by taking the kernel or the range of a certain random matrix. Random matrices arise also in analysis of rates of convergence of computer science algorithms. In both cases, the dimension of the ambient space remains fixed, and one seeks explicit estimates of probabilities in terms of the dimension. For such problems knowing the limit behavior is of little help.

The problems involving estimates for a fixed finite dimension arise in the classical random matrix theory as well. One of the main approaches in deriving the limit laws is based on analysis of the Stieltjes transform of measures [1]. To derive the convergence of Stieltjes transforms, one frequently has to provide explicit bounds on the smallest singular value of a random matrix of a fixed size, which holds with high probability. This need arises, e.g., in derivation of the circular law and the single ring theorem.

These questions led to development of non-asymptotic theory of random matrices, which provides probabilistic bounds for eigenvalues, singular values, etc. for random matrices of a large fixed size. The situation is roughly parallel to that arising for the sums of i.i.d. random variables, where the asymptotic and non-asymptotic results go hand in hand. The asymptotic behavior of the averages of n i.i.d. random variables is governed by the Strong Law of Large Numbers establishing the almost sure convergence to the expectation. Yet, to assert that the average of a large number of random variables is close to the expectation, we need a non-asymptotic version, e.g. Hoeffding inequality. This inequality yields a subgaussian bound for the large deviations (see the details below). Such behavior suggests that the limit distribution of the deviation should be normal, which leads to an asymptotic result, the Central Limit Theorem (CLT). To use the CLT in evaluation of probabilities for random sums, we need its non-asymptotic version, namely the Berry–Esseen Theorem. This theorem provides in turn a crucial step in deriving another fundamental asymptotic result, the Law of Iterated Logarithm.

These notes discuss the methods of the non-asymptotic approach to the random matrix theory. We do not attempt to provide an exhaustive list of references (a reader can check the surveys [5], [27], and [40]). Instead we concentrate on three essentially different examples, with the aim of presenting the methods and results in a maximally self-contained form. This approach inevitably leaves out several important recent developments, such as invertibility of random symmetric matrices [42, 19], applications to the Circular Law [9, 37, 38], and concentration for random determinants [39, 20]. Yet, by restricting ourselves to a few results, we will be able to give a relatively complete picture of the ideas and methods involved in their proofs. We start with introduction to subgaussian random variables in Section 3. In Sections 5-7 we obtain quantitative bounds for invertibility of random matrices with i.i.d. entries. As will be shown in Section 6, the arithmetic structures play a crucial role here. Section 8 studies a question arising in geometric functional analysis. Here the ambient space is Banach, and the approach combines the methods of the previous sections with the functional-analytic considerations. We will also touch upon majorising measures, which are a powerful tool for estimating suprema of random processes. Section 9 contains another quantitative invertibility result. Here we discuss a random unitary or orthogonal perturbation of a fixed matrix. Unlike in the first example, the arithmetic structure plays no role in this problem. The main difficulty is the dependence between the entries of a random matrix, and the method is based on the introduction of perturbations with independent entries.

Acknowledgement. These notes are based in part on the material presented at the workshop "Etats de la Recherche: Probability and geometry in interaction" at Paul Sabatier University in Toulouse, the mini-course given at the Warsaw University, and the Informal Analysis Seminar at Kent State University. The author is grateful to Franck Barthe, Michel Ledoux, Rafal Latala, Krzystof Oleszkiewicz, Michal Wojchehowski, Fedor Nazarov, Dmitry Ryabogin, and Artem Zvavitch for their hospitality. The author also grateful to Fedor Nazarov for careful reading of the manuscript and many suggestions, which led to improvement of the presentation.

2. NOTATION AND BASIC DEFINITIONS

We shall consider random matrices of high order with independent entries. For simplicity, we shall assume that the entries are centered ($\mathbb{E}a_{j,k} = 0$) and identically distributed (both conditions may be relaxed).

Throughout these notes $\|\cdot\|_p$ denotes the ℓ_p norm

$$\|x\|_p = \left(\sum_{j=1}^n |x_j|^p \right)^{1/p}, \qquad 1 \le p < \infty,$$

and B_p^n stands for the unit ball of this norm. The norm of an operator or a matrix will be denoted by $\|\cdot\|$. We use S^{n-1} for the unit Euclidean sphere. If F is a finite set, then $|F|$ denotes the cardinality of F. Letters C, C', c etc. denote absolute constants.

If $N \ge n$ then an $N \times n$ matrix A can be viewed as a mapping of $\mathbb{R}^n$ into $\mathbb{R}^N$. Thus, a random matrix defines a random n-dimensional section of $\mathbb{R}^N$. For geometric applications we need to know that this matrix would not distort the metric too much. Let us formulate it more precisely:

Definition 2.1. Let $N \ge n$ and let A be an $N \times n$ matrix. The condition number of the matrix A is

$$\sigma(A) = \frac{\max_{x \in S^{n-1}} \|Ax\|_2}{\min_{x \in S^{n-1}} \|Ax\|_2}.$$

If $\min_{x \in S^{n-1}} \|Ax\|_2 = 0$, we set $\sigma(A) = \infty$.

The condition number of a matrix can be rewritten in terms of its singular values.

Definition 2.2. Let $N \ge n$ and let A be an $N \times n$ matrix. The singular values of A are the eigenvalues of $(A^*A)^{1/2}$, arranged in the decreasing order: $s_1(A) \ge s_2(A) \ge \ldots \ge s_n(A)$.

The singular values of A are the lengths of the semi-axes of the ellipsoid AB_2^n. The first and the last singular values have a clear functional-analytic meaning:

$$s_1(A) = \left\| A : \mathbb{R}^n \to \mathbb{R}^N \right\|,$$

and

$$s_n(A) = \min_{x \in S^{n-1}} \|Ax\| = 1 / \left\| A^{-1} : A\mathbb{R}^n \to \mathbb{R}^n \right\|,$$

whenever A has the full rank. In this notation $\sigma(A) = s_1(A)/s_n(A)$.

Therefore, to bound the condition number, we have to estimate the first singular value from above, and the last one from below. For matrices with i.i.d. random entries the first singular value is the most robust. It can be estimated using a simple ε-net argument, as will be shown in Proposition 4.4. The last singular value presents a bigger challenge. We will obtain its bounds for "tall" rectangular matrices in Section 4, and for square matrices in Sections 5-7.

3. Subgaussian random variables

In this section we introduce an important class of random variables with strong tail decay properties. This class contains the normal variables, as well as all bounded random variables.

Definition 3.1. Let ξ be a random variable and let $v > 0$. We shall call ξ v-subgaussian if there exist constants C and v such that for any $t > 0$

$$\mathbb{P}(|\xi| > t) \leq Ce^{-vt^2}.$$

A random variable ξ is called centered if $\mathbb{E}\xi = 0$.

If the parameter v is an absolute constant, we call a v-subgaussian random variable subgaussian. We shall assume that the random variable ξ is non-degenerate, i.e. $\mathrm{Var}(\xi) > 0$.

The subgaussian condition can be formulated in a number of different ways.

Theorem 3.2. *Let X be a random variable. The following conditions are equivalent:*

(1) *X is subgaussian;*
(2) $\exists a > 0\ \mathbb{E}e^{aX^2} < +\infty$ (ψ_2-condition)*;*
(3) $\exists B, b > 0\ \forall \lambda \in \mathbb{R}\ \ \mathbb{E}e^{\lambda X} \leq Be^{\lambda^2 b}$ (Laplace transform condition)*;*
(4) $\exists K > 0\ \forall p \geq 1\ \ (\mathbb{E}|X|^p)^{1/p} \leq K\sqrt{p}$ (moment condition)*.*

Moreover, if X is a centered random variable, (3) can be rewritten as

(3)′ $\exists b' > 0\ \forall \lambda \in \mathbb{R}\ \mathbb{E}e^{\lambda X} \leq e^{\lambda^2 b'}$.

Proof. The proof is a series of elementary calculations.
(1) ⇒ (2) Let $a < v$. By the integral distribution formula,

$$\mathbb{E}e^{aX^2} = 1 + \int_0^\infty 2ate^{at^2} \cdot \mathbb{P}(|X| > t)\, dt \leq 1 + \int_0^\infty 2at \cdot Ce^{-(v-a)t^2}\, dt < +\infty.$$

(2) ⇒ (3) Let λ be any real number. Then

$$\mathbb{E}e^{\lambda X} = \mathbb{E}e^{\lambda X - aX^2} e^{aX^2} \leq \sup_{t\in\mathbb{R}} e^{\lambda t - at^2} \cdot \mathbb{E}e^{aX^2} \leq Be^{\lambda^2/4a}.$$

(3) ⇒ (4) Set $\lambda = \sqrt{p}$. Replacing, as before, the the function by its supremum, we get

$$\mathbb{E}|X|^p \leq \sup_{t>0} t^p e^{-\sqrt{p}\,t} \cdot \mathbb{E}e^{\sqrt{p}|X|} \leq \left(\frac{\sqrt{p}}{e}\right)^p \cdot Ce^{pb}.$$

(4) ⇒ (1) Assume first $t \geq eK$. Choose p so that $\frac{K\sqrt{p}}{t} = e^{-1}$.

$$\mathbb{P}(|X| > t) \leq \frac{\mathbb{E}|X|^p}{t^p} \leq \left(\frac{K\sqrt{p}}{t}\right)^p = e^{-p} = e^{-vt^2},$$

where $v = e^{-2}K^{-2}$. This proves (1) for $t \geq eK$. Setting $C = e$ automatically guaranties that (1) holds for $0 < t < eK$ as well.

$(3)'$ We will assume that (3) holds with $B > 1$ since otherwise the statement is trivial. Assume first that X is symmetric. For large values of λ, we can derive (3) with constant $B = 1$ by changing the parameter b. Indeed, set $\lambda_0 = \sqrt{2a}$ and choose $\bar{b} > 0$ so that $Be^{\lambda_0^2 b} \le e^{\lambda_0^2 \bar{b}}$. This guarantees that (3) holds for all λ such that $|\lambda| \ge \lambda_0$ with $B = 1$ and b replaced by $\bar{b}$.

If $\lambda^2 \le 2a$, then by the ψ_2-condition and Holder's inequality,

$$\mathbb{E}e^{\lambda X} = \mathbb{E}\frac{1}{2}(e^{\lambda X} + e^{-\lambda X}) \le \mathbb{E}e^{\lambda^2 X^2/2} \le \left(\mathbb{E}e^{aX^2}\right)^{\lambda^2/2a} \le \exp\left(c\frac{\lambda^2}{2a}\right).$$

Finally, we set $b' = \max(c/2a, \bar{b})$.

If X is merely centered, we use a simple symmetrization. Let X' be an independent copy of X. Then by Jensen's inequality,

$$\mathbb{E}e^{\lambda X} = \mathbb{E}e^{\lambda(X - \mathbb{E}X')} \le \mathbb{E}e^{\lambda(X - X')},$$

where $X - X'$ is a symmetric subgaussian random variable. □

Remark. The ψ_2-condition turns the set of centered subgaussian random variables into a normed space. Define the function $\psi_2 : \mathbb{R} \to \mathbb{R}$ by $\psi_2(t) = \exp(t^2) - 1$. Then for a non-zero random variable set

$$\|X\|_{\psi_2} = \inf\{s > 0 \mid \mathbb{E}\psi_2(X/s) \le 1\}.$$

The subgaussian random variables equipped with this norm form an Orlicz space (see [17] for the details).

To estimate the first singular value, we have to prove a large deviation inequality for a linear combination of independent subgaussian random variables. Note that a linear combination of independent Gaussian random variables is Gaussian. We prove below that a linear combination of independent subgaussian random variables is subgaussian.

Theorem 3.3. *Let $X_1, \dots, X_n$ be independent centered subgaussian random variables. Then for any $a_1, \dots, a_n \in \mathbb{R}$*

$$\mathbb{P}\left(\left|\sum_{j=1}^n a_j X_j\right| > t\right) \le 2\exp\left(-\frac{ct^2}{\sum_{j=1}^n a_j^2}\right).$$

Proof. Set $v_j = a_j / \left(\sum_{j=1}^n a_j^2\right)^{1/2}$. We have to show that the random variable $Y = \sum_{j=1}^n v_j X_j$ is subgaussian. Let us check the Laplace transform condition

(3)′. For any $\lambda \in \mathbb{R}$

$$\mathbb{E}\exp\left(\lambda \sum_{j=1}^{n} v_j X_j\right) = \prod_{j=1}^{n} \mathbb{E}\exp(\lambda v_j X_j)$$
$$\leq \prod_{j=1}^{n} \exp(\lambda^2 v_j^2 b) = \exp\left(\lambda^2 b \sum_{j=1}^{n} v_j^2\right) = e^{\lambda^2 b}.$$

The inequality here follows from (3)′. Note that the fact that the constant in front of the exponent in (3)′ is 1 plays the crucial role here. □

Theorem 3.3 can be used to give a very short proof of a classical inequality due to Khinchin.

Theorem 3.4 (Khinchin)**.** *Let $X_1, \dots, X_n$ be independent centered subgaussian random variables. For any $p \geq 1$ there exist $A_p, B_p > 0$ such that the inequality*

$$A_p \left(\sum_{j=1}^{n} a_j^2\right)^{1/2} \leq \left(\mathbb{E}\left|\sum_{j=1}^{n} a_j X_j\right|^p\right)^{1/p} \leq B_p \left(\sum_{j=1}^{n} a_j^2\right)^{1/2}$$

holds for all $a_1, \dots, a_n \in \mathbb{R}$.

Proof. Without loss of generality, assume that $\left(\sum_{j=1}^{n} a_j^2\right)^{1/2} = 1$.

Let $p \geq 2$. Then by Hölder's inequality

$$\left(\sum_{j=1}^{n} a_j^2\right)^{1/2} = \left(\mathbb{E}\left|\sum_{j=1}^{n} a_j X_j\right|^2\right)^{1/2} \leq \left(\left|\mathbb{E}\sum_{j=1}^{n} a_j X_j\right|^p\right)^{1/p},$$

so $A_p = 1$. By Theorem 3.3, $Y = \sum_{j=1}^{n} a_j X_j$ is a subgaussian random variable. Hence,

$$(\mathbb{E}|Y|^p)^{1/p} \leq C\sqrt{p} =: B_p.$$

This is the right asymptotic as $p \to \infty$.

In the case $1 \leq p \leq 2$ it is enough to prove the inequality for $p = 1$. As before, by Hölder's inequality, we can choose $B_p = 1$. Applying Khinchin's inequality with $p = 3$, we get

$$\mathbb{E}|Y|^2 = \mathbb{E}|Y|^{1/2} \cdot |Y|^{3/2} \leq (\mathbb{E}|Y|)^{1/2} \cdot \left(\mathbb{E}|Y|^3\right)^{1/2} \leq (\mathbb{E}|Y|)^{1/2} \cdot B_3^{3/2} \left(\mathbb{E}|Y|^2\right)^{3/4}.$$

Hence,

$$B_3^{-3} \left(\mathbb{E}|Y|^2\right)^{1/2} \leq \mathbb{E}|Y|.$$ □

4. Invertibility of a rectangular random matrix

We introduce the *ε-net argument*, which will enable us to bound the condition number for a random $N \times n$ matrix with independent entries in the case when $N \gg n$. To simplify the proofs we assume from now on that the entries of the matrix are centered, subgaussian random variables.

