THE CIRCULAR LAW FOR RANDOM REGULAR DIGRAPHS
WITH RANDOM EDGE WEIGHTS

NICHOLAS COOK

Abstract. We consider random $n \times n$ matrices of the form $Y_n = \frac{1}{\sqrt{d}} A_n \circ X_n$, where $A_n$ is the adjacency matrix of a uniform random $d$-regular directed graph on $n$ vertices, with $d = \lfloor pn \rfloor$ for some fixed $p \in (0, 1)$, and $X_n$ is an $n \times n$ matrix of iid centered random variables with unit variance and finite $4 + \eta$-th moment (here $\circ$ denotes the matrix Hadamard product). We show that as $n \to \infty$, the empirical spectral distribution of $Y_n$ converges weakly in probability to the normalized Lebesgue measure on the unit disk.

1. Introduction

1.1. Background. For an $n \times n$ matrix $M$ with complex entries, denote by

$$\mu_M = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(M)}$$  \hspace{1cm} (1.1)

its empirical spectral distribution (ESD), where $\lambda_1(M), \ldots, \lambda_n(M) \in \mathbb{C}$ are the eigenvalues of $M$, counted with multiplicity, and labeled in some arbitrary fashion. When $M$ is a random matrix, $\mu_M$ is a random probability measure. In the following we denote the space of bounded continuous (resp. continuous and compactly supported) real-valued test functions over a space $\mathcal{X}$ by $C_b(\mathcal{X})$ (resp. $C_c(\mathcal{X})$).

Definition 1.1 (Convergence of random measures). Let $(\mu_n)_{n \geq 1}$ be a sequence of random Borel probability measures supported on $\mathcal{X} = \mathbb{R}$ or $\mathbb{C}$, and let $\mu$ be another Borel probability measure on $\mathcal{X}$. We say $\mu_n$ converges to $\mu$ weakly in probability if for all $f \in C_b(\mathcal{X})$ the random variable $\int_{\mathcal{X}} f \, d\mu_n$ converges in probability, i.e. for any $\varepsilon > 0$,

$$\mathbb{P} \left( \left| \int_{\mathcal{X}} f \, d\mu_n - \int_{\mathcal{X}} f \, d\mu \right| > \varepsilon \right) \to 0 \quad \text{as } n \to \infty. \hspace{1cm} (1.2)$$

We say $\mu_n$ converges to $\mu$ weakly almost surely if for all $f \in C_b(\mathcal{X})$ the random variable $\int_{\mathcal{X}} f \, d\mu_n$ converges almost surely to $\int_{\mathcal{X}} f \, d\mu$. We similarly define convergence of $\mu_n$ to $\mu$ vaguely in probability and vaguely almost surely by replacing the space $C_b(\mathcal{X})$ with $C_c(\mathcal{X})$ above.

Date: March 3, 2017.

During the completion of this work the author was partially supported by NSF grant DMS-1266164.
For random matrices with iid entries having finite second moment, the global distribution of the spectrum is asymptotically governed by the circular law:

**Theorem 1.2** (Circular law). Let \( \xi \) be a complex-valued random variable with mean zero and variance one, and for each \( n \) let \( X_n = (\xi_{ij})_{1 \leq i,j \leq n} \) be a matrix whose entries are iid copies of \( \xi \). Then the rescaled ESD \( \mu_{1/\sqrt{n}} X_n \) converges weakly almost surely to \( \mu_{\text{circ}} \), the uniform measure on \( B_{\mathbb{C}}(0,1) \).

In the above form Theorem 1.2 was established by Tao and Vu in [TV10], building on important advances of Girko [Gir84] and Bai [Bai97]. The circular law was earlier established by Mehta for the case that \( \xi \) is a standard complex Gaussian [Meh67], and later by Edelman in the real Gaussian case [Ede97]. Prior to the work [TV10] there were versions of Theorem 1.2 that made more restrictive assumptions on the distribution of \( \xi \) [Bai97,BS10,GT10,PZ10,TV08].

The circular law has been applied to the study of dynamics of networks arising in fields ranging from neuroscience [SCS88,RA06,ARS15] to ecology [May72,AGB+15]. Theorem 1.2 is mathematically interesting as an instance of the universality phenomenon in random matrix theory: in the limit as \( n \to \infty \), the global distribution of the spectrum of \( X_n \) is determined completely by the first two moments of the entries \( \xi_{ij} \). The theorem hence identifies a wide class of matrix ensembles lying in the circular law universality class. For additional background and history on the circular law we point the reader to the survey [BC12].

It is natural to ask whether this universality class extends to include some ensembles not covered by Theorem 1.2, that is, if we can relax the independence, identical distribution, or moment hypotheses on the entries. Informally, one might expect an ensemble \( M_n = (\xi_{ij}) \) to exhibit the circular law if the entries \( \xi_{ij} \) are centered around a common value, have reasonably bounded tails, are not too degenerate (i.e. their distributions do not concentrate too much near deterministic values), and are only weakly correlated.

For iid matrices, the second moment hypothesis is sharp, and the limiting spectral distribution for certain classes of heavy tail matrices was described in [BCC11]. Nevertheless, in [Woo12], Wood showed that the circular law is robust under sparsification: letting \( X_n = (\xi_{ij}) \) be an iid matrix and \( B_n = (b_{ij}) \) a matrix of iid Bernoulli(\( p \)) indicator variables, independent of \( X_n \), the empirical spectral distribution of \( M_n := \frac{1}{\sqrt{np}} B_n \circ X_n \) converges weakly in probability (see Definition 1.1 below) to the circular law, provided that \( p = p(n) \gg n^{\varepsilon - 1} \) for any fixed \( \varepsilon > 0 \). (Here \( \circ \) denotes the matrix Hadamard product, i.e. \( B_n \circ X_n = (b_{ij} \xi_{ij}) \).) The lower bound on \( p \) is near-optimal, as it is easy to see that for \( p \asymp 1/n \) the matrix \( M_n \) will have a linear proportion of trivial columns, so that the limiting spectral distribution will have an atom at zero.

There has been considerable interest in extending the circular law to cover models with dependent entries. In [Ngu14], Nguyen showed that the circular law holds for matrices drawn uniformly (according to volume measure) from the Birkhoff polytope of doubly stochastic matrices. This followed work of Bordenave, Caputo

\[ \text{...} \]
and Chafaï on random stochastic matrices, which still enjoy joint independence of the matrix rows. In [AC15], Adamczak and Chafaï established the circular law for matrices whose distribution over $\mathcal{M}_n(\mathbb{R})$ is a log-concave and unconditional measure, generalizing Edelman’s result for Gaussian matrices [Ede97]. Together with Wolff, the same authors showed in [ACW16] that the circular law holds for a matrix whose entries are exchangeable and have a finite moments of order $20 + \varepsilon$.

A particularly motivating challenge in random matrix theory is to extend universality results to include adjacency matrices of random regular graph models. In the non-Hermitian setting one considers directed graphs (digraphs):}

**Definition 1.3 (rrd matrix).** For $1 \leq d \leq n-1$, denote by $\mathcal{M}_n(d)$ the set of $n \times n$ matrices $M = (m_{ij})$ with $m_{ij} \in \{0,1\}$ for all $i,j \in [n]$, and satisfying the constraint that for all $k \in [n]$,

$$d = \sum_{i=1}^{n} m_{ik} = \sum_{j=1}^{n} m_{kj}.$$  

(1.3)

We refer to a uniform random element $A = A_n \in \mathcal{M}_n(d)$ as a random regular digraph matrix, or rrd matrix.

One can view the rrd matrix $A_n$ as a discrete analogue of the doubly stochastic matrix considered by Nguyen in [Ngu14]. We note also that $A_n$ is not covered by the above-mentioned results in [ACW16]: while the rows and columns of $A_n$ are separately exchangeable, the entries themselves are not.

More generally one can consider random regular digraphs with random edge weights, i.e. matrices of the form $A_n \circ X_n$ with $X_n$ an iid matrix independent from $A_n$. For applications in neuroscience, where random matrices have been used as mathematically tractable models of synaptic matrices [SCS88,RA06,ARS15], random matrices supported on adjacency matrices for directed graphs are of interest as they reflect the sparsity of natural neural networks.

