Spectrum of heavy-tailed elliptic random matrices[*]

Andrew Campbell[†]   Sean O'Rourke[‡]

Abstract

An elliptic random matrix X is a square matrix whose (i,j)-entry X_{ij} is a random variable independent of every other entry except possibly X_{ji}. Elliptic random matrices generalize Wigner matrices and non-Hermitian random matrices with independent entries. When the entries of an elliptic random matrix have mean zero and unit variance, the empirical spectral distribution is known to converge to the uniform distribution on the interior of an ellipse determined by the covariance of the mirrored entries.

We consider elliptic random matrices whose entries fail to have two finite moments. Our main result shows that when the entries of an elliptic random matrix are in the domain of attraction of an α-stable random variable, for $0 < \alpha < 2$, the empirical spectral measure converges, in probability, to a deterministic limit. This generalizes a result of Bordenave, Caputo, and Chafaï for heavy-tailed matrices with independent and identically distributed entries. The key elements of the proof are (i) a general bound on the least singular value of elliptic random matrices under no moment assumptions; and (ii) the convergence, in an appropriate sense, of the matrices to a random operator on the Poisson Weighted Infinite Tree.

Keywords: ellitpic random matrices; Poisson point process; α-stable laws; elliptic law; heavy-tailed entries; singular values: least singular value; empirical spectral measure.
MSC2020 subject classifications: 60B20.
Submitted to EJP on February 8, 2022, final version accepted on September 5, 2022.

1 Introduction

Let $\mathrm{Mat}_n(\mathbb{F})$ be the set of $n \times n$ matrices over the field $\mathbb{F}$. For a matrix $A \in \mathrm{Mat}_n(\mathbb{C})$, the singular values of A are the square roots of the eigenvalues of AA^*, where A^* is the conjugate transpose of A. We let $s_n(A) \leq \cdots \leq s_1(A)$ denote the ordered singular

values of A, and $\lambda_1(A), \ldots, \lambda_n(A) \in \mathbb{C}$ be the eigenvalues of A in no particular order. For a matrix $A \in \mathrm{Mat}_n(\mathbb{C})$ we define the empirical spectral measure

$$\mu_A := \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(A)}$$

and the empirical singular value measure

$$\nu_A := \frac{1}{n} \sum_{i=1}^{n} \delta_{s_i(A)}.$$

These measures are central objects in random matrix theory, and the goal of this paper is to study the asymptotic behavior of the empirical spectral measure for a class of heavy-tailed elliptic random matrices.

Elliptic random matrices can be thought of as interpolating between random matrices whose entries are independent and identically distributed (i.i.d.) and Wigner matrices. We now give a precise definition.

Definition 1.1 (Elliptic Random Matrix). *Let (ξ_1, ξ_2) be a random vector with complex-valued random variable entries, ζ a complex random variable, and $X_n = (X_{ij})_{i,j=1}^n$ be an $n \times n$ random matrix. X_n is an elliptic random matrix if*

  (i) *$\{X_{ii} : 1 \leq i \leq n\} \cup \{(X_{ij}, X_{ji}) : 1 \leq i < j \leq n\}$ is a collection of independent random elements.*
  (ii) *the pairs $\{(X_{ij}, X_{ji})\}_{1 \leq i < j \leq n}$ are independent copies of (ξ_1, ξ_2).*
  (iii) *the diagonal elements $\{X_{ii} : 1 \leq i \leq n\}$ are independent copies of ζ.*

We refer to $(\xi_1, \xi_2), \zeta$ as the atom variables of the matrix X_n.

Elliptic random matrices were originally introduced by Girko [26, 27] in the 1980s, with the name coming from the limit of the empirical spectral measure. When the entries of the matrix have four finite moments, the limiting empirical spectral measures was investigated by Naumov [40]. The general case, when the entries are only assumed to have finite variance, was studied in [42].

Theorem 1.2 (Elliptic law for real random matrices, Theorem 1.5 in [42]). *Let X_n be an $n \times n$ elliptic random matrix with real atom variables $(\xi_1, \xi_2), \zeta$. Assume ξ_1, ξ_2 have mean zero and unit variance, and $\mathbb{E}[\xi_1 \xi_2] =: \rho$ for $-1 < \rho < 1$. Additionally assume ζ has mean zero and finite variance. Then almost surely the empirical spectral measure $\mu_{\frac{1}{\sqrt{n}} X_n}$ of $\frac{1}{\sqrt{n}} X_n$ converges weakly to the uniform probability measure on*

$$\mathcal{E}_\rho := \left\{ z \in \mathbb{C} : \frac{\mathrm{Re}(z)^2}{(1+\rho)^2} + \frac{\mathrm{Im}(z)^2}{(1-\rho)^2} \leq 1 \right\}$$

as $n \to \infty$.

Elliptic random matrices have also been studied in [29, 30, 43, 44].

Our main result, Theorem 1.6, gives an analogous result for heavy-tailed elliptic random matrices, i.e. when $\mathbb{E}|\xi_1|^2$ and $\mathbb{E}|\xi_2|^2$ are both infinite. In 1994, Cizeau and Bouchaud [19] introduced Lévy matrices as a heavy-tailed analogue of the Gaussian Orthogonal Ensemble (GOE). Instead of Gaussian entries, these matrices have entries in the domain of attraction of an α-stable random variable, for $0 < \alpha < 2$. They predicted a deterministic limit μ_α, which depends only on α, for the empirical spectral measures of these matrices when properly scaled. Convergence to a deterministic limit was first proved by Ben Arous and Guionnet [8] and later by Bordenave, Caputo, and Chafaï [12] with an alternative characterization in their study of random Markov matrices.

In [13] Bordenave, Caputo, and Chafaï proved the empirical spectral measure of random matrices with i.i.d. entries in the domain of attraction of a complex α-stable random variable converges almost surely to an isotropic measure on $\mathbb{C}$ with unbounded support. Notably for Lévy matrices the limiting spectral measure inherits the tail behavior of the entries, while the limiting spectral measure of heavy-tailed random matrices with i.i.d. entries has a lighter tail and finite moments of every order.

Much of the work on heavy-tailed random matrices has been on the spectrum of symmetric matrices, either Lévy or sample covariance matrices $[2, 4, 5, 7, 8, 11, 12, 32, 52]$. Motivated by questions of delocalization from physics there has also been considerable work done in studying the eigenvectors of symmetric heavy-tailed matrices $[1, 9, 17–19, 39, 55]$.

As is often the case in random matrix theory most of the work on heavy-tailed random matrices has focused on ensembles where the entries are independent up to symmetry conditions on the matrix. Work on matrices with dependent entries is still limited. Heavy-tailed matrices with normalized rows have been considered for random Markov chains in $[12, 15]$ and sample correlation matrices in $[32]$. In $[6]$ extreme eigenvalue statistics of symmetric heavy-tailed random matrices with m-dependent entries were studied and shown to converge to a Poisson process. This m-dependence can be thought of as a short range dependence meant to model stock returns that depend on stocks from the same sector of size determined by m. To the best of our knowledge there are not any results on non-Hermitian heavy-tailed matrices with long range dependence between entries from different rows outside the random reversible Markov chains studied in $[12]$.

The key question when approaching heavy-tailed elliptic random matrices is how to measure the dependence between ξ_1 and ξ_2. Without two finite moments the covariance between ξ_1 and ξ_2, which was the key parameter in Theorem 1.2, cannot be defined. Similar notions, such as covariation or codifference, exist for α-stable random variables but they do not seem sufficient for our purposes. The difference is that the covariation does not provide as much information for α-stable random vectors as the covariance does for Gaussian random vectors. If $X = (X_1, X_2)$ is a bivariate Gaussian random vector where X_1 and X_2 are standard real Gaussian random variables, then the correlation $\rho = \mathbb{E}[X_1 X_2]$ uniquely determines the distribution of X. Thus one approach to measuring dependence is to find a parameter which uniquely determines the distribution of a properly normalized α-stable random vector. The distribution of an α-stable random vector Y in $\mathbb{R}^n$ is determined uniquely through its characteristic function of the form

$$\mathbb{E}\exp\left(iu^T Y\right) = \begin{cases} \exp\left(-\int_{\mathbb{S}^{n-1}} |u^T s|^\alpha (1 - i\,\mathrm{sign}(u^T s)\tan(\frac{\pi\alpha}{2}))d\theta(s) + iu^T y\right), & \alpha \neq 1 \\ \exp\left(-\int_{\mathbb{S}^{n-1}} |u^T s|(1 + i\,\mathrm{sign}(u^T s)\log|u^T s|)d\theta(s) + iu^T y\right), & \alpha = 1 \end{cases}$$

for a finite measure θ on the unit sphere $\mathbb{S}^{n-1}$ and a deterministic vector y. Y can be translated and scaled so that $y = 0$ and θ is a probability measure uniquely determining the distribution of Y. θ is called the spectral measure of Y, and it turns out to be the appropriate explicit description of the dependence between the entries of Y. The definition of θ can be extended to random variables which are not stable but rather in the domain of attraction of an α-stable random variable, see Definition 1.3.

If the components of Y are independent, then θ is supported entirely on the intersection of the axes and the unit sphere. Intuitively, when considering the mirrored entries of a random matrix, if θ is close to a measure supported on the intersection of the axes and the unit sphere, the entries are close to independent, after scaling. If θ is close to a measure supported on the set $\{(z_1, z_2) : |z_1|^2 + |z_2|^2 = 1, z_1 = \bar{z}_2\}$ then the matrix is close to Hermitian. Numerical simulations seem to reflect this intuition in the spectrum of elliptic random matrices, see Figures 1, 2, and 3.

1.1 Matrix distribution

We will consider elliptic random matrices whose atom variables satisfy the following conditions.

Definition 1.3 (Condition **C1**). *We say the atom variables $(\xi_1, \xi_2), \zeta$ satisfy Condition **C1** if*

(i) *there exists a positive number $0 < \alpha < 2$, a sequence $a_n = \ell(n)n^{1/\alpha}$ for a slowly varying function ℓ (i.e. $\lim_{t \to \infty} \ell(tx)/t = 1$ for all $x > 0$), and a finite measure θ_d on the unit sphere in $\mathbb{C}^2$ such that for all Borel subsets D of the unit sphere with $\theta_d(\partial D) = 0$ and all $r > 0$,*

$$\lim_{n \to \infty} n \mathbb{P}\left(\frac{(\xi_1, \xi_2)}{\|(\xi_1, \xi_2)\|} \in D, \|(\xi_1, \xi_2)\| \geq r a_n \right) = \theta_d(D) m_\alpha([r, \infty)),$$

where m_α is a measure on $(0, \infty)$ with density $f(r) = \alpha r^{-(1+\alpha)}$.

(ii) *there exists a constant $C > 0$ such that $\mathbb{P}(|\zeta| \geq t) \leq Ct^{-\alpha}$ for all $t > 0$.*

We will reserve θ_d to denote the measure on a sphere associated to the atom variables of a heavy-tailed elliptic random matrix; it may be worth noting that d stands for "dependence" and is not a parameter. A key step in the proofs of the main results below relies on the connections between heavy-tailed random variables and Poisson point processes, and Definition 1.3(i) determines the intensity measure of the point process. Heuristically the parameter α determines how heavy-tailed the atom variables are, while the measure θ_d gives the direction (ξ_1, ξ_2) points in when conditioned to have large norm. Definition 1.3(ii) simply rules out the possibility that the diagonal entries are more heavy-tailed than the off-diagonal entries. As it turns out, Condition **C1** is enough to prove convergence of the empirical singular value distribution, see Theorem 1.5. We will need more assumptions to prove convergence of the empirical spectral measure.

Definition 1.4 (Condition **C2**). *We say $(\xi_1, \xi_2), \zeta$ satisfy Condition **C2** if the atom variables satisfy Condition **C1** and if*

(i) *there does not exist $(a, b) \in \mathbb{C}^2 \setminus \{(0,0)\}$ such that*

$$\text{supp}(\theta_d) \subseteq \{(z, w) \in \mathbb{C}^2 : az + bw = 0, |z|^2 + |w|^2 = 1\}.$$

(ii) *$a_n/n^{1/\alpha} \to c$, for some constant $c > 0$ as $n \to \infty$.*

Definition 1.4(i) states that θ_d is not supported on a hyperplane through the origin. This rules out random matrices which should be avoided, such as random matrices that are essentially upper or lower triangular. However, it also rules out real symmetric matrices, suggesting there may be some room for refinement.

1.2 Main results

For simplicity, let $A_n := \frac{1}{a_n} X_n$. Our first result gives the convergence of the singular values of $A_n - zI_n$, which we will denote as $A_n - z$, for $z \in \mathbb{C}$. Here and throughout, I_n denotes the $n \times n$ identity matrix. While interesting on its own, this is the first step in the method of Hermitization to establish convergence of the empirical spectral measure. Throughout we will use $\Rightarrow$ to denote weak convergence of probability measures and convergence in distribution of random variables.

Theorem 1.5 (Singular values of heavy-tailed elliptic random matrices). *Let X_n be an $n \times n$ elliptic random matrix with atom variables $(\xi_1, \xi_2), \zeta$ which satisfy Condition **C1**.*

Then for each $z \in \mathbb{C}$ there exists a deterministic probability measure ν_{z,α,θ_d}, depending only on z, α and θ_d, such that almost surely

$$\nu_{A_n - zI_n} \Rightarrow \nu_{z,\alpha,\theta_d}$$

as $n \to \infty$.

The proof of Theorem 1.5 can be found following Theorem 4.11 in Section 4. Under Condition **C2** we prove the convergence of the empirical spectral measure of A_n.

Theorem 1.6 (Eigenvalues of heavy-tailed elliptic random matrices). *Let X_n be an $n \times n$ elliptic random matrix with atom variables $(\xi_1, \xi_2), \zeta$ which satisfy Condition **C2**. Then there exists a deterministic probability measure μ_{α,θ_d}, depending only on α and θ_d, such that*

$$\mu_{A_n} \Rightarrow \mu_{\alpha,\theta_d}$$

in probability as $n \to \infty$. Moreover for any smooth $\varphi : \mathbb{C} \to \mathbb{C}$ with compact support

$$\int_{\mathbb{C}} \varphi \, d\mu_{\alpha,\theta_d} = \frac{1}{2\pi} \int_{\mathbb{C}} \Delta\varphi(z) \left[\int_0^\infty \log(t) \, d\nu_{\alpha,z,\theta_d} \right] dz, \tag{1.1}$$

where ν_{α,z,θ_d} is as in Theorem 1.5 and dz is the Lebesgue measure on $\mathbb{C}$.

The proof of Theorem 1.6 is given in Subsection 6.5. A distributional equation describing the Stieltjes transform of ν_{α,z,θ_d} is given in Proposition 4.12, which, when combined with (1.1), gives a description of μ_{α,θ_d}.

Remark 1.7. If $\theta_d = 1/2(\theta_1 + \theta_2)$ where θ_1 and θ_2 are probability measures with $\text{supp}(\theta_i) \subseteq \{(z_1, z_2) \in \mathbb{S} : z_i = 0\}$ and $\theta_1(A) = \theta_2(\{(z_1, z_2) : (z_2, z_1) \in A\})$, then ν_{α,z,θ_d} in Theorem 1.5 and μ_{α,θ_d} in Theorem 1.6 are the same measures as in the main results of Bordenave, Caputo, and Chafaï [13]. This can be seen by computing θ_d when ξ_1 and ξ_2 are independent and identically distributed. It is also worth noting that if the matrix X_n is complex Hermitian but not real symmetric, then (ξ_1, ξ_2) will satisfy Condition **C2** (i), and thus Theorem 1.6 holds.

Numerical simulations seem to give weight to the reasoning that the support of θ_d determines how close μ_{α,θ_d} is to being isotropic or supported on the real line. In Figure 3 we see as $\text{supp}(\theta_d)$ moves further from $\{z_1 = z_2\}$ the mass of the spectrum moves further from the real line. A similar phenomenon appears in Figure 2: as $\text{supp}(\theta_d)$ moves further from the intersection of the axes and the sphere the spectrum becomes further from isotropic. We also see in Figure 1 that the tail behavior of the spectrum appears to depend on θ_d and may vary in different directions.

Spectra similar to Figure 1 can be found in [31] where the authors study the spectral measure of $C_1 + iC_2$ where C_1 and C_2 are free random Lévy elements. The authors use tools in free probability to show the spectrum is supported inside the "hyperbolic cross", which appears to be the case for heavy-tailed elliptic random matrices for certain θ_d. The limiting spectral measure in [13] is not contained in a "hyperbolic cross", which suggests $C_1 + iC_2$ is not the heavy-tailed analogue of a circular element, in contrast with the case when C_1 and C_2 are free semicircular elements and $C_1 + iC_2$ is a circular element.

1.3 Further questions

Since our results capture both heavy-tailed Hermitian matrices and heavy-tailed matrices with i.i.d. entries, μ_{α,θ_d} does actually depend on θ_d. However, as can be seen from [13], μ_{α,θ_d} does not depend on every aspect of θ_d.

Question 1.8. What properties of θ_d determine μ_{α,θ_d}?
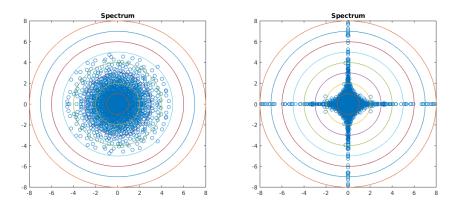
Figure 1: The plot on the left is the spectrum of an $n \times n$ matrix $n^{-1/\alpha}X$ where $\alpha = 1.25$, $n = 2000$, and the entries of X are i.i.d. random variables distributed as $\varepsilon U^{-1/\alpha}$, where ε is uniformly distributed on $\{-1, 1\}$ and U is uniformly distributed on $[0, 1]$. The plot on the right is the spectrum of an $n \times n$ elliptic random matrix $n^{-1/\alpha}X$ where X has atom variables $(U^{-1/\alpha}\cos(w), U^{-1/\alpha}\sin(w)), 1$ with $\alpha = 1.25$, $n = 2000$, U uniformly distributed on $[0, 1]$, and w uniformly distributed on $[0, 2\pi]$. The plot window on the right is cropped to avoid extreme values.
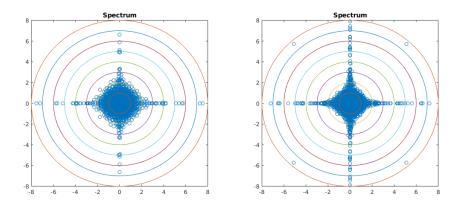

Figure 2: Both plots show the spectrum of an $n \times n$ elliptic random matrix $n^{-1/\alpha}X$ where X has atom variables $(U^{-1/\alpha}\cos(w), U^{-1/\alpha}\sin(w)), 1$ with $\alpha = 1.25$, $n = 2000$, U uniformly distributed on $[0, 1]$, and w uniformly distributed on $\{0, \pi/2, \pi, 3\pi/2\} + [-b\pi/4, b\pi/4]$. For the plot on the left $b = 0.1$, while for the plot on the right $b = 0.5$. Both plot windows are trimmed to avoid extreme values.

In the case when X_n has i.i.d. entries the limiting empirical spectral measure has an exponential tail [13] while in the case when X_n is Hermitian it has the same tail behavior as the entries [8, 12]. This leads us to ask how the tail of μ_{α,θ_d} depends on θ_d.

Question 1.9. How does the tail behavior of μ_{α,θ_d} vary with respect to θ_d?

1.4 Outline

As expected with the empirical spectral measures of non-Hermitian random matrices, we make use of Girko's Hermitization method. However, instead of considering the logarithmic potential directly, we follow the approach of Bordenave, Caputo, and Chafaï
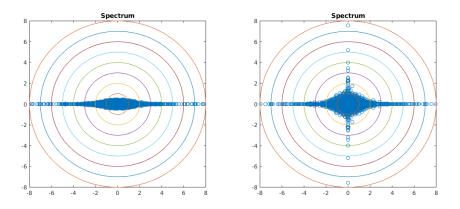
Figure 3: Both plots show the spectrum of an $n \times n$ elliptic random matrix $n^{-1/\alpha}X$ where X has atom variables $(U^{-1/\alpha}\cos(w), U^{-1/\alpha}\sin(w)), 1$ with $\alpha = 1.25$, $n = 2000$, U uniformly distributed on $[0,1]$, and w uniformly distributed on $\{\pi/4, 5\pi/4\} + [-b\pi/4, b\pi/4]$. For the plot on the left $b = 1$, while for the plot on the right $b = 4/3$. Both plot windows are trimmed to avoid extreme values.

[12, 13] by using the objective method of Aldous and Steele [3] to get convergence of the matrices to an operator on Aldous' Poisson Weighted Infinite Tree (PWIT). This objective method approach was expanded by Jung [35] to symmetric light-tailed random matrices, adjacency matrices of sparse random graphs, and symmetric random matrices whose column entries are some combination of heavy-tailed and light-tailed random variables.

In Sections 2 and 3 we give a collection of results and background for approaching the proof of Theorem 1.5. In Section 4 we define the PWIT, establish the local weak convergence of the matrices A_n to an operator associated with the PWIT, and give a proof of Theorem 1.5. In Section 5 we give a very general bound on the least singular value of elliptic random matrices. In Section 6 we establish the uniform integrability of $\log(\cdot)$ against $\nu_{A_n - zI_n}$ and complete the proof of Theorem 1.6. The appendix contains some auxiliary results.

We conclude this section by establishing notation, giving a brief description of Hermitization, and stating some properties of ξ_1 and ξ_2 implied by Condition **C2**.

1.5 Notation

We now establish notation we will use throughout.

Let $[n] := \{1, \ldots, n\}$ denote the discrete interval. For a vector $v = (v_i)_{i=1}^n \in \mathbb{C}^n$ and a subset $I \subset [n]$, we let $v_I := (v_i)_{i \in I} \in \mathbb{C}^I$. Similarly, for an $m \times n$ matrix $A = (A_{ij})_{i \in [m], j \in [n]}$ and $I \subset [m], J \subset [n]$, we define $A_{I \times J} := (A_{ij})_{i \in I, j \in J}$. For a countable set S we will let I_S denote the identity on $\ell^2(S)$. In the case when $S = [n], \ell^2(S) = \mathbb{C}^n$ we will simply write I_n or I if the dimension is clear.

For a vector $x \in \mathbb{C}^n$, $\|x\|$ is the Euclidean norm of x. For a matrix A, A^T is the transpose of A and A^* is the conjugate transpose of A. In addition, $\|A\| = s_1(A)$ is the spectral norm of A and $\|A\|_2$ is the Hilbert–Schmidt norm of A defined by the formula

$$\|A\|_2 = \sqrt{\operatorname{tr}(AA^*)}.$$

We will use the convention that the Kronecker product, $A \otimes B$, of a $k \times k$ matrix A and

an $n \times n$ matrix $B = (b_{ij})_{i,j=1}^n$ is an $nk \times nk$ block matrix

$$A \otimes B = \begin{pmatrix} b_{11}A & \cdots & b_{1n}A \\ \vdots & \ddots & \vdots \\ b_{n1}A & \cdots & b_{nn}A \end{pmatrix}.$$

For a linear, but not necessarily bounded, operator A on a Hilbert space we let $\mathcal{D}(A)$ denote the domain of A. We will often use the shorthand $A + z$ to denote $A + zI$ where I is the identity operator.

For two complex-valued square integrable random variables ξ and ψ, we define the correlation between ξ and ψ as

$$\mathrm{Corr}(\xi, \psi) := \frac{\mathrm{Cov}(\xi, \psi)}{\sqrt{\mathrm{Var}(\xi)\,\mathrm{Var}(\psi)}},$$

where $\mathrm{Cov}(\xi, \psi) := \mathbb{E}[(\xi - \mathbb{E}\xi)\overline{(\psi - \mathbb{E}\psi)}]$ is the covariance between ξ and ψ, and $\mathrm{Var}(\xi) = \mathbb{E}|\xi - \mathbb{E}\xi|^2$ is the variance of ξ. For two random elements X and Y we say $X \stackrel{d}{=} Y$ if X and Y have the same distribution. We will also say a positive random variable Y stochastically dominates a positive random variable Z if for all $x > 0$

$$\mathbb{P}(Y \geq x) \geq \mathbb{P}(Z \geq x).$$

For a topological space E, $\mathcal{B}(E)$ will always denote the Borel σ-algebra of E. $\mathbb{R}_+$ will denote the positive real numbers.

Throughout this paper we will use asymptotic notation (O, o, Θ, etc.) under the assumption that $n \to \infty$. $X = O(Y)$ if $X \leq CY$ for an absolute constant $C > 0$ and all $n \geq C$, $X = o(Y)$ if $X \leq C_n Y$ for $C_n \to 0$, $X = \Theta(Y)$ if $cY \leq X \leq CY$ for absolute constants $C, c > 0$ and all $n \geq C$, and $X \sim Y$ if $X/Y \to 1$.

1.6 Hermitization

Let $\mathcal{P}(\mathbb{C})$ be the set of probability measures on $\mathbb{C}$ which integrate $\log|\cdot|$ in a neighborhood of infinity. For every $\mu \in \mathcal{P}(\mathbb{C})$ the logarithmic potential U_μ of μ on $\mathbb{C}$ is a function $U_\mu : \mathbb{C} \to [-\infty, \infty)$ defined for every $z \in \mathbb{C}$ by

$$U_\mu(z) = \int_{\mathbb{C}} \log|z - w| d\mu(w).$$

In $\mathcal{D}'(\mathbb{C})$ one has $\Delta U_\mu = 2\pi\mu$, where $\mathcal{D}'(\mathbb{C})$ is the set of Schwartz-Sobolev distributions on $\mathbb{C}$ endowed with its usual convergence with respect to all infinitely differentiable functions with compact support.

Lemma 1.10 (Lemma A.1 in [13])**.** *For every $\mu, \nu \in \mathcal{P}(\mathbb{C})$, if $U_\mu = U_\nu$ a.e. then $\mu = \nu$.*

To see the connection between logarithmic potentials and random matrices consider an $n \times n$ random matrix A. If $P(z) = \det(A - zI_n)$ is the characteristic polynomial of A, then

$$U_{\mu_A}(z) = \int_{\mathbb{C}} \log|z - w| d\mu_A(w) = \frac{1}{n}\log|P_A(z)| = \int_0^\infty \log(t) d\nu_{A-z}(t). \tag{1.2}$$

Thus, through the logarithmic potential we can move from a question about eigenvalues of A to singular values of $A - z$. We refer the reader to [16] for more on the logarithmic potential in random matrix theory. One immediate issue is that $\log(\cdot)$ is not a bounded function on $\mathbb{R}_+$, and thus we need more control on the integral of $\log(\cdot)$ with respect to $\{\nu_{A_n - z}\}_{n \geq 1}$.