Recall the definition of an ε-net.

Definition 4.1. Let (T, d) be a metric space. Let $K \subset T$. A set $\mathcal{N} \subset T$ is called an ε-net for K if

$$\forall x \in K \exists y \in \mathcal{N} \ d(x, y) < \varepsilon.$$

A set $\mathcal{S} \subset K$ is called ε-separated if

$$\forall x, y \in \mathcal{S} \quad d(x, y) \geq \varepsilon.$$

These two notions are closely related. Namely, we have the following elementary Lemma.

Lemma 4.2. *Let K be a subset of a metric space (T, d), and let $\mathcal{N} \subset T$ be an ε-net for K. Then*

1. *there exists a 2ε-net $\mathcal{N}' \subset K$ such that $|\mathcal{N}'| \leq |\mathcal{N}|$;*
2. *any 2ε-separated set $\mathcal{S} \subset K$ satisfies $|\mathcal{S}| \leq |\mathcal{N}|$.*
3. *From the other side, any maximal ε-separated set $\mathcal{S}' \subset K$ is an ε-net for K.*

We leave the proof of this lemma for a reader as an exercise.

Lemma 4.3 (Volumetric estimate)**.** *For any $\varepsilon < 1$ there exists an ε-net $\mathcal{N} \subset S^{n-1}$ such that*

$$|\mathcal{N}| \leq \left(\frac{3}{\varepsilon}\right)^n.$$

Proof. Let $\mathcal{N}$ be a maximal ε-separated subset of S^{n-1}. Then for any distinct points $x, y \in \mathcal{N}$

$$\left(x + \frac{\varepsilon}{2} B_2^n\right) \cap \left(y + \frac{\varepsilon}{2} B_2^n\right) = \emptyset.$$

Hence,

$$|\mathcal{N}| \cdot \operatorname{vol}\left(\frac{\varepsilon}{2} B_2^n\right) = \operatorname{vol}\left(\bigcup_{x \in \mathcal{N}} \left(x + \frac{\varepsilon}{2} B_2^n\right)\right) \leq \operatorname{vol}\left(\left(1 + \frac{\varepsilon}{2}\right) B_2^n\right),$$

which implies

$$|\mathcal{N}| \leq \left(1 + \frac{2}{\varepsilon}\right)^n \leq \left(\frac{3}{\varepsilon}\right)^n. \qquad \square$$

Using ε-nets, we prove a basic bound on the first singular value of a random subgaussian matrix:

Proposition 4.4 (First singular value)**.** *Let A be an $N \times n$ random matrix, $N \geq n$, whose entries are independent copies of a subgaussian random variable. Then*

$$\mathbb{P}\big(s_1(A) > t\sqrt{N}\big) \leq e^{-c_0 t^2 N} \qquad \text{for } t \geq C_0.$$

Proof. Let $\mathcal{N}$ be a (1/2)-net in S^{N-1} and $\mathcal{M}$ be a (1/2)-net in S^{n-1}. For any $u \in S^{n-1}$, we can choose a $x \in \mathcal{N}$ such that $\|x-u\|_2 < 1/2$. Then

$$\|Au\|_2 \leq \|Ax\|_2 + \|A\| \cdot \|x-u\|_2 \leq \|Ax\|_2 + \frac{1}{2}\|A\|.$$

This shows that $\|A\| \leq 2\sup_{x\in\mathcal{N}} \|Ax\|_2 = 2\sup_{x\in\mathcal{N}} \sup_{v\in S^{N-1}} \langle Ax, u\rangle$. Approximating v in a similar way by an element of $\mathcal{M}$, we obtain

$$\|A\| \leq 4 \max_{x\in\mathcal{N},\ y\in\mathcal{M}} |\langle x, y\rangle|.$$

By Lemma 4.3, we can choose these nets so that

$$|\mathcal{N}| \leq 6^N, \quad |\mathcal{M}| \leq 6^n.$$

By Theorem 3.3, for every $x \in \mathcal{N}$ and $y \in \mathcal{M}$, the random variable $\langle Ax, y\rangle = \sum_{j=1}^N \sum_{k=1}^n a_{j,k} y_j x_k$ is subgaussian, i.e.,

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

Taking the union bound, we get

$$\begin{aligned}\mathbb{P}\big(\|A\| > t\sqrt{N}\big) &\leq |\mathcal{N}||\mathcal{M}| \max_{x\in N,\, y\in M} \mathbb{P}\big(|\langle Ax, y\rangle| > t\sqrt{N}/4\big) \\ &\leq 6^N \cdot 6^N \cdot C_1 e^{-c_2 t^2 N} \leq C_1 e^{-c_0 t^2 N},\end{aligned}$$

provided that $t \geq C_0$ for an appropriately chosen constant $C_0 > 0$. This completes the proof. □

Proposition 4.4 means that for any $N \geq n$ the first singular value is $O(\sqrt{N})$ with probability close to 1. Thus, the bound for the condition number reduces to a lower estimate of the last singular value.

To obtain it, we prove an easy estimate for a small ball probability of a sum of independent random variables.

Lemma 4.5. *Let $\xi_1, \dots, \xi_n$ be independent copies of a centered subgaussian random variable with variance 1. Then there exists $\mu \in (0,1)$ such that for every coefficient vector $a = (a_1, \dots, a_n) \in S^{n-1}$ the random sum $S = \sum_{k=1}^n a_k \xi_k$ satisfies*

$$\mathbb{P}(|S| < 1/2) \leq \mu.$$

Proof. Let $0 < \lambda < (\mathbb{E}S^2)^{1/2} = 1$. By the Cauchy–Schwarz inequality,

$$\mathbb{E}S^2 = \mathbb{E}S^2 \mathbf{1}_{[\lambda,\lambda]}(S) + \mathbb{E}S^2 \mathbf{1}_{\mathbb{R}\setminus[\lambda,\lambda]}(S) \leq \lambda^2 + \big(\mathbb{E}S^4\big)^{1/2} \mathbb{P}(|S| > \lambda)^{1/2}.$$

This leads to the Paley–Zygmund inequality:

$$\mathbb{P}(|S| > \lambda) \geq \frac{(\mathbb{E}S^2 - \lambda^2)^2}{\mathbb{E}S^4} = \frac{(1-\lambda^2)^2}{\mathbb{E}S^4}.$$

By Theorem 3.3, the random variable S is subgaussian, so by (4), Theorem3.2, $\mathbb{E}S^4 \leq C$. To finish the proof, set $\lambda = 1/2$. □

Lemma 4.5 implies the following invertibility estimate for a fixed vector.

Corollary 4.6. *Let A be a matrix as in Proposition 4.4. Then there exist constants $\eta, \nu \in (0,1)$ such that for every $x \in S^{n-1}$,*

$$\mathbb{P}(\|Ax\|_2 < \eta N^{1/2}) \leq \nu^N.$$

Proof. The coordinates of the vector Ax are independent linear combinations of i.i.d. subgaussian random variables with coefficients $(x_1, \dots, x_n) \in S^{n-1}$. Hence, by Lemma 4.5, $\mathbb{P}(|(Ax)_j| < 1/2) \leq \mu$ for all $j = 1, \dots, N$.

Assume that $\|Ax\|_2 < \eta\sqrt{N}$. Then $|(Ax)_j| < 1/2$ for at least $(1-4\eta^2)N > N/2$ coordinates. If η is small enough, then the number M of subsets of $\{1, \dots, N\}$ with at least $(1-4\eta^2)N$ elements is less than $\mu^{-N/4}$. Then the union bound implies

$$\mathbb{P}(\|Ax\|_2 < \eta N^{1/2}) \leq M \cdot \mu^{N/2} \leq \mu^{N/4}.$$ □

Combining this with the ε-net argument, we obtain the estimate for the smallest singular value of a random matrix, whose dimensions are significantly different.

Proposition 4.7 (Smallest singular value of rectangular matrices)**.** *Let A be an $N \times n$ matrix whose entries are i.i.d. centered subgaussian random variables with variance 1. There exist $c_1, c_2 > 0$ and $\delta_0 \in (0,1)$ such that if $n < \delta_0 N$, then*

$$\mathbb{P}\Big(\min_{x \in S^{n-1}} \|Ax\|_2 \leq c_1 N^{1/2}\Big) \leq e^{-c_2 N}. \tag{4.1}$$

Proof. Let $\varepsilon > 0$ to be chosen later. Let $\mathcal{N}$ be an ε-net in S^{n-1} of cardinality $|\mathcal{N}| \leq (3/\varepsilon)^n$. Let η and ν be the numbers in Corollary 4.6. Then by the union bound,

$$\mathbb{P}\left(\exists y \in \mathcal{N} : \ \|Ay\|_2 < \eta N^{1/2}\right) \leq (3/\varepsilon)^n \cdot \nu^N. \tag{4.2}$$

Let V be the event that $\|A\| \leq C_0 N^{1/2}$ and $\|Ay\|_2 \geq \eta N^{1/2}$ for all points $y \in \mathcal{N}$.

Assume that V occurs, and let $x \in S^{n-1}$ be any point. Choose $y \in \mathcal{N}$ such that $\|y - x\|_2 < \varepsilon$. Then

$$\|Ax\|_2 \geq \|Ay\|_2 - \|A\| \cdot \|x - y\|_2 \geq \eta N^{1/2} - C_0 N^{1/2} \cdot \varepsilon = \frac{\eta N^{1/2}}{2},$$

if we set $\varepsilon = \eta/(2C_0)$. By (4.2) and Proposition 4.4,

$$\mathbb{P}(V^c) \le \big(\nu \cdot (3/\varepsilon)^{n/N}\big)^N + e^{-c'N} \le e^{-c_2 N},$$

if we assume that $n/N \le \delta_0$ for an appropriately chosen $\delta_0 < 1$. This completes the proof. □

5. Invertibility of a square matrix: absolutely continuous entries

Until recently, much less has been known about the behavior of the smallest singular value of a square matrix. In the classic work on numerical inversion of large matrices, von Neumann and his associates used random matrices to test their algorithms, and they speculated that

$$s_n(A) \sim n^{-1/2} \quad \text{with high probability} \tag{5.1}$$

(see [43], pp. 14, 477, 555). In a more precise form, this estimate was conjectured by Smale [29] and proved by Edelman [6] and Szarek [31] for *random Gaussian matrices* A, i.e., those with i.i.d. standard normal entries. Edelman's theorem states that for every $\varepsilon \in (0, 1)$,

$$\mathbb{P}\big(s_n(A) \le \varepsilon n^{-1/2}\big) \sim \varepsilon. \tag{5.2}$$

Conjecture (5.1) for general random matrices was an open problem, unknown even for the *random sign matrices* A, i.e., those whose entries are ± 1 symmetric random variables. The first polynomial bound for the smallest singular value of a random matrix with i.i.d. subgaussian, in particular, ± 1 entries was obtained in [23]. It was proved that for such matrix $s_n(A) \ge Cn^{-3/2}$ with high probability. Following that, Tao and Vu proved that if A is a ± 1 random matrix, then for any $\alpha > 0$ there exists $\beta > 0$ such that $s_n(A) \ge n^{-\beta}$ with probability at least $1 - n^{-\alpha}$. In [24] the conjecture (5.1) is proved in full generality under the fourth moment assumption.

Theorem 5.1 (Invertibility: fourth moment)**.** *Let A be an $n \times n$ matrix whose entries are independent centered real random variables with variances at least 1 and fourth moments bounded by B. Then, for every $\delta > 0$ there exist $\varepsilon > 0$ and n_0 which depend (polynomially) only on δ and B, such that*

$$\mathbb{P}\big(s_n(A) \le \varepsilon n^{-1/2}\big) \le \delta \qquad \text{for all } n \ge n_0.$$

This shows in particular that the median of $s_n(A)$ is at least of order $n^{-1/2}$. To show that $s_n(A) \sim n^{-1/2}$ with high probability, one has to prove a matching lower bound. This was done in [26] for matrices with subgaussian entries and extended in [41] to matrices, whose entries have the finite fourth moment.

Under stronger moment assumptions, more is known about the distribution of the largest singular value, and similarly one hopes to know more about the smallest singular value.

One might then expect that the estimate (5.2) for the distribution of the smallest singular value of Gaussian matrices should hold for all subgaussian matrices. Note however that (5.2) fails for the random sign matrices, since they are singular with positive probability. Estimating the singularity probability for random sign matrices is a longstanding open problem. Even proving that it converges to 0 as $n \to \infty$ is a nontrivial result due to Komlós [16]. Later Kahn, Komlós and Szemerédi [15] showed that it is exponentially small:

$$\mathbb{P}\big(\text{random sign matrix } A \text{ is singular}\big) < c^n \tag{5.3}$$

for some universal constant $c \in (0,1)$. The often conjectured optimal value of c is $1/2 + o(1)$ [15], and the best known value $1/\sqrt{2} + o(1)$ is due to Bourgain, Vu, and Wood [4], (see [33, 35] for earlier results).

Spielman and Teng [30] conjectured that (5.2) should hold for the random sign matrices up to an exponentially small term that accounts for their singularity probability:

$$\mathbb{P}\big(s_n(A) \le \varepsilon n^{-1/2}\big) \le \varepsilon + c^n.$$

We prove Spielman-Teng's conjecture up to a coefficient in front of ε. Moreover, we show that this type of behavior is common for all matrices with subgaussian i.i.d. entries. For a bound for random matrices with general i.i.d. entries see [24].

Theorem 5.2 (Invertibility: subgaussian)**.** *Let A be an $n \times n$ matrix whose entries are independent copies of a centered subgaussian real random variable. Then for every $\varepsilon \ge 0$, one has*

$$\mathbb{P}\big(s_n(A) \le \varepsilon n^{-1/2}\big) \le C\varepsilon + c^n, \tag{5.4}$$

where $C > 0$ and $c \in (0,1)$.

Note that setting $\varepsilon = 0$ we recover the result of Kahn, Komlós and Szemerédi. Also, note that the question whether (5.4) holds for random sign matrices with coefficient $C = 1$ remains open.

