SELECTED TOPICS FROM THE GEOMETRIC THEORY OF RANDOM MATRICES

NICHOLAS COOK

Abstract. These notes provide an introduction to the geometric approach to questions in random matrix theory, focusing on three illustrative results for square random matrices with i.i.d. entries: an upper bound on the operator norm, an upper bound for the norm of the inverse, and universality of the distribution of the smallest singular value. The notes were prepared for three guest lectures I gave at Stanford in Spring quarter of 2017 as part of Amir Dembo's graduate-level course on Large Deviations and Random Matrices. The intended audience is graduate students who have had a course in probability, though the arguments are elementary.

Contents

1. Introduction and overview

1.1. **The geometric approach to random matrix theory.** The geometric theory of random matrices draws on techniques originating from the local theory of Banach spaces; for further background we refer to [Ver18]. An attractive feature is that one often get quantitative bounds at finite n that are within a constant factor of the asymptotic truth, with arguments that are more flexible. For instance, as we shall see below, for X an $n \times n$ matrix with independent uniformly sub-Gaussian entries one can show the operator norm $\|X\|$ is of size $O(\sqrt{n})$ with high probability, using a simple net argument and concentration; for the specific case of i.i.d. entries one can obtain the asymptotic $\sim 2\sqrt{n}$ by a much longer argument using the trace method. Sometimes the geometric methods are even capable of capturing exact asymptotics: for instance, for an i.i.d. matrix with real Gaussian

Date: 29 April 2017, Updated 1 August, 2020.

entries one can show $\mathbb{E}\|X\| \le 2\sqrt{n}$ using Slepian's inequality! – cf. [Ver18]. And, as we'll see in Section 5, they have even been applied to obtain universality for certain "local" statistics, by a very different approach from the popular heat flow [EY17] or Lindeberg [Tao19] methods.

1.2. **Singular values of non-Hermitian matrices.** Let M be an $n \times n$ matrix with complex entries (for simplicity we will mostly work with random matrices having real-valued entries). The ith singular value of M is defined $s_i(M) = \sqrt{\lambda_i(M^*M)}$, and we order them from largest to smallest:

$$s_1(M) \ge \cdots \ge s_n(M) \ge 0.$$

From the Courant–Fischer minimax formula for eigenvalues of Hermitian matrices,

$$\lambda_i(H) = \sup_{W \subset \mathbb{C}^n : \dim W = i} \inf_{u \in W} \langle u, Hu\rangle$$

we have

$$s_1(M) = \Big(\sup_{u \in S^{n-1}} \langle u, M^*Mu\rangle \Big)^{1/2} = \sup_{u \in S^{n-1}} \|Mu\| = \|M\|, \tag{1.1}$$

and

$$s_n(M) = \Big(\inf_{u \in S^{n-1}} \langle u, M^*Mu\rangle \Big)^{1/2} = \inf_{u \in S^{n-1}} \|Mu\|. \tag{1.2}$$

The last expression is zero if M is not invertible; otherwise, from substituting $u = M^{-1}v/\|M^{-1}v\|$ for $v \in S^{n-1}$ we have

$$s_n(M) = \inf_{v \in S^{n-1}} \frac{\|MM^{-1}v\|}{\|M^{-1}v\|} = \|M^{-1}\|^{-1}.$$

The smallest singular value thus provides a quantitative measure of how "well-invertible" M is.

1.3. **Non-Hermitian matrix model.** Throughout these notes we let $M = (\xi_{ij})$ denote an $n \times n$ matrix whose entries are i.i.d. copies of a centered random variable ξ of unit variance. In order to avoid some technicalities and get straight to the main ideas we will assume ξ is almost surely bounded. In fact, in some of the sections below we will specialize to the *i.i.d. sign matrix*, for which ξ is uniform on $\{-1, 1\}$. We will occasionally comment on how these assumptions can be relaxed.

The *quarter-circular law* gives the asymptotics for the empirical distribution of singular values for M. Specifically, it states that for any fixed $0 \le a < b < \infty$,

$$\frac{1}{n}\left| \left\{ i \in [n] : \frac{1}{\sqrt{n}} s_i(M) \in [a, b] \right\} \right| \to \frac{1}{\pi} \int_a^b \sqrt{4 - x^2}\, dx \quad a.s. \tag{1.3}$$

Assuming the rescaled singular values $\frac{1}{\sqrt{n}} s_i(M)$ are roughly evenly spaced within the limiting support $[0, 2]$, the quarter-circular law thus *suggests* that we should have

$$s_1(M) \asymp \sqrt{n}, \qquad s_n(M) \asymp \frac{1}{\sqrt{n}} \tag{1.4}$$

with high probability. The quarter-circular law does not imply such estimates however.[1] For instance, (1.3) allows for a proportion $o(n)$ of the rescaled singular values $\frac{1}{\sqrt{n}}s_i(M)$ diverge to $+\infty$.

The predictions (1.4) turn out to be correct, and we will prove versions of these estimates below. Specifically,

- In Section 2 we prove an upper tail estimate for the largest singular value, which implies that $s_1(M) = O(\sqrt{n})$ with probability $1 - O(e^{-n})$.
- In Section 3 we prove a lower tail estimate for the smallest singular value of the form $\mathbb{P}(s_n(M) \le t/\sqrt{n}) \lesssim t + o(1)$. In particular, $s_n(M) \ge \varepsilon/\sqrt{n}$ with probability $1 - O(\varepsilon)$ for any fixed $\varepsilon > 0$. Our treatment essentially follows an argument from [RV08].
- In Section 5 we sketch parts of the proof from [TV10a] that $\sqrt{n}s_n(M)$ in fact converges in distribution to a random variable with a density.

We will mainly use tools from high dimensional geometry and probability, such as concentration of measure, small ball estimates, and metric entropy bounds. These tools often give correct bounds up to constant factors. Sometimes the correct constants can be obtained by other means with significantly more effort. For instance, it is known [BS98] that $\frac{1}{\sqrt{n}}s_1(M) \to 2$ in probability,[2] but the proof of this result, which is mostly combinatorial, is much more challenging than the proof of Proposition 2.7 below.

1.4. **Notation.** We denote the (Euclidean) unit sphere in $\mathbb{R}^n$ by S^{n-1} and the unit ball by B^n. $\|\cdot\|$ denotes the ℓ^2 norm when applied to vectors and the $\ell^2 \to \ell^2$ operator norm when applied to matrices. We denote other ℓ^p norms on $\mathbb{R}^n$ by subscripts: $\|x\|_p := (\sum_{i=1}^n |x_i|^p)^{(1/p)}$, $\|x\|_\infty := \max_{i\in[n]} |x_i|$. We occasionally write $\mathrm{dist}(x, y) := \|x - y\|$.

C, c, c_0, c', etc. denote absolute constants whose value may change from line to line. We use the standard asymptotic notation $f = O(g)$, $f \lesssim g$, $g \gtrsim f$, all of which mean that $|f| \le Cg$ for some absolute constant $C < \infty$. $f \asymp g$ means $f \lesssim g \lesssim f$. We indicate dependence of the implied constant C on a parameter α with subscripts, *e.g.* $f \lesssim_\alpha g$. We will occasionally write $f = o(g)$, which means $f/g \to 0$ as $n \to \infty$, where n is always the dimension of the random matrix M. For limits with respect to other parameters we write, *e.g.*, $f = o_{\varepsilon\to 0}(g)$. The use of little-$o$ notation will be purely informal; all of the arguments below can be made effective.

---

[1] It does imply the one-sided estimate $s_1(M) \ge (2 - \varepsilon)\sqrt{n}$ a.s. for any fixed $\varepsilon > 0$

[2] In fact, under our assumption that the entries are almost surely bounded we have almost sure convergence. An interleaving argument can be used to show that if the entries only have finite fourth moment, but we additionally couple the sequence of matrices by taking M_n to be the top $n \times n$ corner of an infinite array of i.i.d. variables, then we retain almost sure convergence – see [Tao12].

2. Upper tail for the largest singular value

From (1.1) we can express the upper tail event for the largest singular value as a union of simpler events:

$$\mathbb{P}(\|M\| \geq \lambda) = \mathbb{P}\Big(\bigcup_{u \in S^{n-1}} \big\{ \|Mu\| \geq \lambda \big\} \Big). \tag{2.1}$$

Lemma 2.1 (Upper tail for image of fixed vector)**.** *Let $u \in S^{n-1}$. Then*

$$\mathbb{P}\big(\|Mu\| \geq K\sqrt{n}\big) \leq \exp\left(-cK^2 n\right)$$

for all $K \geq C$, where $C, c > 0$ are absolute constants.

We will use the following standard scalar concentration inequality:

Lemma 2.2 (Azuma–Hoeffding inequality)**.** *Let $\zeta_1, \dots, \zeta_n$ be independent centered random variables with $\|\zeta_j\|_{L^\infty} \leq a_j$ for each $1 \leq j \leq n$. Then for any $\lambda \geq 0$,*

$$\mathbb{P}\left(\left| \sum_{j=1}^n \zeta_j \right| \geq \lambda \right) \leq 2 \exp\left(- \frac{c\lambda^2}{\sum_{j=1}^n a_j^2} \right)$$

for some absolute constant $c > 0$.