Definition 1.11 (Uniform integrability almost surely and in probability)**.** *Let $\{\mu_n\}_{n=1}^\infty$ be a sequence of random probability measures on a measurable space $(T, \mathcal{T})$. We say a measurable function $f : T \to \mathbb{R}$ is uniformly integrable almost surely with respect to $\{\mu_n\}_{n=1}^\infty$ if*

$$\lim_{t \to \infty} \limsup_{n \to \infty} \int_{|f| > t} |f| d\mu_n = 0,$$

with probability one. We say a measurable function $f : T \to \mathbb{R}$ is uniformly integrable in probability with respect to $\{\mu_n\}_{n=1}^\infty$ if for every $\varepsilon > 0$

$$\lim_{t \to \infty} \limsup_{n \to \infty} \mathbb{P}\left(\int_{|f| > t} |f| d\mu_n > \varepsilon \right) = 0.$$

Lemma 1.12 (Lemma 4.3 in [16])**.** *Let $(A_n)_{n \geq 1}$ be a sequence of complex random matrices where A_n is $n \times n$ for every $n \geq 1$. Suppose for Lebesgue almost all $z \in \mathbb{C}$, there exists a probability measure ν_z on $[0, \infty)$ such that*

- *a.s. $(\nu_{A_n - z})_{n \geq 1}$ tends weakly to ν_z*
- *a.s. (resp. in probability) $\log(\cdot)$ is uniformly integrable for $(\nu_{A_n - z})_{n \geq 1}$.*

Then there exists a probability measure $\mu \in \mathcal{P}(\mathbb{C})$ such that

- *a.s. (resp. in probability) $(\mu_{A_n})_{n \geq 1}$ converges weakly to μ*
- *for almost every $z \in \mathbb{C}$,*

$$U_\mu(z) = \int_0^\infty \log(t) d\nu_z(t).$$

1.7 Individual entries and stable random vectors.

We now state some useful properties of ξ_1 and ξ_2 implied by Condition **C2**. First, if $(\xi_1^{(1)}, \xi_2^{(1)}), (\xi_1^{(2)}, \xi_2^{(2)}) \ldots$ are i.i.d copies of (ξ_1, ξ_2), then Condition **C1** (i) guarantees, see Theorem 2 in [49], there exists a sequence b_n such that

$$\frac{1}{a_n} \sum_{i=1}^n (\xi_1^{(i)}, \xi_2^{(i)}) - b_n \Rightarrow Z = (Z_1, Z_2),$$

for some α-stable random vector Z with spectral measure θ_d. Condition **C2** (i) guaranties neither Z_1 nor Z_2 is identically 0. We need the following theorem to get results on stable random vectors.

Theorem 1.13 (Theorem 2.3.9 in [50])**.** *Let $(X_1, \ldots, X_d)$ be an α-stable vector in $\mathbb{R}^d$. Then $(X_1, \ldots, X_k)$ is an α-stable random vector for any $k \leq d$.*

From this, with $k = 1$ for real ξ_1 and $k = 2$ for complex ξ_1, we get that ξ_1 is in the domain of attraction of an α-stable random variable and satisfies

$$\mathbb{P}(|\xi_1| \geq t) = L(t) t^{-\alpha} \tag{1.3}$$

for some slowly varying function L. In addition,

$$\lim_{t \to \infty} \mathbb{P}\left(\frac{\xi_1}{|\xi_1|} \in \cdot \,\middle|\, |\xi_1| \geq t \right) = \theta_1(\cdot) \tag{1.4}$$

for some probability measure θ_1, again see [49]. The same holds for ξ_2 with a possibly different probability measure θ_2. Also note

$$\mathbb{E}[|\xi_i|^p] < \infty \tag{1.5}$$

for all $0 \leq p < \alpha$ and $i = 1, 2$.

2 Poisson point processes and stable distributions

In this section we give a brief review of Poisson Point Processes (p.p.p.) and their relation to the order statistics of random variables in the domain of attraction of an α-stable distribution. See [22, 46, 47], and the references therein for proofs.

2.1 Simple point processes

Throughout this section we will assume $E = \bar{\mathbb{R}}^n \setminus \{0\}$ with the relative topology, where $\bar{\mathbb{R}}^n$ is the one point compactification of $\mathbb{R}^n$, but many of the results can be extended to other topological spaces.

Denote by $\mathcal{M}(E)$ the set of simple point Radon measures

$$\mu = \sum_{x \in D} \delta_x$$

where D is such that $D \cap (B_r(0))^c$ is a finite set for any $r > 0$ and $B_r(0)$ is the ball of radius r around the point 0. Denote by $\mathcal{H}(E)$ the set of supports corresponding to measures in $\mathcal{M}(E)$. The elements of $\mathcal{H}(E)$ are called configurations.

Let $C_K(E)$ denote the set of real-valued continuous functions on E with compact support. The vague topology on $\mathcal{M}(E)$ is the topology where a sequence μ_n converges to μ if for any $f \in C_K(E)$

$$\int_E f d\mu_n \to \int_E f d\mu.$$

$\mathcal{M}(E)$ with the vague topology is a Polish space, and thus complete and metrizable.

If one considers the one-to-one function $I : \mathcal{M}(E) \to \mathcal{H}(E)$ given by $I(\mu) = \mathrm{supp}(\mu)$, then the topology of $\mathcal{M}(E)$ can be pushed forward to $\mathcal{H}(E)$. The vague convergence in $\mathcal{M}(E)$ can be stated in terms of a convergence of the supports. Let $\mu_n \xrightarrow{v} \mu$ and give some labeling $\mathrm{supp}(\mu) = \{x^{(1)}, x^{(2)}, \dots\}$. Then this vague convergence implies there exists labeling $\mathrm{supp}(\mu_n) = \{x_n^{(1)}, x_n^{(2)}, \dots\}$ such that for all k, $x_n^{(k)} \to x^{(k)}$. This description will be particularly useful for our case.

A simple point process N is a measurable mapping from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathcal{M}(E), \mathcal{B}(\mathcal{M}(E)))$, where $\mathcal{B}(\mathcal{M}(E))$ is the Borel σ-algebra defined by the vague topology. Weak convergence of simple point processes is defined by weak convergence of the measures in the vague topology.

2.2 Poisson point process

Definition 2.1. *Let m be a Borel measure on E. A point process N is called a Poisson Point Process (p.p.p.) with intensity measure m if for any pairwise disjoint Borel sets $A_1, \dots, A_n$, the random variables $N(A_1), \dots, N(A_n)$ are independent Poisson random variables with expected values $m(A_1), \dots, m(A_n)$.*

Remark 2.2. *N is a.s. simple if and only if m is non-atomic.*

The next proposition makes clear the connection between point processes and stable random variables. See Proposition 3.21 in [46] for a proof. First, we describe an important point process. Let θ be a finite measure on the unit sphere in $\mathbb{R}^n$, and m_α be a measure with density $\alpha r^{-(\alpha+1)} dr$ on $\mathbb{R}_+$. We let N_α denote the p.p.p. on $\mathbb{R}^n$ with intensity measure $\theta \times m_\alpha$.

Proposition 2.3. *Let $\{\xi_n\}_{n \geq 1}$ be i.i.d. $\mathbb{R}^d$-valued random variables. Assume there exists a finite measure θ on the unit sphere in $\mathbb{R}^d$ such that*

$$\lim_{n \to \infty} n \mathbb{P}\left(\frac{\xi_n}{\|\xi_n\|} \in D, \|\xi_n\| \geq r b_n \right) = \theta(D) m_\alpha([r, \infty)),$$

for every $r > 0$ and all Borel subsets D of the unit sphere with $\theta(\partial D) = 0$, where $b_n = n^{1/\alpha} L(n)$ for $0 < \alpha < 2$ and a slowly varying function L. Then

$$\beta_n := \sum_{i=1}^{n} \delta_{\xi_i/b_n} \Rightarrow N_\alpha$$

as $n \to \infty$.

Remark 2.4. In [22], Davydov and Egorov prove this convergence in ℓ^p-type topologies, under a smoothness assumption on the ξ_n.

2.3 Useful properties of N_α

The following are useful and well known properties of the p.p.p. N_α. Again, see Davydov and Egorov, [22], and the references therein for more information and proofs.

Proposition 2.5. *Let $\{\lambda_i\}$ and $\{w_i\}$ be independent i.i.d. sequences where λ_1 has exponential distribution with mean 1, and w_1 is $\frac{1}{\theta(\mathbb{S}^{d-1})}\theta$ distributed, with θ a finite nonzero measure on $\mathbb{S}^{d-1}$. Define $\Gamma_i = \lambda_1 + \cdots + \lambda_i$. Then, for any $\alpha > 0$,*

$$N_\alpha \overset{d}{=} \sum_{i=1}^{\infty} \delta_{\Gamma_i^{-1/\alpha}(\theta(\mathbb{S}^{d-1}))^{1/\alpha} w_i}.$$

Lemma 2.6. *The p.p.p. N_α has the following properties.*

(1) *Almost surely there are only a finite number of points of $\mathrm{supp}(N_\alpha)$ outside a ball of positive radius centered at the origin.*

(2) *N_α is simple.*

(3) *Almost surely, we can label the points from $\mathrm{supp}(N_\alpha)$ according to the decreasing order of their norms $\mathrm{supp}(N_\alpha) = \{(y_1, y_2, \dots) : \|y_1\| > \|y_2\| > \dots \}$.*

(4) *With probability one, for any $p > \alpha$,*

$$\sum_{i=1}^{\infty} |y_i|^p < \infty.$$

3 Bipartized resolvent matrix

In this section we follow the notation of Bordenave, Caputo, and Chafaï in [13] to define the bipartizations of matrices and operators.

3.1 Bipartization of a matrix

For an $n \times n$ complex matrix A we consider a symmetrized version of ν_{A-zI},

$$\check{\nu}_{A-z} := \frac{1}{2n} \sum_{k=1}^{n} (\delta_{s_k(A-z)} + \delta_{-s_k(A-z)}).$$

Let $\mathbb{C}_+ := \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$ and

$$\mathbb{H}_+ := \left\{ U = \begin{pmatrix} \eta & z \\ \bar{z} & \eta \end{pmatrix} : \eta \in \mathbb{C}_+, z \in \mathbb{C} \right\} \subset \mathrm{Mat}_2(\mathbb{C}).$$

For any $z \in \mathbb{C}, \eta \in \mathbb{C}_+$ and $1 \le i, j \le n$ define the following 2×2 matrices

$$U = U(z, \eta) := \begin{pmatrix} \eta & z \\ \bar{z} & \eta \end{pmatrix} \quad \text{and} \quad B_{ij} := \begin{pmatrix} 0 & A_{ij} \\ \bar{A}_{ji} & 0 \end{pmatrix}.$$

Define the matrix $B \in \operatorname{Mat}_n(\operatorname{Mat}_2(\mathbb{C})) \simeq \operatorname{Mat}_{2n}(\mathbb{C})$ by $B = (B_{ij})_{1 \leq i,j \leq n}$. As an element of $\operatorname{Mat}_{2n}(\mathbb{C})$, B is Hermitian. We call B the bipartization of the matrix A. We define the resolvent matrix in $\operatorname{Mat}_n(\operatorname{Mat}_2(\mathbb{C}))$ by[1]

$$R(U) = (B - U \otimes I_n)^{-1},$$

so that for all i, j, $R(U)_{ij} \in \operatorname{Mat}_2(\mathbb{C})$. For $1 \leq k \leq n$ we write,

$$R(U)_{kk} = \begin{pmatrix} a_k(z,\eta) & b_k(z,\eta) \\ b_k'(z\eta) & c_k(z,\eta) \end{pmatrix}.$$

Letting $B(z) = B - U(z,0) \otimes I_n$ we have

$$R(U) = (B(z) - \eta I_{2n})^{-1}.$$

Let m_μ denote the Stieltjes transform of a probability measure μ on $\mathbb{R}$ defined by

$$m_\mu(\eta) = \int_{\mathbb{R}} \frac{1}{x - \eta} d\mu(x), \qquad \eta \in \mathbb{C}_+.$$

Theorem 3.1 (Theorem 2.1 in [13]). *Let $A \in \operatorname{Mat}_n(\mathbb{C})$. Then $\mu_{B(z)} = \check{\nu}_{A-z}$,*

$$m_{\check{\nu}_{A-z}}(\eta) = \frac{1}{2n} \sum_{k=1}^{n} (a_k(z,\eta) + c_k(z,\eta)),$$

and in $\mathcal{D}'(\mathbb{C})$, the set of Schwartz-Sobolev distributions on $\mathbb{C}$ endowed with its usual convergence with respect to all infinitely differentiable functions with compact support,

$$\mu_A(\cdot) = -\frac{1}{\pi n} \sum_{k=1}^{n} \partial b_k(\cdot, 0).$$

3.2 Bipartization of an operator

Let V be a countable set and let $\ell^2(V)$ denote the Hilbert space defined by the inner product

$$\langle \phi, \psi \rangle := \sum_{u \in V} \bar{\phi}_u \psi_u, \quad \phi_u = \langle \delta_u, \phi \rangle,$$

where δ_u is the unit vector supported on $u \in V$. Let $\mathcal{D}(V)$ denote the dense subset of $\ell^2(V)$ of vectors with finite support. Let $(w_{uv})_{u,v \in V}$ be a collection of complex numbers such that for all $u \in V$,

$$\sum_{v \in V} |w_{uv}|^2 + |w_{vu}|^2 < \infty.$$

We then define a linear operator A with domain $\mathcal{D}(V)$ by

$$\langle \delta_u, A\delta_v \rangle = w_{uv}. \tag{3.1}$$

Let $\hat{V}$ be a set in bijection with V, and let $\hat{v} \in \hat{V}$ be the image of $v \in V$ under the bijection. Let $V^b = V \cup \hat{V}$, and define the bipartization of A as the symmetric operator B on $\mathcal{D}(V^b)$ by

$$\langle \delta_u, B\delta_{\hat{v}} \rangle = \overline{\langle \delta_{\hat{v}}, B\delta_u \rangle} = w_{uv},$$

$$\langle \delta_u, B\delta_v \rangle = \langle \delta_{\hat{u}}, B\delta_{\hat{v}} \rangle = 0.$$

Let $\Pi_u : \ell^2(V^b) \to \operatorname{Span}\{\delta_u, \delta_{\hat{u}}\}$ denote the orthogonal projection onto the span of $\delta_u, \delta_{\hat{u}}$. $\operatorname{Span}\{\delta_u, \delta_{\hat{u}}\}$ is isomorphic to $\mathbb{C}^2$ under the map $\delta_u \mapsto e_1, \delta_{\hat{u}} \mapsto e_2$. Under

---

[1]See Subsection 1.5 for the definition of the Kronecker product.

this isomorphism we may think of $\Pi_u B \Pi_v^*$ as a linear map from $\mathbb{C}^2$ to $\mathbb{C}^2$ with matrix representation

$$\Pi_u B \Pi_v^* = \begin{pmatrix} 0 & w_{uv} \\ \bar{w}_{vu} & 0 \end{pmatrix}.$$

Let $B(z) = B - U(z,0) \otimes I_V$. For simplicity we will denote by $B(z)$ the closure of $B(z)$. Recall the sum of an essentially self-adjoint operator and a bounded self-adjoint operator is essentially self-adjoint, thus if B is (essentially) self-adjoint then $B(z)$ is (essentially) self-adjoint. For $\eta \in \mathbb{C}_+$, $U(z,\eta) \in \mathbb{H}_+$ we define the resolvent operator

$$R(U) := (B(z) - \eta I_{V^b})^{-1},$$

and

$$R(U)_{vv} = \Pi_v R(U) \Pi_v^* = \begin{pmatrix} a_v(z,\eta) & b_v(z,\eta) \\ b_v'(z,\eta) & c_v(z,\eta) \end{pmatrix}. \tag{3.2}$$

Lemma 3.2. *If a_v, b_v, c_v, b_v' are defined by (3.2), then*

- *for each $z \in \mathbb{C}$, $a_v(z,\cdot), c_v(z,\cdot) : \mathbb{C}_+ \to \mathbb{C}_+$,*
- *the functions $a_v(z,\cdot), b_v(z,\cdot), b_v'(z,\cdot), c_v(z,\cdot)$ are analytic on $\mathbb{C}_+$,*
- *and*

$$|a_v| \le (\mathrm{Im}(\eta))^{-1}, \quad |c_v| \le (\mathrm{Im}(\eta))^{-1}, \quad |b_v| \le (2\,\mathrm{Im}(\eta))^{-1}, \text{ and } |b_v'| \le (2\,\mathrm{Im}(\eta))^{-1}.$$

Moreover, if $\eta \in i\mathbb{R}_+$, then a_v, c_v are pure imaginary and $b_v' = \bar{b}_v$.

The first statement follows from the fact that the imaginary part of $(B(z) - \eta I_{V^b})^{-1}$ is positive. See Reed and Simon [45] Theorem VIII.2 for a proof of the second statement and Lemma 2.2 in [13] for the last.

4 Convergence to the Poisson weighted infinite tree

4.1 Operators on a tree

Consider a tree $T = (V, E)$ on a vertex set V with edge set E. We say $u \sim v$ if $\{u, v\} \in E$. Assume if $\{u, v\} \notin E$ then $w_{uv} = w_{vu} = 0$, in particular $w_{vv} = 0$ for all $v \in V$. We consider the operator A defined by (3.1). We begin with useful sufficient conditions for a symmetric linear operator to be essentially self-adjoint, which will be very important for our use.

Lemma 4.1 (Lemma A.3 in [12])**.** *Let $\kappa > 0$ and $T = (V, E)$ be a tree. Assume that for all $u, v \in V$, $w_{uv} = \bar{w}_{vu}$ and if $\{u, v\} \notin E$ then $w_{uv} = w_{vu} = 0$. Assume that there exists a sequence of connected finite subsets $(S_n)_{n \ge 1}$ of V, such that $S_n \subset S_{n+1}$, $\bigcup_n S_n = V$, and for every n and $v \in S_n$,*

$$\sum_{u \notin S_n : u \sim v} |w_{uv}|^2 \le \kappa.$$

Then A is essentially self-adjoint.

Corollary 4.2 (Corollary 2.4 in [13])**.** *Let $\kappa > 0$ and $T = (V, E)$ be a tree. Assume that if $\{u, v\} \notin E$ then $w_{uv} = w_{vu} = 0$. Assume there exists a sequence of connected finite subsets $(S_n)_{n \ge 1}$ of V, such that $S_n \subset S_{n+1}$, $\bigcup_n S_n = V$, and for every n and $v \in S_n$,*

$$\sum_{u \notin S_n : u \sim v} (|w_{uv}|^2 + |w_{vu}|^2) \le \kappa.$$

Then for all $z \in \mathbb{C}$, $B(z)$ is self-adjoint.

The advantages of defining an operator on a tree also include the following recursive formula for the resolvents, see Lemma 2.5 in [13] for a proof.

Proposition 4.3. *Assume B is self-adjoint and let $U = U(z, \eta) \in \mathbb{H}_+$. Then*

$$R(U)_{\varnothing\varnothing} = -\left(U + \sum_{v \sim \varnothing} \begin{pmatrix} 0 & w_{\varnothing v} \\ \bar{w}_{v\varnothing} & 0 \end{pmatrix} \tilde{R}(U)_{vv} \begin{pmatrix} 0 & w_{v\varnothing} \\ \bar{w}_{\varnothing v} & 0 \end{pmatrix}\right)^{-1}, \qquad (4.1)$$

where $\tilde{R}(U)_{kk} := \Pi_v R_{B_v} \Pi_v^*$ where B_v is the bipartization of the operator A restricted to $\ell^2(V_v)$, and V_v is the subtree of V of vertices whose path to $\varnothing$ contains v.

4.2 Local operator convergence

We now define a useful type of convergence.

Definition 4.4 (Local Convergence). *Suppose (A_n) is a sequence of bounded operators on $\ell^2(V)$ and A is a linear operator on $\ell^2(V)$ with domain $\mathcal{D}(A) \supset \mathcal{D}(V)$. For any $u, v \in V$ we say that (A_n, u) converges locally to (A, v), and write*

$$(A_n, u) \to (A, v),$$

if there exists a sequence of bijections $\sigma_n : V \to V$ such that $\sigma_n(v) = u$ and, for all $\phi \in \mathcal{D}(V)$,

$$\sigma_n^{-1} A_n \sigma_n \phi \to A\phi,$$

in $\ell^2(V)$, as $n \to \infty$.

Here we use σ_n for the bijection on V and the corresponding linear isometry defined in the obvious way. This notion of convergence is useful to random matrices for two reasons. First, we will make a choice on how to define the action of an $n \times n$ matrix on $\ell^2(V)$, and the bijections σ_n help ensure the choice of location for the support of the matrix does not matter. Second, local convergence also gives convergence of the resolvent operator at the distinguished points $u, v \in V$. This comes down to the fact that local convergence is strong operator convergence, up to the isometries. See [13] for details.

Theorem 4.5 (Theorem 2.7 in [13]). *Assume $(A_n, u) \to (A, v)$ for some $u, v \in V$. Let B_n be the self-adjoint bipartized operator of A_n. If the bipartized operator B of A is self-adjoint and $\mathcal{D}(V^b)$ is a core for B (i.e., the closure B restricted to $\mathcal{D}(V^b)$ is B), then for all $U \in \mathbb{H}_+$,*

$$R_{B_n}(U)_{uu} \to R_B(U)_{vv}.$$

To apply this to random operators we say that $(A_n, u) \to (A, v)$ in distribution if there exists a sequence of random bijections σ_n such that $\sigma_n^{-1} A_n \sigma_n \phi \to A\phi$ in distribution for every $\phi \in \mathcal{D}(V^b)$.

4.3 Poisson weighted infinite tree (PWIT)

Let ρ be a positive Radon measure on $\mathbb{C}^n \setminus \{0\}$ such that $\rho(\mathbb{C}^n \setminus \{0\}) = \infty$. $\mathbf{PWIT}(\rho)$ is the random weighted rooted tree defined as follows. The vertex set of the tree is identified with $\mathbb{N}^f := \bigcup_{k \in \mathbb{N} \cup \{0\}} \mathbb{N}^k$ by indexing the root as $\mathbb{N}^0 = \varnothing$, the offspring of the root as $\mathbb{N}$ and, more generally, the offspring of some $v \in \mathbb{N}^k$ as $(v1), (v2), \cdots \in \mathbb{N}^{k+1}$. Define T as the tree on $\mathbb{N}^f$ with edges between parents and offspring. Let $\{\Xi_v\}_{v \in \mathbb{N}^f}$ be independent realizations of a Poisson point process with intensity measure ρ. Let $\Xi_\varnothing = \{y_1, y_2, \dots\}$ be ordered such that $\|y_1\| \geq \|y_2\| \geq \cdots$, and assign the weight y_i to the edge between $\varnothing$ and i, assuming such an ordering is possible. More generally assign the weight y_{vi} to the edge between v and vi where $\Xi_v = \{y_{v1}, y_{v2}, \dots\}$ where $\|y_{v1}\| \geq \|y_{v2}\| \geq \cdots$.

Consider a realization of $\mathbf{PWIT}(\theta_d \times m_\alpha)$, with $y_{vk} = (y_{vk}^{(1)}, y_{vk}^{(2)})$. Even though the measure $\theta_d \times m_\alpha$ has a more natural representation in polar coordinates, we let $(y_{vk}^{(1)}, y_{vk}^{(2)})$ be the Cartesian coordinates of y_{vk}. Define an operator A on $\mathcal{D}(\mathbb{N}^f)$ by the formulas

$$\langle \delta_v, A\delta_{vk} \rangle = y_{vk}^{(1)}, \quad \text{and} \quad \langle \delta_{vk}, A\delta_v \rangle = y_{vk}^{(2)} \tag{4.2}$$

and $\langle \delta_v, A\delta_u \rangle = 0$ otherwise. For $0 < \alpha < 2$, we know by Lemma 2.6 that the points in Ξ_v are almost surely square summable for every $v \in \mathbb{N}^f$, and thus A is actually a well defined linear operator on $\mathcal{D}(\mathbb{N}^f)$, though is possibly unbounded on $\ell^2(\mathbb{N}^f)$. Before showing the local convergence of the random matrices A_n to A we will show the bipartization of A is self-adjoint.

Proposition 4.6. *With probability one, for all $z \in \mathbb{C}$, $B(z)$ is self-adjoint, where $B(z)$ is the bipartization of the operator A defined by (4.2).*

We begin with a lemma on point processes, proved in Lemma A.4 from [12], before checking our criterion for self-adjointness.

Lemma 4.7. *Let $\kappa > 0, 0 < \alpha < 2$ and let $0 < x_1 < x_2 < \cdots$ be a Poisson process of intensity 1 on $\mathbb{R}_+$. Define $\tau = \inf\{t \in \mathbb{N} : \sum_{k=t+1}^\infty x_k^{-2/\alpha} \leq \kappa\}$. Then $\mathbb{E}\tau$ is finite and goes to 0 as κ goes to infinity.*

Proof of Proposition 4.6. For $\kappa > 0$ and $v \in \mathbb{N}^f$ define

$$\tau_v = \inf\left\{ t \geq 0 : \sum_{k=t+1}^\infty \|y_{vk}\|^2 \leq \kappa \right\}.$$

Note if N is a homogeneous Poisson process on $\mathbb{R}_+$ with intensity 1 and $f(x) = x^{-1/\alpha}$, then $f(N)$ is Poisson process with intensity measure $\alpha r^{-1-\alpha} dr$. Thus by Lemma 4.7, $\kappa > 0$ can be chosen such that $\mathbb{E}\tau_v < 1$ for any fixed v. Since the random variables $\{\tau_v\}_{v \in \mathbb{N}^f}$ are i.i.d. this κ works for all v. Fix such a κ. We now color the vertices red and green in an effort to build the sets S_n in Corollary 4.2. Put a green color on all vertices v such that $\tau_v \geq 1$ and a red color otherwise. Define the sub-forest T^g of T where an edge between v and vk is included if v is green and $1 \leq k \leq \tau_v$. If the root $\varnothing$ is red let $S_1 = \{\varnothing\}$. Otherwise let $T_\varnothing^g = (V_\varnothing^g, E_\varnothing^g)$ be the subtree of T^g containing $\varnothing$. Let Z_n denote the number of vertices in $T_\varnothing^g$ at a depth n from the root, then Z_n is a Galton-Watson process with offspring distribution $\tau_\varnothing$. It is well known that if $\mathbb{E}\tau_\varnothing < 1$, then the tree is almost surely finite, see Theorem 5.3.7 in [23]. Let $L_\varnothing^g$ be the leaves of the tree $T_\varnothing^g$. Set $S_1 := \bigcup_{v \in L_\varnothing^g} \{vk : 1 \leq k \leq \tau_v\} \bigcup V_\varnothing^g$. It is clear for any $v \in S_1$,

$$\sum_{u \notin S_1 : u \sim v} (|y_{uv}|^2 + |y_{vu}|^2) \leq \kappa.$$

Now define the outer boundary of $\{\varnothing\}$ as $\partial_\tau \{\varnothing\} = \{1, \ldots, \max(\tau_\varnothing, 1)\}$ and for $v = (i_1 \cdots i_k)$ set $\partial_\tau \{v\} = \{(i_1 \cdots i_{k-1}(i_k + 1))\} \cup \{(i_1 \cdots i_k 1), \ldots, (i_1 \cdots i_k \max(\tau_v, 1))\}$. For a connected set S define its outer boundary as

$$\partial_\tau S = \left(\bigcup_{v \in S} \partial_\tau \{v\} \right) \setminus S.$$

Now for each $u_1, \ldots, u_k \in \partial_\tau S_1$ apply the above process to get subtrees $\{T_{u_i}^g = (V_{u_i}^g, E_{u_i}^g)\}_{i=1}^k$ with roots u_i and the leaves of tree $T_{u_i}^g$ denoted by $L_{u_i}^g$. Set

$$S_2 := S_1 \cup \left(\bigcup_{i=1}^k (V_{u_i}^g \cup_{v \in L_{u_i}^g} \{vj : 1 \leq j \leq \tau_v\}) \right).$$

Apply this procedure iteratively to get the sequence of subsets $(S_n)_{n \geq 1}$. Apply Corollary 4.2 to complete the proof. $\square$

4.4 Local convergence

For an $n \times n$ matrix M we aim to define M as a bounded operator on $\ell^2(\mathbb{N}^f)$. For $1 \leq i, j, \leq n$, let $\langle \delta_i, M\delta_j \rangle = M_{ij}$ and $\langle \delta_u, M\delta_v \rangle = 0$ otherwise.