For rrd matrices we have the following augmented version of a conjecture of Bordenave and Chafaï from [BC12] (only case (2) below is stated there).

**Conjecture 1.4 (Bordenave–Chafaï [BC12]).** Let $A_n$ be an rrd matrix.

1. If $d = d_n \leq n/2$ satisfies $d \to \infty$ as $n \to \infty$, then $\mu_{\sqrt{d} A_n}$ converges to the circular law.

2. If $d \geq 3$ is fixed independent of $n$, then $\mu_{A_n}$ converges to the oriented Kesten–McKay law on $\mathbb{C}$, with density given by

$$f_{KM}(w) = \frac{1}{\pi} \frac{d^2(d-1)}{(d^2 - |w|^2)^2} \mathbb{1}_{\{|w| \leq \sqrt{d}\}}.$$  

(1.4)

From the above one obtains results for $n/2 < d \leq n-1$ by considering the complementary matrix with entries $1 - a_{ij}$, which has the same limiting spectral distribution. See Figure 1 for some numerical evidence supporting Conjecture 1.4.

The explicit density in (1.4) can be computed as the Brown measure of the free sum of $d$ Haar unitary operators – see [HL00, Example 5.5]. We note that
in (1.4), if one rescales \( w \) by \( \sqrt{d} \) and sends \( d \to \infty \) the expression converges to the normalized indicator for the unit disk, which gives some evidence for part (1) of the conjecture. By contiguity results (see for instance [Jan95]), to establish Conjecture 1.4 for fixed \( d \) it suffices to consider a different measure, the sum of \( d \) iid uniform permutation matrices. It was shown by Basak and Dembo in [BD13] that the sum of \( d \) iid Haar unitary or orthogonal matrices has limiting spectral distribution given by (1.4), so Conjecture 1.4 posits that their result should hold if the unitaries are restricted to permutation matrices.

The two cases of Conjecture 1.4 parallel known results for undirected regular graphs. Namely, it was shown by McKay in [McK81] that for \( d \geq 3 \) fixed, the limiting spectral distribution of adjacency matrices \( A_n^{\text{Sym}} \) of \( d \)-regular undirected graphs is given by the (explicit) Kesten–McKay law. More recently, it has been shown in [DP12], [TVW13] and [BKY] that if \( d \to \infty \) at certain speeds, the empirical spectral distribution of \( A_n^{\text{Sym}} \) converges to the semicircular law, matching the asymptotic behavior of Wigner matrices (the Hermitian analogue of iid matrices). Moreover, these results show that the convergence of ESDs holds on mesoscopic scales, i.e. on intervals with length shrinking to zero relative to the limiting support of the spectrum, but growing to infinity relative to the mean spacing of eigenvalues. In particular, the work [BKY] has shown convergence at the optimal mesoscopic scale, provided \( d \) grows in the range \( f(n) \ll d \ll \left( \frac{n}{f(n)} \right)^{2/3} \) for some \( f(n) \) growing poly-logarithmically. Very recently, the Kesten–McKay law has been established at the optimal mesoscopic scale for sufficiently large fixed \( d \) [BHY16].

1.2. Main result. In the present work we obtain some partial progress towards Conjecture 1.4 by considering random regular digraphs with random edge weights, i.e. matrices of the form

\[
Y_n = A_n \circ X_n
\]

where \( X_n = (\xi_{ij}) \) is a matrix with iid centered entries of unit variance. One can view this as an rrd matrix with some additional randomness, or alternatively as an iid matrix that has been randomly sparsified to have each row and column supported on \( d \) entries. This is to be compared with the work of Wood [Woo12] discussed above, where the sparsification is performed by iid Bernoulli indicators.

**Theorem 1.5** (Main result). Fix \( p \in (0,1) \), and let \( \xi \in \mathbb{C} \) be a centered random variable with unit variance and \( \mathbb{E}[|\xi|^{4+\eta}] < \infty \) for some fixed \( \eta > 0 \). For each \( n \), let \( A_n \) be a uniform random element of \( \mathcal{M}_n(d) \), where we write \( d = \lfloor pn \rfloor \), and let \( X_n = (\xi_{ij})_{1 \leq i,j \leq n} \) be a matrix of iid copies of \( \xi \), independent of \( A_n \). Denoting \( Y_n = A_n \circ X_n \), we have that the rescaled ESDs \( \mu_n \) converge weakly in probability to \( \mu_{\text{circ}} \), the uniform measure on \( B_{\mathbb{C}}(0,1) \).

Moreover, if we assume \( \xi \) is a standard real Gaussian variable then the rescaled ESDs converge weakly almost surely.
Figure 1. Empirical eigenvalue distributions for simulated 8000 × 8000 rescaled rrd matrices $\frac{1}{\sqrt{d}}A$ for $d = 3$ (top), 10 (middle), and 100 (bottom). Left: scatterplots of eigenvalues, with the unit circle plotted in red for reference. Right: histograms for eigenvalue moduli, with each bin count normalized by $2\pi$ times the distance to the origin. The curves predicted by (1.4) are plotted in red. While the eigenvalue distribution is noticeably more dense near the edge of the support for $d = 3$, for $d = 100$ it is indistinguishable from the uniform distribution on the disk.
Remark 1.6 (Moment assumption on $\xi$). The assumption $\mathbb{E} |\xi|^{4+\eta} < \infty$ is used to ensure that $A_n \circ X_n$ has operator norm of size $O(\sqrt{n})$ with high probability, which is needed in our arguments to obtain control on small singular values. However, we believe it is likely this hypothesis can be eliminated.

Let us denote $\mu_n := \mu \frac{1}{\sqrt{np}} Y_n$. Our general approach to the proof of Theorem 1.5 follows previous works in using Girko’s Hermitization strategy, which reduces the problem of establishing convergence of linear statistics $\int_C f d\mu_n$ to one of establishing convergence of the logarithmic potentials $\int_{\mathbb{R}^+} \log(s) d\nu_{n,z}$ for a.e. $z \in \mathbb{C}$, where

$$\nu_{n,z} := \mu \left( \frac{1}{\sqrt{np}} Y_n - z \right)^* \left( \frac{1}{\sqrt{np}} Y_n - z \right)$$

is the empirical singular value distribution of the scalar-shifted matrix $\frac{1}{\sqrt{np}} Y_n - z$. This in turn essentially reduces to establishing the weak convergence of $\nu_{n,z}$, and proving lower-tail estimates for the small singular values of $\frac{1}{\sqrt{np}} Y_n - z$ in order to argue the measures $\nu_{n,z}$ do not put much mass near the singularity of log at zero.

A key difficulty for both of these problems is the lack of independence among the entries of $A_n$, and to deal with this we combine several results from the literature, as well as a new Wegner-type estimate. For the weak convergence of singular value distributions $\nu_{n,z}$ we can use a conditioning argument of Tran, Vu and Wang [TVW13] to compare with a matrix $\tilde{Y}_n = B_n \circ X_n$, where $B_n$ is an iid Bernoulli matrix of the same expected density as $A_n$. Here we apply an asymptotic for the number of 0–1 matrices with constrained row and column sums due to Canfield and McKay [CM05]. For the bounds on small singular values we leverage the independence of the entries of $X_n$ together with robust connectivity properties of random regular digraphs. Specifically, we use a graph discrepancy result of the author from [Coo16] to prove the rrd matrix $A_n$ satisfies a certain “broad connectivity” condition with high probability. This gives us access to recent results on the invertibility properties of random matrices with independent but non-identically distributed entries having broadly-connected support, due to Rudelson–Zeitouni [RZ16] (for the Gaussian case) and the author [Coob] (for the general case). In the process we prove Theorem 4.5, which gives a Wegner-type bound at essentially optimal scale for the small singular values of random matrices with broadly-connected support.

