Title
The Eigenvalue Spacing of IID Random Matrices and Related Least Singular Value Results

Author
Ge, Stephen

Publication Date
2017

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

The Eigenvalue Spacing of IID Random Matrices and Related Least Singular Value Results

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Mathematics

by

Stephen Cong Ge

2017

ABSTRACT OF THE DISSERTATION

The Eigenvalue Spacing of IID Random Matrices
and Related Least Singular Value Results

by

Stephen Cong Ge
Doctor of Philosophy in Mathematics
University of California, Los Angeles, 2017
Professor Terence Chi-Shen Tao, Chair

This thesis studies the spacing between eigenvalues of random matrices with independent and identically distributed (iid) entries. Tail estimates on the minimum distance between any pair of eigenvalues are proven. In particular, we establish that the spectrum of an iid random matrix is simple with high probability. A key technical result is a new least singular value tail estimate for shifted matrices of the form $A_n - zI_n$, where A_n is an iid random matrix with real entries and z is a complex scalar.

The dissertation of Stephen Cong Ge is approved.

Dimitri Shlyakhtenko

Ali H. Sayed

Terence Chi-Shen Tao, Committee Chair

University of California, Los Angeles

2017

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to thank my advisor Terry Tao for suggesting the problem of eigenvalue spacing and his guidance throughout the research process. I am grateful to Terry for all that he has taught me about mathematics and problem solving through weekly meetings, formal courses, and his books and blog posts.

The UCLA Analysis group has been an integral part of my graduate school experience. I would like to thank professors John Garnett, Rowan Killip, Jim Ralston, and Monica Visan for holding welcoming seminars and for enjoyable discussions during trips to Caltech. I am grateful to Maggie Albert, Maida Bassili, and Martha Contreras for their administrative support. I thank my fellow students Adam Azzam, Nick Cook, Julian Gold, Dohyun Kwon, and Zane Li for many interesting discussions and all their help in general.

Lastly I would like to thank Bonny Lin and my family for their love and never ending support. In particular, I thank my mother for her encouragement of my mathematical pursuits and my father for instilling a lifelong love of tackling difficult problems.

VITA

2012	Phi Beta Kappa
2012	Bachelor of Science (Mathematics), Massachusetts Institute of Technology
2013 - 2017	Teaching Assistant, University of California, Los Angeles
2015	Candidate in Philosophy (Mathematics), University of California, Los Angeles

CHAPTER 1

Introduction

1.1 Background and past results

We first introduce our primary objects of study and classic related results. Let ξ be a fixed random variable and let A_n be an $n \times n$ random matrix with each entry an independent and identically distributed (iid) copy of ξ. We call A_n an *iid matrix* and the sequence $(A_n)_{n=1}^{\infty}$ an *iid ensemble*. We refer to ξ as the *atom random variable*. Common choices for the atom random variable include taking ξ to be the standard normalized real gaussian

$$g_{\mathbf{R}} := N(0,1)$$

or standard normalized complex gaussian, denoted $g_{\mathbf{C}}$, which has real and imaginary parts each an independent copy of $N(0, \frac{1}{2})$. Random sign matrices where ξ is Bernoulli taking the values ± 1 with equal probability are a common discrete example.

The set of eigenvalues, or the spectrum, of A_n is a primary topic in the study of iid matrix ensembles. Let $\{\lambda_k\}_{k=1}^{n}$ be the eigenvalues of A_n and define

$$\mu_n := \frac{1}{n} \sum_{k=1}^{n} \delta_{\frac{1}{\sqrt{n}}\lambda_k}$$

to be the empirical spectral distribution of A_n. A central result for the spectra of iid matrices is the circular law, which has an extensive history with many authors. [Gir84], [Bai97], [PZ10], and [GT10] among others progressively weakened the assumptions on the atom random variables, culminating with the form we state below obtained by Tao and Vu in [TV10b].

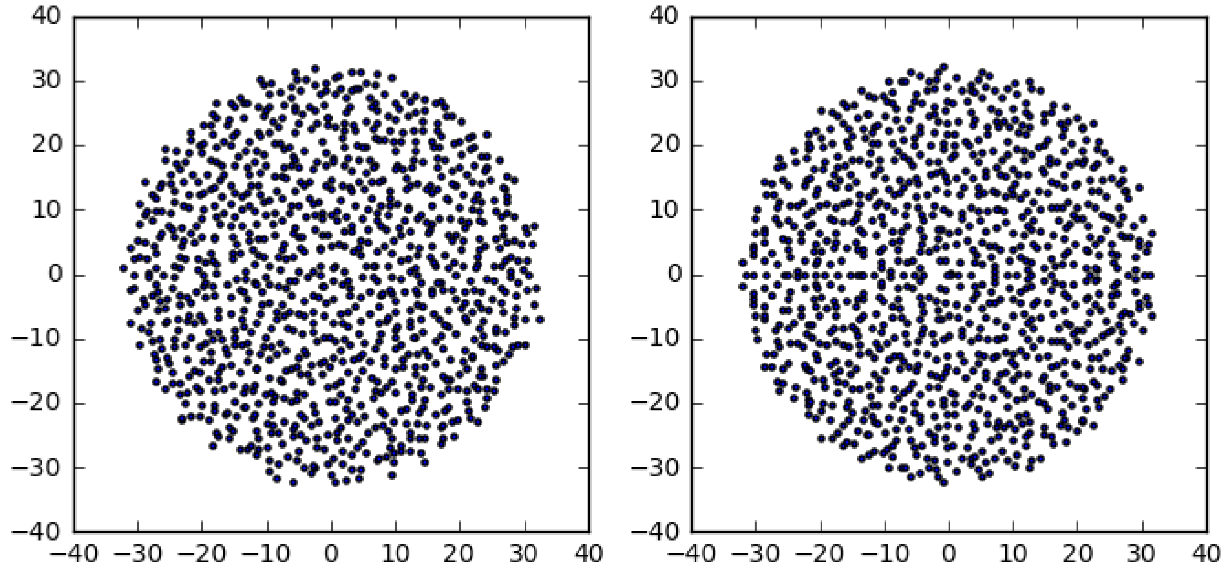

Figure 1.1: Eigenvalues of 256×256 matrices generated using $\xi \sim N_{\mathbf{C}}(0,1)$ (left) and $\xi \sim \text{Bernoulli}(1/2)$ (right)

Theorem 1.1.1. *Let ξ be a random variable with mean 0 and variance 1 and take $(A_n)_{n=1}^{\infty}$ to be an iid ensemble with ξ as the atom random variable. Then: with probability 1, μ_n converges to μ_{circ}, the uniform distribution over the (complex) unit disk as n tends to infinity.*

The circular law is an example of the universality principle in random matrix theory. The limiting behavior of the spectrum of A_n is not affected by the particular distribution of the atom variables ξ, but is instead determined completely by the mean and variance of ξ. The limiting distribution μ_{circ} was first computed explicitly for the complex Ginibre ensemble, where the atom random variable is taken to be $\xi \sim g_{\mathbf{C}}$.

The primary technical tool towards the proof of the circular law in [TV10b] was a lower tail estimate for the least singular value of A_n. Such least singular value results have proven fruitful for other applications in the study of iid random matrices, and the main results in this thesis are another example. Later in this section, we will review a collection of least singular value results and their applications. Before going on to lower tail estimates for the least singular value, we briefly discuss the singularity probability problem, i.e. the extreme case where the least singular value is zero.

1.1.1 Singularity probability

When the atom random variable ξ is gaussian, or any distribution that is absolutely continuous with respect to Lebesgue measure, we have

$$\mathbf{P}(A_n \text{ is singular}) = 0$$

This can be quickly seen from the fact that the subset of $\mathbf{R}^{n\times n}$ or $\mathbf{C}^{n\times n}$ consisting of singular matrices has positive codimension, and hence Lebesgue measure 0. This argument also applies to symmetric or Hermitian matrix ensembles. The situation is markedly different when the atom random variable ξ is discrete. For the rest of this section, we consider the case of random sign matrices, where the entries of A_n are ± 1 Bernoulli. The event that two rows or columns are equal occurs with positive probability, which gives a lower bound for the singularity probability of random sign matrices.

The result that A_n is singular with probability tending to 0 is nontrivial and first established by Komlós in [Kom67]. Exponential bounds of the form

$$\mathbf{P}(A_n \text{ is singular}) \le c^n$$

for $c < 1$ were obtained in [KKS95], [TV07], [BVW10]. The aforementioned works progressively lowered the numerical value of c toward the conjectured $c = 1/2 + o(1)$, which matches the lower bound from considering the event that two rows or columns are parallel. The current record established in [BVW10] is $c = \frac{1}{\sqrt{2}} + o(1)$. The methods in [TV07] and [BVW10] relied on tools from additive combinatorics to illuminate the connections between the arithmetic structure of a vector and concentration probabilities of random walks associated with the vector. Similar methods were later also crucial for the least singular value result that helped establish the circular law in [TV10b].

1.1.2 Least singular value

We now discuss tail bounds for the least singular value of iid random matrices. Recall that for an $n \times n$ matrix A, the singular values are defined to be the eigenvalues of the matrix

$\sqrt{A^*A}$. We will arrange them in nonincreasing order:

$$s_1(A) \geq \cdots \geq s_n(A) \geq 0$$

The two extreme singular values are also commonly characterized as:

$$s_1(A) = \|A\|_{op} = \sup_{\|z\|=1} \|Az\| \text{ and } s_n(A) = \inf_{\|z\|=1} \|Az\|$$

We use $G_{\mathbf{R}}$ or $G_{\mathbf{C}}$ to denote $n \times n$ matrices drawn from the real or complex Ginibre ensemble respectively, i.e. where the atom random variables are gaussian $g_{\mathbf{R}}$ or $g_{\mathbf{C}}$. In [Ede88], Edelman computed the probability density functions of ns_n^2 in both cases, which upon integration give the following lower tail estimates for the least singular value:

$$\mathbf{P}(s_n(G_{\mathbf{R}}) \leq tn^{-1/2}) \leq t \tag{1.1}$$

$$\mathbf{P}(s_n(G_{\mathbf{C}}) \leq tn^{-1/2}) \leq t^2 \tag{1.2}$$

for any $t > 0$. Note the relationship between the bound on the size of s_n and the probability bound. Modulo a constant power of n, it is linear $tn^{-1/2}$ to t for $G_{\mathbf{R}}$ while being quadratic $tn^{-1/2} \mapsto t^2$ for $G_{\mathbf{C}}$. In [SST06], Sankar, Spielman, and Teng extended (1.1) to include an arbitrary shift, showing that

$$\mathbf{P}(s_n(G_{\mathbf{R}} + M) \leq tn^{-1/2}) \leq Ct \tag{1.3}$$

where M is any deterministic $n \times n$ real matrix and $C > 0$ is an absolute constant. More recent advancements have obtained lower tail estimates for ensembles with more general atom distributions, in particular discrete random variables. In [TV10c], Tao and Vu optimized their previous least singular result in [TV10b] and showed:

Theorem 1.1.2. *Let ξ be a mean zero random variable with bounded second moment. Let $\gamma \geq 1/2$ and $A \geq 0$ be constants. Let A be an $n \times n$ random matrix with ξ as atom random variable and let M be a deterministic matrix with $\|M\|_{op} \leq n^\gamma$. Then:*

$$\mathbf{P}(s_n(A + M) \leq n^{-(2A+2)\gamma+1/2}) \leq c(n^{-A+o(1)} + \mathbf{P}(\|M\| \geq n^\gamma))$$

for some absolute constant $c > 0$.

The bound in (1.3) implies that a gaussian random matrix allows for deterministic shifts of any size. However, in [TV10c] it was also established that for more general distributions, the probability bound is impacted by the norm of the shift matrix.

In a series of papers on the smallest singular value of random matrices by Rudelson and Vershynin, the concept of least common denominator was developed to characterize the arithmetic structure of a vector. The following optimal bound was proven in [RV08]:

Theorem 1.1.3. *Let* ξ *be a mean zero, variance* 1*, subgaussian random variable. Let* A *be an* $n \times n$ *matrix with* ξ *as atom variables. Then for any* $t > 0$*:*

$$\mathbf{P}(s_n(A) \leq tn^{-1/2}) \leq Ct + e^{-cn}$$

for absolute constants $C, c > 0$*.*

The least common denominator method was also used to give tail estimates on the smallest singular value of rectangular random matrices in [RV09] and applied to prove a form of delocalization of eigenvectors in [RV16]. Specifically, it was shown that with high probability, every eigenvector of a random matrix has a non-negligible portion of mass located in any subset of its coordinates.

Luh used the least common denominator method in [Luh16] to prove the following least singular value result for complex iid matrices:

Theorem 1.1.4. *Let* $\zeta = \xi + i\xi'$ *be a complex random variable where* ξ, ξ' *are iid copies of a mean* 0*, variance* 1*, subgaussian real valued random variable. Let* A *be a* $n \times n$ *matrix with* ζ *as atom random variable and let* M *be a deterministic complex matrix with* $\|M\|_{op} \leq K\sqrt{n}$*. Then:*

$$\mathbf{P}(s_n(A + M) \leq tn^{-1/2}) \leq Ct^2 + e^{-cn}$$

for absolute constants $K, C, c > 0$*.*

Up to constant factors, the previous two theorems match the bounds (1.1) and (1.2) for gaussian ensembles. The exponential error term is necessary, due to the singularity probability in the discrete case. Combining the above tail estimate (1.1.4) with a covering argument along the real line, it was shown that random matrices with complex entries do not have real eigenvalues. More precisely: with A having atom random variable ζ as above,

$$\mathbf{P}(A \text{ has a real eigenvalue}) \leq e^{-cn}$$

for some $c > 0$.

1.1.3 Towards eigenvalue spacing

The circular law governs the number of eigenvalues at a macroscopic scale: in any fixed region $R \subset \mathbf{C}$, the proportion of eigenvalues out of n total falling in R is given by the area of the intersection of R with the unit disk. This thesis focuses on the eigenvalues at an individual level: the minimal spacing between any two eigenvalues of an iid matrix. More precisely, we study the quantity

$$\Delta(A_n) := \min_{j \neq k} |\lambda_j(A_n) - \lambda_k(A_n)|$$

Before going on to the tail of the minimum gap $\Delta(A_n)$, we briefly discuss the easier problem of bounding

$$\mathbf{P}(A_n \text{ has a repeat eigenvalue})$$

i.e. the case where $\Delta(A_n) = 0$.

The problem of simple spectrum for symmetric Wigner-type ensembles was considered in [TV14], in which the authors prove the following:

Theorem 1.1.5. *Let $M_n = (\xi_{ij})$ be a real symmetric random matrix where ξ_{ij} are jointly independent for $i < j$ and $\xi_{ji} = \xi_{ij}$. With proper distribution assumptions on ξ_{ij}, the spectrum of M_n is simple with probability at least $1 - n^{-B}$ for any fixed $B > 0$.*

The methods in [TV14] and the followup [NTV17] crucially used the Cauchy interlacing law for eigenvalues, which is not available when the random matrices are not constrained to be symmetric or Hermitian. When ξ has a continuous distribution, by a similar argument as previously given for singularity probability we have

$$\mathbf{P}(A_n \text{ has simple spectrum}) = 1$$

On the other hand, a repeat eigenvalue occurs with nonzero probability for random sign matrices. For example, three columns being all multiples of each other implies 0 is a repeat eigenvalue. This possibility occurs with exponentially small probability, which will be an error of acceptable size in the bounds we obtain. For iid matrices with arbitrary normalized atom random variable, we obtain simplicity of the spectrum with high probability as a corollary of a polynomial tail bound for the spacing between eigenvalues.