Proof of Lemma 2.1. Let $R_1, \dots, R_n$ denote the rows of M. For each $i \in [n]$, from Lemma 2.2 we have that for all $\lambda \geq 0$,

$$\mathbb{P}(|R_i \cdot u| \geq \lambda) \leq 2\exp\left(-\frac{c\lambda^2}{\sum_{j=1}^n |u_j|^2} \right) = 2\exp\left(-c\lambda^2\right).$$

(Here we used our assumption that the entries of M are uniformly bounded.) Next we convert this tail bound to an exponential moment bound: for $\alpha > 0$,

$$\begin{aligned} \mathbb{E}\, e^{\alpha |R_i \cdot u|^2} &= \int_0^\infty \mathbb{P}\Big(e^{\alpha |R_i \cdot u|^2} \geq u \Big) du \\ &= \int_0^\infty 2\alpha\lambda e^{\alpha\lambda^2} \mathbb{P}(|R_i \cdot u| \geq \lambda) d\lambda \\ &\leq 4\alpha \int_0^\infty \lambda \exp\left((\alpha - c)\lambda^2 \right) d\lambda \\ &\leq C \end{aligned}$$

where we took $\alpha = c/2$, and $C > 0$ is some constant. By independence of the rows of M we obtain

$$\mathbb{E}\, e^{c\|Mu\|^2/2} \leq C^n.$$

Now the claim follows from Markov's inequality:

$$\mathbb{P}\big(\|Mu\| \geq K\sqrt{n}\big) \leq C^n \exp\left(-\frac{c}{2} K^2 n \right) \leq \exp\left(-\frac{c}{4} K^2 n \right)$$

if K is at least a sufficiently large constant. ∎

We will approximate the supremum of $u \mapsto \|Mu\|$ over the sphere S^{n-1} by the maximum over a finite ε-*net*:

Definition 2.3 (ε-net)**.** Let $S \subset \mathbb{R}^n$ and $\varepsilon > 0$. An ε-*net for* S is a finite subset $\mathcal{N}_\varepsilon \subset S$ such that for all $x \in S$ there exists $y \in \mathcal{N}_\varepsilon$ with $\|x - y\| \leq \varepsilon$.

Remark 2.4. For $\varepsilon > 0$ let $N_\varepsilon(S)$ be the minimal cardinality of an ε-net for S. The numbers $\{N_\varepsilon(S) : \varepsilon > 0\}$ are called the *covering numbers* for S, and $\log N_\varepsilon(S)$ is often called the *metric entropy* of S (at scale ε). Note that $\log N_\varepsilon(S)$ is the number of bits one would need in order to store an "address" for each point in S that is accurate up to an additive error ε.

Lemma 2.5 (Metric entropy of the sphere)**.** *For any $\varepsilon \in (0, 1]$, there is an ε-net $\mathcal{N}_\varepsilon$ for S^{n-1} of size $|\mathcal{N}_\varepsilon| \leq (3/\varepsilon)^n$.*

Proof. We take $\mathcal{N}_\varepsilon \subset S^{n-1}$ to be an ε-separated subset that is maximal with respect to the partial ordering of set-inclusion. We first claim that $\mathcal{N}_\varepsilon$ is an ε-net for S^{n-1}. Indeed, supposing on the contrary that there exists $y \in S^{n-1}$ with $\|x - y\| > \varepsilon$ for all $x \in \mathcal{N}_\varepsilon$, we have that $\mathcal{N}_\varepsilon \cup \{y\}$ is ε-separated in S^{n-1}, which contradicts maximality.

Now to bound the cardinality of $\mathcal{N}_\varepsilon$ we use a volumetric argument. Let $E = \mathcal{N}_\varepsilon + (\varepsilon/2) \cdot B^n$ be the $\varepsilon/2$-neighborhood of $\mathcal{N}_\varepsilon$ in $\mathbb{R}^n$, where we use the notation of set addition and dilation (recall also our notation B^n for the unit ball in $\mathbb{R}^n$). Since $\mathcal{N}_\varepsilon$ is an ε-separated set it follows that E is a disjoint union of Euclidean balls of radius $\varepsilon/2$. On the other hand, E is contained in $(1 + \varepsilon/2) \cdot B^n$, the origin-centered ball of radius $1 + \varepsilon/2$. It follows that

$$|\mathcal{N}_\varepsilon|(\varepsilon/2)^n \operatorname{vol}(B^n) \leq \operatorname{vol}(E) \leq \left(1 + \frac{\varepsilon}{2}\right)^n \operatorname{vol}(B^n)$$

which rearranges to

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

as desired. □

Remark 2.6. A variant of this volumetric argument gives an easy lower bound of 2^{-n} for the maximum density of a sphere packing in $\mathbb{R}^n$ (that is, a union of disjoint balls of radius 1). Despite the efforts of many authors, the best lower bound remains of the form $2^{-n+o(n)}$. (The current record, due to Akshay Venkatesh, is $cn \log\log n \cdot 2^{-n}$ for a certain sequence of dimensions n [Ven13].)

Proposition 2.7 (Upper tail for the largest singular value)**.** *There are constants $C, c > 0$ such that*

$$\mathbb{P}\big(\|M\| \geq K\sqrt{n}\big) \leq \exp\left(-cK^2 n\right)$$

for all $K \geq C$.

Remark 2.8. The results of [BS98] together with Talagrand's concentration inequality can be used to show that in fact

$$\mathbb{P}\big(\|M\| \geq (2 + \varepsilon)\sqrt{n}\big) \leq C \exp\left(-c\varepsilon^2 n\right).$$

Proof. Let $\mathcal{N}$ be a 1/2-net for S^{n-1}. By Lemma 2.5 we may take $|\mathcal{N}| \le 6^n$. Suppose the event $\{\|M\| \ge K\sqrt{n}\}$ holds. Let $u \in S^{n-1}$ such that $\|Mu\| = \|M\|$. There exists $v \in \mathcal{N}$ such that $\|u - v\| \le 1/2$, so

$$\|Mv\| \ge \|Mu\| - \|M(v-u)\| \ge \|M\| - \frac{1}{2}\|M\| \ge \frac{1}{2}K\sqrt{n}.$$

Thus,

$$\mathbb{P}\big(\|M\| \ge K\sqrt{n}\big) \le \mathbb{P}\bigg(\exists v \in \mathcal{N} : \|Mv\| \ge \frac{K}{2}\sqrt{n}\bigg).$$

Applying the union bound and Lemma 2.1,

$$\mathbb{P}\big(\|M\| \ge K\sqrt{n}\big) \le 6^n \exp\big(-c'K^2 n\big) \le \exp\big(-c''K^2 n\big)$$

if K is at least a sufficiently large constant. □

3. Lower tail for the smallest singular value

Now we turn to the problem of lower-bounding the smallest singular value $s_n(M)$. Throughout this section we take M to be the i.i.d. sign matrix (i.e. the ξ_{ij} are uniform in $\{-1, 1\}$). (As it turns out, this is in some sense the most challenging choice of distribution for the entries.)

The smallest singular value turns out to be much more challenging than the largest singular value. Indeed, even showing $s_n(M) \ne 0$ with high probability, *i.e.* that M is non-singular with probability tending to 1, was a non-trivial result of Komlós [Kom67].

The goal of this section is to establish the following:

Theorem 3.1 (Lower tail for the smallest singular value)**.** *For any* $t \ge 0$*,*

$$\mathbb{P}\bigg(s_n(M) \le \frac{t}{\sqrt{n}}\bigg) \lesssim t + \frac{1}{\sqrt{n}}. \tag{3.1}$$

This is a weak version of a result of Rudelson and Vershynin from [RV08] giving the bound $O(t + e^{-cn})$ assuming ξ is *sub-Gaussian* (which holds for the Bernoulli distribution), building on earlier work of Rudelson [Rud08] and Tao–Vu [TV09].

Remark 3.2 (Singularity probability for Bernoulli matrices). Note that (3.1) gives a bound $O(n^{-1/2})$ for the probability that M is singular (i.e. that $s_n(M) = 0$). This recovers a classic result of Komlós for the i.i.d. sign matrix [Kom67, Kom]. The result of [RV08] gives a bound $O(e^{-cn})$ for the singularity probability, which recovers a result of Kahn, Komlós and Szemerédi [KKS95]. An easy lower bound for $\mathbb{P}(\det(M) = 0)$ is given by the probability that the first two rows are equal, which is 2^{-n}. It is a famous conjecture that this is asymptotically the correct base, i.e. $\mathbb{P}(\det(M) = 0) = (\frac{1}{2} + o(1))^n$. Following a breakthrough approach of Tao and Vu using tools from additive combinatorics [TV07], the current record now stands at $\mathbb{P}(\det(M) = 0) \le (\frac{1}{\sqrt{2}} + o(1))^n$, due to Bourgain, Vu and Wood [BVW10].

3.1. **Partitioning the sphere.** We now begin the proof of Theorem 3.1. Recall from (1.2) that

$$s_n(M) = \inf_{u\in S^{n-1}} \|Mu\|.$$

The key will be to split the sphere into two parts and control the infimum over each part by different arguments.