It would be interesting to extend Theorem 1.5 to allow $p = p(n) \to 0$, i.e. to study random matrices supported on sparse random regular digraphs (ideally allowing $p \sim n^{\varepsilon-1}$ for any $\varepsilon > 0$ as in [Woo12]). However, substantial technical issues arise when studying the small singular values for such sparse matrices; in particular, the results we use from [RZ16, Coob] no longer apply.\footnote{We believe that by explicitly quantifying the dependence on sparsity in the steps of [RZ16, Coob] one could obtain results applicable for $p \sim n^{-c}$ for a small absolute constant $c$, but we do not pursue this.} We defer investigation of the sparse case to the forthcoming work [Cooa].
1.3. Outline. The rest of the paper is organized as follows. In Section 2 we give a high level proof of Theorem 1.5 following the Hermitization approach, reducing our task to establishing the weak convergence of empirical singular value distributions (1.5), and lower bounds on small singular values of $\frac{1}{\sqrt{n}} Y_n - z$. The weak convergence of $\nu_{n,z}$ is proved in Section 3. In Section 4, we present two results – Proposition 4.4 and Theorem 4.5 – on the small singular values of random matrices with broadly-connected support, and prove that random regular digraphs satisfy this condition with high probability. In Section 5 we prove Theorem 4.5.

1.4. Notation. We use standard asymptotic notation, always with respect to the limit $n \to \infty$. $f = O(g)$, $f \ll g$ and $g \gg f$ are all synonymous to the statement that $|f| \leq Cg$ for some absolute constant $C > 0$. $f = o(g)$ means $f/g \to 0$ as $n \to \infty$. Dependence of implied constants on parameters is indicated with subscripts. $C, C', c, c_0$, etc. denote constants which may change from line to line.

For $J \subset [n]$, we denote by $\mathcal{C}^J \subset \mathbb{C}^n$ (resp. $S^J \subset S^{n-1}$) the set of vectors (resp. unit vectors) in $\mathbb{C}^n$ supported on $J$. Given a vector $v \in \mathbb{C}^n$, we denote by $v_J \in \mathbb{C}^n$ the projection of $v$ to the coordinate subspace $\mathbb{C}^J$. $\binom{m}{x}$ denotes the family of subsets of $[m]$ of size $|x|$. We write $E_{\rho}$ for the closed Euclidean $\rho$-neighborhood of a set $E \subset \mathbb{C}^n$.

$\mathcal{M}_{n,m}(\mathbb{C})$ denotes the set of $n \times m$ matrices with complex entries, and for square matrices we simply write $\mathcal{M}_{n}(\mathbb{C}) = \mathcal{M}_{n,n}(\mathbb{C})$. For a matrix $A = (a_{ij}) \in \mathcal{M}_{n,m}(\mathbb{C})$ we will sometimes use the notation $A(i,j) = a_{ij}$. For $I \subset [n], J \subset [m]$ with elements $i_1 < \cdots < i_{|I|}$ and $j_1 < \cdots < j_{|J|}$, respectively, $A_{I,J}$ denotes the $|I| \times |J|$ submatrix $(a_{i_k,j_\ell})_{1 \leq k \leq |I|, 1 \leq \ell \leq |J|}$. We often abbreviate $A_{J} := A_{J,J}$. $\| \cdot \|$ denotes the $\ell_2^m \to \ell_2^n$ operator norm when applied to $n \times m$ matrices, and the Euclidean norm when applied to vectors. Other norms are indicated with subscripts.

For a measure $\mu$ on a space $\mathcal{X}$ and a function $f : \mathcal{X} \to \mathbb{C}$, we will sometimes abbreviate $\mu(f) := \int_{\mathcal{X}} f d\mu$.

When there can be no confusion we will suppress the subscript $n$ from the matrices $X_n, A_n, Y_n$ etc.

2. Reduction to estimates on singular values

We follow Girko’s Hermitization method, introduced in [Gir84], and which we briefly review below; see [TV10] or the survey [BC12] for more background. For a probability measure $\mu$ on $\mathbb{C}$ we define the log-potential

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

(2.1)

Let us abbreviate $\mu_n := \mu \frac{1}{\sqrt{n}} Y_n$. The following is taken from [TV10, Theorem 1.20]:

**Lemma 2.1** (Hermitization). Let $\mu$ be a probability measure on $\mathbb{C}$ satisfying $\int_{\mathbb{C}} |z|^2 d\mu(z) < \infty$. The following are equivalent:

(i) $\mu_n$ converges weakly in probability to $\mu$.  

(ii) For almost every \( z \in \mathbb{C} \), \( U_{\mu_n}(z) \) converges in probability to \( U_{\mu}(z) \). Moreover, the same equivalence holds with “in probability” replaced by “almost surely”.

Denoting the ordered singular values of a matrix \( M \in \mathcal{M}_n(\mathbb{C}) \) by \( s_1(M) \geq \cdots \geq s_n(M) \), from the well-known identity

\[
\prod_{i=1}^{n} |\lambda_i(M)| = |\det(M)| = \prod_{i=1}^{n} s_i(M)
\]

we have

\[
U_{\mu_n}(z) = \frac{1}{n} \log |\det \left( \frac{1}{\sqrt{n)pY - zI} \right) | = \frac{1}{n} \sum_{i=1}^{n} \log s_i \left( \frac{1}{\sqrt{n)pY - zI} \right) = \int_{\mathbb{R}^+} \log(s) d\nu_{n,z}(s)
\]

where we have defined

\[
\nu_{n,z} := \frac{1}{n} \sum_{i=1}^{n} \delta_{s_i \left( \frac{1}{\sqrt{n)pY - zI} \right) }.
\]

(2.2)

The log-potential for the uniform measure on the unit disk is

\[
U(z) := \frac{1}{\pi} \int_{B_\mathbb{C}(0,1)} \log |w - z| dw = \begin{cases} \log |z| & \text{if } |z| > 1 \\ -\frac{1}{2}(1 - |z|^2) & \text{otherwise} \end{cases}
\]

(2.3)

From the above lines and Lemma 2.1, to establish Theorem 1.5 it suffices to show that for almost every \( z \in \mathbb{C} \), \( \int_{\mathbb{R}^+} \log(s) d\nu_{n,z}(s) \) converges in probability to \( U(z) \).

As a first step we will establish the following:

**Proposition 2.2** (Weak convergence of \( \nu_{n,z} \)). For all \( z \in \mathbb{C} \), \( \nu_{n,z} \) converges vaguely almost surely to a deterministic probability measure \( \nu_z \) on \( \mathbb{R}^+ \). Furthermore, for all \( z \in \mathbb{C} \) the measures \( \nu_z \) satisfy

\[
\int_{\mathbb{R}^+} \log(s) d\nu_z(s) = U(z)
\]

(2.4)

where \( U(z) \) was defined in (2.3).

In order to deduce from this the convergence of log-potentials \( U_{\mu_n}(z) \), we will apply the above proposition to a truncation \( f \in C_c(\mathbb{R}^+) \) of the function \( s \mapsto \log(s) \). To show that the truncated integral \( \int_{\mathbb{R}^+} f d\nu_{n,z} \) is a good approximation of \( U_{\mu_n}(z) \), we must prove that the measures \( \nu_{n,z} \) uniformly integrate the singularities of \( s \mapsto \log(s) \).

The singularity at \( +\infty \) requires control on a matrix norm of \( \frac{1}{\sqrt{n)pY - zI} \). While the standard approach is to use the Hilbert–Schmidt norm, we will instead apply the following stronger control on the operator norm as we will need it later.
Lemma 2.3 (The largest singular value). Except with probability $O(n^{-\eta/8})$ we have $s_1\left(\frac{1}{\sqrt{np}} Y - zI\right) = O_{z,p}(1)$. If $\xi$ is a standard Gaussian then the probability bound improves to $O(e^{-cn})$ for some constant $c > 0$.

Proof. This is a consequence of a general bound on the expected operator norm of random matrices due to Latała [Lat05] and a standard truncation argument. See the proof of [Coob, Lemma 5.5] for the details. The bound for the Gaussian case is well-known; see for instance [RZ16, Lemma 3.1]. □

The singularity of the logarithm at zero requires lower bounds on the small singular values of $\frac{1}{\sqrt{np}} Y - zI$. It turns out that for $z$ inside the support of the limiting ESD (i.e. $|z| \leq 1 - \varepsilon$), the limiting singular value distributions $\nu_z$ from Proposition 2.2 have a positive density in a neighborhood of zero. Based on this fact (which is not needed for the proof), we might expect that the smallest singular value $s_n\left(\frac{1}{\sqrt{np}} Y - zI\right)$ to live at scale $\sim 1/n$, and in general that the $k$-th smallest singular value would be at scale $\sim k/n$. The following two propositions provide quantitative lower bounds at these scales.