1.2 Main results

The main eigenvalue spacing result in this thesis is the following:

Theorem 1.2.1. *Let ξ be a real valued random variable with mean 0, variance 1, and bounded fourth moment. Let A be an $n \times n$ matrix with each entry an iid copy of ξ. Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $\frac{1}{\sqrt{n}}A$ and*

$$\Delta := \min_{i \neq j} |\lambda_i - \lambda_j|$$

be the minimum gap between any pair of eigenvalues. Then for $s = o(n^{-4+o(1)})$, the event $\{\Delta \geq s\}$ occurs with high probability

Qualitatively, the above theorem implies that iid random matrices have simple spectrum asymptotically almost surely. The asymptotic error can be improved with stronger moment assumptions on the atom random variable. A more precise statement of the main result and additional related corollaries of the method are given in Chapter 3, where the proof can also be found.

The primary technical result needed to establish the eigenvalue spacing theorem above is a tail estimate for the least singular value of a real iid random matrix that has been shifted away from the real line. The new result, the proof of which is given in the following chapter, is as follows:

Theorem 1.2.2. *Let ξ be a real valued random variable with mean 0 and variance 1. Let A be an $n \times n$ matrix with each entry an iid copy of ξ. For every $t \geq 0$ and $\lambda \in \mathbf{C}$ of bounded* *size, we have*

$$\mathbf{P}(s_n(A-\lambda\sqrt{n}I_n) \le tn^{-1/2} \text{ and } \|A\|_{op} \le M\sqrt{n}) \le C\frac{t^2}{\delta} + e^{-cn}$$

where δ is the absolute value of the imaginary part of λ and $M, C, c > 0$ are absolute constants independent of n.

Theorem 1.2.2 can be seen as interpolating between the real and complex gaussian bounds in (1.1), (1.2). When A is shifted a constant amount δ away from the real line, then we have a quadratic t^2 probability bound as in (1.2). On the other hand, for δ of size comparable to t, the probability bound is t as in (1.1).

1.3 Notations and conventions

n will always be the parameter going to infinity in asymptotic notation. In particular, n will be frequently assumed to be sufficiently large. n will usually be related to the dimension of a random matrix. We use $X = O(Y), X \ll Y, Y \gg X, Y = \Omega(X)$ synonymously to denote $X \le CY$ for some positive constant C not depending on n. $X = \Theta(Y)$ denotes that we have both $X \ll Y$ and $X \gg Y$. $X = o(Y)$ means that X/Y tends to 0 as n goes to infinity. Subscripts of constants as in C_γ or $X = O_\gamma(Y)$ will mean that the asymptotic constant may depend on γ. We will usually use $C > 0$ to denote constants that are sufficiently big in order to satisfy parts of the argument and $c > 0$ for constants that are taken sufficiently close to 0.

We will use z to denote complex numbers or vectors. The decomposition $z = x + iy$ will always have x as the real and y as the imaginary parts of z respectively. With $z \in \mathbf{C}^n$, we then have $x, y \in \mathbf{R}^n$.

A will always denote an $n \times n$ matrix. For a scalar λ, we write $A - \lambda$ instead of $A - \lambda I_n$ for brevity. We will use $\{\lambda_j(A)\}_{j=1}^n$ to denote the complex eigenvalues of A, in an arbitrary order. $\{s_j(A)\}_{j=1}^n$ will denote the singular values of A arranged in decreasing order:

$$s_1(A) \ge \cdots \ge s_n(A) \ge 0$$

$\|A\|$ or $\|A\|_{op}$ will denote the $\ell^2 \to \ell^2$ operator norm of the matrix A. The capital letter M will be used for a bound on the size of the operator norm, typically as in $\|A\|_{op} \leq M\sqrt{n}$.

For radius $r > 0$, we use $B(z, r)$ to denote the ball of radius r around z. We will use $S^{n-1}_{\mathbf{R}}$ or $S^{n-1}_{\mathbf{C}}$ to denote the unit sphere in $\mathbf{R}^n$ or $\mathbf{C}^n$ respectively.

CHAPTER 2

A least singular value bound for complex shifted random matrices

2.1 Theorem statement and proof outline

In this section we outline the approach taken to establish the main least singular value tail estimate result. The proof makes up the bulk of the current chapter. We first give some technical assumptions required of the atom random variables. We will use ξ to denote a real valued random variable satisfying the following: there exist $K, p > 0$ such that

$$\sup_{u \in \mathbf{R}} \mathbf{P}(|\xi - u| < 1) \leq 1 - p \tag{2.1}$$

and

$$\mathbf{P}(1 \leq |\xi - \xi'| \leq K) \geq p/2 \tag{2.2}$$

where ξ' is an iid copy of ξ.

Remark. (2.1) is necessary for a crude bound of the form

$$\mathbf{P}(|\xi_1 v_1 + \cdots + \xi_n v_n - z| \leq c) \leq 1 - c \tag{2.3}$$

and (2.2) is necessary for the main small ball probability theorem. Both conditions hold (for some K, p) for ξ having finite non-zero variance (of at least 1).

We will use A to denote an $n \times n$ matrix with each entry an iid copy of ξ. The main theorem in this chapter is the following:

Theorem 2.1.1. *For every $t \geq 0$ and $\lambda = O(1)$, we have*

$$\mathbf{P}(s_n(A - \lambda\sqrt{n}) \leq tn^{-1/2} \text{ and } \|A\|_{op} \leq M\sqrt{n}) \leq O\left(\frac{t^2}{\delta}\right) + e^{-cn}$$

where $\delta > n^{-C}$ is the absolute value of the imaginary part of λ.

We now give a sketch of the argument. The goal is to control

$$\mathbf{P}(s_n(A - \lambda\sqrt{n}) \leq tn^{-1/2})$$

for constant size $\lambda \in \mathbf{C}$. Via standard arguments, we reduce the tail estimate for the smallest singular value to a bound on the distance from one column of $A - \lambda\sqrt{n}$ to the hyperplane spanned by the other $n-1$ columns, i.e.

$$\mathbf{P}(\operatorname{dist}(X_k, H_k) \leq t)$$

where X_k is the kth column and H_k is the span of all columns other than X_k. Taking X^* to be a unit vector orthogonal to H_k, we have $\operatorname{dist}(X_k, H_k) \geq |\langle X^*, X_k\rangle|$ and hence we are interested in bounding

$$\mathbf{P}(|\langle X^*, X_k\rangle| \leq t)$$

The key point is that X^* can be chosen independently of X_k. We use B to denote the first $n-m$ rows of $(A-\lambda\sqrt{n})^T$, where $m = 1$ in this entire chapter, while $m = 2$ later on when we want to bring in the second smallest singular value. The proofs will work for any m that is constant in n. X^* can be seen as an element of the kernel of B. A large part of the argument will be a detailed analysis of the properties of vectors in $\ker(B)$. We will sometimes use R to denote the real part of B, so that $B = R + i\delta\sqrt{n}I_n'$, where I_n' is the first $n-m$ rows of I_n (assuming without loss of generality that δ is positive). We will also frequently need to intersect with the event where $\|B\|_{op} \leq M\sqrt{n}$. The constant M typically comes from an application of finite fourth moment and will be fixed throughout. Some of the parameters may depend on M.

The primary technical result on the way to the least singular value theorem is to show that unit vectors $z = x + iy$ that are in the kernel of B will usually be arithmetically unstructured in a sense to be made more precise later. We want to show that

$$\mathbf{P}(\inf_S \|Bz\| = 0) \leq e^{-cn}$$

where $S \subset S^{n-1}$ is the set of unit vectors with rich arithmetic structure. We will need to divide up S according to several characteristics. For each of these more restricted subsets, we will construct a low entropy net and combine it with a suitable probability bound for a single fixed vector in the net. The first instance of this method is in Section 2.2.1, where we deal with compressible vectors. In Section 2.3, we make precise the notion of arithmetic structure for $z = x + iy$. For fixed z, we will obtain bounds on

$$\mathbf{P}(\|Bz\| \text{ small})$$

that improve the more unstructured z is. In Section 2.4 we construct nets for subsets of S^{n-1} with prescribed arithmetic attributes. In Section 2.5 we combine the nets with the estimates from Section 2.3 to show that $\|Bz\|$ is unlikely to vanish for any arithmetically structured z, which will establish the main structure theorem. The least singular value tail estimate is proven in the last section.

2.2 Null vectors of complex shifted random matrices

In this section we establish some properties of vectors that can appear in the kernel of B. The invertibility for compressible vectors result is from [RV08], [RV09]. We include the proof as a first instance of the method that combines a discrete net with a bound on $\mathbf{P}(\|Bz\| \text{ small})$ for fixed z.

2.2.1 Compressible and incompressible vectors

We first start with subsets of the unit sphere that have very low entropy:

Definition 2.2.1. (Compressible and incompressible vectors)

Fix constants $a, b \in (0, 1)$. We call a vector $v \in S_{\mathbf{C}}^{n-1}$ *compressible* if there is a vector $v' \in \mathbf{C}^n$ with at most an non-zero coordinates such that $\|v - v'\|_2 \leq b$. We call a unit vector *incompressible* if it is not compressible. Let $\mathrm{Comp}(a, b)$ and $\mathrm{Incomp}(a, b)$ denote the sets of compressible and incompressible unit vectors respectively, so that we have $S_{\mathbf{C}}^{n-1} = \mathrm{Comp}(a, b) \sqcup \mathrm{Incomp}(a, b)$. We can define compressible and incompressible real vectors $x \in$

$S_{\mathbf{R}}^{n-1}$ similarly. We will use a subscript $\mathbf{C}$ or $\mathbf{R}$ to distinguish between the complex or real cases.

Lemma 2.2.2. *(Net for compressible vectors)*

There exists a $(2b_z)$-net $\mathcal{N}$ of $Comp_{\mathbf{C}}(a_z, b_z)$ with

$$|\mathcal{N}| \leq \binom{n}{a_z n} \left(\frac{C}{b_z^2}\right)^{a_z n}$$

The exponent of b_z can be lowered from 2 to 1 for a net of $Comp_{\mathbf{R}}$.

Proof. The set of sparse vectors supported on any fixed $a_z n$ coordinates has a (b_z)-net of cardinality at most $(Cb_z^{-2})^{a_z n}$, since we are essentially looking for a net of $S_{\mathbf{C}}^{a_z n}$. Unfixing the nonzero coordinates has an entropy cost of $\binom{n}{a_z n}$. This gives a (b_z)-net of all sparse vectors of size bounded by

$$\binom{n}{a_z n} \left(\frac{C}{b_z^2}\right)^{a_z n}$$

By the definition of compressibility, the same set is a $(2b_z)$-net of $\mathrm{Comp}_{\mathbf{C}}(a_z, b_z)$. □

The bound we have for any single fixed unit vector is the following:

Proposition 2.2.3. *There exists an absolute constant $c > 0$ such that for any unit vector $z \in S_{\mathbf{C}}^{n-1}$, we have*

$$\mathbf{P}(\|Bz\| \leq c\sqrt{n}) \leq e^{-cn}$$

Proof. The argument is standard and given in [RV16], [TV09]. The proof is essentially tensorizing the crude bound (2.3) over the $n - m$ rows of B. □

Combining the above proposition with the small net for compressible vectors from Lemma 2.2.2 gives:

Proposition 2.2.4. *(Invertibility for compressible vectors)*

There exists constants $a_z, b_z \in (0,1)$ and $c > 0$ so that

$$\mathbf{P}(\inf_{z \in Comp_{\mathbf{C}}(a_z, b_z)} \|Bz\| \leq c\sqrt{n} \text{ and } \|B\|_{op} \leq M\sqrt{n}) \leq e^{-cn} \tag{2.4}$$

Proof. Let $\mathcal{N}$ be a $(2b_z)$-net of $\mathrm{Comp}_{\mathbf{C}}(a_z, b_z)$ as constructed in Lemma 2.2.2. Suppose the event in (2.4) occurred. Then we have some z with $\|Bz\|_2 \leq c\sqrt{n}$ and a sparse vector $z' \in \mathcal{N}$ such that $\|z - z'\|_2 \leq 2b_z$. Using the assumption that $\|B\|_{op} \leq M\sqrt{n}$, we have

$$\begin{aligned}
\|Bz'\|_2 &\leq \|Bz - Bz'\|_2 + \|Bz\|_2 \\
&\leq \|B\|_{op} \|z - z'\|_2 + \|Bz\|_2 \\
&\leq M\sqrt{n}2b_z + c\sqrt{n} \\
&\leq (2Mb_z + c)\sqrt{n}
\end{aligned}$$

For a fixed vector z', we have from the previous proposition for some small $c' > 0$

$$\mathbf{P}(\|Bz'\|_2 \leq (2Mb_z + c)\sqrt{n}) \leq e^{-c'n}$$

by picking b_z small enough depending on M and c small enough. By the union bound over all possible $z' \in \mathcal{N}$, we thus have

$$\mathbf{P}\Big(\inf_{z \in \mathrm{Comp}_{\mathbf{C}}(a_z, b_z)} \|Bz\|_2 \leq c\sqrt{n} \text{ and } \|B\|_{op} \leq M\sqrt{n}\Big) \leq \binom{n}{a_z n} \left(\frac{C}{b_z^2}\right)^{a_z n} e^{-c'n}$$

The conclusion follows by picking a_z sufficiently small depending on b_z and c'. □

Since $Bz = 0$ implies $\|Bz\|_2 \leq c\sqrt{n}$, the previous result shows that compressible vectors are rarely in the kernel of B. a_z and b_z will be fixed in the following so that the above result holds. Whenever we refer to $z \in S_{\mathbf{C}}^{n-1}$ being compressible or incompressible, it will be with these parameters.

2.2.2 Reduction to distance problem

We now take a diversion from null vectors to reduce the least singular value tail estimate to controlling the distance between a fixed column of A and the span of the other $n-1$ columns. From the same argument as in Proposition 2.2.4, we have

$$\mathbf{P}\Big(\inf_{z \in \mathrm{Comp}} \left\|(A - \lambda\sqrt{n})z\right\| \leq tn^{-1/2} \text{ and } \|A\|_{op} \leq M\sqrt{n}\Big) \leq e^{-cn}$$

Using the fact that

$$s_n(X) \leq t \Leftrightarrow \inf_{z \in S^{n-1}} \|Xz\| \leq t$$

and decomposing the unit sphere $S_{\mathbf{C}}^{n-1} = \mathrm{Comp} \sqcup \mathrm{Incomp}$, we see that it suffices to prove

$$\mathbf{P}(\inf_{z \in \mathrm{Incomp}} \left\|(A - \lambda\sqrt{n})z\right\| \leq tn^{-1/2} \text{ and } \|A\|_{op} \leq M\sqrt{n}) \leq O\left(\frac{t^2}{\delta}\right) + e^{-cn} \tag{2.5}$$

in order to establish Theorem 2.1.1.