Theorem 4.8. *Let $(P_n)_{n \geq 1}$ be a sequence of uniformly distributed random $n \times n$ permutation matrices, independent of A_n. Then, in distribution, $(P_n A_n P_n^T, 1) \to (A, \varnothing)$ where A is the operator defined by (4.2).*

Remark 4.9. This theorem holds for any sequence of permutation matrices P_n regardless of independence or distribution, but for our use later it is important they are independent of A_n and uniformly distributed. This is to get around the fact the entries of A_n are not exchangeable.

The rest of this subsection is devoted to the proof of Theorem 4.8.

Proof of Theorem 4.8. The procedure follows along the same lines as Bordenave, Caputo, and Chafaï in [12] and [13]. We will define a network as a graph with edge weights taking values in some normed space. To begin let G_n be the complete network on $\{1, \ldots, n\}$ whose weight on edge $\{i, j\}$ equals ξ_{ij}^n for some collection $(\xi_{ij}^n)_{1 \leq i \leq j \leq n}$ of i.i.d. random variables taking values in some normed space. Now consider the rooted network $(G_n, 1)$ with the distinguished vertex 1. For any realization (ξ_{ij}^n), and for any $B, H \in \mathbb{N}$ such that $(B^{H+1} - 1)/(B - 1) \leq n$, we will define a finite rooted subnetwork $(G_n, 1)^{B,H}$ of $(G_n, 1)$ whose vertex set coincides with a B-ary tree of depth H. To this end we partially index the vertices of $(G_n, 1)$ as elements in

$$J_{B,H} := \bigcup_{l=0}^{H} \{1, \ldots, B\}^l \subset \mathbb{N}^f,$$

the indexing being given by an injective map σ_n from $J_{B,H}$ to $V_n := \{1, \ldots, n\}$. We set $I_\varnothing := \{1\}$ and the index of the root $\sigma_n^{-1}(1) = \varnothing$. The vertex $v \in V_n \setminus I_\varnothing$ is given the index $(k) = \sigma_n^{-1}(v), 1 \leq k \leq B$, if $\xi_{1,v}^n$ has the k-th largest norm value among $\{\xi_{1j}^n, j \neq 1\}$, ties being broken by lexicographic order[2]. This defines the first generation, and let I_1 be the union of $I_\varnothing$ and this generation. If $H \geq 2$ repeat this process for the vertex labeled (1) on $V_n \setminus I_1$ to order $\{\xi_{(1)j}^n\}_{j \in V_n \setminus I_1}$ to get $\{11, 12, \ldots, 1B\}$. Define I_2 to be the union of I_1 and this new collection. Repeat again for $(2), (3), \ldots, (B)$ to get the second generation and so on. Call this vertex set $V_n^{B,H} = \sigma_n J_{B,H}$.

For a realization T of $\mathbf{PWIT}(\rho)$, recall we assign the weight y_{vk} to the edge $\{v, vk\}$. Then $(T, \varnothing)$ is a rooted network. Call $(T, \varnothing)^{B,H}$ the finite rooted subnetwork obtained by restricting $(T, \varnothing)$ to the vertex set $J_{B,H}$. If an edge is not present in $(T, \varnothing)^{B,H}$ assign the weight 0. We say a sequence $(G_n, 1)^{B,H}$, for fixed B and H, converges in distribution, as $n \to \infty$, to $(T, \varnothing)^{B,H}$ if the joint distribution of the weights converges weakly.

Let π_n be the permutation on $\{1, \ldots, n\}$ associated to the permutation matrix P_n. We let

$$\xi_{ij}^n = \left(\xi_{ij}^{n,(1)}, \xi_{ij}^{n,(2)} \right) := \left(\frac{X_{\pi_n(i), \pi_n(j)}}{a_n}, \frac{X_{\pi_n(j), \pi_n(i)}}{a_n} \right).$$

We now consider $(G_n, 1)^{B,H}$ with weights ξ_{ij}^n and a realization, T, of $\mathbf{PWIT}(\theta_d \times m_\alpha)$. We aim to show $(G_n, 1)^{B,H}$ converges in distribution to $(T, \varnothing)^{B,H}$, for fixed B, H as $n \to \infty$.

Order the elements of $J_{B,H}$ lexicographically, i.e. $\varnothing \prec 1 \prec 2 \prec \cdots \prec B \prec 11 \prec 12 \prec \cdots \prec B \cdots B$. For $v \in J_{B,H}$ let $\mathcal{O}_v$ denote the offspring of v in $(G, 1)^{B,H}$. By construction $I_\varnothing = \{1\}$ and $I_v = \sigma_n \left(\bigcup_{w \prec v} \mathcal{O}_w \right)$, where $w \prec v$ must be strict in this union. Thus at every

---

[2]To help keep track of notation in this section, note that $v = (w) \in V_n$ if $w \in J_{B,H}$ and $\sigma_n(w) = v$.

step of the indexing procedure we order the weights of neighboring edges not already considered at a previous step. Thus for all v,

$$(\xi^n_{\sigma_n(v),j})_{j \notin I_v} \overset{d}{=} (\xi^n_{1j})_{1 < j \leq n - |I_v|}.$$

Note that by independence, Proposition 2.3 still holds if you take the empirical sum over $\{1, \ldots, n\} \setminus I$ for any fixed finite set I. Thus by Proposition 2.3 the weights from a fixed parent to its offspring in $(G_n, 1)^{B,H}$ converge weakly to those of $(T, \varnothing)^{B,H}$. By independence we can extend this to joint convergence. Because for any fixed n, $(G_n, 1)^{B,H}$ is still a complete network on $V_n^{B,H}$, we must now check the weights connected to vertices not indexed above converge to zero, which is the weight given to edges not in the tree. For $v, w \in J_{B,H}$ define

$$x^n_{v,w} := \xi^n_{\sigma_n(v), \sigma_n(w)}.$$

Also let $\{z^n_{v,w}, v, w \in J_{B,H}\}$ denote independent variables distributed as $\|\xi^n_{12}\|$. Let $E^{B,H}$ denote the set of edges $\{v, w\} \in J_{B,H} \times J_{B,H}$ that do not belong to the finite subtree $(T, \varnothing)^{B,H}$. Because we have sorted out the largest elements, the vector $\{\|x^n_{v,w}\|, \{v, w\} \in E^{B,H}\}$ is stochastically dominated by the vector $Z_n := \{z^n_{v,w}, \{v, w\} \in E^{B,H}\}$ (see [12] Lemma 2.7). Since $J_{B,H}$ is finite, the vector Z_n converges to 0 as $n \to \infty$. Thus $(G, 1)^{B,H} \Rightarrow (T, \varnothing)^{B,H}$.

Let A be the operator associated to $\mathbf{PWIT}(\theta_d \times m_\alpha)$ defined by (4.2). For fixed B, H let $\sigma_n^{B,H}$ be the map σ_n above associated to $(G, 1)^{B,H}$, and arbitrarily extend $\sigma_n^{B,H}$ to a bijection on $\mathbb{N}^f$, where V_n is considered in the natural way as a subset of the offspring of $\varnothing$. From the Skorokhod Representation Theorem we may assume $(G_n, 1)^{B,H}$ converges almost surely to $(T, \varnothing)^{B,H}$. Thus there is a sequences B_n, H_n tending to infinity and $\hat{\sigma}_n := \sigma_n^{B_n, H_n}$ such that for any pair $v, w \in \mathbb{N}^f$, $\xi^n_{\hat{\sigma}_n(v), \hat{\sigma}_n(w)}$ converges almost surely to

$$\begin{cases} y_{vk}, & \text{if } w = vk \text{ for some } k \\ y_{wk}, & \text{if } v = wk \text{ for some } k \\ 0, & \text{otherwise.} \end{cases}$$

Thus almost surely

$$\langle \delta_v, \hat{\sigma}_n^{-1} P_n A_n P_n^T \hat{\sigma}_n \delta_w \rangle = \xi^{n,(i)}_{\hat{\sigma}_n(v), \hat{\sigma}_n(w)} \to \langle \delta_v, A\delta_w \rangle,$$

where $(i) = 1, 2$ depending on whether w is an offspring of v, vice versa, or we take the convention $i = 1$ if neither in which case $\xi^{n,(1)}_{\hat{\sigma}_n(v), \hat{\sigma}_n(w)} \to 0$. To prove the local convergence of operators it is sufficient, by linearity, to prove point wise convergence for any δ_w. For convenience let $\phi_n^w := \hat{\sigma}_n^{-1} P_n A_n P_n^T \hat{\sigma}_n \delta_w$. Thus all that remains to be shown to complete the proof of Theorem 4.8 is that almost surely as $n \to \infty$

$$\sum_{u \in \mathbb{N}^f} |\langle \delta_u, \phi_n^w \rangle - \langle \delta_u, A\delta_w \rangle|^2 \to 0.$$

Since for every u, $\langle \delta_u, \phi_n^w \rangle \to \langle \delta_u, A\delta_w \rangle$ and $\langle \delta_u, A\delta_w \rangle$ is square summable in u it is enough to show that if $u \in \mathbb{N}^f$ are given some indexing by $\mathbb{N}$ $u_1, u_2, \ldots$, then

$$\sup_{n \geq 1} \sum_{i=k}^{\infty} |\langle \delta_{u_i}, \phi_n^w \rangle|^2 \to 0$$

as $k \to \infty$. This follows from the uniform square-integrability of order statistics, see Lemma 2.4 of [12]. This completes the proof. $\square$

4.5 Resolvent matrix

Let R be the resolvent of the bipartized random operator of A. For $U(z,\eta) \in \mathbb{H}_+$, set

$$R(U)_{\varnothing\varnothing} = \begin{pmatrix} a(z,\eta) & b(z,\eta) \\ b'(z,\eta) & c(z,\eta) \end{pmatrix}. \tag{4.3}$$

We have the following result.

Theorem 4.10. *Let P_n, A_n, and A be as in the Theorem 4.8. Since $B(z)$ is almost surely self-adjoint we may almost surely define R, the resolvent of $B(z)$. Let $\hat{R}_n$ be the resolvent of the bipartized matrix of $P_n A_n P_n^T$. For all $U \in \mathbb{H}_+$,*

$$\hat{R}_n(U)_{11} \Rightarrow R(U)_{\varnothing\varnothing}$$

as $n \to \infty$.

Proof. This follows immediately from Theorem 4.8, Proposition 4.6, and Theorem 4.5. $\square$

As functions, the entries of the resolvent matrix are continuous, and by Lemma 3.2 are bounded. Thus

$$\lim_{n\to\infty} \mathbb{E}\hat{R}_n(U)_{11} = \mathbb{E}R(U)_{\varnothing\varnothing}. \tag{4.4}$$

Note that by independence of π_n and A_n

$$\mathbb{E}\hat{R}_n(U)_{11} = \mathbb{E}R_n(U)_{\pi_n^{-1}(1)\pi_n^{-1}(1)} = \mathbb{E}\frac{1}{n}\sum_{i=1}^{n} R_n(U)_{ii}. \tag{4.5}$$

This is the reason for the choice of uniformly distributed P_n independent of A_n.

Theorem 4.11. *For all $z \in \mathbb{C}$, almost surely the measures $\check{\nu}_{A_n-z}$ converge weakly to a deterministic probability measure $\check{\nu}_{\alpha,z,\theta_d}$ whose Stieltjes transform is given by*

$$m_{\check{\nu}_{\alpha,z,\theta_d}}(\eta) = \mathbb{E}a(z,\eta),$$

for $\eta \in \mathbb{C}_+$ and $a(z,\eta)$ in (4.3).

Proof. By Proposition 4.6 for every $z \in \mathbb{C}$, $B(z)$ is almost surely essentially self-adjoint. Thus, using the Borel functional calculus, there exists almost surely a random probability measure $\nu_{\varnothing,z}$ on $\mathbb{R}$ such that

$$a(z,\eta) = \langle \delta_\varnothing, R(U)\delta_\varnothing \rangle = \int_\mathbb{R} \frac{d\nu_{\varnothing,z}(x)}{x-\eta} = m_{\nu_{\varnothing,z}}(\eta).$$

See Theorem VIII.5 in [45] for more on this measure and the Borel functional calculus for unbounded self-adjoint operators. Define R_n as the resolvent matrix of $B_n(z)$, the bipartized matrix of A_n. For $U(z,\eta) \in \mathbb{H}_+$, we write

$$R_n(U)_{kk} = \begin{pmatrix} a_k(z,\eta) & b_k(z,\eta) \\ b'_k(z\eta) & c_k(z,\eta) \end{pmatrix}, \qquad \hat{R}_n(U)_{kk} = \begin{pmatrix} \hat{a}_k(z,\eta) & \hat{b}_k(z,\eta) \\ \hat{b}'_k(z\eta) & \hat{c}_k(z,\eta) \end{pmatrix}.$$

By Theorem 3.1

$$m_{\mathbb{E}\check{\nu}_{A-z}}(\eta) = \mathbb{E}\frac{1}{2n}\sum_{k=1}^{n}(a_k(z,\eta) + c_k(z,\eta)) = \frac{1}{2}\left(\mathbb{E}\hat{a}_1(z,\eta) + \mathbb{E}\hat{c}_1(z,\eta)\right) = \mathbb{E}\hat{a}_1(z,\eta).$$

Thus, by (4.4) and (4.5),

$$\lim_{n\to\infty} m_{\mathbb{E}\check{\nu}_{A-z}}(\eta) = \mathbb{E}a(z,\eta).$$

It follows that $\mathbb{E}\check{\nu}_{A_n-z}$ converges to some deterministic probability measure $\check{\nu}_{z,\alpha,\theta_d} = \mathbb{E}\nu_{\varnothing,z}$. By Lemma A.2 $\check{\nu}_{A_n-z}$ concentrates around its expected value, and thus by the Borel-Cantelli Lemma $\check{\nu}_{A_n-z}$ converges almost surely to $\check{\nu}_{z,\alpha,\theta_d}$. $\square$

Proof of Theorem 1.5. Theorem 1.5 follows immediately from Theorem 4.11. □

We conclude this section with a recursive distributional equation describing $R(U)_{\varnothing\varnothing}$.

Proposition 4.12. *Let θ_d be a probability measure. Let $(w^{(1)}, w^{(2)})$ be a θ_d distributed random vector independent of the matrix given in (4.3). Additionally let $\rho_{z,\eta}$ be the measure on $\mathbb{C}^4 \cong \mathbb{R}^8$ which is the distribution of the random vector*

$$
\begin{pmatrix}
c(z,\eta)|w^{(1)}|^2 \\
b'(z,\eta)w^{(1)}w^{(2)} \\
b(z,\eta)\bar{w}^{(1)}\bar{w}^{(2)} \\
a(z,\eta)|w^{(2)}|^2
\end{pmatrix},
\tag{4.6}
$$

where $a(z,\eta), b(z,\eta), b'(z,\eta)$, and $c(z,\eta)$ are as in (4.3). Then the matrix given in (4.3) satisfies the following recursive distributional equation

$$
\begin{pmatrix}
a(z,\eta) & b(z,\eta) \\
b'(z,\eta) & c(z,\eta)
\end{pmatrix}
\overset{d}{=}
-\left(
\begin{pmatrix}
\eta & z \\
\bar{z} & \eta
\end{pmatrix}
+
\begin{pmatrix}
S_1 & S_2 \\
S_3 & S_4
\end{pmatrix}
\right)^{-1}
\tag{4.7}
$$

where $S = (S_1, S_2, S_3, S_4)$ is an $\alpha/2$-stable random vector with spectral measure is given by the image of the measure $\frac{\Gamma(2-\alpha/2)\cos(\pi\alpha/4)}{1-\alpha/2}\|v\|^{\alpha/2}d\rho_{z,\eta}(v)$ under the map $v \mapsto \frac{v}{\|v\|}$.

The proof of Proposition 4.12 is given in Appendix A.4.

5 Least singular values of elliptic random matrices

Now that we have proven Theorem 1.5, we move on to show that $\log(\cdot)$ is uniformly integrable, in probability, with respect to $\{\nu_{A_n-zI_n}\}_{n\geq 1}$. We begin with a bound on the least singular value of an elliptic random matrix under very general assumptions. This section is entirely self contained.

Theorem 5.1 (Least singular value bound). *Let $X = (X_{ij})$ be an $n \times n$ complex-valued random matrix such that*

(i) (off-diagonal entries) $\{(X_{ij}, X_{ji}) : 1 \leq i < j \leq n\}$ is a collection of independent random tuples,

(ii) (diagonal entries) the diagonal entries $\{X_{ii} : 1 \leq i \leq n\}$ are independent of the off-diagonal entries (but can be dependent on each other),

(iii) there exists $a > 0$ such that the events

$$
\mathcal{E}_{ij} := \{|X_{ij}| \leq a, |X_{ji}| \leq a\}
\tag{5.1}
$$

defined for $i \neq j$ satisfy

$$
b := \min_{i<j}\mathbb{P}(\mathcal{E}_{ij}) > 0, \quad \sigma^2 := \min_{i \neq j}\mathrm{Var}(X_{ij} \mid \mathcal{E}_{ij}) > 0,
$$

and

$$
\rho := \max_{i<j}|\mathrm{Corr}(X_{ij} \mid \mathcal{E}_{ij}, X_{ji} \mid \mathcal{E}_{ji})| < 1.
$$

Then there exists $C = C(a, b, \sigma) > 0$ such that for $n \geq C$, any $M \in \mathrm{Mat}_n(\mathbb{C})$, $s \geq 1$, $0 < t \leq 1$,

$$
\mathbb{P}\left(s_n(X+M) \leq \frac{t}{\sqrt{n}}, s_1(X+M) \leq s\right) \leq C\left(\frac{\log(Cns)}{\sqrt{1-\rho}}\left(\sqrt{s^5 t} + \frac{1}{\sqrt{n}}\right)\right)^{1/4}.
$$

Remark 5.2. The constant C from Theorem 5.1 only depends on a, b, σ and does not depend on ρ. This allows one to apply Theorem 5.1 to cases where ρ depends on n.

While a number of results are known for the smallest singular value of elliptic random matrices with light-tailed entries [30, 40, 42], we are not aware of any results for the heavy-tailed case studied here. Theorem 5.1 can be applied in the special case when X contains i.i.d. heavy-tailed entries, but in this case stronger bounds are known [13, 16], including in the case when X is a rectangular matrix [56]. In fact, Louvaris [37] recently obtained the asymptotic distribution of the smallest singular value for a class of non-Hermitian random matrices with i.i.d. entries following a stable law. Least singular value bounds are also known for several other models of random matrices with dependent entries, e.g., [14, 15, 41], however, none of these results apply to elliptic random matrices.

5.1 Proof of Theorem 5.1

In this section we prove Theorem 5.1. Suppose X and M satisfy the assumptions of the theorem and denote $A := X + M$. We will use A throughout this section and it may be worth noting it is not the A_n of Theorems 1.5 and 1.6. Throughout the section, we allow all constants to depend on a, b, σ without mentioning or denoting this dependence. Constants, however, will not depend on ρ; instead we will state all dependence on ρ explicitly.

For the proof of Theorem 5.1, it suffices to assume that A and every principal submatrix of A is invertible with probability 1. To see this, define $X' := X + \frac{t}{\sqrt{n}} \xi I$, where I is the identity matrix and ξ is a real-valued random variable uniformly distributed on the interval $[-1, 1]$, independent of X. It follows that X' satisfies the assumptions of Theorem 5.1. However, since ξ is continuously distributed, it also follows that $A' := X' + M$ and every principal submatrix of A' is invertible with probability 1. By Weyl's inequality for the singular values (see, for instance, [10, Problem III.6.13]), we find

$$\max_{1 \le k \le n} |s_k(A) - s_k(A')| \le \frac{t}{\sqrt{n}} \le s.$$

Hence, we conclude that

$$\mathbb{P}\left(s_n(A) \le \frac{t}{\sqrt{n}}, s_1(A) \le s\right) \le \mathbb{P}\left(s_n(A') \le \frac{2t}{\sqrt{n}}, s_1(A') \le 2s\right).$$

In other words, it suffices to prove Theorem 5.1 under the additional assumption that A and every principal submatrix of A is invertible. We work under this additional assumption for the remainder of the proof.

5.2 Nets and a decomposition of the unit sphere

Consider a compact set $K \subset \mathbb{C}^n$ and $\varepsilon > 0$. A subset $\mathcal{N} \subset K$ is called an ε-*net* of K if for every point $v \in K$ one has $\text{dist}(v, \mathcal{N}) \le \varepsilon$.

For some real positive parameters $\delta, \tau > 0$ that will be determined later, we define the set of δ-*sparse vectors* as

$$\text{Sparse}(\delta) := \{x \in \mathbb{C}^n : |\text{supp}(x)| \le \delta n\}.$$

We decompose the unit sphere $\mathbb{S}^{n-1}$ into the set of *compressible vectors* and the complementary set of *incompressible* vectors by

$$\text{Comp}(\delta, \tau) := \{x \in \mathbb{S}^{n-1} : \text{dist}(x, \text{Sparse}(\delta)) \le \tau\}$$

and

$$\text{Incomp}(\delta, \tau) := \mathbb{S}^{n-1} \setminus \text{Comp}(\delta, \tau).$$

We have the following result for incompressible vectors.

Lemma 5.3 (Incompressible vectors are spread; Lemma A.3 from [16]). *Let* $x \in \mathrm{Incomp}(\delta, \tau)$. *There exists a subset* $\pi \subset [n]$ *such that* $|\pi| \geq \delta n/2$ *and for all* $i \in \pi$,

$$\frac{\tau}{\sqrt{n}} \leq |x_i| \leq \sqrt{\frac{2}{\delta n}}.$$

5.3 Control of compressible vectors

The case of compressible vectors roughly follows the arguments from [16]. For $i, j \in [n]$, we let C_j denote the j-th column of A and $C_j^{(i)}$ denote the j-th column of A with the i-th entry removed.

Lemma 5.4 (Distance of a random vector to a deterministic subspace). *There exist constants* $\varepsilon, C, c, \delta_0 > 0$ *such that for all* $1 \leq i \leq n$ *and any deterministic subspace* H *of* $\mathbb{C}^{n-1}$ *with* $1 \leq \dim(H) \leq \delta_0 n$, *we have*

$$\mathbb{P}\left(\mathrm{dist}(C_i^{(i)}, H) \leq \varepsilon\sqrt{n}\right) \leq C\exp(-cn).$$

Proof. The proof follows the same arguments as those given in the proof of [16, Theorem A.2]. Fix $1 \leq i \leq n$; the arguments and bounds below are all uniform in i. Recall the definitions of the events $\mathcal{E}_{ij}$ given in (5.1). By the assumptions on $\mathcal{E}_{ij}$, the Chernoff bound gives

$$\mathbb{P}\left(\sum_{j \neq i} \mathbf{1}_{\mathcal{E}_{ji}} \leq \frac{(n-1)b}{2}\right) \leq \exp\left(-\frac{(n-1)b}{8}\right).$$

In other words, with high probability, at least $m := \left\lceil \frac{(n-1)b}{2} \right\rceil$ of the events $\mathcal{E}_{ji}$, $j \neq i$ occur. Thus, it suffices to prove the result by conditioning on the event

$$E_m := \bigcap_{j \in [m], j \neq i} \mathcal{E}_{ji}. \tag{5.2}$$

There are two cases to consider. Either $i \in [m]$ or $i > m$. The arguments for these two cases are almost identical except some notations must be changed slightly to remove the i-th index. For the remainder of the proof, let us only consider the case when $i > m$; the changes required for the other case are left to the reader.

Recall that it suffices to prove the result by conditioning on the event E_m defined in (5.2). In fact, as $i > m$, the definition of the event E_m given in (5.2) can be stated as

$$E_m := \bigcap_{j \in [m]} \mathcal{E}_{ji}.$$

Let $\mathbb{E}_m[\cdot] := \mathbb{E}[\cdot \mid E_m, \mathcal{F}_m]$ denote the conditional expectation given the event E_m and the filtration $\mathcal{F}_m$ generated by X_{ji}, $j > m$, $j \neq i$. Let W be the subspace spanned by H and the vectors

$$u := (0, \ldots, 0, X_{m+1,i}, \ldots, X_{i-1,i}, X_{i+1,i}, \ldots, X_{n,i})$$

and

$$w := (\mathbb{E}_m[X_{1,i}], \ldots, \mathbb{E}_m[X_{m,i}], 0, \ldots, 0).$$

By construction, $\dim(W) \leq \dim(H) + 2$ and W is $\mathcal{F}_m$ measurable. In addition,

$$\mathrm{dist}(C_i^{(i)}, H) \geq \mathrm{dist}(C_i^{(i)}, W) = \mathrm{dist}(Y, W),$$

where

$$Y := (X_{1,i} - \mathbb{E}_m[X_{1,i}], \ldots, X_{m,i} - \mathbb{E}_m[X_{m,i}], 0, \ldots, 0) = C_i^{(i)} - u - w.$$

By assumption, the coordinates $Y_1, \ldots, Y_m$ are independent random variables which satisfy

$$|Y_k| \leq 2a, \qquad E_m[Y_k] = 0, \qquad \mathbb{E}_m|Y_k|^2 \geq \sigma^2$$

for $1 \leq k \leq m$. Thus, since the function $x \mapsto \operatorname{dist}(x, W)$ is convex and 1-Lipschitz, Talagrand's concentration inequality (see for instance [53, Theorem 2.1.13]) yields

$$\mathbb{P}_m(|\operatorname{dist}(Y, W) - \mathbb{E}_m[\operatorname{dist}(Y, W)]| \geq t) \leq C \exp\left(-c\frac{t^2}{a^2}\right) \qquad (5.3)$$

for every $t \geq 0$, where $C, c > 0$ are absolute constants. In particular, this implies that

$$(\mathbb{E}_m \operatorname{dist}(Y, W))^2 \geq \mathbb{E}_m \operatorname{dist}^2(Y, W) - c'a^2 \qquad (5.4)$$

for an absolute constant $c' > 0$. Thus, if P denotes the orthogonal projection onto the orthogonal complement of W, we get

$$
\begin{aligned}
\mathbb{E}_m \operatorname{dist}^2(Y, W) &= \sum_{k=1}^{m} \mathbb{E}_m|Y_k|^2 P_{kk} \\
&\geq \sigma^2 \left(\operatorname{tr} P - \sum_{k=m+1}^{n-1} P_{kk} \right) \\
&\geq \sigma^2((n-1) - \dim(H) - 2 - (n - 1 - m)) \\
&\geq \sigma^2 \left(\frac{(n-1)b}{2} - \delta_0 n - 2 \right).
\end{aligned}
$$

The last term is lower bounded by $c''\sigma^2 n$ for all n sufficiently large by taking $\delta_0 := b/4$. Combining this bound with (5.3) and (5.4) completes the proof. $\qquad \square$

The next bound, which follows as a corollary of Lemma 5.4, will be useful when dealing with compressible vectors.