We shall start with an attempt to apply the ε-net argument. Let us consider an $n \times n$ Gaussian matrix, i.e., a matrix with independent $N(0,1)$ entries. In this case, for any $x \in S^{n-1}$, the vector Ax has independent $N(0,1)$ coordinates, so it is distributed like the standard Gaussian vector in $\mathbb{R}^n$. Hence, for any $t > 0$,

$$\begin{aligned}\mathbb{P}(\|Ax\|_2 \le t\sqrt{n}) &= (2\pi)^{-n/2} \int_{t\sqrt{n}\cdot B_2^n} e^{-\|x\|_2^2/2}\, dx \le (2\pi)^{-n/2} \mathrm{vol}(t\sqrt{n}\cdot B_2^n) \\ &\le (C_1 t)^n.\end{aligned}$$

Fix $\varepsilon > 0$. Let $\mathcal{N}$ be an ε-net in S^{n-1} of cardinality $|\mathcal{N}| \le (3/\varepsilon)^n$. Then by the union bound,

$$\mathbb{P}\left(\exists x \in \mathcal{N} : \ \|Ax\|_2 < tn^{1/2}\right) \le (3/\varepsilon)^n \cdot (C_1 t)^n.$$

To obtain a meaningful estimate we have to require

$$(3/\varepsilon) \cdot (C_1 t) < 1. \tag{5.5}$$

As in Proposition 4.7, we may assume that $\|A\| \le C_0\sqrt{n}$, since the complement of this event has an exponentially small probability. Assume that for any $y \in \mathcal{N}$, $\|Ay\|_2 \ge t\sqrt{n}$. Given $x \in S^{n-1}$, find $y \in \mathcal{N}$ satisfying $\|x - y\|_2 < \varepsilon$. Then

$$\|Ax\|_2 \ge \|Ay\|_2 - \|A\| \cdot \|x - y\|_2 \ge tn^{1/2} - C_0 n^{1/2} \cdot \varepsilon.$$

To obtain a non-trivial lower bound, we have to assume that

$$t > C_0\varepsilon. \tag{5.6}$$

Unfortunately, the system of inequalities (5.5) and (5.6) turns out to be inconsistent, and the ε-net argument fails for the square matrix. Nevertheless, a part of this idea can be salvaged. Namely, if the cardinality of the ε-net satisfies a better estimate

$$|\mathcal{N}| \le (\alpha/\varepsilon)^n \tag{5.7}$$

for a small constant $\alpha > 0$, then (5.5) is replaced by $(\alpha/\varepsilon) \cdot (C_1 t) < 1$, and the system (5.5), (5.6) becomes consistent. While the estimate (5.7) is impossible for the whole sphere, it can be obtained for a small part of it. This becomes the first ingredient of our strategy: small parts of the sphere will be handled by the ε-net argument. However, the "bulk" of the sphere has to be handled differently.

The proof of Theorem 5.2 for random matrices with i.i.d. subgaussian entries having a bounded density is presented below.

5.1. **Conditional argument.** To handle the "bulk", we have to produce an estimate which holds for all vectors in it simultaneously, without taking the union bound. Let $x \in S^{n-1}$ be a vector such that $|x_1| \ge n^{-1/2}$. Denote the columns of the matrix A by $X_1, \dots, X_n$, and let

$$H_j := \operatorname{span}(X_k \mid k \ne j).$$

Then $Ax = \sum_{k=1}^n x_k X_k$, so

$$\|Ax\|_2 \ge \operatorname{dist}(Ax, H_1) = \operatorname{dist}(x_1 X_1, H_1) \ge n^{-1/2} \operatorname{dist}(X_1, H_1). \tag{5.8}$$

Note that the right hand side is independent of x. Therefore it provides a uniform lower bound for all x such that $|x_1| \ge n^{-1/2}$. Since any vector $x \in S^{n-1}$ has a coordinate with absolute value greater than $n^{-1/2}$, we can try to extend this bound to the whole sphere. This approach immediately runs into

a problem: we don't know *a priori* which of the coordinates of x is big. To modify this approach we shall pick a *random* coordinate. To this end we have to know that the random coordinate is big with relatively high probability. This is true for vectors, which look like the vertices of a discrete cube, but is obviously false for vectors with small support, i.e. a small number of non-zero coordinates. This observation leads us to the first decomposition of the sphere:

Definition 5.3 (Compressible and incompressible vectors)**.** Fix $\delta, \rho \in (0,1)$. A vector $x \in \mathbb{R}^n$ is called *sparse* if $|\mathrm{supp}(x)| \le \delta n$. 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 sparse, compressible and incompressible vectors will be denoted by *Sparse*, *Comp* and *Incomp* respectively.

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:

$$\begin{aligned}(5.9)\quad \mathbb{P}\big(s_n(A) \le \varepsilon n^{-1/2}\big) &\le \mathbb{P}\big(\inf_{x \in \mathit{Comp}} \|Ax\|_2 \le \varepsilon n^{-1/2}\big) \\ &\quad + \mathbb{P}\big(\inf_{x \in \mathit{Incomp}} \|Ax\|_2 \le \varepsilon n^{-1/2}\big).\end{aligned}$$

On the set of compressible vectors, we obtain an inequality, which is much stronger than we need.

Lemma 5.4 (Invertibility for compressible vectors)**.** *Let A be a random matrix as in Theorem 5.2, Then there exist $\delta, \rho, c_1, c_2 > 0$ such that*

$$\mathbb{P}\big(\inf_{x \in \mathit{Comp}} \|Ax\|_2 \le c_1 n^{1/2}\big) \le e^{-c_2 n}.$$

Sketch of the proof. Any compressible vectors is close to a coordinate subspace of a small dimension δn. The restriction of our random matrix A onto such a subspace is a random *rectangular* $n \times \delta n$ matrix. Such matrices are well invertible outside of an event of exponentially small probability, provided that δ is small enough (see Proposition 4.7). By taking the union bound over all coordinate subspaces, we deduce the invertibility of the random matrix on the set of compressible vectors. □

We shall fix δ and ρ as in Lemma 5.4 for the rest of the proof.

The incompressible vectors are well spread in the sense that they have many coordinates of the order $n^{-1/2}$. This observation will allow us to realize the scheme described at the beginning of this section.

Lemma 5.5 (Incompressible vectors are spread)**.** *Let $x \in Incomp$. Then there exists a set $\sigma(x) \subseteq \{1, \dots, n\}$ of cardinality $|\sigma(x)| \geq \nu_1 n$ and such that*

$$\frac{\nu_2}{\sqrt{n}} \leq |x_k| \leq \frac{\nu_3}{\sqrt{n}} \qquad \text{for all } k \in \sigma.$$

Here $\nu_1, \nu_2 < 1$ and $\nu_3 > 1$ are constants depending only on the parameters δ, ρ.

We leave the proof of this lemma to the reader.

The main difficulty in implementing the distance bound like (5.8) is to avoid taking the union bound. We achieve this in the proof of the next lemma by a random choice of a coordinate.

Lemma 5.6 (Invertibility via distance)**.** *Let A be a random matrix with i.i.d. entries. Let $X_1, \dots, X_n$ denote the column vectors of A, and let H_k denote the span of all column vectors except the k-th. Then for every $\varepsilon > 0$, one has*

$$\mathbb{P}\Big(\inf_{x \in Incomp} \|Ax\|_2 < \varepsilon \nu_2 n^{-1/2} \Big) \leq \frac{1}{\nu_1} \cdot \mathbb{P}\big(\operatorname{dist}(X_n, H_n) < \varepsilon \big). \tag{5.10}$$

Proof. Denote

$$p := \mathbb{P}\big(\operatorname{dist}(X_k, H_k) < \varepsilon \big).$$

Note that since the entries of the matrix A are i.i.d., this probability does not depend on k. Then

$$\mathbb{E}\Big| \{k : \operatorname{dist}(X_k, H_k) < \varepsilon\} \Big| = np.$$

Denote by U the event that the set $\sigma_1 := \{k : \operatorname{dist}(X_k, H_k) \geq \varepsilon\}$ contains more than $(1 - \nu_1)n$ elements. Then by Chebychev's inequality,

$$\mathbb{P}(U^c) \leq \frac{p}{\nu_1}.$$

Assume that the event U occurs. Fix any incompressible vector x and let $\sigma(x)$ be the set from Lemma 5.5. Then $|\sigma_1| + |\sigma(x)| > (1 - \nu_1)n + \nu_1 n = n$, so the sets σ_1 and $\sigma(x)$ have nonempty intersection. Let $k \in \sigma_1 \cap \sigma(x)$, so

$$|x_k| \geq \nu_2 n^{-1/2} \quad \text{and} \quad \operatorname{dist}(X_k, H_k) \geq \varepsilon.$$

Writing $Ax = \sum_{j=1}^n x_j X_j$, we get

$$\begin{aligned} \|Ax\|_2 &\geq \operatorname{dist}(Ax, H_k) = \operatorname{dist}(x_k X_k, H_k) = |x_k| \operatorname{dist}(X_k, H_k) \\ &\geq \nu_2 n^{-1/2} \cdot \varepsilon. \end{aligned}$$

Summarizing, we have shown that

$$\mathbb{P}\Big(\inf_{x \in Incomp} \|Ax\|_2 < \varepsilon \nu_2 n^{-1/2} \Big) \leq \mathbb{P}(U^c) \leq \frac{p}{\nu_1}.$$

This completes the proof. □

Lemma 5.6 reduces the invertibility problem to a lower bound on the distance between a random vector and a random subspace. Now we reduce bounding the distance to a small ball probability estimate.

Let $X_1, \dots, X_n$ be the column vectors of A. Let X^* be any unit vector orthogonal to $X_1, \dots, X_{n-1}$. We call it a *random normal*. We can choose X^* so that it is a random vector that depends only on $X_1, \dots, X_{n-1}$ and is independent of X_n.

We clearly have

$$\operatorname{dist}(X_n, H_n) \ge |\langle X^*, X_n \rangle|. \tag{5.11}$$

The vectors $X^* =: (a_1, \dots, a_n)$ and $X_n =: (\xi_1, \dots, \xi_n)$ are independent. Condition on the vectors $X_1, \dots, X_{n-1}$. Then the vector X^* can be viewed as fixed, and the problem reduces to the small ball probability estimate for a linear combination of independent random variables

$$\langle X^*, X_n \rangle = \sum_{k=1}^{n} a_k \xi_k.$$

Assume for a moment that the distribution of a random variable ξ is absolutely continuous with bounded density. Then

$$\mathbb{P}(|\xi| < t) \le C' t \quad \text{for any } t > 0. \tag{5.12}$$

This estimate can be extended to a linear combination of independent copies of ξ. Therefore,

$$\mathbb{P}(|\langle X^*, X_n \rangle| < t \mid X^*) \le Ct.$$

Integrating over $X_1, \dots, X_{n-1}$, we obtain

$$\mathbb{P}(|\langle X^*, X_n \rangle| < t) \le Ct.$$

Thus, combining this estimate with Lemma 5.6, we prove that

$$\mathbb{P}\Big(\inf_{x \in Incomp} \|Ax\|_2 < \varepsilon \nu_2 n^{-1/2} \Big) \le C\varepsilon.$$

Then (5.9) and Lemma 5.4 imply Theorem 5.2 in this case.

6. Arithmetic structure and the small ball probability

To prove Theorem 5.2 in the previous section, we used the small ball probability estimate (5.12). However, this estimate does not hold for a general subgaussian random variable, and in particular for any random variable having an atom at 0.

Despite this, a linear combination $\sum_{k=1}^{n} a_k \xi_k$ of independent copies of a subgaussian random variable ξ obeys an estimate similar to (5.12) for a *typical* vector $a = (a_1, \dots, a_n)$. It is easy to see that such estimate is impossible for

all vectors $a \in S^{n-1}$. Indeed, assume that ξ is the random ± 1 variable. Then for

$$a^{(1)} = \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0, \dots, 0 \right), \qquad \mathbb{P}\left(\sum_{k=1}^{n} a_k \xi_k = 0 \right) = \frac{1}{2}.$$

This singular behavior is due to the fact that the vector $a^{(1)}$ is sparse. If we choose the vector a, which is far from the sparse ones, i.e. an incompressible vector, the small ball probability may be significantly improved. Consider for example, the vector

$$a^{(2)} = \left(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}} \right).$$

Then by the Berry–Esséen Theorem,

$$\mathbb{P}\left(\left| \sum_{k=1}^{n} \frac{1}{\sqrt{n}} \xi_k \right| \le t \right) \le C \left(t + \frac{1}{\sqrt{n}} \right)$$

This estimate cannot be improved, since for an even n,

$$\mathbb{P}\left(\sum_{k=1}^{n} \frac{1}{\sqrt{n}} \xi_k = 0 \right) \ge \frac{c}{\sqrt{n}}.$$

The coordinates of the vector $a^{(2)}$ are the same, which results in a lot of cancelations in the random sum $\sum_{k=1}^{n} a_k \xi_k$. If the arithmetic structure of the coordinates of the vector a is less rigid, the small ball probability can be improved even further. For example, for the (not normalized) vector

$$a^{(3)} = \left(\frac{1 + 1/n}{\sqrt{n}}, \frac{1 + 2/n}{\sqrt{n}}, \dots, \frac{1 + n/n}{\sqrt{n}} \right), \qquad \mathbb{P}\left(\sum_{k=1}^{n} a_k \xi_k = 0 \right) \sim n^{-3/2}.$$

Determining the influence of the arithmetic structure of the coordinates of a vector a on the small ball probability for the random sum $\sum_{k=1}^{n} a_k \xi_k$ became known as the *Littlewood–Offord Problem*. It was investigated by Littlewood and Offord, Erdös, Sárcözy and Szeméredi, etc. Recently Tao and Vu [36] put forward the inverse Littlewood–Offord theorems, stating that the large value of the small ball probability implies a rigid arithmetic structure. The inverse Littlewood–Offord theorems are extensively discussed in [34], see also [21] for current results in this direction. We will need a result of this type for the conditional argument to compensate for the lack of the bound (5.12).

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 embeds. This length is conveniently expressed as the least common denominator of a, defined as follows:

$$\operatorname{lcd}(a) := \inf \Big\{ \theta > 0 : \ \theta a \in \mathbb{Z}^n \setminus \{0\} \Big\}.$$

For the vector $a^{(2)}$,

$$\operatorname{lcd}(a^{(2)}) = \sqrt{n} \sim 1 \Big/ \mathbb{P}\left(\sum_{k=1}^{n} a_k \xi_k = 0 \right).$$

A similar phenomenon occurs for the vector $a^{(3)}$:

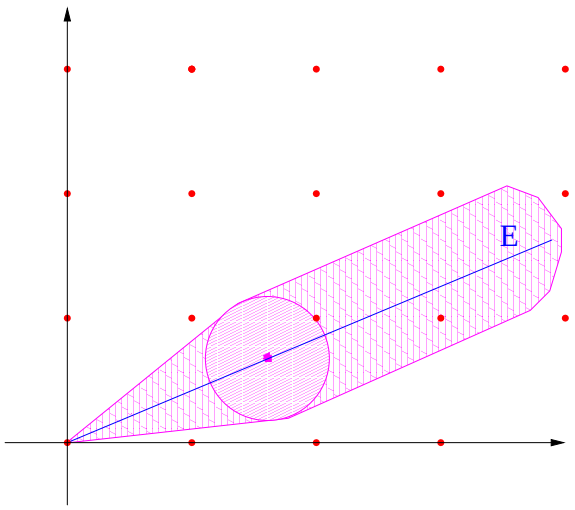
$$\operatorname{lcd}(a^{(3)}) = n^{3/2} \sim 1 \Big/ \mathbb{P}\left(\sum_{k=1}^{n} a_k \xi_k = 0 \right).$$

This suggests that the least common denominator of the sequence controls the small ball probability. However, in the case when $t > 0$, or when the random variable ξ is not purely discrete, the precise inclusion $\theta a \in \mathbb{Z}^n \setminus \{0\}$ loses its meaning. It should be relaxed to measure the closeness of the vector θa to the integer lattice. This leads us to the definition of the *essential least common denominator*.