A form of the following Lemma was given in [RV08]:

Lemma 2.2.5. *(Invertibility via distance)*

Let $X_1, \ldots, X_n$ denote the column vectors of $A - \lambda\sqrt{n}$ and H_k denote the span of all column vectors except X_k. Then for every $a, b \in (0,1)$ and $t > 0$, we have

$$\mathbf{P}(\inf_{z \in \mathit{Incomp}(a,b)} \left\|(A - \lambda\sqrt{n})z\right\| < tbn^{-1/2} \textit{ and } \|A\|_{op} \leq M\sqrt{n})$$
$$\leq \frac{1}{an}\sum_{k=1}^{n} \mathbf{P}(\mathrm{dist}(X_k, H_k) < t \textit{ and } \|A\|_{op} \leq M\sqrt{n})$$

In Appendix A we prove a modified form extended to include the second smallest singular value that is necessary for our later eigenvalue spacing application.

Since the quantities $\mathbf{P}(\mathrm{dist}(X_k, H_k) < t \text{ and } \|A\|_{op} \leq M\sqrt{n})$ are symmetrical in k, to prove (2.5) it suffices to establish the following bound:

$$\mathbf{P}(\mathrm{dist}(X_n, H_n) < t \text{ and } \|A\|_{op} \leq M\sqrt{n}) \leq O\left(\frac{t^2}{\delta}\right) + e^{-cn} \tag{2.6}$$

2.2.3 Two-dimensionality of incompressible random normal

We now return back to null vectors of B. Recall that we are assuming our real random matrix is shifted by a complex λ with imaginary part equal to δ, which is bounded below. We next show that incompressible $z = x + iy$ in the kernel of B cannot be "almost real", i.e. z cannot be rotated so that either its real or imaginary part vanishes. Geometrically, this means that the coordinates of z when plotted in the complex plane cannot lie on a line through the origin.

Lemma 2.2.6. *(Two-dimensionality of incompressible random normal)*

Suppose $z \in S_{\mathbf{C}}^{n-1}$ is incompressible with $Bz = 0$ and $\|B\|_{op} \leq Mn^{1/2}$. Let $z = x + iy$ with $x, y \in \mathbf{R}^n$. We then have $\|x\|_2 \geq c\delta$ and $\|y\|_2 \geq c\delta$ for some small constant $c > 0$

(depending on a_z, b_z). Note that this applies for any rotation $e^{i\theta}z$ also, since $Bz = 0$ if and only if $B(e^{i\theta}z) = 0$. Thus we have $\|\cos(\theta)x - \sin(\theta)y\|_2 \geq c\delta$ for any θ.

Proof. Let R be the real part of B. The real part of the equation $Bz = 0$ is

$$Rx = \delta\sqrt{n}y'$$

where y' denotes the first $n - m$ entries of y. Solving for y' gives:

$$y' = \frac{Rx}{\delta\sqrt{n}}$$

Now we use the fact that $z = x + iy$ is a unit vector to get:

$$\begin{aligned}
1 &= \|x\|_2^2 + \|y\|_2^2 \\
&= \|x\|_2^2 + \|y'\|_2^2 + \sum_{k=m+1}^{n} |y_k|^2 \\
&= \|x\|_2^2 + \left\|\frac{Rx}{\delta\sqrt{n}}\right\|_2^2 + \sum_{k=m+1}^{n} |y_k|^2 \\
&\leq \|x\|_2^2 + \left(\frac{M\sqrt{n}}{\delta\sqrt{n}}\right)^2 \|x\|_2^2 + \sum_{k=m+1}^{n} |y_k|^2
\end{aligned}$$

where we've used R being the real part of B and $\|B\|_{op} \leq M\sqrt{n}$ for the last inequality. Rearranging we get

$$\|x\|_2^2 \geq \frac{1 - \sum_{k=m+1}^{n} |y_k|^2}{1 + \frac{M^2}{\delta^2}}$$

We are assuming that z is incompressible, so in particular the sum of the last m coordinates of y, $\sum_{k=m+1}^{n} |y_k|^2$, is bounded above by $1 - b_z^2$. The numerator $1 - \sum_{k=m+1}^{n} |y_k|^2$ is then bounded below by a constant, and hence we can conclude $\|x\|_2 \geq c\delta$ for some small c depending on M and b_z. Repeating the above for the imaginary part of the equation $Bz = 0$ gives the claim for $\|y\|_2$. The properties of z being incompressible and z being in the kernel of B are not affected by a rotation $z \mapsto e^{i\theta}z$, so we get the final assertion. □

2.2.4 Vectors with compressible real part

Proposition 2.2.4 showed that a complex vector that is compressible is unlikely to be in the kernel of B. We next show that it is also unlikely that $z = x + iy$, where $\frac{x}{\|x\|} \in \text{Comp}_{\mathbf{R}}$, is in the kernel of B.

Proposition 2.2.7. *Let $r \in [c\delta, 1/2]$ be a fixed scale for the size of x. We can pick constants a_x, b_x (independently of r) so that if*

$$S_r := \left\{ z = x + iy \in Incomp_{\mathbf{C}}(a_z, b_z) : r < \|x\| \leq 2r, \frac{x}{\|x\|} \in Comp_{\mathbf{R}}(a_x, b_x) \right\}$$

we have

$$\mathbf{P}(\inf_{z \in S_r} \|Bz\| = 0 \text{ and } \|B\|_{op} \leq M\sqrt{n}) \leq e^{-cn}$$

for some small $c > 0$.

Proof. The first case is if $r \geq C\delta$ for some big constant C. Let

$$S'_r := \left\{ x \in \mathbf{R}^n : r < \|x\| \leq 2r, \frac{x}{\|x\|} \in \text{Comp}_{\mathbf{R}}(a_x, b_x) \right\}$$

Since the real part of the equation $Bz = 0$ is $Rx = \delta\sqrt{n}y'$, the event

$$\left\{ Bz = 0 \text{ for some } z \in S_r \text{ and } \|B\|_{op} \leq M\sqrt{n} \right\}$$

implies $\|Rx\| \leq \delta\sqrt{n}$ for some $x \in S'_r$ and $\|R\|_{op} \leq M\sqrt{n}$. Thus it suffices to check

$$\mathbf{P}(\inf_{x \in S'_r} \|Rx\| \leq \delta\sqrt{n} \text{ and } \|R\|_{op} \leq M\sqrt{n}) \leq e^{-cn} \tag{2.7}$$

S'_r has a $O(b_x r)$-net $\mathcal{N}$ of size

$$|\mathcal{N}| \leq \binom{n}{a_x n} \left(\frac{C}{b_x}\right)^{a_x n} \cdot \frac{2}{b_x}$$

This net of S'_r comes from discretizing the typical net of $\text{Comp}_{\mathbf{R}}(a_x, b_x)$, which has size bounded by $\binom{n}{a_x n} \left(\frac{C}{b_x}\right)^{a_x n}$. For each sparse vector s in the net of $\text{Comp}_{\mathbf{R}}(a_x, b_x)$, add $2/b_x$ multiples of s to $\mathcal{N}$ (those with length between r and $2r$).

Now suppose the event in (2.7) occurred. We would then have a vector $x \in S'_r$ with $\|Rx\| \leq \delta\sqrt{n}$. Let $s \in \mathcal{N}$ be the closest element in the net to x, so that $\|s - x\|_2 \leq 2b_x r$. We then have

$$\begin{aligned} \|Rs\| &\leq \|Rs - Rx\| + \|Rx\| \\ &\leq M\sqrt{n}2b_x r + \delta\sqrt{n} \end{aligned}$$

Dividing by r, we get

$$\left\|R\frac{s}{r}\right\| \leq M2b_x\sqrt{n} + \frac{\delta}{r}\sqrt{n} \leq (2Mb_x + \frac{1}{C})\sqrt{n}$$

where we've used the assumption that $r \geq C\delta$. Note that the vector s/r has length at least 1/2. By Proposition 2.2.3, the probability that the above occurs (for a single fixed s) is at most e^{-cn}, picking b_x small enough depending on M and C in $r \geq C\delta$ big enough. Using the union bound and the size of $\mathcal{N}$, we get a bound of e^{-cn} for (2.7) after adjusting a_x small enough.

Now we are left with the case $r \in [c\delta, C\delta]$. In this situation we need to use both the real and imaginary parts of the equation $Bz = 0$:

$$\begin{aligned} Rx &= \delta\sqrt{n}y' \\ Ry &= -\delta\sqrt{n}x' \end{aligned}$$

Suppose that x is $2b_x r$ close to a sparse vector s_x that is supported on the first $a_x n$ coordinates (we will pay an entropy cost of $\binom{n}{a_x n}$ at the end). Let $\mathcal{N}_x$ denote a $(2b_x r)$-net of such x (consisting of sparse vectors supported on the first $a_x n$ coordinates). The cardinality of $\mathcal{N}_x$ may be bounded by

$$\left(\frac{C}{b_x}\right)^{a_x n}\frac{2}{b_x}$$

Next let us freeze the first $a_x n$ columns of B. Since each $s_x \in \mathcal{N}_x$ is supported only on the first $a_x n$ coordinates, $\frac{1}{\delta\sqrt{n}}Rs_x$ is a deterministic vector (first columns of B are frozen) that approximates y'. Let $s_y \in \mathbf{R}^n$ have the first $n - m$ coordinates equal to $\frac{1}{\delta\sqrt{n}}Rs_x$ and the last m coordinates approximate those of y's up to b_x accuracy. The set $\mathcal{N}'$ of such pairs

$s_x + is_y$ has cardinality bounded by

$$\left(\frac{C}{b_x}\right)^{a_x n} \frac{2}{b_x} \frac{1}{(b_x)^m}$$

Now given $x+iy$ such that $Rx = \delta\sqrt{n}y'$, we have $s_x + is_y$ with $\|x - s_x\| \le 2b_x r$ and the first $n-m$ coordinates of s_y equal to $\frac{1}{\delta\sqrt{n}} R s_x$. Let us compute how small $\|s_y - y\|$ can be:

$$\begin{aligned}
\|s_y - y\| &\le \left\| \frac{1}{\delta\sqrt{n}} R(s_x - x) \right\| + m b_x \\
&\le M \frac{\sqrt{n}}{\delta\sqrt{n}} \|s_x - x\| + m b_x \\
&\le M \frac{b_x r}{\delta} + m b_x \\
&\le (M\frac{r}{\delta} + m) b_x
\end{aligned}$$

The mb_x summand comes from the approximation of the last m coordinates. Note that we are assuming r is comparable to δ, so the above is $O(b_x)$. In summary: in the case where $r \in [c\delta, C\delta]$, the event that $Bz = 0$ for some $z = x + iy \in S_r$ implies that there exists $s_x + is_y \in \mathcal{N}'$ such that $\|s_x - x\| = O(b_x r)$ and $\|s_y - y\| = O(b_x)$. Note that since $z = x + iy$ is incompressible and x is compressible, $\|y\|$ is bounded below by (say) $b_z/2$, assuming that b_x is small enough. For this $s_x + is_y$ we have

$$\begin{aligned}
\left\|R s_y + \delta\sqrt{n} s_x'\right\| &= \left\|R s_y + R y - (R y + \delta\sqrt{n} x') + \delta\sqrt{n}(x' - s_x')\right\| \\
&\le \|R(s_y - y)\| + \delta\sqrt{n}\, \|x' - s_x'\| \\
&\le M\sqrt{n} b_x + \delta\sqrt{n} r b_x
\end{aligned}$$

We've used $Ry + \delta\sqrt{n}x' = 0$. We can guarantee that $\|s_y\|$ is bounded below, say by $b_z/4$, by picking b_x small enough. Moreover, the above bound can be made less than $cb_z\sqrt{n}$ by picking b_x small enough. The probability that the above occurs for a single fixed $s_x + is_y$ is e^{-cn} (there is still a $(n-m) \times (n - a_x n)$ random matrix). Multiplying the above by the size of $\mathcal{N}'$ gives a probability bound of

$$\left(\frac{C}{b_x}\right)^{a_x n} \frac{2}{b_x} \frac{1}{(b_x)^m} e^{-cn}$$

The last steps are to take the expectation over the frozen first $a_x n$ columns, and then multiply by the entropy cost from picking $a_x n$ out of n columns, which gives a final probability bound of

$$\binom{n}{a_x n}\left(\frac{C}{b_x}\right)^{a_x n}\frac{2}{b_x}\frac{1}{(b_x)^m}e^{-cn}$$

The above can be made less than $e^{-c'n}$ by picking a_x small enough. □

Note that by Lemma 2.2.6, incompressible $z = x + iy$ in the kernel of B must have $\|x\|$ bounded below by $c\delta$. We are assuming $\delta = \Omega(n^{-C})$ and hence taking the union bound over all dyadic values of r between $c\delta$ and 1 only gives an extra multiplicative factor of $O(\log n)$. This leads to:

Corollary 2.2.8. *We can pick constants a_x, b_x so that if*

$$S := \left\{ z = x + iy \in Incomp_{\mathbf{C}}(a_z, b_z) : \frac{x}{\|x\|} \in Comp_{\mathbf{R}}(a_x, b_x) \right\}$$

we have

$$\mathbf{P}(\inf_{z\in S} \|Bz\| = 0 \text{ and } \|B\|_{op} \le M\sqrt{n}) \le e^{-cn}$$

a_x and b_x will be fixed from now on so that the conclusion of the above Corollary holds. Whenever the compressibility or incompressibility of a real vector is discussed, it will be with these parameters.

2.3 Least common denominator and statement of main structure theorem

To treat the case of incompressible vectors in the kernel of B, we now introduce the notion of least common denominator developed by Rudelson and Vershynin. We give the definition and some useful lemmas from [RV16].

2.3.1 The least common denominator

Definition 2.3.1. Fix $L, \gamma > 0$.

- For vectors $v \in \mathbf{R}^n$, define the least common denominator (LCD) of v as

$$D(v) = D(v; L, \gamma) := \inf \left\{ \theta \in \mathbf{R}_{>0} : \mathrm{dist}(\theta v, \mathbf{Z}^n) < \gamma L \sqrt{\log_+ \frac{\|\theta v\|_2}{L}} \right\}$$

- For matrices $V \in \mathbf{R}^{m \times n}$, define the LCD of V as

$$D(V) = D(V; L, \gamma) := \inf \left\{ \|\theta\|_2 : \theta \in \mathbf{R}^m, \mathrm{dist}(V^T\theta, \mathbf{Z}^n) < \gamma L \sqrt{\log_+ \frac{\|V^T\theta\|_2}{L}} \right\}$$

- For complex vectors $z = x + iy \in \mathbf{C}^n$ where $x, y \in \mathbf{R}^n$, we define the LCD of z to be the LCD of $\begin{pmatrix} x^T \\ y^T \end{pmatrix}$ as an element of $\mathbf{R}^{2 \times n}$.