Proposition 2.4 (The smallest singular value). For any $t > 0$,

$$\mathbb{P}\left(s_n(Y - z\sqrt{np}I) \leq tn^{-1/2}\right) \ll_{z, p} t + n^{-1/2}. \tag{2.5}$$

Moreover, if $\xi$ is a standard Gaussian variable then we have that for any $K > 0$,

$$\mathbb{P}\left(s_n(Y - z\sqrt{np}I) \leq n^{-K-1/2}\right) \ll_{z, p, K} n^{-K}. \tag{2.6}$$

Proposition 2.5 (Local Wegner estimate). There are constants $a_0, a_1, a_2 > 0$ depending only on $p, z$ such that for all $\alpha \in (0, 1)$, except with probability $O(\exp(-a_0 n^{\alpha}))$, for all $i \in [n^\alpha, a_1 n]$ we have

$$s_{n-i}\left(\frac{1}{\sqrt{np}} Y - zI\right) \geq a_2 \frac{i}{n}. \tag{2.7}$$

We defer the proofs to Sections 4 and 5, and conclude the proof of Theorem 1.5 on Propositions 2.2, 2.4, 2.5 and Lemma 2.3. We will first argue the general case; at the end we will discuss the necessary modifications for the Gaussian case.

Fix $z \in \mathbb{C}$. Here we allow implied constants to depend on $\eta, p$ and $z$. Our aim is to show that for any $\varepsilon > 0$,

$$\mathbb{P}\left(|\nu_{n,z}(\log) - \nu_z(\log)| > \varepsilon\right) = o(1) \tag{2.8}$$

with $\nu_z$ as in Proposition 2.2. From Lemma 2.3 we have

$$\mathbb{P}(\nu_{n,z}([C_0, \infty)) = 0) = 1 - o(1) \tag{2.9}$$

for some $C_0 > 0$ sufficiently large depending only on $z$. In particular, with Proposition 2.2 this implies that $\nu_z$ is supported on $[0, C_0)$. For $\delta > 0$ small,
let $f_\delta \in C_c(\mathbb{R}_+)$ satisfy
\[ f_\delta(s) = \begin{cases} 
0 & s \in [0, \delta/2] \cup [2C_0, \infty) \\
\log(s) & s \in [\delta, C_0] 
\end{cases} \]
and take $f_\delta$ to be linearly increasing and decreasing on the intervals $[\delta/2, \delta]$ and $[C_0, 2C_0]$, respectively. From Proposition 2.2 we have that for any fixed $\delta > 0$, $\nu_{n,z}(f_\delta) \rightarrow \nu_z(f_\delta)$ almost surely, so from (2.9) it only remains to show that with high probability,
\[ \int_0^\delta |\log(s)| d\nu_{n,z}(s) = O(\delta^{0.9}) \quad (2.10) \]
(say) for all $\delta > 0$ sufficiently small.

For ease of notation we write $s_i$ for $s_i(\frac{1}{\sqrt{np}}Y - zI)$. Fix $\alpha \in (0,1)$ arbitrarily. We write left hand side of (2.10) as
\[ \frac{1}{n} \sum_{i=0}^{n-1} |\log(s_{n-i})| \mathbf{1}(s_{n-i} \leq \delta) = \frac{1}{n} \sum_{i=0}^{[n\alpha]} |\log(s_{n-i})| + \frac{1}{n} \sum_{i>n\alpha} |\log(s_{n-i})| \mathbf{1}(s_{n-i} \leq \delta). \]
\[ (2.11) \]
By (2.5) in Proposition 2.4, except with probability $O(n^{-1/2})$, the first term on the right hand side is bounded by $O(n^{1-\alpha} \log n) = o(1)$. By Proposition 2.5, taking $\delta < a_1 a_2$ we have except with probability $O(\exp(-a_0 n^\alpha))$, the second term is bounded by
\[ \frac{1}{n} \sum_{1 \leq i < \delta n/a_2} \left| \log \left( \frac{a_2 i}{n} \right) \right| \ll \frac{1}{n} \sum_{1 \leq i < \delta n/a_2} \left( \frac{a_2 i}{n} \right)^{-0.1} \ll \int_0^{2\delta/a_2} s^{-0.1} ds = O(\delta^{0.9}) \]
which completes the proof.

Now for the Gaussian case, our aim is to show $\nu_{n,z}((\log) \rightarrow \nu_z((\log)$ almost surely. By the Borel–Cantelli lemma it suffices to establish (2.8) with a bound that is summable in $n$ for any fixed $\varepsilon > 0$. Fixing $\varepsilon > 0$, we only need to verify that the $o(1)$ term in (2.9) is summable in $n$, and that (2.10) holds with probability $1 - O(n^{-2})$, say. For the former, by Lemma 2.3 we can replace the bound $o(1)$ in (2.9) with $O(e^{-cn})$. For the latter we only need to replace the application of (2.5) with (2.6), taking $K = 2$. This completes the proof for the Gaussian case of Theorem 1.5.

3. Weak convergence of singular value distributions

In this section we prove Proposition 2.2. Specifically, we show that for any $f \in C_c(\mathbb{R}_+) \text{ and any } \varepsilon > 0$,
\[ \mathbb{P}(|\nu_{n,z}(f) - \nu_z(f)| > \varepsilon) \ll_p \exp \left(-c_0 pn^2 \right) \quad (3.1) \]
for some $c_0 = c_0(\varepsilon, f) > 0$, where here and in the remainder of the section we allow implied constants to depend on $f$ and $\varepsilon$. The desired result will then follow from the Borel–Cantelli lemma.
Following an approach of Tran, Vu and Wang from [TVW13], we compare the signed rrd matrix $Y$ with an iid matrix taking the form

$$\tilde{Y} = (b_{ij}\xi_{ij})_{1 \leq i,j \leq n} = B \circ X$$

where the variables $b_{ij}$ are Bernoulli($p$) indicator variables (recall that $d = \lfloor pn \rfloor$), jointly independent of each other and of the matrix $X$. Note that the entries of $\frac{1}{\sqrt{p}} \tilde{Y}$ are iid and centered with unit variance. Writing $\tilde{\nu}_{n,z} := \frac{1}{n} \sum_{i=1}^{n} \delta_{s_{i}}(\frac{1}{\sqrt{pn}} \tilde{Y} - zI)$, it follows from a result of Dozier and Silverstein [DS07] that for all $z \in \mathbb{C}$ the measures $\tilde{\nu}_{n,z}$ converge weakly almost surely to a deterministic probability measure $\nu_{z}$; see [PZ10, Lemma 3] for the verification that (2.4) holds with $\mu = \mu_{circ}$. In particular, we have that for any $f \in C_{b}((R_{+})$,

$$|\nu_{z}(f) - \mathbb{E} \tilde{\nu}_{n,z}(f)| = o(1) \quad (3.2)$$

so it suffices to show

$$P(|\nu_{n,z}(f) - \mathbb{E} \tilde{\nu}_{n,z}(f)| > \varepsilon) \ll \exp\left(-c_{0}pn^{2}\right). \quad (3.3)$$

Define the event

$$E_{n,d} := \{ B \in M_{n}(d) \} \quad (3.4)$$

that the Bernoulli matrix $B$ is the adjacency matrix of a $d$-regular digraph. Note that $B|E_{n,d} = A$. Hence, to bound an event $\{ A \in M_{0} \}$ for some set $M_{0} \subset M_{n}(d)$, we can bound the corresponding event $\{ B \in M_{0} \}$ and apply the bound

$$P(A \in M_{0}) = \frac{P(B \in M_{0} \cap M_{n}(d))}{P(E_{n,d})} \leq \frac{P(B \in M_{0})}{P(E_{n,d})} \quad (3.5)$$

together with a lower bound on $P(E_{n,d})$. For this we have the following result of Canfield and McKay from [CM05].