Definition 3.3. We denote the *support* of a vector $x \in \mathbb{R}^n$ by

$$\operatorname{supp}(x) := \{j \in [n] : x_j \neq 0\}. \tag{3.2}$$

For $0 < \theta \leq n$ the set of θ-sparse vectors is defined

$$\operatorname{Sparse}(\theta) := \{x \in \mathbb{R}^n : |\operatorname{supp}(x)| \leq \theta\}. \tag{3.3}$$

For $\delta, \varepsilon \in (0,1)$, define the set of (δ,ε)-*compressible* unit vectors to be the ε-neighborhood in S^{n-1} of the δn-sparse vectors:

$$\operatorname{Comp}(\delta,\varepsilon) := S^{n-1} \cap (\operatorname{Sparse}(\delta n) + \varepsilon \cdot B^n) \tag{3.4}$$

and the complementary set of (δ,ε)-*incompressible* unit vectors

$$\operatorname{Incomp}(\delta,\varepsilon) := S^{n-1} \setminus \operatorname{Comp}(\delta,\varepsilon). \tag{3.5}$$

In the remainder of the proof we will show, roughly speaking, that for sufficiently small fixed choices of δ, ε,

$$\inf_{u\in \mathrm{Comp}} \|Mu\| \gtrsim \sqrt{n} \quad \text{and} \quad \inf_{u \in \mathrm{Incomp}} \|Mu\| \gtrsim \frac{1}{\sqrt{n}}$$

with high probability. Specifically, we will prove the following two propositions:

Proposition 3.4 (Invertibility over compressible vectors)**.** *There are sufficiently small constants $\delta, \varepsilon, c_1, c_2 > 0$ such that*

$$\mathbb{P}\left(\inf_{u\in \mathrm{Comp}(\delta,\varepsilon)} \|Mu\| \leq c_1\sqrt{n}\right) \leq e^{-c_2 n}. \tag{3.6}$$

Proposition 3.5 (Invertibility over incompressible vectors)**.** *For $\delta, \varepsilon > 0$ sufficiently small and for all $t \geq 0$,*

$$\mathbb{P}\left(\inf_{u\in \mathrm{Incomp}(\delta,\varepsilon)} \|Mu\| \leq \frac{t}{\sqrt{n}}\right) \lesssim_{\delta,\varepsilon} t + \frac{1}{\sqrt{n}}. \tag{3.7}$$

Now to conclude the proof of Theorem 3.1, let $\delta, \varepsilon, c_1, c_2 > 0$ be small constants such that (3.6) and (3.7) holds. Let $t \geq 0$. We may assume $t \leq 1$. By the union bound,

$$\begin{aligned}\mathbb{P}\Big(s_n(M) \le \frac{t}{\sqrt{n}}\Big) &= \mathbb{P}\Big(\inf_{u\in S^{n-1}} \|Mu\| \le \frac{t}{\sqrt{n}}\Big) \\ &\le \mathbb{P}\Big(\inf_{u\in \mathrm{Comp}(\delta,\varepsilon)} \|Mu\| \le \frac{t}{\sqrt{n}}\Big) + \mathbb{P}\Big(\inf_{u\in \mathrm{Incomp}(\delta,\varepsilon)} \|Mu\| \le \frac{t}{\sqrt{n}}\Big) \\ &\lesssim t + \frac{1}{\sqrt{n}} + e^{-c_2 n} \\ &\lesssim t + \frac{1}{\sqrt{n}}\end{aligned}$$

where in the third line we used that $t/\sqrt{n} \le 1/\sqrt{n} \le c_1\sqrt{n}$ for n sufficiently large.

3.2. **Anti-concentration estimates.** A central ingredient of our proof of the upper tail bound for the operator norm in Proposition 2.7 was concentration of measure, in the form of the Azuma–Hoeffding inequality (Lemma 2.2). It turns out that to prove lower bounds on the smallest singular value of the square i.i.d. sign matrix M we will need to apply the dual notion of an *anti-concentration* (or “small ball”) estimate. For a scalar random variable Z we define the *Lévy concentration function*

$$\mathcal{L}(Z,r) := \sup_{a\in\mathbb{R}} \mathbb{P}(|Z-a| \le r). \tag{3.8}$$

We now state two anti-concentration estimates for random walks $X\cdot v$. The proofs are deferred to Section 4. The first, Lemma 3.6 below, is a rough form of the Berry–Esséen theorem, which gives a quantitative comparison between the distribution of $X\cdot v$ and a Gaussian. (In fact, Lemma 3.6 readily follows from the Berry–Esséen theorem, though this is in some sense cheating as the Berry–Esséen theorem itself is proved by a refined version of the arguments of this section. Furthermore, Corollary 4.2 opens the way to much stronger anti-concentration bounds that will be discussed in Section 4.2, and which cannot be deduced from the Berry–Esséen theorem.)

Lemma 3.6 (Berry–Esséen-type anti-concentration for random walks)**.** *Let $v \in \mathbb{R}^n$ and let $X = (\xi_1,\dots,\xi_n) \in \{-1,1\}^n$ be a vector of i.i.d. uniform signs. Then for all $t \ge 0$,*

$$\mathcal{L}(X\cdot v, r) \lesssim \frac{r + \|v\|_\infty}{\|v\|}. \tag{3.9}$$

Lemma 3.6 gives a bound of the form $\mathcal{L}(X\cdot v,r) \lesssim r + o(1)$ for unit vectors v with $\|v\|_\infty = o(1)$. However, this bound is trivial for unit vectors v having a sufficiently large constant amount of mass in one coordinate (such as a standard basis vector). For such vectors we will use the following crude bound.

Lemma 3.7 (Crude anti-concentration for random walks)**.** *Let $u \in S^{n-1}$ and let $X = (\xi_1,\dots,\xi_n) \in \{-1,1\}^n$ be a vector of i.i.d. uniform signs. There is a* *constant* $c_0 > 0$ *such that*

$$\mathcal{L}(X \cdot u, c_0) \le 1/2.$$

We prove these lemmas in Section 4.

3.3. **Compressible vectors.** In this subsection we prove Proposition 3.4. We follow a similar approach to the proof of Proposition 2.7. The following is the analogue of Lemma 2.1 in our setting.

Lemma 3.8 (Lower tail for image of fixed vector)**.** *There is a constant* $c > 0$ *such that for any fixed* $u \in S^{n-1}$,

$$\mathbb{P}\big(\|Mu\| \le c\sqrt{n}\big) \le e^{-cn}.$$

Proof. For $t \ge 0$ let

$$\mathcal{E}_t = \{\|Mu\| \le t\sqrt{n}\} = \left\{ \sum_{i=1}^{n} |R_i \cdot u|^2 \le t^2 n \right\}.$$

Let $\beta \in (0, 1/2]$ to be chosen later. On $\mathcal{E}_t$ we have that $|R_i \cdot u| \le t/\beta$ for at least $(1-\beta^2)n$ values of $i \in [n]$ (from Markov's inequality). Since the rows of M are exchangeable, we can spend a factor $\binom{n}{(1-\beta^2)n}$ to assume these are the first $(1-\beta^2)n$ rows:

$$\mathbb{P}(\mathcal{E}_t) \le \binom{n}{(1-\beta^2)n} \mathbb{P}\left(\bigwedge_{1 \le i \le (1-\beta^2)n} |R_i \cdot u| \le t/\beta \right).$$

By independence and row exchangeability,

$$\mathbb{P}(\mathcal{E}_t) \le \binom{n}{(1-\beta^2)n} \mathbb{P}(|R_1 \cdot u| \le t/\beta)^{(1-\beta^2)n}.$$

Taking $t = c_0\beta$, by Lemma 3.7,

$$\mathbb{P}(\mathcal{E}_t) \le \binom{n}{(1-\beta^2)n} (1-c_0)^{(1-\beta^2)n} = \binom{n}{\beta^2 n} (1-c_0)^{(1-\beta^2)n} \le (e/\beta^2)^{\beta^2 n} e^{-c_0 n/2}$$

where in the last inequality we used the bound $\binom{n}{k} \le (en/k)^k$. The claim now follows by taking β a sufficiently small constant (which also fixes $t = c_0/\beta$). ☐

Proof of Proposition 3.4. For $J \subset [n]$ let S^J denote the set of unit vectors supported on J. By Lemma 2.5, for each $J \in \binom{[n]}{\delta n}$ there exists $\mathcal{N}_J$ an ε-net of S^J of size $|\mathcal{N}_J| \le (3/\varepsilon)^{|J|}$. Put

$$\mathcal{N} := \bigcup_{J \in \binom{[n]}{\delta n}} \mathcal{N}_J.$$

For $K > 0$ write

$$\mathcal{B}_K = \{\|M\| \le K\sqrt{n}\}.$$

Claim 3.9. Let $K > 0$, $\delta, \varepsilon \in (0,1)$. Then

$$\mathcal{B}_K \cap \big\{\exists u \in \mathrm{Comp}(\delta, \varepsilon) : \|Mu\| \le \varepsilon K\sqrt{n}\big\} \subset \big\{\exists v \in \mathcal{N} : \|Mv\| \le 4\varepsilon K\sqrt{n}\big\}.$$

We first complete the proof of Proposition 3.4 on the above claim. From Proposition 2.7 we have that $\mathbb{P}(\mathcal{B}_K) \geq 1 - e^{-n}$ if K is a sufficiently large constant. Fixing such K, it suffices to show

$$\mathbb{P}\big(\mathcal{B}_K \cap \big\{\exists u \in \mathrm{Comp}(\delta,\varepsilon) : \|Mu\| \leq \varepsilon K\sqrt{n}\big\}\big) \leq e^{-cn}. \tag{3.10}$$

Let $\varepsilon, \delta > 0$ to be chosen as sufficiently small constants. From the above claim and the union bound,

$$\mathbb{P}\big(\mathcal{B}_K \cap \big\{\exists u \in \mathrm{Comp}(\delta,\varepsilon) : \|Mu\| \leq \varepsilon K\sqrt{n}\big\}\big) \leq \sum_{v \in \mathcal{N}} \mathbb{P}\big(\|Mv\| \leq 4\varepsilon K\sqrt{n}\big). \tag{3.11}$$

Now from Lemma 3.8 we have

$$\mathbb{P}\big(\|Mv\| \leq 4\varepsilon K\sqrt{n}\big) \leq e^{-cn}$$

taking $\varepsilon = c_1/K$ for a sufficiently small constant $c_1 > 0$. Thus, right hand side of (3.11) is bounded by

$$|\mathcal{N}|e^{-cn} \leq \binom{n}{\delta n}(3/\varepsilon)^n e^{-cn} \leq \exp\left(\delta n \log \frac{C}{\delta\varepsilon} - cn\right) \leq e^{-cn/2}$$

taking δ sufficiently small.