Fix a parameter $\gamma \in (0,1)$. For $\alpha > 0$ define

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

The requirement that the distance is smaller than $\gamma \|\theta a\|_2$ forces us to consider only non-trivial integer points as approximations of θa – only those in a small aperture cone around the direction of a (see the picture below).



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. This choice would allow us to conclude that the least common denominator of any incompressible vector is of order at least $\sqrt{n}$. Let us formulate this statement precisely.

Lemma 6.1. *There exist constants $\gamma > 0$ and $\lambda > 0$ depending only on the compressibility parameters δ, ρ such that any incompressible vector a satisfies $\mathrm{LCD}_\alpha(a) \geq \lambda\sqrt{n}$.*

Proof. Assume that a is an incompressible vector, and let $\sigma(a)$ be the set defined in Lemma 5.5. If $\mathrm{LCD}_\alpha(a) < \lambda\sqrt{n}$, then

$$\|\theta a - z\|_2 < \gamma\theta < \gamma\lambda\sqrt{n} \qquad \text{for some } \theta \in (0, \lambda\sqrt{n}),\ z \in \mathbb{Z}^n.$$

Let $I(a)$ be the set of all $j \in \{1, \dots, n\}$ such that

$$|\theta a_j - z_j| < \frac{2\gamma\lambda}{\nu_1}.$$

The previous inequality implies that $|I(a)| > (1 - \nu_1/2)n$. Therefore, for the set $J(a) = I(a) \cap \sigma(a)$, we have

$$|J(a)| > \frac{\nu_1}{2} n.$$

For any $j \in J(a)$, we have

$$|z_j| < \theta|a_j| + \frac{2\gamma\lambda}{\nu_1} < \lambda\sqrt{n} \cdot \frac{\nu_3}{\sqrt{n}} + \frac{2\gamma\lambda}{\nu_1} < 1,$$

provided that λ is chosen so that $\lambda\left(\nu_3 + \frac{2\gamma}{\nu_2}\right) < 1$. Since $z \in \mathbb{Z}$, this means that $z_j = 0$. Finally, this implies

$$\|\theta a - z\|_2 \geq \left(\sum_{j \in J(a)} \theta^2 a_j^2\right)^{1/2} > \theta\nu_2\sqrt{\frac{\nu_1}{2}} > \gamma\theta$$

for $\gamma < \nu_2\sqrt{\nu_1/2}$. This contradicts the assumption that $\mathrm{LCD}_\alpha(a) < \lambda\sqrt{n}$. □

We fix γ satisfying Lemma 6.1 for the rest of the proof.

The following theorem gives a bound on the small ball probability for a random sum in terms of the additive structure of a. The less structure a has, the bigger its least common denominator is, and the smaller the small ball probability is.

Theorem 6.2 (Small ball probability)**.** *Let $\xi_1, \dots, \xi_n$ be independent copies of a centered subgaussian random variable ξ of unit variance. Consider a sequence $a = (a_1, \dots, a_n) \in S^{n-1}$. Then, for every $\alpha > 0$, and for*

$$\varepsilon \geq \frac{(4/\pi)}{\mathrm{LCD}_\alpha(a)},$$

we have

$$\mathbb{P}\left(\left|\sum_{k=1}^n a_k\xi_k\right| \leq \varepsilon\right) \leq C\varepsilon + Ce^{-c\alpha^2}.$$

We shall prove more than is claimed in the Theorem. Instead of the small ball probability we shall bound a parameter, which controls the concentration of a random variable around any fixed point.

Definition 6.3. The *Lévy concentration function* of a random variable S is defined for $\varepsilon > 0$ as

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

The proof of the Theorem uses the Fourier-analytic approach developed by Halász [14], [13].

We start with the classical Lemma of Esséen, which estimate the Lévy concentration function in terms of the characteristic function of a random variable.

Lemma 6.4. *Let Y be a real-valued random variable. Then*

$$\sup_{v \in \mathbb{R}} \mathbb{P}(|Y - v| \le 1) \le C \int_{-2}^{2} |\phi_Y(\theta)| \, d\theta,$$

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

Proof. Let $\psi = \chi_{[-1,1]} * \chi_{[-1,1]}$ and let $f = \hat{\psi}$:

$$f(t) = \left(\frac{2 \sin t}{t} \right)^2.$$

Then both $f \in L_1(\mathbb{R})$ and $\psi \in L_1(\mathbb{R})$, so f satisfies the Fourier inversion formula. Note also, that $f(t) \ge c$ whenever $|t| \le 1$. Therefore,

$$\begin{aligned}
&\mathbb{P}(|X - v| \le 1) = \mathbb{E} \chi_{[-1,1]}(X - v) \le \frac{1}{c} \mathbb{E} f(X - v) \\
&= \frac{1}{c} \mathbb{E} \left(\frac{1}{2\pi} \int_{\mathbb{R}} \psi(\theta) e^{i\theta(X - v)} \, d\theta \right) \le \frac{1}{2\pi c} \int_{\mathbb{R}} \psi(\theta) |\mathbb{E} e^{i\theta(X - v)}| \, d\theta \\
&\le \frac{1}{\pi c} \int_{-2}^{2} |\mathbb{E} e^{i\theta X}| \, d\theta.
\end{aligned}$$

The last inequality follows from $\operatorname{supp}(\psi) = [-2, 2]$ and $\psi(x) \le 2$. □

Proof of Theorem 6.2. To make the proof more transparent, we shall assume that ξ is the random ± 1 variable. The general case is considered in [24].

Let $S = \sum_{j=1}^{n} a_j \xi_j$. Applying Esséen's Lemma to the random variable $Y = S/\varepsilon$, we obtain

$$\mathcal{L}(S, \varepsilon) \le C \int_{-2}^{2} |\phi_S(\theta/\varepsilon)| \, d\theta = C \int_{-2}^{2} \prod_{j=1}^{n} |\phi_j(\theta/\varepsilon)| \, d\theta, \tag{6.1}$$

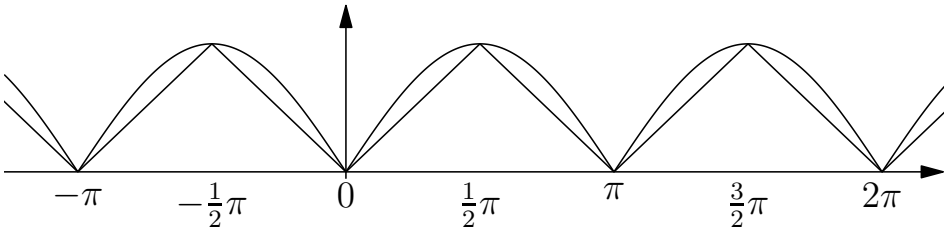
where

$$\phi_j(t) = \mathbb{E} \exp(i a_j \xi_j t) = \cos(a_j t).$$

The last equality in (6.1) follows from the independence of $\xi_j,\ j = 1, \dots, n$. The inequality $|x| \le \exp(-\frac{1}{2}(1 - x^2))$, which is valid for all $x \in \mathbb{R}$, implies

$$|\phi_j(t)| \le \exp\left(-\frac{1}{2}\sin^2(a_j t)\right) \le \exp\left(-\frac{1}{2}\min_{q\in\mathbb{Z}} |\frac{2}{\pi}a_j t - q|^2\right).$$

In the last inequality we estimated the absolute value of the sinus by a piecewise linear function, see the picture below.



Combining the previous inequalities, we get

$$\mathcal{L}(S,\varepsilon) \le C\int_{-2}^{2} \exp\left(-\frac{1}{2}\sum_{j=1}^{n}\min_{q\in\mathbb{Z}}\left|\frac{2}{\pi}a_j\cdot\frac{\theta}{\varepsilon} - q\right|^2\right) d\theta \tag{6.2}$$

$$= C\int_{-2}^{2} \exp(-h^2(\theta)/2)\, d\theta,$$

where

$$h(\theta) = \min_{p\in\mathbb{Z}^n}\left\|\frac{2}{\pi\varepsilon}\cdot\theta a - p\right\|_2.$$

Since by the assumption, $4/(\pi\varepsilon) \le \mathrm{LCD}_\alpha(a)$, the definition of the least common denominator implies that for any $\theta \in [-2, 2]$,

$$h(\theta) \ge \min(\gamma\frac{2\theta}{\pi\varepsilon}\,\|a\|_2, \alpha).$$

Recall that $\|a\|_2 = 1$. Then the previous inequality implies

$$\exp(-h^2(\theta)/2) \le \exp\left(-\left(\frac{2\gamma}{\pi\varepsilon}\theta\right)^2\Big/2\right) + \exp(-\alpha^2/2).$$

Substituting this into (6.2) we complete the proof. □

To apply the previous result for random matrices we shall combine it with the following *Tensorization Lemma*.

Lemma 6.5 (Tensorization)**.** *Let $\zeta_1, \dots, \zeta_m$ be independent real random variables, and let $K, \varepsilon_0 \ge 0$. Assume that for each k*

$$\mathbb{P}(|\zeta_k| < \varepsilon) \le K\varepsilon \qquad \textit{for all } \varepsilon \ge \varepsilon_0.$$

Then

$$\mathbb{P}\Big(\sum_{k=1}^m \zeta_k^2 < \varepsilon^2 m\Big) \le (CK\varepsilon)^m \qquad \text{for all } \varepsilon \ge \varepsilon_0,$$

where C is an absolute constant.

Proof. Let $\varepsilon \ge \varepsilon_0$. We have

$$\mathbb{P}\Big(\sum_{k=1}^m \zeta_k^2 < \varepsilon^2 m\Big) = \mathbb{P}\Big(m - \frac{1}{\varepsilon^2}\sum_{k=1}^m \zeta_k^2 > 0\Big) \le \mathbb{E}\exp\Big(m - \frac{1}{\varepsilon^2}\sum_{k=1}^m \zeta_k^2\Big)$$

$$= e^m \prod_{k=1}^m \mathbb{E}\exp(-\zeta_k^2/\varepsilon^2). \tag{6.3}$$

By Fubini's theorem,

$$\mathbb{E}\exp(-\zeta_k^2/\varepsilon^2) = \mathbb{E}\int_0^{\exp(-\zeta_k^2/\varepsilon^2)} ds = \int_0^1 \mathbb{P}\big(\exp(-\zeta_k^2/\varepsilon^2) > s\big)\, ds$$
$$= \int_0^\infty 2ue^{-u^2}\, \mathbb{P}(\zeta_k < \varepsilon u)\, du.$$

For $u \in (0,1)$, we have $\mathbb{P}(|\zeta_k| < \varepsilon u) \le \mathbb{P}(|\zeta_k| < \varepsilon) \le K\varepsilon$. This and the assumption of the lemma yields

$$\mathbb{E}\exp(-\zeta_k^2/\varepsilon^2) \le \int_0^1 2ue^{-u^2} K\varepsilon\, du + \int_1^\infty 2ue^{-u^2} K\varepsilon u\, du \le CK\varepsilon.$$

Putting this into (6.3) yields

$$\mathbb{P}\Big(\sum_{k=1}^m \zeta_k^2 < \varepsilon^2 m\Big) \le e^m (CK\varepsilon)^m.$$

This completes the proof. □

Combining Theorem 6.2 and Lemma 6.5 yields the multidimensional small ball probability estimate similar to the one we had for absolutely continuous random variable.

Lemma 6.6 (Invertibility on a single vector)**.** *Let A' be an $m \times n$ random matrix, whose entries are independent copies of a centered subgaussian random variable ξ. Then for any $\alpha > 0$, for every vector $x \in S^{n-1}$, and for every $t \ge 0$, satisfying*

$$t \ge \max\left(\frac{(4/\pi)}{\mathrm{LCD}_\alpha(x)},\ e^{-c\alpha^2}\right),$$

one has

$$\mathbb{P}\big(\|A'x\| < tn^{1/2}\big) \le (Ct)^m.$$

7. Putting all ingredients together

Now we have developed all necessary tools to prove the central result.

Theorem. 5.2. (Invertibility: subgaussian) *Let A be an $n \times n$ matrix whose entries are independent copies of a centered subgaussian real random variable. Then for every $\varepsilon \geq 0$ one has*

$$\mathbb{P}\big(s_n(A) \leq \varepsilon n^{-1/2}\big) \leq C\varepsilon + c^n,$$

where $C > 0$ and $c \in (0,1)$.

Recall that we have divided the unit sphere into compressible and incompressible vectors (see Definition 5.3 and inequality (5.9)), and proved that the first term in (5.9) is exponentially small. Applying Lemma 5.6 and (5.11), we reduced the estimate for the second term to the bound for

$$p(\varepsilon) := \mathbb{P}\big(|\langle X^*, X_n\rangle| \leq \varepsilon\big),$$

where X_n is the n-th column of the matrix A, and X^* is a unit vector orthogonal to the first $n-1$ columns. To complete the proof, we have to show that

$$p(\varepsilon) \leq C\varepsilon, \tag{7.1}$$

whenever $\varepsilon \geq e^{-cn}$. Here $\langle X^*, X_n\rangle = \sum_{j=1}^n X_j^* \xi_j$, where $X^* = (X_1^*, \dots, X_n^*)$. Throughout the rest of the proof set

$$\alpha = \beta\sqrt{n}, \tag{7.2}$$

where $\beta > 0$ is a small absolute constant, which will be chosen at the end of the proof. If $\mathrm{LCD}_\alpha(X^*) \geq e^{cn}$, then (7.1) follows from Theorem 6.2. Therefore, our problem has been further reduced to

Theorem 7.1 (Random normal)**.** *Let $X_1, \dots, X_{n-1}$ be random vectors whose coordinates are independent copies of a centered subgaussian random variable ξ. Consider a unit vector X^* orthogonal to all these vectors. There exist constants $c, c' > 0$ such that*

$$\mathbb{P}\big(\mathrm{LCD}_\alpha(X^*) < e^{cn}\big) \leq e^{-c'n}.$$

Intuitively, the components of a random vector should be arithmetically incommensurate to the extent that their essential LCD is exponential in n. This is rather obvious for a random vector uniformly distributed over the sphere. However, the distribution of the random normal X^* is more involved, and it requires some work to confirm this intuition.