**Lemma 3.1.** Assume $\min(d, n - d) \gg n/\log n$. Then

$$P(E_{n,d}) = (1 + o(1))\sqrt{2\pi d(n - d)} \exp\left(-n \log \left(\frac{2\pi d(n - d)}{n}\right)\right). \quad (3.6)$$

In particular, if $d = \lfloor pn \rfloor$ for some fixed $p \in (0, 1)$, then

$$P(E_{n,d}) \geq \exp(-O_{p}(n \log n)). \quad (3.7)$$

By the above lemma, (3.5) and (3.3), it suffices to show that for any $f \in C_{c}(\mathbb{R}_{+})$ and any $\varepsilon > 0$,

$$P(|\tilde{\nu}_{n,z}(f) - \mathbb{E} \tilde{\nu}_{n,z}(f)| > \varepsilon) \ll \exp(-c_{0}pn^{2}) \quad (3.8)$$

for some $c_{0}(\varepsilon, f) > 0$.

We have hence reduced our task to proving a concentration bound for linear statistics of the singular value distribution of the perturbed iid matrix $\frac{1}{\sqrt{np}} \tilde{Y} - zI$. For this task we have the following lemma, which follows from the work of Guionnet and Zeitouni in [GZ00].
Finally, we note that \( H = (h_{ij})_{1 \leq i, j \leq n} \) be a Hermitian random matrix, and assume the variables on and above the diagonal are jointly independent and that \( |h_{ij}| \leq K/\sqrt{n} \) uniformly in \( i, j \) for some \( K \in (0, \infty) \).

Let \( f : \mathbb{R} \to \mathbb{R} \) be an \( L \)-Lipschitz function supported on a compact interval \( I \subset \mathbb{R} \), and let \( H_0 \) be an arbitrary deterministic \( n \times n \) Hermitian matrix. Then for any \( \delta > 0 \),
\[
\mathbb{P} \left( |\mu_{H+H_0}(f) - \mathbb{E}\mu_{H+H_0}(f)| \geq \delta \right) \leq \frac{C|I|}{\delta} \exp \left( -\frac{cn^2\delta^4}{K^2L^2|I|^2} \right)
\]  
for some absolute constants \( C, c > 0 \).

**Proof.** For the case that \( H_0 = 0 \), this follows directly from [GZ00, Theorem 1.3(a)]. For the general case, the only part of the argument that needs modification is in the proof of their Theorem 1.1(a), where we need to show that for \( f : \mathbb{R} \to \mathbb{R} \) convex and \( L \)-Lipschitz function on the space of Hermitian matrices. However, this follows directly from their Lemma 1.2 and the fact that the convexity and Lipschitz properties are invariant under translations \( H \mapsto H + H_0 \). The rest of the proofs of [GZ00, Theorem 1.1(a)] and [GZ00, Theorem 1.3(a)] apply with no modification. \( \square \)

Fix \( \varepsilon > 0 \) and \( f \in C_c(\mathbb{R}) \). To apply the above concentration estimate to the measures \( \tilde{\nu}_{n,z} \), we recall the linearization approach to the study of singular value distributions of random matrices. From \( \tilde{Y} \) we form a \( 2n \times 2n \) Hermitian matrix
\[
H(z) = \frac{1}{\sqrt{np}} \begin{pmatrix} 0 & \tilde{Y} - z\sqrt{n}pI \\ \tilde{Y} - z\sqrt{n}pI^* & 0 \end{pmatrix}.
\]  
(3.10)

It is routine to verify that the \( 2n \) eigenvalues of \( H(z) \), counted with multiplicity, are \( \{\pm s_i(\frac{1}{\sqrt{np}}\tilde{Y} - zI)\}_{i=1}^n \). In terms of empirical spectral distributions, \( \mu_{H(z)} \) is the symmetrization across the origin of the measure \( \tilde{\nu}_{n,z} \) on \( \mathbb{R}_+ \). From (3.8), it now suffices to show
\[
\mathbb{P}(|\mu_{H(z)}(f) - \mathbb{E}\mu_{H(z)}(f)| > \varepsilon) \ll \exp(-c_0pn^2).
\]  
(3.11)

As \( f \) is uniformly continuous, there exists \( \delta = \delta(\varepsilon, f) > 0 \) such that \( |f(s) - f(t)| \leq \varepsilon/10 \) whenever \( |s - t| \leq \delta \). By linear interpolation on a \( \delta \)-mesh of the support of \( f \) we may find a function \( f_\varepsilon \in C_c(\mathbb{R}) \) with Lipschitz constant \( O(\varepsilon/\delta) = O(\varepsilon, f)(1) \), and such that \( \|f - f_\varepsilon\|_\infty \leq \varepsilon/10 \). It now suffices to show
\[
\mathbb{P}(|\mu_{H(z)}(f_\varepsilon) - \mathbb{E}\mu_{H(z)}(f_\varepsilon)| > \varepsilon/2) \ll \exp(-c_0pn^2).
\]  
(3.12)

Finally, we note that \( H(z) \) has the form \( H + H_0 \) as in Lemma 3.2, with \( K = O(1/\sqrt{p}) \), \( H = H(0) \) and
\[
H_0 = \begin{pmatrix} 0 & -zI \\ -z^*I & 0 \end{pmatrix}.
\]

Applying that lemma yields (3.12), and hence (3.1), which completes the proof.
4. Bounds on small singular values

In this section we deduce Propositions 2.4 and 2.5 from two general results on the small singular values of random matrices with independent entries: results of Rudelson–Zeitouni [RZ16] and the author [Coob] on the smallest singular value, and a new Wegner-type bound on “moderately small” singular values.

We begin by recalling some definitions and notation from [Coob]. The following allows us to quantify the dependence of our bounds on the distribution of the matrix entries.

Definition 4.1 (Spread random variable). Let ξ be a complex random variable and let κ ≥ 1. We say that ξ is κ-spread if

\[ \text{Var} \left[ \xi \mathbb{1}(|\xi - \mathbb{E}\xi| \leq \kappa) \right] \geq \frac{1}{\kappa}. \] (4.1)

Remark 4.2. It follows from the monotone convergence theorem that any random variable ξ with non-zero second moment is κ-spread for some κ < ∞. Furthermore, if ξ is centered with unit variance and finite pth moment µp for some p > 2, then it is routine to verify that ξ is κ-spread with κ = 3(3µp)p/(p−2), say.

Our results on small singular values assume the matrix Σ = (σij) of standard deviations satisfies a certain expansion-type condition originally formulated in [RZ16]. To state it we will need some graph-theoretic notation. To a non-negative n × m matrix Σ = (σij) we associate a bipartite graph ΓΣ = ([n], [m], EΣ), with (i, j) ∈ EΣ if and only if σij > 0. For a row index i ∈ [n] we denote by

\[ N_{\Sigma}(i) = \{ j \in [m] : \sigma_{ij} > 0 \} \] (4.2)

its neighborhood in ΓΣ. Thus, the neighborhood of a column index j ∈ [m] is denoted N_{\Sigma}^T(j). For I ⊂ [n] and δ ∈ (0, 1), define the set of δ-broadly connected neighbors of I as

\[ N_{\Sigma}^\delta(I) = \{ j \in [m] : |N_{\Sigma}^T(j) \cap I| \geq \delta |I| \}. \] (4.3)

Given sets of row and column indices I ⊂ [n], J ⊂ [m], we define the associated edge count

\[ e_{\Sigma}(I, J) := |\{(i, j) \in [n] \times [m] : \sigma_{ij} > 0\}|. \] (4.4)

We will generally work with the graph that only puts an edge (i, j) when σij exceeds some fixed cutoff parameter σ0 > 0. Thus, we denote by

\[ \Sigma(\sigma_0) = (\sigma_{ij}1_{\sigma_{ij} \geq \sigma_0}) \] (4.5)

the matrix which thresholds out entries smaller than σ0.

Definition 4.3 (Random matrix with broadly connected profile). Let Σ = (σij) and Z = (zij) be deterministic n × m matrices with σij ∈ [0, 1] and zij ∈ C for all i, j. We assume that Σ satisfies the following expansion property: for some σ0, δ, ν ∈ (0, 1) we have

1. |N_{\Sigma(\sigma_0)}(i)| ≥ δm for all i ∈ [n];
(2) \(|N_{\Sigma(\sigma_0)^T}(j)| \geq \delta n\) for all \(j \in [m]\);
(3) \(|N_{\Sigma(\sigma_0)^T}(J)| \geq \min(n, (1 + \nu)|J|)\) for all \(J \subset [m]\).