It remains to establish Claim 3.9. By definition, for any $u \in \mathrm{Comp}(\delta,\varepsilon)$ there exists $w \in \mathrm{Sparse}(\delta n)$ such that $\|u - w\| \leq \varepsilon$. In particular, $|1 - \|w\|| \leq \varepsilon$. Writing $w' = w/\|w\|$, it follows that $\|w - w'\| \leq \varepsilon$, and by the triangle inequality $\|u - w'\| \leq 2\varepsilon$. Let $J \in \binom{[n]}{\delta n}$ such that $\mathrm{supp}(w') \subset J$. There exists $v \in \mathcal{N}_J \subset \mathcal{N}$ such that $\|w' - v\| \leq \varepsilon$. By another application of the triangle inequality, $\|u - v\| \leq 3\varepsilon$.

Thus, if $\mathcal{B}_K$ holds and if $u \in \mathrm{Comp}(\delta,\varepsilon)$ is such that $\|Mu\| \leq \varepsilon K\sqrt{n}$, then letting $v \in \mathcal{N}$ such that $\|u - v\| \leq 3\varepsilon$ as above, we have

$$\|Mv\| \leq \|Mu\| + \|M(u - v)\| \leq \varepsilon K\sqrt{n} + 3\varepsilon\|M\| \leq 4\varepsilon K\sqrt{n}.$$

The claim follows. □

Remark 3.10. It is easy to check by very minor modifications of the above proof, the conclusion of Proposition 3.4 also holds with M replaced by $M^{(i)}$ for any $i \in [n]$, where $M^{(i)}$ denotes the $n-1 \times n$ matrix obtained by removing the ith row from M. We will use this fact in the following subsection.

3.4. **Incompressible vectors.** In this section we establish Proposition 3.5.

Lemma 3.11 (Incompressible vectors are spread)**.** *Let $u \in \mathrm{Incomp}(\delta,\varepsilon)$.*

(1) There exists $L_0 \subset [n]$ with $|L| \geq \delta n$ such that $|u_j| \geq \varepsilon/\sqrt{n}$ for all $j \in L_0$.
(2) There exists $L \subset [n]$ with $|L| \geq \delta n/2$ such that $|u_j| \in [\frac{\varepsilon}{\sqrt{n}}, \frac{2}{\sqrt{\delta n}}]$ for all $j \in L$.

Proof. For (1) we take L_0 to be the set of the largest δn coordinates of u. Then u_{L_0}, the projection of u to $\mathbb{R}^{L_0}$, is δn-sparse. If $|u_j| < \varepsilon/\sqrt{n}$ for some $j \in L_0$, then

$|u_j| < \varepsilon/\sqrt{n}$ for all $j \in [n] \setminus L_0$ and so $\|u - u_{L_0}\| < \varepsilon$. This implies u is within distance ε of a δn-sparse vector, a contradiction.

Now for (2), since $u \in S^{n-1}$, $|u_j| > 2/\sqrt{\delta n}$ for at most $\delta n/4$ values of $j \in [n]$ (by Markov's inequality). We can thus obtain the desired set L by removing at most $\delta n/4$ bad elements from L_0. □

A key challenge for establishing Proposition 3.5 is to deal with the fact that we are lower bounding an infimum over the uncountable set $\mathrm{Incomp}(\delta, \varepsilon)$. Whereas in Propositions 2.7 and 3.4 we were able to pass to a finite net and take a union bound, here we will not have sufficiently strong tail bounds to beat the cost of such a union bound. (Note that whereas the covering number of $\mathrm{Comp}(\delta, \varepsilon)$ was essentially $\exp(n\delta \log(1/\delta\varepsilon))$, for $\mathrm{Incomp}(\delta, \varepsilon)$ it is essentially $O(1/\varepsilon)^n$, the covering number of the sphere.)

Instead we will use an averaging argument to reduce to considering the event that a fixed column of M is close to the span of the remaining columns. This argument was first used to control the smallest singular value by Rudelson and Vershynin [RV08]; a similar idea was used by Komlós in the context of the invertibility problem for Bernoulli matrices [Kom].

First we give a naïve version of the argument. Denote the columns of M by $X_1, \dots, X_n$, and for $i \in [n]$ denote $V_i := \mathrm{span}\{X_j : j \neq i\}$. Suppose our aim is to control the event that

$$s_n(M) = \inf_{u \in S^{n-1}} \|Mu\| \le \frac{t}{\sqrt{n}}$$

by the event that a $\mathrm{dist}(X_i, V_i)$ is small for some i. Let $u \in S^{n-1}$ be arbitrary. By the pigeonhole principle there exists $i \in [n]$ such that $|u_i| \ge 1/\sqrt{n}$. Since $\|Mu\| \ge \left\|\mathrm{Proj}_{V_i^\perp}(Mu)\right\|$, by expanding Mu as $\sum_{j=1}^n u_j X_j$ we see that

$$\|Mu\| \ge \left\|\mathrm{Proj}_{V_i^\perp}\left(\sum_{j=1}^n u_j X_j\right)\right\| = |u_i| \left\|\mathrm{Proj}_{V_i^\perp}(X_i)\right\| \ge \frac{1}{\sqrt{n}} \mathrm{dist}(X_i, V_i).$$

Applying the union bound,

$$\begin{aligned} \mathbb{P}\left(\inf_{u \in S^{n-1}} \|Mu\| \le \frac{t}{\sqrt{n}}\right) &\le \sum_{i=1}^n \mathbb{P}(\mathrm{dist}(X_i, V_i) \le t) \\ &= n\mathbb{P}(\mathrm{dist}(X_1, V_1) \le t) \qquad (3.12) \end{aligned}$$

where in the second line we used that the columns of M are exchangeable. However, the best bound we will be able to achieve below for $\mathbb{P}(\mathrm{dist}(X_1, V_1) \le t)$ is of the form $\lesssim t + n^{-1/2}$, so the union bound over n events is too costly. (By using more advanced Littlewood–Offord bounds such as Theorem 4.3 it is possible to reduce the term $n^{-1/2}$ to size e^{-cn}, at which point (3.12) is acceptable after reducing t, but this leads to a bound of the form $s_n(M) \gtrsim n^{-3/2}$ with high probability, which is not the correct scale $n^{-1/2}$.)

The following improves on the above naïve argument by specializing to incompressible vectors and using Lemma 3.11.

Lemma 3.12. *Let $\delta, \varepsilon \in (0,1)$ and $t \geq 0$. Then*

$$\mathbb{P}\bigg(\inf_{u\in \mathrm{Incomp}(\delta,\varepsilon)} \|Mu\| \leq \frac{t}{\sqrt{n}}\bigg) \leq \frac{1}{\delta n}\sum_{i=1}^n \mathbb{P}\bigg(\mathrm{dist}(X_i, V_i) \leq \frac{t}{\varepsilon}\bigg). \tag{3.13}$$

Proof. For any $u \in S^{n-1}$, writing $Mu = \sum_{j=1}^n u_j X_j$, we see that for any $i \in [n]$,

$$\|Mu\| \geq |u_i| \|\mathrm{Proj}_{V_i^\perp}(X_i)\| = |u_i| \,\mathrm{dist}(X_i, V_i).$$

If $u \in \mathrm{Incomp}(\delta, \varepsilon)$, by the proof of Lemma 3.11 we have that $|u_i| \geq \varepsilon/\sqrt{n}$ for all $i \in L_0$ for some $L_0 \subset [n]$ of size at least δn. Thus,

$$\|Mu\| \geq \frac{\varepsilon}{\sqrt{n}} \mathrm{dist}(X_i, V_i)$$

for all $i \in L_0$. Denoting the event on the left hand side of (3.13) by $\mathcal{E}(t)$, we have that on $\mathcal{E}(t)$, $\{\mathrm{dist}(X_i, V_i) \leq t/\varepsilon\}$ holds for all $i \in L_0$. By double counting,

$$\sum_{i=1}^n \mathbb{P}\bigg(\mathrm{dist}(X_i, V_i) \leq \frac{t}{\varepsilon}\bigg) \geq \delta n \mathbb{P}(\mathcal{E}(t))$$

and the claim follows. □

Now we conclude the proof of Proposition 3.5. Let $\delta, \varepsilon > 0$ be constants to be taken sufficiently small, and let $t \geq 0$. For the remainder of the proof we allow implied constants to depend on δ, ε. By Lemma 3.12 (adjusting t by a constant factor) and the fact that the columns of M are exchangeable, it suffices to show

$$\mathbb{P}(\mathrm{dist}(X_1, V_1) \leq t) \lesssim t + \frac{1}{\sqrt{n}}. \tag{3.14}$$

Note that we may assume t is smaller than any fixed constant.

Let $\mathcal{G}$ denote the event that for all $u \in \mathrm{Comp}(\delta, \varepsilon)$, $u^{\mathsf{T}} M^{(1)} \neq 0$, where $M^{(1)}$ denotes the $n \times (n-1)$ matrix obtained by removing the first column from M. By Proposition 3.4 and Remark 3.10, $\mathcal{G}$ holds with probability $1 - O(e^{-cn})$ if $\delta, \varepsilon > 0$ are sufficiently small constants. Fixing such δ, ε, we may restrict to $\mathcal{G}$ for the remainder of the proof. Note that by independence of X_1 from the remaining columns this restriction does not affect the distribution of X_1.