Proof. Let A' be the $(n-1) \times n$ matrix with rows $X_1^T, \dots, X_{n-1}^T$. Then $X^* \in \mathrm{Ker}(A')$. The matrix A' has i.i.d. entries. We start with using the

decomposition similar to (5.9):

$$\begin{aligned}&\mathbb{P}\big(\exists X^* \in S^{n-2} \quad \mathrm{LCD}_\alpha(X^*) < e^{cn} \text{ and } A'X^* = 0\big)\\ &\le \mathbb{P}\big(\exists X^* \in \mathit{Comp} \quad A'X^* = 0\big)\\ &\quad + \mathbb{P}\big(\exists X^* \in \mathit{Incomp} \quad \mathrm{LCD}_\alpha(X^*) < e^{cn} \text{ and } A'X^* = 0\big).\end{aligned}$$

Lemma 5.4 implies that the first term in the right hand side does not exceed e^{-cn}. Formally, we have to reprove this lemma for $(n-1)\times n$ matrices, instead of the $n \times n$ ones, but the proof extends to this case without any changes.

To bound the second term, we introduce a new decomposition of the sphere. Recall that by Lemma 6.1, any incompressible vector a satisfies $\mathrm{LCD}_\alpha(a) \ge \lambda\sqrt{n}$. For $D > 0$, set

$$S_D = \{x \in S^{n-1} \mid D \le \mathrm{LCD}_\alpha(x) \le 2D\}.$$

It is enough to prove that

$$\mathbb{P}(\exists x \in S_D \ A'x = 0) \le e^{-n}.$$

whenever $\lambda\sqrt{n} \le D \le e^{cn}$. Indeed, the statement of the Theorem will then follow by taking the union bound over $D = 2^k$ for $k \le cn$.

To this end, we shall use the ε-net argument to bound $\|A'x\|$ below. For a fixed $x \in S_D$, the required estimate follows from substituting the bound $\mathrm{LCD}_\alpha(x) \ge D$ in Lemma 6.6:

$$\mathbb{P}\big(\|A'x\|_2 < tn^{1/2}\big) \le (Ct)^{n-1}, \tag{7.3}$$

provided $t \ge \frac{(4/\pi)}{D}$. To estimate the size of the ε-net we use the bound for the essential least common denominator again. The 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 7.2 (Nets of level sets)**.** *There exists a $(4\alpha/D)$-net in S_D of cardinality at most $(CD/\sqrt{n})^n$.*

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

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

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

$$\|D(x)x - p\|_2 < \alpha. \tag{7.4}$$

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

$$\left\| x - \frac{p}{\|p\|_2} \right\|_2 < \frac{2\alpha}{D}. \tag{7.5}$$

On the other hand, by (7.4) and using $\|x\|_2 = 1$, $D(x) \le 2D$ and $4\alpha/D \le 1$, we obtain

$$\|p\|_2 < D(x) + \alpha \le 2D + \alpha \le 3D. \tag{7.6}$$

Inequalities (7.5) and (7.6) show that the set

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

is a $(2\alpha/D)$-net of S_D. Recall that, by a known volumetric argument, the number of integer points in $B(0, 3D)$ is at most $(1 + 9D/\sqrt{n})^n \le (CD/\sqrt{n})^n$ (where in the last inequality we used that by the definition of the level set, $D > c_0\sqrt{n}$ for all incompressible vectors). Finally, we can find a $(4\alpha/D)$-net of the same cardinality, which lies in S_D. □

Now we can complete the ε-argument. Recall that by Proposition 4.4,

$$\mathbb{P}(s_1(A') \ge C_0\sqrt{n}) \le e^{-cn}.$$

Therefore, in order to complete the proof, it is enough to show that the event

$$\mathcal{E} := \Big\{ \exists x \in S_D \ \ A'x = 0 \text{ and } \|A'\| \le C_0\sqrt{n} \Big\}$$

has probability at most e^{-n}.

Assume that $\mathcal{E}$ occurs, and let $x \in S_D$ be such that $A'x = 0$. Let $\mathcal{N}$ be the $(4\alpha/D)$-net constructed in Lemma 7.2. Choose $y \in \mathcal{N}$ such that $\|x - y\| < 4\alpha/D$. Then by the triangle inequality,

$$\|A'y\|_2 \le \|A'\| \cdot \|x - y\|_2 < C_0\sqrt{n} \cdot \frac{4\alpha}{D} = 4C_0\beta\frac{n}{D},$$

if we recall that $\alpha = \beta\sqrt{n}$. Set $t = 4C_0\beta\sqrt{n}/D$. Combining the estimate (7.3) for this t with the union bound, we obtain

$$\begin{aligned}\mathbb{P}(\mathcal{E}) &\le \mathbb{P}(\exists y \in \mathcal{N} \ \ \|A'y\|_2 \le t\sqrt{n}) \le |\mathcal{N}| \cdot (Ct)^{n-1} \le \left(\frac{CD}{\sqrt{n}}\right)^n \cdot (Ct)^{n-1} \\ &\le \left(\frac{CD}{\sqrt{n}}\right) \cdot (4CC_0\beta)^{n-1}.\end{aligned}$$

Since $D \le e^{cn}$, we can choose the constant β so that the right hand side of the previous inequality will be less than e^{-n}. The proof of Theorem 5.2 is complete.

□

8. Short Khinchin inequality

Let $1 \le p < \infty$. Recall that $\|\cdot\|_p$ denotes the standard ℓ_p norm, and B_p^n its unit ball.

Let $X \in \mathbb{R}^n$ be a vector with independent centered random ± 1 coordinates, i.e. a random vertex of the discrete cube $\{-1,1\}^n$. The classical Khinchin inequality, Theorem 3.4, asserts that for any $p \ge 1$ and for any vector $a \in \mathbb{R}^n$, $\big(\mathbb{E}|\langle a, X\rangle|^p\big)^{1/p} \sim_p \|a\|_2$. This equivalence can be obtained if one averages not over the whole discrete cube, but over some small part of it. The problem how small should this set be was around since mid-seventies. More precisely,

> *Let $p \ge 1$. Find constants α_p, β_p and a set $V \subset \{-1,1\}$ of a small cardinality such that*
>
> $$\alpha_p \|a\|_2 \le \left(\frac{1}{|V|} \sum_{x \in V} |\langle a, x\rangle|^p \right)^{1/p} \le \beta_p \|a\|_2$$
>
> *for any $a \in \mathbb{R}^n$.*

Deterministic constructions of sets V of reasonably small cardinality are unknown. Therefore, we shall construct the set V probabilistically. Namely, we choose $N = N(n,p)$ and consider N independent copies $X_1, \dots, X_N$ of the random vector X. If $N \ll 2^{n/2}$, in particular, if N is polynomial in n, all vectors $X_1, \dots, X_N$ are distinct with high probability. The problem thus is reduced to showing that with high probability, any vector $a \in \mathbb{R}^n$ satisfies

$$\alpha_p \|a\|_2 \le \left(\frac{1}{N} \sum_{j=1}^{N} |\langle a, X_j\rangle|^p \right)^{1/p} \le \beta_p \|a\|_2 . \tag{8.1}$$

This problem can be recast in the language of random matrices. Let A be the $N \times n$ matrix with rows $X_1, \dots, X_N$. Then the inequality above means that A defines a nice isomorphic embedding of ℓ_2^n into ℓ_p^N.

As in the proof of the original Khinchin inequality, we consider cases $p = 1$ and $p > 2$ separately.

8.1. Short Khinchin inequality for $p = 1$.

In this case we derive the inequality (8.1) in a more general setup. Assume that the coordinates of the vector X are i.i.d. centered subgaussian variables. Then Proposition 4.4 combined with the inequality $\big\|A : \ell_2^n \to \ell_1^N\big\| \le \sqrt{N} \cdot \big\|A : \ell_2^n \to \ell_2^N\big\|$ yields the following

Proposition 8.1. *Let A be an $N \times n$ random matrix, $N \ge n$, whose entries are independent copies of a subgaussian random variable. Then*

$$\mathbb{P}\big(\big\|A : \ell_2^n \to \ell_1^N\big\| > tN \big) \le e^{-c_0 t^2 N} \qquad \text{for } t \ge C_0.$$

This implies the second inequality in (8.1) with $\beta_1 = C_0$, so (8.1) is reduced to the first inequality. To establish it we apply the random matrix machinery developed in the previous sections. Without loss of generality, we may assume that $n \le N \le 2n$, because we are looking for small values of N. Then the following Theorem shows that the short Khinchin inequality holds for any $N \ge n$ with α_1 depending only on the ratio of N/n.

Theorem 8.2. *Let n, N be natural numbers such that $n \le N \le 2n$. Let τ be a centered subgaussian random variable of variance 1. Let A be an $N \times n$ matrix, whose entries are independent copies of τ. Set $m = N - n + 1$. Then for any $\varepsilon > 0$*

$$\mathbb{P}\left(\exists x \in S^{n-1} \ \ \|Ax\|_1 < \varepsilon m\right) \le \left(\frac{CN}{m} \cdot \varepsilon\right)^m + C\exp(-cn).$$

Proof. Adding to the entries of A a small multiples of independent $N(0,1)$ variables, we may assume that the entries of A are absolutely continuous, so the matrix A is of a full rank almost surely.

We start with an elementary lemma from linear algebra.

Lemma 8.3. *Let $N > n$ and let $A : \mathbb{R}^n \to \mathbb{R}^N$ be a random matrix with absolutely continuous entries. Let $x \in S^{n-1}$ be a vector for which $\|Ax\|_1$ attains the minimal value. Then*

$$|\operatorname{supp}(Ax)| = N - n + 1$$

almost surely.

Proof. Let $E = A\mathbb{R}^n$ and let $K = B_1^N \cap E$. Set $y = Ax/\|Ax\|_1$. Since the function $g : S^{n-1} \to (0,\infty)$, $g(u) = \|Au\|_1$ attains the minimum at $u = x$, the function $f : K \to (0,\infty)$, $f(z) = \|A^{-1}|_E\, z\|_2$ attains the maximum over K at $z = y$. The convexity of $\|\cdot\|_2$ implies that y is an extreme point of K. Since K is the intersection of the octahedron B_1^N with an n-dimensional subspace, this means that $|\operatorname{supp} y| \le N - n + 1$. Finally, since the entries of A are absolutely continuous, any coordinate subspace $F \subset R^N$, whose dimension does not exceed $N - n$, satisfies $E \cap F = \{0\}$ a.s. Therefore, $|\operatorname{supp} y| = N - n + 1$. □

This Lemma allows us to reduce the minimum of $\|Ax\|_1$ over the whole sphere S^{n-1} to a certain finite subset of it. To each subset $J \subset \{1,\dots,N\}$ of cardinality $m = N - n + 1$ corresponds a unique pair of extreme points x_J and $-x_J$ of K such that $\sum_{j\in J} |x_J(j)| = 1$ and $x_J(j) = 0$ whenever $j \notin J$. Let $A_{J'}$ be the matrix consisting of the rows of A, whose numbers belong to $J' = \{1,\dots,N\} \setminus J$. The vector $y_J \in S^{n-1}$ such that $Ay_J = tx_J$ for some $t > 0$ is uniquely defined by the matrix $A_{J'}$ via the condition $A_{J'}y_J = 0$. By Lemma

8.3,

$$\min\{\|Ay\|_1 \mid y \in S^{n-1}\} = \min\{\|Ay_J\|_1 \mid J \subset \{1,\dots,N\},\ |J| = m\}.$$

To finish the proof, we estimate each $\|Ay_J\|_1$ below and apply the union bound. Fix a set $J \subset \{1,\dots,N\}$ of cardinality m. Denote the rows of the matrix $A_{J'}$ by $X_1^T,\dots,X_{n-1}^T$. Applying Theorem 7.1 to the vectors $X_1,\dots,X_{n-1}$, we conclude that

$$\mathbb{P}\big(\operatorname{LCD}_\alpha(y_J) < e^{cn}\big) \le e^{-c'n}. \tag{8.2}$$

Conditioning on the matrix $A_{J'}$, we may regard the vector y_J as fixed. Denote a row of the matrix A_J by Y^T, so the coordinates of $A_J y_J$ are distributed like $\langle Y, y_J\rangle$. If $\operatorname{LCD}_\alpha(y_J) \ge e^{cn}$, then by Theorem 6.2

$$\mathbb{P}(|\langle Y, y_J\rangle| \le \varepsilon \mid A_{J'}) \le C\varepsilon,$$

whenever $\varepsilon > Ce^{-cn}$. Then taking expectation over $A_{J'}$ and using (8.2) yields

$$\mathbb{P}(|\langle Y, y_J\rangle| \le \varepsilon) \le C\varepsilon + Ce^{-cn} + e^{-c'n}$$

for any $\varepsilon > 0$. Coordinates $\zeta_j,\ j \in J$ of the vector $A_J y_J$ are i.i.d. random variables. Tensorization Lemma 6.5 can be easily reproved for $\sum|\zeta_j|$ instead of $\sum \zeta_j^2$. In this form it implies

$$\mathbb{P}(\|Ay_J\|_1 \le \varepsilon m) = \mathbb{P}(\|A_J y_J\|_1 \le \varepsilon m) \le \big(C\varepsilon + Ce^{-cn}\big)^m$$

for any $\varepsilon > 0$. Finally, taking the union bound over all sets J, we obtain

$$\begin{aligned}\mathbb{P}(\exists J\ |J| = m,\ \|Ay_J\|_1 \le \varepsilon m) &\le \binom{N}{m} \cdot \big(C\varepsilon + Ce^{-cn}\big)^m \\ &\le \left(\frac{CN}{m} \cdot \varepsilon\right)^m + Ce^{-c''n}. \qquad \square\end{aligned}$$

Assume now that N is in a fixed proportion to n, and define δ by $N = (1+\delta)n$. Then Theorem 8.2 implies that, with high probability, the short Khinchin inequality holds for N independent subgaussian vectors with constant $\alpha_1 = c\delta^2$. To see this, set $\varepsilon = \frac{m}{2CN}$ to make the right-hand side of the inequality in Theorem 8.2 non-trivial.

Theorem 8.2 proves more than the short Khinchin inequality. Combining it with Proposition 4.4, we show that

$$\forall x \in \mathbb{R}^n \quad \varepsilon\delta n \|x\|_2 \le \|Ax\|_1 \le \sqrt{N}\,\|Ax\|_2 \le C'n\,\|x\|_2\,. \tag{8.3}$$

with probability greater than $1 - C\exp(-cn) - (\varepsilon/\bar{c}\delta)^{\delta n}$. This immediately yields a lower bound for the smallest singular value of a rectangular random matrix.

Corollary 8.4. *Let $n, N, \delta, A, \varepsilon$ be as above. Then the smallest singular value of A is bounded below by $\varepsilon\delta \cdot \sqrt{n}$ with probability at least $1 - C\exp(-cn) - (\varepsilon/\bar{c}\delta)^{\delta n}$.*

This bound is not sharp for small δ. The optimal estimate, valid for all n, N and ε, was recently obtained in [25].