Corollary 5.5. *There exist $\varepsilon, C, c, \delta_0 > 0$ such that for any deterministic subset $\pi \subset [n]$ with $|\pi| \leq \delta_0 n$ and any deterministic $u \in \mathbb{C}^n$, we have*

$$\mathbb{P}\left(\min_{i \in \pi} \operatorname{dist}(C_i, H_i) \leq \varepsilon\sqrt{n}\right) \leq Cn \exp(-cn),$$

where $H_i := \operatorname{Span}\left(\{C_j : j \in \pi, j \neq i\} \cup \{u\}\right)$.

Proof. We will apply Lemma 5.4 to control $\operatorname{dist}(C_i, H_i)$. To this end, define $u^{(i)}$ to be the vector u with the i-th entry removed, and set $H_i^{(i)} := \operatorname{Span}\left(\{C_j^{(i)} : j \in \pi, j \neq i\} \cup \{u^{(i)}\}\right)$. Then

$$\operatorname{dist}(C_i, H_i) \geq \operatorname{dist}(C_i^{(i)}, H_i^{(i)})$$

for all $1 \leq i \leq n$. Note that $H_i^{(i)}$ is independent of $C_i^{(i)}$. Hence, conditioning on $H_i^{(i)}$ and applying Lemma 5.4, we find the existence of $\varepsilon, C, c, \delta_0$ such that for any $\pi \subset [n]$ with $|\pi| \leq \delta_0 n$ and all $i \in \pi$

$$\mathbb{P}\left(\operatorname{dist}(C_i, H_i) \leq \varepsilon\sqrt{n}\right) \leq \mathbb{P}\left(\operatorname{dist}(C_i^{(i)}, H_i^{(i)}) \leq \varepsilon\sqrt{n}\right) \leq C \exp(-cn).$$

Therefore, by the union bound,

$$\mathbb{P}\left(\min_{i \in \pi} \operatorname{dist}(C_i, H_i) \leq \varepsilon\sqrt{n}\right) \leq \sum_{i \in \pi} \mathbb{P}\left(\operatorname{dist}(C_i, H_i) \leq \varepsilon\sqrt{n}\right) \leq Cn \exp(-cn),$$

and the proof is complete. $\qquad \square$

Let ε, δ_0 be as in Corollary 5.5. From now on we set

$$\tau := \frac{1}{4} \min \left\{ 1, \frac{\varepsilon}{s\sqrt{\delta}} \right\}. \tag{5.5}$$

(Recall that $s \geq 1$ is the upper bound for $s_1(X + M)$ specified in the statement of Theorem 5.1.) In particular, this definition implies that $\tau \leq 1/4$. The parameter δ is still to be specified. Right now we only assume that $\delta < \delta_0$.

Lemma 5.6 (Control of compressible vectors). *There exist constants $C, c, \delta > 0$ such that for any deterministic vector $u \in \mathbb{C}^n$ and any $s \geq 1$*

$$\mathbb{P}\left(\inf_{x \in \mathrm{Comp}(\delta, \tau)} \|Ax - u\| \leq \frac{\varepsilon}{2\sqrt{\delta}}, s_1(A) \leq s \right) \leq C \exp(-cn).$$

Proof. Let $0 < \delta < \delta_0$ be a constant to be chosen later. We decompose $\mathrm{Comp}(\delta, \tau)$ as

$$\mathrm{Comp}(\delta, \tau) = \bigcup_{\pi \subset [n]: |\pi| = \lfloor \delta n \rfloor} S_\pi,$$

where

$$S_\pi := \{ x \in \mathrm{Comp}(\delta, \tau) : \mathrm{dist}(x, \mathrm{Sparse}_\pi(\delta)) \leq \tau \}$$

and

$$\mathrm{Sparse}_\pi(\delta) := \{ y \in \mathrm{Sparse}(\delta) : \mathrm{supp}(y) \subset \pi \}.$$

So by the union bound,

$$\mathbb{P}\left(\inf_{x \in \mathrm{Comp}(\delta, \tau)} \|Ax - u\| \leq \frac{\varepsilon}{2\sqrt{\delta}}, s_1(A) \leq s \right) \tag{5.6}$$

$$\leq \sum_{\pi \in [n]: |\pi| = \lfloor \delta n \rfloor} \mathbb{P}\left(\inf_{x \in S_\pi} \|Ax - u\| \leq \frac{\varepsilon}{2\sqrt{\delta}}, s_1(A) \leq s \right).$$

Fix $\pi \subset [n]$ with $|\pi| = \lfloor \delta n \rfloor$, and suppose there exists $x \in S_\pi$ such that $\|Ax - u\| \leq \frac{\varepsilon}{2\sqrt{\delta}}$ and $s_1(A) \leq s$. Then there exists $y \in \mathrm{Sparse}_\pi(\delta)$ with $\|x - y\| \leq \tau$. In particular, this implies that $\|y\| \geq 3/4$. In addition, we have

$$\frac{\varepsilon}{2\sqrt{\delta}} \geq \|Ax - u\| \geq \|Ay - u\| - \tau\|A\| \geq \|Ay - u\| - \frac{\varepsilon}{4\sqrt{\delta}}$$

by the assumption that $\|A\| \leq s$ and the definition of τ (5.5). Hence, we obtain

$$\|Ay - u\| \leq \frac{3\varepsilon}{4\sqrt{\delta}}. \tag{5.7}$$

We now bound $\|Ay - u\|$ from below. Indeed, we have

$$\|Ay - u\|^2 = \left\| \sum_{i \in \pi} C_i y_i - u \right\|^2 \geq \max_{i \in \pi} |y_i|^2 \mathrm{dist}^2(C_i, H_i), \tag{5.8}$$

where $H_i := \mathrm{Span}\left(\{ C_j : j \in \pi, j \neq i \} \cup \{u\} \right)$. In addition, we bound

$$\max_{i \in \pi} |y_i|^2 \mathrm{dist}^2(C_i, H_i) \geq \min_{i \in \pi} \mathrm{dist}^2(C_i, H_i) \frac{1}{|\pi|} \sum_{i \in \pi} |y_i|^2 \geq \min_{i \in \pi} \mathrm{dist}^2(C_i, H_i) \left(\frac{3}{4} \right)^2 \frac{1}{\delta n}.$$

Thus, combining (5.7) and (5.8) with the bound above, we find

$$\min_{i \in \pi} \mathrm{dist}(C_i, H_i) \leq \varepsilon \sqrt{n}.$$

To conclude, we have shown that

$$\mathbb{P}\left(\inf_{x \in S_\pi} \|Ax - u\| \le \frac{\varepsilon}{2\sqrt{\delta}}, s_1(A) \le s\right) \le \mathbb{P}\left(\min_{i \in \pi} \operatorname{dist}(C_i, H_i) \le \varepsilon\sqrt{n}\right).$$

In view of Corollary 5.5, there exist $C, c > 0$ such that for any $\pi \in [n]$ with $|\pi| = \lfloor \delta n \rfloor$, we have

$$\mathbb{P}\left(\inf_{x \in S_\pi} \|Ax - u\| \le \frac{\varepsilon}{2\sqrt{\delta}}, s_1(A) \le s\right) \le Cn \exp(-cn).$$

Returning to (5.6), we conclude that

$$\begin{aligned}
\mathbb{P}\left(\inf_{x \in \operatorname{Comp}(\delta, \tau)} \|Ax - u\| \le \frac{\varepsilon}{2\sqrt{\delta}}, s_1(A) \le s\right) &\le \binom{n}{\lfloor \delta n \rfloor} Cn \exp(-cn) \\
&\le \left(\frac{ne}{\lfloor \delta n \rfloor}\right)^{\lfloor \delta n \rfloor} Cn \exp(-cn) \\
&\le C' \exp(-c'n)
\end{aligned}$$

for some constants $C', c' > 0$ by taking δ sufficiently small (in terms of c). $\qquad\square$

We now fix δ to be the constant from Lemma 5.6. Thus, δ and τ are now completely determined. We will also need the following corollary of Lemma 5.6.

Corollary 5.7. *There exist constants $C, c > 0$ such that for any deterministic vector $u \in \mathbb{C}^n$ and any $s \ge 1$*

$$\mathbb{P}\left(\inf_{x/\|x\| \in \operatorname{Comp}(\delta, \tau)} \frac{\|Ax - u\|}{\|x\|} \le \frac{\varepsilon}{4\sqrt{\delta}}, s_1(A) \le s\right) \le Cs \exp(-cn). \tag{5.9}$$

Proof. The proof is based on the arguments given in [57]. We first note that if $u = 0$, then the claim follows immediately from Lemma 5.6. Assume $u \ne 0$. Let $\mathcal{E}$ denote the event on the left-hand side of (5.9) whose probability we would like to bound. Suppose that $\mathcal{E}$ holds. Then there exists $x_0 := x/\|x\| \in \operatorname{Comp}(\delta, \tau)$ and $u_0 := u/\|x\| \in \operatorname{Span}(u)$ such that $\|Ax_0 - u_0\| \le \frac{\varepsilon}{4\sqrt{\delta}}$ and $s_1(A) \le s$. In particular, this implies that

$$\|u_0\| \le \|Ax_0 - u_0\| + \|Ax_0\| \le \frac{\varepsilon}{4\sqrt{\delta}} + s.$$

Let $\mathcal{N}$ be a $\frac{\varepsilon}{4\sqrt{\delta}}$-net of the real interval $[-\frac{\varepsilon}{4\sqrt{\delta}} - s, \frac{\varepsilon}{4\sqrt{\delta}} + s]$. In particular, we can choose $\mathcal{N}$ so that

$$|\mathcal{N}| \le C's \tag{5.10}$$

for some constant $C' > 0$. Here, we have used the assumption that $s \ge 1$. In particular, there exists $c_0 \in \mathcal{N}$ such that

$$\left\| \frac{u}{\|x\|} - c_0 \frac{u}{\|u\|} \right\| \le \frac{\varepsilon}{4\sqrt{\delta}}.$$

By the triangle inequality, this implies that

$$\left\| Ax_0 - c_0 \frac{u}{\|u\|} \right\| \le \frac{\varepsilon}{2\sqrt{\delta}}.$$

To conclude, we have shown that

$$\mathbb{P}(\mathcal{E}) \le \mathbb{P}\left(\inf_{c_0 \in \mathcal{N}} \inf_{x_0 \in \operatorname{Comp}(\delta, \tau)} \left\| Ax_0 - c_0 \frac{u}{\|u\|} \right\| \le \frac{\varepsilon}{2\sqrt{\delta}}, s_1(A) \le s\right),$$

and thus, by the union bound, we have

$$\mathbb{P}(\mathcal{E}) \le \sum_{c_0 \in \mathcal{N}} \mathbb{P}\left(\inf_{x_0 \in \mathrm{Comp}(\delta,\tau)} \left\| Ax_0 - c_0 \frac{u}{\|u\|} \right\| \le \frac{\varepsilon}{2\sqrt{\delta}}, s_1(A) \le s \right).$$

The claim now follows by the cardinality bound for $\mathcal{N}$ in (5.10) and Lemma 5.6. □

When dealing with incompressible vectors we will need the following corollary.

Corollary 5.8. *There exists constants $C, c > 0$, such that, for any $u \in \mathbb{C}^n$ with $u \ne 0$, and any $s \ge 1$, we have*

$$\mathbb{P}\left(\frac{A^{-1}u}{\|A^{-1}u\|} \in \mathrm{Comp}(\delta,\tau), s_1(A) \le s \right) \le Cs\exp(-cn).$$

Proof. If $x := A^{-1}u$, then $(Ax - u)/\|x\| = 0$. Thus, we have

$$\mathbb{P}\left(\frac{A^{-1}u}{\|A^{-1}u\|} \in \mathrm{Comp}(\delta,\tau), s_1(A) \le s \right)$$
$$\le \mathbb{P}\left(\inf_{x/\|x\| \in \mathrm{Comp}(\delta,\tau)} \frac{\|Ax - u\|}{\|x\|} = 0, s_1(A) \le s \right).$$

The conclusion now follows from (5.9). □

5.4 Anti-concentration bounds

In order to handle incompressible vectors, we will need several anti-concentration bounds. The main idea is to use the rate of convergence from the Berry–Esseen Theorem to obtain the estimates. This idea appears to have originated in [36] and has been used previously in many works including [16, 36, 48].

Lemma 5.9 (Small ball probability via Berry–Esseen; Lemma A.6 from [16])**.** *There exists $C > 0$ such that if $Z_1, \ldots, Z_n$ are independent centered complex-valued random variables, then for all $t \ge 0$,*

$$\sup_{z \in \mathbb{C}} \mathbb{P}\left(\left| \sum_{i=1}^n Z_i - z \right| \le t \right) \le C\left(\frac{t}{\sqrt{\sum_{i=1}^n \mathbb{E}|Z_i|^2}} + \frac{\sum_{i=1}^n \mathbb{E}|Z_i|^3}{\left(\sum_{i=1}^n \mathbb{E}|Z_i|^2\right)^{3/2}} \right).$$

We begin the following anti-concentration bound for sums involving dependent random variables.

Lemma 5.10. *Let $\{(\xi_i, \psi_i) : 1 \le i \le n\}$ be a collection of independent complex-valued random tuples, and assume there exist $a, b, \sigma > 0$ such that the events*

$$\mathcal{E}_i := \{|\xi_i| \le a, |\psi_i| \le a\}$$

for $1 \le i \le n$ satisfy

$$b \le \min_{1 \le i \le n} \mathbb{P}(\mathcal{E}_i), \qquad \sigma^2 \le \min_{1 \le i \le n} \mathrm{Var}(\xi_i \mid \mathcal{E}_i), \qquad \sigma^2 \le \min_{1 \le i \le n} \mathrm{Var}(\psi_i \mid \mathcal{E}_i).$$

In addition, assume there exists $\rho < 1$ such that

$$\max_{1 \le i \le n} |\mathrm{Corr}(\xi_i \mid \mathcal{E}_i, \psi_i \mid \mathcal{E}_i)| \le \rho. \tag{5.11}$$

Then for any $\delta, \tau \in (0, 1)$, any $w = (w_i)_{i=1}^n \in \mathrm{Incomp}(\delta, \tau)$, any $w' = (w'_i)_{i=1}^n \in \mathbb{C}^n$ with $\|w'\| \le 1$, any $J \subset [n]$ with $|J| \ge n(1 - \delta/4)$, and any $t \ge 0$, we have

$$\sup_{z \in \mathbb{C}} \mathbb{P}\left(\left| \sum_{i \in J} (\xi_i w_i + \psi_i w'_i) - z \right| \le t\tau \right)$$
$$\le \frac{C}{\sigma\sqrt{\delta b(1-\rho)}} \sqrt{\left\lceil \frac{1}{2}\log_2\left(\frac{2}{\tau^2\delta}\right) \right\rceil \left\lceil \log_2\left(\frac{\sqrt{n}}{\tau}\right) \right\rceil} \left(t + \frac{a}{\sqrt{n}} \right) + \exp(-\delta nb/32),$$

where $C > 0$ is an absolute constant.

Proof. The proof is based on the arguments from [16]. Since $w \in \text{Incomp}(\delta, \tau)$ and $|J| \geq n(1 - \delta/4)$, Lemma 5.3 implies the existence of $\pi \subset J$ such that $|\pi| \geq \delta n/4$ and

$$\frac{\tau}{\sqrt{n}} \leq |w_i| \leq \sqrt{\frac{2}{\delta n}}$$

for all $i \in \pi$. By conditioning on the random variables ξ_i, ψ_i for $i \notin \pi$ and absorbing their contribution into the constant z, it suffices to bound

$$\sup_{z \in \mathbb{C}} \mathbb{P}\left(\left| \sum_{i \in \pi} (\xi_i w_i + \psi_i w_i') - z \right| \leq t\tau \right).$$

We now proceed to truncate the random variables ξ_i, ψ_i for $i \in \pi$. Indeed, by the Chernoff bound, it follows that

$$\sum_{i \in \pi} \mathbf{1}_{\mathcal{E}_i} \geq \frac{|\pi| b}{2} \geq \frac{\delta b n}{8}$$

with probability at least $1 - \exp(-\delta b n/32)$. Therefore, taking $m := \lceil \delta b n/8 \rceil$, it suffices to show

$$\sup_{z \in \mathbb{C}} \mathbb{P}_m \left(\left| \sum_{i=1}^{m} (\xi_i w_i + \psi_i w_i') - z \right| \leq t\tau \right)$$
$$\leq \frac{C}{\sigma \sqrt{\delta b (1 - \rho)}} \sqrt{\left\lceil \frac{1}{2} \log_2 \left(\frac{2}{\tau^2 \delta} \right) \right\rceil \left\lceil \log_2 \left(\frac{\sqrt{n}}{\tau} \right) \right\rceil} \left(t + \frac{a}{\sqrt{n}} \right),$$

where $\mathbb{P}_m(\cdot) := \mathbb{P}(\cdot | E_m, \mathcal{F}_m)$ is the conditional probability given $\mathcal{F}_m$, the σ-algebra generated by all random variables except $\xi_1, \ldots, \xi_m, \psi_1, \ldots, \psi_m$, and the event

$$E_m := \left\{ |\xi_i| \leq a, |\psi_i| \leq a : 1 \leq i \leq m \right\} \bigcap \left\{ \frac{\tau}{\sqrt{n}} \leq |w_i| \leq \sqrt{\frac{2}{\delta n}} : 1 \leq i \leq m \right\}.^{[3]}$$

By centering the random variables (and absorbing the expectations into the constant z), it suffices to bound

$$\sup_{z \in \mathbb{C}} \mathbb{P}_m \left(\left| \sum_{i=1}^{m} [(\xi_i - \mathbb{E}_m[\xi_i]) w_i + (\psi_i - \mathbb{E}_m[\psi_i]) w_i'] - z \right| \leq t\tau \right).$$

We again reduce to the case where we only need to consider a subset of the coordinates of w and w' which are roughly comparable. Indeed, as the random tuples $(\xi_1, \psi_1), \ldots, (\xi_m, \psi_m)$ are jointly independent under the probability measure $\mathbb{P}_m$, we can condition on any subset of them (and again absorb their contribution into the constant z); hence, for any subset $I \subset [m]$, we have

$$\sup_{z \in \mathbb{C}} \mathbb{P}_m \left(\left| \sum_{i=1}^{m} [(\xi_i - \mathbb{E}_m[\xi_i]) w_i + (\psi_i - \mathbb{E}_m[\psi_i]) w_i'] - z \right| \leq t\tau \right) \tag{5.12}$$
$$\leq \sup_{z \in \mathbb{C}} \mathbb{P}_m \left(\left| \sum_{i \in I} [(\xi_i - \mathbb{E}_m[\xi_i]) w_i + (\psi_i - \mathbb{E}_m[\psi_i]) w_i'] - z \right| \leq t\tau \right).$$

---

[3]When defining this event, we consider the w_i to be constant random variables.

We now choose the subset I in a sequence of two steps. First, define

$$L := 2 \left\lceil \frac{1}{2} \log_2 \left(\frac{2}{\delta \tau^2} \right) \right\rceil,$$

and for $1 \leq j \leq L$ set

$$I_j := \left\{ 1 \leq i \leq m : \frac{2^{j-1} \tau}{\sqrt{n}} \leq |w_i| < \frac{2^j \tau}{\sqrt{n}} \right\}.$$

By construction, $I_1, \ldots, I_L$ partition the index set $[m]$. Hence, by the pigeonhole principle, there exists j such that $|I_j| \geq m/L$. Second, we partition the set I_j as follows. Define

$$K := 2 \left\lceil \log_2 \left(\frac{\sqrt{n}}{\tau} \right) \right\rceil,$$

and for $1 \leq k \leq K$ set

$$I_{j,k} := \left\{ i \in I_j : \frac{2^{k-1} \tau}{\sqrt{n}} \leq |w_i'| < \frac{2^k \tau}{\sqrt{n}} \right\}$$

and define

$$I_{j,0} := \left\{ i \in I_j : |w_i'| < \frac{\tau}{\sqrt{n}} \right\}.$$

As $\|w'\| \leq 1$ by assumption, the sets $I_{j,0}, I_{j,1}, \ldots, I_{j,K}$ form a partition of I_j. By the pigeonhole principle, there exists k such that

$$|I_{j,k}| \geq \frac{m}{L(K+1)} \geq \frac{m}{2LK}. \tag{5.13}$$

Applying (5.12) to the set $I_{j,k}$, it now suffices to show

$$\sup_{z \in \mathbb{C}} \mathbb{P}_m \left(\left| \sum_{i \in I_{j,k}} [(\xi_i - \mathbb{E}_m[\xi_i]) w_i + (\psi_i - \mathbb{E}_m[\psi_i]) w_i'] - z \right| \leq t\tau \right) \tag{5.14}$$

$$\leq \frac{C\sqrt{LK}}{\sigma \sqrt{\delta b (1 - \rho)}} \left(t + \frac{a}{\sqrt{n}} \right)$$

for some absolute constant $C > 0$.

We will apply Lemma 5.9 to obtain (5.14). For $i \in I_{j,k}$, define

$$Z_i := (\xi_i - \mathbb{E}_m[\xi_i]) w_i + (\psi_i - \mathbb{E}_m[\psi_i]) w_i'.$$

Then

$$q^2 := \sum_{i \in I_{j,k}} \mathbb{E}_m |Z_i|^2$$

$$\geq \sum_{i \in I_{j,k}} \left[|w_i|^2 \operatorname{Var}_m(\xi_i) + |w_i'|^2 \operatorname{Var}_m(\psi_i) - 2\rho |w_i| |w_i'| \sqrt{\operatorname{Var}_m(\xi_i) \operatorname{Var}_m(\psi_i)} \right]$$

by assumption (5.11). Thus, we deduce that

$$q^2 \geq (1 - \rho) \sum_{i \in I_{j,k}} \left[|w_i|^2 \operatorname{Var}_m(\xi_i) + |w_i'|^2 \operatorname{Var}_m(\psi_i) \right]$$

$$\geq (1 - \rho) \sigma^2 \sum_{i \in I_{j,k}} \left[|w_i|^2 + |w_i'|^2 \right]. \tag{5.15}$$

In addition,

$$\sum_{i \in I_{j,k}} \mathbb{E}_m |Z_i|^3 \le 2a\tau \left(\frac{2^j + 2^k}{\sqrt{n}} \right) q^2.$$

Hence, Lemma 5.9 implies the existence of an absolute constant $C > 0$ such that

$$\sup_{z \in \mathbb{C}} \mathbb{P}_m \left(\left| \sum_{i \in I_{j,k}} [(\xi_i - \mathbb{E}_m[\xi_i]) w_i + (\psi_i - \mathbb{E}_m[\psi_i]) w_i'] - z \right| \le t\tau \right)$$
$$\le \frac{C\tau}{q} \left(t + \frac{a(2^j + 2^k)}{\sqrt{n}} \right). \tag{5.16}$$

We complete the proof by considering two separate cases. First, if $k = 0$, then using (5.13) and (5.15), we obtain

$$q^2 \ge \sigma^2 (1 - \rho) \sum_{i \in I_{j,k}} |w_i|^2 \ge \sigma^2 (1 - \rho) |I_{j,k}| \frac{2^{2j-2} \tau^2}{n} \ge \sigma^2 (1 - \rho) \delta b \frac{2^{2j} \tau^2}{64 L K}.$$

Hence returning to (5.16), we find that

$$\sup_{z \in \mathbb{C}} \mathbb{P}_m \left(\left| \sum_{i \in I_{j,k}} [(\xi_i - \mathbb{E}_m[\xi_i]) w_i + (\psi_i - \mathbb{E}_m[\psi_i]) w_i'] - z \right| \le t\tau \right)$$
$$\le \frac{C' \sqrt{LK}}{\sigma \sqrt{\delta b (1 - \rho)}} \left(t + \frac{a}{\sqrt{n}} \right), \tag{5.17}$$

where $C' > 0$ is an absolute constant. Similarly, if $1 \le k \le K$, then

$$q^2 \ge \sigma^2 (1 - \rho) |I_{j,k}| \frac{\tau^2}{n} \left(2^{2j-2} + 2^{2k-2} \right) \ge \sigma^2 (1 - \rho) \delta b \tau^2 \frac{2^{2j} + 2^{2k}}{64 L K}.$$

In this case, we again apply (5.16) to deduce the existence of an absolute constant $C'' > 0$ such that

$$\sup_{z \in \mathbb{C}} \mathbb{P}_m \left(\left| \sum_{i \in I_{j,k}} [(\xi_i - \mathbb{E}_m[\xi_i]) w_i + (\psi_i - \mathbb{E}_m[\psi_i]) w_i'] - z \right| \le t\tau \right)$$
$$\le \frac{C'' \sqrt{LK}}{\sigma \sqrt{\delta b (1 - \rho)}} \left(t + \frac{a}{\sqrt{n}} \right). \tag{5.18}$$

Combining (5.17) and (5.18), we obtain the bound (5.14) (with the absolute constant $C := \max\{C', C''\}$), and the proof of the lemma is complete. $\qquad \square$

Lastly, we will need the following technical anti-concentration bound which is similar to Lemma 5.10.

Lemma 5.11. *Let $\{(\xi_i, \psi_i) : 1 \le i \le n\} \cup \{(\xi_i', \psi_i') : 1 \le i \le n\}$ be a collection of independent complex-valued random tuples with the property that (ξ_i, ψ_i) has the same distribution as (ξ_i', ψ_i') for $1 \le i \le n$, and assume there exist $a, b, \sigma > 0$ such that the events*

$$\mathcal{E}_i := \{|\xi_i| \le a, |\psi_i| \le a\}$$

for $1 \le i \le n$ satisfy

$$b \le \min_{1 \le i \le n} \mathbb{P}(\mathcal{E}_i), \qquad \sigma^2 \le \min_{1 \le i \le n} \mathrm{Var}(\xi_i \mid \mathcal{E}_i), \qquad \sigma^2 \le \min_{1 \le i \le n} \mathrm{Var}(\psi_i \mid \mathcal{E}_i).$$

In addition, let $\delta_1, \ldots, \delta_n$ be i.i.d. $\{0,1\}$-valued random variables with $\mathbb{E}\delta_i = c_{oo} \in (0,1]$, and assume $\delta_1, \ldots, \delta_n$ are independent of the random tuples $\{(\xi_i, \psi_i), (\xi_i', \psi_i') : 1 \leq i \leq n\}$. Then for any $\delta, \tau \in (0,1)$, any $w = (w_i)_{i=1}^n \in \mathrm{Incomp}(\delta, \tau)$, any $w' = (w_i')_{i=1}^n \in \mathbb{C}^n$ with $\|w'\| \leq 1$, and any $t \geq 0$, we have

$$
\sup_{z \in \mathbb{C}} \mathbb{P}\left(\left| \sum_{i=1}^n (\delta_i \xi_i w_i + \delta_i \xi_i' w_i') - z \right| \leq t\tau \right)
$$

$$
\leq \frac{C}{\sigma \sqrt{c_{oo} \delta b^2}} \sqrt{\left\lceil \frac{1}{2} \log_2\left(\frac{2}{\tau^2 \delta} \right) \right\rceil \left\lceil \log_2\left(\frac{\sqrt{n}}{\tau} \right) \right\rceil} \left(t + \frac{a}{\sqrt{n}} \right) + \exp(-\delta n c_{oo} b^2/16),
$$

where $C > 0$ is an absolute constant.