We condition on the columns $X_2, \dots, X_n$ to fix the subspace V_1, and let $v \in V_1^\perp$ be a unit normal vector. In other words, $0 = v^* M^{(1)}$, so by our restriction to $\mathcal{G}$ it follows that $v \in \mathrm{Incomp}(\delta, \varepsilon)$. By Lemma 3.11 there exists $L \subset [n]$ with $|L| \geq \delta n/2$ and

$$\frac{\varepsilon}{\sqrt{n}} \leq |v_j| \leq \frac{2}{\sqrt{\delta n}} \qquad \forall j \in L.$$

Writing v_L for the projection of v to the coordinate subspace $\mathbb{R}^L$, we have

$$\|v_L\|^2 \geq \frac{\delta n}{2}\frac{\varepsilon^2}{n} = \frac{1}{2}\varepsilon^2 \delta, \qquad \|v_L\|_\infty \leq \frac{2}{\sqrt{\delta n}}.$$

Applying Lemma 3.6, we have that for any $r \geq 0$,

$$\mathcal{L}(X_1 \cdot v_L, r) = \sup_{a\in\mathbb{R}} \mathbb{P}(|X_1 \cdot v_L - a| \leq r) \lesssim \frac{r + \|v_L\|_\infty}{\|v_L\|} \lesssim r + \frac{1}{\sqrt{n}}.$$

Now since $|X_1 \cdot v| \leq \mathrm{dist}(X_1, V_1)$, by conditioning on the variables $\{\xi_{1j} : j \notin L\}$ we find

$$\mathbb{P}(\mathrm{dist}(X_1, V_1) \leq t) \leq \mathbb{P}(|X_1 \cdot v| \leq t) \lesssim t + \frac{1}{\sqrt{n}}$$

as desired. This completes the proof of Proposition 3.5, and hence of Theorem 3.1.

4. Anti-concentration for scalar random walks

4.1. **The Fourier-analytic approach.** In this section we prove Lemmas 3.6 and 3.7, which are estimates on the Lévy concentration function (defined in (3.8)) for scalar random variables of the form $X \cdot v$, with $X = (\xi_1, \dots, \xi_n)$ a vector of i.i.d. uniform Bernoulli signs and $v \in \mathbb{R}^n$ a fixed vector. In the literature $X \cdot v$ is often referred to as a *random walk* with *steps vector* v. The problem of estimating the concentration function $\mathcal{L}(X \cdot v, r)$ for general random vectors X with i.i.d. coordinates is a continuous version of the classical *Littlewood–Offord problem*, which in its original form was to estimate $\mathcal{L}(X \cdot v, 0)$ for the case that the ξ_j are uniform random signs. See [TV10b, Chapter 7] and the survey [NV13] for a more complete overview of the Littlewood–Offord theory and its applications in random matrix theory.

While we focus here on the case that the components of X are Bernoulli signs, everything in this section extends to more general classes of complex-valued random variables; see [TV10b, RV10, TV08, RV08].

The main tool in this game is Fourier analysis, via the following lemma. Here and in the sequel we abbreviate $e(x) := \exp(2\pi i x)$.

Lemma 4.1 (Esséen concentration inequality)**.** *Let Z be a real-valued random variable and let $s > 0$. Then*

$$\mathcal{L}(Z, s) \lesssim s \int_{|t|\leq 1/s} |\mathbb{E}\, e(tZ)| dt.$$

Proof. By replacing Z with Z/s and change of variables it suffices to show

$$\mathcal{L}(Z, 1) \lesssim \int_{-1}^{1} |\mathbb{E}(tZ)| dt.$$

Fixing $a \in \mathbb{R}$ arbitrarily, it suffices to show

$$\mathbb{P}(|Z - a| \leq 1) \lesssim \int_{-1}^{1} |\mathbb{E}(tZ)| dt \tag{4.1}$$

where the implied constant is independent of a.

Recall the Fourier transform of a function $f \in L_1(\mathbb{R})$ is defined

$$\hat{f}(t) := \int_{\mathbb{R}} f(x)\overline{e(xt)} dx.$$

We claim there is a non-negative function f such that

$$f(x) \geq c1_{[-1,1]}(x), \qquad 0 \leq \hat{f}(t) \leq 1_{[-1,1]}(t) \tag{4.2}$$

for some absolute constant $c > 0$, with the inequalities holding pointwise for $x, t \in \mathbb{R}$. Indeed, setting $g(t) := 1_{[-\frac{1}{4},\frac{1}{4}]}(t)$, one easily computes that

$$\hat{g}(x) = \frac{\sin(\pi x/2)}{\pi x}.$$

Then putting $h(t) := g * g(t)$, clearly $0 \leq h(t) \leq 1_{[-1,1]}(t)$, and

$$\check{h}(x) = |\hat{g}(x)|^2 \geq c1_{[-1,1]}(x)$$

for some constant $c > 0$. Thus, (4.2) hold with $f = \hat{h}$.

We can now bound the left hand side of (4.1) by

$$\mathbb{E}\, 1_{[-1,1]}(Z - a) \lesssim \mathbb{E}\, f(Z - a) = \mathbb{E} \int_{\mathbb{R}} e(at)\hat{f}(t)e(tZ)dt,$$

where in the last equality we used the Fourier inversion formula. Applying Fubini's theorem and taking the modulus of the integrand gives

$$\mathbb{P}(|Z - a| \leq 1) \lesssim \int_{\mathbb{R}} |\hat{f}(t)||\, \mathbb{E}(tZ)|dt \lesssim \int_{-1}^{1} |\, \mathbb{E}\, e(tZ)|dt$$

where in the final bound we applied (4.2). □

Corollary 4.2. *Let $v \in \mathbb{R}^n$ and let $X = (\xi_1, \dots, \xi_n) \in \{-1, 1\}^n$ be a vector of i.i.d. uniform signs. For all $r \geq 0$,*

$$\begin{aligned} \mathcal{L}(X \cdot v, r) &\lesssim \int_{-1}^{1} \exp\left(-c \sum_{j=1}^{n} \|2v_j t/r\|_{\mathbb{R}/\mathbb{Z}}^2\right) dt \\ &= \int_{-1}^{1} \exp\left(-c\, \mathrm{dist}^2\left(\frac{2t}{r} v, \mathbb{Z}^n\right)\right) dt. \end{aligned}$$

Here $\|x\|_{\mathbb{R}/\mathbb{Z}}$ denotes the distance of $x \in \mathbb{R}$ to the nearest integer.

Proof. Applying Lemma 4.1 (with $s = 1$),

$$\mathcal{L}(X \cdot v, r) = \mathcal{L}\left(\frac{1}{r} X \cdot v, 1\right) \lesssim \int_{-1}^{1} |\, \mathbb{E}\, e((t/r)X \cdot v)|dt.$$

By independence of the components of X,

$$\mathbb{E}\, e((t/r)X \cdot v) = \prod_{j=1}^{n} \mathbb{E}\, e(tv_j\xi_j/r) = \prod_{j=1}^{n} \frac{1}{2}(e(tv_j/r) - e(-tv_j/r)) = \prod_{j=1}^{n} \cos(2\pi v_j t/r).$$

We next apply the pointwise bound

$$|\cos(\pi x)| \leq \exp\left(-c\|x\|_{\mathbb{R}/\mathbb{Z}}^2\right) \tag{4.3}$$

valid for $x \in \mathbb{R}$ and a sufficiently small absolute constant $c > 0$. Combining the previous lines,

$$\mathcal{L}(X \cdot v, r) \lesssim \int_{-1}^{1} \prod_{j=1}^{n} |\cos(2\pi v_j t/r)| dt \leq \int_{-1}^{1} \exp\left(-c \sum_{j=1}^{n} \|2v_j t/r\|_{\mathbb{R}/\mathbb{Z}}^2\right) dt$$

as desired. □

We now use Corollary 4.2 to complete the proofs of Lemmas 3.6 and 3.7.

Proof of Lemma 3.6. By replacing v with $v/\|v\|$ and r with $r/\|v\|$ we may assume $\|v\| = 1$. Let $r \geq C\|v\|_\infty$ for some absolute constant $C > 0$ to be chosen sufficiently large. It suffices to show

$$\mathcal{L}(X \cdot v, r) \lesssim r. \tag{4.4}$$

From Corollary 4.2,

$$\mathcal{L}(X \cdot v, r) \lesssim \int_{-1}^{1} \exp\left(-c \sum_{j=1}^{n} \|2v_j t/r\|_{\mathbb{R}/\mathbb{Z}}^2\right) dt.$$

Now for $t \in [-1, 1]$, for each $j \in [n]$,

$$|2v_j t/r| \leq 2\|v\|_\infty / r \leq 2/C$$

by our assumption on r. Thus, taking $C > 4$ we have $\|2v_j t/r\|_{\mathbb{R}/\mathbb{Z}} = |2v_j t/r|$ for each $j \in [n]$, so

$$\mathcal{L}(X \cdot v, r) \lesssim \int_{-1}^{1} \exp\left(-c' t^2 \|v\|^2 / r^2\right) dt \leq \int_{\mathbb{R}} \exp\left(-c' t^2 / r^2\right) dt \lesssim r$$

where in the last bound we used that v is a unit vector and performed the Gaussian integral. This gives (4.4) and completes the proof. □

Proof of Lemma 3.7. Let $c_0 > 0$ to be chosen sufficiently small. We first address the case that $\|u\|_\infty > 2c_0$. Without loss of generality assume $|u_1| > 2c_0$. Then by conditioning on $\xi_2, \ldots, \xi_n$ we see

$$\mathcal{L}(X \cdot u, c_0) \leq \mathcal{L}(u_1 \xi_1, c_0) \leq 1/2$$

where we used that ξ_1 is uniformly distributed on the 2-separated set $\{-1, 1\}$.