We say that a matrix \(\Sigma\) satisfying conditions (1)–(3) is \((\sigma_0, \delta, \nu)\)-broadly connected. Let \(X = (\xi_{ij})\) be an \(n \times m\) matrix with independent entries, all identically distributed to a complex random variable \(\xi\) with mean zero and variance one. Put

\[ M = \Sigma \circ X + Z = (\sigma_{ij} \xi_{ij} + z_{ij})_{i,j=1}^n \]  

(4.6)

where \(\circ\) denotes the matrix Hadamard product. We refer to \(\Sigma\) as the (standard deviation) profile for \(M\). Without loss of generality, we assume throughout that \(\xi\) is \(\kappa\)-spread for some fixed \(\kappa \geq 1\).

The following proposition collects known bounds on the smallest singular value for random matrices with a broadly connected profile.

**Proposition 4.4** (Smallest singular value: broadly connected profile). Let \(M = \Sigma \circ X + Z\) be an \(n \times n\) matrix as in Definition 4.3. Let \(K_0 \geq 1\). For any \(t \geq 0\),

\[ P \left( s_n(M) \leq \frac{t}{\sqrt{n}}, \|M\| \leq K_0 \sqrt{n} \right) \ll t + \frac{1}{\sqrt{n}} \]  

(4.7)

where the implied constant depends only on \(K_0, \delta, \nu, \sigma_0\) and \(\kappa\). If we further assume the entries of \(X\) are standard real Gaussians, then

\[ P \left( s_n(M) \leq \frac{t}{\sqrt{n}}, \|M\| \leq K_0 \sqrt{n} \right) \ll t + e^{-cn} \]  

(4.8)

where the implied constant and \(c\) depend only on \(K_0, \delta, \nu, \sigma_0\).

**Proof.** (4.7) follows from [Coob, Theorem 1.13], while (4.8) follows from [RZ16, Theorem 2.3] (together with (2.3) and the triangle inequality for the operator norm). \(\square\)

We will prove the following theorem in Section 5.

**Theorem 4.5** (Wegner-type estimate: broadly connected profile). Let \(M = \Sigma \circ X + Z\) be an \(n \times n\) matrix as in Definition 4.3. There are constants \(a_0, a_1, a_2 > 0\) depending (polynomially) on \(\kappa, \sigma_0, \delta, \nu, K_0\) such that

\[ P \left( \exists k \in [n^{2\varepsilon}, a_1 n] : s_{n-k}(M) < a_2 \frac{k}{\sqrt{n}}, \|M\| \leq K_0 \sqrt{n} \right) \ll \exp \left( -a_0 n^{\varepsilon} \right) \]  

(4.9)

where the implied constant depends on \(\kappa, \sigma_0, \delta, \nu, K_0\) and \(\varepsilon\).

To apply these results to the case that \(\Sigma\) is an rrd matrix we show the following:

**Proposition 4.6** (Random regular digraphs are broadly connected). Let \(p \in (0, 1)\) and let \(A\) be a uniform random element of \(\mathcal{M}_{n}(\lfloor np \rfloor)\). Then \(A\) is \((1, p/2, p/8)\)-broadly connected with probability \(1 - O_p(n^{-K})\) for any \(K > 0\).
To prove this we need the following result on edge counts in random regular digraphs, which is an immediate consequence of Corollary 1.9 in [Coo16]. (In the present setting the probability bound below can easily be improved, but it is sufficient for our purposes.)

**Lemma 4.7** (No sparse patches). For \( \varepsilon_0 > 0 \) let \( \mathcal{G}(\varepsilon_0) \) be the event that for all \( I, J \subset [n] \) with \( |I|, |J| \geq \varepsilon_0 n \), \( e_A(I, J) \geq \frac{1}{2} p |I||J| \). For all \( \varepsilon_0 > 0 \) we have

\[
\mathbb{P}(\mathcal{G}(\varepsilon_0)) = 1 - O_{p, \varepsilon_0}(e^{-c \log^2 n}).
\]

**Proof of Proposition 4.6.** We may assume that \( n \) is sufficiently large depending on \( p \). A clearly satisfies conditions (1) and (2) in Definition 4.3 with \( \delta = p/2 \) since every row and column has support \( \lfloor np \rfloor \). Now we verify condition (3). Let \( J \subset [n] \), and abbreviate \( I_0 = I_0(J) = N_{A^T}^{(p/2)}(J) \). Since each column of \( A \) has support \( \lfloor np \rfloor \), we have

\[
|I_0||J| + (p/2)n|J| < |I_0||J| + (p/2)n|J|
\]

which rearranges to

\[
|I_0| > \frac{1}{2} pn - O(1) \gg pn.
\]

Thus, condition (3) is satisfied (deterministically) for any \( J \) of size at most \( cpn \) for a sufficiently small absolute constant \( c > 0 \). On the other hand, if \( |J| > n(1 - p/4) \) then we must have \( |N(i) \cap J| \geq (p/2)n \) for all \( i \in [n] \), so condition (3) also holds deterministically for any \( J \) of size at least \( n(1 - p/4) \). So we may assume \( cpn \leq |J| \leq (1 - p/4)n \). In particular,

\[
(1 + 2\nu)|J| = (1 + p/4)|J| \leq n \tag{4.10}
\]

Suppose that \( |I_0(J)| < (1 + \nu)|J| \). Then

\[
|I_0^c| > n - (1 + \nu)|J| = \nu|J| + n - (1 + 2\nu)|J| \geq \nu|J|
\]

and

\[
e_A(I_0^c, J) = \sum_{i \in I_0^c} |N(i) \cap J| < \frac{1}{2} p |I_0^c||J|.
\]

Thus, on the event that \( |I_0(J)| < (1 + \nu)|J| \) for some \( J \subset [n] \), there exists \( I \subset [n] \) such that \( |I| \gg p|J| \) and \( e_A(I, J) < \frac{1}{2} p |I||J| \). But this means the event \( \mathcal{G}(c'p^2) \) holds for some constant \( c' > 0 \) sufficiently small, and the result follows from Lemma 4.7. \( \square \)
Now we deduce Propositions 2.4 and 2.5 from Proposition 4.4 and Theorem 4.5. Fix $p \in (0, 1)$, $z \in \mathbb{C}$, let $A, X, Y$ be as in Proposition 2.4, and write $Z := -z\sqrt{mp}I$, $M = Y + Z$. Conditional on any realization of $A$, by Lemma 2.3 we may restrict to the event $\{\|M\| \leq K_0\sqrt{n}\}$ for some $K_0 = O_{p,z}(1)$. By Proposition 4.6 we may further condition on $A$ lying in the event that it is $(p/2, p/8)$-broadly connected. Propositions 2.4 and 2.5 now follow from Proposition 4.4 and Theorem 4.5 with $\Sigma = A$.

5. Control of moderately small singular values

In this section we prove Theorem 4.5. While the traditional approach to obtaining such estimates (going back to the work of Bai [Bai97]) is based on quantitative control on the rate of convergence of Stieltjes transforms, in [TV10] Tao and Vu introduced a simpler and more direct geometric argument. The first key element of their approach is the so-called “inverse second moment identity” ((5.5) below), which relates the size of the singular values of a rectangular matrix to the distances between rows and the span of the remaining rows.

The second key element is a high probability lower bound for the distance between a random row vector $R$ and a fixed subspace $W$ of moderately large codimension $k$ (in our case of size $k \sim n^{\varepsilon}$). This distance can be expressed in the form $\|P_{W^\perp}R\|$, where $P_{W^\perp}$ is the matrix for projection to the orthogonal complement of $W$. If $R$ has iid centered components with unit variance, then

$$E \text{dist}(R, W)^2 = E \|P_{W^\perp}R\|^2 = E R^T P_{W^\perp} R = \text{tr}(P_{W^\perp}) = k. \quad (5.1)$$

A high probability lower bound on $\text{dist}(R, W)$ is then deduced from concentration of measure – in this case Talagrand’s isoperimetric inequality (after a truncation).