A celebrated theorem of Kashin [12] states that a random n-dimensional section of the standard octahedron B_1^N of dimension $N = \lfloor (1+\delta)n \rfloor$ is close to the section of the inscribed ball $(1/\sqrt{N})B_2^N$. The optimal estimates for the diameter of a random section of the octahedron were obtained by Garnaev and Gluskin [7]. Recently the attention was attracted to the question whether the almost spherical sections of the octahedron can be generated by simple random matrices, in particular by a random ± 1 matrix. A general result proved in [18] implies that if $N = \lfloor (1+\delta)n \rfloor$ with $\delta \geq c/\log n$, then a random $N \times n$ matrix with independent subgaussian entries generates a section of the octahedron B_1^N which is not far from the ball with probability exponentially close to 1. For random ± 1 matrices this result was improved by Artstein-Avidan at al. [2], who proved a polynomial type estimate for the diameter of a section for $\delta \geq Cn^{-1/10}$. Using (8.3) we obtain a polynomial estimate for the diameter of sections for smaller values of δ.

Corollary 8.5. *Let n, N be natural numbers such that $n < N < 2n$. Denote $\delta = (N-n)/n$. Let ξ be a centered subgaussian random variable. Let A be an $N \times n$ matrix, whose entries are independent copies of ξ and let $E = A\mathbb{R}^n$. Then for any $\varepsilon > 0$*

$$\mathbb{P}\left(\forall y \in E,\ \|y\|_1 \leq \sqrt{N}\,\|y\|_2 \leq \frac{c}{\varepsilon\delta}\,\|y\|_1 \right) \geq 1 - C\exp(-cn) - (\varepsilon/\bar{c}\delta)^{\delta n}.$$

Note that to make the probability bound non-trivial, we have to assume that $\varepsilon = c'\delta$ for some $0 < c' < \bar{c}$. In this case the corollary means that a random n-dimensional subspace E satisfies

$$\frac{1}{\sqrt{N}} B_2^N \cap E \subset B_1^N \cap E \subset \left(\frac{c}{\delta^2}\right) \cdot \frac{1}{\sqrt{N}} B_2^N.$$

This inclusion remains non-trivial as long as $\left(\frac{c}{\delta^2}\right) < \sqrt{N}$, i.e., as long as $\delta > cN^{-1/4}$.

8.2. **Short Khinchin inequality for $p > 2$.** The case $p > 2$ requires a completely different approach. In this case the assumption that the coordinates of the random vector X are independent becomes unnecessary. We shall assume instead that X is isotropic and subgaussian. The first property means that for any $y \in S^{n-1}$

$$\mathbb{E}\langle X, y\rangle^2 = 1,$$

while the second means that for any $y \in S^{n-1}$ the random variable $\langle X, y\rangle$ is centered subgaussian. By Theorem 3.3, any random vector with independent centered subgaussian coordinates of variance 1 is isotropic subgaussian. This includes, in particular, an appropriately scaled random vertex of the discrete cube $\{-1,1\}^n$.

We prove the following Theorem [10].

Theorem 8.6. *Let X be an isotropic subgaussian vector in $\mathbb{R}^n$. Let $X_1, \dots, X_N$ be independent copies of X. Then for any $p > 2$ and any $N \geq n^{p/2}$, the inequality*

$$c \|y\|_2 \leq \left(\frac{1}{N} \sum_{j=1}^{N} |\langle y, X_j\rangle|^p \right)^{1/p} \leq C\sqrt{p}\, \|y\|_2$$

holds with high probability for all $y \in \mathbb{R}^n$.

Proof. As in the classical Khinchin inequality, the first inequality in Theorem 8.6 is easy. Denote, as before, by A the $N \times n$ matrix with rows $X_1, \dots, X_N$. Assume that n is large enough, so that $N \geq n^{p/2} \geq \delta_0^{-1} n$, where δ_0 is the constant from Proposition 4.7. Combining this Proposition with the inequality $\|y\|_2 \leq N^{1/2-1/p} \cdot \|y\|_p$, valid for all $y \in \mathbb{R}^N$, we obtain

$$\mathbb{P}\big(\min_{x \in S^{n-1}} \|Ax\|_p \leq c_1 N^{1/p} \big) \leq e^{-c_2 N},$$

which establishes the left inequality with probability exponentially close to 1.

To prove the second inequality, we use the method of majorizing measures, or generic chaining, developed by Talagrand [32]. Let $\{X_t\}_{t \in T}$ be a real-valued random process, i.e., a collection of interdependent random variables, indexed by some set T. In the setup below, we can assume that T is finite or countable, eliminating the question of measurability of $\sup_{t \in T} X_t$. We shall call the process $\{X_t\}_{t \in T}$ centered if $\mathbb{E} X_t = 0$ for all $t \in T$.

Definition 8.7. Let (T, d) be a metric space. A random process $\{X_t\}_{t \in T}$ is called subgaussian with respect to the metric d if for any $t, s \in T$, $t \neq s$ the random variable $(X_t - X_s)/d(t,s)$ is subgaussian. A random process $\{G_t\}_{t \in T}$ is called Gaussian with respect to the metric d if for any finite set $F \subset T$ the joint distribution of $\{G_t\}_{t \in F}$ is Gaussian, and for any $t, s \in T$, $t \neq s$ $(G_t - G_s)/d(t,s)$ is $N(0,1)$ random variable.

We use the following fundamental result due to Talagrand.

Theorem 8.8 (Majorizing Measure Theorem)**.** *Let (T, d) be a metric space, and let $\{G_t\}_{t \in T}$ be a Gaussian random process with respect to the metric d. For any centered random process $\{X_t\}_{t \in T}$, which is subgaussian with respect to the same metric,*

$$\mathbb{E} \sup_{t \in T} X_t \leq C\, \mathbb{E} \sup_{t \in T} G_t.$$

For $(s, y) \in \mathbb{R}^N \times \mathbb{R}^n$ define the random variable $X_{s,y}$ by

$$X_{s,y} = \sum_{j=1}^{N} s_j \langle X_j, y \rangle.$$

Then for any $T \subset B_2^N \times B_2^n$, the random process $\{X_{s,y}\}_{(s,y)\in T}$ is subgaussian with respect to the Euclidean metric. Indeed, for any $(s, y), (s', y') \in T$,

$$X_{s,y} - X_{s',y'} = \sum_{j=1}^{N} \Big((s_j - s'_j)\langle X_j, y \rangle + s'_j \langle X_j, y - y' \rangle \Big).$$

Let $\lambda \in \mathbb{R}$. Since the vector X is centered subgaussian, for any $z \in \mathbb{R}^N$ $\exp(\lambda \langle X, z \rangle) \le \exp(C\lambda^2 \|z\|_2^2)$. Hence, using independence of X_j and applying Cauchy–Schwartz inequality, we get

$$\begin{aligned}
&\mathbb{E} \exp\big(\lambda(X_{s,y} - X_{s',y'})\big) \\
&= \prod_{j=1}^{N} \mathbb{E}\Big[\exp\big(\lambda(s_j - s'_j)\langle X_j, y \rangle\big) \cdot \exp\big(\lambda s'_j \langle X_j, y - y' \rangle\big) \Big] \\
&\le \prod_{j=1}^{N} \exp\big(2C\lambda^2 ((s_j - s'_j)^2 \|y\|_2^2)\big) \cdot \prod_{j=1}^{N} \exp\big(2C\lambda^2 (s_j'^2 \|y - y'\|_2^2)\big) \\
&\le \exp\big(2C\lambda^2 (\|s - s'\|_2^2 + \|y - y'\|_2^2)\big).
\end{aligned}$$

By Theorem 3.2 this means that the random variable

$$\frac{X_{s,y} - X_{s',y'}}{\|(s, y) - (s', y')\|_2}$$

is subgaussian.

Let Y and Z be independent standard Gaussian vectors in $\mathbb{R}^n$ and $\mathbb{R}^N$ respectively. Set

$$G_{s,y} = \langle s, Z \rangle + \langle y, Y \rangle.$$

Then for any $T \subset \mathbb{R}^N \times \mathbb{R}^n$, $\{G_{s,y}\}_{(s,y)\in T}$ is a Gaussian process with respect to the Euclidean metric. Let $1/p + 1/p^* = 1$, and set $T = B_{p^*}^N \times B_2^n \subset B_2^N \times B_2^n$. By the Majorizing Measure Theorem

$$\mathbb{E} \sup_{(s,y)\in T} X_{s,y} \le C\, \mathbb{E} \sup_{(s,y)\in T} G_{s,y}.$$

Therefore,

$$\mathbb{E}\sup_{y\in B_2^n}\left(\frac{1}{N}\sum_{j=1}^N |\langle X_j,y\rangle|^p\right)^{1/p} = \frac{1}{N^{1/p}}\mathbb{E}\sup_{s\in B_{p^*}^N}\sup_{y\in B_2^n}\sum_{j=1}^N s_j\langle X_j,y\rangle$$
$$\le \frac{C}{N^{1/p}}\mathbb{E}\sup_{s\in B_{p^*}^N}\sup_{y\in B_2^n} G_{s,y} = \frac{C}{N^{1/p}}\big(\mathbb{E}\,\|Z\|_p + \mathbb{E}\,\|Y\|_2\big)$$
$$\le C\left(\sqrt{p}+\frac{\sqrt{n}}{N^{1/p}}\right).$$

Since $N \ge n^{p/2}$, the last expression does not exceed $C'\sqrt{p}$. To complete the proof we combine this estimate of the expectation with Chebyshev's inequality. □

Note that Theorem 8.6 implies that the matrix A formed by the vectors $X_1,\dots,X_N$ defines a subspace of ℓ_p^N which is close to Euclidean. This, in particular, means that the bound $N \ge n^{p/2}$ is optimal (see e.g., [8] for details).

9. Random unitary and orthogonal perturbations

The need for probabilistic bounds for the smallest singular value of a random matrix from a certain class arises in many intrinsic problems of the random matrix theory. Such bounds are the standard step in many proofs based on the convergence of Stieltjes transforms of the empirical measures to the Stieltjes transform of the limit measure. One of the examples, where such bounds become necessary is the Circular Law. The proof of this law requires the lower bound on the smallest singular value of a random matrix with i.i.d. entries, which was obtained above. Another setup, where such bounds become necessary, is provided by the Single Ring Theorem of Guionnet, Krishnapur and Zeitouni [11]. The proof of this theorem deals with another natural class of random matrices, namely random unitary or orthogonal perturbations of a fixed matrix.

Let us consider the complex case first. Let D be a fixed $n\times n$ matrix, and let U be a random matrix uniformly distributed over the unitary group $U(n)$. In this case the solution of the qualitative invertibility problem is trivial, since the matrix $D+U$ is non-singular with probability 1. This can be easily concluded by considering the determinant of $D+U$. The determinant, however, provides a poor tool for studying the quantitative invertibility problem. In regard to this problem we will prove the following theorem.

Theorem 9.1. *Let D be an arbitrary $n\times n$ matrix, $n\ge 2$. Let U be a random matrix uniformly distributed over the unitary group $U(n)$. Then*

$$\mathbb{P}(s_n(D+U)\le t) \le t^c n^C \quad \textit{for all } t>0.$$

Here C and c are absolute constants.

An important feature of Theorem 9.1 is its independence of the matrix D. This independence is essential for the Single Ring Theorem.

The statement similar to Theorem 9.1 fails in the real case, i.e., for random matrices distributed over the orthogonal group. Indeed, suppose that n is odd. If $-D, U \in SO(n)$, then $-D^{-1}U \in SO(n)$ has the eigenvalue 1, and the matrix $D + U = D(D^{-1}U + I_n)$ is singular. Therefore, if U is uniformly distributed over $O(n)$, then $s_n(D + U) = 0$ with probability at least $1/2$. Nevertheless, it turns out that this is essentially the only obstacle to the extension of Theorem 9.1 to the orthogonal case.

Theorem 9.2 (Orthogonal perturbations)**.** *Let D be a fixed $n \times n$ real matrix, $n \geq 2$. Assume that*

$$\|D\| \leq K, \quad \inf_{V \in O(n)} \|D - V\| \geq \delta \tag{9.1}$$

for some $K \geq 1$, $\delta \in (0, 1)$. Let U be a random matrix uniformly distributed over the orthogonal group $O(n)$. Then

$$\mathbb{P}(s_n(D + U) \leq t) \leq t^c (Kn/\delta)^C, \quad t > 0.$$

Similarly to the complex case, this bound is uniform over all matrices D satisfying (9.1). This condition is relatively mild: in the case when $K = n^{C_1}$ and $\delta = n^{-C_2}$ for some constants $C_1, C_2 > 0$, we have

$$\mathbb{P}(s_n(D + U) \leq t) \leq t^c n^C, \quad t > 0,$$

as in the complex case. It is possible that the condition $\|D\| \leq K$ can be eliminated from the Theorem 9.2. However, this is not crucial because such condition already appears in the Single Ring Theorem.

The problems we face in the proofs of Theorems 9.1 and 9.2 are significantly different from those appearing in Sections 5, 7. In the case of the independent entries the argument was based on the analysis of the small ball probability $\mathbb{P}(\|Ax\|_2 < t)$ or $\mathbb{P}(\|Ax\|_1) < t$ for a fixed vector x. As shown in Section 6, the decay of this probability as $t \to 0$ is determined by the arithmetic structure of the coordinates of x. In contrast to this, the arithmetic structure plays no role in Theorems 9.1 and 9.2. The difficulty lies elsewhere, namely in the lack of independence of the entries of the matrix. We will have to introduce a set of the independent random variables artificially. These variables have to be chosen in a way that allows one to express tractably the smallest singular value in terms of them. To illustrate this approach, we present the proof of Theorem 9.1 below. The proof of Theorem 9.2 starts with the similar ideas, but requires new and significantly more delicate arguments. We refer the reader to [28] for the details.

Proof of Theorem 9.1. Throughout the proof we fix $t > 0$ and introduce several small and large parameters depending on t. The values of such parameters will be chosen of orders t^a, where $0 < a < 1$ for the small parameters, and t^{-b}, $0 < b < 1$ for the large ones. This would allow us to introduce an hierarchy of parameters, and disregard the terms corresponding to the smaller ones. Also, note that we have to prove Theorem 9.1 only for $t < n^{-C'}$ for a given constant C', because for larger values of t its statement can be made vacuous by choosing a large constant C. This observation would allow us to use bounds of the type $\sqrt{n}t^a \le t^{a'}$ whenever $a < a'$ are constants.

For convenience of a reader, we include a special paragraph entitled "Choice of the parameters" in the analysis of each case. In these paragraphs we list the constraints that the small and large parameters must satisfy, as well as the admissible numerical values of those parameters. These paragraphs will be printed in sans-serif and can be omitted on the first reading.