Now assume $\|u\|_\infty \leq 2c_0$. By Corollary 4.2 and change of variable,

$$\mathcal{L}(X \cdot u, c_0) \lesssim c_0^{1/2} \int_{|t| \leq c_0^{-1/2}} \exp\left(-c \sum_{j=1}^{n} \|2c_0^{-1/2} u_j t\|_{\mathbb{R}/\mathbb{Z}}^2\right) dt.$$

Now by our assumption on u, for all $t \in [-1, 1]$ and $j \in [n]$ we have

$$|2c_0^{-1/2} u_j t| \leq 4c_0^{1/2}.$$

Thus, taking c_0 sufficiently small we have $\|2c_0^{-1/2}u_jt\|_{\mathbb{R}/\mathbb{Z}} = |2c_0^{-1/2}u_jt|$, so

$$\begin{aligned}\mathcal{L}(X\cdot u, c_0) &\lesssim c_0^{1/2}\int_{|t|\le c_0^{-1/2}} \exp\left(-c'\sum_{j=1}^n |c_0^{-1/2}u_jt|^2\right)dt\\ &\le c_0^{1/2}\int_{\mathbb{R}}\exp\left(-c't^2/c_0\right)dt\\ &\lesssim c_0.\end{aligned}$$

The claim now follows by taking c_0 sufficiently small. □

4.2. **Improved Littlewood–Offord bounds and arithmetic structure.** In this section we return to the problem discussed in Section 4 of estimating the concentration function $\mathcal{L}(X\cdot v, r)$ for X a random vector with i.i.d. coordinates and $v \in \mathbb{R}^n$ a fixed vector of "steps". We present a result of Rudelson and Vershynin from [RV08] (see also [RV10] and references therein) which gives improvements on the Berry–Esséen-type bound of Lemma 3.6 under additional information on the *arithmetic structure* among the components of v.

Lemma 3.6 states that for any unit vector v,

$$\mathcal{L}(X\cdot v, r) \lesssim r \qquad \text{for all } r \ge \|v\|_\infty. \tag{4.5}$$

One can interpret this as saying that the discrete random variable $X\cdot v$ behaves like a variable with bounded density above scales $\gtrsim \|v\|_\infty$. Our starting point for proving this was Corollary 4.2, which gave

$$\mathcal{L}(X\cdot v, r) \lesssim \int_{-1}^1 \exp\left(-c\,\mathrm{dist}^2\left(\frac{2t}{r}v, \mathbb{Z}^n\right)\right)dt. \tag{4.6}$$

The lower bound on r in (4.5) allowed us to argue that for all $t\in[-1,1]$ the closest lattice point to the vector $(2t/r)v$ is the origin, and the right hand side of (4.6) was then controlled by a Gaussian integral.

We would like to obtain improvements of (4.5) of the form

$$\mathcal{L}(X\cdot v, r) \lesssim r \qquad \text{for all } r \ge r_0(v) \tag{4.7}$$

where $r_0(v)$ can be much smaller than $\|v\|_\infty$. The bound (4.6) suggests that the optimal choice for $r_0(v)$ will involve the distance between dilations of v and the lattice $\mathbb{Z}^n$. Specifically, we will obtain a good estimate on $\mathcal{L}(X\cdot v, r)$ for unit vectors v for which the dilations $\{\theta v : \theta \ge 0\}$ stay well-separated from the lattice $\mathbb{Z}^n$ up to a fairly large value of θ.

We formalize this as follows. For $v\in\mathbb{R}^n$ and a fixed *accuracy level* $\alpha > 0$, define the *essential least common denominator*

$$\mathrm{LCD}_\alpha(v) := \inf\left\{\theta > 0 : \mathrm{dist}(\theta v, \mathbb{Z}^n) \le \min\left(\frac{1}{10}\|\theta v\|, \alpha\sqrt{n}\right)\right\}. \tag{4.8}$$

In terms of this quantity we can obtain the following refinement of Lemma 3.6 due to Rudelson and Vershynin [RV08, RV10].

Theorem 4.3 (Lévy concentration function via additive structure, [RV10])**.** *Let $X = (\xi_1, \dots, \xi_n)$ be a vector of i.i.d. centered random variables satisfying $p := \mathcal{L}(\xi_1, 1) < 1$. Then for every $v \in S^{n-1}$ and every $\alpha > 0$,*

$$\mathcal{L}(X \cdot v, r) \lesssim r + \frac{1}{\mathrm{LCD}_\alpha(v)} + \exp\left(-c\alpha^2 n\right), \qquad r \geq 0 \tag{4.9}$$

where the implied constant and $c > 0$ depend only on p.

Proof. We consider only the case that the variables ξ_j are uniform Bernoulli signs; for the general case see [RV08]. Fix $\alpha > 0$. We may assume

$$r > \frac{2}{\mathrm{LCD}_\alpha(v)}. \tag{4.10}$$

It follows that for every $t \in [-1, 1]$,

$$\frac{2t}{r} \leq \frac{2}{r} < \mathrm{LCD}_\alpha(v).$$

Thus, by definition of $\mathrm{LCD}_\alpha(v)$,

$$\mathrm{dist}\left(\frac{2t}{r} v, \mathbb{Z}^n\right) > \min\left(\frac{1}{5}\left\|\frac{t}{r} v\right\|, \alpha\sqrt{n}\right) \quad \forall t \in [-1, 1]. \tag{4.11}$$

Inserting this in (4.5) gives

$$\mathcal{L}(X \cdot v, r) \lesssim \int_{-1}^{1} \exp\left(-c' t^2 / r^2\right) dt + \exp\left(-c\alpha^2 n\right) \lesssim r + \exp\left(-c\alpha^2 n\right)$$

as desired. □

Theorem 4.3 improves on Lemma 3.6 for vectors $v \in S^{n-1}$ for which $\mathrm{LCD}_\alpha(v) \gg \sqrt{n}$ for some fixed accuracy level $\alpha > 0$. We now consider some examples. For the examples we assume α is fixed and n is sufficiently large.

- If v is a standard basis vector, i.e. $v = e_j$ for some $1 \leq j \leq n$, then $\mathrm{LCD}_\alpha(v)$ is of constant order (specifically equal to 10/11).
- If v is the constant vector $v = \frac{1}{\sqrt{n}}(1, \dots, 1)$ then $\mathrm{LCD}_\alpha(v) \asymp \sqrt{n}$, and we recover the estimate of Lemma 3.6 in this case.
- If $v = \frac{c}{\sqrt{n}}(1 + \frac{1}{n}, 1 + \frac{2}{n}, \dots, 1 + \frac{n}{n})$ (where c is a normalizing constant) we have $\mathrm{LCD}_\alpha(v) \asymp n^{3/2}$, and in this case Theorem 4.3 improves on Lemma 3.6.

The following theorem together with Theorem 4.3 leads to an improvement of the $O(n^{-1/2})$ term in Theorem 3.1 to $O(e^{-cn})$.

Theorem 4.4. *Let $X_1, \dots, X_{n-1}$ be i.i.d. random vectors in $\mathbb{R}^n$ whose components are i.i.d. centered* sub-Gaussian *variables with unit variance. Let $u \in S^{n-1}$ be a unit normal vector to $H = \mathrm{span}(X_1, \dots, X_{n-1})$. If $\alpha > 0$ is a sufficiently small constant then*

$$\mathbb{P}(\mathrm{LCD}_\alpha(u) < e^{cn}) \leq e^{-cn}$$

for some constant $c > 0$.

See [RV08] for the proof.

5. Universality for the smallest singular value

In this section we sketch the proof of a result of Tao and Vu from [TV10a] that the distribution of the smallest singular value of a square matrix M with i.i.d. centered entries of unit variance (and satisfying an additional moment hypothesis) asymptotically matches that of a Gaussian matrix. In recent years there have been several results of this type establishing universality for local eigenvalue statistics of random matrices; we refer to [Tao19, EY17] and references therein. Many of the proofs proceed by comparison with a Gaussian matrix, for which the spectral statistics of interest can be computed explicitly. One such comparison approach for Hermitian random matrices is to use the so-called *Dyson Brownian motion*. Roughly speaking, one considers a stationary diffusion on the space of Hermitian matrices with initial condition H. In the long-time limit the matrix converges to a Gaussian matrix G; on the other hand, one can show that the spectral statistics have changed by a negligible amount, and universality follows. Dyson Brownian motion thus provides a nice intuition for why Gaussian matrices asymptotically give the right answers.

Besides diffusion processes, there is another phenomenon of a geometric flavor that naturally gives rise to Gaussian distributions:

(Generic) low-dimensional projections of high dimensional measures (satisfying certain hypotheses) are approximately multivariate Gaussian vectors.