- For subspaces $E \subseteq \mathbf{R}^n$, define the LCD of E as

$$D(E) = D(E, L) := \inf \{D(v, L) : v \in E, \|v\|_2 = 1\}$$

We can similarly define the LCD of complex subspaces $E \subseteq \mathbf{C}^n$

Remark. We will make use of the fact that $\frac{1}{x}\sqrt{\log_+ x}$ has a maximum of $\frac{1}{\sqrt{2e}}$ at $x = \sqrt{e}$. γ will be used to adjust the value of the maximum according to other parameters.

A vector having low LCD should be interpreted as having rich arithmetic structure. Our primary aim is to show that with high probability, every vector in the kernel of B will have exponentially large LCD.

2.3.2 Main small ball probability theorem

Theorem 2.3.2. *[RV16] Consider a random vector $\xi = (\xi_1, \ldots, \xi_n)$, where ξ_k are iid copies of ξ. Let $V \in \mathbf{R}^{m \times n}$ be a fixed matrix. Then for every $L \geq \sqrt{\frac{8m}{\gamma p}}$, we have*

$$\sup_{x \in \mathbf{R}^m} \mathbf{P}(\|V\xi - x\|_2 \leq t\sqrt{m}) \leq \frac{(CL/\sqrt{m})^m}{\det(VV^T)^{1/2}} \left(t + \frac{\sqrt{m}}{D(V)}\right)^m$$

for all $t \geq 0$. p in the condition on L is from the initial assumptions (2.1), (2.2) on ξ.

Remark. If $\xi_1, \ldots, \xi_n$ are standard normal random variables, then we have

$$\sup_{x \in \mathbf{R}^m} \mathbf{P}(\|V\xi - x\|_2 \leq t\sqrt{m}) \leq O_m\left(\frac{t^m}{\det(VV^T)^{1/2}}\right)$$

for any V and $t \geq 0$. Thus we can think of $1/D(V)$ as the smallest level at which the small ball concentration probabilities for any random variable satisfying ξ's assumptions match those of the gaussian.

In the $m = 1$ case we have:

Corollary 2.3.3. *Consider a random vector $\xi = (\xi_1, \ldots, \xi_n)$, where ξ_k are iid copies of ξ. Let $v = (v_1, \ldots, v_n) \in \mathbf{R}^n$ be a fixed vector. Then for every $L \geq \sqrt{\frac{8}{\gamma p}}$, we have*

$$\sup_{x \in \mathbf{R}} \mathbf{P}(|\xi_1 v_1 + \cdots + \xi_n v_n - x| \leq t) \leq \frac{CL}{\|v\|_2} \left(t + \frac{1}{D(v)} \right)$$

for any $t \geq 0$.

The small ball probability theorem controls concentration of quantities of the form $\xi_1 v_1 + \ldots + \xi_n v_n = \xi \cdot v$. We would like to control $\|Bz\|$. For each row of B dotted with z, we can use the small ball probability theorem. The next lemma from [RV16] tensorizes the bounds on individual rows together to control $\|Bz\|$.

Lemma 2.3.4. *(Tensorization)*

Let $Z = (Z_1, \ldots, Z_n)$ be a random vector in $\mathbf{C}^n$ with independent coordinates. Assume that there exists numbers $t_0, M \geq 0$ such that

$$\sup_{u \in \mathbf{C}} \mathbf{P}(\|Z_j - u\|_2 \leq t) \leq M(t + t_0)$$

for all j and $t \geq 0$. Then for some absolutely constant $C > 0$,

$$\sup_{u \in \mathbf{C}^n} \mathbf{P}(\|Z - u\|_2 \leq t\sqrt{n}) \leq [CM(t + t_0)]^n$$

for all $t \geq 0$.

An important fact relating the LCD to compressibility is the following:

Proposition 2.3.5. *(LCD of real incompressible vectors)*

Let $x \in S^{n-1}_{\mathbf{R}}$ be incompressible with fixed parameters $a_x, b_x > 0$. Then if we pick γ to be a small multiple of b_x, we have $D(x; L, \gamma) \geq \frac{1}{2}\sqrt{a_x n}$.

Proof. Let $\sigma := \left\{k : |x_k| \leq \frac{1}{\sqrt{a_x n}}\right\}$. Since x is a unit vector, we must have $|\sigma| \geq n - a_x n$. Let x_σ denote the projection of x onto the coordinates in σ. Since $|\sigma| \geq n - a_x n$ and x is incompressible, we must have $\|x_\sigma\|_2 \geq b_x$. By definition of the LCD, we have some $p \in \mathbf{Z}^n$ such that

$$\|D(x)x - p\|_2 < \gamma L \sqrt{\log_+ \frac{\|D(x)x\|_2}{L}}$$

which implies via projection to the σ coordinates

$$\|D(x)x_\sigma - p_\sigma\|_2 \leq \|D(x)x - p\|_2 < \gamma L \sqrt{\log_+ \frac{\|D(x)x\|_2}{L}}$$

If we assume that $D(x) < \frac{1}{2}\sqrt{a_x n}$, then $|D(x)x_k| < \frac{1}{2}$ for each $k \in \sigma$, and hence $p_\sigma = 0$ is the optimal choice to minimize distance to the integer lattice. This gives

$$\|D(x)x_\sigma\|_2 < \gamma L \sqrt{\log_+ \frac{\|D(x)x\|_2}{L}}$$

The left hand side is at least $D(x)b_x$. We can insure that the above inequality is never satisfied by picking γ to be a small constant multiple of b_x. This lets us conclude that $D(x) \geq \frac{1}{2}\sqrt{a_x n}$. □

γ will be fixed from now on, so that the conclusion of the above proposition holds.

2.3.3 Statement of main structure theorem and level sets

Theorem 2.3.6. *(Structure theorem for kernels of random matrices)*

Let B be as before. There exists a small constant $c > 0$ such that taking γ to be a small constant and $L = \sqrt{\frac{8m}{\gamma p}}$ in the definition of LCD, we have:

$$\mathbf{P}(D(\ker B, L) \leq D_0 \text{ and } \|B\|_{op} \leq M\sqrt{n}) \leq e^{-cn}$$

where

$$D_0 = \min(c\sqrt{n}e^{cn/(2m+1)}, Le^{cn/(\gamma L)^2}) \tag{2.8}$$

The proof of Theorem 2.3.6 will occupy the remainder of this section and the next two sections. The precise form of D_0 comes from technical aspects of the proof. The theorem states that the LCD of a vector in $\ker B$ is typically at least exponentially (D_0) big.

2.3.4 Small ball probabilities depending on LCD and real-imaginary correlation

Due to the fact that we are considering complex unit vectors, we need to introduce another characteristic that will help us construct small nets in the next section.

Definition 2.3.7. (Real-imaginary correlation)

For $z = x + iy \in S_{\mathbf{C}}^{n-1}$, let $V = \begin{pmatrix} x^T \\ y^T \end{pmatrix}$ and define

$$d(z) := \det(VV^T)^{1/2} = (\|x\|_2^2 \|y\|_2^2 - (x \cdot y)^2)^{1/2}$$

The real-imaginary correlation $d(z)$ is related to the minimal length of the real part of any rotation $e^{i\theta}z$. The calculation can be found in Proposition B.0.1 in the appendix. From Lemma 2.2.6 we then have:

Corollary 2.3.8. *If $z = x + iy$ satisfies the assumptions of Lemma 2.2.6, then*

$$d(z) \geq c\delta$$

for some small $c > 0$.

Proof. Combining Proposition B.0.1 and Lemma 2.2.6 gives

$$\frac{1}{2} - \frac{1}{2}\sqrt{1 - 4d(z)^2} \geq (c\delta)^2$$

for some small $c > 0$. After rearranging, the above implies

$$d(z)^2 \geq (c\delta)^2 - (c\delta)^4$$

and thus the result holds (with an adjusted value of c). □

We will be considering the two-dimensional LCD of $z = x + iy$, and hence the length of x becomes key. The elements $z = x + iy$ of the unit sphere will be partitioned according to the length of x, the LCD $D(z)$, and the real-imaginary correlation $d(z)$. Note that by Lemma 2.2.6, we may assume $\|x\| \geq c\delta$.

Definition 2.3.9. (Level sets)

Fix a scale $r \in [c\delta, 1]$ for $\|x\|$. Take $D \in [c\sqrt{n}/r, D_0]$ as a scale for the LCD and let $d_0 = C \cdot \max(\frac{\gamma L}{D}\sqrt{\log_+ \frac{Dr}{L}}, \frac{\sqrt{n}r}{D})$.

- (Genuinely complex z)

  For $d_0 \leq d \leq 1$ define

$$S_{D,d,r} := \{z = x + iy \in \text{Incomp} : r \leq \|x\| \leq 2r, D_2(z) = D(x), \\ \frac{x}{\|x\|} \in \text{Incomp}, D \leq D_2(z) \leq 2D, d \leq d(z) \leq 2d\}$$

- (Essentially real z)

  Define

$$S_{D,d_0,r} := \{z = x + iy \in \text{Incomp} : r \leq \|x\| \leq 2r, D_2(z) = D(x), \\ \frac{x}{\|x\|} \in \text{Incomp}, D \leq D(z) \leq 2D, d(z) \leq d_0\}$$

Note that we have different parameters (a_z, b_z) and (a_x, b_x) for the incompressibility of complex z and real $\frac{x}{\|x\|}$ respectively. We are able to assume the two-dimensional LCD of z is exhibited by the real part x since there is freedom to rotate by a phase $e^{i\theta}$. More details are provided when we wrap up the proof of Theorem 2.3.6. We can assume $D \geq c\sqrt{n}/r$ by Corollary 2.2.8 and Proposition 2.3.5. To close the current section we state improved versions of Proposition 2.2.3 given additional arithmetic constraints on the vector z, namely z residing in one of the level sets above.

Proposition 2.3.10. *(Genuinely complex z)*

For any single fixed $z \in S_{D,d,r}$, we have

$$\mathbf{P}(\|Bz\|_2 \leq t\sqrt{n-m}) \leq \left(\frac{C}{d}(t + \frac{1}{D})^2\right)^{n-m}$$

for some absolute constant $C > 0$

Proof. Consider a row B_j of B and the quantity $B_j z$. We may write $B_j z = \xi_j x + i\xi_j y + a_j$ where a_j is some deterministic complex number and ξ_j is the jth (random) column of A.

Theorem 2.3.2 then gives

$$\begin{aligned}\mathbf{P}(\|B_j z\|_2 \le t\sqrt{2}) &\le \frac{(CL/\sqrt{m})^2}{d(z)}\left(t + \frac{\sqrt{2}}{D_2(z)}\right)^2 \\ &\le \frac{C}{d}(t + \frac{1}{D})^2\end{aligned}$$

where we have used the bounded ranges for $d(z)$ and $D_2(z)$ in the definition of $S_{D,d,r}$. Using the above for each of the $n-m$ rows and tensorization Lemma 2.3.4 gives the claimed result (with an adjusted $C > 0$). □

For $z = x + iy$ in the essentially real level sets $S_{D,d_0,r}$, we have similarly:

Proposition 2.3.11. *(Essentially real z)*

For any single fixed $z = x + iy \in S_{D,d_0,r}$, we have

$$\mathbf{P}(\left\|Rx - \delta\sqrt{n}y'\right\| \le t\sqrt{n-m}) \le \left(\frac{C}{r}(t + \frac{1}{D})\right)^{n-m}$$

for some absolute constant $C > 0$

Proof. This follows from Corollary 2.3.3 and an application of tensorization Lemma 2.3.4 analogous to that in the previous Proposition. □

2.4 Construction of nets for level sets

The next step is to discretize the level sets using the LCD in order to construct suitably small nets. We start with the genuinely complex case.

2.4.1 Genuinely complex case

Theorem 2.4.1. *Let $\mu > 0$ be a fixed small constant. There exists a $(4\frac{\mu\sqrt{n}}{D})$-net of $S_{D,d,r}$ with cardinality bounded by*

$$\frac{C^{2n}D^{2n+1}d^{n-1}}{\mu^{n+1}\sqrt{n}^{2n+1}}$$

Proof. By the definition of LCD and the assumption $D_2(z) = D(x)$, we have $p \in \mathbf{Z}^n$ such that

$$\|D_2(z)x - p\|_2 < \gamma L \sqrt{\log_+ \frac{\|D(z)x\|_2}{L}}$$

This gives that the length of p is comparable to Dr:

$$\begin{aligned}\|p\|_2 &\leq \|D(z)x\| + \frac{\gamma L}{\|D(z)x\|}\sqrt{\log_+ \frac{\|D(z)x\|}{L}} \cdot \|D(z)x\| \\ &\leq 2D2r + cDr \\ &\leq CDr\end{aligned}$$

and similarly:

$$\begin{aligned}\|p\|_2 &\geq \|D(z)x\| - \frac{\gamma L}{\|D(z)x\|}\sqrt{\log_+ \frac{\|D(z)x\|}{L}} \cdot \|D(z)x\| \\ &\geq Dr - cDr \\ &\geq cDr\end{aligned}$$

Next we will deduce a dependence of y on p that will help to approximate y. By definition, $d(z) = \det\left(\begin{pmatrix} x^T \\ y^T \end{pmatrix}\begin{pmatrix} x & y \end{pmatrix}\right) = s_1(V)s_2(V)$, where $s_1 \geq s_2$ are the singular values of V. Since z is a unit vector, one of x or y must have length at least $\frac{1}{\sqrt{2}}$. Since x, y are the rows of V, we must have $s_1(V) \geq \frac{1}{\sqrt{2}}$. This gives the following upper bound on $s_2(V)$:

$$s_2(V) \leq d(z)/s_1(V) \leq 2\sqrt{2}d \tag{2.9}$$

Now consider the matrix $W = \begin{pmatrix} p^T \\ D(z)y^T \end{pmatrix}$ and the difference in operator norms

$$\|D(z)V - W\|_{op} = \left\|\begin{pmatrix} D(z)x^T - p^T \\ 0 \end{pmatrix}\right\|_{op} \leq \gamma L \sqrt{\log_+ \frac{D(z)x}{L}}$$

By Weyl's inequalities for singular values and (2.9), we can thus deduce

$$s_2(W) \leq s_2(D(z)V) + \gamma L\sqrt{\log_+ \frac{D(z)x}{L}} \leq 4\sqrt{2}Dd + \gamma L \sqrt{\log_+ \frac{D(z)x}{L}}$$

Expressing $\det(WW^T)^{1/2}$ in two ways, we have the identity

$$\|p\|_2 \cdot \left\|P_{p^\perp} D(z)y\right\| = \det(WW^T)^{1/2} = s_1(W)s_2(W)$$

$s_1(W) \leq \|p\| + \|D(z)y\|$ and hence we have

$$\left\|P_{p^\perp} D(z)y\right\| \leq (1 + \frac{\|D(z)y\|}{\|p\|})s_2(W)$$

Using $\|p\| \geq cDr$, $\|D(z)y\| \leq 2D$, and the above bound for $s_2(W)$, we get

$$\left\|P_{p^\perp} D(z)y\right\| \leq (1+\frac{4}{r})(4\sqrt{2}Dd+\gamma L\sqrt{\log_+ \frac{\|D(z)x\|}{L}}) \leq C(\frac{Dd}{r}+\frac{\gamma L}{r}\sqrt{\log_+ \frac{\|D(z)x\|}{L}}) \quad (2.10)$$