Lemma 5.11 is very similar to Lemma 5.10. The statement of the lemma is somewhat unusual since the hypotheses involve the variables $\xi_i, \psi_i, \xi_i', \psi_i'$, but the conclusion only involves the random variables ξ_i, ξ_i'. This is to match the assumptions of Theorem 5.1. The proof of Lemma 5.11 presented below follows the same framework as the proof of Lemma 5.10.

Proof of Lemma 5.11. Since $w \in \mathrm{Incomp}(\delta, \tau)$, Lemma 5.3 implies the existence of $\pi \subset [n]$ such that $|\pi| \geq \delta n/2$ and

$$
\frac{\tau}{\sqrt{n}} \leq |w_i| \leq \sqrt{\frac{2}{\delta n}}
$$

for all $i \in \pi$. By conditioning on the random variables $\xi_i, \psi_i, \xi_i', \psi_i', \delta_i$ for $i \notin \pi$ and absorbing their contribution into the constant z, it suffices to bound

$$
\sup_{z \in \mathbb{C}} \mathbb{P}\left(\left| \sum_{i \in \pi} (\delta_i \xi_i w_i + \delta_i \xi_i' w_i') - z \right| \leq t\tau \right).
$$

We now proceed to truncate the random variables $\xi_i, \psi_i, \xi_i', \psi_i'$ for $i \in \pi$. Define the events

$$
\mathcal{E}_i' := \{|\xi_i'| \leq a, |\psi_i'| \leq a\},
$$

and observe that $\mathcal{E}_i'$ is independent of $\mathcal{E}_i$ for $1 \leq i \leq n$. By the Chernoff bound, it follows that

$$
\sum_{i \in \pi} \mathbf{1}_{\mathcal{E}_i \cap \mathcal{E}_i' \cap \{\delta_i = 1\}} \geq \frac{c_{oo} |\pi| b^2}{2} \geq \frac{c_{oo} \delta b^2 n}{4}
$$

with probability at least $1 - \exp(-c_{oo} \delta b^2 n/16)$. Therefore, taking $m := \lceil c_{oo} \delta b^2 n/4 \rceil$, it suffices to bound

$$
\sup_{z \in \mathbb{C}} \mathbb{P}_m\left(\left| \sum_{i=1}^m (\delta_i \xi_i w_i + \delta_i \xi_i' w_i') - z \right| \leq t\tau \right) = \sup_{z \in \mathbb{C}} \mathbb{P}_m \left(\left| \sum_{i=1}^m (\xi_i w_i + \xi_i' w_i') - z \right| \leq t\tau \right)
$$

$$
\tag{5.19}
$$

where $\mathbb{P}_m(\cdot) := \mathbb{P}(\cdot | E_m, \mathcal{F}_m)$ is the conditional probability given $\mathcal{F}_m$, the σ-algebra generated by all random variables except $\xi_1, \ldots, \xi_m$, $\psi_1, \ldots, \psi_m$, $\xi_1', \ldots, \xi_m'$, $\psi_1', \ldots, \psi_m'$, $\delta_1, \ldots, \delta_m$, and the event

$$
E_m := \bigcap_{i=1}^m \{|\xi_i| \leq a, |\psi_i| \leq a, |\xi_i'| \leq a, |\psi_i'| \leq a, \delta_i = 1\} \bigcap_{i=1}^m \left\{ \frac{\tau}{\sqrt{n}} \leq |w_i| \leq \sqrt{\frac{2}{\delta n}} \right\}.
$$

Here, we have exploited the fact that on the event E_m, $\delta_i = 1$ for $i \in [m]$, and so all factors of δ_i have been replaced by 1 on the right-hand side of (5.19). By centering the

random variables (and absorbing the expectations into the constant z), it suffices to bound

$$\sup_{z \in \mathbb{C}} \mathbb{P}_m \left(\left| \sum_{i=1}^m [(\xi_i - \mathbb{E}_m[\xi_i])w_i + (\xi_i' - \mathbb{E}_m[\xi_i'])w_i'] - z \right| \leq t\tau \right).$$

We again reduce to the case where we only need to consider a subset of the coordinates of w and w' which are roughly comparable. Indeed, as the random vectors $(\xi_1, \psi_1, \xi_1', \psi_1'), \ldots, (\xi_m, \psi_m, \xi_m', \psi_m')$ are jointly independent under the probability measure $\mathbb{P}_m$, we can condition on any subset of them (and again absorb their contribution into the constant z); hence, for any subset $I \subset [m]$, we have

$$\sup_{z \in \mathbb{C}} \mathbb{P}_m \left(\left| \sum_{i=1}^m [(\xi_i - \mathbb{E}_m[\xi_i])w_i + (\xi_i' - \mathbb{E}_m[\xi_i'])w_i'] - z \right| \leq t\tau \right) \tag{5.20}$$

$$\leq \sup_{z \in \mathbb{C}} \mathbb{P}_m \left(\left| \sum_{i \in I} [(\xi_i - \mathbb{E}_m[\xi_i])w_i + (\xi_i' - \mathbb{E}_m[\xi_i'])w_i'] - z \right| \leq t\tau \right).$$

We now choose the subset I in a sequence of two steps as was done in the proof of Lemma 5.10. First, define

$$L := 2 \left\lceil \frac{1}{2} \log_2 \left(\frac{2}{\delta \tau^2} \right) \right\rceil,$$

and for $1 \leq j \leq L$ set

$$I_j := \left\{ 1 \leq i \leq m : \frac{2^{j-1}\tau}{\sqrt{n}} \leq |w_i| < \frac{2^j \tau}{\sqrt{n}} \right\}.$$

By construction, $I_1, \ldots, I_L$ partition the index set $[m]$. Hence, by the pigeonhole principle, there exists j such that $|I_j| \geq m/L$. Second, we partition the set I_j as follows. Define

$$K := 2 \left\lceil \log_2 \left(\frac{\sqrt{n}}{\tau} \right) \right\rceil,$$

and for $1 \leq k \leq K$ set

$$I_{j,k} := \left\{ i \in I_j : \frac{2^{k-1}\tau}{\sqrt{n}} \leq |w_i'| < \frac{2^k \tau}{\sqrt{n}} \right\}$$

and define

$$I_{j,0} := \left\{ i \in I_j : |w_i'| < \frac{\tau}{\sqrt{n}} \right\}.$$

As $\|w'\| \leq 1$ by assumption, the sets $I_{j,0}, I_{j,1}, \ldots, I_{j,K}$ form a partition of I_j. By the pigeonhole principle, there exists k such that

$$|I_{j,k}| \geq \frac{m}{L(K+1)} \geq \frac{m}{2LK}. \tag{5.21}$$

Applying (5.20) to the set $I_{j,k}$, it now suffices to show

$$\sup_{z \in \mathbb{C}} \mathbb{P}_m \left(\left| \sum_{i \in I_{j,k}} [(\xi_i - \mathbb{E}_m[\xi_i])w_i + (\xi_i' - \mathbb{E}_m[\xi_i'])w_i'] - z \right| \leq t\tau \right) \tag{5.22}$$

$$\leq C \frac{\sqrt{LK}}{\sigma \sqrt{c_{oo} \delta b^2}} \left(t + \frac{a}{\sqrt{n}} \right)$$

for some absolute constant $C > 0$.

We will apply Lemma 5.9 to obtain (5.22). For $i \in I_{j,k}$, define

$$Z_i := (\xi_i - \mathbb{E}_m[\xi_i])w_i + (\xi_i' - \mathbb{E}_m[\xi_i'])w_i'.$$

Then

$$
\begin{aligned}
q^2 &:= \sum_{i \in I_{j,k}} \mathbb{E}_m |Z_i|^2 \\
&= \sum_{i \in I_{j,k}} \left[|w_i|^2 \operatorname{Var}_m(\xi_i) + |w_i'|^2 \operatorname{Var}_m(\xi_i') \right] \\
&\geq \sigma^2 \sum_{i \in I_{j,k}} \left[|w_i|^2 + |w_i'|^2 \right]
\end{aligned}
\tag{5.23}
$$

since ξ_i and ξ_i' are independent under the measure $\mathbb{P}_m$ for $1 \leq i \leq n$. In addition,

$$\sum_{i \in I_{j,k}} \mathbb{E}_m |Z_i|^3 \leq 2a\tau \left(\frac{2^j + 2^k}{\sqrt{n}} \right) q^2.$$

Hence, Lemma 5.9 implies the existence of an absolute constant $C > 0$ such that

$$
\sup_{z \in \mathbb{C}} \mathbb{P}_m \left(\left| \sum_{i \in I_{j,k}} [(\xi_i - \mathbb{E}_m[\xi_i])w_i + (\xi_i' - \mathbb{E}_m[\xi_i'])w_i'] - z \right| \leq t\tau \right)
$$
$$
\leq \frac{C\tau}{q} \left(t + \frac{a(2^j + 2^k)}{\sqrt{n}} \right).
\tag{5.24}
$$

We complete the proof by considering two separate cases. First, if $k = 0$, then using (5.21) and (5.23), we obtain

$$q^2 \geq \sigma^2 \sum_{i \in I_{j,k}} |w_i|^2 \geq \sigma^2 |I_{j,k}| \frac{2^{2j-2}\tau^2}{n} \geq \sigma^2 c_{oo} \delta b^2 \frac{2^{2j}\tau^2}{32LK}.$$

Hence returning to (5.24), we find that

$$
\sup_{z \in \mathbb{C}} \mathbb{P}_m \left(\left| \sum_{i \in I_{j,k}} [(\xi_i - \mathbb{E}_m[\xi_i])w_i + (\xi_i' - \mathbb{E}_m[\xi_i'])w_i'] - z \right| \leq t\tau \right)
$$
$$
\leq \frac{C'\sqrt{LK}}{\sigma\sqrt{c_{oo}\delta b^2}} \left(t + \frac{a}{\sqrt{n}} \right),
\tag{5.25}
$$

where $C' > 0$ is an absolute constant. Similarly, if $1 \leq k \leq K$, then

$$q^2 \geq \sigma^2 |I_{j,k}| \frac{\tau^2}{n} \left(2^{2j-2} + 2^{2k-2} \right) \geq \sigma^2 c_{oo} \delta b^2 \tau^2 \frac{2^{2j} + 2^{2k}}{32LK}.$$

In this case, we again apply (5.24) to deduce the existence of an absolute constant $C'' > 0$ such that

$$
\sup_{z \in \mathbb{C}} \mathbb{P}_m \left(\left| \sum_{i \in I_{j,k}} [(\xi_i - \mathbb{E}_m[\xi_i])w_i + (\xi_i' - \mathbb{E}_m[\xi_i]')w_i'] - z \right| \leq t\tau \right)
$$
$$
\leq \frac{C''\sqrt{LK}}{\sigma\sqrt{c_{oo}\delta b^2}} \left(t + \frac{a}{\sqrt{n}} \right).
\tag{5.26}
$$

Combining (5.25) and (5.26), we obtain the bound (5.22) (with the absolute constant $C := \max\{C', C''\}$), and the proof of the lemma is complete. $\qquad\square$

5.5 Incompressible vectors

In order to control the set of incompressible vectors, we will require the following averaging estimate.

Lemma 5.12 (Invertibility via average distance; Lemma A.4 from [16])**.** *Let A be a random matrix taking values in $\mathrm{Mat}_n(\mathbb{C})$ with columns $C_1, \dots, C_n$. For any $1 \leq k \leq n$, let $H_k := \mathrm{Span}\{C_i : i \neq k\}$. Then, for any $t \geq 0$,*

$$\mathbb{P}\left(\min_{x \in \mathrm{Incomp}(\delta, \tau)} \|Ax\| \leq \frac{t\tau}{\sqrt{n}} \right) \leq \frac{2}{\delta n} \sum_{k=1}^{n} \mathbb{P}(\mathrm{dist}(C_k, H_k) \leq t).$$

Let $A = X + M$ be the matrix from Theorem 5.1. Let $C_1, \dots, C_n$ be the columns of A and $H_k := \mathrm{Span}\{C_i : i \neq k\}$ be as in Lemma 5.12. Our main result for controlling the set of incompressible vectors is the following.

Lemma 5.13. *There exists $C, c > 0$ such that for every $1 \leq k \leq n$, any $t > 0$ and any $s \geq 1$,*

$$\mathbb{P}(\mathrm{dist}(C_k, H_k) \leq t, s_1(A) \leq s) \leq C \left(\frac{\log(Csn)}{\sqrt{1-\rho}} \left(s^2 \sqrt{t} + \frac{1}{\sqrt{n}} \right) \right)^{1/4}.$$

The rest of the subsection is devoted to the proof of Lemma 5.13. We complete the proof of Theorem 5.1 in Subsection 5.6. We will also need the following result based on [57, Proposition 5.1] and [30, Statement 2.8].

Lemma 5.14 (Distance problem via bilinear forms)**.** *Let $A = (A_{ij}) \in \mathrm{Mat}_n(\mathbb{C})$, let $C_1, \dots, C_n$ denote the columns of A, and fix $1 \leq k \leq n$. Let $H_k := \mathrm{Span}\{C_i : i \neq k\}$, u be the k-th row of A with the k-th entry removed, v be C_k with the k-th entry removed, and let B be the $(n-1) \times (n-1)$ submatrix of A formed from removing the k-th row and k-th column. If B is invertible, then*

$$\mathrm{dist}(C_k, H_k) \geq \frac{|A_{kk} - uB^{-1}v|}{\sqrt{1 + \|uB^{-1}\|^2}}. \tag{5.27}$$

Proof. The proof presented below is based on the arguments given in [30, 57]. By permuting the rows and columns, it suffices to assume that $k = 1$. Let $h \in \mathbb{S}^{n-1}$ denote any normal to the hyperplane H_1. Then

$$\mathrm{dist}(C_1, H_1) \geq |h^* C_1|.$$

We decompose

$$C_1 = \begin{pmatrix} A_{11} \\ v \end{pmatrix}, \quad h = \begin{pmatrix} h_1 \\ g \end{pmatrix},$$

where $h_1 \in \mathbb{C}$ and $g \in \mathbb{C}^{n-1}$. Then

$$\mathrm{dist}(C_1, H_1) \geq |h^* C_1| = |\bar{h}_1 A_{11} + g^* v|. \tag{5.28}$$

Since h is orthogonal to the columns of the matrix $\begin{pmatrix} u \\ B \end{pmatrix}$, we find

$$0 = h^* \begin{pmatrix} u \\ B \end{pmatrix} = \bar{h}_1 u + g^* B,$$

and hence

$$g^* = -\bar{h}_1 u B^{-1}.$$

Returning to (5.28), we have

$$\mathrm{dist}(C_1, H_1) \geq |h_1| |A_{11} - uB^{-1}v|. \tag{5.29}$$

In addition,

$$1 = \|h\|^2 = |h_1|^2 + \|g\|^2 = |h_1|^2(1 + \|uB^{-1}\|^2),$$

and so

$$|h_1|^2 = \frac{1}{1 + \|uB^{-1}\|^2}. \tag{5.30}$$

The conclusion now follows from (5.29) and (5.30). $\qquad\square$

Our study of the bilinear form $uB^{-1}v$ is based on the following general result, which will allow us to introduce some additional independence into the problem to deal with the fact that u and v are dependent. Similar decoupling techniques have also appeared in [20, 21, 28, 30, 57]. The lemma below is based on [57, Lemma 8.4] for quadratic forms.

Lemma 5.15 (Decoupling lemma). *Let* $M \in \mathrm{Mat}_n(\mathbb{C})$, *and let* $x = (x_i)_{i=1}^n, y = (y_i)_{i=1}^n$ *be random vectors in* $\mathbb{C}^n$ *such that* $\{(x_i, y_i) : 1 \le i \le n\}$ *is a collection of independent random tuples. Let* (x', y') *denote an independent copy of* (x, y). *Then for every subset* $\pi \subset [n]$ *and* $t \ge 0$, *we have*

$$\sup_{z \in \mathbb{C}} \mathbb{P}(|x^{\mathrm{T}} M y - z| \le t)^2$$
$$\le \mathbb{P}_{x,y,x',y'}\left(|x_\pi^{\mathrm{T}} M_{\pi \times \pi^c}(y_{\pi^c} - y'_{\pi^c}) + (x_{\pi^c} - x'_{\pi^c})^{\mathrm{T}} M_{\pi^c \times \pi} y_\pi + z_0| \le 2t\right),$$

where z_0 *is a random variable whose value is determined by* $M_{\pi^c \times \pi^c}, x_{\pi^c}, y_{\pi^c}, x'_{\pi^c}, y'_{\pi^c}$.

The proof of Lemma 5.15 is based on the following decoupling bound from [51, 57].

Lemma 5.16 (Lemma 8.5 from [57]). *Let* ξ *and* ψ *be independent random vectors, and let* ψ' *be an independent copy of* ψ. *Let* $\mathcal{E}(\xi, \psi)$ *be an event which is determined by the values of* ξ *and* ψ. *Then*

$$\mathbb{P}(\mathcal{E}(\xi, \psi))^2 \le \mathbb{P}(\mathcal{E}(\xi, \psi) \cap \mathcal{E}(\xi, \psi')).$$

Proof of Lemma 5.15. Let ξ be the random vector formed by the tuples $\{(x_i, y_i) : i \in \pi\}$, and let ψ be the random vector formed from the tuples $\{(x_i, y_i) : i \notin \pi\}$. Then ξ and ψ are independent by supposition, and we can apply Lemma 5.16. To this end, let $\tilde{x}, \tilde{y}$ be random vectors in $\mathbb{C}^n$ defined by

$$\tilde{x}_\pi := x_\pi, \quad \tilde{x}_{\pi^c} := x'_{\pi^c}, \quad \tilde{y}_\pi := y_\pi, \quad \tilde{y}_{\pi^c} := y'_{\pi^c}.$$

An application of Lemma 5.16 yields

$$\mathbb{P}(|x^{\mathrm{T}} M y - z| \le t)^2 \le \mathbb{P}_{x,y,\tilde{x},\tilde{y}}(|x^{\mathrm{T}} M y - z| \le t, |\tilde{x}^{\mathrm{T}} M \tilde{y} - z| \le t)$$
$$\le \mathbb{P}_{x,y,\tilde{x},\tilde{y}}(|x^{\mathrm{T}} M y - \tilde{x}^{\mathrm{T}} M \tilde{y}| \le 2t),$$

where the last inequality follows from the triangle inequality. We now note that

$$x^{\mathrm{T}} M y - \tilde{x}^{\mathrm{T}} M \tilde{y} = x_\pi^{\mathrm{T}} M_{\pi \times \pi^c}(y_{\pi^c} - \tilde{y}_{\pi^c}) + (x_{\pi^c} - \tilde{x}_{\pi^c})^{\mathrm{T}} M_{\pi^c \times \pi} y_\pi + z_0,$$

where z_0 depends only on $M_{\pi^c \times \pi^c}, x_{\pi^c}, y_{\pi^c}, \tilde{x}_{\pi^c}, \tilde{y}_{\pi^c}$. Since $\tilde{x}_{\pi^c} = x'_{\pi^c}$ and $\tilde{y}_{\pi^c} = y'_{\pi^c}$, the claim follows. $\qquad\square$

We now turn to the proof of Lemma 5.13. The arguments presented here follow the general framework of [57, Section 8.3]. Fix $1 \le k \le n$; the arguments and bounds below will all be uniform in k. Let u be the k-th row of A with the k-th entry removed. Let v be the k-th column of A with the k-th entry removed, and let B be the $(n-1) \times (n-1)$

matrix formed by removing the k-th row and k-th column from A. In view of Lemma 5.14, it suffices to prove that

$$\mathbb{P}\left(\frac{|A_{kk} - uB^{-1}v|}{\sqrt{1 + \|uB^{-1}\|^2}} \leq t, s_1(A) \leq s\right) \leq C\left(\frac{\log(Csn)}{\sqrt{1-\rho}}\left(s^2\sqrt{t} + \frac{1}{\sqrt{n}}\right)\right)^{1/4} \tag{5.31}$$

for some constants $C, c > 0$. Our argument is based on applying Lemma 5.15 to decouple the bilinear form $uB^{-1}v$ and then applying our anti-concentration bounds from Subsection 5.4 to bound the resulting expressions. We divide the proof of (5.31) into a number of sub-steps.

5.5.1 Step 1: constructing a random subset π

Following [57], we decompose $[n-1]$ into two random subsets π and π^c. Let $\delta_1, \ldots, \delta_{n-1}$ be i.i.d. $\{0,1\}$-valued random variables, independent of X, with $\mathbb{E}\delta_i = c_{oo}/2$, where c_{oo} is a constant defined by

$$c_{oo} := \delta/8$$

and $\delta \in (0,1)$ was previously fixed. We then define $\pi := \{i \in [n-1] : \delta_i = 0\}$. By the Chernoff bound, it follows that

$$|\pi^c| \leq c_{oo}n \tag{5.32}$$

with probability at least $1 - C'_{oo}\exp(-c'_{oo}n)$ for some constants $C'_{oo}, c'_{oo} > 0$.

5.5.2 Step 2: estimating $\|B^{-1}\|_2$

Lemma 5.17 below will allow us to estimate the denominator appearing on the right-hand side of (5.27). To this end, let (u', v') be an independent copy of (u, v), also independent of X.

Lemma 5.17. *There exist constants $C, c > 0$ such that, for any $s \geq 1$, the random matrix B has the following properties with probability at least $1 - Cs\exp(-cn)$. If $s_1(B) \leq s$, one has:*

(i) for any $t_0 \geq 0$, with probability at least $1 - C\log(Cns)\left(st_0 + n^{-1/2}\right)$ in u, u', π,

$$\|(u-u')_{\pi^c}B^{-1}_{\pi^c \times [n-1]}\| \geq t_0\|B^{-1}\|_2.$$

(ii) for any $t_0 \geq 0$, with probability at least $1 - C\log(Cns)\left(st_0 + n^{-1/2}\right)$ in v, v', π,

$$\|B^{-1}_{[n-1] \times \pi^c}(v-v')_{\pi^c}\| \geq t_0\|B^{-1}\|_2.$$

In order to prove the lemma, we will need the following elementary result.

Lemma 5.18 (Sums of dependent random variables; Lemma 8.3 from [57]). *Let $Z_1, \ldots, Z_n$ be arbitrary non-negative random variables (not necessarily independent), and $p_1, \ldots, p_n$ be non-negative numbers such that*

$$\sum_{j=1}^{n} p_j = 1.$$

Then, for every $t \geq 0$, we have

$$\mathbb{P}\left(\sum_{j=1}^{n} p_j Z_j \leq t\right) \leq 2\sum_{j=1}^{n} p_j\mathbb{P}(Z_j \leq 2t).$$

Proof of Lemma 5.17. By Corollary 5.8 and the union bound, under the assumption $s_1(B) \le s$, we have

$$x_j := \frac{B^{-1}e_j}{\|B^{-1}e_j\|} \in \mathrm{Incomp}(\delta, \tau), \qquad y_j := \frac{e_j^* B^{-1}}{\|e_j^* B^{-1}\|} \in \mathrm{Incomp}(\delta, \tau)$$

for $1 \le j \le n-1$ with probability at least $1 - Cs \exp(-cn)$. Here, $e_1, \dots, e_{n-1}$ are the standard basis elements of $\mathbb{C}^{n-1}$. Fix a realization of B for which this property holds. We will prove that both properties hold with the desired probability for this fixed realization of B.

For (i), we note that

$$\|(u-u')_{\pi^c} B^{-1}_{\pi^c \times [n-1]}\|^2 = \sum_{j=1}^{n-1} |(u-u')_{\pi^c} B^{-1}_{\pi^c \times [n-1]} e_j|^2$$
$$= \sum_{j=1}^{n-1} |(u-u')_{\pi^c} (x_j)_{\pi^c}|^2 \|B^{-1}e_j\|^2.$$

Taking $p_j := \|B^{-1}e_j\|^2 / \|B^{-1}\|_2^2$, we see that $\sum_{j=1}^{n-1} p_j = 1$, and hence

$$\mathbb{P}_{u,u',\pi} \left(\|(u-u')_{\pi^c} B^{-1}_{\pi^c \times [n-1]}\| \le t_0 \|B^{-1}\|_2 \right)$$
$$\le \mathbb{P}_{u,u',\pi} \left(\sum_{j=1}^{n-1} |(u-u')_{\pi^c} (x_j)_{\pi^c}|^2 p_j \le t_0^2 \right)$$
$$\le 2 \sum_{j=1}^{n-1} p_j \mathbb{P}_{u,u',\pi}(|(u-u')_{\pi^c} (x_j)_{\pi^c}|^2 \le 2t_0^2)$$
$$\le 2 \sup_{w \in \mathrm{Incomp}(\delta,\tau)} \mathbb{P}_{u,u',\pi}(|(u-u')_{\pi^c} w_{\pi^c}| \le \sqrt{2}t_0)$$

by Lemma 5.18. Recalling our choice of δ, τ (5.5), and c_{oo}, the claim now follows from the anti-concentration bound given in Lemma 5.11.

The proof of (ii) is similar. Indeed, we have

$$\|B^{-1}_{[n-1] \times \pi^c} (v-v')_{\pi^c}\|^2 = \sum_{j=1}^{n-1} |(y_j)_{\pi^c} (v-v')_{\pi^c}|^2 \|e_j^* B^{-1}\|^2.$$

Applying Lemma 5.18 with $p_j := \|e_j^* B^{-1}\|^2 / \|B^{-1}\|_2^2$, we conclude that

$$\mathbb{P}_{v,v',\pi} \left(\|B^{-1}_{[n-1] \times \pi^c} (v-v')_{\pi^c}\| \le t_0 \|B^{-1}\|_2 \right)$$
$$\le 2 \sum_{j=1}^{n-1} p_j \mathbb{P}_{v,v',\pi} \left(|(y_j)_{\pi^c} (v-v')_{\pi^c}| \le \sqrt{2}t_0 \right)$$
$$\le 2 \sup_{w \in \mathrm{Incomp}(\delta,\tau)} \mathbb{P}_{v,v',\pi} \left(|w_{\pi^c}^{\mathrm{T}} (v-v')_{\pi^c}| \le \sqrt{2}t_0 \right).$$

As before, the conclusion now follows from the anti-concentration bound given in Lemma 5.11. $\qquad \square$

5.5.3 Step 3: working on the appropriate events

We have one last preparatory step before we can apply the decoupling lemma, Lemma 5.15. In this step, we define the events we will need to work on for the remainder of the

proof. To this end, define the events

$$\mathcal{B}_A := \{s_1(A) \leq s\}, \qquad \mathcal{B}_B := \{s_1(B) \leq s\}, \qquad \mathcal{B}_u := \{\|u\| \leq s\}.$$

We note that $\mathcal{B}_A \subset \mathcal{B}_B$ and $\mathcal{B}_A \subset \mathcal{B}_u$ since u and B are sub-matrices of A. Consider the random vectors

$$w := (u - u')_{\pi^c} B^{-1}_{\pi^c \times [n-1]}, \qquad w' := B^{-1}_{[n-1] \times \pi^c} (v - v')_{\pi^c}. \tag{5.33}$$

It is possible that $w = 0$ or $w' = 0$, although we will show that these events happen with small probability momentarily.