To simplify the argument, we will also assume that $\|D\| \le K$, as in Theorem 9.2. The proof of Theorem 9.1 without this assumption can be found in [28].

9.1. **Decomposition of the sphere and introduction of local and global perturbations.** We have to bound $s_n(U + D)$, which is the minimum of $\|(D + U)x\|_2$ over the unit sphere. For every $x \in S^{n-1}$, there is a coordinate x_j with $|x_j| \ge 1/\sqrt{n}$. Hence, the union bound yields

$$\mathbb{P}(s_n(D + U) \le t) \le \sum_{j=1}^{n} \mathbb{P}\left(\inf_{x \in S_j} \|(U + D)x\|_2 \le t \right),$$

where

$$S_j = \left\{ x \in S^{n-1} \mid |x_j| \ge 1/\sqrt{n} \right\}.$$

All terms on the right hand side of the inequality above can be estimated in the same way. So, without loss of generality we will consider the case $j = 1$. Note that the application of the crude union bound here may have increased the probability estimate of Theorem 9.1 n times. This, however, is unimportant, since we allow the coefficient n^C anyway.

The proof of the theorem reduces to the estimate of

$$\mathbb{P}\left(\inf_{x \in S_1} \|(U + D)x\|_2 \le t \right). \tag{9.2}$$

The structure of the set S_1 gives a special role to the first coordinate. This will be reflected in our choice of independent random variables. If $R, W \in U(n)$ are any matrices, and V is uniformly distributed over $U(n)$, then the matrix $U = V^{-1}R^{-1}W$ is uniformly distributed over $U(n)$ as well. Hence, if we assume that the matrices R and W are random and independent of V, then this property would remain valid for U. The choice of the distributions of R

and W is in our hands. Set

$$R = \operatorname{diag}(r, 1, \dots, 1),$$

where r is a random variable uniformly distributed over $\{z \in \mathbb{C} \mid |z| = 1\}$. This is a "global" perturbation, since we will need the values of r, which are far from 1. The matrix W will be "local", i.e., it will be a small perturbation of the identity matrix. Let $\varepsilon > 0$ be a "small" parameter, and set $W = \exp(\varepsilon S)$, where S is an $n \times n$ skew-symmetric matrix, i.e. $S^* = -S$. Although he matrix W is unitary, the dependence of its entries on the entries of S is hard to trace. To simplify the structure, we consider the linearization of W,

$$W_0 = I + \varepsilon S.$$

The matrix W_0 is not unitary, but its distance to the group $U(n)$ is at most $\|W - W_0\| \le \varepsilon^2 \|S\|^2$. Thus, for any $x \in S_1$,

$$\begin{aligned} \|(D+U)x\|_2 = \|(D + V^{-1}R^{-1}W)x\|_2 &= \|(RVD + W)x\|_2 \\ &\ge \|(RVD + W_0)x\|_2 - \|W - W_0\| \\ &\ge \|(RVD + I + \varepsilon S)x\|_2 - \varepsilon^2 \|S\|^2. \end{aligned}$$

We will use S to introduce a collection of independent random variables. Set

$$S = \begin{bmatrix} \sqrt{-1}\, s & -Z^T \\ Z & 0 \end{bmatrix} \tag{9.3}$$

where $s \sim N_{\mathbb{R}}(0,1)$ and $Z \sim N_{\mathbb{R}}(0, I_{n-1})$ are independent real-valued standard normal random variable and vector respectively. Clearly, S is skew-Hermitian. If K_0 is a "large" parameter, $K_0 = t^{-b_0}$, then by Proposition 4.4,

$$\mathbb{P}(\|Z\|_2 \ge K_0\sqrt{n}) \le \exp(-c_0 K_0^2 n) \le t$$

for all sufficiently small $t > 0$. This means that $\|S\|^2 \le K_0^2 n$ with probability close to 1. Disregarding an event of a small probability, we reduce the problem to obtaining a lower bound for

$$\inf_{x \in S_1} \|(RVD + I + \varepsilon S)x\|_2 ,$$

provided that the bound we obtain is of order at least ε. Indeed, we may assume that $K_0^2 n \varepsilon^2 \ll \varepsilon$, if ε is chosen small enough.

Choice of the parameters. The second order term $2\varepsilon^2 \|S^2\|$ should not affect the estimate of $\mathbb{P}(\inf_{x \in S_1} \|Ax\| \le t)$. To guarantee it, we require that

$$K_0^2 n \varepsilon^2 \le t/2.$$

Also, to bound the probability by a power of t, we have to assume that

$$\exp(-c_0 K_0^2 n) \le t^c$$

for some $c > 0$. Both inequalities are satisfied for small t if $\varepsilon = t^{0.6}$ and $K_0 = t^{-0.05}$.

Starting from this moment we will condition on the matrix V and evaluate the conditional probability with respect to the random matrices R and S. The original random structure will be lost after this conditioning. However, we introduced a new independent structure in the form of the matrices R and S, and it will be easier to manipulate. Each of the matrices R and S alone is insufficient to obtain any meaningful estimate. Nevertheless, the combination of these two sources of randomness, a local perturbation S and a global perturbation R, produces enough power to conclude that $RVD + I + \varepsilon S$ is typically well invertible, and this leads to the proof of Theorem 9.1.

Summarizing the previous argument, we conclude that our goal is to bound

$$\mathbb{P}(\inf_{x \in S_1} \|Ax\|_2 \le t),$$

where

$$A = RVD + I + \varepsilon S =: \begin{bmatrix} A_{11} & Y^T \\ X & B^T \end{bmatrix}, \tag{9.4}$$

$X, Y \in \mathbb{C}^{n-1}$, B is an $(n-1) \times (n-1)$ matrix, and $\varepsilon = t^a$. Here we decomposed the matrix A separating the first coordinate to emphasize its special role. For future reference we write A in terms of the components of the matrix VD, and random variables r, s, and Z exposing the dependence on these random parameters:

$$A = \begin{bmatrix} A_{11} & Y^T \\ X & B^T \end{bmatrix} = \begin{bmatrix} ra + 1 + \sqrt{-1}\, \varepsilon s & (rv - \varepsilon Z)^T \\ u + \varepsilon Z & B^T \end{bmatrix}. \tag{9.5}$$

Here $a \in \mathbb{C}$, $u, v \in \mathbb{C}^{n-1}$, and the matrix B are independent of r, s, and Z. After conditioning on V, we can treat them as constants.

The further strategy takes into account the properties of the matrix B. Depending on the invertibility properties of this matrix, we condition on some of the random variables r, s, and Z, and use the other ones to show that A is well-invertible with high probability.

9.2. **Case 1: B is poorly invertible.** Assume that $s_n(B) \le \lambda_1 \varepsilon$, where λ_1 is another "small" parameter ($\lambda_1 = t^{a_1}$ for $0 < a_1 < 1$). In this case we will condition on r and s, and rely on Z to obtain the probability bound. We know that there exists a vector $\tilde{w} \in S^{n-2}$ such that $\|B\tilde{w}\|_2 \le \lambda_1 \varepsilon$. Let $x \in S_1$ be arbitrary. We can express it as

$$x = \begin{bmatrix} x_1 \\ \tilde{x} \end{bmatrix}, \quad \text{where } |x_1| \ge \frac{1}{\sqrt{n}}.$$

Set

$$w = \begin{bmatrix} 0 \\ \tilde{w} \end{bmatrix} \in \mathbb{C}^n.$$

Using the decomposition of A given in (9.4), we obtain

$$\begin{aligned}\|Ax\|_2 \geq |w^T Ax| &= \left| \begin{bmatrix} 0 & \tilde{w}^T \end{bmatrix} \begin{bmatrix} A_{11} & Y^T \\ X & B^T \end{bmatrix} \begin{bmatrix} x_1 \\ \tilde{x} \end{bmatrix} \right| \\ &= |x_1 \cdot \tilde{w}^T X + \tilde{w}^T B^T \tilde{x}| \\ &\geq |x_1| \cdot |\tilde{w}^T X| - \|B\tilde{w}\|_2 \quad \text{(by the triangle inequality)} \\ &\geq \frac{1}{\sqrt{n}} |\tilde{w}^T X| - \lambda_1 \varepsilon \quad \text{(using } |x_1| \geq 1/\sqrt{n}\text{)}.\end{aligned}$$

By the representation (9.5), $X = u + \varepsilon Z$, where $u \in \mathbb{C}^{n-1}$ is a vector independent of Z. Taking the infimum over $x \in S_1$, we obtain

$$\inf_{x \in S_1} \|Ax\|_2 \geq \frac{1}{\sqrt{n}} |\tilde{w}^T u + \varepsilon \tilde{w}^T Z| - \lambda_1 \varepsilon.$$

Recall that $\tilde{w}$, u are fixed vectors, $\|\tilde{w}\|_2 = 1$, and $Z \sim N_{\mathbb{R}}(0, I_{n-1})$. Then $\tilde{w}^T Z = \gamma$ is a complex normal random variable of variance 1: $\mathbb{E}|\gamma|^2 = 1$. This means that $\mathbb{E}\big(\mathrm{Re}(\gamma)\big)^2 \geq 1/2$ or $\mathbb{E}\big(\mathrm{Im}(\gamma)\big)^2 \geq 1/2$. A quick density calculation yields the following bound on the conditional probability:

$$\mathbb{P}_Z \left\{ |\tilde{w}^T u + \varepsilon \tilde{w}^T Z| \leq 2\lambda_1 \varepsilon \sqrt{n} \right\} \leq C\lambda_1 \sqrt{n}.$$

Therefore, a similar bound holds unconditionally. Thus, combining the previous estimates, we conclude that in case when $s_n(B) \leq \lambda_1 \varepsilon$, and if ε and λ_1 are chosen so that $\lambda_1 \varepsilon \geq t$, we have

$$\begin{aligned}\mathbb{P}(\inf_{x \in S_1} \|Ax\|_2 \leq t) &\leq \mathbb{P}(\frac{1}{\sqrt{n}} |\tilde{w}^T X| - \lambda_1 \varepsilon \leq t) \\ &\mathbb{P}\left\{ |\tilde{w}^T u + \varepsilon \tilde{w}^T Z| \leq 2\lambda_1 \varepsilon \sqrt{n} \right\} \leq C\lambda_1 \sqrt{n} = C\sqrt{n} \cdot t^{a_1}.\end{aligned}$$

Choice of the parameters. The constraint

$$\lambda_1 \varepsilon \geq t,$$

appearing in this case, holds if we take $\lambda_1 = t^{0.1}$.

9.3. **Case 2: B is nicely invertible.** Assume that $s_n(B) \geq \lambda_2$, where $\lambda_2 = t^{a_2}$ is a "small" parameter. In this case, we will also use only the local perturbation, however the crucial random variable will be different. We will condition on r and Z, and use the dependence on s to derive the conclusion of the theorem.

Set

$$M = \begin{bmatrix} 1 & 0 \\ 0 & (B^T)^{-1} \end{bmatrix},$$

then $\|M\| \leq \lambda_2^{-1}$. Therefore,

$$\inf_{x \in S_1} \|Ax\|_2 \geq \lambda_2 \inf_{x \in S_1} \|MAx\|_2 .$$

The matrix MA has the following block representation:

$$MA = \begin{bmatrix} A_{11} & Y^T \\ (B^T)^{-1}X & I_{n-1} \end{bmatrix}.$$

Recall that we assumed that $\|D\| \le K$ where K is a constant. Combining this with the already used inequality $\|Z\|_2 \le K_0\sqrt{n}$, which holds outside of the event of exponentially small probability, we conclude that

$$\|Y\|_2 \le \|v\|_2 + \varepsilon \|Z\|_2 \le 2K$$

if $\varepsilon K_0\sqrt{n} \le K$. To bound $\inf_{x\in S_1} \|Ax\|_2$, we use an observation that

$$\begin{bmatrix} 1 & -Y^T \end{bmatrix} \cdot \begin{bmatrix} Y^T \\ I_{n-1} \end{bmatrix} = 0.$$

This implies that for every $x \in S_1$,

$$\begin{aligned} \|MAx\|_2 &\ge \frac{1}{\|[1 \ -Y^T]\|_2} \cdot \left| \begin{bmatrix} 1 & -Y^T \end{bmatrix} MA \begin{bmatrix} x_1 \\ \tilde{x} \end{bmatrix} \right| \\ &\ge \frac{1}{2K} \cdot |A_{11} - Y^T(B^T)^{-1}X| \cdot |x_1| \\ &\ge \frac{1}{2K\sqrt{n}} \cdot |A_{11} - Y^T(B^T)^{-1}X|. \end{aligned}$$

The right hand side of this inequality does not depend on x, so we can take the infimum over $x \in S_1$ in the left hand side. Combination of the previous two inequalities reads

$$\inf_{x\in S_1} \|Ax\|_2 \ge \frac{\lambda_2}{2K\sqrt{n}} \cdot |A_{11} - Y^T(B^T)^{-1}X|$$

Recall that according to (9.5), $A_{11} = \sqrt{-1}\varepsilon s + d$, where s is a real $N(0,1)$ random variable, and d is independent of s. Conditioning on everything, but s, we can treat d and $Y^T(B^T)^{-1}X$ as constants. An elementary estimate using the normal density yields

$$\mathbb{P}_s(|A_{11} - Y^T(B^T)^{-1}X| \le \mu) \le C\frac{\mu}{\varepsilon} \quad \text{for all } \mu > 0.$$

Applying this estimate with $\mu = \frac{2K\sqrt{n}}{\lambda_2} \cdot t$ and integrating over the other random variables, we obtain

$$\mathbb{P}(\inf_{x\in S_1} \|Ax\|_2 \le t) \le C\frac{2K\sqrt{n}}{\lambda_2\varepsilon} \cdot t \le C'\sqrt{n} \cdot t^c$$

for some $c > 0$ if λ_2 is chosen appropriately.

Choice of the parameters. The inequality

$$\frac{1}{\lambda_2\varepsilon} \cdot t \le t^c, \ c > 0$$

holds with $c = 0.2$ if we set $\lambda_2 = t^{0.2}$. The constraint

$$\varepsilon K_0 \sqrt{n} \le K,$$

appearing above, is satisfied since we have chosen $\varepsilon = t^{0.6}$ and $K_0 = t^{-0.05}$.

One can try to tweak the parameters λ_1, λ_2, and ε to cover all possible scenarios. This attempt, however, is doomed to fail since the system of the constraints becomes inconsistent. Indeed, to include all matrices B in Cases 1 and 2, we have to choose $\lambda_2 \le \lambda_1 \varepsilon$. With this choice,

$$\frac{t}{\lambda_2 \varepsilon} \ge \frac{t}{\lambda_1 \varepsilon^2} > 1,$$

because of the constraint $K_0^2 n \varepsilon^2 \le t/2$. This forces us to consider the intermediate case.