Now from the "genuinely complex" assumption

$$d \geq \frac{\gamma L}{D}\sqrt{\log_+ \frac{\|Dr\|}{L}}$$

we get that $\frac{Dd}{r}$ dominates in (2.10). Dividing by $D(z)$, we get the following for the component of y that is not along p:

$$\left\|P_{p^\perp} y\right\| \leq \frac{Cd}{r}$$

Now we are ready to construct the $O(\frac{\mu\sqrt{n}}{D})$-net of $S_{D,d,r}$. Given $x + iy \in S_{D,d,r}$, we have $p \in \mathbf{Z}^n \cap B(0, CDr)$ such that $p/D(z)$ approximates x well, namely:

$$\left\|x - \frac{p}{D(z)}\right\| < \frac{\gamma L}{D(z)}\sqrt{\log_+ \frac{\|D(z)x\|_2}{L}} \leq \frac{\mu\sqrt{n}}{D}$$

where the second inequality comes from the second expression in (2.8). For each $p \in \mathbf{Z}^n \cap B(0, CDr)$, we will work with the $O(\frac{D}{\mu\sqrt{n}})$ discrete multiples αp that approximate $p/D(z)$ up to $\frac{\mu\sqrt{n}}{D}$ accuracy. Counting the number of lattice points in $B(0, CDr)$ along with this extra discretization factor, we get a bound of

$$\frac{CD}{\mu\sqrt{n}}\left(\frac{CDr}{\sqrt{n}}\right)^n \quad (2.11)$$

for the total number of αp we have to consider. For each αp, we have

$$\left\|P_{p^\perp} y\right\| \leq \frac{Cd}{r}$$

and hence y has to lie in a cylinder in the direction of p. The typical volume bound gives a $\frac{\mu\sqrt{n}}{D}$-net of all possible y with size bounded by

$$\frac{CD}{\mu\sqrt{n}}\left(\frac{CDd}{\mu\sqrt{n}r}\right)^{n-1} \tag{2.12}$$

Note that we have $d/r = \Omega(\frac{\sqrt{n}}{D})$ from the second expression in the definition of d_0. The first $\frac{CD}{\mu\sqrt{n}}$ is from approximating in the one dimension along p. Combining (2.11) and (2.12), we see that our net for $S_{D,d,r}$ has the claimed cardinality bound. □

2.4.2 Essentially real case

Theorem 2.4.2. *Let $\mu > 0$ be a fixed small constant. There exists a set $\mathcal{N}$ with cardinality bounded by*

$$\frac{C^{2n+1}D^{n+2}r^n}{\mu^2\sqrt{n}^{n+2}}$$

such that for every $z = x + iy \in S_{D,d_0,r}$, there is $u + iv \in \mathcal{N}$ such that $\|x-u\| \leq 2\frac{\mu\sqrt{n}}{D}$ and $\|y - v\| \leq \frac{\mu\sqrt{n}}{Dr}$.

Remark. The approximation in this theorem is different than before, but it is necessary to maintain a small size for the net. The rougher approximation for y will be not be problematic due to the equation $Rx = \delta\sqrt{n}y'$.

Proof. There are two sub cases, due to d_0 being the max of two expressions. Suppose

$$d(z) < C\frac{\gamma L}{D}\sqrt{\log_+\frac{Dr}{L}}$$

Much of the analysis remains the same. In particular, we will not need to change any of the work done in the genuinely complex case for p. The bound for the portion of y along p is changed. The second term now dominates in (2.10) and hence we have

$$\left\|P_{p^\perp}y\right\| \leq \frac{C\gamma L}{Dr}\sqrt{\log_+\frac{Dr}{L}} \leq \frac{\mu\sqrt{n}}{Dr}$$

where as before $D \leq D_0$ and the second expression in (2.8). In this case, we see that using the same αp's as in the genuinely complex case, for each αp, we need to pair it with at most

$$\frac{CDr}{\mu\sqrt{n}}C^{n-1}$$

y's to approximate up to $\frac{\mu\sqrt{n}}{Dr}$ accuracy, and hence the total size of the approximating set is bounded by

$$\frac{CD}{\mu\sqrt{n}}\left(\frac{CDr}{\sqrt{n}}\right)^n \frac{CDr}{\mu\sqrt{n}} C^{m-1}$$

The last case is if

$$d(z) \geq C\frac{L}{D}\sqrt{\log_+ \frac{Dr}{L}} \text{ and } d(z) < C\frac{\sqrt{n}r}{D}$$

In this case the only thing that changes from the genuinely complex case is that instead of (2.12), the volume bound gives a $\frac{\mu\sqrt{n}}{D}$-net of all possible y with size bounded by

$$\frac{CD}{\mu\sqrt{n}} C^{m-1}$$

In this case we have the same approximation as in the genuinely complex case and a total bound of

$$\frac{CD}{\mu\sqrt{n}}\left(\frac{CDr}{\sqrt{n}}\right)^n \frac{CDr}{\mu\sqrt{n}} C^{m-1}$$

for the cardinality. In all cases, we have the claimed cardinality bound and approximation. □

2.5 Combining small ball probability bounds and nets to finish proof of main structure theorem

In this section we combine the small ball probability bounds from the end of Section 3 and nets constructed in the previous section to prove that B is unlikely to have a null vector in each of the level sets.

2.5.1 Genuinely complex case

Theorem 2.5.1. *(Genuinely complex case)*

$$\mathbf{P}(\inf_{z \in S_{D,d,r}} \|Bz\| = 0 \text{ and } \|B\|_{op} \leq M\sqrt{n}) \leq e^{-cn}$$

Proof. From Theorem 2.4.1, we have a $(4\frac{\mu\sqrt{n}}{D})$-net $\mathcal{N}$ of $S_{D,d,r}$ of cardinality at most

$$\frac{C^{2n}D^{2n+1}d^{n-1}r}{\mu^{n+1}\sqrt{n}^{2n+1}}$$

Now suppose that $Bz = 0$ for some $z = x + iy \in S_{D,d,r}$. Let $z' \in \mathcal{N}$ be such that $\|z - z'\| \leq \frac{4\mu\sqrt{n}}{D}$. From Proposition 2.3.10, we have

$$\mathbf{P}(\|Bz'\|_2 \leq t\sqrt{n-m}) \leq \left(\frac{C}{d}(t + \frac{1}{D})^2\right)^{n-m} \tag{2.13}$$

Now suppose that $\|Bz\| = 0$ occurred for some $z \in S_{D,d,r}$ (and $\|B\|_{op} \leq M\sqrt{n}$), and $z' \in \mathcal{N}$ satisfies $\|z - z'\| \leq \frac{4\mu\sqrt{n}}{D}$, we then have

$$\|Bz'\| = \|Bz' - Bz\| \leq \|B\|\,\|z - z'\| \leq M\sqrt{n}\frac{\mu\sqrt{n}}{D} \leq \frac{M\mu n}{D} \leq \frac{\nu\sqrt{n}\sqrt{n-m}}{D}$$

where the last inequality is by picking μ to be a small constant multiple of ν. Taking $t = \frac{\nu\sqrt{n}}{D}$ in (2.13) and using the union bound over $\mathcal{N}$ we get:

$$\begin{aligned}\mathbf{P}(\inf_{z\in S_{D,d,r}} \|Bz\| = 0 \text{ and } \|B\|_{op} \leq M\sqrt{n}) &\leq \mathbf{P}(\inf_{z'\in\mathcal{N}} \|Bz'\|_2 \leq \frac{\nu\sqrt{n}\sqrt{n-m}}{D}) \\ &\leq \frac{C^{2n}D^{2n+1}d^{n-1}r}{\mu^{n+1}\sqrt{n}^{2n+1}} \frac{C^{n-m}(\nu\sqrt{n})^{2(n-m)}}{d^{n-m}D^{2(n-m)}} \\ &\leq \frac{C^{3n-m}D^{2m+1}d^{m-1}r}{\sqrt{n}^{2m+1}} \frac{\nu^{2(n-m)}}{\mu^{n+1}} \\ &\leq C_1^n \frac{\nu^{2(n-m)}}{\mu^{n+1}}\end{aligned}$$

The last inequality is by our assumption $D \leq D_0$ and (2.8). $n+1$ copies of ν will cancel with the μ's in the denominator (and leave another C^{n+1}, since we picked μ to be a small multiple of ν). This leaves ν^{n-m-1} that we can adjust as small as necessary depending on C_1 so that the total probability is bounded by e^{-cn}. □

2.5.2 Essentially real case

Theorem 2.5.2. *(Essentially real case)*

$$\mathbf{P}(\inf_{z\in S_{D,d_0,r}} \|Bz\| = 0 \textit{ and } \|B\|_{op} \leq M\sqrt{n}) \leq e^{-cn}$$

Proof. Let $z = x + iy \in S_{D,d_0,r}$ and suppose that we are in the event $Bz = 0$ and $\|B\|_{op} \le M\sqrt{n}$. The real part of the equation $Bz = 0$ says

$$Rx = \delta\sqrt{n}y'$$

Theorem 2.4.2 gives a set $\mathcal{N}$ with cardinality bounded by

$$\frac{C^{2n+1}D^{n+2}r^n}{\mu^2\sqrt{n}^{n+2}}$$

such that for every $z = x + iy \in S_{D,d_0,r}$, there is $u + iv \in \mathcal{N}$ such that $\|x - u\| \le 2\frac{\mu\sqrt{n}}{D}$ and $\|y - v\| \le \frac{\mu\sqrt{n}}{Dr}$. We thus have

$$\begin{aligned}
\left\|Ru - \delta\sqrt{n}v'\right\| &= \left\|Ru - Rx + \delta\sqrt{n}(y' - v')\right\| \\
&\le \|R\|\,\|u - x\| + \delta\sqrt{n}\,\|y' - v'\| \\
&\le M\sqrt{n}\frac{2\mu\sqrt{n}}{D} + \delta\sqrt{n}\frac{\mu\sqrt{n}}{Dr} \\
&\le O(\frac{\mu n}{D})
\end{aligned}$$

where for the last inequality we've used $r \ge c\delta$. From Proposition 2.3.11 we have

$$\mathbf{P}(\left\|Ru - \delta\sqrt{n}v'\right\| \le t\sqrt{n-m}) \le \left(\frac{Ct}{r}\right)^{n-m} \tag{2.14}$$

Picking μ to be a small multiple of ν again, we see that $Bz = 0$ and $\|B\|_{op} \le M\sqrt{n}$ implies that there is some $u + iv \in \mathcal{N}$ such that

$$\left\|Ru - \delta\sqrt{n}v'\right\| \le \frac{\nu\sqrt{n}\sqrt{n-m}}{D}$$

Taking $t = \frac{\nu\sqrt{n}}{D}$ in (2.14) and using the union bound over $\mathcal{N}$ we get:

$$\begin{aligned}
\mathbf{P}(\inf_{z \in S_{D,d_0,r}} \|Bz\| = 0 \text{ and } \|B\|_{op} \le M\sqrt{n}) &\le \mathbf{P}(\inf_{u+iv \in \mathcal{N}} \left\|Ru - \delta\sqrt{n}v'\right\| \le \frac{\nu\sqrt{n}\sqrt{n-m}}{D}) \\
&\le \frac{C^{2n+1}D^{n+2}r^n}{\mu^2\sqrt{n}^{n+2}}\left(\frac{C\nu\sqrt{n}}{Dr}\right)^{n-m} \\
&\le \frac{C_1^n D^{m+2}r^m}{\sqrt{n}^{m+2}}\frac{\nu^{n-m}}{\mu^2} \\
&\le C_2^n\frac{\nu^{n-m}}{\mu^2}
\end{aligned}$$

We have again used the bound $D \le D_0$ to simplify. The result now follows by taking ν small enough so that the above is bounded by e^{-cn}. □

2.5.3 Proof of Theorem 2.3.6

We are now ready to prove the main structure theorem by combining the small ball probability estimates and discrete nets from the previous sections.

Proof. First fix a level $1 \leq D \leq D_0$ and define the level set

$$S_D := \{z \in \text{Incomp} : D \leq D_2(z) \leq 2D\}$$

Note that $Bz = 0$ is equivalent for $B(e^{i\theta}z) = 0$ for any phase θ. Moreover, the LCD is also rotation invariant, i.e. $D_2(z) = D_2(e^{i\theta}z)$ for any phase θ. Thus we may assume that the LCD of z is exhibited by the real part x. More precisely, the event $Bz = 0$ for some $z \in S_D$ is equivalent to $Bz = 0$ for some $z \in S'_D$, where

$$S'_D := \{z = x + iy \in \text{Incomp} : D \leq D_2(z) \leq 2D, D_2(z) = D(x)\}$$

We can decompose S'_D into the previously defined genuinely complex and essentially real level sets $S_{D,d,r}$ and $S_{D,d_0,r}$ for different dyadic levels of d and r. Note that since z is incompressible, by Lemma 2.2.6 and Corollary 2.3.8 and how we've restricted $\delta = \Omega(n^{-C})$, there are only $\log n$ dyadic values of d and r we have to range over. We also need to exclude the vectors with compressible real part in the result of Corollary 2.2.8. Let us denote that set S_{comp}. Using the results of Corollary 2.2.8, Theorem 2.5.1, and Theorem 2.5.2, we have:

$$\begin{aligned}
\mathbf{P}(\inf_{S_D} \|Bz\| = 0 \text{ and } \|B\|_{op} \leq M\sqrt{n}) &= \mathbf{P}(\inf_{S'_D} \|Bz\| = 0 \text{ and } \|B\|_{op} \leq M\sqrt{n}) \\
&\leq \mathbf{P}(\inf_{z \in S_{comp}} \|Bz\| = 0 \text{ and } \|B\|_{op} \leq M\sqrt{n}) \\
&+ \sum_r \sum_d \mathbf{P}(\inf_{z \in S_{D,d,r}} \|Bz\| = 0 \text{ and } \|B\|_{op} \leq M\sqrt{n}) \\
&\leq e^{-cn} + O(\log n)^2 e^{-cn} \\
&\leq e^{-c'n}
\end{aligned}$$

The final step is to vary over the dyadic values of D in the range $[1, D_0]$. Since D_0 is exponential in n, this incurs a multiplicative $O(n)$ cost. Adjusting the constant c again, we get the conclusion of Theorem 2.3.6 $\square$

2.6 Distance problem via small ball probability and structure theorem

With the structure theorem in hand, we next prove the distance bound (2.6) and complete the proof of the least singular value tail bound Theorem 2.1.1. Let X^* denote a unit vector that is orthogonal to $X_1, \ldots, X_{n-1}$. X^* can be chosen so that it only depends on the first $n-1$ columns, and in particular so that it is independent of X_n. As before there is an ambiguity in the phase of X^* that we make an arbitrary choice of. We have

$$\operatorname{dist}(X_n, H_n) \geq |\langle X^*, X_n\rangle|$$

and hence it suffices to show

$$\mathbf{P}(|\langle X^*, X_n\rangle| < t \text{ and } \|A\|_{op} \leq M\sqrt{n}) \leq O\left(\frac{t^2}{\delta}\right) + e^{-cn}$$

We will condition on the first $n-1$ columns, so that X^* may be viewed as a fixed vector. The probability of $|\langle X^*, X_n\rangle| < t$ will then be controlled via the small ball probability theorem.