Let $t_0 > 0$. Consider the event

$$\|B^{-1}\|_2 \leq \frac{1}{t_0} \min\{\|w\|, \|w'\|\}. \tag{5.34}$$

By Lemma 5.17, we find

$$\mathbb{P}_{B,u,u',v,v',\pi}((5.34) \text{ holds } \vee \mathcal{B}_B^c) \geq 1 - C\log(csn)(st_0 + n^{-1/2}) - Cs\exp(-cn)$$

for some constants $C, c > 0$. In particular, since $\|B^{-1}\|_2 > 0$, it follows that when the event in (5.34) occurs, it must be the case that w and w' are both nonzero. In order to avoid several different cases later in the proof, let us define ω and ω' as follows. If w is nonzero, we take $\omega := w$, and if w is zero, we define ω to be a fixed vector in $\mathrm{Incomp}(\delta, \tau)$. We define ω' analogously in terms of w'. It follows that on the event (5.34), we have

$$\omega = w, \qquad \omega' = w'. \tag{5.35}$$

Next, consider the event

$$\frac{\omega}{\|\omega\|} \in \mathrm{Incomp}(\delta, \tau), \qquad \frac{\omega'}{\|\omega'\|} \in \mathrm{Incomp}(\delta, \tau). \tag{5.36}$$

Let us fix an arbitrary realization of u, v, u', v' and a realization of π which satisfies (5.32). We will apply Corollary 5.8 to control the event in (5.36). Indeed, we only need to consider the cases when $w \neq 0$ or $w' \neq 0$. In these cases, it follows that $\omega = w$ or $\omega' = w'$. Let us suppose this is the case. Then $\omega = (u - u')P_{\pi^c}B^{-1}$ or $\omega' = B^{-1}P_{\pi^c}(v - v')$, where P_{π^c} is an orthogonal projection onto those coordinates specified by π^c. Thus, from Corollary 5.8, we deduce that

$$\mathbb{P}_B((5.36) \text{ holds } \vee \mathcal{B}_B^c \mid u, v, u', v', \pi \text{ satisfies } (5.32)) \geq 1 - C's\exp(-c'n)$$

for some constants $C', c' > 0$. Combining the probabilities above, we conclude that

$$\mathbb{P}_{B,u,v,u',v',\pi}\left(((5.32), (5.34), (5.36) \text{ hold}) \vee \mathcal{B}_B^c\right)$$
$$\geq 1 - C''\log(C''sn)(st_0 + n^{-1/2}) - C''s\exp(-c''n)$$
$$=: 1 - p_0$$

for some constants $C'', c'' > 0$.

It follows that there exists a realization of π that satisfies (5.32) and such that

$$\mathbb{P}_{B,u,v,u',v'}\left(((5.34), (5.36) \text{ hold}) \vee \mathcal{B}_B^c\right) \geq 1 - p_0.$$

We fix such a realization of π for the remainder of the proof. Using Fubini's theorem, we deduce that the random matrix B has the following property with probability at least $1 - \sqrt{p_0}$:

$$\mathbb{P}_{u,v,u',v'}\left(((5.34), (5.36) \text{ hold}) \vee \mathcal{B}_B^c \mid B\right) \geq 1 - \sqrt{p_0}.$$

Since the event $\mathcal{B}_B$ depends only on B and not on u, v, u', v', it follows that the random matrix B has the following property with probability at least $1 - \sqrt{p_0}$: either $\mathcal{B}_B^c$ holds, or

$$\mathcal{B}_B \text{ holds and } \mathbb{P}_{u,v,u',v'}((5.34), (5.36) \text{ hold} \mid B) \geq 1 - \sqrt{p_0}. \tag{5.37}$$

5.5.4 Step 4: decoupling

Recall that we are interested in bounding $\mathbb{P}_{B,u,v,A_{kk}}(\mathcal{E} \wedge \mathcal{B}_A)$, where

$$\mathcal{E} := \left\{ \frac{|A_{kk} - uB^{-1}v|}{\sqrt{1 + \|uB^{-1}\|^2}} \leq t \right\}.$$

We first observe that

$$\mathbb{P}_{B,u,v,A_{kk}}(\mathcal{E} \wedge \mathcal{B}_A) \leq \mathbb{P}_{B,u,v,A_{kk}}(\mathcal{E} \wedge \mathcal{B}_B \wedge \mathcal{B}_u).$$

On the event $\mathcal{B}_u$, we have

$$\|uB^{-1}\| \leq \|u\|\|B^{-1}\| \leq s\|B^{-1}\|_2.$$

In addition, if $s_1(B) \leq s$, then $\|B^{-1}\|_2 \geq 1/s$. Hence, on the event $\mathcal{B}_B \wedge \mathcal{B}_u$,

$$1 + \|uB^{-1}\|^2 \leq 1 + s^2\|B^{-1}\|_2^2 \leq 2s^2\|B^{-1}\|_2^2.$$

Thus, we obtain

$$\mathbb{P}_{B,u,v,A_{kk}}(\mathcal{E} \wedge \mathcal{B}_A) \leq \mathbb{P}_{B,u,v,A_{kk}}(\mathcal{E}' \wedge \mathcal{B}_B),$$

where

$$\mathcal{E}' := \left\{ |A_{kk} - uB^{-1}v| \leq \sqrt{2}ts\|B^{-1}\|_2 \right\}.$$

Thus, we find

$$\mathbb{P}_{B,u,v,A_{kk}}(\mathcal{E} \wedge \mathcal{B}_A) \leq \mathbb{P}_{B,u,v,A_{kk}}(\mathcal{E}' \wedge (5.37) \text{ holds}) + \mathbb{P}_{B,u,v,A_{kk}}(\mathcal{B}_B \wedge (5.37) \text{ fails}).$$

The last probability is bounded above by $\sqrt{p_0}$ by the previous step. We conclude that

$$\mathbb{P}_{B,u,v,A_{kk}}(\mathcal{E} \wedge \mathcal{B}_A) \leq \sup_{\substack{B \text{ satisfies } (5.37) \\ A_{kk} \in \mathbb{C}}} \mathbb{P}_{u,v}(\mathcal{E}' \mid B, A_{kk}) + \sqrt{p_0}.$$

We now begin to work with the random vectors u', v' (recall that (u', v') are independent of X). To do so, we will work on a larger probability space which also includes the random vectors u', v'. Indeed, computing the probability above on the larger space which includes u', v', we conclude that

$$\mathbb{P}_{B,u,v,A_{kk}}(\mathcal{E} \wedge \mathcal{B}_A) \leq \sup_{\substack{B \text{ satisfies } (5.37) \\ A_{kk} \in \mathbb{C}}} \mathbb{P}_{u,v,u',v'}(\mathcal{E}' \mid B, A_{kk}) + \sqrt{p_0}.$$

For the remainder of the proof, we fix a realization of B which satisfies (5.37) and fix an arbitrary realization of A_{kk}. By supposition, both B and A_{kk} are independent of u, v, u', v'. It remains to bound the probability

$$p_1 := \sup_{z \in \mathbb{C}} \mathbb{P}_{u,v,u',v'}(\mathcal{E}'_z),$$

where

$$\mathcal{E}'_z := \left\{ |z - uB^{-1}v| \leq \sqrt{2}ts\|B^{-1}\|_2 \right\}.$$

To bound p_1, we apply the decoupling lemma, Lemma 5.15. Indeed, by Lemma 5.15,

$$p_1^2 \leq \mathbb{P}_{u,v,u',v'}(\mathcal{E}''),$$

where

$$\mathcal{E}'' := \left\{ |u_\pi B_{\pi \times \pi^c}^{-1}(v - v')_{\pi^c} + (u - u')_{\pi^c} B_{\pi^c \times \pi}^{-1} v_\pi + z_0| \leq 2\sqrt{2}st\|B^{-1}\|_2 \right\}$$

and z_0 is a complex number depending only on $B_{\{\pi^c \times \pi^c\}}$, $u_{\pi^c}, v_{\pi^c}, u'_{\pi^c}, v'_{\pi^c}$. Using (5.37) (where the conditioning on B is no longer required since B is now fixed), we find

$$p_1^2 \leq \mathbb{P}_{u,v,u',v'}(\mathcal{E}'' \wedge (5.34), (5.36) \text{ hold}) + \sqrt{p_0},$$

and hence

$$p_1^2 \leq \mathbb{P}_{u,v,u',v'}(\mathcal{E}''' \wedge (5.35), (5.36) \text{ hold}) + \sqrt{p_0},$$

where

$$\mathcal{E}''' := \left\{ |u_\pi w'_\pi + w_\pi v_\pi + z_0| \leq 2\sqrt{2}\frac{st}{t_0} \max\{\|w\|, \|w'\|\} \right\};$$

here, we used the fact that on the event (5.34), the event (5.35) holds.

5.5.5 Step 5: applying the anti-concentration bounds

Recall that w, w' depend only on $u_{\pi^c}, v_{\pi^c}, u'_{\pi^c}, v'_{\pi^c}$. In addition, u_π and v_π are independent of these random vectors. Let us fix a realization of the random vectors $u_{\pi^c}, v_{\pi^c}, u'_{\pi^c}, v'_{\pi^c}$ which satisfy (5.35) and (5.36). This completely determines w and w'; moreover, z_0 is also completely determined. Therefore, we conclude that

$$p_1^2 \leq \sup_{\substack{w,w' \text{ satisfy } (5.35),(5.36) \\ z_0 \in \mathbb{C}}} \mathbb{P}_{u_\pi, v_\pi}\left(|u_\pi w'_\pi + w_\pi v_\pi + z_0| \leq 2\sqrt{2}\frac{st}{t_0} \max\{\|w\|, \|w'\|\}\right) + \sqrt{p_0}.$$

In order to bound this first term on the right-hand side, we will apply the anti-concentration bound given in Lemma 5.10. Without loss of generality, let us assume that $\max\{\|w\|, \|w'\|\} = \|w\|$. Then dividing through by $\|w\|$, we find that

$$p_1^2 \leq \sup_{\substack{w,w' \text{ satisfy } (5.35),(5.36) \\ z_0 \in \mathbb{C}}} \mathbb{P}_{u_\pi, v_\pi}\left(\left|u_\pi \frac{w'_\pi}{\|w\|} + \frac{w_\pi}{\|w\|} v_\pi + z_0\right| \leq 2\sqrt{2}\frac{st}{t_0}\right) + \sqrt{p_0},$$

and

$$\frac{\|w'\|}{\|w\|} \leq 1.$$

In view of Lemma 5.10 (where we recall that $|\pi| \geq n(1 - \delta/8)$ due to (5.32)), we conclude that

$$p_1^2 \leq C''' \frac{1}{\sqrt{1-\rho}} \log(C'''ns)\left(\frac{s^2 t}{t_0} + \frac{1}{\sqrt{n}}\right) + \sqrt{p_0}$$

for some constant $C''' > 0$.

5.5.6 Step 6: completing the proof

Combining the bounds from the previous steps, we obtain

$$\mathbb{P}_{B,u,v,A_{kk}}(\mathcal{E} \wedge \mathcal{B}_A) \leq p_1 + \sqrt{p_0}.$$

We now proceed to simplify the expression to obtain (5.31). We still have the freedom to chose $t_0 > 0$; let us take $t_0 := \sqrt{t}$. In addition, we may assume that the expression

$$C''' \log(C'''ns)\left(s^2\sqrt{t} + \frac{1}{\sqrt{n}}\right) \tag{5.38}$$

is less than one as the bound is trivial otherwise. In particular, this implies that $s \leq \exp(\sqrt{n})$. Among others, this means that the error term $C''s \exp(-c''n)$ can be absorbed

into terms of the form (5.38) (by increasing the constant C''' if necessary). After some simplification, the bound for p_1 obtained in the previous step (with the substitution $t_0 := \sqrt{t}$) yields

$$\mathbb{P}_{B,u,v,A_{kk}}(\mathcal{E} \wedge \mathcal{B}_A) \leq C \left(\frac{\log(Csn)}{\sqrt{1-\rho}} \left(s^2\sqrt{t} + \frac{1}{\sqrt{n}} \right) \right)^{1/4}$$

for some constant $C > 0$. This completes the proof of (5.31), and hence the proof of Lemma 5.13 is complete.

5.6 Proof of Theorem 5.1

In this subsection, we complete the proof of Theorem 5.1. Indeed, for any $s \geq 1$ and any $0 < t \leq 1$, we have

$$\mathbb{P}\left(s_n(A) \leq \frac{t}{\sqrt{n}}, s_1(A) \leq s \right) \leq \mathbb{P}\left(\min_{x \in \mathbb{S}^{n-1}} \|Ax\| \leq \frac{t}{\sqrt{n}}, s_1(A) \leq s \right)$$

$$\leq \mathbb{P}\left(\min_{x \in \text{Incomp}(\delta,\tau)} \|Ax\| \leq \frac{t}{\sqrt{n}}, s_1(A) \leq s \right) \quad (5.39)$$

$$+ \mathbb{P}\left(\min_{x \in \text{Comp}(\delta,\tau)} \|Ax\| \leq \frac{1}{\sqrt{n}}, s_1(A) \leq s \right)$$

due to our decomposition of the unit sphere into compressible and incompressible vectors. It remains to bound each of the terms on the right-hand side.

For the incompressible vectors, we combine Lemmas 5.12 and 5.13 to find that, for any $t > 0$,

$$\mathbb{P}\left(\min_{x \in \text{Incomp}(\delta,\tau)} \|Ax\| \leq \frac{t\tau}{\sqrt{n}}, s_1(A) \leq s \right) \leq \frac{2}{\delta n} \sum_{k=1}^{n} \mathbb{P}(\text{dist}(C_k, H_k) \leq t, s_1(A) \leq s)$$

$$\leq \frac{2C}{\delta} \left(\frac{\log(Csn)}{\sqrt{1-\rho}} \left(s^2\sqrt{t} + \frac{1}{\sqrt{n}} \right) \right)^{1/4}$$

for some constant $C > 0$. Recalling the definitions of δ and τ (5.5), we conclude that

$$\mathbb{P}\left(\min_{x \in \text{Incomp}(\delta,\tau)} \|Ax\| \leq \frac{t}{\sqrt{n}}, s_1(A) \leq s \right) \leq C' \left(\frac{\log(C'sn)}{\sqrt{1-\rho}} \left(\sqrt{s^5 t} + \frac{1}{\sqrt{n}} \right) \right)^{1/4} \quad (5.40)$$

for some constant $C' > 0$. For compressible vectors, Lemma 5.6 implies the existence of constants $C'', c'' > 0$ such that

$$\mathbb{P}\left(\min_{x \in \text{Comp}(\delta,\tau)} \|Ax\| \leq \frac{1}{\sqrt{n}}, s_1(A) \leq s \right) \leq C'' \exp(-c''n). \quad (5.41)$$

Combining (5.40) and (5.41) with (5.39), we conclude that, for any $s \geq 1$ and any $0 < t \leq 1$,

$$\mathbb{P}\left(s_n(A) \leq \frac{t}{\sqrt{n}}, s_1(A) \leq s \right) \leq C' \left(\frac{\log(C'sn)}{\sqrt{1-\rho}} \left(\sqrt{s^5 t} + \frac{1}{\sqrt{n}} \right) \right)^{1/4} + C'' \exp(-c''n)$$

$$\leq C''' \left(\frac{\log(C'''sn)}{\sqrt{1-\rho}} \left(\sqrt{s^5 t} + \frac{1}{\sqrt{n}} \right) \right)^{1/4}$$

for some constant $C''' > 0$, where the second inequality follows from the fact that the first error term dominates the second for all n sufficiently large. The proof of Theorem 5.1 is complete.

6 Singular values of A_n and uniform integrability

6.1 Tightness

We begin with a bound on the largest singular values of $A_n - z$.

Lemma 6.1. *If Condition **C1** holds, there exists $r > 0, C > 0$ such that the following hold.*

- *For all $z \in \mathbb{C}$, there exists $C_z > 0$ such that almost surely*

$$\limsup_{n \to \infty} \int_0^\infty t^r d\nu_{A_n - z}(t) < C_z \text{ and thus } (\nu_{A_n - z})_{n \geq 1} \text{ is tight.}$$

- *Almost surely*

$$\limsup_{n \to \infty} \int_{\mathbb{C}} |w|^r d\mu_{A_n}(w) < C \text{ and thus } (\mu_{A_n})_{n \geq 1} \text{ is tight..}$$

Proof. We follow the approach of Lemma 3.1 in [13]. The tightness follows from the moment bound and Markov's inequality. The moment bound on μ_{A_n} follows from the bound on ν_{A_n} and Weyl's inequality, Lemma A.7. One has from Lemma A.3, $s_k(A_n - z) \leq s_k(A_n) + |z|$ for every $1 \leq k \leq n$ and thus using the fact for any $x, y \geq 0, (x + y)^r \leq 2^r(x^r + y^r)$ we can assume $z = 0$. We aim to work with matrices with independent entries, and thus decompose $A_n = U_n + L_n$ where L_n is strictly lower triangular, and U_n is upper triangular. Note for all $0 \leq k \leq n - 1$, we have by Lemma A.3

$$s_{1+k}(A_n) \leq s_{1+\lfloor k/2 \rfloor}(U_n) + s_{1+\lceil k/2 \rceil}(L_n).$$

We now restrict r such that $0 < r \leq 2$. Thus

$$\int_0^\infty t^r d\nu_{A_n}(t) \leq 8 \left[\int_0^\infty t^r d\nu_{U_n}(t) + \int_0^\infty t^r d\nu_{L_n}(t) \right].$$

We show only

$$\limsup_{n \to \infty} \int_0^\infty t^r d\nu_{U_n}(t) < \infty \qquad \text{a.s.}$$

as the proof that

$$\limsup_{n \to \infty} \int_0^\infty t^r d\nu_{L_n}(t) < \infty \qquad \text{a.s.}$$

follows in the exact same way.

By the Schatten bound, Lemma A.8,

$$\int_0^\infty t^r d\nu_{U_n}(t) \leq Z_n := \frac{1}{n} \sum_{i=1}^n Y_{n,i}$$

where

$$Y_{n,i} = \left(\sum_{j=i}^n a_n^{-2} |X_{ij}|^2 \right)^{r/2}.$$

For every $1 \leq i < j \leq n$ we let X'_{ji} be a copy of ξ_1, independent of all X_{ij} and all other X'_{ji}. In addition let

$$Y'_{n,i} := \left(\sum_{j=i}^n a_n^{-2} |X_{ij}|^2 + \sum_{j=1}^{i-1} a_n^{-2} |X'_{ij}|^2 \right)^{r/2}$$

then

$$Z_n \leq \frac{1}{n} \sum_{i=1}^n Y'_{n,i},$$

since $Y_{n,i} \leq Y'_{n,i}$. The proof then follows exactly as in Lemma 3.1 of [13]. $\qquad \square$

6.2 Distance from a row and a vector space

Throughout the rest of this section we assume the atom variables of X_n satisfy Condition **C2**. The proof of Proposition 3.3 from [13] can be adapted in a straight forward way to get Proposition 6.2 below. We give a brief explanation of the changes to the proof needed for entries that are independent but not necessarily identically distributed.

Proposition 6.2. *Let $0 < \gamma < 1/2$, and R be the i-th row of $a_n(A_n - z)$ with the i-th entry set to zero. There exists $\delta > 0$ depending on α, γ such that for all d-dimensional subspaces W of $\mathbb{C}^n$ with $n - d \geq n^{1-\gamma}$, one has*

$$\mathbb{P}\left(\operatorname{dist}(R, W) \leq n^{(1-2\gamma)/\alpha}\right) \leq e^{-n^\delta}.$$

Proof. Assume R is the i-th row of $a_n(A_n - z)$ with the i-th entry set to zero. If $X^{(i)}$ is the i-th row of X_n with the i-th entry set to zero, we have

$$\operatorname{dist}(R, W) \geq \operatorname{dist}(X^{(i)}, W_1)$$

where $W_1 = \operatorname{Span}(W, e_i)$. Note the entries of $X^{(i)}$ are independent, but have two potentially different distributions, in contrast with [13] where the entries are independent and identically distributed. However, Lemma A.9 can be applied to either distribution. Under Condition **C2** the slowly varying function, $L(t)$, in (1.3) is bounded and $L(t) \to c > 0$ as $t \to \infty$ for both entries. To adapt the proof of Proposition 3.3 in [13] apply Lemma A.9 to the entries of $X^{(i)}$ to get uniform bounds on the truncated moments without assuming the entries are identically distributed. $\square$

We now give some results for stable random variables which will be helpful. For $0 < \beta < 1$, let $Z = Z^{(\beta)}$ denote the one-sided positive β-stable distribution such that for all $s \geq 0$,

$$\mathbb{E}\exp(-sZ) = \exp(-s^\beta). \tag{6.1}$$

Recall for $y, m > 0$,

$$y^{-m} = \Gamma(m)^{-1} \int_0^\infty x^{m-1} e^{-xy} dx.$$

Thus for all $m > 0$,

$$\mathbb{E}[Z^{-m}] = \Gamma(m)^{-1} \int_0^\infty x^{m-1} e^{-x^\beta} dx, \tag{6.2}$$

and if $Z_1, \ldots, Z_n$ are i.i.d. copies of Z and $w_1, w_2, \ldots, w_n$ are non-negative real numbers then

$$\sum_{i=1}^n w_i Z_i \stackrel{d}{=} \left(\sum_{i=1}^n w_i^\beta\right)^{1/\beta} Z_1. \tag{6.3}$$

Lemma 6.3 (Lemma 3.5 in [13]). *Let $0 < \alpha < 2$ and Y be a random variable such that $t^\alpha \mathbb{P}(|Y| \geq t) \to c$ as $t \to \infty$ for some $c > 0$. Then there exists $\varepsilon > 0$ and $p \in (0,1)$ such that the random variable $|Y|^2$ dominates stochastically the random variable εDZ, where $\mathbb{P}(D = 1) = 1 - \mathbb{P}(D = 0) = p$ is a Bernoulli random variable, $Z = Z^{(\alpha/2)}$ and D and Z are independent.*

Lemma 6.4. *Let $X^{(i)}$ be the i-th row of the matrix X_n, with the i-th entry set to zero. Let $w_j \in [0, 1]$ be numbers such that $w(n) := \sum_{j=1}^n w_j \geq n^{1/2+\varepsilon}$ for some $\varepsilon > 0$. Let $Z = Z^{(\beta)}$ with $\beta = \alpha/2$. Then there exists $\delta > 0$ and a coupling of $X^{(i)}$ and Z such that*

$$\mathbb{P}\left(\sum_{j=1}^n w_j |X_{ij}|^2 \leq \delta w(n)^{1/\beta} Z\right) \leq Ce^{-cn^\delta}$$

for constants $C, c > 0$.

Proof. Let $D = (D_j)_{j=1}^i$ and $D' = (D_j)_{j=i+1}^n$ be two independent vectors of i.i.d. Bernoulli random variables given by Lemma 6.3 for $Y = X_{21}$ with parameter p and $Y = X_{12}$ with parameter p' respectively. We know from Lemma 6.3 there exists $\varepsilon' > 0$, such that for independent random variables Z_j satisfying (6.1) such that for every j, $w_j|X_{ij}|^2$ stochastically dominates $\varepsilon' w_j D_j Z_j$. Then there exists a coupling (see Lemma 2.12 in [33]) such that

$$\mathbb{P}\left(\sum_{j=1}^n w_j|X_{ij}|^2 \geq \varepsilon' \sum_{j=1}^n w_j D_j Z_j\right) = 1.$$

Let $\mathbf{a} = (a_1, \ldots, a_n) \in \{0,1\}^n$, and $A_{\mathbf{a}}$ be the event $D_j = a_j$ for all j. Then define the random variable Z pointwise on $A_{\mathbf{a}}, \mathbf{a} \neq \mathbf{0}$

$$Z(\omega) := \frac{\sum_{j=i}^n w_j a_j Z_j(\omega)}{\left(\sum_{j=1}^n w_j^\beta a_j\right)^{1/\beta}},$$

for $\omega \in A_{\mathbf{a}}$ and $Z(\omega) = Z_1(\omega)$ on $A_{\mathbf{0}}$. From (6.3), we see Z satisfies (6.1) and the distribution of Z does not depend on $D_1, \ldots, D_n$. Thus it is sufficient to show there exists $\varepsilon'' > 0$ such that

$$\mathbb{P}\left(\sum_{j=1}^n w_j^\beta D_j \leq \varepsilon'' w(n)\right) \leq Ce^{-cn^{\varepsilon''}}.$$

Note $w_j^\beta \geq w_j$, and thus $\mathbb{E}\sum_{j=1}^n w_j^\beta D_j \geq \min(p, p')w(n)$. Therefore for $0 < \varepsilon'' < \min(p, p')$,

$$\mathbb{P}\left(\sum_{j=1}^n w_j^\beta D_j \leq \varepsilon'' w(n)\right)$$
$$\leq \mathbb{P}\left(|\sum_{j=1}^n (w_j^\beta D_j - \mathbb{E}w_j^\beta D_j)| \geq (\min(p,p') - \varepsilon'')w(n)\right)$$
$$\leq 2e^{-\frac{1}{2}(\min(p,p') - \varepsilon'')^2 w(n)^2/n},$$

where the last bound follows from Hoeffding's inequality. $\square$

We now give another bound on the distance between a row of A_n and a deterministic subspace.

Proposition 6.5. *Take $0 < \gamma < \alpha/4$. Let R be the i-th row of $a_n(A_n - z)$ with the i-th entry set to zero. There exists an event E such that for any d-dimensional subspace W of $\mathbb{C}^n$ with $n - d \geq n^{1-\gamma}$, we have for sufficiently large n*

$$\mathbb{E}[\text{dist}^{-2}(R, W); E] \leq c(n-d)^{-2/\alpha} \quad \text{and} \quad \mathbb{P}(E^c) \leq cn^{-\frac{1}{2} + \gamma(\frac{2}{\alpha} - \frac{1}{2})},$$

where $c > 0$ is an absolute constant which does not depend on the choice of row.