This heuristic was exploited by Tao and Vu to establish the following:

Theorem 5.1 (Universality for the smallest singular value [TV10a])**.** *Let ξ be a real-valued random variable satisfying $\mathbb{E}\,\xi = 0, \mathbb{E}\,|\xi|^2 = 1$, and $\mathbb{E}\,|\xi|^{C_0} < \infty$ for some sufficiently large constant C_0. Let $M = (\xi_{ij})$ be an $n \times n$ matrix whose entries are i.i.d. copies of ξ, and let G be as in Theorem 5.3. Then for any fixed $t \geq 0$,*

$$\mathbb{P}\bigg(s_n(M)^2 \leq \frac{t}{n}\bigg) = \mathbb{P}\bigg(s_n(G)^2 \leq \frac{t}{n}\bigg) + o(1).$$

Remark 5.2. [TV10a] obtains the same result in the complex setting, where M has complex entries with independent real and imaginary parts, and G has standard complex Gaussian entries. They also obtain universality for the joint law of the smallest k singular values, for any fixed k. Finally, the above is a weak formulation of their result, which gives quantitative errors in place of the $o(1)$ terms (of the form $O(n^{-c})$ for a constant $c > 0$ sufficiently small depending on C_0).

For the Gaussian case the limiting distribution was worked out explicitly by Edelman [Ede88]:

Theorem 5.3. *Let G be an $n \times n$ matrix with i.i.d. standard (real) Gaussian entries. For any fixed $t \geq 0$,*

$$\mathbb{P}\left(s_n(G)^2 \leq \frac{t}{n}\right) = \int_0^t \frac{1+\sqrt{x}}{2\sqrt{x}} \exp\left(-\frac{x}{2}+\sqrt{x}\right) dx + o(1).$$

We now discuss some of the steps of Theorem 5.1. We denote the columns of M by $X_1, \ldots, X_n$. For $0 \leq k \leq n-1$ we denote $W_{>k} := \operatorname{span}(X_{k+1}, \ldots, X_n)$. The basic idea is that the inverse of M is effectively of low rank (consider inverting the singular values distributed according to the quarter-circular law). In particular, one can estimate its norm by that of a randomly sampled small submatrix. One can check that for M this corresponds to estimating its smallest singular value by that of a small matrix obtained by projecting a randomly sampled collection of columns to the span of the remaining columns. We summarize these steps in the following; see [TV10a] for the proof. In these notes we focus on how the projection operation gives rise to the universal Gaussian distribution.

Proposition 5.4. *There is a constant $c > 0$ depending only on C_0 such that the following holds. Set $m = \lfloor n^{500/C_0} \rfloor$, and let A be the $n \times m$ matrix with columns $X_1, \ldots, X_m$. Let U be an $n \times m$ matrix whose columns $u_1, \ldots, u_n$ are an orthonormal basis for $W_{>m}^{\perp}$, chosen measurably with respect to the sigma algebra generated by $X_{k+1}, \ldots, X_n$ (say according to Haar measure). Put $\widehat{M} = U^{\mathsf{T}} A$. Then for all fixed $t \geq 0$,*

$$\mathbb{P}\left(s_m(\widehat{M})^2 \leq \frac{t-o(1)}{m}\right) - o(1) \\ \leq \mathbb{P}\left(s_n(M)^2 \leq \frac{t}{n}\right) \leq \mathbb{P}\left(s_m(\widehat{M})^2 \leq \frac{t+o(1)}{m}\right) + o(1).$$

Thus, our aim is now to show that $\widehat{M}$ is approximately an $m \times m$ Gaussian matrix. Writing $\widehat{M} = (Y_1 \cdots Y_m)$, we have $Y_j = U^{\mathsf{T}} X_j$. The main point is that the matrix U acts as a low-dimensional projection; moreover, it is independent of the columns of A. Heuristically, we expect that for "generic" projections U, the vector Y_j will be approximately Gaussian. Indeed, the Central Limit theorem is the special case $m = 1$ and $u = \frac{1}{\sqrt{n}}(1, \ldots, 1)$. Note this can't hold for *any* projection – consider for instance the projection to the first m coordinates. However, if the components the columns of U are sufficiently spread out then we expect to have $Y_j \approx N(0, I_m)$. This is formalized in the following:

Proposition 5.5 (Berry–Esséen theorem for frames)**.** *Let $\xi_1, \ldots, \xi_n$ be i.i.d. centered variables of unit variance and finite third moment. Let $1 \leq m \leq n$, and let $v_1, \ldots, v_n$ be a tight frame for $\mathbb{R}^m$, i.e.*

$$\sum_{i=1}^n v_i v_i^{\mathsf{T}} = I_m.$$

Let K be such that $\|v_i\| \le K$ for all $1 \le i \le n$. Put $Y = \sum_{i=1}^n \xi_i v_i$, and let $Z \sim N(0, I_m)$ be a standard Gaussian vector. For any Borel set $\Omega \subset \mathbb{R}^m$ and any $\varepsilon > 0$,

$$\mathbb{P}\big(Z \in \Omega_\varepsilon^-\big) - O(m^{5/2}\varepsilon^{-3}K) \le \mathbb{P}(Y \in \Omega) \le \mathbb{P}\big(Z \in \Omega_\varepsilon^+\big) + O(m^{5/2}\varepsilon^{-3}K)$$

where Ω_ε^+ denotes the ε-neighborhood of Ω in the ℓ^∞ metric, and Ω_ε^- is the complement of the ε-neighborhood of Ω^c in the ℓ^∞ metric.

Now we show how Proposition 5.5 applies to the matrix $\widehat{M}$ constructed in Proposition 5.4. Let $v_1^\mathsf{T}, \dots, v_n^\mathsf{T}$ be the rows of U. Since the columns of U are orthonormal, $I_m = U^\mathsf{T} U$. On the other hand we have $U^\mathsf{T} U = \sum_{i=1}^n v_i v_i^\mathsf{T}$. Thus, the rows of U form a tight frame for $\mathbb{R}^m$. By construction, the columns of $\widehat{M}$ are independent, and are given by

$$Y_j = U^\mathsf{T} X_j = \sum_{i=1}^n \xi_{ij} v_i, \qquad j = 1, \dots, m.$$

We then want to apply Proposition 5.5 to deduce that each of the columns $Y_1, \dots, Y_m$ is approximately a standard Gaussian vector in $\mathbb{R}^m$, and for this we will want to take K small. Specifically, it will be enough to show

$$\max_{i\in[n], j\in[m]} |v_i(j)| \le n^{-c} \quad \text{with probability } 1 - o(1) \tag{5.1}$$

for some absolute constant $c > 0$. From this it follows that we can take $K = \sqrt{m} n^{-c}$ in Proposition 5.5. Then it is possible to argue that

$$\mathbb{P}\left(s_m(\widehat{M})^2 \le \frac{t}{m}\right) = \mathbb{P}\left(s_m(G)^2 \le \frac{t}{m}\right) + o(1)$$

for any fixed $t \ge 0$, where G is an $m \times m$ matrix whose entries are i.i.d. standard Gaussians. See [TV10a] for the details.

Here we will only show how to prove (5.1). Specifically we will prove

Proposition 5.6 (Normal vectors are delocalized)**.** *Let m be as in Proposition 5.4, with C_0 sufficiently large. Set $B = (X_{m+1} \cdots X_n)$. Then with probability $1 - o(1)$, for any $v \in S^{n-1}$ such that $v^\mathsf{T} B = 0$ we have $\|v\|_\infty \le n^{-c_0}$, where $c_0 > 0$ is a sufficiently small absolute constant.*

To prove this we will need two lemmas.

Lemma 5.7 (Distance of a random vector to a fixed subspace)**.** *Let $X = (\xi_1, \dots, \xi_n)$ have i.i.d. centered components of unit variance and bounded by 100 almost surely (we could replace 100 with any other fixed constant). Let $V \subset \mathbb{R}^n$ be a fixed subspace of dimension d. Then for some constant $c > 0$ we have*

$$\mathbb{P}\Big(\mathrm{dist}(X, V) \le c\sqrt{n-d}\Big) \lesssim e^{-c(n-d)}.$$

(In fact one can adjust the proof to get two-sided concentration around $\sqrt{n-d}$, but we only need a lower bound.)

Proof. It is not hard to show that $x \mapsto \operatorname{dist}(X, V)$ is a convex and 1-Lipschitz function with respect to the Euclidean metric. Write $D = \operatorname{dist}(X, V)$. From Talagrand's inequality it follows that

$$\mathbb{P}(|D - M| \geq t) \leq 4 \exp\left(-ct^2\right)$$

where M is any median for D.

Now we argue that $M \approx \sqrt{\mathbb{E}\, D^2}$. First,

$$|\,\mathbb{E}\, D - M| \leq \mathbb{E}\,|D - M| = \int_0^\infty \mathbb{P}(|D - M| \geq t)dt = O(1).$$

Next,

$$\mathbb{E}\, D^2 - (\mathbb{E}\, D)^2 = \mathbb{E}\,|D - \mathbb{E}\, D|^2 \leq \int_0^\infty \mathbb{P}\big(|D - \mathbb{E}\, D|^2 \geq t\big)dt = O(1).$$

Thus, $M = \sqrt{\mathbb{E}\, D^2} + O(1)$.