The main new difficulty for proving Theorem 4.5 stems from the fact that the entries of $M$ are not identically distributed, in which case the computation (5.1) breaks down, and in general the expected distance between a row and a fixed subspace can be quite small. However, from a result in [Coob] (Proposition 5.3 below), it turns out that under the broad connectivity assumption the orthogonal complement of a large number of rows is in “generic position” in a certain sense. Specifically, any unit normal vector for a large number of rows is highly incompressible (see (5.7) below). This can be used to obtain a lower bound on $E \text{dist}(R, W)^2$ that only loses a constant factor from the identity (5.1). (Observe that in the iid case the only information we used about $W$ was its dimension.)

We turn to the details. The following reduces our task to obtaining lower bounds on the distance between a fixed row and the span of most of the other rows.

**Lemma 5.1.** Let $M \in \mathcal{M}_n(\mathbb{C})$ and $0 \leq k \leq n - 1$. Put $m = n - \lfloor k/2 \rfloor$. Denote the rows of $M$ by $R_1, \ldots, R_n$, and for $i \in [m]$ denote

$$R_{-i} := \text{span} \{R_j : j \in [m] \setminus \{i\}\}. \quad (5.2)$$
We have
\[ s_{n-k}(M) \geq \sqrt{\frac{k}{n}} \min_{i \in [m]} \text{dist}(R_i, R_{-i}). \]  
(5.3)

**Proof.** We follow the argument from [TV10]. Denote \( M' = M_{(m)} \times [n] \), the matrix obtained by removing the last \( \lceil k/2 \rceil \) rows from \( M \). By the Cauchy interlacing law,
\[ s_{n-k}(M) \geq s_{n-k}(M'). \]  
(5.4)

On the other hand, from the inverse second moment identity (cf. [TV10, Lemma A.4]) we have
\[ \sum_{i=1}^{n} s_i(M')^{-2} = \sum_{i=1}^{M} \text{dist}(R_i, R_{-i})^{-2} \]  
(5.5)

and so
\[ m \left( \min_{i \in [m]} \text{dist}(R_i, R_{-i}) \right)^{-2} \geq \sum_{i=1}^{n \lceil k/2 \rceil} \text{dist}(R_i, R_{-i})^{-2} \]
\[ \geq \sum_{i=n-k}^{M} s_i(M')^{-2} \]
\[ \geq \frac{k}{2} s_{n-k}(M')^{-2}. \]

(5.3) now follows from the above and (5.4) (noting that \( m \geq n/2 \)). \( \square \)

Our next task is to show how the distances \( \text{dist}(R_i, R_{-i}) \) can be controlled from below if we have certain structural information on the normal vectors of \( R_{-i} \). We recall the following definitions from [RZ16, Coob]. For \( m \geq 1 \) and \( \theta, \rho \in (0, 1) \) we define the set of **compressible vectors**
\[ \text{Comp}_m(\theta, \rho) := S^{m-1} \cap \bigcup_{J \in \binom{[m]}{\theta m}} (C_J)_{\rho} \]  
(5.6)

and the complementary set of **incompressible vectors**
\[ \text{Incomp}_m(\theta, \rho) := S^{m-1} \setminus \text{Comp}(\theta, \rho) \]  
(5.7)

(recall our notational conventions from Section 1.4). That is, \( \text{Comp}_m(\theta, \rho) \) is the set of unit vectors within (Euclidean) distance \( \rho \) of a vector supported on at most \( \theta m \) coordinates.

**Lemma 5.2** (Distance of a random vector to an incompressible subspace). Let \( \xi \) be a centered complex-valued random variable with unit variance, and let \( X = (\xi_1, \ldots, \xi_n) \) be a vector of iid copies of \( \xi \). Let \( \sigma = (\sigma_1, \ldots, \sigma_n) \in [0, 1]^n \), and put \( R = X \circ \sigma = (\sigma_j \xi_j)_{j=1}^{n} \). Suppose that for some \( \delta, \sigma_0 \in (0, 1) \) we have
\[ |L| := |\{ j \in [n] : \sigma_j \geq \sigma_0 \}| \geq \delta n. \]
Let $\varepsilon \in (0, 1)$ and let $V \subset \mathbb{C}^n$ be a subspace of dimension $n - k$. Suppose that for some $\rho > 0$ and any unit vector $u \in V^\perp$ we have $u \in \text{Incomp}_n((1 - \frac{\delta}{2}), \rho)$. Then for any fixed $v \in \mathbb{C}^n$ we have

$$\mathbb{P}\left(\text{dist}(R + v, V) \leq c\rho \sqrt{\delta k}\right) = O_\varepsilon(\exp(-c\sigma_0^2 \rho^2 \delta k/n^\varepsilon))$$

(5.8)

if

$$\frac{Cn^\varepsilon}{\sigma_0^2 \rho^2 \delta} \leq k \leq n - 1$$

(5.9)

for a sufficiently large constant $C > 0$.

**Proof.** Fix $\varepsilon \in (0, 1)$. We may assume $n$ is sufficiently large depending on $\varepsilon$. By replacing $V$ with $\text{span}(V, v)$ and $k$ with $k - 1$ we may take $v = 0$.

Towards an application of concentration of measure, we first perform a truncation. By Chebyshev’s inequality, for all $j \in \mathbb{P}(|\xi_j| \geq n^{\varepsilon/2}) \leq n^{-\varepsilon}$. It follows from Hoeffding’s inequality that

$$\mathbb{P}\left(\{|j \in [n] : |\xi_j| \leq n^{\varepsilon/2}\} \geq n - n^{1-\varepsilon/2}\right) \geq 1 - \exp(-cn^{1-\varepsilon}).$$

(5.10)

Put $m = [n - n^{1-\varepsilon/2}]$, and for $J \in \binom{m}{n}$ denote the event

$$\mathcal{E}_J := \{|\xi_j| \leq n^{\varepsilon/2} \forall j \in J\}.$$  

(5.11)

It suffices to obtain control of the lower tail of $\text{dist}(R, V)$ conditional on $\mathcal{E}_J$ that is uniform in the choice $J$. For $j \in [n]$ let

$$\lambda := \mathbb{E}(\xi_j | |\xi_j| \leq n^{\varepsilon/2}), \quad \xi_j' := \xi_j - \lambda$$

and write $R' = (\xi_1', \ldots, \xi_n')$. We have

$$\tau^2 := \mathbb{E}\left(|\xi_j'|^2 | |\xi_j| \leq n^{\varepsilon/2}\right) \geq 1/2.$$

Fix $J \in \binom{m}{n}$. Condition on a realization of $\{\xi_j\}_{j \notin J}$, and write $\mathbb{P}_J(\cdot)$ and $\mathbb{E}_J(\cdot)$ for probability and expectation conditional on $\{\xi_j\}_{j \notin J}$. Let $W = \text{span}(V, R_{[n] \setminus J}, \lambda \sigma_J)$; note that $W$ is deterministic under the conditioning on $\{\xi_j\}_{j \notin J}$. Then $\dim(W) \leq \dim(V) + 2$ and

$$\text{dist}(R, V) \geq \text{dist}(R', W).$$

(5.12)

Note that $\text{dist}(R', W) = \|P_{W^\perp} R\|$, where $P_{W^\perp}$ is the orthogonal projection to $W^\perp$. Letting $u_1, \ldots, u_{k-2}$ be an arbitrary set of orthonormal vectors in $W^\perp$, we have

$$\mathbb{E}_J \text{dist}(R', W)^2 \geq \sum_{i=1}^{k-2} \mathbb{E}|R' \cdot u_i|^2 = \tau^2 \sum_{i=1}^{k-2} \sum_{j=1}^n |u_j^i \sigma_j|^2 \geq \frac{1}{2} \sigma_0^2 \sum_{i=1}^{k-2} \|u_i\|_L^2.$$  

(5.13)

For each $1 \leq i \leq k - 2$, since $u_i \in \text{Incomp}_n((1 - \frac{\delta}{2}), \rho)$ there is a set $J_i \subset [n]$ with $|J_i| \geq (1 - \frac{\delta}{2})n$ and $|u_j^i| \geq \rho/\sqrt{n}$ for all $j \in J_i$. Indeed, if this were not the case then the projection of $u_i$ to its largest $(1 - \frac{\delta}{2})n$ coordinates would be a
It shows that under the broad connectivity hypothesis, the matrix
\[ n/4.3 \text{ with (Compressible vectors) Proposition 5.3} \]
is invertible on compressible vectors, even after the removal of some rows.
\[ \square \]
and the result follows from the lower bound in (5.9).