Proof. We follow the approach of the proof of Proposition 3.7 in [13]. The only difference with Proposition 3.7 in [13] is the entries here are independent but not necessarily identically distributed. Assume that R is the i-th row of $a_n(A_n - z)$ with the i-th entry set to zero. Note

$$\text{dist}(R, W) \geq \text{dist}(X^{(i)}, W_1), \tag{6.4}$$

where $W_1 = \text{Span}(W, e_i)$, e_i is the i-th basis vector, and $X^{(i)} = (X_{ij})_{1 \leq j \leq n}$. Though the i-th entry of $X^{(i)}$ is not necessarily zero, inequality (6.4) still holds since the subspace spanned by e_i is contained in W_1. Let $\mathcal{J}$ denote the set of indices j such that $|X_{ij}| \leq a_n$.

It is a straight forward application of the Chernoff bound to show there exists $\delta > 0$ such that
$$\mathbb{P}(|\mathcal{J}| < n - \sqrt{n}) \le e^{-n^\delta}.$$
We begin by showing for any set $J \subset \{1, \dots, n\}$ such that $|J| \ge n - \sqrt{n}$,
$$\mathbb{E}[\operatorname{dist}^{-2}(R, W); E_J | \mathcal{J} = J] \le c(n - d)^{-2/\alpha},$$
for some event E_J satisfying $\mathbb{P}(E_J^c | \mathcal{J} = J) \le cn^{-\frac{1}{2} + \gamma(\frac{2}{\alpha} - \frac{1}{2})}$. Without loss of generality assume $J := \{1, \dots, n'\}$ with $n' \ge n - \sqrt{n}$. Let π_J be the orthogonal projection onto the first n' canonical basis vectors. Let $W_2 = \pi_J(W_1)$, set
$$W' = \operatorname{Span}\left(W_2, \mathbb{E}(\pi_J(X^{(i)}) | \mathcal{J} = J)\right).$$
Note $d - \sqrt{n} \le \dim(W') \le \dim(W_1) + 1 \le d + 2$. Define
$$Y = \pi_J(X^{(i)}) - \mathbb{E}(\pi_J(X^{(i)}) | \mathcal{J} = J).$$
One has
$$\operatorname{dist}(R, W) \ge \operatorname{dist}(Y, W').$$
Y is a mean zero vector under $\mathbb{P}(\cdot | \mathcal{J} = J)$. Note $W' \subseteq \pi_J(\mathbb{C}^n)$ and $Y \in \pi_J(\mathbb{C}^n)$, so we will work with both as objects in only $\pi_J(\mathbb{C}^n) \simeq \mathbb{C}^{n'}$ and not the larger vector space $\mathbb{C}^n$. Let P be the orthogonal projection matrix onto $(W')^\perp$ in $\mathbb{C}^{n'}$. Note $\operatorname{tr} P = n' - \dim(W')$ satisfies for sufficiently large n
$$2(n - d) \ge \operatorname{tr} P \ge \frac{1}{2}(n - d). \tag{6.5}$$

By construction
$$\mathbb{E}(\operatorname{dist}^2(Y, W') | \mathcal{J} = J) = \mathbb{E}\left[\sum_{j,k=1}^{n'} Y_j P_{jk} \bar{Y}_k \Big| \mathcal{J} = J\right] = \sum_{j=1}^{n'} P_{jj} \mathbb{E}[|Y_j|^2 | \mathcal{J} = J]. \tag{6.6}$$

Let $S := \sum_{j=1}^{n'} P_{jj} |Y_j|^2$. Before beginning note by Lemma A.9 there exists $C > 0$ such that $\mathbb{E}[|Y_j|^2 | \mathcal{J} = J] \le C a_n^2 / n$ for $1 \le j \le n'$. Thus[4]

$$\mathbb{E}(|\operatorname{dist}^2(Y, W') - S|^2 | \mathcal{J} = J) = \mathbb{E}\left(\left|\sum_{k \ne j} Y_j P_{jk} \bar{Y}_k\right|^2 \Big| \mathcal{J} = J\right)$$
$$\le \frac{2C a_n^4}{n^2} \sum_{j,k} |P_{jk}|^2$$
$$= \frac{2C a_n^4}{n^2} \|P\|_2^2$$
$$= \frac{2C a_n^4}{n^2} \operatorname{tr}(P^* P)$$
$$= \frac{2C a_n^4}{n^2} \operatorname{tr}(P).$$

Thus
$$\mathbb{E}(|\operatorname{dist}^2(Y, W') - S|^2 | \mathcal{J} = J) = O\left(a_n^4 \frac{n - d}{n^2}\right). \tag{6.7}$$

---

[4]We believe there is a small typo in the bound of $\mathbb{E}(|\operatorname{dist}^2(Y, W') - S|^2 | \mathcal{J} = J)$ in the proof of Proposition 3.7 in [13]

Let Z be as in Lemma 6.4. Set $w_j = P_{jj}$ and for $\varepsilon > 0$, consider the event

$$\Gamma_J := \left\{ \sum_{j=1}^{n'} w_j |X_{ij}|^2 \geq \varepsilon (n-d)^{1/\beta} Z \right\},$$

where $\beta = \alpha/2$. From Lemma 6.4 there exists a coupling of $X_{i1}, \ldots, X_{i,n'}$ and Z such that

$$\mathbb{P}(\Gamma_J^c) \leq C e^{-cn^{\delta}}, \tag{6.8}$$

for some $\delta > 0$ and some choice of $\varepsilon > 0$. Since $(a-b)^2 \geq a^2/2 - b^2$ for $a, b \in \mathbb{R}$ we have $S \geq \frac{1}{2} S_a - S_b$ where

$$S_a := \sum_{j=1}^{n'} w_j |X_{ij}|^2,$$

and

$$S_b := \sum_{j=1}^{n'} w_j \mathbb{E}[|X_{ij}| \big| X_{ij} \leq a_n]^2.$$

From Lemma A.9 and (6.5) one has

$$S_b \leq h^{(\alpha)}(n, d)$$

where $h^{(\alpha)}(n, d) = \Theta((n-d)a_n^2/n^2)$ if $\alpha \in (0, 1]$ and $h^{(\alpha)}(n, d) = \Theta((n-d))$ if $\alpha \in (1, 2)$. Let G_J^1 be the event that $S_a \geq 3 S_b$. There exists some c_0 such that

$$\mathbb{P}((G_J^1)^c \cap \Gamma_J | \mathcal{J} = J) \leq \mathbb{P}(Z \leq c_0 (n-d)^{-1/\beta} h^{(\alpha)}(n, d) | \mathcal{J} = J).$$

From the assumption $n - d \geq n^{1-\gamma}$ with $0 < \gamma < \alpha/4$ we have $(n-d)^{-1/\beta} h^{(\alpha)}(n, d) \leq C n^{-\varepsilon_0}$ for some $C, \varepsilon_0 > 0$. From here, using the bound (6.2) on the negative second moment of Z, it is straightforward to show that for every $p > 0$ there exists a constant κ_p such that

$$\mathbb{P}((G_J^1)^c \cap \Gamma_J | \mathcal{J} = J) \leq \kappa_p n^{-p}. \tag{6.9}$$

Set $\tilde{\Gamma}_J = G_J^1 \cap \Gamma_J$. On $\tilde{\Gamma}_J$, $S \geq \frac{1}{6} S_a \geq \frac{\varepsilon}{6}(n-d)^{2/\alpha} Z$, and therefore

$$\mathbb{E}[S^{-2}; \tilde{\Gamma}_J | \mathcal{J} = J] \leq c_1 (n-d)^{-4/\alpha} \mathbb{E}[Z^{-2}].$$

and thus using again the negative second moment bound on Z,

$$\mathbb{E}[S^{-2}; \tilde{\Gamma}_J | \mathcal{J} = J] = O\left((n-d)^{-4/\alpha} \right). \tag{6.10}$$

Let G_J^2 be the event $\{\mathrm{dist}^2(Y, W') \geq S/2\}$. Note using Markov's inequality and the Cauchy-Schwarz inequality leads to

$$\mathbb{P}\left(\mathrm{dist}^2(Y, W') \leq S/2; \tilde{\Gamma}_J | \mathcal{J} = J \right) \leq \mathbb{P}\left(\frac{|\mathrm{dist}^2(Y, W') - S|}{S} \geq 1/2; \tilde{\Gamma}_J | \mathcal{J} = J \right)$$

$$\leq 2\mathbb{E}\left[\frac{|\mathrm{dist}^2(Y, W') - S|}{S}; \tilde{\Gamma}_J | \mathcal{J} = J \right] \tag{6.11}$$

$$\leq 2\sqrt{\mathbb{E}\left[|\mathrm{dist}^2(Y, W') - S|^2 | \mathcal{J} = J \right] \mathbb{E}[S^{-2}; \tilde{\Gamma}_J | \mathcal{J} = J]}.$$

Then by (6.7), (6.10), and (6.11)

$$\mathbb{P}\left((G_J^2)^c \cap \tilde{\Gamma}_J | \mathcal{J} = J \right) = O\left(a_n^2 n^{-1} (n-d)^{\frac{1}{2} - \frac{2}{\alpha}} \right). \tag{6.12}$$

By the Cauchy-Schwarz inequality

$$\mathbb{E}\left[\operatorname{dist}^{-2}(X,W); G_J^2 \cap \tilde{\Gamma}_J | \mathcal{J}=J\right] \leq 2\mathbb{E}\left[S^{-1}; \tilde{\Gamma}_J | \mathcal{J}=J\right] = O\left((n-d)^{-2/\alpha}\right).$$

To conclude take $E_J = G_J^2 \cap \tilde{\Gamma}_J$. Then

$$\mathbb{P}(E_J^c | \mathcal{J}=J) \leq \mathbb{P}(\Gamma_J^c | \mathcal{J}=J) + \mathbb{P}((G_J^1)^c \cap \Gamma_J | \mathcal{J}=J) + \mathbb{P}\left((G_J^2)^c \cap \tilde{\Gamma}_J | \mathcal{J}=J\right).$$

It is then straightforward to show using (6.8), (6.9), and (6.12)

$$\mathbb{P}(E_J^c | \mathcal{J}=J) = O\left(n^{-\frac{1}{2}+\gamma(\frac{2}{\alpha}-\frac{1}{2})}\right).$$

Take $E = \bigcup_{J \in \mathbf{J}} E_J \cap \{\mathcal{J}=J\}$ where $\mathbf{J} = \{J \subseteq [n] : |J| \geq n - \sqrt{n}\}$ to complete the proof. $\qquad\square$

6.3 Application of Theorem 5.1

We now show that Theorem 5.1 can be applied to A_n to bound $s_n(A_n - z)$ from below with high probability. The difficulty is connecting the hypothesis on the spectral measure in Condition **C2** to the correlation of truncated random variables in the statement of Theorem 5.1.

Theorem 6.6. *For all $z \in \mathbb{C}$, there exists $C, r > 0$ such that*

$$\mathbb{P}\left(s_n(A_n - z) \leq Cn^{-r}\right) = o(1).$$

It is worth noting $o(1)$ can be improved to $n^{-\varepsilon}$ for some sufficiently small $\varepsilon > 0$. The proof of Theorem 6.6 will be an application of Theorem 5.1. First we need a bound on the operator norm, and hence the largest singular value, of $A_n - z$.

Lemma 6.7. *For every $z \in \mathbb{C}$, there exists $C > 0$ depending on the distribution of the entries and z such that for any $k > 1/\alpha$ and n sufficiently large,*

$$\mathbb{P}(\|X_n - a_n z\| \geq n^k) \leq \frac{C}{n^{((k-1)\alpha/2)-2}}.$$

Proof. Since

$$\|X_n - a_n z\| \leq \|X_n\| + a_n|z| \leq \|X_n\| + C_z n^{1/\alpha},$$

it is sufficient to bound $\mathbb{P}\left(\|X_n\| \geq cn^k\right)$, for some $0 < c < 1$ and n sufficiently large. Note

$$\|X_n\|^2 \leq n^2 \max_{1 \leq i,j \leq n} \{|X_{ij}|^2\}.$$

Thus, by the union bound,

$$\begin{aligned}
\mathbb{P}\left(\|X_n\| \geq cn^k\right) = \mathbb{P}\left(\|X_n\|^2 \geq c^2 n^{2k}\right) \\
\leq \mathbb{P}\left(\max_{1 \leq i,j \leq n}\{|X_{ij}|^2\} \geq c^2 n^{2(k-1)}\right) \\
\leq \mathbb{P}\left(\bigcup_{1 \leq i,j \leq n}\{|X_{ij}|^2 \geq c^2 n^{2(k-1)}\}\right) \\
\leq n^2 \max_{i,j=1,2}\left\{\mathbb{P}\left(|X_{ij}|^2 \geq c^2 n^{2(k-1)}\right)\right\} \\
\leq n^2 \frac{C'\mathbb{E}|X_{12}|^{\alpha/2}}{n^{2(k-1)\alpha/4}},
\end{aligned}$$

where the last inequality follows from Markov's inequality and (1.5). $\qquad\square$

Proof of Theorem 6.6. Clearly X_n satisfies conditions (i) and (ii) of Theorem 5.1, so it only remains to check condition (iii) before we can apply Theorem 5.1. Let $\mathcal{E}_{ij}^n := \{|X_{ij}| < \log(n)a_n, |X_{ji}| < \log(n)a_n\}$. Since $\{(X_{ij}, X_{ji}) : 1 \leq i < j \leq n\}$ is a collection i.i.d. random tuples we focus on showing for some n, $\mathcal{E}_{12}^n$ satisfies the desired conditions. From the tail bounds on X_{12}, and X_{21},

$$\mathbb{P}(\mathcal{E}_{12}^n) \geq 1 - \frac{C}{\log(n)^\alpha n}.$$

$\mathrm{Var}(X_{12}|\mathcal{E}_{12}^n) = 0$ if and only if X_{12} is constant on $\mathcal{E}_{12}^n$, and hence on any subset of $\mathcal{E}_{12}^n$. Thus if $\mathrm{Var}(X_{12}|\mathcal{E}_{12}^{n_0}) > 0$ for some n_0, then $\mathrm{Var}(X_{12}|\mathcal{E}_{12}^n) > 0$ for all $n \geq n_0$. Since X_{12} is non-constant, there must be some n sufficiently large such that X_{12} is non-constant on $\mathcal{E}_{12}^n$. Thus $\mathrm{Var}(X_{12}|\mathcal{E}_{12}^n) > 0$ for all n sufficiently large. The same argument follows for $\mathrm{Var}(X_{21}|\mathcal{E}_{12}^n)$.

Now assume for all n sufficiently large,

$$|\mathrm{Corr}\left(X_{12}|\mathcal{E}_{12}^n, X_{21}|\mathcal{E}_{12}^n\right)| = 1.$$

Then there exists $\theta_n \in [0, 2\pi)$ such that on $\mathcal{E}_{12}^n$,

$$X_{12} - \mathbb{E}[X_{12}|\mathcal{E}_{12}^n] = e^{i\theta_n} \sqrt{\frac{\mathrm{Var}(X_{12}|\mathcal{E}_{12}^n)}{\mathrm{Var}(X_{21}|\mathcal{E}_{12}^n)}} \left[X_{21} - \mathbb{E}[X_{21}|\mathcal{E}_{12}^n]\right].$$

For $\alpha < 1$, by Lemma A.9

$$\frac{\mathbb{E}[|X_{12}|\mathbf{1}_{\{|X_{12}|\leq t\}}]}{\left(\frac{\alpha}{1-\alpha}L(t)t^{1-\alpha}\right)} \to 1,$$

as $t \to \infty$, for some slowly varying function L. We assumed the atom variables satisfy Condition **C2** (ii), specifically $a_n \sim cn^{1/\alpha}$ and thus $L(t) \to c'$ for some constant $c' > 0$ as $t \to \infty$. For $\alpha \geq 1$, $\mathbb{E}[|X_{12}|\mathbf{1}_{\{|X_{12}|\leq t\}}]$ is dominated by t^r for any $r > 0$. Thus on $\mathcal{E}_{12}^n$

$$X_{12} = R_n X_{21} + C_n$$

where R_n is a sequence of complex numbers and

$$\lim_{n\to\infty} \frac{C_n}{a_n} = 0, \tag{6.13}$$

since $\sqrt{\frac{\mathrm{Var}(X_{12}|\mathcal{E}_{12}^n)}{\mathrm{Var}(X_{21}|\mathcal{E}_{12}^n)}}$ is slowly varying by Lemma A.9. If R_n has a bounded subsequence R_{n_k}, we shall take the corresponding sequences a_{n_k}, and n_k and for simplicity denote all three by R_n, a_n, and n. If not, then we note on $\mathcal{E}_{12}^n$

$$X_{21} = (R_n)^{-1} X_{12} + C_n'$$

where $(R_n)^{-1}$ is bounded, and (6.13) still holds for C_n'. Thus we will assume R_n is bounded and we take a subsequence which converges to R. Let $r > 0$ and B be a Borel subset of the unit sphere in $\mathbb{C}^2$ such that $\theta_d(\partial B) = 0$. Note

$$\theta_d(B)m_\alpha([r, \infty) = \lim_{n\to\infty} n\mathbb{P}\left(\frac{(X_{12}, X_{21})}{\|(X_{12}, X_{21})\|} \in B, \|(X_{12}, X_{21})\| \geq ra_n\right) =$$

$$\lim_{n\to\infty} n\mathbb{P}\left(\frac{(X_{12}, X_{21})}{\|(X_{12}, X_{21})\|} \in B, \|(X_{12}, X_{21})\| \geq ra_n, \mathcal{E}_{12}^n\right) \tag{6.14}$$

$$+ \lim_{n\to\infty} n\mathbb{P}\left(\frac{(X_{12}, X_{21})}{\|(X_{12}, X_{21})\|} \in B, \|(X_{12}, X_{21})\| \geq ra_n, (\mathcal{E}_{12}^n)^c\right),$$

Heavy-tailed elliptic random matrices

and

$$\lim_{n\to\infty} n\mathbb{P}\left(\frac{(X_{12},X_{21})}{\|(X_{12},X_{21})\|} \in B, \|(X_{12},X_{21})\| \geq ra_n, (\mathcal{E}_{12}^n)^c\right)$$
$$\leq \lim_{n\to\infty} n\mathbb{P}((\mathcal{E}_{12}^n)^c)$$
$$= \lim_{n\to\infty} n\mathbb{P}(|X_{12}| \geq \log(n)a_n, \text{or}|X_{21}| \geq \log(n)a_n)$$
$$\leq \lim_{n\to\infty} Cn(\log(n)a_n)^{-\alpha}$$
$$= \lim_{n\to\infty} C'n(\log(n)n^{1/\alpha})^{-\alpha}$$
$$= 0. \tag{6.15}$$

Thus we will consider only

$$\lim_{n\to\infty} n\mathbb{P}\left(\frac{(X_{12},X_{21})}{\|(X_{12},X_{21})\|} \in B, \|(X_{12},X_{21})\| \geq ra_n, \mathcal{E}_{12}^n\right). \tag{6.16}$$

Define the set A as

$$A = \{(z,w) \in \mathbb{C}^2 : z = Rw, |z|^2 + |w|^2 = 1\},$$

where R is the limit of R_n. Let $(z_1, z_2) \in \mathbb{C}^2$ be such that $(z_1, z_2) \notin A$ and $|z_1|^2 + |z_2|^2 = 1$. Let O be a small open neighborhood of (z_1, z_2) in $\mathbb{C}^2$ such that $A \cap \bar{O} = \emptyset$. We will consider the limit

$$\lim_{n\to\infty} n\mathbb{P}\left(\frac{(X_{12},X_{21})}{\|(X_{12},X_{21})\|} \in O, \|(X_{12},X_{21})\| \geq ra_n, \mathcal{E}_{12}^n\right). \tag{6.17}$$

Before we deal with this limit note that on $\mathcal{E}_{12}^n$

$$\frac{(X_{12},X_{21})}{\|(X_{12},X_{21})\|} = \frac{(RX_{21},X_{21})}{\|(X_{12},X_{21})\|} + \frac{((R_n-R)X_{21},0)}{\|(X_{12},X_{21})\|} + \frac{(C_n,0)}{\|(X_{12},X_{21})\|}, \tag{6.18}$$

$$|\|(RX_{21},X_{21})\| - \|((R_n-R)X_{21},0)\| - \|(C_n,0)\|| \leq \|(X_{12},X_{21})\|, \tag{6.19}$$

and

$$\|(X_{12},X_{21})\| \leq \|(RX_{21},X_{21})\| + \|((R_n-R)X_{21},0)\| + \|(C_n,0)\|. \tag{6.20}$$

We aim to show the unit vector in (6.17) is approaching the bad set A, which will lead to a contradiction of Condition **C2**. To this end note by factoring out $\|(RX_{21},X_{21})\|$ from (6.19) and (6.20), we get

$$\|(X_{12},X_{21})\| \geq \|(RX_{21},X_{21})\| \left|1 - \frac{\|((R_n-R)X_{21},0)\| + |C_n|}{\|(RX_{21},X_{21})\|}\right|,$$

and

$$\|(X_{12},X_{21})\| \leq \|(RX_{21},X_{21})\| \left[1 + \frac{\|((R_n-R)X_{21},0)\| + |C_n|}{\|(RX_{21},X_{21})\|}\right].$$

Since $|C_n| = o(a_n)$, it follows from (6.20) that if $\|(X_{12},X_{21})\| \geq ra_n$, then $\|(RX_{21},X_{21})\| \geq ca_n$ for some $c > 0$. It then follows that on $\{\|(X_{12},X_{21})\| \geq ra_n, \mathcal{E}_{12}^n\}$

$$\|(X_{12},X_{21})\| \geq \|(RX_{21},X_{21})\| \left[1 - o(1)\right],$$

and

$$\|(X_{12},X_{21})\| \leq \|(RX_{21},X_{21})\| \left[1 + o(1)\right].$$

Thus

$$
\begin{aligned}
\lim_{n\to\infty} n\mathbb{P}&\left(\frac{(X_{12}, X_{21})}{\|(X_{12}, X_{21})\|} \in O, \|(X_{12}, X_{21})\| \geq ra_n, \mathcal{E}_{12}^n\right)\\
&= \lim_{n\to\infty} n\mathbb{P}\left(\frac{(RX_{21}, X_{21})}{\|(RX_{21}, X_{21})\|} + o(1) \in O, \|(X_{12}, X_{21})\| \geq ra_n, \mathcal{E}_{12}^n\right) \qquad (6.21)\\
&= 0,
\end{aligned}
$$

since $\frac{(RX_{21}, X_{21})}{\|(RX_{21}, X_{21})\|} + o(1)$ is a small perturbation of a vector in A which is disjoint from O. From (6.14), (6.15), and (6.21) we see

$$
\lim_{n\to\infty} n\mathbb{P}\left(\frac{(X_{12}, X_{21})}{\|(X_{12}, X_{21})\|} \in O, \|(X_{12}, X_{21})\| \geq ra_n\right) = 0.
$$

Thus for arbitrary $(z_1, z_2) \in \mathbb{C}^2$, $(z_1, z_2) \notin \mathrm{supp}(\theta_d)$ and $\mathrm{supp}(\theta_d) \subseteq A$, a contradiction of Condition **C2** (i). Therefore there exists arbitrarily large n such that

$$
|\mathrm{Corr}\left(X_{12}|\mathcal{E}_{12}^n, X_{21}|\mathcal{E}_{12}^n\right)| < 1.
$$

From the above we see Theorem 5.1 may be applied to $A_n - z$, which combining with Lemma 6.7 completes the proof of Theorem 6.6. $\qquad\square$

6.4 Uniform integrability

For $0 < \delta < 1$ we define $K_\delta = [\delta, \delta^{-1}]$. We aim to show that $\log(\cdot)$ is uniformly integrable in probability with respect to $\{\nu_{A-z}\}_{n\geq 1}$, i.e. for all $\varepsilon > 0$

$$
\lim_{\delta\to 0}\lim_{n\to\infty}\mathbb{P}\left(\int_{K_\delta^c} |\log(x)|d\nu_{A_n-z}(x) > \varepsilon\right) = 0. \qquad (6.22)
$$

From Lemma 6.1 there exists a constant $c_0 > 0$ such that with probability 1

$$
\limsup_{n\to\infty}\int_1^\infty |\log(x)|d\nu_{A_n-z}(x) < c_0.
$$

From this, the part of the integral in (6.22) over $[\delta^{-1}, \infty)$ is not a concern. Thus it suffices to prove that for every sequence δ_n converging to 0,

$$
\frac{1}{n}\sum_{i=0}^{n-1}\mathbf{1}_{\{s_{n-i}\leq\delta_n\}}\log s_{n-i}^{-2}
$$

converges to 0 in probability. By Theorem 6.6 we may, with probability $1 - o(1)$, lower bound s_{n-i} by cn^{-r} for all i. Take $0 < \gamma < \alpha/4$ to be fixed later. Using the polynomial lower bound for $0 \leq i \leq n^{1-\gamma}$, it follows that it is sufficient to prove that

$$
\frac{1}{n}\sum_{i=\lfloor n^{1-\gamma}\rfloor}^{n-1}\mathbf{1}_{\{s_{n-i}\leq\delta_n\}}\log s_{n-i}^{-2}
$$

converges in probability to 0.

We next aim to show there exists an event F_n such that for some $\delta > 0$ and $c > 0$

$$
\mathbb{P}(F_n^c) \leq c\exp(-n^\delta),
$$

and

$$
\mathbb{E}[s_{n-i}^{-2}|F_n] \leq c\left(\frac{n}{i}\right)^{\frac{2}{\alpha}+1},
$$

for $\lfloor n^{1-\gamma} \rfloor \leq i \leq n-1$. First to see why this implies convergence in probability to zero, note

$$\mathbb{P}(s_{n-i} \leq \delta_n) \leq \mathbb{P}(F_n^c) + c\delta_n^2 \left(\frac{n}{i}\right)^{1+2/\alpha}.$$

It follows there exists a sequence $\varepsilon_n = \delta_n^{1/(\frac{2}{\alpha}+1)}$ tending to zero such that $\mathbb{P}(s_{n-\lfloor n\varepsilon_n \rfloor} \leq \delta_n)$ converges to 0. Hence it is sufficient to prove, given F_n,

$$\frac{1}{n} \sum_{i=\lfloor n^{1-\gamma} \rfloor}^{\lfloor \varepsilon_n n \rfloor} \log s_{n-i}^{-2} \tag{6.23}$$

converges to zero in probability. Using the negative second moment bound on F_n and the concavity of $\log$ we have

$$\mathbb{E}\left[\frac{1}{n} \sum_{i=\lfloor n^{1-\gamma} \rfloor}^{\lfloor \varepsilon_n n \rfloor} \mathbf{1}_{\{s_{n-i} \leq \delta_n\}} \log s_{n-i}^{-2} \big| F_n\right] \leq \frac{1}{n} \sum_{i=\lfloor n^{1-\gamma} \rfloor}^{\lfloor \varepsilon_n n \rfloor} \log \mathbb{E}[s_{n-i}^{-2}|F_n]$$

$$\leq \frac{c_1}{n} \sum_{i=1}^{\lfloor \varepsilon_n n \rfloor} \log\left(\frac{n}{i}\right)$$

$$\to 0$$

where the last sum converges to zero by a Riemann Sum approximation. Thus, by Markov's inequality we obtain convergence (given F_n) of (6.23) in probability to zero.