We will next establish some properties of a typical random normal vector X^*. The first is that X^* should be incompressible, which will give us two-dimensionality of the random normal. The second is that the LCD of X^* will be exponentially big, which will give us small ball concentration down to an exponentially small scale for t. Recall D_0 from Theorem 2.3.6 is exponentially big.

Proposition 2.6.1. *(Typical properties of X^*)*

$$\mathbf{P}((X^* \in Comp(a_z, b_z) \text{ or } LCD(X^*) \leq D_0) \text{ and } \|A\|_{op} \leq M\sqrt{n}) \leq e^{-cn} \tag{2.15}$$

Proof. Ruling out compressible vectors follows from a similar argument as Proposition 2.2.4. Theorem 2.3.6 rules out X^* having small LCD. □

Now for a fixed X^* that is incompressible and has large LCD, the small ball probability theorem implies

Proposition 2.6.2. *Let $X^* \in S_{\mathbf{C}}^{n-1}$ be a fixed (complex) unit vector such that $d(X^*) \geq c\delta$ and $D_2(X^*) \geq D_0$. Then:*

$$\mathbf{P}(|\langle X^*, X_n\rangle| < t) \leq O\left(\frac{t^2}{\delta}\right) + e^{-cn}$$

Proof. X_n is a random vector $(\xi_1, \dots, \xi_n)$ shifted by a constant $i\delta e_n$. By Theorem 2.3.2, we then have

$$\begin{aligned}\mathbf{P}(|\langle X^*, X_n\rangle| < t) &\leq O(\frac{1}{d(X^*)}(t + \frac{1}{D_2(X^*)})^2) \\ &\leq O\left(\frac{t^2}{\delta}\right) + e^{-cn}\end{aligned}$$

□

Now we can piece everything together for:

Lemma 2.6.3. *(Distance bound)*

Let $X_1, \dots, X_n$ denote the columns of the random matrix $A - \lambda\sqrt{n}I_n$ and H_n be the hyperplane spanned by the first $n-1$ columns. For every $t \geq 0$, we have

$$\mathbf{P}(\text{dist}(X_n, H_n) < t \text{ and } \|A\|_{op} \leq M\sqrt{n}) \leq O(\frac{t^2}{\delta} + e^{-cn})$$

Proof. We condition on a realization of the first $n-1$ columns $X_1, \dots, X_{n-1}$ such that the normal vector X^* is incompressible and has large LCD. Denote the expectation with respect to the first $n-1$ columns $\mathbf{E}_{1,\dots,n-1}$ and the probability with respect to the last column $\mathbf{P}_n$. Let $\mathcal{E}$ denote the event that X^* satisfies $d(X^*) \geq c\delta$ and $D_2(X^*) \geq D_0$.We then have

$$\begin{aligned}\mathbf{P}(|\langle X^*, X_n\rangle| < \text{ and } \|A\|_{op} \leq M\sqrt{n}) &\leq \mathbf{E}_{1,\dots,n-1}\mathbf{P}_n(|\langle X^*, X_n\rangle| < t \text{ and } \mathcal{E}) + \mathbf{P}(\mathcal{E}^c) \\ &\leq O(\frac{t^2}{\delta} + e^{-cn}) + e^{-cn} \\ &\leq O(\frac{t^2}{\delta} + e^{-cn})\end{aligned}$$

Incompressible X^* must have $d(X^*) \geq c\delta$ by Lemma 2.2.6. That fact lets us control $\mathbf{P}(\mathcal{E}^c)$ to be exponentially small via Proposition 2.6.1. Proposition 2.6.2 gives the bound on $\mathbf{P}_n(|\langle X^*, X_n\rangle| < t \text{ and } \mathcal{E})$ for a fixed realization of the first $n-1$ columns. □

CHAPTER 3

Eigenvalue spacing of iid random matrices

3.1 Theorem statement and proof outline

In this section we outline the approach taken to establish the main eigenvalue spacing results. We will make the same assumptions on the atom random variables ξ as in the previous chapter, in addition to assuming that ξ is centered. A will be used to denote an $n \times n$ matrix with each entry an iid copy of ξ. The main theorem in this chapter is the following:

Theorem 3.1.1. *Let $\lambda_1, \ldots, \lambda_n$ be the eigenvalues of $\frac{1}{\sqrt{n}}A$ and let*

$$\Delta := \min_{i \neq j} |\lambda_i - \lambda_j|$$

be the minimum gap between any two distinct eigenvalues. Let $s > n^{-C}$ be a scale for spacing of polynomial size. We then have the following tail estimate for eigenvalue spacing:

$$\mathbf{P}(\Delta < s) = O(\delta n^{2+o(1)} + \frac{s^2 n^{4+o(1)}}{\delta^2}) + e^{-cn} + \mathbf{P}(\|A\|_{op} > Mn^{1/2})$$

where $\delta > s$ is a scale at which eigenvalues near the real line are to be separated from each other.

The large operator norm term can be controlled via moment assumptions on ξ. For example, imposing a bounded fourth moment implies

$$\mathbf{P}(\|A\|_{op} > Mn^{1/2}) = o(1)$$

and a subgaussian assumption implies that

$$\mathbf{P}(\|A\|_{op} > Mn^{1/2}) \leq 2e^{-n}$$

where the constant M depends on the bound on the fourth moment or subgaussian moment. See [YBK88], [RV08] and the references within. Qualitatively, Theorem 3.1.1 implies that iid random matrices have simple spectrum asymptotically almost surely:

Corollary 3.1.2. *Let A be an $n \times n$ iid random matrix where the atom random variables are centered, have variance at least 1, and have bounded fourth moment. Let Δ be the minimum gap between any two distinct eigenvalues of $\frac{1}{\sqrt{n}}A$. Taking $\delta = n^{-(2+o(1))}$ and $s = \delta n^{-(2+o(1))}$ in Theorem 3.1.1 gives*

$$\mathbf{P}(\Delta < n^{-(4+o(1))}) = o(1)$$

which implies

$$\mathbf{P}(A \text{ has simple spectrum}) = 1 - o(1)$$

We now give an outline of the chapter, starting with a brief sketch of the argument for Theorem 3.1.1. The intermediate goal is to establish a bound for the probability that there are two eigenvalues near a fixed $z \in \mathbf{C}$, i.e.

$$\mathbf{P}(\lambda_i, \lambda_j \in B(z,s) \text{ for } i \neq j)$$

Many applications of least singular value results involve using a bound on

$$\mathbf{P}(s_n(A - z\sqrt{n}) \leq sn^{-1/2})$$

to control or detect the presence of (at least) one eigenvalue near z. We will deduce bounds on the two smallest singular values s_{n-1} and s_n of $\frac{1}{\sqrt{n}}A - z$ from the presence of two eigenvalues near z. In contrast with the Hermitian case, eigenvectors of iid matrices may not be orthogonal. The process involves accounting for the possibility that the associated eigenvectors of $\lambda_i, \lambda_j \in B(z,s)$ are almost parallel, leading to a range of bounds for s_{n-1}, s_n. We then need to use a version of the least singular value bound from the previous chapter extended to give joint tail estimates for s_{n-1} and s_n. The final step is to use a covering argument to deduce Theorem 3.1.1 from the bounds on the event where there are two eigenvalues of $\frac{1}{\sqrt{n}}A$ near a fixed center $z \in \mathbf{C}$. We conclude by stating further eigenvalue spacing results that are obtainable via similar arguments.

3.2 Two eigenvalues near a fixed center

In this section we consider a local version of the eigenvalue spacing problem. Instead of the entire spectrum, we prove bounds on the event that there are two eigenvalues close to a fixed point $z \in \mathbf{C}$. More precisely, the estimates we will establish are:

Proposition 3.2.1.

$$\mathbf{P}(\lambda_i, \lambda_j \in B(z,\delta) \text{ and } \|A\|_{op} \leq M\sqrt{n}) = O(\delta^2 n^{2+o(1)}) + e^{-cn} \tag{3.1}$$

for fixed real $z = O(1)$.

$$\mathbf{P}(\lambda_i, \lambda_j \in B(z,s) \text{ and } \|A\|_{op} \leq M\sqrt{n}) = O(\frac{s^4 n^{4+o(1)}}{\delta^2}) + e^{-cn} \tag{3.2}$$

for fixed complex $z = O(1)$ with imaginary part bounded below by $c\delta$.

3.2.1 Accounting for non-orthogonality of eigenvectors

Fix $z = O(1)$ and suppose λ_i, λ_j are within s of z. We will denote $N := \frac{1}{\sqrt{n}}A - z$. With the assumptions $z = O(1)$ and $\|A\|_{op} \leq M\sqrt{n}$, we then have $\|N\|_{op} = O(1)$. Ideally, the eigenvectors associated to λ_i, λ_j would be orthogonal, and then we would immediately have $s_n(N), s_{n-1}(N) = O(s)$. That may not necessarily be the situation, as A is not constrained to be symmetric. Indeed, [MC98], [MC00] give that eigenvectors of a complex Ginibre matrix become more correlated as their associated eigenvalues get closer. Nevertheless, we can convert the event of two eigenvalues being close to fixed $z = O(1)$ into the following information on orthogonal vectors:

Lemma 3.2.2. *Suppose $\lambda_i, \lambda_j \in B(z,s)$ for some $1 \leq i < j \leq n$ and $\|A\|_{op} \leq M\sqrt{n}$. Then there exists orthogonal unit vectors $v, w \in \mathbf{C}^n$ and $a = O(1)$ such that $Nv = (\lambda_i - z)v$ and $Nw = (\lambda_j - z)w + av$. In particular we have $\|Nv\| = O(s)$ and $\|Nw - av\| = O(s)$*

Proof. First suppose that $\lambda_i \neq \lambda_j$. Let v_i, v_j be eigenvectors corresponding to λ_i, λ_j. Take $v = \frac{v_i}{\|v_i\|}$ and let w be a unit vector orthogonal to v such that v, w span the same plane as

v_i, v_j. We then have

$$\|Nv\| = \left\|(\frac{1}{\sqrt{n}}A - z)v\right\| = \|(\lambda_i - z)v\| = O(s)$$

by our assumption on $|\lambda_i - z|$ and v being a unit vector. Now write $w = a_i v_i + a_j v_j$. Then

$$Nw = a_i(\lambda_i - z)v_i + a_j(\lambda_j - z)v_j = (\lambda_j - z)w + a_i(\lambda_i - \lambda_j)v_i = (\lambda_j - z)w + av$$

taking $a = a_i(\lambda_i - \lambda_j)\|v_i\|$. Arguing as above, we get $\|Nw - av\| = O(s)$ as well. Finally, we have $av = (N - (\lambda_j - z))w$. Since $\|N\|_{op} = O(1)$, $|\lambda_j - z| = O(s)$, and v, w are unit vectors, we have $a = O(1)$. Now suppose $\lambda_i = \lambda_j$. If this repeat eigenvalue has geometric multiplicity 2 or greater, then the above argument still applies since we have distinct eigenvectors v_i, v_j as before. Now suppose $\lambda_i = \lambda_j = \lambda$ and the geometric multiplicity is 1. Using the Jordan canonical form, we have v_i, v_j such that

$$Nv_i = (\lambda - z)v_i \text{ and } Nv_j = (\lambda v_j + 1 \cdot v_i) - zv_j = (\lambda - z)v_j + v_i$$

where the $1 \cdot v_i$ comes from the off-diagonal 1. Writing $w = a_i v_i + a_j v_j$ again, we now have

$$Nw = (\lambda - z)w + a_j v_i$$

and hence we take $a = a_j$ this time and finish as above. □

3.2.2 Reduction to singular values tail estimates

Let $\mathcal{E}$ be the event in the conclusion of Lemma 3.2.2. We will next deduce bounds on the two smallest singular values of N given that $\mathcal{E}$ occurs.

Proposition 3.2.3. *Suppose $\mathcal{E}$ occurs for $a = O(1)$. There are two cases corresponding to the range of a:*

- *If $a = O(s)$, then*

$$s_n(N) = O(s) \text{ and } s_{n-1}(N) = O(s)$$

- *If $a = \Omega(s)$, then*

$$s_n(N) = O\left(\frac{s^2}{|a|}\right) \text{ and } s_{n-1}(N) = O(|a|)$$

Proof. First suppose $a = O(s)$, which is akin to the situation where the eigenvectors v and w are orthogonal. We have $\|Nv\| = O(s)$ and $\|Nw\| \leq \|Nw - av\| + \|av\| = O(s)$. Since v and w are orthogonal, this implies that both $s_n(N) = O(s)$ and $s_{n-1}(N) = O(s)$ as claimed.

Now suppose a is in the range from s to 1, i.e. $a = \Omega(s)$. $s_{n-1}(N) = O(|a|)$ since $\|N\|_{op} = O(|a|)$ on the plane spanned by v and w. For the least singular value, we can bound

$$s_n(N) \leq s_2(N|_{\text{span}(v,w)}) \leq \text{dist}(Nv, \text{span}(Nw))$$

Recall that $Nv = (\lambda_i - z)v$ and $Nw = (\lambda_j - z)w + av$, where v and w are orthogonal. Combined with the fact that $\lambda_i, \lambda_j \in B(z, s)$, we may then bound

$$\text{dist}(Nv, \text{span}(Nw)) \leq \frac{|\lambda_i - z|\,|\lambda_j - z|}{|a|} = O\left(\frac{s^2}{|a|}\right)$$

□

3.2.3 Proof of Proposition 3.2.1

We will use the following extension of Theorem 2.1.1:

Theorem 3.2.4. *For any $t_2 \geq t_1 \geq 0$ and $\lambda = O(1)$, we have*

$$\mathbf{P}(s_n(A - \lambda\sqrt{n}) \leq t_1 n^{-1/2}, s_{n-1}(A - \lambda\sqrt{n}) \leq t_2 n^{-1/2} \text{ and } \|A\|_{op} \leq M\sqrt{n}) \leq O\left(\frac{t_1^2 t_2^2}{\delta^2}\right) + e^{-cn}$$

where δ is the absolute value of the imaginary part of λ.

The proof is given in Appendix A. We will also require the following result:

Theorem 3.2.5. *For any $t_2 \geq t_1 \geq 0$ and real $\lambda = O(1)$, we have*

$$\mathbf{P}(s_n(A - \lambda\sqrt{n}) \leq t_1 n^{-1/2}, s_{n-1}(A - \lambda\sqrt{n}) \leq t_2 n^{-1/2} \text{ and } \|A\|_{op} \leq M\sqrt{n}) \leq O(t_1 t_2) + e^{-cn}$$

Theorem 3.2.5 is not stated explicitly in the literature. Using the same process as in Appendix A applied to the results in [RV09] will give the above statement.