Then for any \( 0 < \theta \leq (1 - \frac{\delta}{4}) \min(\frac{n}{m}, 1) \),
\[ \mathbb{P}(\|M\| \leq K_0 \sqrt{n}, \exists u \in \text{Comp}_n(\theta, \rho) : \|Mu\| \leq \rho K_0 \sqrt{n}) \ll \exp \left(-c_0 \delta^2 n\right) \] (5.18)
where \( c_0 > 0 \) depends only on \( \kappa \) and the implied constant depends only on \( \kappa, \sigma_0, \delta, \nu \) and \( K_0 \).

We will need the following technical result, which is Proposition 3.1 from [Coob].
It shows that under the broad connectivity hypothesis, the matrix \( M \) is well-invertible on compressible vectors, even after the removal of some rows.

**Proposition 5.3 (Compressible vectors).** Let \( M = \Sigma \circ X + Z \) be as in Definition 4.3 with \( n/2 \leq m \leq 2n \). Let \( K_0 \geq 1 \). There exist \( \theta_0(\kappa, \sigma_0, \delta, K_0) > 0 \) and \( \rho(\kappa, \sigma_0, \delta, \nu, K_0) > 0 \) such that the following holds. Assume

1. \( |\mathcal{N}_{\Sigma(\sigma_0)^T}(j)| \geq \delta n \) for all \( j \in [m] \);

2. \( |\mathcal{N}_{\Sigma(\sigma_0)^T}(J)| \geq \min((1 + \nu)|J|, n) \) for all \( J \subset [m] \) with \( |J| \geq \theta_0 m \).

Then for any \( 0 < \theta \leq (1 - \frac{\delta}{4}) \min(\frac{n}{m}, 1) \),
\[ \mathbb{P}(\|M\| \leq K_0 \sqrt{n}, \exists u \in \text{Comp}_n(\theta, \rho) : \|Mu\| \leq \rho K_0 \sqrt{n}) \ll \exp \left(-c_0 \delta^2 n\right) \] (5.18)
Remark 5.4. In [Coob, Proposition 3.1] is stated under a more general distributional hypothesis for $\xi$; see [Coob, Lemma 2.5].

Proof of Theorem 4.5. For the duration of the proof we restrict to the event \( \{ \|M\| \leq K_0\sqrt{n} \} \). We may assume that \( n \) is sufficiently large depending on \( \kappa, \varepsilon, \sigma_0, \delta, \nu, \) and \( K_0 \). By the union bound it suffices to show that for \( a_2 \) sufficiently small depending on \( \kappa, \sigma_0, \delta, \nu, K_0 \),

\[
P( s_{n-k}(M) \leq a_2 k/\sqrt{n} ) = O(n \exp(-a_0 n^{\varepsilon})) \tag{5.19}
\]

for arbitrary fixed \( k \in [n^{2\varepsilon}, a_1 n] \). By Lemma 5.1 and another application of the union bound, after modifying \( a_2 \) by a constant factor it suffices to show

\[
P( \text{dist}(R_i, V_I) \leq a_2 \sqrt{k} ) = O(\exp(-a_0 n^{\varepsilon})) \tag{5.20}
\]

where \( V_I = \text{span}\{R_i, i \in I\} \) for an arbitrary fixed subset \( I \subset [n] \) with \( |I| = n - \lfloor k/2 \rfloor - 1 =: n' \) and arbitrary fixed \( i \in [n] \setminus I \).

Fix such \( I \subset [n] \) and \( i \in [n] \setminus I \). Let \( \theta_0 \) be as in Proposition 5.3 and let \( J \subset [n] \) with \( |J| \geq \theta_0 n \). Denoting \( \Sigma := \Sigma(\sigma_0)_{|J|\times|J|} \), by the assumption that \( \Sigma \) is \((\sigma_0, \delta, \nu)\)-broadly connected we have

\[
|\lambda_{\Sigma_{J}}^{(\delta)}(J)| \geq \min(n', (1+\nu)|J| - \lfloor k/2 \rfloor - 1) \geq \min(n', (1+\nu/2)|J|)
\]

taking \( a_1 \leq c\nu \theta_0 \) for a sufficiently small constant \( c > 0 \). Applying Proposition 5.3 to \( M_{I\times[n]} \) (with \( (n', n) \) in place of \( (n, m) \) and \((\delta/2, \nu/2) \) in place of \((\delta, \nu) \)) we have that except with probability \( O(\exp(-\frac{1}{4}c_0 \delta \sigma_0^2 n)) \),

\[
\|M_{I\times[n]}u\| \geq \rho K_0 \sqrt{n} \quad \forall u \in \text{Comp}_n \left( \left( 1 - \frac{\delta}{4} \right) \frac{n'}{n}, \rho \right) \tag{5.21}
\]

for some \( \rho = \rho(\kappa, \sigma_0, \delta, \nu, K_0) > 0 \). In particular, taking \( a_1 \) smaller if necessary, we have

\[
\left( 1 - \frac{\delta}{4} \right) \frac{n'}{n} \geq \left( 1 - \frac{\delta}{4} \right) \left( 1 - \frac{a_1}{2} \right) \geq 1 - \frac{\delta}{2}
\]

and so

\[
S^{n-1} \cap V_I^\perp = \ker(M_{I\times[n]}) \subset \text{Incomp}_n(1 - \delta/2, \rho). \tag{5.22}
\]

We may condition on the event that (5.21) holds. Conditional on \( M_{I\times[n]} \), we apply Lemma 5.2 with \( R + v = R_i \) and \( V = V_I \) to obtain (5.20) as desired. \( \square \)

References

[AC15] Radosław Adamczak and Djalil Chafaï. Circular law for random matrices with unconditional log-concave distribution. Commun. Contemp. Math., 17(4):1550020 (22 pages), 2015.

[ACW16] Radosław Adamczak, Djalil Chafaï, and Pawel Wolff. Circular law for random matrices with exchangeable entries. Random Structures Algorithms, 48(3):454–479, 2016.

[AGB+15] S. Allesina, J. Grilli, G. Barabási, S. Tang, J. Aljadeff, and A. Maritan. Predicting the stability of large structured food webs. Nature communications, 6, 2015.
[Ngu14] Hoi H. Nguyen. Random doubly stochastic matrices: the circular law. *Ann. Probab.*, 42(3):1161–1196, 2014.

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

[RA06] Kanaka Rajan and LF Abbott. Eigenvalue spectra of random matrices for neural networks. *Physical review letters*, 97(18):188104, 2006.

[RZ16] Mark Rudelson and Ofer Zeitouni. Singular values of Gaussian matrices and permanent estimators. *Random Structures Algorithms*, 48(1):183–212, 2016.

[SCS88] H. Sompolinsky, A. Crisanti, and H.-J. Sommers. Chaos in random neural networks. *Phys. Rev. Lett.*, 61(3):259–262, 1988.

[Tal96] Michel Talagrand. A new look at independence. *Ann. Probab.*, 24(1):1–34, 1996.

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

[TV10] Terence Tao and Van H. Vu. Random matrices: universality of ESDs and the circular law. *Ann. Probab.*, 38(5):2023–2065, 2010. With an appendix by Manjunath Krishnapur.

[TVW13] Linh V. Tran, Van H. Vu, and Ke Wang. Sparse random graphs: eigenvalues and eigenvectors. *Random Structures Algorithms*, 42(1):110–134, 2013.

[Woo12] Philip Matchett Wood. Universality and the circular law for sparse random matrices. *Ann. Appl. Probab.*, 22(3):1266–1300, 2012.

Department of Mathematics, Stanford University

E-mail address: nickcook@stanford.edu