Now we compute $\mathbb{E}\, D^2$. Let $\Pi_{V^\perp}$ denote the matrix for projection to $V^\perp$. We have

$$\mathbb{E}\, D^2 = \mathbb{E}\, \|\Pi_{V^\perp} X\|^2 = \mathbb{E} \sum_{i,j=1}^n \Pi_{V^\perp}(i,j)\xi_i\xi_j = \operatorname{tr} \Pi_{V^\perp} = \dim(V^\perp) = n - d.$$

Thus,

$$\begin{aligned}
\mathbb{P}\Big(D \leq c\sqrt{n-d}\Big) &\leq \mathbb{P}\Big(D - M \leq -(1-c)\sqrt{n-d} + O(1)\Big) \\
&\leq \mathbb{P}\bigg(D - M \leq \frac{1}{2}\sqrt{n-d}\bigg) \\
&\leq 4 \exp\left(-c(n-d)^2\right)
\end{aligned}$$

where we took $c > 0$ sufficiently small. (Note that we can assume $n - d$ is larger than any fixed constant, which allowed us to bound the $O(1)$ term by $\frac{1}{10}\sqrt{n-d}$, say.) □

In the proof of Proposition 5.6 we will want to locate a large subspace on which a large square sub-matrix of M has small norm. Note that for an i.i.d. matrix, by the quarter circular law we have

$$|\{i : s_i(M) \leq \varepsilon\sqrt{n}\}| \asymp \varepsilon n$$

for any fixed $\varepsilon \in (0, 2)$. The following lemma, which we state without proof, extends the lower bound to ε decaying with n.

Lemma 5.8 (Many small singular values)**.** *There is a sufficiently small constant $c > 0$ such that with probability $1 - O(e^{-n^c})$,*

$$|\{i : s_i(M) \leq n^{1/2-c}\}| \gtrsim n^{1-c}.$$

Remark 5.9. The above is a weak version of the *local Marchenko–Pastur law*, which basically states that the Marchenko–Pastur distribution gives an accurate estimate for the number of singular values $s_i(M)$ in intervals of length down to $n^{-1/2}\log^{O(1)} n$, which is optimal up to poly-logarithic factors (note we have not scaled M to have a limiting singular value distribution supported on a compact interval). See [ESYY12, Section 8].

Proof of Proposition 5.6. Let $c_0 > 0$ to be chosen sufficiently small. By the union bound it suffices to show that with probability $1 - o(1/n)$, for any fixed $i \in [n]$ and any $v \in S^{n-1}$ such that $v^\mathsf{T} B = 0$, we have $|v_i| \le n^{-c_0}$. By exchangeability of the rows of B it suffices to show this for $i = 1$.

We denote the rows of B by $R_1, \dots, R_n$. We let B' be the $n-1 \times n-m$ submatrix with rows $R_2, \dots, R_n$, and $\widetilde{M}$ the $n-m \times n-m$ submatrix with rows $R_{m+1}, \dots, R_n$. Recall from Proposition 5.4 that $m = \lfloor n^{500/C_0} \rfloor$ where C_0 is as large as we please. In particular, $n - m = n(1 - o(1))$.

First we will find a large subspace V on which B' has small norm. Applying Lemma 5.8 to $\widetilde{M}$, it follows that with probability $1 - O(n^{-10})$ there exists a subspace $V_0 \subset \mathbb{R}^{n-m}$ such that $\dim V_0 \gtrsim n^{1-c}$ and $\|\widetilde{M}|_{V_0}\| \le n^{1/2-c}$. Indeed, V_0 is simply the space spanned by the eigenvectors of $\widetilde{M}^*\widetilde{M}$ associated to the eigenvalues $s_i(\widetilde{M})^2$ with $s_i(\widetilde{M}) \le n^{1/2-c}$. Now set

$$V = V_0 \cap \operatorname{span}(R_2, \dots, R_m)^\perp.$$

Then $\dim V \ge \dim V_0 - m \gtrsim n^{1-c}$, and by construction we have $\|B'|_V\| \le n^{1/2-c}$.

Now suppose $v \in S^{n-1}$ satisfies $v^\mathsf{T} B = 0$ and $|v_1| \ge n^{-c_0}$. Then

$$0 = v_1 R_1 + v'^\mathsf{T} B'$$

where $v' := (v_2, \dots, v_n)$. Let Π_V be the matrix for projection to the subspace V constructed above. Multiplying on the right by Π_V in the above equality, rearranging and taking norms gives

$$\|v_1 R_1 \Pi_V\| = \left\| v'^\mathsf{T} B' \Pi_V \right\|. \tag{5.2}$$

The right hand side is bounded by

$$\|B' \Pi_V\| \le n^{1/2-c} \tag{5.3}$$

with probability $1 - O(n^{-10})$. For the left hand side we have

$$\|v_1 R_1 \Pi_V\| \ge n^{-c_0} \|R_1 \Pi_V\| = n^{-c_0} \operatorname{dist}(R_1, V^\perp).$$

Now we condition on a realization of $R_2, \dots, R_n$ for which the bound (5.3) holds. This fixes the subspace V. Applying Lemma 5.7 we have

$$\operatorname{dist}(R_1, V^\perp) \gtrsim \sqrt{\dim V} \gtrsim n^{(1-c)/2} \tag{5.4}$$

with probability $1 - O(e^{-n^{c'}})$ for some constant $c' > 0$. Intersecting the good events that (5.4) and (5.3) hold and putting the bounds together, we conclude that with probability $1 - O(n^{-10})$,

$$n^{-c_0} n^{(1-c)/2} \leq n^{1/2-c}.$$

Taking c_0 sufficiently small depending on c gives a contradiction, and the result follows. □

6. Further reading

For further reading on the topics covered here one may consult:

(1) Sections 2.3 and 2.7 of Tao's text on random matrix theory [Tao12] (also available on his blog).
(2) The ICM lecture notes of Rudelson and Vershynin [RV10].
(3) The paper [TV10a] of Tao and Vu establishing universality for small singular values.

All of these sources are clearly written, with the first two assuming little background in random matrix theory, so the reader is encouraged to consult them for further information or clarification.

References

[BS98] Z. D. Bai and J. W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. *Ann. Probab.*, 26(1):316–345, 1998.

[BVW10] Jean Bourgain, Van H. Vu, and Philip Matchett Wood. On the singularity probability of discrete random matrices. *J. Funct. Anal.*, 258(2):559–603, 2010.

[Ede88] Alan Edelman. Eigenvalues and condition numbers of random matrices. *SIAM J. Matrix Anal. Appl.*, 9(4):543–560, 1988.

[ESYY12] László Erdős, Benjamin Schlein, Horng-Tzer Yau, and Jun Yin. The local relaxation flow approach to universality of the local statistics for random matrices. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 48(1):1–46, 2012.

[EY12] László Erdős and Horng-Tzer Yau. Universality of local spectral statistics of random matrices. *Bulletin, new series, of the American Mathematical Society*, 49(3):377–414, 2012.

[EY17] László Erdős and Horng-Tzer Yau. *A dynamical approach to random matrix theory*, volume 28 of *Courant Lecture Notes in Mathematics*. Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2017.

[KKS95] Jeff Kahn, János Komlós, and Endre Szemerédi. On the probability that a random ±1-matrix is singular. *J. Amer. Math. Soc.*, 8(1):223–240, 1995.

[Kom] János Komlós. Circulated manuscript, 1977. Edited version available online at: http://www.math.rutgers.edu/∼komlos/01short.pdf.

[Kom67] J. Komlós. On the determinant of (0, 1) matrices. *Studia Sci. Math. Hungar*, 2:7–21, 1967.

[NV13] Hoi H. Nguyen and Van H. Vu. Small ball probability, inverse theorems, and applications. In *Erdös centennial*, volume 25 of *Bolyai Soc. Math. Stud.*, pages 409–463. János Bolyai Math. Soc., Budapest, 2013.

[Rud08] Mark Rudelson. Invertibility of random matrices: norm of the inverse. *Ann. of Math. (2)*, 168(2):575–600, 2008.

[RV08] Mark Rudelson and Roman Vershynin. The Littlewood-Offord problem and invertibility of random matrices. *Adv. Math.*, 218(2):600–633, 2008.

[RV10] Mark Rudelson and Roman Vershynin. Non-asymptotic theory of random matrices: extreme singular values; proceedings of the international congress of mathematicians. volume iii. pages 1576–1602, 2010.

[Tao12] Terence Tao. *Topics in random matrix theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.

[Tao19] Terence Tao. Least singular value, circular law, and Lindeberg exchange. In *Random matrices*, volume 26 of *IAS/Park City Math. Ser.*, pages 461–498. Amer. Math. Soc., Providence, RI, 2019.

[TaV14] *Random matrices: The universality phenomenon for Wigner ensembles; Proceedings of Symposia in Applied Mathematics, Volume 72 : Modern Aspects of Random Matrix Theory*, volume 72 of *Proceedings of Symposia in Applied Mathematics*. American Mathematical Society, 2014.

[TV07] Terence Tao and Van H. Vu. On the singularity probability of random Bernoulli matrices. *J. Amer. Math. Soc.*, 20(3):603–628, 2007.

[TV08] Terence Tao and Van H. Vu. Random matrices: the circular law. *Commun. Contemp. Math.*, 10(2):261–307, 2008.

[TV09] Terence Tao and Van H. Vu. Inverse Littlewood-Offord theorems and the condition number of random discrete matrices. *Ann. of Math. (2)*, 169(2):595–632, 2009.

[TV10a] Terence Tao and Van Vu. Random matrices: the distribution of the smallest singular values. *Geom. Funct. Anal.*, 20(1):260–297, 2010.

[TV10b] Terence Tao and Van H. Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010. Paperback edition [of MR2289012].

[TV15] Terence Tao and Van Vu. Random matrices: Universality of local spectral statistics of non-hermitian matrices. *Annals of probability*, 43(2):782–874, 2015.

[Ven13] Akshay Venkatesh. A note on sphere packings in high dimension. *International mathematics research notices*, (7):1628–1642, 2013.

[Ver18] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2018.

Department of Mathematics, Duke University, Durham, NC 27708, USA
Email address: `nickcook@math.duke.edu`