Proof. We are now ready to prove (3.2). Suppose $\|A\|_{op} \leq M\sqrt{n}$ and we have distinct eigenvalues $\lambda_i, \lambda_j \in B(z, s)$ for fixed complex $z = O(1)$ with imaginary part bounded below by $c\delta$. This implies that one of the conclusions in Proposition 3.2.3 occurs. By a union bound over the dyadic possibilities for $|a|$ in the range (cs, C), we have

$$\mathbf{P}(\lambda_i, \lambda_j \in B(z,s) \text{ and } \|A\|_{op} \leq M\sqrt{n}) \leq \mathbf{P}(s_n(N), s_{n-1}(N) = O(s) \text{ and } \|A\|_{op} \leq M\sqrt{n})$$
$$+ \sum_{|a| \in \mathcal{D}} \mathbf{P}(s_n(N) = O\left(\frac{s^2}{|a|}\right), s_{n-1}(N) = O(|a|) \text{ and } \|A\|_{op} \leq M\sqrt{n})$$

where $\mathcal{D}$ denotes the dyadic values in the range (cs, C), with its cardinality bounded by

$$|\mathcal{D}| \leq |\log(s)| \leq O(\log(n))$$

For any of the summands, Theorem 3.2.4 gives a probability bound of

$$O\left(\frac{s^4 n^4}{\delta^2}\right) + e^{-cn}$$

A union bound over the dyadic possibilities for $|a|$ in the range (ct, C) gives the claim in (3.2).

$$\mathbf{P}(\lambda_i, \lambda_j \in B(z,s) \text{ and } \|A\|_{op} \leq M\sqrt{n}) \leq |\log(s)|\,(O\left(\frac{s^4 n^4}{\delta^2}\right) + e^{-cn}) \leq O\left(\frac{s^4 n^{4+o(1)}}{\delta^2}\right) + e^{-cn}$$

An analogous argument using Theorem 3.2.5 gives (3.1).

□

3.3 A covering argument and further corollaries

We next reduce Theorem 3.1.1 to (3.1), (3.2), which will establish the main eigenvalue spacing result.

Proof. Suppose we are outside the event where the operator norm $\|A\|_{op}$ is bigger than $M\sqrt{n}$. Let $\mathcal{B}$ be a ball of $O(1)$ bounded radius that contains all the eigenvalues of $\frac{1}{\sqrt{n}}A$. First cover $\mathbf{R} \cap \mathcal{B}$ with balls of radius δ centered at real z_α such that if $|\lambda_i - \lambda_j| \leq \delta/100$ and either λ_i or λ_j has imaginary part less than $\delta/100$, then there is some α such that $\lambda_i, \lambda_j \in B(z_\alpha, \delta)$. Since $\mathcal{B}$ has $O(1)$ radius, this covering takes $O(\delta^{-1})$ many balls, i.e. $O(\delta^{-1})$ many centers z_α.

Next cover the rest of B with balls of radius s centered at z_β with $|\Im z_\beta| \geq t/1000$ such that if $|\lambda_i - \lambda_j| \leq s/100$ and both λ_i, λ_j have imaginary part at least $\delta/100$ in absolute value, then $\lambda_i, \lambda_j \in B(z_\beta, s)$ for some z_β. This takes $O(s^{-2})$ distinct centers z_β.

Now we have

$$\mathbf{P}(\Delta < s) \leq \mathbf{P}(\|A\|_{op} > Mn^{1/2}) + \sum_\alpha \mathbf{P}(\lambda_i, \lambda_j \in B(z_\alpha, \delta) \text{ and } \|A\|_{op} \leq M\sqrt{n})$$
$$+ \sum_\beta \mathbf{P}(\lambda_i, \lambda_j \in B(z_\beta, s) \text{ and } \|A\|_{op} \leq M\sqrt{n})$$

Combining (3.1), (3.2) along with the entropy bounds of $O(\delta^{-1})$ for the number of α and $O(s^{-2})$ for the number of β gives

$$\mathbf{P}(\Delta < s) \leq O(\delta n^{2+o(1)} + \frac{s^2 n^{4+o(1)}}{\delta^2}) + e^{-cn} + \mathbf{P}(\|A\|_{op} > Mn^{1/2})$$

□

3.3.1 Further results

We now state some eigenvalue spacing results that are readily obtained from arguments similar to those in this chapter. The first is by applying the covering argument only to the real line, as opposed to a complex ball. We can obtain a bound on the event that A has two real eigenvalues within t of each other.

Theorem 3.3.1. *Let A satisfy the same assumptions as in Theorem 3.1.1 and $t > 0$. Let $\mathcal{E}(t)$ denote the event that A has two real eigenvalues within t of each other. Then:*

$$\mathbf{P}(\mathcal{E}(t)) = O(tn^{2+o(1)}) + e^{-cn} + \mathbf{P}(\|A\|_{op} > Mn^{1/2}) \tag{3.3}$$

Since complex eigenvalues of real matrices must come in pairs, we also have the same bound in the following:

Theorem 3.3.2. *Let A satisfy the same assumptions as in Theorem 3.1.1 and $t > 0$. Let $\mathcal{F}(t)$ denote the event that A has a complex conjugate pair of eigenvalues within t of the real* *line. Alternatively, we can characterize $\mathcal{F}(t)$ as the event that A has a complex eigenvalue with imaginary part strictly in the range $(0, t)$. Then:*

$$\mathbf{P}(\mathcal{F}(t)) = O(tn^{2+o(1)}) + e^{-cn} + \mathbf{P}(\|A\|_{op} > Mn^{1/2}) \tag{3.4}$$

Assuming the entries of A have bounded fourth moment and taking $t = o(n^{-2+o(1)})$ in (3.4), we get that with high probability, A has only strictly real eigenvalues within $o(n^{-2+o(1)})$ of the real line.

The right side of Figure 3.1 demonstrates the effect of complex conjugate pairs of eigenvalues being repelled from the real line. In the same figure we see the contrast between the case where the atom random variables take on complex values versus purely real values.

When the atom random variables have independent real and imaginary part, we can use the argument in this chapter along with a least singular value result whose bound holds uniformly in the shift $z\sqrt{n}I_n$. In particular, the bound does not deteriorate when approaching the real line. Applying the preceding covering argument with Theorem 1.1.4 extended appropriately to the two smallest singular values, we get the following version of eigenvalue spacing for iid matrices where the atom variables have independent real and imaginary part:

Theorem 3.3.3. *Let A be an $n \times n$ iid matrix with atom random variables that satisfy the same assumptions required for Theorem 1.1.4. Let Δ denote the minimum eigenvalue spacing of A. For $s > 0$ we have:*

$$\mathbf{P}(\Delta < s) = O(s^2 n^{4+o(1)}) + e^{-cn}$$

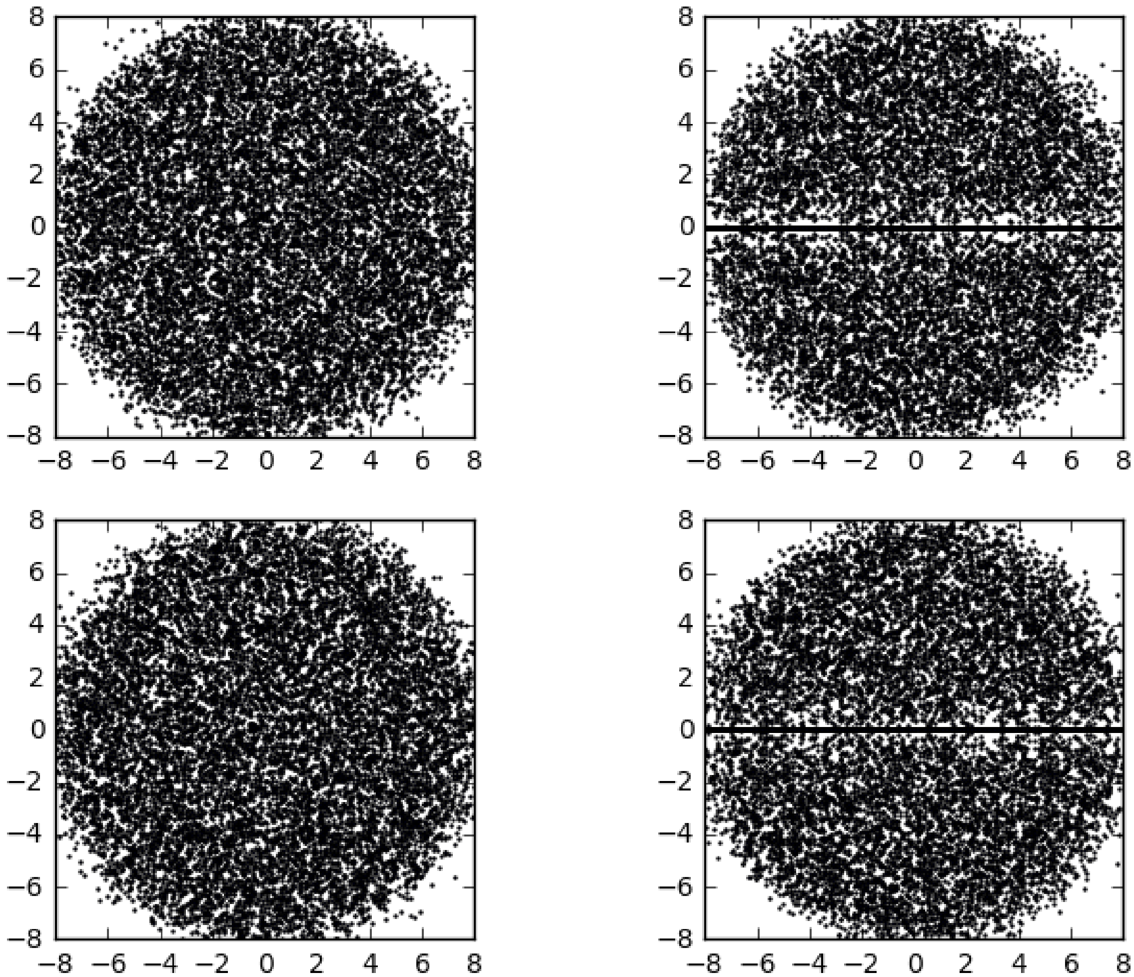

Figure 3.1: Eigenvalues of 200 independent 64×64 matrices generated using $\xi \sim g_{\mathbf{C}}$ (top-left), $\xi \sim g_{\mathbf{R}}$ (top-right), $\xi \sim \mathrm{Unif}(S^1)$ (bottom-left), and $\xi \sim \mathrm{Bernoulli}(1/2)$ (bottom-right)

APPENDIX A

Joint tail estimates for two smallest singular values

In this appendix we extend the least singular value theorem established in Chapter 2 to also include s_{n-1}. The approach will be similar to that of before: first using compressible and incompressible vectors to reduce to a distances problem and then applying a small ball probability result to the last two columns after conditioning on the rest of the matrix. The key new ingredient is a version of the invertibility via distances lemma accounting for two smallest singular values.

A.1 Reduction to distances problem

Let A be an $n \times n$ matrix with each entry an iid copy of ξ, where ξ satisfies the same assumptions as in Chapter 2. The main result in this appendix is the following:

Theorem A.1.1. *For any $t_2 \geq t_1 \geq 0$ and $\lambda = O(1)$, we have*

$$\mathbf{P}(s_n(A-\lambda\sqrt{n}) \leq t_1 n^{-1/2}, s_{n-1}(A-\lambda\sqrt{n}) \leq t_2 n^{-1/2} \text{ and } \|A\|_{op} \leq M\sqrt{n}) \leq O\left(\frac{t_1^2 t_2^2}{\delta^2}\right) + e^{-cn}$$

where $\delta > n^{-C}$ is the absolute value of the imaginary part of λ.

Recall that for fixed parameters $a, b \in (0,1)$, a unit vector z is compressible if there is a sparse vector z' with at most an non-zero coordinates such that $\|z - z'\|_2 \leq b$. A unit vector is incompressible if it is not compressible. Every incompressible vector z must have at least an indices k where $|z_k| \geq bn^{-1/2}$. Otherwise we have $\|z - z'\| \leq b$, where z' is the projection of z onto the indices where $|z_k| \geq bn^{-1/2}$.

Call a plane W incompressible if all of the unit vectors on W are incompressible. The event where $s_{n-1}(A - \lambda\sqrt{n}) \leq t_2 n^{-1/2}$ occurs implies that there is a plane W on which

$\|(A-\lambda\sqrt{n})|_W\|_{op} \leq t_2 n^{-1/2}$. If there is a compressible vector $z \in W$, then we would have $\|(A-\lambda\sqrt{n})z\| \leq t_2 n^{-1/2}$ for some compressible z. Moreover, since we are assuming that $t_1 \leq t_2$, we can now break the event in Theorem A.1.1 into two parts:

$$
\begin{aligned}
&\mathbf{P}(s_n(A-\lambda\sqrt{n}) \leq t_1 n^{-1/2}, s_{n-1}(A-\lambda\sqrt{n}) \leq t_2 n^{-1/2} \text{ and } \|A\|_{op} \leq M\sqrt{n}) \\
&\leq \mathbf{P}(\inf_{z\in \text{Incomp}} \left\|(A-\lambda\sqrt{n})z\right\| \leq \frac{t_1}{\sqrt{n}}, \inf_{\text{incomp } W} \left\|(A-\lambda\sqrt{n})|_W\right\|_{op} \leq \frac{t_2}{\sqrt{n}} \text{ and } \|A\|_{op} \leq M\sqrt{n}) \\
&+ \mathbf{P}(\inf_{z\in \text{Comp}} \left\|(A-\lambda\sqrt{n})z\right\| \leq t_2 n^{-1/2} \text{ and } \|A\|_{op} \leq M\sqrt{n})
\end{aligned}
$$

Proposition 2.2.4 controls the second term above to be exponentially small. The next steps are to go toward a new invertibility via distances lemma for the purpose of bounding the first summand. We will use the following form of Lemma 2.2.5:

Lemma A.1.2. *Let X be an $n\times n$ random matrix and $X_1,\dots,X_n$ denote the column vectors of X. Let H_k denote the span of the $n-1$ columns excluding X_k. Let $\mathcal{E}$ be an arbitrary event. Then for every $a, b \in (0,1)$ and $t_1 > 0$, we have*

$$\mathbf{P}(\inf_{z\in Incomp(a,b)} \|Xz\|_{op} < t_1 b n^{-1/2} \text{ and } \mathcal{E}) \leq \frac{1}{an}\sum_{k=1}^{n} \mathbf{P}(\text{dist}(X_j, H_k) < t_1 \text{ and } \mathcal{E})$$

Lemma A.1.2 is given in [RV08] and as remarked in [TV10c], the intersection with an arbitrary event E does not affect the proof. A straightforward modification of the argument gives:

Lemma A.1.3. *Let X be an $n\times n$ random matrix and $X_1,\dots,X_n$ denote the column vectors of X. Fix an index k. For $j\neq k$, let H_{jk} denote the span of the $n-2$ columns excluding X_j and X_k. Let $\mathcal{E}$ be an arbitrary event. Then for every $a, b \in (0,1)$ and $t_2 > 0$, we have*

$$\mathbf{P}(\inf_{incomp\ W} \|X|_W\|_{op} < t_2 b n^{-1/2} \text{ and } \mathcal{E}) \leq \frac{1}{an}\sum_{j\neq k} \mathbf{P}(\text{dist}(X_j, H_{jk}) < t_2 \text{ and } \mathcal{E})$$

Proof. We give the proof without intersecting with the arbitrary event $\mathcal{E}$ to ease notation. Suppose there is an incompressible plane W on which X's operator norm is at most $t_2 b n^{-1/2}$.