Now we finish with the construction of such an event F_n. Let B_n be the matrix formed by the first $n - \lfloor i/2 \rfloor$ rows of $a_n(A_n - z)$. If $s_1' \geq s_2' \geq \cdots \geq s_{n-\lfloor i/2 \rfloor}'$ are the singular values of B_n, then by Cauchy interlacing

$$s_{n-i} \geq \frac{s_{n-i}'}{a_n}. \tag{6.24}$$

By the Tao-Vu negative second moment lemma, Lemma A.4, we have

$$s_1'^{-2} + \cdots s_{n-\lfloor i/2 \rfloor}'^{-2} = \text{dist}_1^{-2} + \cdots \text{dist}_{n-\lfloor i/2 \rfloor}^{-2} \tag{6.25}$$

where dist_j is the distance from the j-th row of B_n to the subspace spanned by the other rows. Using (6.24) and (6.25) we get

$$\frac{i}{2} s_{n-i}^{-2} \leq a_n^2 \sum_{j=i}^{n-\lfloor i/2 \rfloor} \text{dist}_j^{-2}.$$

Now let dist_j' be the distance from the j-th row of B_n with its j-th entry removed and the subspace spanned by the other rows minus their j-th entry. Since $\text{dist}_j \geq \text{dist}_j'$ we have

$$\frac{i}{2} s_{n-i}^{-2} \leq a_n^2 \sum_{j=i}^{n-\lfloor i/2 \rfloor} \text{dist}_j'^{-2} \tag{6.26}$$

Let $F_{n,i}$ be the event that for all $1 \leq j \leq n - \lfloor i/2 \rfloor$, $\text{dist}_j' \geq (n-1)^{(1-2\gamma)/\alpha}$. Since the span of all but 1 row of B_n is at most $n - i/2$, we can use Proposition 6.2 to obtain, for sufficiently large n,

$$\mathbb{P}(F_{n,i}^c) \leq \exp(-(n-1)^{\delta'}),$$

for some $\delta' > 0$ (after a union bound). Let $F_n = \bigcap_{i=0}^{n^{1-\gamma}} F_{n,i}$, then

$$\mathbb{P}(F_n^c) \leq \exp(-(n-1)^{\delta}),$$

for some $\delta > 0$.

We now aim to show the desired negative second moment bound on F_n. Recall from Proposition 6.5 there exists an event E_j independent from the rows $i \neq j$ of B_n without their i-th entry such that $\mathbb{P}(E_j^c) \leq n^{-\frac{1}{2}+\gamma(\frac{2}{\alpha}-\frac{1}{2})}$ and for any subspace W with dimension $d < n - n^{1-\gamma}$ one has

$$\mathbb{E}[\text{dist}(R, W)^{-2}; E_j] \leq c(n-d)^{-2/\alpha}$$

where R is the j-th row of B_n with the j-th entry removed. Thus

$$\mathbb{E}[\text{dist}_j'^{-2}; E_j] = O(i^{-2/\alpha}),$$

for $i \leq j \leq n - \lfloor i/2 \rfloor$.

By taking the lower bound of $\text{dist}_j' \geq (n-1)^{(1-2\gamma)/\alpha}$ on $E_j^c \cap F_n$, we get

$$\mathbb{E}[\text{dist}_j'^{-2}; F_n] \leq c_2 \left(i^{-2/\alpha} + n^{-\frac{1}{2}+\gamma(\frac{2}{\alpha}-\frac{1}{2})-2(1-2\gamma)/\alpha} \right).$$

So for γ, not dependent on i, sufficiently small one has

$$\mathbb{E}[\text{dist}_j'^{-2}; F_n] \leq c_3 i^{-2/\alpha}.$$

Thus by (6.26)

$$\mathbb{E}[s_{n-i}^{-2}; F_n] \leq c_3 a_n^2 n i^{-(1+2/\alpha)}.$$

The result then follows from the assumption $a_n \sim cn^{1/\alpha}$.

6.5 Proof of Theorem 1.6

We have shown that almost surely $\nu_{A_n-zI_n}$ converges weakly to ν_{z,α,θ_d} and that $\log(\cdot)$ is uniformly integrable in probability with respect to $(\nu_{A_n-zI_n})_{n\geq 1}$. Thus by Lemma 1.12, μ_{A_n} converges weakly to some deterministic probability measure μ_{α,θ_d} in probability.

A Additional lemmas

A.1 Concentration

Lemma A.1 (McDiarmid's inequality, Theorem 3.1 in [38]). *Let $X_1, X_2, \ldots, X_n$ be independent random variables taking values in $R_1, R_2, \ldots, R_n$ respectively. Let $F : R_1 \times \cdots \times R_n \to \mathbb{C}$ be a function with the property that for every $1 \leq i \leq n$ there exists a $c_i > 0$ such that*

$$|F(x_1, x_2, \ldots, x_i, \ldots, x_n) - F(x_1, x_2, \ldots, x_i', \ldots, x_n)| \leq c_i$$

for all $x_j \in R_j$, $x_i' \in R_i$ for $1 \leq j \leq n$. Then for any $t > 0$,

$$\mathbb{P}(|F(X) - \mathbb{E}F(X)| \geq \sigma t) \leq C \exp(-ct^2)$$

for absolute constants $C, c > 0$ and $\sigma^2 := \sum_{i=1}^n c_i^2$.

For the following lemma we need a way of breaking up an elliptic random matrix X into independent pieces. For a matrix $M = (m_{ij})_{i,j=1}^n$ we let the k-wedge of M be the $n \times n$ matrix C_k with entries $c_{ij} = m_{ij}$ if either $i = k$ and $j \geq i$ or $j = k$ and $i \geq j$, with $c_{ij} = 0$ otherwise. Note

$$M_n = \sum_{k=1}^n C_k,$$

and $\text{rank}(C_k) \leq 2$ for all $1 \leq k \leq n$. Recall the total variation norm of a function $f : \mathbb{R} \to \mathbb{R}$ is given by

$$\|f\|_{TV} = \sup \sum_{k \in \mathbb{Z}} |f(x_{k+1}) - f(x_k)|,$$

where the supremum runs over all sequences $(x_k)_{k \in \mathbb{Z}}$ with $x_{k+1} \geq x_k$.

Lemma A.2. *Let M be a $n \times n$ random matrix with independent wedges. Then for any $f : \mathbb{R} \to \mathbb{R}$ going to 0 at $\pm\infty$ with $\|f\|_{TV} \le 1$ and every $t \ge 0$,*

$$\mathbb{P}\left(\left|\int f d\nu_M - \mathbb{E}\int f d\nu_M\right| \ge t\right) \le C \exp(-cnt^2),$$

for absolute constants $C, c > 0$.

Proof. If $A, B \in \mathrm{Mat}_n(\mathbb{C})$ let F_A and F_B be the cumulative distribution functions of ν_A and ν_B. By the Lidskii inequality (see Theorem 3.3.16 in [34]) for singular values

$$\|F_A - F_B\|_\infty \le \frac{\mathrm{rank}(A - B)}{n}.$$

Assume f is smooth. Integrating by parts we get

$$\left|\int f d\nu_A - \int f d\nu_B\right| = \left|\int_\mathbb{R} f'(t)(F_A(t) - F_B(t))dt\right| \le \frac{\mathrm{rank}(A - B)}{n}\|f\|_{TV}. \tag{A.1}$$

Since $\left|\int f d\nu_A - \int f d\nu_B\right|$ depends on only finitely many points for any f, we can extend the previous inequality to any function of finite total variation.

Now fix $f : \mathbb{R} \to \mathbb{R}$ going to 0 at $\pm\infty$ with $\|f\|_{TV} \le 1$. Let $\mathcal{C}_k$ be the space of all k-wedges and $F_f : \mathcal{C}_1 \times \mathcal{C}_2 \times \cdots \times \mathcal{C}_n$ be the function defined by

$$F_f(C_1, C_2, \ldots, C_n) = \int f d\nu_A$$

where A is the matrix with k-wedge C_k. By (A.1)

$$|F_f(C_1, C_2, \ldots, C_i, \ldots, C_n) - F_f(C_1, C_2, \ldots, C_i', \ldots, C_n)| \le \frac{2}{n},$$

and thus by McDiarmid's inequality, Lemma A.1, and the definition of F_f,

$$\mathbb{P}\left(\left|\int f d\nu_M - \mathbb{E}\int f d\nu_M\right| \ge t\right) \le C \exp(-cnt^2). \qquad \square$$

A.2 Singular value estimates

Lemma A.3 (See [34], Chapter 3)**.** *If A and B are $n \times n$ complex matrices then*

$$s_1(AB) \le s_1(A)s_1(B) \text{ and } s_1(A + B) \le s_1(A) + s_1(B),$$

$$\max_{1 \le k \le n} |s_k(A + B) - s_k(A)| \le \|B\|,$$

$$s_{i+j-1}(A + B) \le s_i(A) + s_j(B)$$

for $1 \le i, j \le n$ and $i + j \le n + 1$. In addition,

$$\max_{1 \le i \le n} |s_i(A) - s_i(B)| \le s_1(A - B).$$

Lemma A.4 (Tao-Vu Negative Second Moment, [54] Lemma A.4)**.** *If A is a full rank $n' \times n$ complex matrix with rows $R_1, \ldots, R_{n'}$ and $R_{-i} := \mathrm{Span}\{R_j : j \ne i\}$, then*

$$\sum_{i=1}^{n'} s_i(A)^{-2} = \sum_{i=1}^{n'} \mathrm{dist}(R_i, R_{-i})^{-2}.$$

Lemma A.5 (Cauchy interlacing, see [34])**.** *Let A be an $n \times n$ complex matrix. If B is an $n' \times n$ matrix obtained by deleting $n - n'$ rows from A, then for every $1 \le i \le n'$,*

$$s_i(A) \ge s_i(B) \ge s_{i+n-n'}(A).$$

Lemma A.6 (See Remark 1 in [24])**.** *Let A be an $n \times n$ matrix and let $\mathrm{Re}(A) = (A + A^*)/2$ denote the real part of A. If $s_1 \ge s_2 \ge \cdots \ge s_n$ and $\lambda_1 \ge \lambda_2 \ge \cdots \ge \lambda_n$ denote the singular values of A and eigenvalues of $\mathrm{Re}(A)$ respectively, then*

$$\lambda_j \le s_j \tag{A.2}$$

for every $1 \le j \le n$.

Lemma A.7 (Weyl's inequality, [58], see also Lemma B.5 in [13])**.** *For every $n \times n$ complex matrix A with eigenvalues ordered as $|\lambda_1(A)| \ge |\lambda_2(A)| \ge \cdots \ge |\lambda_n(A)|$ one has*

$$\prod_{i=1}^{k} |\lambda_i(A)| \le \prod_{i=1}^{k} s_i(A) \text{ and } \prod_{i=k}^{n} s_i(A) \le \prod_{i=k}^{n} |\lambda_i(A)|$$

for all $1 \le k \le n$. Moreover for $r > 0$

$$\sum_{k=1}^{n} |\lambda_k(A)|^r \le \sum_{k=1}^{n} s_k(A)^r.$$

Lemma A.8 (Schatten Bound, see proof of Theorem 3.32 in [59])**.** *Let A be an $n \times n$ complex matrix with rows $R_1, \ldots, R_n$. Then for every $0 < r \le 2$,*

$$\sum_{i=1}^{n} s_k(A)^r \le \sum_{i=1}^{r} \|R_k\|^r.$$

A.3 Moments of stable distributions

Lemma A.9 (See [13] and [25] Theorem VIII.9.2)**.** *Let Z be a positive random variable such that for every $t > 0$,*

$$\mathbb{P}(Z \ge t) = L(t)t^{-\alpha}$$

for some slowly varying function L and some $\alpha \in (0, 2)$. Then for every $p > \alpha$,

$$\lim_{t \to \infty} \frac{\mathbb{E}[Z^p \mathbf{1}_{\{Z \le t\}}]}{c(p)L(t)t^{p-\alpha}} \to 1,$$

where $c(p) := \alpha/(p - \alpha)$.

A.4 Proof of Proposition 4.12

From Proposition 4.3 one has

$$\begin{pmatrix} a(z,\eta) & b(z,\eta) \\ b'(z,\eta) & c(z,\eta) \end{pmatrix} = -\left(\begin{pmatrix} \eta & z \\ \bar{z} & \eta \end{pmatrix} + \sum_{k=1}^{\infty} \begin{pmatrix} 0 & y_k^{(1)} \\ \bar{y}_k^{(2)} & 0 \end{pmatrix} \tilde{R}(U)_{kk} \begin{pmatrix} 0 & y_k^{(2)} \\ \bar{y}_k^{(1)} & 0 \end{pmatrix} \right)^{-1}.$$

We will consider the point process $\{(y_k^{(1)}, y_k^{(2)})\}_{k \ge 1}$ in the polar form $\{(r_k, w_k)\}_{k \ge 1}$ where $\{r_k\}_{k \ge 1}$ is a Poisson point process with intensity measure m_α and $w_1, w_2, \ldots$ is a collection of independent identically θ_d distributed random variables independent of $\{r_k\}_{k \ge 1}$. We will denote the coordinates of w_k by $(w_k^{(1)}, w_k^{(2)})$. In polar coordinates the recursive equation becomes

$$\begin{pmatrix} a(z,\eta) & b(z,\eta) \\ b'(z,\eta) & c(z,\eta) \end{pmatrix} = -\left(\begin{pmatrix} \eta & z \\ \bar{z} & \eta \end{pmatrix} + \sum_{k=1}^{\infty} r_k^2 \begin{pmatrix} c_k(z,\eta)|w_k^{(1)}|^2 & b'_k(z,\eta)w_k^{(1)}w_k^{(2)} \\ b_k(z,\eta)\bar{w}_k^{(1)}\bar{w}_k^{(2)} & a_k(z,\eta)|w_k^{(2)}|^2 \end{pmatrix} \right)^{-1},$$

$$\tag{A.3}$$

where

$$\begin{pmatrix} a_k(z,\eta) & b_k(z,\eta) \\ b'_k(z,\eta) & c_k(z,\eta) \end{pmatrix} := \tilde{R}(U)_{kk}.$$

To complete the proof that the matrix in (4.3) satisfies the distributional equation we need the following lemma, which is essentially Theorem 10.3 in [8].

Lemma A.10. *Let $\{r_k\}_{k \geq 1}$ be a Poisson point process with intensity measure m_α for $\alpha \in (0,2)$ and $v_1, v_2, \ldots$ be a collection of bounded independent and identically ν distributed random vectors for some probability measure ν. Then*

$$\sum_{k=1}^{\infty} r_k^2 v_k \stackrel{d}{=} S, \tag{A.4}$$

where S is an $\frac{\alpha}{2}$-stable random vector with spectral measure Γ_ν, where Γ_ν is the measure on the unit sphere obtained by the image of the measure $\frac{\Gamma(2-\alpha/2)\cos(\pi\alpha/4)}{1-\alpha/2}\|v\|^{\alpha/2}d\nu(v)$ under the map $v \mapsto \frac{v}{\|v\|}$.

Proof. Let $(X_k)_{k \geq 1}$ be a sequence of i.i.d. non-negative random variables such that

$$\mathbb{P}(X_1 \geq u) = \frac{L(u)}{u^{\alpha/2}}, \tag{A.5}$$

for some slowly varying function L, and define the normalizing constants

$$b_n := \inf\left\{b : \mathbb{P}(X \geq b) \leq \frac{1}{n}\right\}. \tag{A.6}$$

From Theorem 10.3 in [8], $\frac{1}{b_n}\sum_{k=1}^{n} X_k v_k$ converges in distribution to S. It also follows from Proposition 2.3 that

$$\sum_{k=1}^{n} \delta_{X_k v_k/b_n} \Rightarrow \sum_{k=1}^{\infty} \delta_{r_k^2 v_k}, \tag{A.7}$$

as $n \to \infty$. It then follows from the almost sure uniform summability of the sequence $(X_k)_{k \geq 1}$ that

$$\frac{1}{b_n}\sum_{k=1}^{n} X_k v_k \Rightarrow \sum_{k=1}^{\infty} r_k^2 v_k, \tag{A.8}$$

as $n \to \infty$. $\qquad\square$

An application of Lemma A.10 to the series in (A.3) gives (4.7).

References

[1] Amol Aggarwal, Patrick Lopatto, and Jake Marcinek, *Eigenvector statistics of Lévy matrices*, The Annals of Probability **49** (2021), no. 4, 1778–1846. MR4260468

[2] Amol Aggarwal, Patrick Lopatto, and Horng-Tzer Yau, *GOE statistics for Lévy matrices*, Journal of the European Mathematical Society **23** (2021), no. 11, 3707–3800. MR4310816

[3] David Aldous and J. Michael Steele, *The objective method: Probabilistic combinatorial optimization and local weak convergence*, Probability on Discrete Structures, Encyclopaedia of Mathematical Sciences, vol. 110, Springer-Verlag, New York, 2004, pp. 1–72. MR2023650

[4] Antonio Auffinger, Gérard Ben Arous, and Sandrine Péché, *Poisson convergence for the largest eigenvalues of heavy tailed random matrices*, Ann. Inst. Henri Poincaré, Probab. Stat. **45** (2009), no. 3, 589–610. MR2548495

[5] Antonio Auffinger and Si Tang, *Extreme eigenvalues of sparse, heavy tailed random matrices*, Stochastic Processes and their Applications **126** (2016), no. 11, 3310–3330. MR3549709

[6] Bojan Basrak, Yeonok Cho, Johannes Heiny, and Paul Jung, *Extreme eigenvalue statistics of m-dependent heavy-tailed matrices*, Annales de l'Institut Henri Poincaré, Probabilités et Statistiques **57** (2021), no. 4, 2100–2127. MR4415391

[7] Serban Belinschi, Amir Dembo, and Alice Guionnet, *Spectral measure of heavy tailed band and covariance random matrices*, Communications in Mathematical Physics **289** (2009), no. 3, 1023–1055. MR2511659

[8] Gérard Ben Arous and Alice Guionnet, *The spectrum of heavy tailed random matrices*, Communications in Mathematical Physics **278** (2007), no. 3, 715–751. MR2373441

[9] Florent Benaych-Georges and Alice Guionnet, *Central limit theorem for eigenvectors of heavy tailed matrices*, Electron. J. Probab. **19** (2014), 27 pp. MR3227063

[10] Rajendra Bhatia, *Matrix analysis*, Graduate Texts in Mathematics, vol. 169, Springer-Verlag, New York, 1997. MR1477662

[11] G Biroli, J.-P Bouchaud, and M Potters, *On the top eigenvalue of heavy-tailed random matrices*, Europhysics Letters (EPL) **78** (2007), no. 1, 10001. MR2371333

[12] Charles Bordenave, Pietro Caputo, and Djalil Chafaï, *Spectrum of large random reversible Markov chains: heavy-tailed weights on the complete graph*, The Annals of Probability **39** (2011), no. 4, 1544–1590. MR2857250

[13] Charles Bordenave, Pietro Caputo, and Djalil Chafaï, *Spectrum of non-Hermitian heavy tailed random matrices*, Communications in Mathematical Physics **307** (2011), no. 2, 513–560. MR2837123

[14] Charles Bordenave, Pietro Caputo, and Djalil Chafaï, *Spectrum of Markov generators on sparse random graphs*, Comm. Pure Appl. Math. **67** (2014), no. 4, 621–669. MR3168123

[15] Charles Bordenave, Pietro Caputo, Djalil Chafaï, and Daniele Piras, *Spectrum of large random Markov chains: Heavy-tailed weights on the oriented complete graph*, Random Matrices: Theory and Applications **06** (2017), no. 02, 1750006. MR3656964

[16] Charles Bordenave and Djalil Chafaï, *Around the circular law*, Probab. Surveys **9** (2012), 1–89. MR2908617

[17] Charles Bordenave and Alice Guionnet, *Localization and delocalization of eigenvectors for heavy-tailed random matrices*, Probability Theory and Related Fields **157** (2013), no. 3-4, 885–953. MR3129806

[18] Charles Bordenave and Alice Guionnet, *Delocalization at small energy for heavy-tailed random matrices*, Communications in Mathematical Physics **354** (2017), no. 1, 115–159. MR3656514

[19] P. Cizeau and J. P. Bouchaud, *Theory of Lévy matrices*, Phys. Rev. E **50** (1994), 1810–1822.

[20] Kevin P. Costello, *Bilinear and quadratic variants on the Littlewood-Offord problem*, Israel Journal of Mathematics **194** (2013), no. 1, 359–394. MR3047075

[21] Kevin P. Costello, Terence Tao, and Van Vu, *Random symmetric matrices are almost surely nonsingular*, Duke Math. J. **135** (2006), no. 2, 395–413. MR2267289

[22] Youri Davydov and Vladimir Egorov, *On convergence of empirical point processes*, Statistics & Probability Letters **76** (2006), no. 17, 1836–1844. MR2271178

[23] Rick Durrett, *Probability: Theory and examples*, 4 ed., Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2010. MR2722836

[24] Ky Fan and A. J. Hoffman, *Some metric inequalities in the space of matrices*, Proceedings of the American Mathematical Society **6** (1955), no. 1, 111–116. MR67841

[25] William Feller, *An introduction to probability theory and its applications. Vol. II.*, Second edition, John Wiley & Sons Inc., New York, 1971. MR0270403

[26] VL Girko, *Elliptic law*, Theory of Probability & Its Applications **30** (1986), no. 4, 677–690. MR816278

[27] VL Girko, *The elliptic law: ten years later I*, Random Operators and Stochastic Equations **3** (1995), no. 3, 257–302. MR1354817

[28] F. Götze, *Asymptotic expansions for bivariate von Mises functionals*, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **50** (1979), no. 3, 333–355. MR554550

[29] F. Götze, A. Naumov, and A. Tikhomirov, *On a generalization of the elliptic law for random matrices*, Acta Physica Polonica B **46** (2015), no. 9, 1737–1745. MR3403838

[30] F. Götze, A. Naumov, and A. Tikhomirov, *On minimal singular values of random matrices with correlated entries*, Random Matrices: Theory and Applications **04** (2015), no. 02, 1550006. MR3356884

[31] Ewa Gudowska-Nowak, A. Jarosz, Maciej Nowak, and Gábor Papp, *Towards non-hermitian random Lévy matrices*, Acta Physica Polonica B **38** (2007). MR2372777

[32] Johannes Heiny and Jianfeng Yao, *Limiting distributions for eigenvalues of sample correlation matrices from heavy-tailed populations*, arXiv:2003.03857

[33] Remco van der Hofstad, *Random graphs and complex networks*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 1, Cambridge University Press, 2016. MR3617364

[34] Roger Horn and Charles R Johnson, *Topics in matrix analysis*, Cambridge University Press, 1991. MR1091716

[35] Paul Jung, *Lévy-Khintchine random matrices and the poisson weighted infinite skeleton tree*, Transactions of the American Mathematical Society **370** (2014). MR3717992

[36] A. E. Litvak, A. Pajor, M. Rudelson, and N. Tomczak-Jaegermann, *Smallest singular value of random matrices and geometry of random polytopes*, Adv. Math. **195** (2005), no. 2, 491–523. MR2146352

[37] Michail Louvaris, *Universality of the least singular value and singular vector delocalization for levy non-symmetric matrices*, arXiv:2204.06399

[38] Colin McDiarmid, *Concentration*, Probabilistic Methods for Algorithmic Discrete Mathematics (Michel Habib, Colin McDiarmid, Jorge Ramirez-Alfonsin, and Bruce Reed, eds.), Algorithms and Combinatorics, Springer-Verlag, New York, 1998. MR1678578

[39] Cécile Monthus, *Localization transition in random Lévy matrices: multifractality of eigenvectors in the localized phase and at criticality*, Journal of Statistical Mechanics: Theory and Experiment **2016** (2016), no. 9, 093304. MR3562241

[40] Alexey Naumov, *Elliptic law for real random matrices*, arXiv:1201.1639

[41] Hoi H. Nguyen, *Random doubly stochastic matrices: the circular law*, Ann. Probab. **42** (2014), no. 3, 1161–1196. MR3189068

[42] Hoi H. Nguyen and Sean O'Rourke, *The elliptic law*, International Mathematics Research Notices **2015** (2014), no. 17, 7620–7689. MR3403996

[43] Sean O'Rourke and David Renfrew, *Central limit theorem for linear eigenvalue statistics of elliptic random matrices*, Journal of Theoretical Probability **29** (2015), no. 3, 1121–1191. MR3540493

[44] Sean O'Rourke, David Renfrew, Alexander Soshnikov, and Van Vu, *Products of independent elliptic random matrices*, Journal of Statistical Physics **160** (2015), no. 1, 89–119. MR3357969

[45] Michael Reed and Barry Simon, *Methods of modern mathematical physics*, vol. 1, Academic Press, 1972. MR0493419

[46] Sidney I. Resnick, *Extreme values, regular variation and point processes*, 1 ed., Springer Series in Operations Research and Financial Engineering, Springer-Verlag, New York, 1987. MR900810

[47] Sidney I. Resnick, *Heavy-tailed phenomemena*, 1 ed., Springer Series in Operations Research and Financial Engineering, Springer-Verlag, New York, 2007. MR2271424

[48] Mark Rudelson and Roman Vershynin, *The Littlewood-Offord problem and invertibility of random matrices*, Adv. Math. **218** (2008), no. 2, 600–633. MR2407948

[49] GN Sakovich, *A single form for the conditions for attraction to stable laws*, Theory of Probability & Its Applications **1** (1956), no. 3, 322–325.

[50] Gennady Samorodnitsky and Murad S Taqqu, *Stable non-gaussian random processes*, Stochastic Modeling, Chapman & Hall, New York, NY, 1994. MR1280932

[51] Alexander Sidorenko, *A correlation inequality for bipartite graphs*, Graphs and Combinatorics **9** (1993), no. 2, 201–204. MR1225933

[52] Alexander Soshnikov, *Poisson statistics for the largest eigenvalues of Wigner random matrices with heavy tails*, Electronic Communications in Probability **9** (2004), 82–91. MR2081462

[53] Terence Tao, *Topics in random matrix theory*, Graduate Studies in Mathematics, vol. 132, American Mathematical Society, Providence, RI, 2012. MR2906465

[54] Terence Tao, Van Vu, and Manjunath Krishnapur, *Random matrices: Universality of ESDs and the circular law*, Ann. Probab. **38** (2010), no. 5, 2023–2065. MR2722794

[55] E. Tarquini, G. Biroli, and M. Tarzia, *Level statistics and localization transitions of Lévy matrices*, Physical Review Letters **116** (2016), no. 1. MR3555687

[56] Konstantin E. Tikhomirov, *The smallest singular value of random rectangular matrices with no moment assumptions on entries*, Israel J. Math. **212** (2016), no. 1, 289–314. MR3504328

[57] Roman Vershynin, *Invertibility of symmetric random matrices*, Random Structures & Algorithms **44** (2014), no. 2, 135–182. MR3158627

[58] Hermann Weyl, *Inequalities between the two kinds of eigenvalues of a linear transformation*, Proceedings of the National Academy of Sciences **35** (1949), no. 7, 408–411. MR30693

[59] Xinghzi Zhan, *Matrix inequalities*, Lecture Notes in Mathematics, vol. 1790, Springer-Verlag, New York, 2002. MR1927396