9.4. **Case 3, intermediate: B is invertible, but not nicely invertible.** Assume that $\lambda_1 \varepsilon \le s_n(B) \le \lambda_2$ with λ_2, λ_1 defined in Cases 1 and 2. This is the most delicate case. Here we will have to rely on both local and global perturbations. We proceed like in Case 2 by multiplying Ax from the left by a vector which eliminates the dependence on all coordinates of x, except the first one. To this end, note that

$$\begin{bmatrix} 1 & -Y^T (B^T)^{-1} \end{bmatrix} \cdot \begin{bmatrix} Y^T \\ B^T \end{bmatrix} = 0.$$

Hence, for any $x \in S_1$,

$$\begin{aligned} \|Ax\|_2 &\ge \frac{1}{\left\| \begin{bmatrix} 1 & -Y^T (B^T)^{-1} \end{bmatrix} \right\|_2} \left| \begin{bmatrix} 1 & -Y^T (B^T)^{-1} \end{bmatrix} \cdot \begin{bmatrix} A_{11} & Y^T \\ X & B^T \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \tilde{x} \end{bmatrix} \right| \\ &\ge \frac{1}{1 + \|Y^T (B^T)^{-1}\|_2} \left| (A_{11} - Y^T (B^T)^{-1} X) x_1 \right| \\ &\ge \frac{1}{1 + \|Y^T (B^T)^{-1}\|_2} |A_{11} - Y^T (B^T)^{-1} X| \cdot \frac{1}{\sqrt{n}}. \end{aligned}$$

Since the right hand side is independent of x, we can take the infimum over $x \in S_1$.

Note that $Y^T (B^T)^{-1}$ is independent of s, see (9.5). We consider two subcases. If $\left\| Y^T (B^T)^{-1} \right\|_2 \le \lambda_2^{-1}$, then

$$\inf_{x \in S_1} \|Ax\|_2 \ge \frac{\lambda_2}{2\sqrt{n}} |A_{11} - Y^T (B^T)^{-1} X|,$$

and we can finish the proof exactly like in Case 2, by conditioning on everything except s, and estimating the probability with respect to s.

The second subcase requires more work. Assume that $\left\| Y^T (B^T)^{-1} \right\|_2 \ge \lambda_2^{-1}$. Then the inequality above yields

$$\inf_{x \in S_1} \|Ax\|_2 \ge \frac{1}{2\sqrt{n}\, \|Y^T (B^T)^{-1}\|_2} |A_{11} - Y^T (B^T)^{-1} X|.$$

Since we do not have a satisfactory upper bound for $\left\|Y^T(B^T)^{-1}\right\|_2$, we cannot rely on A_{11} to estimate the small ball probability. The second term in the numerator looks more promising, because it contains the same vector $Y^T(B^T)^{-1}$. This term, however, is difficult to analyze, since the random vectors X and Y are dependent. A simplification of both numerator and denominator would allow us to get rid of this dependence.

We start with analyzing the denominator. By (9.5), $Y = rv - \varepsilon Z$, so

$$\left\|Y^T(B^T)^{-1}\right\|_2 \le \left\|v^T(B^T)^{-1}\right\|_2 + \varepsilon \left\|Z^T(B^T)^{-1}\right\|_2.$$

As in the previous cases, disregarding an event of a small probability, we can assume that $\|Z\|_2 \le K_0\sqrt{n}$. Then by the assumption on $s_n(B)$,

$$\varepsilon \left\|Z^T(B^T)^{-1}\right\|_2 \le \frac{\varepsilon K_0\sqrt{n}}{s_n(B)} \le \frac{K_0\sqrt{n}}{\lambda_1}.$$

The parameters K_0, λ_1, and λ_2 can be chosen so that $\frac{K_0\sqrt{n}}{\lambda_1} \le \lambda_2^{-1}/2$. Then, since $\left\|Y^T(B^T)^{-1}\right\|_2 \ge \lambda_2^{-1}$, we conclude that

$$\left\|Y^T(B^T)^{-1}\right\|_2 \le 2\left\|v^T(B^T)^{-1}\right\|_2$$

and

$$\inf_{x\in S_1} \|Ax\|_2 \ge \frac{1}{4\sqrt{n}\,\|v^T(B^T)^{-1}\|_2} \cdot |A_{11} - Y^T(B^T)^{-1}X|.$$

The denominator here is independent of our random parameters.

Now we pass to the analysis of the numerator. From (9.5) follows that $A_{11} - Y^T(B^T)^{-1}X = \alpha r + \beta$ is a linear function of r with coefficients α and β, which depend on other random parameters. This representation would allow us to filter out several complicated terms in $A_{11} - Y^T(B^T)^{-1}X$ by using the global perturbation r.

Let $\lambda_3 > 0$ be a "small" parameter: $\lambda_3 = t^{a_3}$. Condition on everything, except r. Since r is uniformly distributed over the unit circle in $\mathbb{C}$, an easy density calculation yields

$$\mathbb{P}_r(|\alpha r + b| \ge \lambda_3|\alpha|) \ge 1 - C\lambda_3. \tag{9.6}$$

Taking the expectation with respect to the other random variables shows that the same bound holds unconditionally. Thus, disregarding the event of a small probability $C\lambda_3$, we obtain that $|A_{11} - Y^T(B^T)^{-1}X| \ge \lambda_3|\alpha|$. The coefficient α in turn can be represented as follows: $\alpha = \alpha' - \varepsilon v^T(B^T)^{-1}Z$, where $a' \in \mathbb{C}$ is independent of Z. Incorporating this into the bound above, we obtain

$$\inf_{x\in S_1} \|Ax\|_2 \ge \frac{\lambda_3}{4\sqrt{n}\,\|v^T(B^T)^{-1}\|_2}|\alpha' - \varepsilon v^T(B^T)^{-1}Z|.$$

Using the global perturbation allowed us to simplify the numerator and expose its dependence on the local perturbation variable Z. We will finish the proof using this local perturbation.

Set $h^T = v^T(B^T)^{-1} / \left\|v^T(B^T)^{-1}\right\|_2$ and recall that $h \in \mathbb{C}^{n-1}$ is independent of Z. Conditioning on everything except Z, we see that

$$g := \frac{\alpha'}{\|v^T(B^T)^{-1}\|_2} - \varepsilon h^T Z = \text{const} + \varepsilon\gamma',$$

where γ' is a complex normal random variable of unit variance: $\mathbb{E}|\gamma'|^2 = 1$. Hence, as before, for any $\mu > 0$

$$\mathbb{P}_Z(|g| \le \mu) \le C\mu/\varepsilon,$$

and integrating over other random variables, we conclude that the same estimate holds unconditionally. Combining this inequality with the previous one and recalling that we dropped an event of probability $C\lambda_3$ while using (9.6), we obtain

$$\mathbb{P}(\inf_{x\in S_1} \|Ax\|_2 \le t) \le \mathbb{P}\left(|g| \le \frac{4\sqrt{n}}{\lambda_3}t\right) + C\lambda_3 \le C\frac{4\sqrt{n}}{\lambda_3\varepsilon}t + C\lambda_3 \le C'\sqrt{n}t^{c'}$$

for some $c' > 0$. Choosing appropriate constants a and a_3 in $\varepsilon = t^a$ and $\lambda_3 = t^{a_3}$ finishes the proof in this case and completes the proof of Theorem 9.1.

Choice of the parameters. The analysis of this case requires the following two constraints:

$$\frac{K_0\sqrt{n}}{\lambda_1} \le \frac{\lambda_2^{-1}}{2} \quad \text{and} \quad \frac{t}{\lambda_3\varepsilon} + \lambda_3 \le t^{c'},\ c > 0.$$

The first one is satisfied with the choice $K_0 = t^{-0.05},\ \lambda_1 = t^{0.1}, \lambda_2 = t^{0.2}$ that we made above. To satisfy the second one, set $\lambda_3 = t^{0.2}$. □

We made no effort to optimize the dependence on t and n in the proof above. It would be interesting to find the optimal bound here. Another interesting question, suggested by Djalil Chafai, is to analyze the behavior of the smallest singular value of the matrix $D + U$ where U is uniformly distributed over a discrete subgroup of the unitary group. The case of the permutation group may be of special interest, because of its relevance for random graph theory. This question may require a combination of tools from Sections 5–9, since both obstacles, the arithmetic structure and the lack of independence, make an appearance here.

References

[1] G. W. Anderson, A. Guionnet, O. Zeitouni, An introduction to random matrices. Cambridge Studies in Advanced Mathematics, 118. Cambridge University Press, Cambridge, 2010. xiv+492 pp.

[2] 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.

[3] 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.

[4] J. Bourgain, V. Vu, P. Wood, *On the singularity probability of discrete random matrices*, J. Funct. Anal. 258 (2010), no. 2, 559–603.

[5] K. R. Davidson, S. J. Szarek, Local operator theory, random matrices and Banach spaces. Handbook of the geometry of Banach spaces, Vol. I, 317366, North-Holland, Amsterdam, 2001.

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

[7] A. Yu. Garnaev, E. D. Gluskin, The widths of a Euclidean ball (Russian), Dokl. Akad. Nauk SSSR 277 (1984), 1048–1052. English translation: Soviet Math. Dokl. 30 (1984), 200–204.

[8] A. A. Giannopoulos, V. D. Milman, *Concentration property on probability spaces*, Adv. Math. 156 (2000), no. 1, 77–106.

[9] F. Götze, A. Tikhomirov, *The circular law for random matrices*, Ann. Probab. 38 (2010), 1444–1491.

[10] O. Guedon, M. Rudelson, L_p *moments of random vectors via majorizing measures*, Adv. Math. 208 (2007), no. 2, 798–823.

[11] A. Guionnet, M. Krishnapur, O. Zeitouni, *The single ring theorem*, Ann. of Math. (2) 174 (2011),1189–1217.

[12] B. Kashin, The widths of certain finite-dimensional sets and classes of smooth functions, (Russian), Izv. Akad. Nauk SSSR Ser. Mat., 41 (1977), 334–351.

[13] G. Halász, *On the distribution of additive arithmetic functions*, Acta Arith. 27 (1975), 143–152

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

[15] J. Kahn, J. Komlós, E. Szemerédi, *On the probability that a random* ± 1*-matrix is singular*, J. Amer. Math. Soc. 8 (1995), no. 1, 223–240

[16] J. Komlós, *On the determinant of* $(0, 1)$ *matrices*, Studia Sci. Math. Hungar. 2 (1967), 7–21

[17] M. Ledoux and M. Talagrand, *Probability in Banach spaces*, Springer, 1991.

[18] A. Litvak, A. Pajor, M. Rudelson, N. Tomczak-Jaegermann, R. Vershynin, *Euclidean embeddings in spaces of finite volume ratio via random matrices*, J. Reine Angew. Math. 589 (2005), 1–19.

[19] H.H. Nguyen, *Inverse Littlewood-Offord problems and The Singularity of Random Symmetric Matrices*, arXiv:1101.3074.

[20] H. H. Nguyen, V. Vu, *Random matrices: law of the determinant*, to appear in Annal of Probability.

[21] H. H. Nguyen, V. Vu, *Optimal inverse Littlewood-Offord theorems*, Adv. Math. 226 (2011), no. 6, 52985319.

[22] M. Rudelson, *Lower estimates for the singular values of random matrices*, Compt. Rendus Math. de L'Academie des Sciences 342 (2006), no. 4, 247–252.

[23] M. Rudelson, *Invertibility of random matrices: norm of the inverse*, Ann. of Math. (2) 168 (2008), no. 2, 575–600.
[24] M. Rudelson, R. Vershynin, *The Littlewood-Offord Problem and invertibility of random matrices*, Adv. Math. 218 (2008), no. 2, 600–633.
[25] M. Rudelson, R. Vershynin, *The smallest singular value of a random rectangular matrix* , Comm. Pure Appl. Math. 62 (2009), no. 12, 1707–1739.
[26] M. Rudelson, R. Vershynin, *The least singular value of a random square matrix is O(n-1/2)*, C. R. Math. Acad. Sci. Paris 346 (2008), no. 15–16, 893–896.
[27] M. Rudelson, R. Vershynin, *Non-asymptotic theory of random matrices: extreme singular values*. Proceedings of the International Congress of Mathematicians. Volume III, 1576–1602, Hindustan Book Agency, New Delhi, 2010.
[28] M. Rudelson, R. Vershynin, *Invertibility of random matrices: unitary and orthogonal permutations*, arXiv:1206.5180.
[29] S. Smale, *On the efficiency of algorithms of analysis*, Bull. Amer. Math. Soc. (N.S.) 13 (1985), 87–121
[30] D. Spielman, S.-H. Teng, *Smoothed analysis of algorithms*. Proceedings of the International Congress of Mathematicians, Vol. I (Beijing, 2002), 597–606, Higher Ed. Press, Beijing, 2002
[31] S. Szarek, *Condition numbers of random matrices*, J. Complexity 7 (1991), no. 2, 131–149.
[32] M. Talagrand, *Majorizing measures: the generic chaining*, Ann. Probab. 24 (1996), no. 3, 1049–1103.
[33] T. Tao, V. Vu, *On random ± 1 matrices: singularity and determinant*, Random Structures and Algorithms 28 (2006), 1–23.
[34] T. Tao, V. Vu, Additive combinatorics. Cambridge Studies in Advanced Mathematics, 105. Cambridge University Press, Cambridge, 2006.
[35] T. Tao, V. Vu, *On the singularity probability of random Bernoulli matrices*, J. Amer. Math. Soc. 20 (2007), no. 3, 603–628.
[36] T. Tao, V. Vu, *Inverse Littlewood-Offord theorems and the condition number of random discrete matrices*, Ann. of Math. (2) 169 (2009), no. 2, 595–632.
[37] T. Tao, V. Vu, *Random matrices: the circular law*, Commun. Contemp. Math. 10 (2008), no. 2, 261–307.
[38] T. Tao, V. Vu, *Random matrices: universality of ESDs and the circular law. With an appendix by Manjunath Krishnapur*, Ann. Probab. 38 (2010), no. 5, 2023-2065.
[39] T. Tao, V. Vu, *A central limit theorem for the determinant of a Wigner matrix*, Adv. Math. 231 (2012), no. 1, 74–101.
[40] R. Vershynin, *Introduction to the non-asymptotic analysis of random matrices.* In: Compressed Sensing, Theory and Applications, ed. Y. Eldar and G. Kutyniok. Cambridge University Press, 2012. pp. 210–268.
[41] R. Vershynin, *Spectral norm of products of random and deterministic matrices*, Probab. Theory Related Fields 150 (2011), no. 3–4, 471–509.
[42] R. Vershynin, *Invertibility of symmetric random matrices*, arXiv:1102.0300v4, Random Structures and Algorithms, to appear.
[43] J. von Neumann, *Collected works. Vol. V: Design of computers, theory of automata and numerical analysis*. General editor: A. H. Taub. A Pergamon Press Book The Macmillan Co., New York 1963

Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA
E-mail address: rudelson@umich.edu