Consider $z \in W$ with $z_k = 0$. For any $j \neq k$, we have

$$\begin{aligned} \|Xz\| &\geq \operatorname{dist}(Xz, H_{jk}) \\ &= \operatorname{dist}(z_j X_j + z_k X_k, H_{jk}) \\ &= |z_j| \operatorname{dist}(X_j, H_{jk}) \end{aligned}$$

This implies that for any $j \neq k$, we have either

$$|z_j| < bn^{-1/2} \text{ or } \operatorname{dist}(X_j, H_{jk}) < t_2$$

Otherwise, $\|Xz\| \geq bn^{-1/2} t_2$, in contradiction with our assumption that X's operator norm on W is strictly less than $t_2 bn^{-1/2}$. The incompressibility of z implies that there are at least an indices j where $|z_j| \geq bn^{-1/2}$. On these indices, we must have $\operatorname{dist}(X_j, H_{jk}) < t_2$. In summary, we have deduced that the existence of an incompressible plane W with $\|X|_W\|_{op} < t_2 bn^{-1/2}$ implies that there are at least an indices j such that $\operatorname{dist}(X_j, H_{jk}) < t_2$. Markov's inequality now gives:

$$\begin{aligned} \mathbf{P}(|\{j : \operatorname{dist}(X_j, H_{jk}) < t_2\}| > an) &\leq \frac{\mathbf{E}\, |\{k : \operatorname{dist}(X_j, H_{jk}) < t_2\}|}{an} \\ &= \frac{1}{an} \sum_{j \neq k} \mathbf{P}(\operatorname{dist}(X_j, H_{jk}) < t_2) \end{aligned}$$

□

Combining the previous two lemmas gives:

Lemma A.1.4. *Let X be an $n \times n$ random matrix and $X_1, \ldots, X_n$ denote the column vectors of X. Let H_k denote the span of the $n-1$ columns excluding X_k and for $j \neq k$, let H_{jk} denote the span of the $n-2$ columns excluding X_j and X_k. Let $\mathcal{E}$ be an arbitrary event. Then for every $a, b \in (0,1)$ and $t_2 \geq t_1 \geq 0$, we have*

$$\begin{aligned} &\mathbf{P}\Big(\inf_{z \in Incomp(a,b)} \|Xz\| < t_1 bn^{-1/2}, \inf_{incomp\ W} \|X|_W\|_{op} < s_2 bn^{-1/2} \text{ and } \mathcal{E}\Big) \\ &\qquad \leq \frac{1}{a^2 n^2} \sum_{k=1}^{n} \sum_{j \neq k} \mathbf{P}(\operatorname{dist}(X_k, H_k) < t_1, \operatorname{dist}(X_j, H_{jk}) < t_2 \text{ and } \mathcal{E}) \end{aligned}$$

Proof. First take the arbitrary event in Lemma A.1.2 as

$$\mathcal{F} = \left\{ \inf_{\text{incomp } W} \|X|_W\|_{op} < t_2 b n^{-1/2} \text{ and } \mathcal{E} \right\}$$

to get

$$\mathbf{P}(\inf_{z \in \text{Incomp}(a,b)} \|Xz\| < t_1 b n^{-1/2} \text{ and } \mathcal{F}) \leq \frac{1}{an} \sum_{k=1}^{n} \mathbf{P}(\text{dist}(X_k, H_k) < t_1 \text{ and } \mathcal{F}) \tag{A.1}$$

Now for each k, take the arbitrary event as

$$\mathcal{F}_k = \{\text{dist}(X_k, H_k) < t_1 \text{ and } \mathcal{E}\}$$

in Lemma A.1.3 to get

$$\begin{aligned} \mathbf{P}(\text{dist}(X_k, H_k) < t_1 \text{ and } \mathcal{F}) &= \mathbf{P}(\inf_{\text{incomp } W} \|X|_W\|_{op} < t_2 b n^{-1/2} \text{ and } \mathcal{F}_k) \\ &\leq \frac{1}{an} \sum_{j \neq k} \mathbf{P}(\text{dist}(X_j, X_{jk}) < t_2 \text{ and } \mathcal{F}_k) \\ &= \frac{1}{an} \sum_{j \neq k} \mathbf{P}(\text{dist}(X_k, H_k) < t_1, \text{dist}(X_j, H_{jk}) < t_2 \text{ and } \mathcal{E}) \end{aligned}$$

Combining the above with A.1 gives the result. □

By the preceding Lemma and the compressible/incompressible decomposition, the proof of Theorem A.1.1 is reduced to establishing the following distances bound:

Proposition A.1.5. *Let A be an $n \times n$ matrix satisfying the same assumptions as in Theorem A.1.1. Let $X_1, \ldots, X_n$ be the columns of $A - \lambda\sqrt{n}$. Let H_k denote the span of the $n-1$ columns excluding X_k and for $j \neq k$, let H_{jk} denote the span of the $n-2$ columns excluding X_j and X_k. Then for every $\lambda = O(1)$ and $t_2 \geq t_1 \geq 0$, we have*

$$\mathbf{P}(\text{dist}(X_n, H_n) < t_1, \text{dist}(X_{n-1}, H_{n-1n}) < t_2 \textit{ and } \|A\|_{op} \leq M\sqrt{n}) \leq O\left(\frac{t_1^2 t_2^2}{\delta^2}\right) + e^{-cn}$$

where δ is the absolute value of the imaginary part of λ.

A.2 Distances problem via small ball probability and structure theorem

Proof. We condition on a realization of the first $n-2$ columns so that H_{n-1n}, H_n, and normal vectors are fixed. Denote the expectation with respect to the first $n-2$ columns $\mathbf{E}_{1,\dots,n-2}$ and the probability with respect to the last two columns $\mathbf{P}_{n-1n}$. $\operatorname{dist}(X_n, H_n) < t_1$ implies that that there is some vector X normal to H_n such that

$$|\langle X, X_n\rangle| < t_1$$

and similarly we have some vector X' normal to H_{n-1n} such that

$$|\langle X', X_{n-1}\rangle| < t_2$$

Let B denote the $n-2\times n$ matrix with rows equal to the first $n-2$ columns of $A-\lambda\sqrt{n}$. Let $\mathcal{E}$ denote the event that every unit vector X^* in the kernel of B has $d(X^*)\geq c\delta$ and $D_2(X^*)\geq D_0$.

$$\begin{aligned}
&\mathbf{P}(\operatorname{dist}(X_n, H_n) < t_1, \operatorname{dist}(X_{n-1}, H_{n-1n}) < t_2 \text{ and } \|A\|_{op}\leq M\sqrt{n})\\
&\leq \mathbf{P}(|\langle X, X_n\rangle|\leq t_1 \text{ and } |\langle X', X_{n-1}\rangle|\leq t_2 \text{ for some } X, X'\in\ker B \text{ and } \|B\|_{op}\leq M\sqrt{n})\\
&\leq \mathbf{E}_{1,\dots,n-2}\mathbf{P}_{n-1n}(|\langle X, X_n\rangle|\leq t_1, |\langle X', X_{n-1}\rangle|\leq t_2 \text{ for some } X, X'\in\ker B \text{ and } \mathcal{E}) + \mathbf{P}(\mathcal{E}^c)\\
&\leq \mathbf{E}_{1,\dots,n-2}\mathbf{P}_{n-1n}(|\langle X, X_n\rangle|\leq t_1, |\langle X', X_{n-1}\rangle|\leq t_2 \text{ for some } X, X'\in\ker B \text{ and } \mathcal{E}) + e^{-cn}
\end{aligned}$$

where we have used Proposition 2.6.1 to control the probability of $\mathcal{E}^c$. The event $\mathcal{E}$ does not depend on any of the last two columns. Therefore, by independence of X_{n-1}, X_n and Proposition 2.6.2 we have

$$\begin{aligned}
&\mathbf{P}_{n-1n}(|\langle X, X_n\rangle|\leq t_1 \text{ and } |\langle X', X_{n-1}\rangle|\leq t_2 \text{ for some } X, X'\in\ker B \text{ and } \mathcal{E})\\
=&\mathbf{P}_{n-1}(|\langle X', X_{n-1}\rangle|\leq t_2 \text{ for } X'\in\ker B \text{ and } \mathcal{E})\cdot\mathbf{P}_n(|\langle X, X_n\rangle|\leq t_1 \text{ for } X'\in\ker B \text{ and } \mathcal{E})\\
\leq&O(\frac{t_2^2}{\delta}+e^{-cn})\cdot O(\frac{t_1^2}{\delta}+e^{-cn})
\end{aligned}$$

The result now follows from taking the expectation over the first $n-2$ columns. □

APPENDIX B

An elementary computation for real-imaginary correlation

Recall that for a complex vector $z = x + iy \in S_{\mathbf{C}}^{n-1}$, the real-imaginary correlation is defined to be

$$d(z) := (\|x\|_2^2 \|y\|_2^2 - (x \cdot y)^2)^{1/2} = \sqrt{\det\left(\begin{pmatrix} x^T \\ y^T \end{pmatrix} \begin{pmatrix} x & y \end{pmatrix}\right)}$$

$d(z)$ is invariant under rotations, i.e. $d(e^{i\theta} z) = d(z)$ for any θ. The following proposition relates $d(z)$ to the extremal lengths of the real and imaginary parts of rotations $e^{i\theta} z$.

Proposition B.0.1. *Let $z = x + iy \in S_{\mathbf{C}}^{n-1}$ have real-imaginary correlation $d(z)$. As θ varies, the minimal value of $\left\|Re(e^{i\theta} z)\right\|_2^2$ is*

$$\frac{1}{2} - \frac{1}{2}\sqrt{1 - 4d(z)^2}$$

and the corresponding maximal value is

$$\frac{1}{2} + \frac{1}{2}\sqrt{1 - 4d(z)^2}$$

Proof. We have

$$\mathrm{Re}(e^{i\theta} z) = \cos(\theta)x - \sin(\theta)y$$

and hence the extremal lengths of $\left\|\mathrm{Re}(e^{i\theta} z)\right\|_2$ correspond to the singular values of the matrix that has x and y as columns:

$$\begin{pmatrix} x & y \end{pmatrix}$$

The eigenvalues of

$$\begin{pmatrix} x^T \\ y^T \end{pmatrix} \begin{pmatrix} x & y \end{pmatrix} = \begin{pmatrix} \|x\|_2^2 & x \cdot y \\ x \cdot y & \|y\|_2^2 \end{pmatrix}$$

satisfy

$$\lambda^2 - (\|x\|_2^2 + \|y\|_2^2)\lambda + (\|x\|_2^2 \, \|y\|_2^2 - (x \cdot y)^2) = 0$$

or equivalently

$$\lambda^2 - \lambda + d(z)^2 = 0$$

The claim follows from solving the above quadratic. $\square$

REFERENCES

[AGZ10] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.

[Bai97] Z. D. Bai. "Circular law." *Ann. Probab.*, **25**(1):494–529, 1997.

[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.

[EKS94] Alan Edelman, Eric Kostlan, and Michael Shub. "How many eigenvalues of a random matrix are real?" *J. Amer. Math. Soc.*, **7**(1):247–267, 1994.

[Gir84] V. L. Girko. "The circular law." *Teor. Veroyatnost. i Primenen.*, **29**(4):669–679, 1984.

[GT10] Friedrich Götze and Alexander Tikhomirov. "The circular law for random matrices." *Ann. Probab.*, **38**(4):1444–1491, 2010.

[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.

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

[Lat05] RafałLatał a. "Some estimates of norms of random matrices." *Proc. Amer. Math. Soc.*, **133**(5):1273–1282, 2005.

[Luh16] Kyle Luh. "Complex Random Matrices have no Real Eigenvalues." 09 2016.

[MC98] B. Mehlig and J. T. Chalker. "Eigenvector correlations in non-Hermitian random matrix ensembles." *Ann. Phys. (8)*, **7**(5-6):427–436, 1998.

[MC00] B. Mehlig and J. T. Chalker. "Statistical properties of eigenvectors in non-Hermitian Gaussian random matrix ensembles." *J. Math. Phys.*, **41**(5):3233–3256, 2000.

[Meh04] Madan Lal Mehta. *Random matrices*, volume 142 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, third edition, 2004.

[NTV17] Hoi Nguyen, Terence Tao, and Van Vu. "Random matrices: tail bounds for gaps between eigenvalues." *Probab. Theory Related Fields*, **167**(3-4):777–816, 2017.

[NV11] Hoi Nguyen and Van Vu. "Optimal inverse Littlewood-Offord theorems." *Adv. Math.*, **226**(6):5298–5319, 2011.

[PZ10] Guangming Pan and Wang Zhou. "Circular law, extreme singular values and potential theory." *J. Multivariate Anal.*, **101**(3):645–656, 2010.

[Rud06] Mark Rudelson. "Lower estimates for the singular values of random matrices." *C. R. Math. Acad. Sci. Paris*, **342**(4):247–252, 2006.

[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.

[RV09] Mark Rudelson and Roman Vershynin. "Smallest singular value of a random rectangular matrix." *Comm. Pure Appl. Math.*, **62**(12):1707–1739, 2009.

[RV15] Mark Rudelson and Roman Vershynin. "Delocalization of eigenvectors of random matrices with independent entries." *Duke Math. J.*, **164**(13):2507–2538, 2015.

[RV16] Mark Rudelson and Roman Vershynin. "No-gaps delocalization for general random matrices." *Geom. Funct. Anal.*, **26**(6):1716–1776, 2016.

[SST06] Arvind Sankar, Daniel A. Spielman, and Shang-Hua Teng. "Smoothed analysis of the condition numbers and growth factors of matrices." *SIAM J. Matrix Anal. Appl.*, **28**(2):446–476 (electronic), 2006.

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

[TV06] Terence Tao and Van Vu. *Additive combinatorics*, volume 105 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2006.

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

[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 Vu. "Random matrices: universality of ESDs and the circular law." *Ann. Probab.*, **38**(5):2023–2065, 2010. With an appendix by Manjunath Krishnapur.

[TV10c] Terence Tao and Van Vu. "Smooth analysis of the condition number and the least singular value." *Math. Comp.*, **79**(272):2333–2352, 2010.

[TV14] Terence Tao and Van Vu. "Random matrices have simple spectrum." 12 2014.

[TV15] Terence Tao and Van Vu. "Random matrices: universality of local spectral statistics of non-Hermitian matrices." *Ann. Probab.*, **43**(2):782–874, 2015.

[Ver14] Roman Vershynin. "Invertibility of symmetric random matrices." *Random Structures Algorithms*, **44**(2):135–182, 2014.

[YBK88] Y. Q. Yin, Z. D. Bai, and P. R. Krishnaiah. "On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix." *Probab. Theory Related Fields*, **78**(4):509–521, 1988.